Zero or One, Up or Down

Statistical Inference for Binary Data with Applications in Sensitivity
Analysis of Energetic Materials

Dennis Christensen

Supervisors:
Nils Lid Hjort and Erik Unneberg

Department of Mathematics

Abstract

This thesis reports novel advances in the study of binary phenomena, with sensitivity analysis of energetic materials as a key application. Its contributions are two-fold. Firstly, a new method is presented, called permutation counting, for estimating marginal likelihoods in Bayesian nonparametric models for binary data. By exploiting the symmetries of exchangeable data sequences, this method yields a new unbiased and strongly consistent estimator of the marginal likelihood. An explicit algorithm for computing the importance weights in polynomial time is provided, and numerous examples with real data are given. The algorithm is also efficiently implemented in an accompanying software package. The second half of the thesis focuses on aspects of frequentist parametric inference in the context of estimating sensitivities of energetic materials. New results pertaining to the asymptotic theory of sequential experimental designs are proved, thus justifying the construction of confidence intervals via large sample theory in this setting. These theoretical results are accompanied by extensive simulation studies, along with applications with real data. In particular, we demonstrate that remnants of the high explosive amatol are more sensitive to impact than previously believed in the literature. Based on our results, we strongly recommend updating NATO's official protocol for sensitivity measurements.

Sammendrag

I denne avhandlingen rapporteres det nye fremskritt innen studiet av binære fenomener, med spesielt fokus på å anvende dem til å bestemme følsomheten til energetiske materialer. Avhandlingen er todelt. Først presenteres en ny metode basert på telling av permutasjoner for å estimere marginalsannsynligheter innen Bayesianske ikke-parametriske modeller for binære data. Ved å utnytte symmetriene til utskiftbare datafølger gir denne metoden en ny forventningsrett og sterkt konsistent estimator for marginalsannsynligheten, ved hjelp av viktighetsutvalg (*importance sampling*). Det blir introdusert en eksplisitt algoritme for å beregne viktighetsvektene i polynom tid, og algoritmen blir anvendt på flere eksempler med ekte data. Det hører også til programvare hvor algoritmen er effektivt implementert. Den andre delen av avhandlingen setter søkelys på aspekter rundt frekventistisk parametrisk inferens innen estimering av følsomheten til energetiske materialer. Det presenteres nye resultater innen asymptotikken til sekvensiell forsøksplanlegging, hvilket, i denne sammenhengen, rettferdiggjør å konstruere konfidensintervaller basert på teori av store utvalg. Disse teoretiske resultatene kombineres med omfangsrike simuleringsstudier og anvendelser på eksperimentelle data. Vi demonstrerer blant annet at krigsetterlatenskaper av høyeksplosivet amatol er mer følsomme for slag enn det litteraturen har gitt uttrykk for. Basert på resultatene våre anbefaler vi på det sterkeste en oppdatering av NATOs offisielle prosedyrer for følsomhetstesting.

Preface

I still remember vividly my first lecture in real analysis in Oxford back in 2014, which professor Hilary Priestley commenced by dramatically declaring that the course we were about to take could be summarised by a single symbol: $\mathbb{R}$. If I were to try the same exercise for this thesis, I would say that it can be summarised by two symbols: 0 and 1. It was not necessarily my original intention to focus explicitly on binary data when I first began my doctoral studies at the University of Oslo in 2021, or even when I started working for the Norwegian Defence Research Establishment (FFI) in 2018. However, upon immersing myself in the study of sensitivity analysis of energetic materials, it gradually became evident to me that the statistical analysis of such data contains a surfeit of rich mathematics whilst simultaneously being highly useful for FFI.

I first wish to extend my sincerest gratitude to my two supervisors: Nils Lid Hjort and Erik Unneberg. Nils, whom I first met as my lecturer in a course on Bayesian nonparametrics in 2019, has been a tremendous source of inspiration throughout the course of the PhD. I am continually astounded by his extraordinary breadth and depth of knowledge, particularly in, but certainly not excluded to, the realm of statistics. I am thankful for the many weird and wonderful concepts he has introduced me to, and for the enthusiasm with which he has done so. It was a particular pleasure to sing with him on multiple occasions in the "Blindern Stunt- and Pop-up Choir". As for Erik, I am immensely grateful for all our motivating conversations, both of work-related and extracurricular nature. As a chemist with an expertise in the study of energetic materials, Erik has enabled me to place the statistical analysis contained in the present thesis in a wider context, without which this thesis would have been much weaker and less significant. With a passion for writing and language, Erik has also provided tremendously useful support in preparing manuscripts and presentations. I also had the privilege of experiencing his supervision skills on multiple arenas, such as when we left work early one hot summer day to play golf together.

Next, I want to thank all my co-authors: Per August Jarval Moen, Tomas Lunde Jensen, John Fredrik Moxnes, Eirik Høyheim, Emil Aas Stoltenberg, Geir Petter Novik and Sebastian Teigen Nygård. Without exception, working with them has been extremely an instructive and stimulating experience, and I feel very privileged to have been able to work with such a wide variety of scientists and researchers. Furthermore, I thank my directors at FFI, Hege Jødahl, Ingebjørg Kåsen and Ivar Sollien, for the flexibility they have granted me in working on my PhD. Hege, in particular, has been exceptionally generous in allowing me to travel to present my work and to start new collaborations with others. Thanks also to the

academic faculty at the University of Oslo for providing an intellectually inspiring environment in which to work.

Finally, a warm thanks to my closest family and friends, near and far, for always showing me love and support.

List of Figures

List of Figures

List of Tables

List of Tables

Contents

VII Increased impact sensitivity in ageing high explosives; analysis of Amatol extracted from explosive remnants of war **173**

VIII Unlevel playing field: Evidence of ethnic discrimination in the access to children's football from a field experiment in Norway **193**

Contents

Thesis

But let your communication be, Yea, yea; Nay, nay: for whatsoever is more than these cometh of evil.

<div align="right">Matt. 5:37</div>

Hun er vor højeste glæde i verden, og – (sænker stemmen) hun er også vor dybeste sorg, Gregers.

<div align="right">Hjalmar Ekdal, Henrik Ibsen's *Vildanden*</div>

Jeg har skrevet to taler, far. En er grøn, og en er gul. Og du kan selv vælge hvad for en, det skal være.

<div align="right">Christian, Thomas Vinterberg's *Festen*</div>

I get up, I get down.

<div align="right">Yes (Anderson and Howe), *Close to the Edge*</div>

Chapter 1

Introduction

The world is full of binary data. This will come as no surprise to the modern reader, as all digital information is stored, transmitted and communicated in terms of zeros and ones. However, the idea of binary phenomena constitutes a much more fundamental concept to the natural and social world than merely that of digital information. In medicine, we record whether patients are healthy or sick; in engineering, we record whether physical systems are intact or shattered; in sociology, we record whether job applications are successful or not; in machine learning, we train models to predict whether a given picture is of a cat or a dog. In fact, it was the epistemological study of the uncertainty related to whether certain events would occur or not (i.e. binary outcomes) by philosophers like David Hume and Thomas Bayes that led to the birth of probability theory as a mathematical discipline in the first place. It should come as no surprise that the analysis of binary data, or tasks like binary classification, has been a key application in the field of statistics ever since the birth of the subject. Like many areas of mathematical statistics, a particularly appealing aspect of the analysis of such data is the versatility of the mathematics involved. Indeed, a mathematical model for predicting whether a patient is infected or immune to a certain disease may later be adapted by an engineer to model whether a physical system is intact or shattered. For the statistician, the data arising in these two settings are nearly identically interpreted, with little regard to the application they are serving.

Having appreciated the generality of the themes involved in binary responses, let us shift our attention to the key application of the present thesis, namely the analysis of the sensitivity of energetic materials. These are molecules or compounds which release large amounts of chemical energy quickly upon decomposition, and are a key ingredient in explosives, propellants and pyrotechnics, both in civil and military sector. A proper understanding of the sensitivities of such materials is crucial for ensuring safety at each step of the life cycle of explosives, which includes production, manufacture, transport, usage and destruction. This is particularly true for explosive remnants of war (ERW) and dumped ammunition, whose hazard properties are much less studied than newly manufactured explosives. There already exist millions of tonnes of ERW and dumped ammunition worldwide, and the volume is rapidly increasing with the horrific ongoing wars of aggression in Ukraine and Gaza. It is worth noting that most victims of spontaneous detonation of ageing explosives in ERW are civilians.

Table 1.1: Data from fallhammer experiments on amatol extracted from explosive remnants of World War II.

Height ($\log_{10}$ cm)	1.05	1.10	1.15	1.20	1.25	1.30	1.35	1.40	1.45	1.50	1.55
# Reactions	0	1	2	2	5	2	2	2	1	1	1
# Trials	1	3	3	6	6	3	3	2	1	1	1

The sensitivity of an energetic material refers to its susceptibility to ignite under different forms of external stimuli, such as physical impact, friction, electric discharge or heat. To avoid too much generality, let us focus on the former, namely impact. To measure impact sensitivity, an apparatus known as a fallhammer is used, in which a weight of known mass is repeatedly dropped from different heights onto samples of the explosive of interest. For each drop, the operator observes whether an explosion occurred or not. Hence, the outcomes are binary in nature (no explosion versus explosion), and depend on a single covariate, namely the height from which the weight is dropped. To introduce some notation, let $t_1, \ldots, t_n \in \mathbb{R}$ denote these heights (usually on a log-scale in the sensitivity testing literature) and let $y_1, \ldots, y_n \in \{0, 1\}$ denote the binary outcomes, where $y_i = 0$ means no explosion and $y_i = 1$ means explosion. A real dataset which was obtained in the research involved in **Paper VII** is given in Table 1.1. Here, $n = 30$ drops were conducted on a sample of amatol, a common explosive during World War I and World War II, extracted from explosive remnants of World War II and analysed at the Norwegian Defence Research Establishment (FFI).

Some key questions for the statistician to answer include:

- At what height is there a 50% probability of an explosion occurring? What about a 99.9% probability? These heights are commonly referred to as h_{50} and $h_{99.9}$, respectively.

- How confident should we be in our estimates of quantiles such as h_{50} or h_{99}?

- How should we choose the heights $t_1, \ldots, t_n$?

From a mathematical point of view, the third point in particular is of great interest, as it marks a divergence from the standard assumptions of regression problems usually encountered in statistics. Indeed, the standard assumption in regression (including binary regression, i.e. regression with binary outcomes) is to view the covariate values $t_1, \ldots, t_n$ as fixed, or, as realisations of independent and identically distributed (i.i.d.) random variables. However, in sensitivity testing, the ith height is usually determined from the results of the previous $i-1$ measurements, meaning that the sequence $t_1, t_2, \ldots$ is in fact a stochastic process. A very simple experimental design, which is one of the main points of study in **Paper V** and **Paper VI**, is the Bruceton method. Introduced by Dixon and Mood (1948) and named after the Bruceton Research Center in Pittsburgh, Pennsylvania, this simple design is imposed both by NATO's and the U.S. Department of Defense's standard procedures for sensitivity testing of explosives (NATO, 1999; NATO, 2009a; NATO, 2009b; U.S. Department of Defense, 2001). The design depends

on two parameters, chosen by the researcher before the testing begins. First, the researcher needs to decide on the first input t_1. Second, the researcher chooses a step size $d > 0$, determining how far up or down the weight should be set between consecutive drops. The design is defined by the inductive rule

$$t_i = \begin{cases} t_{i-1} - d & \text{if } y_{i-1} = 1, \\ t_{i-1} + d & \text{if } y_{i-1} = 0, \end{cases} \qquad (1.1)$$

for $i = 2, \ldots, n$. That is, if we observe an explosion, we go down one step at the next measurement, and vice versa. We now see the source of the second half of the thesis' title, namely *up or down*. As a result of the up-and-down motion of the Bruceton design, the covariate sequence $t_1, t_2, \ldots$ is a time-homogeneous Markov chain, whose large-sample properties have been thoroughly studied (Derman, 1957; Durham and Flournoy, 1994; Flournoy, 2002). This is also the main topic of **Paper V**. Although other more efficient experimental designs have been proposed since (Joseph, 2004; Langlie, 1965; Neyer, 1994; Wu and Tian, 2014), the Bruceton design remains a simple, nonparametric alternative which is widely used in both civil and military industry.

We conclude this chapter by considering some other key applications in which binary responses naturally arise, other than sensitivity testing. A classic example is that of current status data, in which we wish to monitor when individuals in a population transition from a state A to another state B. For instance, suppose we want to measure the onset of menarche in a population, i.e. the age at which girls experience their first menstruation. We would do this by asking a sample of n girls (of ages ranging from 8 to 20, say), whether they have experienced menarche yet. If the girls' ages are $t_1, \ldots, t_n$, then their answers $y_1, \ldots, y_n \in \{0, 1\}$ are binary indicators from which the onset of menarche is estimated. A similar example involving current status data, monitoring immunity to the rubella virus, is considered in **Paper I**. Although current status data problems are indeed very similar to sensitivity testing, there is a crucial mathematical difference in how the inputs $t_1, \ldots, t_n$ arise. In current status data problems, t_i is the age of the ith individual in the sample, and it is natural to assume that $t_1, \ldots, t_n$ are independently sampled from some latent distribution, and in particular, that the responses $y_1, \ldots, y_n$ are conditionally independent given $t_1, \ldots, t_n$. In sensitivity testing, on the other hand, t_i is the height of the ith experiment, and is chosen by the researcher. As mentioned in the previous paragraph, this typically means that the heights t_i are chosen according to some experimental design (depending on the outcomes y_i). Crucially, the responses $y_1, \ldots, y_n$ are not conditionally independent given $t_1, \ldots, t_n$.

A further application from sociology, which is the main topic of **Paper VIII**, is to predict the outcome of applications in situations such as employment, housing and participation in recreational activities. Here, the main question is to estimate whether attributes such as the applicant's sex, age or ethnicity affect the probability of their application being successful. In order to do this, a field experiment is needed, in which a large number n of essentially equal fictional applications are sent out, which differ only in the aforementioned attributes, *mutatis mutandis*. For example, if we wish to measure whether the applicant's

ethnicity affects whether the application is successful, half of the applications would be signed by a name signalling a certain ethnicity, whilst the other half would be signed with a name signalling another. Letting t_i denote the vector of covariates (e.g. sex, age, ethnicity etc.) of the ith applicant, the binary outcome y_i is determined by whether they receive a response or not within a pre-determined time limit. In **Paper VIII**, such an experiment was conducted to investigate whether parents from an ethnic minority in Norway had a lower chance of receiving a reply when enquiring on behalf of their ten-year-old child about a trial football practice.

Having introduced the concept of binary data and the significance of sensitivity testing of energetic materials in particular, we will now look closer at the question of choice of methodology in the context of mathematical statistics. In the next chapter, we will investigate four of the main approaches to statistical inference, with a particular emphasis on how this affects the analysis of binary data problems.

Chapter 2

Approaches to statistical inference

Like any field of science, mathematical statistics may be approached by multiple perspectives, all of which have their own strengths and weaknesses. Although individual statisticians may have a favourite angle from which to approach problems, awareness of the possibilities and limitations of the different approaches to the subject is a crucial part of reaching scientific maturity.

2.1   The big picture (resolution: 2 x 2)

Although there exist many useful ways in which to split up the subject of statistics, we shall focus on two dichotomies which prove particularly useful for establishing a meta-narrative for the present thesis. Firstly, we shall separate between *frequentist* and *Bayesian* statistics, which, loosely speaking, constitutes a difference in the mathematical formulation of ignorance, and, consequently, in how information is inferred from data. Secondly, we shall separate between *parametric* and *nonparametric* statistics, which constitutes a difference in which mathematical models are under consideration. Considering all four possible combinations of these subfields, we end up with the 2×2 grid in Table 2.2. This meta-perspective is useful as it allows us to narrow in on the main focus of the present thesis: namely the shaded regions of Table 2.2, *frequentist parametrics* and *Bayesian nonparametrics*. Whilst the first two papers concern methodological developments in the latter perspective, the remaining articles focus on the former, and contain both methodological advances as well as considerations of real-world applications. The only slight exception to this rule is **Paper III**, which, in addition to a frequentist parametric regression analysis, also contains a comparison between the frequentist parametric and Bayesian parametric viewpoints. In the following chapters, we shall cover each of the four approaches outlined in Table 2.2. Naturally, the chapters on frequentist parametrics and Bayesian nonparametrics will serve as a preamble for the submitted manuscripts, but an introduction to the two remaining perspectives is also vital for understanding the context in which this thesis' contributions reside. The following chapters do by no means serve as complete accounts, but rather introduce some key concepts needed for contextualising the articles.

Table 2.2: The big picture of statistical inference, differentiating between frequentist versus Bayesian inference and parametric versus nonparametric modelling. The shaded regions indicate the focus of the present thesis.

	Frequentist	Bayesian
Parametric	Frequentist parametrics	Bayesian parametrics
Nonparametric	Frequentist nonparametrics	Bayesian nonparametrics

Chapter 3

Frequentist parametrics

We begin with the domain of frequentist parametrics, which comprises the majority of the articles written. In particular, we shall introduce the idea of confidence curves, which provide a useful graphical summary how confident we should be in our parameter estimates. Also known as classical statistics, frequentist statistics form the largest and most standard approach to mathematical statistics currently. The underlying idea is that data are probabilities may be inferred from frequencies of occurrences in observations. That is, the frequency of the occurrence of an event directly reflects the probability of that event occurring.

3.1 Maximum likelihood estimation and confidence intervals

Suppose we want to infer properties of a parameter of interest $\theta \in \Theta$, living in some parameter space Θ. In the parametric domain, Θ is assumed to be finite-dimensional. We wish to infer properties of θ from observed *data*, say $y = (y_1, \ldots, y_n) \in \mathcal{Y}$. We view the observations y_i as realisations of random variables Y_i. In the simplest case, the variables Y_i are assumed to be independent and identically distributed (i.i.d.), whose density (or mass function, if they are discrete) will be denoted $f_\theta(y)$. The combined information from the model choice and the observed data is summarised by the likelihood function

$$L_n(\theta) = \prod_{i=1}^{n} f_\theta(y_i).$$ (3.1)

However, in practice, it turns out to be easier to work with the log likelihood function, given by

$$\ell_n(\theta) = \log L_n(\theta) = \sum_{i=1}^{n} \log f_\theta(y_i).$$

The maximum likelihood estimator $\widehat{\theta}$ is the maximiser of either of these functions,

$$\widehat{\theta}_n = \arg\max_\theta L(\theta) = \arg\max_\theta \ell(\theta).$$ (3.2)

In most cases we will simply write $\widehat{\theta} = \widehat{\theta}_n$ as the number of observations in the data is implicitly known. The maximum likelihood estimator $\widehat{\theta}$ is widely used due to its appealing asymptotic properties. In particular, for i.i.d. data, letting θ_0 denote the true parameter value, we have that[1]

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathrm{N}\left(0, \mathcal{I}(\theta_0)^{-1}\right) \tag{3.3}$$

where

$$\mathcal{I}(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta\partial\theta^\top}\log f_\theta(Y)\right]$$

is the Fisher information of the model. In the case of regression, a similar result holds, but here, the Fisher information takes the form

$$\mathcal{I}_n(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta\partial\theta^\top}\sum_{i=1}^{n}\log f_\theta(Y_i \mid z_i)\right],$$

where $z_1, \ldots, z_n$ are the observed covariates. The corresponding asymptotic result for regression then asserts that $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathrm{N}(0, \mathcal{I}(\theta_0)^{-1})$, where now $\mathcal{I}(\theta)$ is the probability limit of $\mathcal{I}_n(\theta)$ as $n \to \infty$.

The convergence result (3.3) yields a natural way of constructing confidence intervals (CIs) for parameters of interest. Indeed, let $\xi = g(\theta) \in \mathbb{R}$ be a focus parameter. Suppose we want to construct a $100(1-\gamma)\%$ CI for ξ, where $0 < 1-\gamma < 1$ is our desired confidence level. Let $\Phi(a) = (2\pi)^{-1/2}\int_{-\infty}^{a}\exp\{-x^2/2\}\,\mathrm{d}x$ denote the standard normal cumulative distribution function (c.d.f.). Then applying the delta method to (3.3) and appealing to functional invariance[2] of the maximum likelihood principle, we obtain

$$\sqrt{n}\left(g(\widehat{\theta}) - g(\theta_0)\right) \xrightarrow{d} \mathrm{N}\left(0, \nabla g(\theta_0)^\top \mathcal{I}(\theta_0)^{-1}\nabla g(\theta_0)\right).$$

Hence, we obtain $\widehat{\xi} \pm z_{\gamma/2}\widehat{\sigma}/\sqrt{n}$ as our $100(1-\gamma)\%$ CI for ξ, where $z_{\gamma/2} = \Phi^{-1}(1-\gamma/2)$ and

$$\widehat{\sigma}^2 = \nabla g(\widehat{\theta})\mathcal{I}(\widehat{\theta})^{-1}\nabla g(\widehat{\theta}).$$

Example 1. Let us study the data given in Table 1.1 via a standard model in frequentist statistics, namely *probit regression*. Let $\phi(a) = (2\pi)^{-1/2}\exp\{-a^2/2\}$ denote the standard normal probability density function (p.d.f.). We impose the model $Y_i \sim \mathrm{Bernoulli}(p_i)$, independently for $i = 1, \ldots, n$, where $p_i = \Phi(\alpha + \beta x_i) = \Phi(\theta^\top z_i)$. Here, $\theta = (\alpha, \beta)^\top$ are the model parameters and $z_i = (1, x_i)^\top$ are the covariate vectors. Hence, $\mathbb{P}(Y_i = 1) = \Phi(\alpha + \beta x_i) = 1 - \mathbb{P}(Y_i = 0)$. The log-likelihood takes the form

$$\ell_n(\theta) = \sum_{i=1}^{n}\left\{y_i\log\Phi(\theta^\top z_i) + (1-y_i)\log[1 - \Phi(\theta^\top z_i)]\right\},$$

---

[1]Strictly speaking, all convergence statements should be of the form $X_n \xrightarrow{d} X$, where X is the random variable in the limit, not a distribution. We shall, however, abuse notation slightly and write statements like $X_n \xrightarrow{d} \mathrm{N}(0,1)$ for brevity.

[2]Here, functional invariance refers to the fact that under the parametrisation $\xi = g(\theta)$, the maximum likelihood estimator for ξ is found by simply applying the function g to that of θ, i.e. $\widehat{\xi} = g(\widehat{\theta})$.

which is concave in θ, and may thus easily be optimised numerically using e.g. the Newton–Raphson method. Doing so for the data in Table 1.1, we obtain the estimates $\widehat{\alpha} = -7.319$ and $\widehat{\beta} = 6.172$. In the context of binary regression, researchers are commonly particularly interested in estimating the median of the dosage-response curve, i.e. the value x at which we have $\Phi(\alpha + \beta x) = 1/2$. For the probit regression model (and indeed for many other standard model choices), this value is given by $\xi = g(\alpha, \beta) = -\alpha/\beta$. For the data in Table 1.1, $\widehat{\xi} = -\widehat{\alpha}/\widehat{\beta} = 1.186$. In Figure 3.1, we see the resulting fitted curve.

To obtain the asymptotic distribution of $\widehat{\theta}$, we first compute the Fisher information, which in this case takes the form

$$\mathcal{I}_n(\theta) = \sum_{i=1}^{n} \frac{\phi(\theta^\top z_i)^2}{\Phi(\theta^\top z_i)[1 - \Phi(\theta^\top z_i)]} z_i z_i^\top.$$

Given the data in Table 1.1, the asymptotic covariance of $\widehat{\theta}$ is estimated by

$$V = \mathcal{I}_n(\widehat{\theta})^{-1} = \begin{pmatrix} 11.046 & -8.932 \\ -8.932 & 7.266 \end{pmatrix}.$$

Now, letting $\xi = g(\alpha, \beta) = -\alpha/\beta$, we have $\nabla g(\alpha, \beta) = (-1/\beta, \alpha/\beta^2)^\top$, and so the asymptotic variance of $\widehat{\xi}$ is estimated by

$$\widehat{\sigma}^2 = \frac{1}{\widehat{\beta}^2} \left\{ V_{11} + 2\widehat{\xi} V_{12} + (\widehat{\xi})^2 V_{22} \right\} = 0.002.$$

Hence, we obtain $[1.141, 1.231]$ as our 95% CI for ξ.

The above procedure is commonly referred to constructing CIs via the delta method. Although this is arguably the best known method for creating generic CIs, it should be noted that there exist other alternatives which are based on other large-sample results than (3.3). Another highly general recipe is to instead construct CIs via Wilks' theorem (Wilks, 1938), the main steps of which we shall now go through. To begin, define the *profile log-likelihood* ℓ_{prof} by

$$\ell_{\text{prof}}(\xi) = \sup \left\{ \ell(\theta) : g(\theta) = \xi \right\}.$$

Next, let $\widehat{\ell} = \ell(\widehat{\theta})$ to be the maximum value the log-likelihood function takes overall, and define the *deviance function* $D(\xi)$ to be twice the difference between these two quantities, $D(\xi) = 2\left\{ \widehat{\ell} - \ell_{\text{prof}}(\xi) \right\}$. Wilks' theorem now asserts that, under model conditions,

$$D(\xi) \xrightarrow{d} \chi_1^2,$$

where χ_p^2 denotes the chi-squared distribution with p degrees of freedom. Note that this theorem can be interpreted as a likelihood-ratio test between the alternative hypothesis $H_1 : g(\theta) = \xi$ against the natural null hypothesis $H_0 : g(\theta) \in \mathbb{R}$, and the manifold parametrised by $g(\theta) = \xi$ reduces the dimension of the parameter space Θ by one dimension, yielding a chi-squared distribution with a single degree of freedom. The resulting $100(1 - \gamma)\%$ CI resulting from Wilks' theorem is the set of values ξ which satisfy $D(\xi) < \Gamma^{-1}(1 - \gamma) = z_{\gamma/2}^2$, where Γ denotes the c.d.f. of
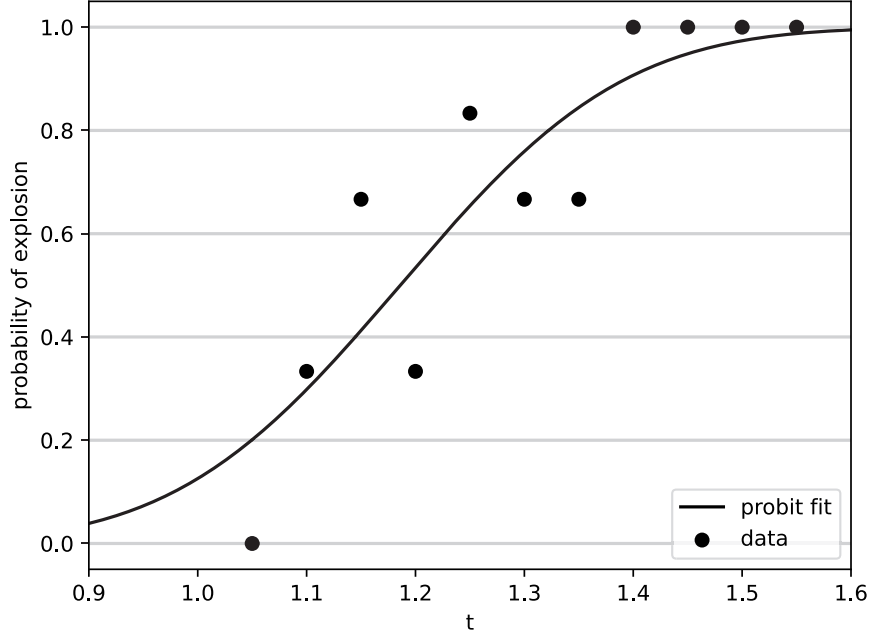
Figure 3.1: A fitted probit curve based on the $n = 30$ observations from Table 1.1.

a χ_1^2 distribution. In general, this CI will be different from that obtained via the delta method. In particular, there is no guarantee that it will be symmetric about the maximum likelihood estimator $g(\widehat{\theta})$.

In the case of estimating quantiles of a sensitivity curve in binary regression, a further method still is to invoke Fieller's theorem (Fieller, 1954), which relies on the following observation. Suppose $(\widehat{\alpha}, \widehat{\beta})^\top$ is a two-dimensional normally distributed vector with mean $(\alpha, \beta)^\top$ and 2×2 covariance matrix V. Letting $\xi = -\alpha/\beta$ and $\widehat{\xi} = -\widehat{\alpha}/\widehat{\beta}$, we have, by properties of the normal distribution, that

$$\xi\widehat{\beta} + \widehat{\alpha} \sim \mathrm{N}(0, V_{11} + 2\xi V_{12} + \xi^2 V_{22}).$$

Consequently,

$$\frac{(\xi\widehat{\beta} + \widehat{\alpha})^2}{V_{11} + 2\xi V_{12} + \xi^2 V_{22}} \sim \chi_1^2. \tag{3.4}$$

Hence, our $100(1 - \gamma)\%$ CI will be the set of values ξ satisfying $(\xi\widehat{\beta} + \widehat{\alpha})^2/\{V_{11} + 2\xi V_{12} + \xi^2 V_{22}\} < \Gamma^{-1}(1 - \gamma) = z_{\gamma/2}^2$. Equivalently,

$$\left(\widehat{\beta}^2 - z_{\gamma/2}^2 V_{22}\right) \xi^2 + 2\left(\widehat{\alpha}\widehat{\beta} - z_{\gamma/2}^2 V_{12}\right) \xi + \widehat{\alpha} - z_{\gamma/2}^2 V_{11} < 0, \tag{3.5}$$

a quadratic inequality in ξ, whose set of solutions will depend on whether the leading coefficient is non-negative, i.e. whether the parabola described by (3.5) is convex or concave. If $\widehat{\beta}^2 - z_{\gamma/2}^2 V_{22} < 0$, then said parabola is concave, and so the CI will either be the union of two disjoint unbounded intervals or the entire real

line. However, in most cases, it will be the case that $\widehat{\beta}^2 - z_{\gamma/2}V_{22} > 0$, and so the CI will be a bounded interval whose endpoints can be explicitly computed as

$$\widehat{\xi} + \frac{g}{1-g}\left(\widehat{\xi} + \frac{V_{12}}{V_{22}}\right) \pm \frac{z_{\gamma/2}}{\widehat{\beta}(1-g)}\sqrt{V_{11} + 2\widehat{\xi}V_{12} + \widehat{\xi}^2 V_{22} - g\left(V_{11} - \frac{V_{12}^2}{V_{22}}\right)},$$

where $g = z_{\gamma/2}^2 V_{22}\widehat{\beta}^2 < 1$. Again, this interval is in general not symmetric about $\widehat{\xi}$. Note also that the interval will be empty if the expression under the radical is negative.

It is worth pointing out that unlike the delta method, the result (3.4) is exact, and does not rely on use of the delta method. That is, CIs obtained via Fieller's theorem only rely on asymptotic normality of $\widehat{\theta}$ itself, and no further approximations. A further interpretation of Fieller's theorem, which is explained in **Paper IV**, shows that it may be perceived as a hybrid between the delta method and the likelihood ratio test. This connection is also explained by Schweder and Hjort (2016). In short, if we apply the likelihood ratio test via the profile log-likelihood *after* the asymptotic normality assumption has been made (rather than using the exact log likelihood function as we do in the vanilla likelihood ratio test), then we recover Fieller's theorem.

3.2 Confidence curves

Having seen three different recipes for constructing CIs for a focus parameter in the binary regression setting, we now shift our attention to confidence curves (CCs), which provide an appealing and illuminating graphical summary of the confidence we ought to have in different point estimates. Schweder and Hjort (2016, Chapter 4) explain CCs as a consequence of inference via confidence distributions, the study of which may serve as a revivification of fiducial inference, as originally explored by Fisher (1935). However, in the present thesis, we shall merely apply CCs as a useful graphical tool for frequentist inference.

The basic idea of CCs is to plot all CIs obtained when letting the confidence level range between 0 and 1. For a given $\gamma \in (0,1)$ on the y-axis, we plot the corresponding $100(1-\gamma)\%$ CI obtained on the x-axis. In doing so, we combine an infinite collection of confidence intervals in a single plot, obtaining a cusp-like curve (see Figure 3.2 for examples). At the nadir of a CC, where the confidence level is nil, all CIs collapse to a single point, namely the maximum likelihood estimate. As the confidence level increases, the CIs become wider and asymptotically cover the entire real line, which happens at 100% confidence.

Example 2. Let us now continue our analysis of the data from Table 1.1. Previously, we calculated a 95% CI for $\xi = -\alpha/\beta$ using the delta method. We will now compute CCs for ξ using two of the above methods, namely the delta method and Fieller's theorem. Since the construction of CCs via Wilks' theorem is much less numerically stable, we omit it in this analysis. The curves are plotted in Figure 3.2. This figure also illustrates how we would read off a 95% CI, using either the delta method or Fieller's theorem. For example, if we want to read off a

15
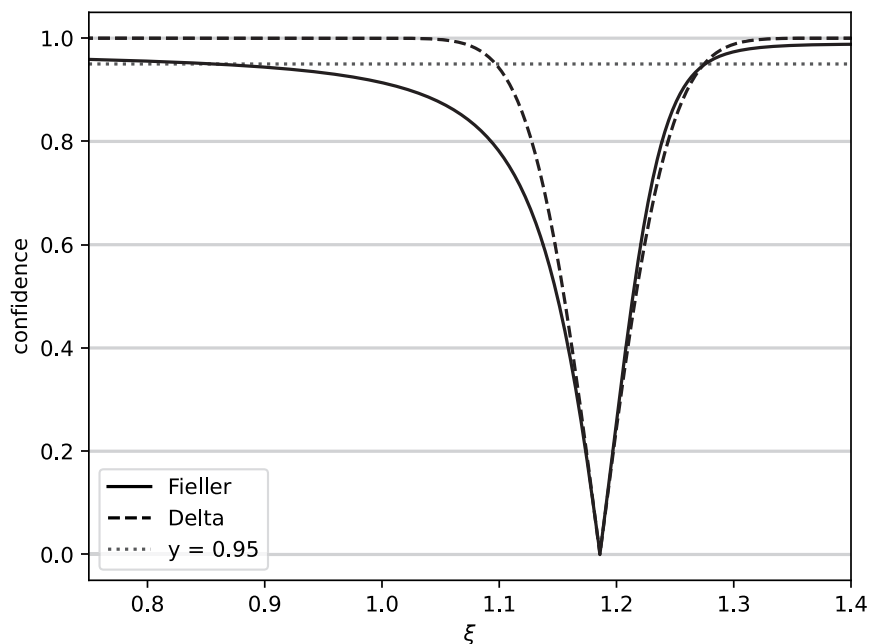
Figure 3.2: Confidence curves for ξ via the delta method and Fieller's theorem. Note that, given normality, the former is approximate and the latter is exact.

95% CI using the delta method, we simply locate where the line $y = 0.95$ intersects the CC constructed using the delta method. Note that doing so, we recover the interval given previously. From Figure 3.2, we also see explicitly that Fieller's theorem yields asymmetric CIs, whereas the delta method does not. In particular, the CIs derived using Fieller's theorem are more conservative regarding their lower bounds than those derived using the delta method. For instance, Fieller's theorem yields $[0.857, 1.276]$ as its 95% CI for ξ. Simulation studies consistently show that this is due to the delta method being over-confident, see Abdelbasit and Plackett (1983), Cox (1990), Faraggi et al. (2003) and Sitter and Wu (1993). **Paper IV** reports a simulation study specifically investigating data obtained via the Bruceton design (1.1) and reaches similar conclusions.

Chapter 4

Frequentist nonparametrics

Having seen the basic recipe for plain frequentist parametric inference, at least in the case of binary regression, we shall now explore the corresponding nonparametric approach. The general story of nonparametric inference and estimation techniques comprises an extremely broad literature which stretches well beyond the scope of this thesis, so we shall focus our attention on the binary regression setting specifically. This story began in 1955, when Ayer et al. (1955) provided a recipe for maximising the log-likelihood

$$\ell_n(F) = \sum_{i=1}^{n} \{y_i \log F(t_i) + (1 - y_i) \log[1 - F(t_i)]\} \tag{4.1}$$

over the class of *all* distribution functions F. For the remainder of this chapter we shall denote this maximiser by $\widehat{F}$. Since (4.1) only depends on F via the values taken at the points $t_1, \ldots, t_n$, $\widehat{F}$ is only unique up to the values it takes in between these points. However, any valid solution $\widehat{F}$ must be a valid distribution function, i.e. non-decreasing and right-continuous. The maximisation of (4.1) can therefore be thought of as a maximisation problem under shape constraints. A common choice is to insist that $\widehat{F}$ should be a step function (as done by Groeneboom and Jongbloed (2014)) or to linearly interpolate between the t_i (as done by Bhattacharya and Kong (2007)).

4.1 Binary responses and censored data

In order to understand further how the discovery of Ayer et al. (1955) relates to other nonparametric estimators, it is now time that we consider the relation between binary regression problems and censored data. In statistics, we call an observation censored if it takes the form of an event $\{T \in A\}$ for some set A, rather than an exact value, say $T = t$. Whenever data are censored, any inferential procedure must take into account that the exact value of the variable of interest has not been observed, but rather that it is only known that it has landed in the set A. Censored data are most frequently encountered in the subject of survival analysis, in which one aims to estimate the time it takes for a group of individuals to go from a state 0 to another state 1 (e.g. time until death for a group of patients). In any such study monitoring a group of patients, censoring will naturally occur.

Indeed, for all the patients who survive the study, the time of death is by definition not observed, and so we only know that it will occur at some point after a certain time t. Such observations are called right censored, since the censoring set A takes the form $[t, \infty)$.

Censoring and binary responses are naturally related in the following way. Suppose we have made a number of measurements at the inputs $t_1, \ldots, t_n \in \mathbb{R}$. Now let $X_1, \ldots, X_n$ be latent i.i.d. variables, each of which distributed according to the distribution emitted by the c.d.f. F. That is, $\mathbb{P}(X_i \leq a) = F(a)$, independently for all $i = 1, \ldots, n$. Then if we define the variables Y_i to be the binary indicators $Y_i = \mathbb{1}\{X_i \leq t_i\}$ for $i = 1, \ldots, n$, then we recover the binary regression model. Indeed, defining the Y_i this way yields

$$\mathbb{P}(Y_i = 1) = \mathbb{P}(X_i \leq t_i) = F(t_i),$$

as required. Hence, any binary regression problem may equivalently be interpreted as a problem in which all observations are either left or right censored. In fact, the title of the original article by Ayer et al. (1955) includes the phrase "incomplete information", referring to the fact that the observations are not exact. Three years after this work was published, it was discovered by Kaplan and Meier (1958) that there also exists a canonical nonparametric estimator (called the Kaplan–Meier estimator) in the case where each of the observations is either exact or right censored. Due to the prevalence of such data in medicine research, the Kaplan–Meier estimator is one of the most successful and substantial inventions in all of mathematical statistics. In the time of writing this thesis, the article by Kaplan and Meier (1958) has been cited more than 65,000 times. A further improvement was made by Turnbull (1974), who showed that both the nonparametric maximum likelihood estimator by Ayer et al. (1955) and the Kaplan–Meier estimator can be seen as special cases of a class of nonparametric estimators given censored data.

Example 3. We will now analyse the data from Table 1.1 using the nonparametric maximum likelihood estimator. The results are plotted in Figure 4.3. Note that both the step function and the partially linear function are equally valid realisations of $\widehat{F}$, as $\widehat{F}$ is only unique up to its behaviour on the intervals between the observed values t_i.

4.2 Asymptotics

We will conclude this chapter by considering the large-sample properties of the nonparametric maximum likelihood estimator, which were first established by Groeneboom and Wellner (1992). Like in the parametric setting, the nonparametric estimator also exhibits asymptotic normality. However, rather than the standard convergence rate of $\sqrt{n}$, like we saw in (3.3), we now get a rate of $n^{1/3}$. This behaviour is indeed typical for nonparametric estimators, and reflects the balance between an estimator's flexibility and efficiency. If a parametric model is known to be true, then a parametric estimator (like the standard maximum likelihood estimator) is nearly always guaranteed to more
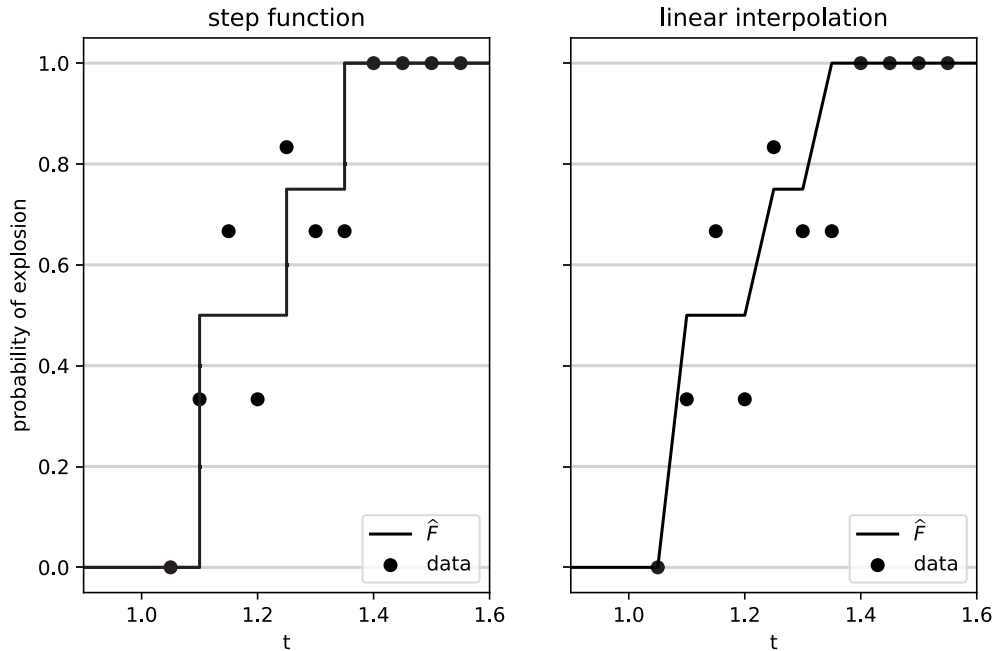
Figure 4.3: The nonparametric maximum likelihood estimator, realised either as a step function or a partially linear function, based on the $n = 30$ observations from Table 1.1.

efficient than any nonparametric approach. However, parametric models may be misspecified and are thus not as flexible as nonparametric approaches.

The main asymptotic result for $\widehat{F}$ is summarised by Groeneboom and Jongbloed (2014, Theorem 3.7). We restate this theorem here.

Theorem 4.2.1. *Let $X_1, X_2, \ldots$ be independent random variables with c.d.f. F_0 and let $T_1, T_2, \ldots$ be independent random variables with c.d.f. G. Insist also that these sequences are independent of each other. For all $i \in \mathbb{N}$, let $Y_i = \mathbb{1}\{X_i \leq T_i\}$ and suppose the data observed are $\{(T_i, Y_i)\}_{i=1}^{n}$. Let $t_0 \in \mathbb{R}$ be any point such that $0 < F_0(t_0), G(t_0) < 1$. Furthermore, assume that F_0 and G are both differentiable at the point t_0 and that their derivatives $f_0(t_0)$ and $g(t_0)$ are strictly positive. Then*

$$n^{1/3}\left(\widehat{F}(t_0) - F_0(t_0)\right) \xrightarrow{d} \{4F_0(t_0)[1 - F_0(t_0)]f_0(t_0)/g(t_0)\}^{1/3} 2Z$$

where Z follows Chernoff's distribution. That is, $Z = \arg\max_t\{B(t) - t^2\}$, where $B(t)$ is two-sided Brownian motion with $B(0) = 0$.

Much like the standard asymptotic theory of parametric maximum likelihood estimation, Theorem 4.2.1 is a fascinating and remarkably general result. The fact that such a precise statement can be made with virtually no assumptions regarding the behaviour of the functions F_0 and G apart from local differentiable is certainly a triumphant feat. Nevertheless, the theorem has its limitations. In particular, note that it only asserts the large-sample behaviour of $\widehat{F}(t_0)$ in a specific point t_0, and so it does not directly yield a convergence theorem for the estimated median
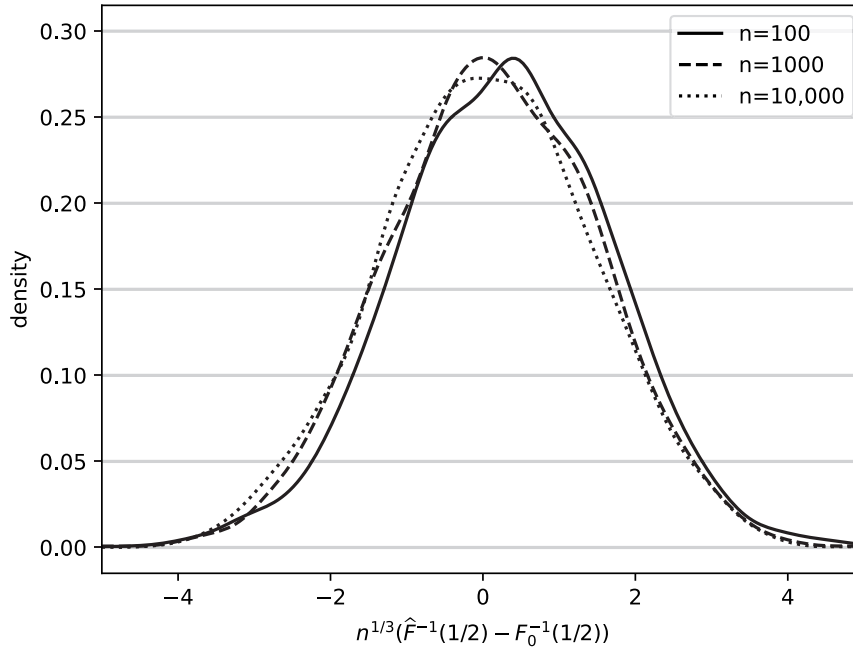
Figure 4.4: Indications of cube root asymptotics for $\widehat{F}^{-1}(1/2)$.

$\widehat{F}^{-1}(1/2)$, for example. It is worth mentioning here that Bhattacharya and Kong (2007) have in fact established a limiting distribution for $\widehat{F}^{-1}(1/2)$, but their setup is rather different to that of Theorem 4.2.1. Indeed, rather than assuming that the values $t_1, t_2, \ldots$ are sampled i.i.d. from F_0, they insist that $t_1, t_2, \ldots$ all take values in a finite set of k values $v_1, \ldots, v_k$, and that approximately the same number of binary trials are performed at each of these as the total number of trials tends to infinity. Using linear interpolation between the inputs for $\widehat{F}$, they derive, under certain regularity conditions, asymptotic normality of $\widehat{F}^{-1}(q)$, for $0 < q < 1$, with a normal limit and a convergence rate of $\sqrt{n_c}$, where n_c is the number of binary trials performed at the input v_c closest to the true quantile $F_0^{-1}(q)$. The fact that we recover square root asymptotics here, like we saw in Chapter 3, deserves a moment of appreciation. It seems that by insisting that the inputs all take one of finitely many possible values $v_1, \ldots, v_k$, we somehow collapse the problem back to the parametric realm and recover square root asymptotics. This is true in a very real sense; when there are only finitely many inputs $v_1, \ldots, v_k$, the nonparametric maximum likelihood estimation can be seen as the simultaneous estimation of the values $(F(v_1), \ldots, F(v_k))$ under the constraint that $0 \leq F(v_1) \leq \cdots \leq F(v_k)$. That is, there is a natural parametric formulation of the problem (also keeping in mind that we interpolate linearly between the values $v_1, \ldots, v_k$). When the inputs $t_1, t_2, \ldots$ are simply assumed to be i.i.d. samples from F_0, however, as is the case in Theorem 4.2.1, this parametric formulation fails and we are forced to think genuinely nonparametrically, as the number of parameters of the model grows with the number of observations.

It would probably be possible to extend Theorem 4.2.1 to also get results for the estimated quantiles, but as far as the author is aware, this has not been done yet. By performing a simple simulation study, however, we can see that any such theorem will most likely need cube root asymptotics, much like Theorem 4.2.1. With the X_i i.i.d. from a Student's t-distribution with three degrees of freedom and the T_i i.i.d. from a Uniform$[0, 1]$ distribution, we simulate samples of size $S = 3000$ of the variable $n^{1/3}(\widehat{F}^{-1}(1/2) - F_0^{-1}(1/2))$ for $n = 100$, $n = 1000$ and $n = 10,000$. The results of this simulation are given in Figure 4.4, in which we clearly see that these three distributions are of similar scale. Further work is nevertheless needed to find the limiting distribution in explicit form.

Chapter 5

Bayesian parametrics

We shall now shift our attention to Bayesian statistics. Although everything in the present chapter will be explained in the context of parametric models, the concepts introduced here will also be crucial when we come to the subject of Bayesian nonparametrics in Chapter 6.

5.1   Prior and posterior distributions

Although the differences between frequentist and Bayesian statistics may be formulated as a philosophical problem in epistemology, the mathematical difference between the two perspectives may be summarised as follows: In Bayesian statistics, unknown parameters are treated as random variables, whereas in frequentist statistics, they are not. Returning to the notation introduced in Chapter 3, let $\theta \in \Theta$ be a vector of parameters whose properties we want to infer, residing in a finite-dimensional space Θ. In the next chapter, on Bayesian nonparametrics, we shall also include the possibility of Θ being an infinite-dimensional space, such that θ can be for example a random probability measure, or a random function. Since the parameters θ are unknown, we treat them as a random variable, and assign to it a probability distribution $\pi(\theta)$, called the prior. The word prior is used because the distribution $\pi(\theta)$ reflects our believes about θ *prior* to having observed any data. There is no fixed recipe for choosing π correctly, and different problems will require different priors. When eliciting a prior distribution, however, there are some key concepts the researcher should keep in mind. Once such concept is that of domain expertise. Statisticians should always confer with experts to understand what kind of behaviour of θ is reasonable to expect. Another concept to keep in mind is whether the goal of the statistical inference is really to answer a key question, and, if so, to make sure the prior is as neutral as possible with respect to it. For example, if θ has $[0,1]$ as its domain and the key question at hand is to infer whether $\theta > 0.99$, then a uniform prior would be biased.

Having chosen a suitable prior π, we then observe data $y = (y_1, \ldots, y_n) \in \mathcal{Y}$. The key question for the Bayesian statistician at this point is: *How have our beliefs about θ changed after having observed y?* This question is answered by the posterior distribution, $\pi(\theta \,|\, y)$, which conditions on the observed data. By Bayes' theorem,

$$\pi(\theta \,|\, y) = \frac{\pi(\theta)\pi(y \,|\, \theta)}{\pi(y)}, \tag{5.1}$$

where $\pi(y \,|\, \theta) = L_n(\theta)$ is the likelihood function from (3.1). Since conditioning on θ is well-defined in the Bayesian paradigm, as θ itself is a random variable, we use the notation $\pi(y \,|\, \theta)$ rather than $L_n(\theta)$. The denominator $\pi(y) = \int \pi(\theta')\pi(y \,|\, \theta)\,\mathrm{d}\theta'$ is called the *marginal likelihood* and is usually an intractable integral. As we shall see shortly, the marginal likelihood is a key ingredient in Bayesian model selection and model averaging, which are the main themes of **Paper I** and **Paper II**. From (5.1) and the relation $\pi(y) = \int \pi(\theta')\pi(y \,|\, \theta')\,\mathrm{d}\theta'$, we see that the posterior is entirely defined in terms of the prior $\pi(\theta)$ and the likelihood $\pi(y \,|\, \theta)$. Thus, in Bayesian statistics, the specification of these two objects is collectively referred to as the choice of a model.

Due to its intractability, direct evaluations of the marginal likelihood are typically avoided when computing the posterior density $\pi(\theta \,|\, y)$. One way to do this is to choose a conjugate prior, i.e. a prior distribution of the same functional form (as a function of θ) as the likelihood. For example, let $y_1, \ldots, y_n \,|\, \theta$ be independent Poisson(θ) variables, so that

$$\pi(y \,|\, \theta) = \prod_{i=1}^{n} \frac{\theta^y e^{-\theta}}{y!} \propto \theta^{n\bar{y}} e^{-n\theta},$$

where $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$. Then, as a function of θ, $\pi(y \,|\, \theta)$ takes the same functional form as a Gamma density, so choose a Gamma(a, b) prior for θ, with $\pi(\theta) = \{b^a / \Gamma(a)\}\theta^{a-1}e^{-b\theta}$ for $\theta \geq 0$. Then we have

$$\pi(y \,|\, \theta) \propto \pi(\theta) \times \pi(y \,|\, \theta) \propto \theta^{a-1}e^{-b\theta} \times \theta^{n\bar{y}}e^{-n\theta} = \theta^{a+n\bar{y}-1}e^{-\theta(b+n)}.$$

As this must be a normalised density, we do not need to work out the normalisation constant, so the above forces $\theta \,|\, y \sim \text{Gamma}(a + n\bar{y}, b + n)$.

5.2 Markov chain Monte Carlo (MCMC)

Although conjugate priors avoid evaluations of the marginal likelihood, their use is quite limited. Ideally, we would like to be able to carry out posterior inference regardless of the choice of prior. This is to a large extent achieved by Markov chain Monte Carlo (MCMC) methods, which have been a tremendous success for the field of Bayesian statistics over the preceding half-century. The basic idea of MCMC is to construct a Markov chain $\theta_1, \theta_2, \ldots$ (with a transition kernel not depending on the marginal likelihood) whose long-term behaviour mimics that of the posterior distribution. That is, the chain is ergodic and has the posterior distribution as its stationary distribution. Although there exist many MCMC algorithms, the simplest and one of the most versatile is the Metropolis–Hastings algorithm (Hastings, 1970; Metropolis et al., 1953). The algorithm requires a choice of *proposal distribution* $q(\theta' \,|\, \theta_s)$ which samples a new proposed value θ'

given that the current state of the chain is θ_s at iteration $s \in \mathbb{N}$. This proposed value θ' is then accepted, so that $\theta_{s+1} = \theta'$, with probability

$$\eta(\theta' \,|\, \theta_s) = \min\left\{1, \frac{\pi(\theta' \,|\, y)q(\theta_s \,|\, \theta')}{\pi(\theta_s \,|\, y)q(\theta' \,|\, \theta_s)}\right\}.$$

Otherwise, θ' is rejected and $\theta_{s+1} = \theta_s$. This process is repeated to obtain a chain $\theta_1, \theta_2, \dots$. The first state θ_1 is usually just sampled directly from the prior, $\theta_1 \sim \pi(\theta)$. At first glance, it seems like the marginal likelihood is still needed, as $\eta(\theta' \,|\, \theta_s)$ intrinsically depends on $\pi(y)$ through the posterior densities. However, upon expanding we see that

$$\frac{\pi(\theta' \,|\, y)}{\pi(\theta_s \,|\, y)} = \frac{\pi(y \,|\, \theta')\pi(\theta')/\pi(y)}{\pi(y \,|\, \theta_s)\pi(\theta_s)/\pi(y)} = \frac{\pi(y \,|\, \theta')\pi(\theta')}{\pi(y \,|\, \theta_s)\pi(\theta_s)},$$

so that all evaluations of the marginal likelihood are bypassed. In Algorithm 1, the full Metropolis–Hastings algorithm is given as pseudocode.

Algorithm 1 Metropolis-Hastings

Require: Sample size S
1: $\theta_1 \sim \pi(\theta)$
2: $\mathcal{S} \leftarrow \{\theta_1\}$ ▷ Initialise sample
3: **for** $s = 1, \dots, S - 1$ **do**
4:    $\theta' \sim q(\theta' \,|\, \theta_s)$ ▷ Generate proposal
5:    Calculate acceptance probability

$$\eta(\theta' \,|\, \theta) \leftarrow \frac{\pi(\theta')q(\theta_s \,|\, \theta')}{\pi(\theta_s)q(\theta' \,|\, \theta_s)}$$

6:    $u \sim \text{Uniform}[0, 1]$
7:    **if** $u \leq \eta(\theta' \,|\, \theta_s)$ **then**
8:      $\theta_{s+1} \leftarrow \theta'$ ▷ Accept proposal
9:    **else**
10:      $\theta_{s+1} \leftarrow \theta_s$ ▷ Reject proposal
11:    **end if**
12:    $\mathcal{S} \leftarrow \mathcal{S} \cup \{\theta_{s+1}\}$ ▷ Update sample
13: **end for**

To explore the properties of the Metropolis–Hastings algorithm further, we will need to describe its kernel. The kernel of a Markov chain[1] $\theta_1, \theta_2, \dots$ is the conditional density k satisfying

$$\mathbb{P}(\theta_{s+1} \in A \,|\, \theta_s = \theta) = \int_A k(\theta, \theta')\,\mathrm{d}\theta'.$$

By the presence of the accept/reject step in the Metropolis–Hastings algorithm, its kernel takes the form

$$k(\theta, \theta') = \eta(\theta' \,|\, \theta)q(\theta' \,|\, \theta) + [1 - \eta(\theta)]\,\delta_\theta(\theta'), \tag{5.2}$$

---

[1]We are assuming here that the chain is homogeneous, so that the kernel does not depend on which iteration the chain is at.

where

$$\eta(\theta) = \int \eta(\theta' \mid \theta) q(\theta' \mid \theta) \, d\theta'$$

and δ_θ denotes the degenerate distribution with unit mass at the point θ. It is not difficult to show that the Metropolis–Hastings kernel k satisfies *detailed balance* with respect to the posterior distribution $\pi(\theta \mid y)$. That is, $\pi(\theta \mid y)k(\theta, \theta') = \pi(\theta' \mid y)k(\theta', \theta)$. Indeed, the second term of (5.2) clearly satisfies this relation. For the first term,

$$
\begin{aligned}
\pi(\theta \mid y)\eta(\theta' \mid \theta)q(\theta' \mid \theta) &= \pi(\theta \mid y)q(\theta' \mid \theta) \min\left\{1, \frac{\pi(\theta' \mid y)q(\theta \mid \theta')}{\pi(\theta \mid y)q(\theta' \mid \theta)}\right\} \\
&= \min\left\{\pi(\theta \mid y)q(\theta' \mid \theta), \pi(\theta' \mid y)q(\theta \mid \theta')\right\} \\
&= \pi(\theta' \mid y)q(\theta \mid \theta') \min\left\{1, \frac{\pi(\theta \mid y)q(\theta' \mid \theta)}{\pi(\theta' \mid y)q(\theta \mid \theta')}\right\} \\
&= \pi(\theta' \mid y)\eta(\theta \mid \theta')q(\theta \mid \theta'),
\end{aligned}
$$

as required. In order to deduce ergodicity of the chain yielded by the Metropolis–Hastings algorithm, we also need to verify that it is aperiodic and irreducible. The former condition is automatically guaranteed as there is always a positive probability that the proposal is rejected and $\theta_{s+1} = \theta_s$. The latter is usually guaranteed by choosing an appropriate proposal which yields an irreducible chain (for example, if $\Theta = \mathbb{R}^p$, a Gaussian proposal will always work). Hence, the ergodic theorem for continuous-state space Markov chains applies to the Metropolis–Hastings algorithm, with the posterior as the stationary limit distribution. In practice, we need to run convergence diagnostics to be confident that the chain has converged, such as verifying a low correlation in trace plots, a quick drop-off in autocorrelation, a sufficiently large effective sample size, and similar results across different runs of the chain with different initial states. We also delete a small proportion of the initial states of the chain (before the chain has reached its equilibrium), called the burn-in.

In addition to merely sampling from the posterior distribution of $\theta \mid y$, the output of an MCMC algorithm can also be used to simulate the predictive distribution. That is, the distribution of the next point in the data sequence, conditioned on the observations that came before. Letting y' denote the next outcome, we have

$$\pi(y' \mid y) = \int_\Theta \pi(y' \mid \theta)\pi(\theta \mid y) \, d\theta, \tag{5.3}$$

where $\pi(y' \mid \theta)$ is the single-observation likelihood for y' and $\pi(\theta \mid y)$ is the posterior. After running an MCMC algorithm, we may simulate this distribution by generating a new observation y' (from the observation model) given each sample of θ in the posterior.

Example 4. We return to the data from Table 1.1. In the Bayesian framework, we use the same probit model as in Example 1, but we also impose independent Uniform$[-20, 20]$ priors for α and β. Running the Metropolis–Hastings algorithm with a Gaussian proposal (with scale $\sigma = 0.1$) for $S = 10^7$ iterations and a burn-in rate of 5%, we obtain the posterior densities of α, β and $\xi = -\alpha/\beta$ drawn
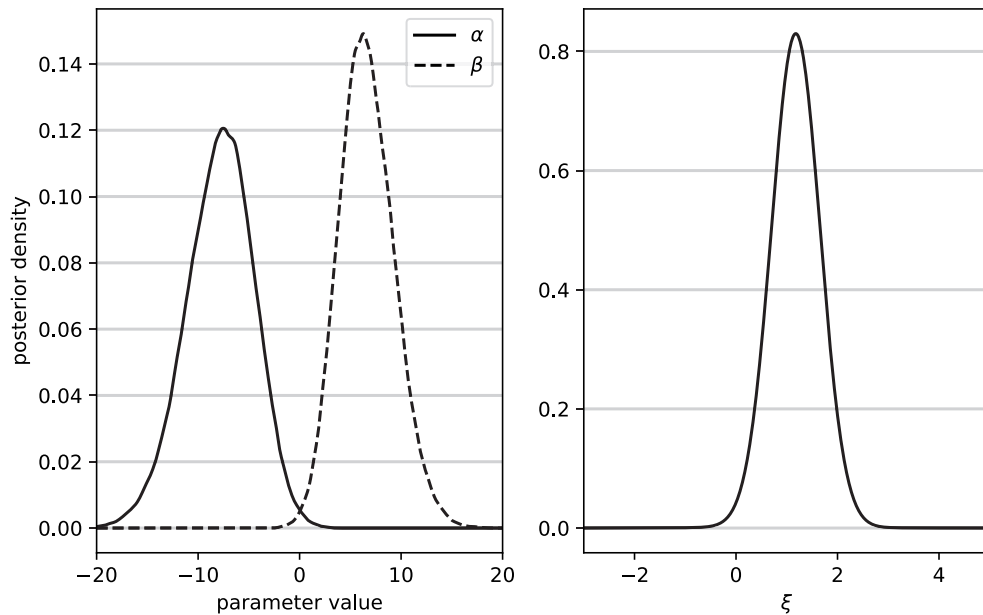
Figure 5.5: Posterior densities of α, β and $\xi = -\alpha/\beta$ given the $n = 30$ observations from Table 1.1.

in Figure 5.5. Standard convergence diagnostics tools indicate that the MCMC algorithm has converged, with an effective sample size of more than 1000 for both α and β. Note how the posterior distributions peak around the maximum likelihood estimates obtained in the frequentist analysis from Chapter 3. In Figure 5.6, we see the predictive mean of the probit model, along with a 95% credibility band, obtained empirically from the MCMC output.

The fact that the distributions in the above example look like normal distributions is no coincidence. In fact, this alludes to one of the major theorems of Bayesian statistics, namely the Bernstein–von Mises theorem. Letting θ_0 denote the true parameter value, this asserts that under mild regularity conditions, we have that

$$\sqrt{n}(\theta - \widehat{\theta}) \,|\, y \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1}),$$

in probability (where "in probability" in this case refers to the distribution of the data sequence $y_1, y_2, \dots$). Hence, the Bernstein–von Mises provides a mirroring story to the asymptotic normality result in the frequentist case that we saw earlier in Chapter 3. Hence, when working within a fixed model, there is asymptotic agreement between the frequentist and the Bayesian statistician, and the prior information is "washed out" as the number of observations increases.
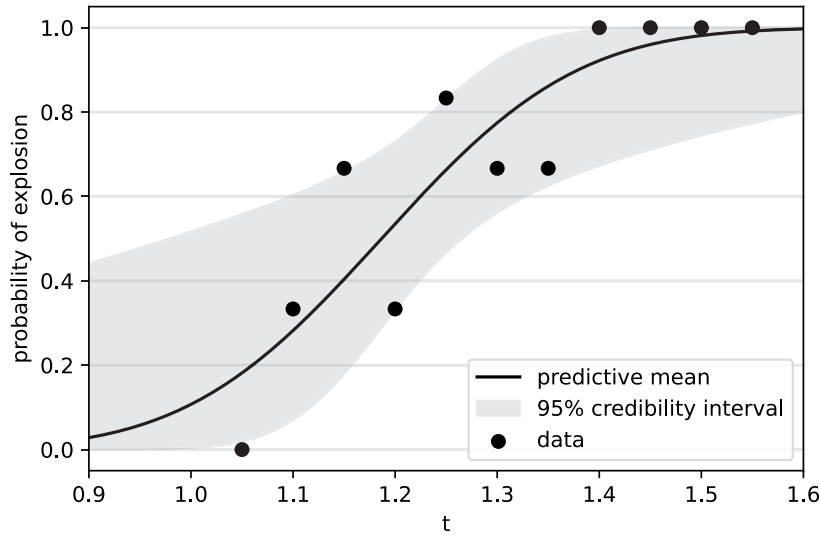
Figure 5.6: The predictive distribution of the next outcome given the $n = 30$ observations from Table 1.1.

5.3 Bayesian model comparison

The Bayesian inferential procedure outlined so far has all taken place within a single model choice. That is, only a single prior $\pi(\theta)$ and observation model $\pi(y \,|\, \theta)$ are under consideration. Within this realm of Bayesian statistics, the Bernstein–von Mises theorem asserts agreement between the frequentist and Bayesian points of view. However, the frequentist and Bayesian stories begin to diverge significantly once we consider multiple models, either for model selection or model averaging. Although it is possible to consider infinite ensembles of models simultaneously, we shall restrict ourselves to the situation in which there are k models under consideration. We introduce a new parameter m, residing in the model space $\mathcal{M} = \{1, \dots, k\}$, deciding which model under which to work. Now, m is an unknown parameter, and is therefore treated as a random variable in the Bayesian framework. This involves choosing prior probabilities $p_j = \mathbb{P}(m = j)$ for $j = 1, \dots, k$. Having chosen a model m, we condition on m in the remainder of the inferential pipeline. That is, the within-model prior takes the form $\pi_j(\theta_j) = \pi(\theta_j \,|\, m = j)$, with $\theta_j \in \Theta_j$, for $j = 1, \dots, k$. Likewise, the within-model likelihood is $\pi_j(y \,|\, \theta_j) = \pi(y \,|\, \theta_j, m = j)$ for $j = 1, \dots, k$. By the law of total probability, the full model-averaged prior for θ is given by

$$\pi(\theta) = \sum_{j=1}^{k} p_j \pi_j(\theta_j),$$

a mixture of the within-model priors. By Bayes' theorem, the posterior probability $p_j^\star$ for model j is given by

$$p_j^\star = \mathbb{P}(m = j \mid y) = \frac{p_j \pi_j(y)}{\pi(y)},$$

where

$$\pi_j(y) = \int_{\Theta_j} \pi_j(\theta) \pi_j(y \mid \theta_j) \, \mathrm{d}\theta_j$$

is the within-model marginal likelihood and

$$\pi(y) = \sum_{j=1}^{k} p_j \pi_j(y)$$

is the full model-averaged marginal likelihood.

The above calculations have a series of interesting consequences. First of all, it illustrates how model selection and/or averaging in the Bayesian framework is mathematically equivalent to two levels of posterior inference. For this reason, MacKay (1992) refers to within-model and across-model inference as the first and second levels of inference, respectively. One can ask what the point of such a distinction would be, as the two levels seem quite similar, purely mathematically speaking. The distinction is useful because unlike in the first level of inference, the second level is fundamentally different in the Bayesian and frequentist perspectives. Indeed, unlike the asymptotic equivalence of maximum likelihood estimation and posterior inference due to the Bernstein-von Mises theorem, it is precisely in tasks such as model comparison and hypothesis testing that there is a critical distinction between Bayesian and frequentist statistics. For instance, in the Bayesian perspective, model comparison and hypothesis testing are mathematically equivalent procedures, and by choosing a uniform prior for the models, $p_j = 1/k$, the statistician can be neutral in the prior. This is not possible in frequentist hypothesis testing, where a null hypothesis must be chosen.

The preceding derivations also show how the dichotomy between model selection and model averaging naturally arises in Bayesian statistics. In terms of the posterior model probabilities, model selection corresponds to choosing the model with the larges value $p_j^\star$, whereas model averaging corresponds to weighting the different models under consideration by their posterior model probabilities $p_j^\star$. The pros and cons of model selection versus model averaging deserves a mention. It is well-established in the machine learning literature that model averaging, usually referred to as ensembling, increases predictive accuracy. That is, a suitably weighted average of multiple trained models nearly always outperforms any single model, in terms of minimising the objective loss. However, this increase in predictive accuracy usually comes at the cost of reducing interpretability. To see why this is the case, consider a simple scenario in which physicists want to know whether there is a linear relationship between two parameters x and y of interest, say $y = ax + b$. They have gathered data $\{(x_i, y_i)\}_{i=1}^{n}$ and want to compare the constant model $y = \mathrm{const}$ (in which there is no linear relationship present) versus the linear model $y = ax + b$. With a neutral prior on these two model choices, the

posterior model probabilities come out as 0.1 and 0.9 for the constant and linear models, respectively. Although the average of these two models can be shown to have better predictive accuracy, it is of little use to the physicists to know that 90% of the time, there is a linear relationship present, and 10% of the time, there is not. In this scenario, the prime motivation is not just to predict future outcomes, but also to interpret the results. Therefore, model selection is more useful, thus asserting a linear relationship, which had the larger posterior model probability, after all. The compromise between accuracy and interpretability is still a vital topic for the interplay between statistics and machine learning and has been ever since it was so pertinently addressed by Breiman (2001).

Finally, the aforementioned derivations tell us that the marginal likelihood is the main ingredient for all Bayesian model comparison. Accurate estimation of the marginal likelihood is therefore a crucial topic in Bayesian statistics. Although particular strategies exist when the mathematical elements involved are sufficiently well-behaved, there is no general recipe. This is in contrast to how MCMC serves as a nearly universal recipe for posterior inference.

Let us conclude this chapter by considering two methods for estimating the marginal likelihood. Both of these appear in **Paper II** as benchmark approaches. The first, called the naive approach, is given by

$$\widehat{\pi}_{\text{naive}}(y) = \frac{1}{S} \sum_{s=1}^{S} \pi(y \,|\, \theta^{(s)}), \tag{5.4}$$

where $\theta^{(1)}, \ldots, \theta^{(S)} \sim \pi(\theta)$, independently. By noting that

$$\mathbb{E}_{\theta}[\pi(y \,|\, \theta)] = \int_{\Theta} \pi(y \,|\, \theta) \pi(\theta) \, \mathrm{d}\theta = \pi(y),$$

we see that $\widehat{\pi}_{\text{naive}}(y)$ is unbiased, and therefore, by the law of large numbers, consistent. However, it is rarely used in practice, as it is relatively unstable, which manifests in practice as a large variance. Its instability can be understood in loose terms by considering how the prior distribution and the likelihood function typically distribute their densities. In most applications, the prior $\pi(\theta)$ is relatively flat, covering a large space. In contrast, the likelihood $\pi(y \,|\, \theta)$ (here thought of as a function of θ), is typically close to zero in nearly all parameter space except for a small region of large contribution. Thus, when sampling independently from the prior, which is the case in the naive estimator, the few samples which land in the region of large likelihood-contribution will tend to dominate the sum in (5.4).

Although many different estimators of the marginal likelihood have been proposed in the literature, we shall consider a simple alternative by Meng and Wong (1996). This particular estimator is a near-optimal choice amongst a family of estimators (with respect to a certain optimisation criterion), called bridge estimators. It takes the form

$$\widehat{\pi}_{\text{bridge}}(y) = \frac{\sum_{s=1}^{S} \pi(y \,|\, \theta^{(s,1)})^{1/2}}{\sum_{s=1}^{S} \pi(y \,|\, \theta^{(s,2)})^{-1/2}}, \tag{5.5}$$

where the samples in the numerator and denominator are drawn independently from the prior $\pi(\theta)$ and the posterior $\pi(\theta \,|\, y)$, respectively. Note that unlike the

naive estimator, $\widehat{\pi}_{\text{bridge}}(y)$ depends on a posterior sample (usually generated using MCMC).

Let us conclude by verifying that $\widehat{\pi}_{\text{bridge}}(y)$ is indeed a sensible estimator for the marginal likelihood. We do this by taking the estimator of the numerator and denominator of (5.5) separately and verify that we recover $\pi(y)$. Expanding the denominator, we have

$$
\begin{aligned}
\mathbb{E}_{\theta \mid y}\left[\pi(y \mid \theta)^{-1/2}\right] &= \int_{\Theta} \pi(y \mid \theta)^{-1/2} \pi(\theta \mid y) \, \mathrm{d}\theta \\
&= \int_{\Theta} \pi(y \mid \theta)^{-1/2} \frac{\pi(y \mid \theta)\pi(\theta)}{\pi(y)} \, \mathrm{d}\theta \\
&= \frac{1}{\pi(y)} \int_{\Theta} \pi(y \mid \theta)^{1/2} \pi(\theta) \, \mathrm{d}\theta \\
&= \frac{1}{\pi(y)} \mathbb{E}_{\theta}\left[\pi(y \mid \theta)^{1/2}\right],
\end{aligned}
$$

as required.

Chapter 6

Bayesian nonparametrics

We now turn to the last of our four approaches to statistical inference under consideration, namely Bayesian nonparametrics. Although there existed antecedent ideas of nonparametric constructions within Bayesian inference already in the 1960s, the subject firmly came to life with the introduction of Dirichlet processes by Ferguson (1973). Before defining Dirichlet processes and examining some of their elementary properties, let us spend a moment thinking through what desiderata we would have for a Bayesian nonparametric process. Suppose we have observed data $y = (y_1, \ldots, y_n)$, and we model these as i.i.d. observations $y_1, \ldots, y_n \,|\, P \sim P$, where P is some probability measure inducing the distribution of the y_i. That is, $\mathbb{P}(y_i \in A) = P(A)$ for all measurable sets A. As Bayesian statisticians, we would like to treat P itself as a random object, and associate a prior process to it. Rather than restricting P to a specific parametric form, as we did in Chapter 5, we would now like to impose a prior process on the entire space of random probability measures. What properties would we desire that such a distribution should have? Ferguson puts forth two criteria, summarised here as follows:

1. **Large support:** The prior process must be able to cover a wide range of possible probability measures. That is, given a certain probability measure Q, the process must, with positive probability, be able to generate measures arbitrarily close to Q (measured by some appropriate metric).

2. **Tractable updating theorems:** The process must have a tractable posterior form. That is, the process for $P \,|\, y$ should be available in closed form.

At first glance, these two desiderata look reasonable. Indeed, if the first criterion is not met, then we might as well restrict ourselves to the parametric regime, which is known to be more efficient than the nonparametric (see e.g. Miller Jr (1983) and the introduction to Doss (1994)). The latter criterion also has merit, since a process is of little use unless we are actually able to carry out posterior inference with it. However, with the advent of modern MCMC algorithms (see Chapter 5), it is possible to simulate posteriors which are not mathematically available in closed form. Hence, the original requirement of explicit posteriors, although highly reasonable in 1973, is arguably too restrictive and has thus *de facto* been

replaced by a weaker requirement, namely that it should be possible to simulate the posterior process. Indeed, in recent papers introducing new nonparametric processes, inference is usually taken care of by means of an MCMC algorithm (Hjort and Walker, 2009; Lijoi et al., 2020).

6.1 Dirichlet processes

Having briefly discussed what properties we would want a nonparametric process to possess, let us shift our attention to the definition of Dirichlet processes. First, recall that the k-dimensional Dirichlet distribution $\mathrm{Dir}(a_1, \ldots, a_k)$ with parameters $a_1, \ldots, a_k > 0$ emits the density

$$f(x_1, \ldots, x_k; a_1, \ldots, a_k) = \frac{\Gamma(a_1 + \cdots + a_k)}{\prod_{j=1}^k \Gamma(a_j)} \prod_{j=1}^k x_j^{a_j - 1}, \quad \text{for } (x_1, \ldots, x_k) \in \Delta_k,$$

where $\Delta_k = \{(x_1, \ldots, x_k) \in \mathbb{R}^k : \sum_{j=1}^k x_j = 1, x_j \geq 0 \text{ for } j = 1, \ldots, k\}$ is the standard k-simplex. Due to the restriction imposed by $\sum_{j=1}^k x_j = 1$, the k-Dirichlet distribution is really determined by the first $k-1$ coordinates $x_1, \ldots, x_{k-1}$, since we can uniquely recover $x_k = 1 - x_1 - \cdots - x_{k-1}$.

We are now ready to define Dirichlet processes. Let $(\Omega, \mathcal{F})$ be a measurable space, let P_0 be a fixed probability measure on $(\Omega, \mathcal{F})$ and let $a > 0$. We say that a random probability measure P is distributed according to a *Dirichlet process with base measure P_0 and concentration parameter a* if, for all finite measurable partitions[1] $A_1, \ldots, A_k$ of Ω, we have that

$$(P(A_1), \ldots, P(A_k)) \sim \mathrm{Dir}(aP_0(A_1), \ldots, aP_0(A_k)). \tag{6.1}$$

When this is the case, we write $P \sim \mathrm{Dir}(aP_0)$. Ferguson (1973) proves that Dirichlet processes indeed exist (meaning, they are uniquely determined by the property (6.1)) and that they have large support.

Let us consider the simplest possible partition $\{A, \Omega \backslash A\}$ with the aim of gaining some intuition as to how the base measure P_0 and the concentration parameter a ought to be interpreted. By (6.1), the random variable $P(A)$ is a random variable following a $\mathrm{Beta}(aP_0(A), a(1 - P_0(A)))$ distribution, so that

$$\mathbb{E}\, P(A) = P_0(A), \quad \mathrm{Var}\, P(A) = \frac{1}{a+1} P_0(A)(1 - P_0(A)),$$

which tells us two things: (i) our prior guess for P is in fact P_0, and (ii) a reflects our confidence in this belief. That is, a large value of a corresponds to high confidence in the prior mean P_0.

In light of the aforementioned criterion 2, Ferguson proves the following result:

---

[1] Recall that a partition of a set is a collection of non-empty disjoint subsets whose union is the whole set.

Theorem 6.1.1. *Let $P \sim \mathrm{Dir}(aP_0)$ and let $y_1, \ldots, y_n \mid P \sim P$, independently. Then*

$$P \mid y_1, \ldots, y_n \sim \mathrm{Dir}(aP_0 + \sum_{i=1}^{n} \delta_{y_i}),$$

where δ_y denotes the degenerate distribution with all its mass at the point y.

This theorem is indeed highly elegant, as it is able to classify the entire behaviour of the posterior process in a single equation. It essentially tells us that the Dirichlet process is the conjugate prior for the nonparametric observation model $y_1, \ldots, y_n \mid P \sim P$. Due to the simplicity of the expression obtained in Theorem 6.1.1, we are able to derive properties of the posterior process straightaway. For instance, it follows directly that

$$\widehat{P}(A) = \mathbb{E}[P(A) \mid y_1, \ldots, y_n] = \frac{a}{a+n} P_0(A) + \frac{1}{a+n} \sum_{i=1}^{n} \mathbb{1}\{y_i \in A\},$$

from which we see that the prior guess P_0 will carry less and less weight as the sample size increases, and the empirical rate $(1/n) \sum_{i=1}^{n} \mathbb{1}\{y_i \in A\}$ will dominate. Further, we have

$$\mathrm{Var}\left(P(A) \mid y_1, \ldots, y_n\right) = \frac{1}{a+n+1} \widehat{P}(A)\left(1 - \widehat{P}(A)\right) \to 0$$

as $n \to \infty$, so that the posterior variance decreases with the sample size.

6.2 Stick-breaking

Before we look at how a nonparametric process such as the Dirichlet process can be used to model a distribution from binary data, we will look at one more fundamental property of the Dirichlet process, namely its stick-breaking representation. Let $\xi_1, \xi_2, \cdots \sim P_0$ independently, let $B_1, B_2, \cdots \sim \mathrm{Beta}(1, a)$ independently and define

$$W_1 = B_1, \quad W_2 = (1 - B_1)B_2, \quad W_3 = (1 - B_1)(1 - B_2)B_3, \ldots$$

Then

$$\sum_{j=1}^{\infty} W_j \delta_{\xi_j} \sim \mathrm{Dir}(aP_0) \quad \text{almost surely.} \tag{6.2}$$

This result has an immediate consequence, namely that if $P \sim \mathrm{Dir}(aP_0)$, then P is almost surely discrete, even if P_0 induces a continuous distribution. Furthermore, stick-breaking has been very useful in the development of simulation techniques for more complicated models based on the Dirichlet process, such as Dirichlet process mixture models, commonly abbreviated DPMMs (Kalli et al., 2009; Walker, 2007). The same holds for situations in which we have censored data, of which binary outcomes are a special case, as we recall from Chapter 4. Doss (1994) derived an MCMC algorithm for posterior inference for a Dirichlet process prior given censored data in which the stick-breaking representation is repeatedly

invoked. Let us briefly go through the mechanisms of this technique. Let $P \sim \mathrm{Dir}(aP_0)$ and $X_1, \ldots, X_n \,|\, P \sim P$ independently. We model the binary outcomes $y_1, \ldots, y_n$ given inputs $t_1, \ldots, t_n$ as realisations of binary indicators of the form $Y_i = \mathbb{1}\{X_i \leq t_i\}$. From our discussion regarding the connection between binary outcomes and censored data from Chapter 4, we know that this is equivalent to not observing the X_i directly, but only that $X_i \in A_i$, where $A_i = (-\infty, t_i]$ if $y_i = 1$ and $A_i = (t_i, \infty)$ if $y_i = 0$. Hence, posterior inference amounts to simulating the process $P \,|\, \{X_1 \in A_1, \ldots, X_n \in A_n\}$, rather than the standard form $P \,|\, \{X_1 = x_1, \ldots, X_n = x_n\}$. The algorithm by Doss (1994) achieves this by alternating between the two following steps:

1. Sample $X_1, \ldots, X_n \,|\, P \sim P$ independently, conditioned on $X_1 \in A_1, \ldots, X_n \in A_n$,

2. Sample $P \,|\, X_1, \ldots, X_n \sim \mathrm{Dir}(aP_0 + \sum_{i=1}^n X_i)$,

The stick-breaking representation is used in the second step. In practice, it is not possible to simulate infinitely many variables $B_1, B_2, \ldots$ or $\xi_1, \xi_2, \ldots$, and therefore, for most applications, we have to use a finite truncation of the stick breaking representation, say

$$P \approx \sum_{j=1}^K \tilde{W}_j \delta_{\xi_j},$$

for some large number K, where $\tilde{W}_j = W_j / \sum_{j=1}^K W_j$. How large K should be chosen depends on the problem at hand, but naturally, we want that $\sum_{j=1}^K W_j$ is close to 1. However, by cleverly treating K as a random variable, Doss (1994) is able to carry out the above steps exactly, with no approximations.

Example 5. Let us return to the dataset from Table 1.1 one final time where we impose a Dirichlet process prior on the sensitivity distribution. The base measure is chosen to be uniform between the minimum and maximum of the t_i, i.e. $\mathrm{Uniform}[1.05, 1.55]$, and the concentration parameter is $a = 1.0$. Using the successive substitution sampling algorithm by Doss (1994) with $M = 500$ iterations, we obtain the posterior distribution drawn in Figure 6.7. Here, we see both the prior and posterior means and realisations from both processes. By inspection, we see that the posterior variance is reduced.

6.3 Model comparison in Bayesian nonparametrics

We conclude this section, and thus the entire exploration of approaches to statistical inference for binary data models, with a discussion about computation of marginal likelihoods in Bayesian nonparametrics. Let proc denote a nonparametric process for generating a random probability distribution (e.g. a Dirichlet process). To avoid too much generality, let us restrict ourselves to cases like binary response models, where the marginal likelihood takes the form of a probability, $\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n)$, where $X_1, \ldots, X_n \,|\, P \sim P$ independently and $P \sim$ proc. We will first look at how this can be achieved when the generating process proc yields a
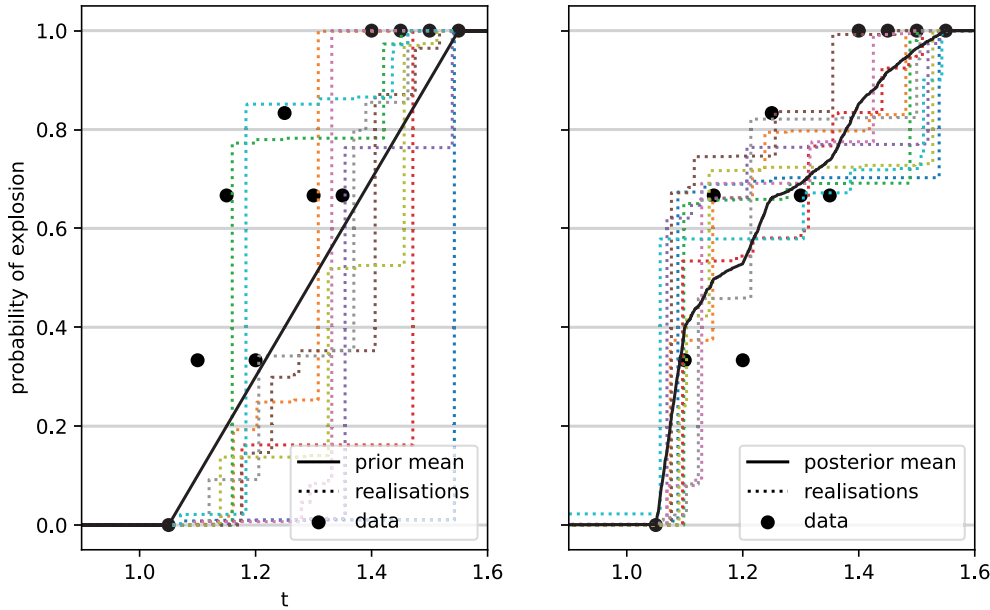
Figure 6.7: Prior and posterior Dirichlet processes based on the $n = 30$ observations from Table 1.1.

stick-breaking representation for P, like we saw for the Dirichlet process. Although the following construction essentially extends to any process with a stick-breaking representation (Gil-Leyva and Mena, 2023; Hjort, 1990; Paisley et al., 2010), we shall focus on the expression (6.2). To employ this in practice, we have to use a finite truncation as mentioned above, say $P \approx \sum_{j=1}^{K} \tilde{W}_j \delta_{\xi_j}$. Employing this truncation and conditioning on P, we have

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n \mid P) = \prod_{i=1}^{n} P(A_i) = \prod_{i=1}^{n} \sum_{j:\xi_j \in A_i} \tilde{W}_j.$$

Hence, we obtain the following estimator for the marginal likelihood,

$$\widehat{\mathbb{P}}_{\text{stick}} = \frac{1}{S} \sum_{s=1}^{S} \prod_{i=1}^{n} \sum_{j:\xi_j^{(s)} \in A_i} \tilde{W}_j^{(s)}, \tag{6.3}$$

where $\xi_j^{(s)}$ and $\tilde{W}_j^{(s)}$ are generated as described above for ξ_j and $\tilde{W}_j$, independently for $s = 1, \ldots, S$.

Although (6.3) is a useful alternative when stick-breaking is available, it cannot be applied when P does not have such a representation. Let us finally briefly discuss the challenge of estimating marginal likelihoods for generic nonparametric processes. Since we wish to estimate a probability, namely $\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n)$, where $X_i \mid P \sim P$ independently for $i = 1, \ldots, n$ and $P \sim$ proc, the first

and most naive approach would be to estimate it simply using frequencies of these events,

$$\widehat{\mathbb{P}} = \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}\{X_1^{(s)} \in A_1, \ldots, X_1^{(s)} \in A_n\},$$

where $X_1^{(s)}, \ldots, X_n^{(s)} \mid P^{(s)} \sim P^{(s)}$ independently and $P^{(s)} \sim \text{proc}$, independently for $s = 1, \ldots, S$. Although this estimator unbiased and strongly consistent, it is unfortunately useless in practice. This is because even for moderately large n (≥ 50, say), we will never experience that $X_1^{(s)} \in A_1, \ldots, X_n^{(s)} \in A_n$ when computing $\widehat{\mathbb{P}}$. This estimator will thus always return zero. The main topic of **Paper I** and **Paper II** is to remedy this problem by exploiting that when $X_1, \ldots, X_n \mid P \sim P$, the X_i are *exchangeable*, meaning that the distribution of the vector $(X_1, \ldots, X_n)$ is invariant under permuting its components. That is, the distribution of $X_1, \ldots, X_n$ does not depend on the ordering of the sequence $X_1, \ldots, X_n$. This allows us to loosen the criterion $X_1^{(s)} \in A_1, \ldots, X_n^{(s)} \in A_n$, effectively turning the rejection sampling algorithm yielding $\widehat{\mathbb{P}}$ into an importance sampling algorithm, where the importance weights are computed by counting a certain number of permutations.

Chapter 7

Paper summaries

7.1 Paper I

This paper's primary concern is the problem of estimating the marginal likelihood of a Bayesian nonparametric model given binary response data. In the article, we introduce the idea of *permutation counting*, which is a new importance sampling technique specifically aimed at this task. Let $P \sim$ proc be a random probability distribution, modelled to generate the latent random variables, i.e. $X_1, \ldots, X_n \mid P \sim P$, independently. As we saw in the end of Chapter 6, it is, in general, not possible to estimate the marginal likelihood simply by generating from the observation model and counting frequencies of events. However, by exploiting the symmetry introduced by exchangeability of the sequence $X_1, X_2, \ldots,$ this approach can be modified to yield an estimator that actually works in practice. In the notation of Chapter 6, let $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ and let[1] $A = A_1 \times \cdots \times A_n \subset \mathbb{R}^n$. Letting S_n denote the group[2] of n-permutations, define the *permutation number* of X with respect to A by

$$w(X; A) = \#\{\sigma \in S_n : \sigma(X) \in A\},$$

where $\sigma(X) = (X_{\sigma(1)}, \ldots, X_{\sigma(n)})$. That is, $w(X; A)$ is the number of permutations σ with the property that when we permute the components of X according to σ, we end up in the set A. In the paper, it is shown that $\mathbb{E}\, w(X; A) = n! \times \mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n)$, thus yielding the modified estimator

$$\widehat{\mathbb{P}}_{\text{perm}} = \frac{1}{S} \sum_{s=1}^{S} \frac{1}{n!} w(X^{(s)}, A), \tag{7.1}$$

where $X_1^{(s)}, \ldots, X_n^{(s)} \mid P^{(s)}$ independently and $P^{(s)} \sim$ proc, independently for $s = 1, \ldots, S$. Note that this modified estimator only requires that it is possible to

---

[1]In **Paper I**, the symbol $\mathcal{B}$ is used rather than A.

[2]Recall that an n-permutation is a bijection from the set $\{1, \ldots, n\}$ to itself, and that the set of such permutations forms a group under functional composition.

generate a sample $X_1, \ldots, X_n \,|\, P \sim P$. In particular, it does not rely on properties of the Dirichlet process or require a stick-breaking representation. It is proved that the variance of $\widehat{\mathbb{P}}_{\text{perm}}$ is smaller than that of the naive frequency counting estimator.

Via simulations and real data problems, it is, perhaps even more importantly, illustrated that $\widehat{\mathbb{P}}_{\text{perm}}$ works in practice. Namely, even for large values of n, there is a sufficiently large proportion of contributing terms in (7.1) and that the importance weights $w(X^{(s)}; A)/n!$ are reasonably balanced, so that the estimator is able to obtain accurate results with a large effective sample size.

Regarding actually computing permutation numbers, it is shown that these can be represented as permanents of a certain class of binary matrices. Although the problem of computing permanents is in general #P-complete, it is proved that the permanents arising in the specific context of the paper are computable in polynomial time. This is done by exhibiting an explicit algorithm, an implementation of which is available at the paper's accompanying Github repository.

The new estimator is applied in two problems: the first is a nonparametric bioassay benchmark problem with a Dirichlet process prior. Here, other methods, namely those explained in Chapter 6, are available, and it is shown that permutation counting performs as well as existing methods. In the second problem, real rubella seroprevalence data are analysed using a Pitman–Yor multinomial process.

7.2   Paper II

Christensen, D.; Moen, P. A. J. **perms**: Likelihood-free estimation of marginal likelihoods for binary response data in Python and R. *Submitted for publication*, 2024.

This paper presents a software implementation of permutation counting, introduced in **Paper I**. The software, titled perms, is written in the programming language C and implemented both as an R package and a Python library. The paper goes through the extensive optimisation efforts made in the writing of the software, and how this translates to efficiency in practice. Unlike the computer code used for the examples in **Paper I**, perms is able to handle datasets of magnitude up to a few thousand, rather than a few hundred. In addition, perms is easier to use, with minimal effort required by the user of the software.

After going through the idea of permutation counting and the importance sampling technique introduced in **Paper I**, the article shows how perms is used in practice by considering a tractable toy problem. This is done independently in both R and Python, with replicable code given for both. The paper then moves on to more realistic applications. First, the marginal likelihood of a Bayesian multivariate logistic regression model is estimated using perms, and its performance is compared with existing methods. Next, a larger data problem is considered, in which the marginal likelihood of a nonparametric bioassay model with a Dirichlet process prior is estimated using perms and stick-breaking as

outlined in Chapter 6. Although perms requires neither a tractable likelihood function nor a stick-breaking representation, these examples are included to compare the accuracy and computational cost of perms with alternative methods. In both problems, perms achieves the same accuracy as the state of the art alternatives with the same number of samples. However, perms is slower, which alludes to a natural trade-off between flexibility and computational cost.

Next, the computational details of the software are presented, and the techniques used to increase computational speed are summarised. In particular, it is explained how perms exploits sparsity of the data structures arising in the computation of permutation numbers. A simulation study comparing perms with the Python implementation accompanying **Paper I** is presented, where it is shown that perms is substantially faster. This has the consequence that perms is able to handle more complicated problems with larger datasets, and an example of such a problem, namely a novel Bayesian nonparametric changepoint analysis task, is solved using perms in the conclusion of the paper.

7.3 Paper III

Jensen, T. L.; Moxnes, J. F.; Unneberg, E.; Christensen, D. Models for predicting impact sensitivity of energetic materials based on the trigger linkage hypothesis and Arrhenius kinetics. *Journal of Molecular Modeling* **26**, 2020. DOI: 10.1007/s00894-019-4269-z

This paper considers models for predicting impact sensitivity of energetic materials from quantum chemical properties. Under the trigger linkage hypothesis, which asserts that the decomposition of an energetic material is initiated by breaking the weakest bond. In the compounds considered in the article, this weakest bond is an $A-NO_2$ bond, where A denotes either a carbon, a nitrogen or an oxygen atom. By Arrhenius' theory of kinetics, the rate of reaction r is then related to the activation energy E_a (i.e. the energy required to convert the reactant to the transition state) by

$$r = c[A - NO_2]^n \exp\left(-\frac{E_a}{RT}\right),$$

where $[A - NO_2]$ is the molar concentration of $A - NO_2$, c is a proportionality constant (the pre-exponential factor), n is the reaction order, R is the molar gas constant and T is the absolute temperature. Motivated by the fact that r has been documented to correlate with impact sensitivity, and, working on a log scale, we model[3] ξ and $\log r$ to be inversely proportional. Here, ξ is the median of the sensitivity curve (on a log scale), commonly referred to as h_{50} in the literature. In doing so, we obtain the basis for the models under consideration in the paper, namely

$$\xi = a + b\frac{E_a}{T},$$

where a and b are the parameters of the model, to be estimated from data. Unfortunately, it is both difficult and time consuming to calculate E_a and T

---

[3]In the notation used in the article, we have $\xi = \log I$.

directly. Therefore, the article considers a selection of techniques of approximating these.

The most successful approach is to approximate E_a by the bond dissociation energy (BDE) and T by the detonation temperature T_{ex}. Here, the BDE is the energy required to break the trigger-linkage. It is worth noting that the BDE and E_a will coincide if the energy increases monotonically as the bond is broken.

Three families of molecules are considered in the paper: nitroaromatics, nitramines and nitrate esters. For the former family, the aforementioned model $\xi = a + b(\text{BDE}/T_{ex})$ gives a good fit, with a coefficient of determination (R^2) of 0.81. The model is used to predict the impact sensitivity of 1,3,5-triamino-2,4,6-trinitrobenzene (TATB), whose impact sensitivity is only bounded above in the existing literature. For nitramines, the fit is not as good, alluding to the fact that other parameters than BDE and temperature of detonation are required to predict impact sensitivity accurately for these compounds. For the nitrate esters, we get a very good fit, but this analysis is only based on seven data points and is therefore of limited value.

All regression analysis in the paper is done twice: both in a frequentist and Bayesian framework. This is of particular use in the aforementioned nitrate esters regression, in which the frequentist analysis yields a significant linear relationship ($p = 0.003$), but when Bayesian model comparison is employed, a constant model is preferred. This illustrates how model complexity is automatically penalised in the Bayesian framework, a phenomenon sometimes called the Occam's razor of Bayesian inference.

7.4   Paper IV

Christensen, D.; Unneberg, E.; Høyheim, E.; Jensen, T. L.; Hjort, N. L. Improved measurements of impact sensitivities of energetic materials. In *Proceedings of the 25th International Seminar on New Trends in Research of Energetic Materials (NTREM)*, Institute of Energetic Materials, University of Pardubice, Czechia, 2023.

The main issue this paper addresses is how one should construct confidence intervals (CIs) for quantiles of an impact sensitivity curve under the probit model when employing the Bruceton design. The three main alternatives are via the delta method, via the likelihood ratio test and via Fieller's theorem, as explained in Chapter 3. This question is dealt with via simulations, where the step size is varied and different values of the sample size n are considered. The quality of the three aforementioned approaches to constructing CIs are compared by means of their empirical coverage probabilities. The simulations consistently show that the delta method is the poorest performing method, and that Fieller's theorem yields the best performance overall.

In the simulations conducted in the paper, CIs for both $h_{50} = \xi$ and h_{99} are considered. Comparing the results achieved with a sample size of $n = 30$ and $n = 100$, it is concluded that $n = 30$ is insufficient to achieve any useful results for h_{99}. It is, however, concluded that $n = 30$ measurements suffice if one is

only interested in constructing a CI for h_{50}, as these perform rather well in the simulations. It is important to keep in mind that all of these results assume that the probit model is correct.

In addition to computing CIs, the article introduces the idea of confidence curves (CCs), which were discussed in Section 3.2. Based on experiments conducted at FFI, the CCs for h_{50} and h_{99} for cyclotetramethylene-tetranitramine (HMX) are drawn. These plots are rather instructive, as they explicitly show how Fieller's theorem can yield non-central CIs, and that the CIs for h_{99} are much wider than those for h_{50}.

7.5 Paper V

Christensen, D.; Stoltenberg, E. A.; Hjort, N. L. Sequential experimental designs in regression: Theory for the Bruceton and Langlie designs. *Submitted for publication*, 2023.

Motivated by the simulations conducted in **Paper IV**, this article formally verifies asymptotic normality of the maximum likelihood estimator when a sequential design like the Bruceton design is employed. When this is the case, the covariate sequence[4] $X_1, X_2, \ldots$ is a stochastic process, and the standard theory for i.i.d. regression does not apply. Letting $f_\theta(y \mid x)$ denote the regression model, parametrised by θ, we assume that the covariate sequence $X_1, X_2, \ldots$ is chosen via a sequential experimental design for the form

$$X_i = h(\{X_j, Y_j\}_{1 \leq j \leq i-1}),$$

so that the ith measurement X_i is allowed to depend on the data obtained up to that point. The asymptotic properties of one such design, namely the famous Robbins–Monro procedure, has been studied extensively, but nearly all the results for this design rely the covariate sequence converging in probability to the target of interest. In the Bruceton design and many other designs alike, this is not the case, and so a separate theory must be developed.

Letting θ_0 denote the true parameter value, the sequential design implies that the variables

$$J_n = -\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial^2 \log f_{\theta_0}(Y_i \mid X_i)}{\partial \theta_0 \partial \theta_0^\top} \,\middle|\, X_i \right]$$

are not i.i.d. The paper therefore uses martingale theory to establish a large-sample result of the form

$$\sqrt{n}(\theta - \theta_0) \xrightarrow{d} \mathrm{N}(0, J^{-1}),$$

where J is the probability limit of J_n. This theorem requires the sequential design function h to be chosen such that the sequence J_n behaves sufficiently well. In particular, we require that the process J_n converges in probability to an invertible matrix J.

---

[4]We use the notation of the paper here, where $X_1, X_2, \ldots$ denotes the sequence of covariates. This is different from the notation of the preceding chapters, in which the notation $t_1, t_2, \ldots$ was used.

The second half of the article is to verify explicitly that J_n converges in the Bruceton design and in a Markovian version of the Langlie design. By appealing to Markov chain theory, it is proved that the process $X_1, X_2, \ldots$ is *ergodic*, which in this sense implies that any empirical average of the form $(1/n) \sum_{i=1}^{n} g(X_i)$ converges in probability, provided g is bounded. In particular, this means that J_n converges in probability in the cases of logistic and probit regression. For the Bruceton design, this result is proved straight forwardly via discrete state space Markov chain theory (namely via Foster's theorem). For the Markovian version of the Langlie design, the kernel is chosen to be continuous to ensure irreducibility of the chain, and so arguments from continuous state-space Markov chain theory are applied. More specifically, ergodicity of the chain is proved by establishing Harris recurrence, which again is verified by exhibiting a petite recurrent set. A key component of the proof is to appeal to compactness of the state space $[0, 1]$ of the Langlie design.

7.6 Paper VI

Christensen, D.; Novik, G. P.; Unneberg, E. Estimating sensitivity with the Bruceton method: Setting the record straight. *Submitted for publication*, 2024.

The main purpose of this paper is to clear up many of the misconceptions persisting in the energetic materials industry regarding how sensitivity estimates are to be obtained when employing the Bruceton design. The source of this confusion stems from the fact that in the original article by Dixon and Mood (1948) in which the Bruceton design is first introduced, an approximation scheme for computing the maximum likelihood estimates is also proposed. Only requiring pen-and-paper calculations, this approximation was indeed useful in 1948. However, as the Bruceton design has been adapted by NATO for research involving sensitivity testing, so has the approximation been adapted, even though nowadays, the maximum likelihood estimates can easily be computed using standard spreadsheet software like Microsoft Excel or virtually any general purpose programming language, like R or Python. As Dixon and Mood (1948) explain, their approximation depends on certain numerical criteria being satisfied by the data obtained, and its use should be avoided when these are not satisfied. This has, however, led to the unfortunate situation that many researchers believe their results are somehow "not valid", when really it is only the use of the 1948 approximation that is invalid. Thus, perfectly useful datasets (for which the maximum likelihood estimates exist and can be computed) are often discarded.

To examine how often the approximation does not apply even though the maximum likelihood estimates exist, a simulation study is performed. Here, it is found that a substantial proportion of the datasets generated are deemed "not valid" by said approximation, even in cases where the maximum likelihood estimates exist for nearly 100% of these datasets. In particular, a small choice of the step size d seems to decrease the probability of the approximation being valid, even though there is no such negative effect on the existence of the maximum likelihood estimates.

To illustrate further the information lost by applying the 1948 approximation, five datasets from **Paper VII** are analysed using it. All of them are deemed as "not valid" and should be, according to NATO's guidelines (called STANAGs) on the matter, discarded. It is worth pointing out that some of these datasets contain as many as $n = 70$ measurements.

The paper also criticises that there are no instructions on how to create confidence intervals in the STANAGs, and consequently, how little attention is paid to the construction of confidence intervals in the energetic materials literature. The paper summarises the theoretical foundations laid in **Paper V** and the simulation studies from **Paper IV**, providing advice on how researchers should construct confidence intervals in sensitivity experiments.

7.7 Paper VII

This paper reports the results of experiments measuring the impact sensitivity of amatol extracted from explosive remnants of war. A mixture of TNT and ammonium nitrate, amatol was developed by the British during World War I in response to a shortage of shortage of munition shells and high explosives, an event called the shell crisis of 1915. Due to its many desirable properties as a high explosive and low production cost, amatol was rapidly adapted as an explosive filling and was widely used right up until the very end of World War II, when TNT became plentiful. As a result of having been used as a substitute for TNT by multiple countries for several decades, excessive supply of amatol can be found today at ammunition dumping sites and in explosive remnants of war.

All samples of amatol analysed in the paper were collected in Finnmark county during national explosive ordnance disposal clearance operations. They are all of German origin and were produced before 1945. The key aim of the study conducted in the paper is to investigate whether the sensitivity properties of the samples collected were altered as a result of being exposed to nature for nearly 80 years. If no alteration had occurred, we would expect the impact sensitivity to be the same as what is recorded for amatol in the literature. That is, similar to or slightly less sensitive than TNT, for which it is well documented that[5] h_{50} is circa 30 Joules.

To measure the impact sensitivity of the samples, fallhammer experiments were conducted in FFI's laboratories. The data obtained were analysed using the methods outlined in Chapter 3, namely with a probit model on a log scale, and with confidence intervals computed via Fieller's theorem. The analysis shows clearly that only one out of the five samples coincided with the impact sensitivity of TNT. The remaining samples were significantly more sensitive, with their h_{50}

---

[5]Here, the value h_{50} is measured in energy, not on a log scale.

values ranging between 7.5 and 15.4 Joules. In the course of the experiments, explosions were observed with impact as low as 6.2 Joules.

Although the paper does not address precisely why the substances analysed have become so more sensitive to impact over time, it firmly demonstrates that this is the case and that the standard references for TNT are of no use for amatol extracted from explosive remnants of war. This in turn ought to have an affect on risk assessment and policy making regarding whether further action should be taken to locate and destroy such materials. In particular, explosive ordnance disposal operators should take the findings of the paper into account when encountering ordnance potentially containing amatol.

7.8 Paper VIII

Nygård, S. T.; Christensen, D. Unlevel playing field: Evidence of ethnic discrimination in the access to children's football from a field experiment in Norway. *Submitted for publication*, 2023.

This paper reports the findings of a field experiment whose purpose was to investigate whether people from ethnic minorities are discriminated against in the access to participation in children's sports. In the experiment, $n = 949$ emails were sent out to local children's football clubs all across Norway. Each email appeared to be from a parent asking about his or her ten-year-old child participating in a trial practice session. The emails only differed in whether (i) the sender appeared male or female, (ii) the child was referred to as male or female and (iii) the sender's name signalled membership of an ethnic majority or minority. The ethnic majority names were chosen to be common Norwegian names, whereas for the minority group, common Somali and Pakistani names were chosen. All eight combinations of the above three dichotomous choices were represented equally in the pool of emails sent. Apart from these key differences, the textual content of the emails were identical. The number of emails receiving a reply within two weeks were then counted in each group.

Although this application is rather different to that of the previous seven papers, many of the statistical tools are in fact quite similar. Indeed, note that the indicator of whether an email received a reply is binary, and therefore all the tools from sensitivity testing are largely applicable. The responses are all modelled using standard logistic regression. To avoid issues with multiplicative interaction terms (Braumoeller, 2004), regression is performed separately for the minority and majority groups. Four models are considered, depending on whether the regression should include covariates, and whether the regression should be performed separately for each county. The best model (decided by comparing the Akaike information criterion (AIC)) includes the covariates, with a single regression model for all counties. The response rates are estimated using maximum likelihood theory, and confidence curves are drawn (using likelihood profiling) for the majority response rate, the minority response rate and the difference of these two.

The results show that the sex of the sender or their child has no significant effect on the response rate, but that there is a significant difference ($p < 0.001$) between the response rates for the ethnic majority and minority groups, standing at 84.24% and 70.19%, respectively. In addition to the above three features, the proportion of non-Western immigrants living in the municipality of the football club is also included as a covariate, based on previous studies indicating that this may have an effect on the response rates in this kind of field experiment. This proportion is found to correlate positively ($p = 0.006$) with the response rate of the minority group, but not with that of the majority group. This coincides with the ideas of contact theory, in which it is postulated that an increase in ethnic diversity leads to less prejudice.

Chapter 7. Paper summaries

Chapter 8

Discussion

In this chapter, we elaborate further on some of the key topics of the papers listed. In particular, we outline possibilities of further research. This discussion is not meant to be exhaustive, but rather to explore ideas of further development of some selected key concepts introduced.

8.1 Permutation counting

As the main topic of both **Paper I** and **Paper II**, permutation counting is one of the chief themes of the present thesis. Although the technique is applied in a multitude of different problems in both papers, there are various directions of further study in which potential improvements of said technique can be made. Firstly, it is worth exploring whether permutation counting can be applied to other types of censored data, such as interval censoring or even perhaps more general forms of incomplete observations. Although it is mentioned at the end of **Paper I** that the algorithm for computing the permutation numbers $w(X; A)$ does not extend to cases with interval censoring, it might be possible to approximate them, rather than computing them exactly. As permutation numbers can be realised as permanents of binary matrices, a natural first step to take would be to apply some of the approximation schemes for permanents which have been developed over the last preceding decades (Chertkov and Yedidia, 2013; Jerrum and Sinclair, 1989; Karmarkar et al., 1993). However, as permutation counting requires thousands upon thousands of such permanents to be computed efficiently, it is not certain that all these methods will be sufficiently fast. Fortunately, permutation counting is entirely parallelisable, which helps tremendously.

Secondly, note that the importance sampler (7.1) yielded by permutation counting naively samples from the prior as its proposal distribution at each iteration. This is likely to be quite wasteful, and further research into combining permutation counting with cleverer proposal distributions is encouraged. In particular, sequential Monte Carlo techniques (see Cappé et al. (2007) for an overview) may be particularly useful for this purpose, as they have been for approximate Bayesian computation (ABC, see Marin et al. (2012) for a useful review). Tools like these may open permutation counting to further domains, such as problems with high-dimensional data. Although a simple example with

multidimensional features was analysed in **Paper II**, it is improbable that the naive sampler (7.1) is able to handle much more complicated problems.

Finally, it remains to be seen whether the idea of exploiting symmetries introduced by exchangeability can be used in other areas of statistics. Indeed, exchangeability is a standard assumption in many models, also ones without censoring. Appealing to permutations of exchangeable data sequences can sometimes lead to fruitful nonparametric tools, of which the permutation test is a key example. The question remains whether there are other applications where we simply do not benefit from permuting our observations, but also from *counting* in how many ways this can be done in order to satisfy a certain criterion.

8.2 Predicting sensitivity

The main topic of **Paper III**, namely to predict the sensitivity of energetic materials purely from their quantum chemical properties, is a rich area of study with many successful efforts (see the references in **Paper III**). In order to improve on the model considered in **Paper III**, namely $\xi = a + b(\mathrm{BDE}/T_{\mathrm{ex}})$, various directions of future study are worth exploring. To give one example, it is reasonable to hypothesise that the above model works better for nitroaromatics than nitramines due to the former group being a more homogeneous set of compounds than the latter, as can be seen in Appendix B. Indeed, by definition, at least one of the nitro groups of a nitroaromatic compound is attached to a benzene ring, whereas this is not the case for nitramines (in fact, there need not be an aromatic ring present at all in a nitramine). Furthermore, it has indeed been documented that the BDE serves as an accurate approximation of the activation energy for nitroaromatic compounds (Khrapkovskii et al., 2013). However, for nitramines, the situation is typically more complicated, for instance due to multiple decomposition routes (Samseth, 2022).

Furthermore, a complete description of the explanatory factors of a compound's sensitivity properties is most certainly required to contain parameters beyond the molecular level. Indeed, as cited in **Paper III**, higher level properties such as crystal morphology and crystal defects are known to affect sensitivity. When building models which take these into account, a key challenge is how to combine them with the purely quantum chemical properties of the molecules themselves.

8.3 Sequential experimental designs

Motivated by the simulation study presented in **Paper IV**, the asymptotic properties of the probit model under sequential designs are verified in **Paper V**. This kind of work can naturally be extended to the study of other sequential designs than the Bruceton design, of which there exist multiple examples (Joseph, 2004; Neyer, 1994; Wu and Tian, 2014).

Arguably, the most well studied such sequential procedure is the Robbins–Monro design, and there are many lessons to be learned from this work when working with other designs (see Lai (2003) for an excellent review). In particular,

in designs in which the covariate sequence converges to the point of interest (which, in the case of binary responses, is the median of the distribution), care must be taken to make sure that the sequence does not converge too quickly. Indeed, this is the main point raised at the discussion of **Paper V**. This makes intuitive sense in the case of one-dimensional linear regression, where there are two parameters: the intercept α and the linear coefficient β. If all the measurements are made at the same input, i.e. the covariate sequence is constant, then it is not possible to estimate α and β simultaneously from data. The same problem arises if the covariate sequence converges too quickly to a single point. As several experimental designs in the sensitivity testing literature aim to converge to the median ξ, verifying their large sample properties is of crucial importance for the study of sensitivity testing.

8.4 Measuring sensitivity in practice

A considerable proportion of both **Paper IV**, **VI** and **VII** is devoted to improvements to the current recommendations for measuring sensitivities of energetic materials. In particular, **Paper VI** explains why the current protocols from NATO and the U.S. Department of Defense are outdated, and that the approximation introduced by Dixon and Mood (1948) should be replaced by plain maximum likelihood estimation. This issue is also raised in **Paper VII**, since the approximation would have deemed all datasets "not valid" even though the maximum likelihood estimates existed and confidence intervals could be derived without issue.

It is important here to realise that **Paper VI** does not call for discarding the Bruceton method altogether. Indeed, the simulation studies and theoretical developments from **Paper IV** and **Paper V** suggest that the up-and-down experimental design is well-founded. The design is also appealing in the effect of it being model-free, i.e. the design does not depend on the choice of statistical model. In contrast, most of the more sophisticated designs that have been developed since 1948, like the Neyer design and the 3pod, rely on model assumptions (Neyer, 1994; Wu and Tian, 2014). Simulation studies or theoretical results establishing how well these perform compared to the Bruceton design under model misspecification is encouraged. If the researcher has little knowledge beforehand of whether the probit model is going to yield a good fit, it is perhaps not advisable to employ an experimental design based on explicit model assumptions. In cases like these, it is worth exploring whether an entirely nonparametric approach would yield better results. This opens the door for new research questions to be asked. For instance, suppose we combine the Bruceton up-and-down design with the nonparametric maximum likelihood estimator (see Chapter 4). If we wanted to construct confidence intervals for the median $\xi = F^{-1}(1/2)$ in this case, we would first need to study the asymptotic properties of the estimator $\widehat{F}^{-1}(1/2)$ specifically for the Bruceton design. That is, we may not apply the results of Groeneboom and Wellner (1992) or Bhattacharya and Kong (2007) directly, as these assume that a different experimental design is being used. Interestingly, the up-and-down design is also widely employed in bioassay experiments and medicine, but here

it is uncommon to impose a parametric model and to estimate the median ξ via maximum likelihood estimation. Rather, the median is estimated directly from the covariate sequence, with the justification that the maximum likelihood estimator may not always exist (Derman, 1957; Durham and Flournoy, 1994; Flournoy, 2002). It would be useful to compare this approach with that of maximum likelihood estimation.

Finally, it is worth mentioning that although the main focus of the papers in this thesis has been on impact sensitivity, there are similar questions to ask for other types of sensitivity, such as friction or shock.

Bibliography

Abdelbasit, K. M. and R. L. Plackett (1983). 'Experimental design for binary data'. In: *Journal of the American Statistical Association* 78, pp. 90–98. DOI: 10.1080/01621459.1983.10477936.

Anderson, J. and S. Howe (1972). *Close to the Edge.* Atlantic Recording Corporation.

Ayer, M. et al. (1955). 'An empirical distribution function for sampling with incomplete information'. In: *The Annals of Mathematical Statistics* 26, pp. 641–647. DOI: 10.1214/aoms/1177728423.

Bhattacharya, R. and M. Kong (2007). 'Consistency and asymptotic normality of the estimated effective doses in bioassay'. In: *Journal of Statistical Planning and Inference* 137, pp. 643–658. DOI: 10.1016/j.jspi.2006.06.027.

Braumoeller, B. F. (2004). 'Hypothesis testing and multiplicative interaction terms'. In: *International Organization* 58, pp. 807–820. DOI: 10.1017/S0020818304040251.

Breiman, L. (2001). 'Statistical modeling: The two cultures (with comments and a rejoinder by the author)'. In: *Statistical Science* 16, pp. 199–231. DOI: 10.1214/ss/1009213726.

Cappé, O., S. J. Godsill and E. Moulines (2007). 'An overview of existing methods and recend advances in sequential Monte Carlo'. In: *Proceedings of the International Conference on Grey Systems and Intelligent Services.* Vol. 95. IEEE, pp. 899–924.

Chertkov, M. and A. B. Yedidia (2013). 'Approximating the permanent with fractional belief propagation'. In: *Journal of Machine Learning Research* 14, pp. 2029–2066.

Cox, C. (1990). 'Fieller's theorem, the likelihood and the delta method'. In: *Biometrics* 46, pp. 709–817. DOI: 10.2307/2532090.

Derman, C. (1957). 'Non-parametric up-and-down experimentation'. In: *The Annals of Mathematical Statistics* 28, pp. 795–798. DOI: 10.1214/aoms/1177706895.

Dixon, W. J. and A. M. Mood (1948). 'A method for obtaining and analyzing sensitivity data'. In: *Journal of the American Statistical Association* 43, pp. 109–126. DOI: 10.1080/01621459.1948.10483254.

Doss, H. (1994). 'Bayesian nonparametric estimation for incomplete data via successive substitution sampling'. In: *The Annals of Statistics* 22, pp. 1763–1786. DOI: 10.1214/aos/1176325756.

Durham, S. D. and N. Flournoy (1994). 'Random walks for quantile estimation'. In: *Statistical Decision Theory and Related Topics*. Ed. by S. S. Gupta and J. O. Berger. Springer, pp. 457–476. DOI: 10.1007/978-1-4612-2618-5_36.

Faraggi, D., P. Izikson and B. Reiser (2003). 'Confidence intervals for the 50 per cent response dose'. In: *Statistics in Medicine* 22, pp. 1977–1988. DOI: 10.1002/sim.1368.

Ferguson, T. S. (1973). 'A Bayesian analysis of some nonparametric problems'. In: *The Annals of Statistics* 1, pp. 209–230. DOI: 10.1214/aos/1176342360.

Fieller, E. C. (1954). 'Some problems in interval estimation'. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 16, pp. 175–185. DOI: 10.1111/j.2517-6161.1954.tb00159.x.

Fisher, R. A. (1935). 'The fiducial argument in statistical inference'. In: *Annals of Eugenics* 5, pp. 391–398. DOI: 10.1111/j.1469-1809.1935.tb02120.x.

Flournoy, N. (2002). 'Up-and-down designs'. In: *Encyclopedia of Environmetrics*. Ed. by A. H. El-Shaarawi and W. W. Piegorsch. Vol. 4. John Wiley & Sons, pp. 2297–2302. DOI: 10.1002/9781118445112.stat07338.

Gil-Leyva, M. F. and R. H. Mena (2023). 'Stick-breaking processes with exchangeable length variables'. In: *Journal of the American Statistical Association* 118, pp. 537–550. DOI: 10.1080/01621459.2021.1941054.

'Gospel of Matthew' (1611). In: *King James Bible*.

Groeneboom, P. and G. Jongbloed (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge University Press.

Groeneboom, P. and J. A. Wellner (1992). *Information bounds and nonparametric maximum likelihood estimation*. Birkhäuser Verlag.

Hastings, W. K. (1970). 'Monte Carlo sampling methods using Markov chains and their applications'. In: *Biometrika* 57, pp. 97–109. DOI: 10.1093/biomet/57.1.97.

Hjort, N. L. (1990). 'Nonparametric Bayes estimators based on Beta processes in models for life history data'. In: *The Annals of Statistics* 18, pp. 1259–1294. DOI: 10.1214/aos/1176347749.

Hjort, N. L. and S. G. Walker (2009). 'Quantile pyramids for Bayesian nonparametrics'. In: *The Annals of Statistics* 37, pp. 105–131. DOI: 10.1214/07-AOS553.

Ibsen, H. (1884). *Vildanden*. Gyldendal.

Jerrum, M. and A. Sinclair (1989). 'Approximating the permanent'. In: *SIAM Journal on Computing* 18, pp. 1149–1178. DOI: 10.1137/0218077.

Joseph, V. R. (2004). 'Efficient Robbins–Monro procedure for binary data'. In: *Biometrika* 91, pp. 461–470. DOI: 10.1093/biomet/91.2.461.

Kalli, M., J. E. Griffin and S. G. Walker (2009). 'Slice sampling mixture models'. In: *Statistics and Computing* 21, pp. 93–105. DOI: 10.1007/s11222-009-9150-y.

Kaplan, E. L. and P. Meier (1958). 'Nonparametric estimation from incomplete observations'. In: *Journal of the American Statistical Association* 53, pp. 457–481. DOI: 10.1080/01621459.1958.10501452.

Karmarkar, N. et al. (1993). 'A Monte–Carlo algorithm for estimating the permanent'. In: *SIAM Journal of Computing* 22, pp. 284–293. DOI: 10.1137/0222021.

Khrapkovskii, G. M. et al. (2013). 'Theoretical study of substituents effect on C–NO2 bond strength in mono substituted nitrobenzenes'. In: *Computational and Theoretical Chemistry* 1017, pp. 7–13. DOI: 10.1016/j.comptc.2013.04.013.

Lai, T. L. (2003). 'Stochastic approximation'. In: *The Annals of Statistics* 31, pp. 391–406. DOI: 10.1214/aos/1051027873.

Langlie, H. J. (1965). *A reliability test for "one-shot" items*. Tech. rep. U-1792. Newport Beach, CA: Aeronutronic Division, Ford Motor Company.

Lijoi, A., I. Prünster and T. Rigon (2020). 'The Pitman–Yor multinomial process for mixture modelling'. In: *Biometrika* 107, pp. 891–906. DOI: 10.1093/biomet/asaa030.

MacKay, D. J. C. (1992). 'Bayesian interpolation'. In: *Neural Computation* 4, pp. 415–447. DOI: 10.1162/neco.1992.4.3.415.

Marin, J.-M. et al. (2012). 'Approximate Bayesian computational methods'. In: *Statistics and Computing* 22, pp. 1167–1180. DOI: 10.1007/s11222-011-9288-2.

Meng, X.-L. and W. H. Wong (1996). 'Simulating ratios of normalizing constants via a simple identity: A theoretical exploration'. In: *Statistica Sinica* 6, pp. 831–860.

Metropolis, N. et al. (1953). 'Equation of state calculations by fast computing machines'. In: *The Journal of Chemical Physics* 21, pp. 1087–1092. DOI: 10.1063/1.1699114.

Miller Jr, R. G. (1983). 'What price Kaplan–Meier?' In: *Biometrics* 39, pp. 1077–1081. DOI: 10.2307/2531341.

Neyer, B. T. (1994). 'A D-optimality-based sensitivity test'. In: *Technometrics* 36, pp. 61–70. DOI: 10.1080/00401706.1994.10485401.

North Atlantic Treaty Organization (NATO) (1999). *STANAG 4489 - Explosives, impact sensitivity tests*.

— (2009a). *STANAG 4487 - Explosives, friction sensitivity tests*.

— (2009b). *STANAG 4488 - Explosives, shock sensitivity tests*.

Paisley, J. W. et al. (2010). 'A stick-breaking construction for the Beta process'. In: *Proceedings of the International Conference on Machine Learning*. PMLR, pp. 847–854.

Samseth, Jarl Øystein (2022). 'Which bond cleavage is the explosive's impact sensitivity most sensitive to?' In: *Proceedings of the 51st International Annual Conference of the Fraunhofer ICT*. Fraunhofer ICT.

Schweder, T. and N. L. Hjort (2016). *Confidence, Likelihood, Probability*. Cambridge University Press.

Sitter, R. R. and C. F. J. Wu (1993). 'On the accuracy of Fieller intervals for binary response data'. In: *Journal of the American Statistical Association* 88, pp. 1021–1025. DOI: 10.1080/01621459.1993.10476370.

Turnbull, B. W. (1974). 'Nonparametric estimation of a survivorship function with doubly censored data'. In: *Journal of the American Statistical Association* 69, pp. 169–173. DOI: 10.1080/01621459.1974.10480146.

U.S. Department of Defense (2001). *MIL-STD-1751A - Safety and performance tests for the qualification of explosives (high explosives, propellants, and pyrotechnics)*.

Vinterberg, T. and M. Rukov (1998). *Festen*. Nimbus Film.

Walker, S. G. (2007). 'Sampling the Dirichlet mixture model with slices'. In: *Communications in Statistics - Simulation and Computation* 36, pp. 45–54. DOI: 10.1080/03610910601096262.

Wilks, S. S. (1938). 'The large-sample distribution of the likelihood ratio for testing composite hypotheses'. In: *The Annals of Mathematical Statistics* 9, pp. 60–62. DOI: 10.1214/aoms/1177732360.

Wu, C. F. J. and Y. Tian (2014). 'Three-phase optimal design of sensitivity experiments'. In: *Journal of Statistical Planning and Inference* 149, pp. 1–15. DOI: 10.1016/j.jspi.2013.10.007.

Paper I

Inference for Bayesian nonparametric models with binary response data via permutation counting

Inference for Bayesian Nonparametric Models with Binary Response Data via Permutation Counting

Dennis Christensen[*,†]

Abstract. Since the beginning of Bayesian nonparametrics in the early 1970s, there has been a wide interest in constructing models for binary response data. Such data arise naturally in problems dealing with bioassay, current status data and sensitivity testing, and are equivalent to left and right censored observations if the inputs are one-dimensional. For models based on the Dirichlet process, inference is possible via Markov chain Monte Carlo (MCMC) simulations. However, there exist multiple processes based on different principles, for which such MCMC-based methods fail. Examples include logistic Gaussian processes and quantile pyramids. These require MCMC for posterior inference given exact observations, and thus become intractable when the data comprise both left and right censored observations. Here we present a new importance sampling algorithm for nonparametric models given exchangeable binary response data. It can be applied to any model from which samples can be generated, or even only approximately generated. The main idea behind the algorithm is to exploit the symmetries introduced by exchangeability. Calculating the importance weights turns out to be equivalent to evaluating the permanent of a certain class of $(0, 1)$-matrix, which we prove can be done in polynomial time by deriving an explicit algorithm.

MSC2020 subject classifications: Primary 62N01, 62G05; secondary 15A15.

Keywords: Bayesian nonparametrics, binary response data, current status data, bioassay, permanents, importance sampling, binary classification.

1 Introduction

In many statistical applications, we only observe a Bernoulli random variable indicating whether a real-valued latent variable is below or above a certain threshold. Examples include problems in current status data and bioassay, where we aim to estimate a distribution P governing the probability that an individual has transitioned from state 0 to state 1 before time t (Albert and Chib, 1993; Keiding et al., 1996; Groeneboom and Jongbloed, 2014). Another example is sensitivity testing, in which we repeatedly choose an impact level E of energy and then observe whether a physical system is intact or broken after the impact. Such methods are used for studying the sensitivity of explosives or a material's resistance to stress (Dixon and Mood, 1948; Neyer, 1994; Christensen, 2022).

Since the beginning of Bayesian nonparametrics, there has been an interest in such

---

[*]Department of Mathematics, University of Oslo, Oslo, Norway.
[†]Norwegian Defence Research Establishment (FFI), Kjeller, Norway, dennis.christensen@ffi.no

binary response problems. Following the introduction of the Dirichlet process by Ferguson (1973), Antoniak (1974) showed that the posterior distribution of a Dirichlet process given censored data is a mixture of Dirichlet processes, and applied this to bioassay. Dirichlet processes with binary response data were further studied in Bhattacharya (1981); Kuo (1988); Gelfand and Kuo (1991); Doss (1994); Newton and Zhang (1999). These methods rely on one or more of the following particularly useful properties of the Dirichlet process: its conjugate posterior representation (Ferguson, 1973), its explicit marginal distribution (Antoniak, 1974) and its stick-breaking representation (Sethuraman, 1994). These three properties have also allowed for the development of Markov chain Monte Carlo (MCMC) sampling methods for Dirichlet process mixture models (DPMMs) (Ferguson, 1983; Lo, 1984). In the terminology of Papaspiliopoulos and Roberts (2008), this may be achieved either with marginal MCMC methods (Escobar and West, 1995; Neal, 2000) or conditional MCMC methods (Walker, 2007; Kalli et al., 2011). Due to the tractability of the mixture components, such MCMC techniques also apply when dealing with binary response data (see Paulon et al. (2020) for a recent application with current status data and dependent censoring). In principle, this is not only true for DPMMs, but for any mixture model with tractable marginal mixture distributions, such as normalised random measures with independent increments (NRMIs) (Regazzini et al., 2003; Lijoi et al., 2005, 2007), or with stick-breaking representations (Hjort, 1990; Paisley et al., 2010; Ishwaran and James, 2001). On the semiparametric side, Bayesian inference for the proportional hazards model (Cox, 1972) with current status data has been studied via Gibbs sampling (Cai et al., 2011) and expectation maximisation (Wang et al., 2015).

Although models based on Dirichlet processes are largely applicable to problems with binary response data, many nonparametric models are not. A notable example of this is the logistic Gaussian process (Leonard, 1978; Lenk, 1988, 1991) and the Gaussian process density sampler (Murray et al., 2008). In these models, the likelihood function will contain an integral of a Gaussian process due to the censoring. Thus, a direct implementation of MCMC-based inference is not feasible. Furthermore, since there is no conjugate posterior representation for such models, alternatives such as successive substitution sampling (Doss, 1994) are also out of reach. The same is true for other model choices in Bayesian nonparametrics, such as quantile pyramids (Hjort and Walker, 2009), normalised infinitely divisible multinomial (NIDM) processes (Lijoi et al., 2019) and Pitman-Yor multinomial processes (Lijoi et al., 2020). These examples are in line with Orbanz and Teh's (2011) prediction that over time, more Bayesian nonparametric models which are not based on the Dirichlet process will continue to arise.

In this paper, we introduce a new importance sampling algorithm which enables full Bayesian inference for models with exchangeable binary response data. The construction is highly general and applies to any model from which a data sample can be simulated. In particular, it does not rely on the tractable properties of Dirichlet processes. Unlike approximate methods such as the approximate Bayesian computation (ABC) rejection sampling algorithm (see Marin et al. (2012) for a review), our new simulation algorithm converges to the true posterior distribution, not just an approximation of it. As is illustrated in our simulation case study (see Section 4), this exact convergence result also holds when it is only possible to sample from a finite-dimensional truncation of the

model, as studied by Muliere and Secchi (1995); Campbell et al. (2019); Arbel et al. (2019); Lijoi et al. (2019, 2020).

The key to the new algorithm is to exploit the symmetry introduced by exchangeability of the data, and then essentially to correct for this exploitation by multiplying by an appropriate importance weight. Calculating the weight turns out to be equivalent to evaluating the permanent of a $(0, 1)$-matrix, that is, a matrix whose entries are all either 0 or 1. For a general such matrix, this is known to be a #P-complete problem (Valiant, 1979). However, for the matrices arising in our setting, we are able to derive an explicit algorithm which computes their permanents in polynomial time. Code for implementing this new algorithm can be found in the publicly available GitHub repository.

The remainder of the paper is structured as follows. In Section 2, we set up the problem and introduce the importance sampling algorithm. We show how to calculate the marginal likelihood and how to carry out posterior inference. In Section 3, we derive an algorithm for calculating the importance weights in polynomial time. Next, in Section 4, we apply the new importance sampling algorithm to experiments with both simulated and real data. The theory is then extended in Section 5 to problems with multidimensional inputs. Finally, we briefly discuss extensions, limitations and consistency in Section 6.

2 Construction

2.1 Model

Let $([0, \infty), \mathscr{F})$ be the measurable space of non-negative real numbers equipped with the Borel σ-algebra. We use $[0, \infty)$ as our sample space to more conveniently illustrate the theory, although everything also applies to $\mathbb{R}$ or a real bounded interval. Let $P \sim \pi(\cdot)$ be a random probability distribution on $([0, \infty), \mathscr{F})$. Then P induces a random cumulative distribution function (cdf) F on $[0, \infty)$. Our binary data $y = (y_1, \ldots, y_n) \in \{0, 1\}^n$ is assumed to be generated by

$$y_i \mid F \sim \text{Bernoulli}(F(t_i)),$$

independently for $i = 1, \ldots, n$, for some known thresholds $t_1, \ldots, t_n$. That is, $\pi(y_i \mid F) = F(t_i)^{y_i} \{1 - F(t_i)\}^{1-y_i}$. We also write $\pi(y)$ for the marginal distribution of y, having marginalised over F.

It is useful to introduce the latent variables $x = (x_1, \ldots, x_n) \in [0, \infty)^n$ with $x_i \mid P \sim P$ independently for $i = 1, \ldots, n$. That is, $\pi(x_i \mid P) = P$. Then our binary variables y_i can be seen as indicator variables, $y_i = \mathbb{1}_{x_i \leq t_i}$. Let $\pi(x)$ be the marginal distribution of x, marginalising over P. Note that the x_i need not be marginally independent. However, they will always form an exchangeable sequence. The same is true for the y_i. Since $\pi(\cdot)$ may both refer to the distribution of P and the marginal distributions of x or y, we will make it clear from context which distribution is in use.

We shall need to introduce some notation. Given $y_1, \ldots, y_n$, let

$$\mathcal{B}_i = \begin{cases} [0, t_i] & \text{if } y_i = 1, \\ (t_i, \infty) & \text{if } y_i = 0, \end{cases} \tag{2.1}$$

for $i = 1, \ldots, n$. Thus, observing y is equivalent to observing that $x_i \in \mathcal{B}_i$ for all $i = 1, \ldots, n$. Now, let $n_0 = \#\{i \mid y_i = 1\}$. Then, by exchangeability, we may without loss of generality order the y_i so that $y_1 = \cdots = y_{n_0} = 1$ and $y_{n_0+1} = \cdots = y_n = 0$, and further so that $t_1 \leq \cdots \leq t_{n_0}$ and $t_{n_0+1} \leq \cdots \leq t_n$. Note that this also induces an ordering of the sets $\mathcal{B}_1, \ldots, \mathcal{B}_n$. We write $\mathcal{B} = \mathcal{B}_1 \times \cdots \times \mathcal{B}_n$, so that observing y is equivalent to observing that $x \in \mathcal{B}$. For now, we will assume that there are no repeated values amongst the thresholds $t_1, \ldots, t_n$. Later on, in Section 2.2, we show how to account for situations where we have repeated values amongst them.

2.2 Estimating the marginal likelihood

Our first objective is to estimate the marginal likelihood $\pi(y) = \mathbb{P}(x \in \mathcal{B})$ of the model. In addition to being valuable in its own right, this will also guide how to perform posterior inference for P in general, to be covered in Section 2.3. For brevity of notation, define the measure $\mathscr{P}$ on $([0, \infty)^n, \mathscr{F}^n)$ by $\mathscr{P}(\mathcal{A}) = \mathbb{P}(x \in \mathcal{A})$. The marginal likelihood will be estimated via an importance sampling algorithm, exploiting the symmetries present as a result of the x_i being exchangeable.

Let $\mathbb{1}$ denote the indicator function, so $\mathbb{1}_{\mathcal{B}}(x)$ returns 1 if $x \in \mathcal{B}$ and 0 otherwise. Consider first the following naive estimator.

$$\hat{\mathscr{P}}_T(\mathcal{B}) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}_{\mathcal{B}}(x^{(t)}), \tag{2.2}$$

where $x^{(t)} \sim \pi(\cdot)$ independently for $t = 1, \ldots, T$.

By the law of large numbers, $\hat{\mathscr{P}}_T(\mathcal{B})$ is indeed a consistent estimator for the marginal likelihood $\mathscr{P}(\mathcal{B})$. However, in practice, we will never experience that $x \in \mathcal{B}$ if n is even moderately large, so $\hat{\mathscr{P}}_T(\mathcal{B})$ will always just be zero. This is also true even if parallel computing is employed, as the acceptance probability decreases exponentially with n. In order to adjust it to yield something practically feasible, we will have to loosen the condition that $x \in \mathcal{B}$ by replacing $\mathcal{B}$ with a larger space. We do this by exploiting the symmetries of the measure $\mathscr{P}$ due to the exchangeability of the x_i.

The group S_n of n-permutations acts on $[0, \infty)^n$ via permutations of indices. Specifically, for $x = (x_1, \ldots, x_n) \in [0, \infty)^n$, we write $\sigma(x) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for the result of hitting x with the permutation $\sigma \in S_n$. Similarly, S_n acts on $\mathscr{F}^n$ via permutations, and we write $\sigma(\mathcal{B}) = \mathcal{B}_{\sigma(1)} \times \cdots \times \mathcal{B}_{\sigma(n)}$ for $\mathcal{B} = \mathcal{B}_1 \times \cdots \times \mathcal{B}_n \in \mathscr{F}^n$. We define the *orbit*[1] $\mathrm{Orb}(\mathcal{B})$ of $\mathcal{B}$ to be the set

$$\mathrm{Orb}(\mathcal{B}) = \bigcup_{\sigma \in S_n} \sigma(\mathcal{B}).$$

---

[1] Strictly speaking, this is the union of the orbit, where the orbit is usually defined as $\{\sigma(\mathcal{B}) \mid \sigma \in S_n\}$.
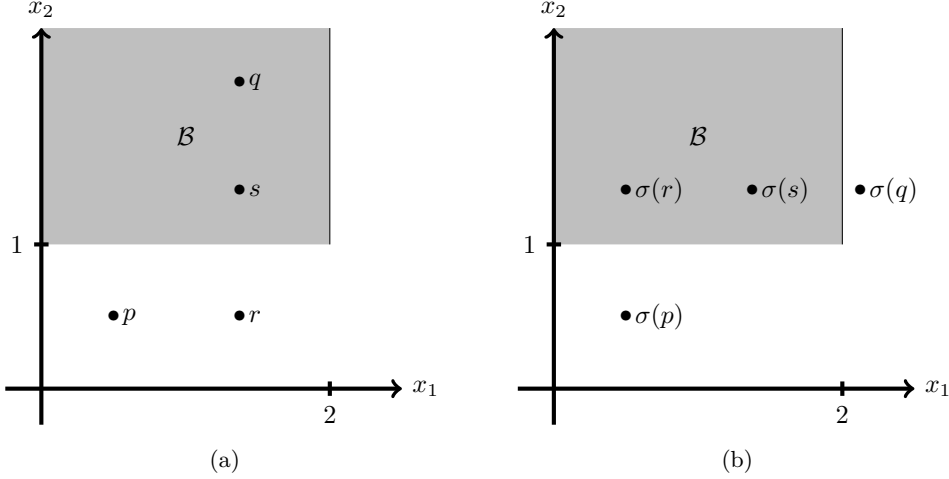
Figure 1: A two-dimensional example of calculating permutation numbers.

Next, we define the *permutation number $w(x; \mathcal{B})$ of x with respect to $\mathcal{B}$* as

$$w(x; \mathcal{B}) = \#\{\sigma \in S_n \mid \sigma(x) \in \mathcal{B}\}.$$

Note that $0 \leq w(x; \mathcal{B}) \leq n!$ for all x, and furthermore that $w(x; \mathcal{B}) = 0$ if and only if $x \notin \mathrm{Orb}(\mathcal{B})$, that is, if and only if $\sigma(x) \notin \mathcal{B}$ for all permutations $\sigma \in S_n$.

Example. *We show how to calculate permutation numbers in a simple two–dimensional example. Consider the set $\mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2 = [0, 2] \times (1, \infty)$, drawn in Figure 1a. In order to calculate the permutation numbers of the points p, q, r, s, we hit these points with the nontrivial permutation $\sigma \in S_2$, as shown in Figure 1b. That is, we reflect them across the diagonal $x_1 = x_2$. Now, $p \notin \mathcal{B}$ and $\sigma(p) \notin \mathcal{B}$, so $w(p; \mathcal{B}) = 0$. Next, $q \in \mathcal{B}$ and $\sigma(q) \notin \mathcal{B}$, so $w(q; \mathcal{B}) = 1$. Similarly, $r \notin \mathcal{B}$ and $\sigma(r) \in \mathcal{B}$, so $w(r; \mathcal{B}) = 1$. Finally, $s \in \mathcal{B}$ and $\sigma(s) \in \mathcal{B}$, so $w(s; \mathcal{B}) = 2$. Note in particular that $w(r; \mathcal{B}) > 0$ even though $r \notin \mathcal{B}$.*

In Section 3, we will derive an algorithm for efficiently computing the permutation numbers $w(x; \mathcal{B})$. For now, we shall show how they can be used to construct an importance sampling algorithm as an alternative to (2.2). Consider the modified estimator

$$\hat{\mathscr{P}}_T^{\mathrm{IS}}(\mathcal{B}) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{n!} w(x^{(t)}; \mathcal{B}), \tag{2.3}$$

where $x^{(t)} \sim \pi(\cdot)$ independently for $t = 1, \ldots, T$. This is essentially an importance sampling estimator with proposal distribution $x \sim \pi(\cdot)$ and weights $W(x) = w(x; \mathcal{B})/n!$.

Proposition 2.1. *$\hat{\mathscr{P}}_T^{\mathrm{IS}}(\mathcal{B})$ is an unbiased and consistent estimator for the marginal likelihood $\mathscr{P}(\mathcal{B})$.*

Proof. We have that $w(x; \mathcal{B}) = \#\{\sigma \in S_n \mid \sigma(x) \in \mathcal{B}\} = \#\{\sigma \in S_n \mid x \in \sigma(\mathcal{B})\}$, so that taking expectations, we get

$$\mathbb{E}\left[\frac{1}{n!} w(x; \mathcal{B})\right] = \frac{1}{n!} \int_{\mathrm{Orb}(\mathcal{B})} w(x; \mathcal{B}) \, \mathrm{d}\pi(x) = \frac{1}{n!} \sum_{\sigma \in S_n} \int_{\sigma(\mathcal{B})} \mathrm{d}\pi(x)$$

$$= \frac{1}{n!} \sum_{\sigma \in S_n} \mathscr{P}(\sigma(\mathcal{B})) = \mathscr{P}(\mathcal{B}),$$

where we have used exchangeability for the final equality. This proves that the estimator is unbiased. Applying the law of large numbers to (2.3) establishes consistency. □

The benefit of calculating $\hat{\mathscr{P}}_T^{\mathrm{IS}}(\mathcal{B})$ rather than the naive estimate $\hat{\mathscr{P}}_T(\mathcal{B})$ is that we only require x to land in $\mathrm{Orb}(\mathcal{B})$, which is a much larger set than $\mathcal{B}$. In practice, this means that we get way more contributing samples when calculating (2.3) rather than (2.2).

Given x and $\mathcal{B}$, it is not, a priori, easy to determine whether $x \in \mathrm{Orb}(\mathcal{B})$. However, by considering the order statistics of x, we can establish an easily verifiable criterion.

Definition 2.1. *Let $x \in [0, \infty)^n$ be fixed and let $\sigma \in S_n$ be any n-permutation such that $x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}$. That is, $\sigma(x)$ are the order statistics of x. We say that x is $\mathcal{B}$-admissible if $\sigma(x) \in \mathcal{B}$.*

Proposition 2.2. *Let $x \in [0, \infty)^n$. Then $w(x; \mathcal{B}) > 0$ if and only if x is $\mathcal{B}$-admissible.*

Proof. See the Supplementary Material (Christensen, 2023). □

Proposition 2.2 simplifies computations significantly. For instance, when calculating $\hat{\mathscr{P}}_T^{\mathrm{IS}}(\mathcal{B})$, we now have a simple criterion for checking whether $x \in \mathrm{Orb}(\mathcal{B})$. Namely, we check whether $\sigma(x) \in \mathcal{B}$, where σ is as in Definition 2.1.

Example. *We illustrate the volume of $\mathrm{Orb}(\mathcal{B})$ via a simple simulation study. For $n = 1, \ldots, 300$, we let $0 < r_1 < \ldots r_n < 1$ be uniformly spaced and simulated $u_1, \ldots, u_n \sim \mathrm{Uniform}[0, 1]$ independently. As in Section 2.1, we write $n_0 = \#\{i \mid u_i \leq r_i\}$ and let $t_1 < \cdots < t_{n_0}$ be those r_i satisfying $u_i \leq r_i$. Similarly, we let $t_{n_0+1} < \cdots < t_n$ be those r_i satisfying $u_i > r_i$. This defines the set $\mathcal{B}$.*

Two experiments were conducted, one where $x_1, \ldots, x_n \sim \mathrm{Uniform}[0, 1]$ and another where $x_1, \ldots, x_n \sim \mathrm{Beta}(2, 2)$, independently. The probability that $x \in \mathcal{B}$ is given by

$$\mathbb{P}(x \in \mathcal{B}) = \prod_{i=1}^{n_0} F(t_i) \times \prod_{i=n_0+1}^{n} \{1 - F(t_i)\},$$

where F denotes the cdf of the Uniform and Beta distribution in the first and second experiment, respectively. In either case, this probability decays exponentially and gets vanishingly small as n gets large. In order to compare $\mathbb{P}(x \in \mathcal{B})$ with the probability $\mathbb{P}(x \in \mathrm{Orb}(\mathcal{B}))$, we repeatedly simulated copies of x a total of 1000 times and counted
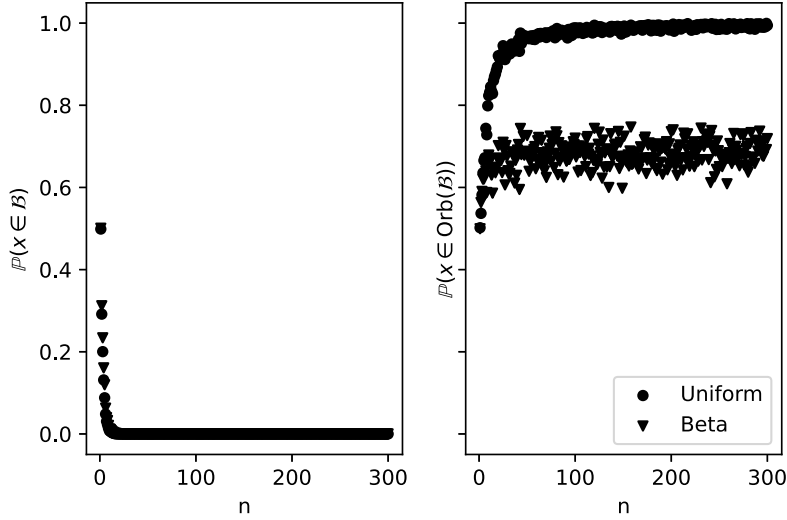
Figure 2: Empirical estimates of $\mathbb{P}(x \in \mathcal{B})$ and $\mathbb{P}(x \in \mathrm{Orb}(\mathcal{B}))$ for $x_1, \ldots, x_n \sim$ Uniform$[0, 1]$ independently and $x_1, \ldots, x_n \sim \mathrm{Beta}(2, 2)$ independently.

how many times we observed $x \in \mathcal{B}$ and how many times we observed $x \in \mathrm{Orb}(\mathcal{B})$ (using Proposition 2.2). We did this 100 times for each n and averaged the results, which are plotted in Figure 2. We note that in both experiments, we get a much higher acceptance proportion when working with $\mathrm{Orb}(\mathcal{B})$. Also, this proportion does not seem to decrease with n. This makes sense intuitively, since the number of permutations increases exponentially with n.

Remark. *The quality of the estimate $\hat{\mathscr{P}}_T^{\mathrm{IS}}(\mathcal{B})$ is diagnosed by calculating the* effective sample size *(*ESS*), given by*

$$\mathrm{ESS} = \frac{\left(\sum_{t=1}^T w(x^{(t)}; \mathcal{B})/n!\right)^2}{\sum_{t=1}^T \left(w(x^{(t)}; \mathcal{B})/n!\right)^2} = \frac{\left(\sum_{t=1}^T w(x^{(t)}; \mathcal{B})\right)^2}{\sum_{t=1}^T w(x^{(t)}; \mathcal{B})^2}.$$

In (2.3), it may be useful not to treat T as a fixed sample size, but rather to keep adding terms until the ESS *reaches a fixed, predetermined value.*

In principle, $\hat{\mathscr{P}}_T^{\mathrm{IS}}(\mathcal{B})$ can also be computed in cases where all observations are only right or left censored. That is, in cases where $n_0 = 0$ or $n_0 = n$. However, for such problems, we experience that the ESS *increases too slowly. By studying the geometry of the situation, we can gain some insight into why this is the case. The* ESS *will be small if only a few weights dominate. Now, if all the $\mathcal{B}_i$ extend to the right, say, then samples $x^{(t)}$ with all $x_i^{(t)}$ far to the right will yield large permutation numbers $w(x^{(t)}; \mathcal{B})$. Indeed, the maximum value $w(x^{(t)}; \mathcal{B}) = n!$ is attainable if the $x_i^{(t)}$ are sufficiently far to the right. However, when both left and right censored observations are present, the $\mathcal{B}_i$ are*

no longer nested, and so the weights tend to be more uniformly spread out, making a larger effective sample size obtainable.

Remark. *Some of the thresholds t_i may be equal. Let $r_1 \leq \cdots \leq r_l$ be the unique values of the set $\{t_1, \ldots, t_n\}$ and for each $j = 1, \ldots, l$, let $a_j = \#\{i \mid t_i = r_j\}$ and $b_j = \#\{i \mid t_i = r_j, i \leq n_0\}$. That is, a_j is the number of trials conducted at input r_j and b_j is the number of successes. The observations are now a sequence of binomial variables, and so the marginal likelihood takes the form $\mathbb{P}(x \in \mathcal{B}) \times \prod_{j=1}^{l} \binom{a_j}{b_j}$. Hence, if we have repeated trials, we may simply redefine $w(x; \mathcal{B})$ to be $w(x; \mathcal{B}) \times \prod_{j=1}^{l} \binom{a_j}{b_j}$ and carry out our analysis as normal. We continue without explicitly multiplying by this factor in our notation, but it should be kept in mind that the permutation numbers are multiplied by this factor if the data include repeated trials.*

Remark. *In (2.3), the samples come from the prior. Although the simulation study above indicates that this naive approach works sufficiently well, other choices of proposal distribution may be more efficient and increase performance. Examples of methods that might do so include sequential Monte Carlo (see Cappé et al. (2007) for a review), defensive mixture proposal distributions (Hesterberg, 1995) and population Monte Carlo (Cappé et al., 2004). For the sake of simplicity, in the present paper, we shall only consider the case where the samples are drawn from the prior.*

We conclude this section by showing that our estimate $\hat{\mathscr{P}}_T^{\mathrm{IS}}(\mathcal{B})$ yields a smaller variance than the naive estimate $\hat{\mathscr{P}}_T(\mathcal{B})$.

Proposition 2.3. *We have that*

$$\mathrm{Var}\left(\frac{1}{n!} w(x; \mathcal{B})\right) = \mathrm{Var}\left(\mathbb{1}_{\mathcal{B}}(x)\right) + \mathscr{P}(\mathcal{B}) - \frac{1}{n!} \sum_{\sigma \in S_n} \mathscr{P}(\sigma(\mathcal{B}) \cup \mathcal{B}).$$

In particular,

$$\mathrm{Var}\left(\frac{1}{n!} w(x; \mathcal{B})\right) \leq \mathrm{Var}\left(\mathbb{1}_{\mathcal{B}}(x)\right).$$

Proof. Using the same reasoning as in Proposition 2.1, we have that

$$w(x; \mathcal{B})^2 = \#\{(\sigma, \tau) \in S_n^2 \mid \sigma(x), \tau(x) \in \mathcal{B}\} = \#\{(\sigma, \tau) \in S_n^2 \mid x \in \sigma(\mathcal{B}) \cap \tau(\mathcal{B})\},$$

and so

$$\mathbb{E}\left[\left(\frac{1}{n!} w(x; \mathcal{B})\right)^2\right] = \frac{1}{(n!)^2} \int_{\mathrm{Orb}(\mathcal{B})} w(x; \mathcal{B})^2 \, \mathrm{d}\pi(x) = \frac{1}{(n!)^2} \sum_{\sigma, \tau \in S_n} \int_{\sigma(\mathcal{B}) \cap \tau(\mathcal{B})} \mathrm{d}\pi(x)$$

$$= \frac{1}{n!} \sum_{\sigma \in S_n} \mathscr{P}(\sigma(\mathcal{B}) \cap \mathcal{B}) = \frac{1}{n!} \sum_{\sigma \in S_n} \{\mathscr{P}(\sigma(\mathcal{B})) + \mathscr{P}(\mathcal{B}) - \mathscr{P}(\sigma(\mathcal{B}) \cup \mathcal{B})\}$$

$$= 2\mathscr{P}(\mathcal{B}) - \frac{1}{n!} \sum_{\sigma \in S_n} \mathscr{P}(\sigma(\mathcal{B}) \cup \mathcal{B}).$$

Subtracting $\mathbb{E}[(1/n!)w(x; \mathcal{B})]^2$ from both sides yields the first result. For the second, note that $\mathscr{P}(\sigma(\mathcal{B}) \cup \mathcal{B}) \geq \mathscr{P}(\mathcal{B})$ for all $\sigma \in S_n$. □

2.3 Posterior inference

We now extend the results from the previous section to a general importance sampling algorithm targeting the posterior distribution of P given that $x \in \mathcal{B}$. Let $\theta = \theta(P) \in \mathbb{R}$ be some quantity related to P, such as a specific cdf value $F(t) = P([0, t])$ or a quantile $F^{-1}(q)$. Then consider the estimator

$$\hat{\theta}_T^{\mathrm{IS}} = \frac{\sum_{t=1}^T \theta^{(t)} w(x^{(t)}; \mathcal{B})}{\sum_{t=1}^T w(x^{(t)}; \mathcal{B})}, \tag{2.4}$$

where $\theta^{(t)} = \theta(P^{(t)})$, $x^{(t)} \sim P^{(t)}$ and $P^{(t)} \sim \pi(\cdot)$, independently for $t = 1, \ldots, T$.

Proposition 2.4. *The statistic $\hat{\theta}_T^{\mathrm{IS}}$ is a consistent estimator for the posterior mean*

$$\mathbb{E}[\theta \mid x \in \mathcal{B}].$$

Proof. The expression (2.4) is precisely the self-normalised importance sampling estimator targeting

$$\frac{\int_{[0,\infty)^n \times \mathbb{R}} \theta w(x; \mathcal{B}) \, \mathrm{d}\pi(x, \theta)}{\int_{[0,\infty)^n \times \mathbb{R}} w(x; \mathcal{B}) \, \mathrm{d}\pi(x, \theta)} = \frac{\int_{\mathrm{Orb}(\mathcal{B}) \times \mathbb{R}} \theta w(x; \mathcal{B}) \, \mathrm{d}\pi(x, \theta)}{\int_{\mathrm{Orb}(\mathcal{B}) \times \mathbb{R}} w(x; \mathcal{B}) \, \mathrm{d}\pi(x, \theta)} = \frac{\int_{\mathcal{B} \times \mathbb{R}} \theta \, \mathrm{d}\pi(x, \theta)}{\int_{\mathcal{B} \times \mathbb{R}} \mathrm{d}\pi(x, \theta)}$$

$$= \int_{\mathbb{R}} \theta \, \mathrm{d}\pi(\theta \mid x \in \mathcal{B}) = \mathbb{E}[\theta \mid x \in \mathcal{B}],$$

as required. □

3 Permutation numbers

We now outline how to calculate the permutation numbers $w(x; \mathcal{B})$. For the remainder of this section, assume that x is $\mathcal{B}$-admissible, so we know that $w(x; \mathcal{B}) > 0$. The first step of the derivation is to express the permutation number $w(x; \mathcal{B})$ as the *permanent* of a $(0, 1)$-matrix.

Definition 3.1. *Let $A = (a_{ij})$ be an $m \times n$ matrix where $m \leq n$. Let $S_{n,m}$ denote the set of all m-permutations of the set $\{1, \ldots, n\}$. The* permanent $\mathrm{perm}(A)$ *of A is defined by*

$$\mathrm{perm}(A) = \sum_{\tau \in S_{n,m}} \prod_{i=1}^m a_{i,\tau(i)} \tag{3.1}$$

Note that the permanent is defined for rectangular matrices, not just square ones. In order to express the permutation number $w(x; \mathcal{B})$ as the permanent of a matrix, we make the following definition. Given $x \in [0, \infty)^n$, the *matching matrix* $A = (a_{ij})$ of x is the $n \times n$ $(0, 1)$-matrix defined by

$$a_{ij} = \begin{cases} 1 & \text{if } x_i \in \mathcal{B}_j \\ 0 & \text{if } x_i \notin \mathcal{B}_j. \end{cases}$$

Lemma 3.1. *Let A be the matching matrix of x. Then*

$$w(x; \mathcal{B}) = \operatorname{perm}(A).$$

The key ingredient in the proof is to count the number of matchings in a bipartite graph. If the reader is unfamiliar with these notions, we recommend reading Bollobás (1979, Chapter 3).

Proof of Lemma 3.1. Let $\mathcal{G} = (V, E)$ be the bipartite graph with vertex set $V = \{x_1, \ldots, x_n\} \cup \{\mathcal{B}_1, \ldots, \mathcal{B}_n\}$ and edges $E = \{(x_i, \mathcal{B}_j) \in V^2 \mid x_i \in \mathcal{B}_j\}$. Then the permutation number $w(x; \mathcal{B})$ is equal to the number of bijections $f : \{x_1, \ldots, x_n\} \to \{\mathcal{B}_1, \ldots, \mathcal{B}_n\}$ such that $x_i \in f(x_i)$. That is, the number of perfect matchings in $\mathcal{G}$. But this is precisely the permanent of the biadjacency matrix of $\mathcal{G}$. That is, the permanent of the matrix A. □

Permanents are notoriously difficult to compute. Unlike the closely related determinant function (which is obtained by multiplying each term in (3.1) by $\operatorname{sign}(\tau)$), the permanent function is not multiplicative, and thus we cannot employ Gaussian elimination to compute permanents in polynomial time. In general, computing permanents of $(0, 1)$-matrices is known to be a #P-complete problem (Valiant, 1979). The fastest general formula known for $(0, 1)$-matrices is that by Ryser (1963), which requires $O(2^{n-1}n)$ operations for an $n \times n$ matrix. More recently, Huh (2022) has presented an efficient quantum algorithm for estimating permanents.

A $(0, 1)$-matrix A is said to be *convertible* if there exists a matrix A' obtained by changing the signs of some of the entries in A such that $\operatorname{perm}(A) = \det(A')$. This means that $\operatorname{perm}(A)$ can be computed in polynomial time. Little (1975) provided a classification of all convertible matrices. Namely, a matrix A is convertible if and only if it can be realised as the biadjacency matrix of a bipartite graph $\mathcal{G}$ which does not contain an even subdivision J of the complete bipartite graph $K_{3,3}$ such that $\mathcal{G} - V(J)$ has a perfect matching. It is easy to construct examples of matching matrices which violate this criterion. For example, let $x = (1, 1, 1, 3)$ and $\mathcal{B} = [0, 2]^3 \times (2, \infty)$. If $\mathcal{G}$ denotes the graph constructed in the proof of Lemma 3.1, then $\mathcal{G} \cong K_{3,3} + K_{1,1}$, where the plus denotes disjoint union. Thus, $\mathcal{G}$ contains $K_{3,3}$ as a subgraph and still contains a perfect matching once this subgraph has been removed. Consequently, the corresponding matching matrix is not convertible.

The above example demonstrates that we cannot use convertibility to compute the permanents of matching matrices. However, we will show that matching matrices belong to a larger class of matrices, which we shall call *block rectangular*. We will then prove that the permanent of a block rectangular matrix can be calculated in polynomial time.

3.1 Block rectangular matrices

We begin with the definition of block rectangular matrices.

Definition 3.2. *Let k, m, n be natural numbers, where $k, m \leq n$. Let $\alpha \in \mathbb{Z}_{>0}^k$ and $\beta, \gamma \in \mathbb{Z}_{\geq 0}^{k-1}$ be three integer-valued vectors such that*

$$\sum_{r=1}^k \alpha_r = n, \quad \sum_{r=1}^{k-1} \beta_r \leq m, \quad \sum_{r=1}^{k-1} \gamma_r \leq m \tag{3.2}$$

and

$$\sum_{r=1}^t \beta_r \leq \sum_{r=1}^t \alpha_r, \quad \sum_{r=t}^{k-1} \gamma_r \leq \sum_{r=t+1}^k \alpha_r \tag{3.3}$$

for all $t = 1, \ldots, k$. The block rectangular matrix *$M = (m_{ij})$ associated to α, β, γ, m is the $m \times n$ $(0, 1)$-matrix such that $m_{ij} = 1$ if and only if there exists $t \in \{1, \ldots, k\}$ such that*

$$\sum_{r=1}^{t-1} \beta_r < i \leq m - \sum_{r=t}^{k-1} \gamma_r, \quad \sum_{s=1}^{t-1} \alpha_s < j \leq \sum_{s=1}^t \alpha_s. \tag{3.4}$$

We say that a matrix M is block rectangular *if there exist α, β, γ, m such that M is the block rectangular matrix associated to α, β, γ, m. Note that we suppress k and n in the definition as these are implicitly defined through α.*

Example. *We consider three examples of constructing a block rectangular matrix from its associated parameters, as well as an example of a matrix which is not block rectangular.*

(a) *Let $\alpha = (1, 3, 1, 1, 1), \beta = (0, 1, 2, 1), \gamma = (1, 1, 1, 0)$ and $m = 7$. This choice of α, β, γ satisfies conditions (3.2) and (3.3). Constructing the matrix from these parameters is done as follows. Firstly, $n = \sum_{i=1}^5 \alpha_i = 7$, which, together with $m = 7$, determines the dimensions of the matrix. We consider each rectangular block separately. Letting $t = 1$ in (3.4), we obtain that $0 < i \leq 4$ and $0 < j \leq 1$. Repeating this step for $t = 2, \ldots, T$ establishes the dimensions of all the rectangular blocks, resulting in the matrix shown in Figure 3a.*

(b) *A block rectangular matrix need not be square. Let $\alpha = (3, 2, 2, 1), \beta = (1, 0, 2), \gamma = (1, 1, 1), m = 6$. Then α, β, γ satisfy conditions (3.2) and (3.3), but $n = \sum_{r=1}^4 \alpha_r = 8 > 6$, so the resulting matrix, shown in Figure 3b, is not square.*

(c) *In the two examples above, we have $m = \sum_{r=1}^{k-1} \beta_r + \sum_{r=1}^{k-1} \gamma_r$. This need not be the case. Indeed, let $\alpha = (2, 3, 1, 2), \beta = (1, 0, 2), \gamma = (0, 1, 1), m = 7$. Then α, β, γ satisfy condition (3.2) and (3.3), but $\sum_{r=1}^3 \beta_r + \sum_{r=1}^3 \gamma_r = 5 < 7$. This means that the matrix contains $7 - 5 = 2$ rows of ones, as can be seen in Figure 3c. This example illustrates why we need to include m as a separate parameter in order to describe the matrix uniquely.*

(d) *The matrix in Figure 3d is not block rectangular. Indeed, if it were, then it would be associated with the parameters $\alpha = (2, 2, 2, 2), \beta = (2, 3, 1), \gamma = (3, 1, 0), m = 8$. But then we have that $\alpha_1 + \alpha_2 = 4 < 5 = \beta_1 + \beta_2$, which violates condition (3.3).*

$$
\begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}
$$

(a)

$$
\begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

(b)

$$
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}
$$

(c)

$$
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{pmatrix}
$$

(d)

Figure 3: The matrices from the example above with the contours of the rectangular blocks highlighted.

As the definition stands, it is possible for a block rectangular matrix to contain a row of zeros. Indeed, consider for instance the matrix A parametrised by $\alpha = (1,1,1)$, $\beta = (1,0)$, $\gamma = (0,1)$, $m = 1$. Then A is the 1×3 zero matrix since no i,j,t will satisfy condition (3.4). We say that a block rectangular matrix A is *complete* if it does not contain a row of zeros.

Note that multiple parametrisations will give rise to the same block rectangular matrix. Indeed, we can always subdivide a rectangular block into more rectangular blocks of equal heights. In our notation, this would mean that for some $r \in \{1, \ldots, k{-}1\}$, we have $\beta_r = \gamma_r = 0$. However, by insisting that $k = \dim(\alpha)$ should always be minimal, we obtain a unique parametrisation for every block rectangular matrix. We refer to this as the *minimal parametrisation.*

There are examples of block rectangular matrices which cannot be realised as matching matrices. For example, let $\alpha = (1,1,1), \beta = (1,1), \gamma = (1,1), m = 3$. Then A is the 3×3 identity matrix, which is not a matching matrix. However, we have the following converse result.

Proposition 3.1. *Let x be $\mathcal{B}$-admissible and let A be the matching matrix of x. Then, after permuting its columns if necessary, A is a complete block rectangular matrix.*

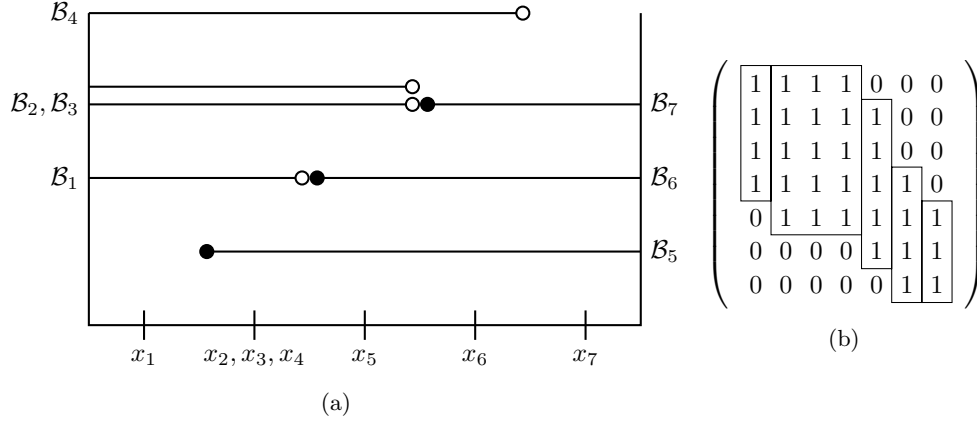*Proof.* See the Supplementary Material (Christensen, 2023). $\qquad\qquad\qquad\square$

(a)

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

(b)

Figure 4: Example of a matching of x and $\mathcal{B}$. (a) Pictorial representation of x and $\mathcal{B}$. The x_i are marked along the horizontal axis. The $\mathcal{B}_i$ are drawn along the same axis, and stacked vertically for visual clarity. (b) The resulting matching matrix, with the contours of the rectangular blocks highlighted.

Example. *Suppose we have four left-censored observations, with $s_1 = 2, s_2 = s_3 = 3, s_4 = 4$, and three right-censored observations, with $t_5 = 1, t_6 = 2, t_7 = 3$. Next, let $x = (x_1, \ldots, x_7) = (0.5, 1.5, 1.5, 1.5, 2.5, 3.5, 4.5)$. See Figure 4a. To construct the matching matrix A, we first consider x_1. We observe that $x_1 \in \mathcal{B}_1, \ldots, \mathcal{B}_4$, but not $\mathcal{B}_5, \mathcal{B}_6$ or $\mathcal{B}_7$. Hence the first column of A consists of four ones followed by three zeros. We continue this way for all the x_i, which results in the matrix in Figure 4b. Note that A is block rectangular. In fact, we recognise it as the matrix from Figure 3a.*

We are now ready to state the main result of this section.

Theorem 3.1. *Let A be an $m \times n$ complete block rectangular matrix. Then there exists an implementable algorithm for computing $\mathrm{perm}(A)$, whose computational complexity grows polynomially with n.*

Proof. See the Supplementary Material (Christensen, 2023). □

Thus, with Theorem 3.1, we are able to compute the permutation numbers needed for the estimators (2.3) and (2.3). Code for computing permanents can be found in the publicly available GitHub repository. For a reasonably large value of n, say $n = 200$, this new approach is able to compute tens of thousands of permanents of $n \times n$ block rectangular matrices within a few hours. This is in contrast with more general approaches, such as the aforementioned Ryser's formula, which would not be able to handle even a single matrix of this dimension. In the following section, we illustrate the efficiency of the new approach with experiments.

4 Experiments

We now look at two simulation studies and a real data example in order to illustrate the performance of the new estimator. The first simulation study is a tractable bioassay problem involving a Dirichlet process model. Such models were first studied by Antoniak (1974). This problem is included to verify that the new algorithm agrees with existing methods. More precisely, we will compare it with the successive substitution sampling (SSS) algorithm introduced by Doss (1994). In the second simulation study, we employ the new algorithm to fit a quantile pyramids model, introduced by Hjort and Walker (2009), to binary response data. This is an example of a process from which data samples may only be simulated approximately, but where our new algorithm nevertheless works exactly. Finally, we employ a Pitman-Yor multinomial process model (Lijoi et al., 2020) to real seroprevalence data, originally studied by Keiding et al. (1996). All code was run on a computer running Windows 11 Pro with an Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz and 16GB DDR4 RAM.

4.1 Simulations

A tractable Dirichlet process problem

For the first simulation study, we used the data reported in Table 1. This data was generated by simulating $n = 100$ points $u_1, \ldots, u_n \sim \frac{1}{3}\mathcal{N}(-2, 0.7^2) + \frac{2}{3}\mathcal{N}(1, 0.7^2)$, a mixture of two normal distributions, and observing whether these points were below or above the respective thresholds in Table 1. Thus, for example, since the number of trials at the threshold -3 was 10, the number of successes there refers to the quantity $\#\{i \in \{1, \ldots, 10\} \mid u_i \leq -3\}$. In the prior we let P be distributed according to a Dirichlet process with concentration parameter $\alpha = 1$ and base measure $\mathcal{N}(0, 1)$. The prior mean and individual realisations of the prior process are plotted in Figure 5. Using (2.3), the log marginal likelihood was calculated to be -12.861. Note that we have repeated thresholds in the data set. Repeating this calculation ten times yielded a standard deviation of 0.0137, showing that the estimate is stable. On average, it took $T = 438,606$ iterations to yield an ESS of 2000. Out of these, an average number of 411,837 yielded a vanishing permanent which could immediately be discarded. The average computation time for calculating the permanents was 6 minutes and 33 seconds. The slowest run took 6 minutes and 57 seconds.

Due to the posterior tractability of the Dirichlet process, the posterior process can be simulated directly, for instance via the SSS method, introduced by Doss (1994). Table 2 show how the new importance sampling algorithm compares with the SSS method by comparing the values of the posterior mean at various quantiles. As we can see, the two methods are in agreement. Figure 5 shows plots of the posterior mean, calculated using the two different methods, along with individual realisations of the posterior process. This plot further verifies the agreement of the two approaches, illustrating that the new algorithm indeed converges to the posterior process.

Threshold	Number of successes	Number of trials
−3	0	10
−2.33	0	10
−1.67	2	10
−1	1	10
−0.33	4	10
0.33	6	10
1	9	10
1.67	10	10
2.33	10	10
3	10	10

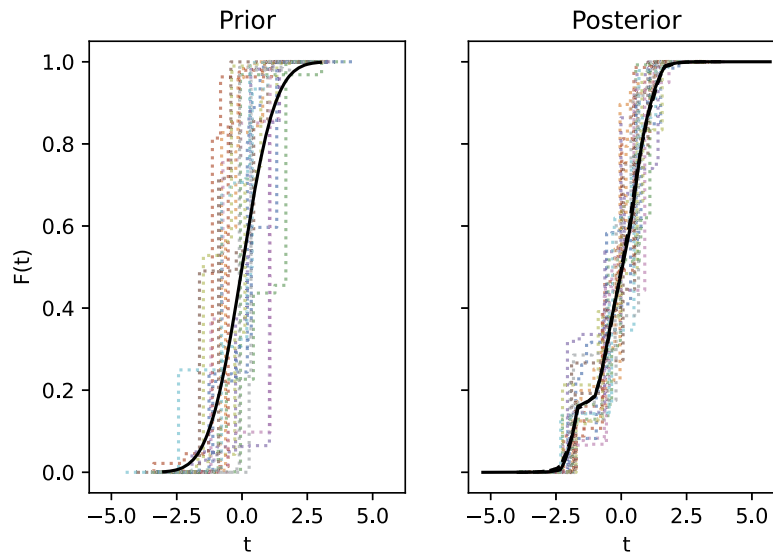Table 1: The thresholds, number of successes and number of trials for the Dirichlet process simulations.

Figure 5: Prior and posterior estimates of the Dirichlet process model given $n = 100$ binary response data points. On the left, the solid curve is the prior mean and the dotted curves are realisations of the prior process. On the right, the solid and dashed curves are the posterior mean as obtained via permutation counting or successive substitution sampling, respectively, and the dotted curves are realisations of the posterior process.

Quantile pyramids simulations

Our next simulation study is a problem in which we wish to fit a quantile pyramids model, given binary response data. Such models were first studied by Hjort and Walker (2009), and provide an appealing alternative to Pólya trees, since they avoid the specification of a partition of the sample space. More specifically, we model P as a Beta

q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
SSS	-1.851	-0.949	-0.572	-0.283	0.015	0.305	0.525	0.784	1.176
PC	-1.842	-0.950	-0.576	-0.290	0.040	0.349	0.558	0.789	1.130

Table 2: Posterior estimates of $\mathbb{E}[F^{-1}(q) \mid x \in \mathcal{B}]$ for different values of q, calculated via successive substitution sampling (SSS) and the new importance sampling algorithm based on permutation counting (PC).

quantile pyramid with parameters $(\frac{1}{2}a_m, \frac{1}{2}a_m)$, where $a_m = cm^3$ and $c = 2.5$. This is the same model as that considered in the simulation study by Hjort and Walker (2009). Given exact observations, the posterior process is intractable and so MCMC-based inference is required. As a result, there is, a priori, no straightforward way of simulating the posterior process given censored data. However, our new importance sampling algorithm will circumvent this issue.

As can be seen from (2.3), we require that it is possible to generate samples $x_i \sim P$. For the quantile pyramids model, this can only be done approximately. Indeed, the process is realised by simulating the quantiles $F^{-1}(j/2^K)$, $j = 1, \ldots, 2^K - 1$, for some finite number K. Increasing the value of K increases the precision of the realisation. If K were allowed to be infinite, then F would, by absolute continuity, be uniquely determined. Thus, for $i = 1, \ldots, n$, we could sample $x_i \sim P$ by first sampling uniform variables $u_i \sim \text{Unif}[0, 1]$ and then letting $x_i = F^{-1}(u_i)$. In practice, K is a finite number, and so this approach cannot determine the exact value of the x_i. However, by finding the numbers j_i such that $u_i \in (j_i/2^K, (j_i + 1)/2^K]$, we know that $x_i \in (F^{-1}(j_i/2^K), F^{-1}((j_i + 1)/2^K)]$. Thus, by increasing the value of K if necessary, we can make these intervals arbitrarily fine and thus know for certain whether $x_i \in \mathcal{B}_j$ for $j = 1, \ldots, n$. That is, we can sample the value of $w(x; \mathcal{B})$ *exactly*, even though x was only simulated approximately. As a result, the convergence results for the importance sampling algorithm still hold exactly.

The synthetic data were simulated as follows. For $n = 100$, we let $0 < r_1 < \cdots < r_n < 1$ be equally spaced points on the unit interval $[0, 1]$, and simulated $u_1, \ldots, u_n \sim \text{Beta}(1/2, 1)$ independently. Thus, the true underlying distribution is also the same as in the original simulation study undertaken by Hjort and Walker (2009). Writing $n_0 = \#\{i \mid u_i \leq r_i\}$, we let $t_1 < \cdots < t_{n_0}$ be those r_i such that $u_i \leq r_i$. Similarly, we let $t_{n_0+1} < \cdots < t_n$ be those r_i such that $u_i > r_i$. As in Section 2.1, we then let $\mathcal{B} = \mathcal{B}_1, \ldots, \mathcal{B}_n$, where $\mathcal{B}_i = [0, t_i]$ for $i = 1, \ldots, n_0$ and $\mathcal{B}_i = (t_i, 1]$ for $i = n_0 + 1, \ldots, n$.

Using (2.3), we calculated the log marginal likelihood to be -53.698. Performing this calculation 10 times yielded a standard deviation of 0.013, showing that the estimate is stable. On average, it took $T = 29{,}965$ iterations to obtain an effective sample size of 2000. Out of these, an average number of 8227 yielded vanishing permanents which could be discarded immediately. The average computation time for calculating the permanents was 6 minutes and 25 seconds. The slowest run took 8 minutes and 8 seconds. In Figure 6, we plot prior and posterior cdfs given the simulated data. We see that the posterior estimate has moved closer to the true cdf, and that the posterior variance is smaller than that of the prior. Indeed, Kolmogorov-Smirnov distances from the prior and the posterior means to the ground truth are 0.25 and 0.094, respectively.
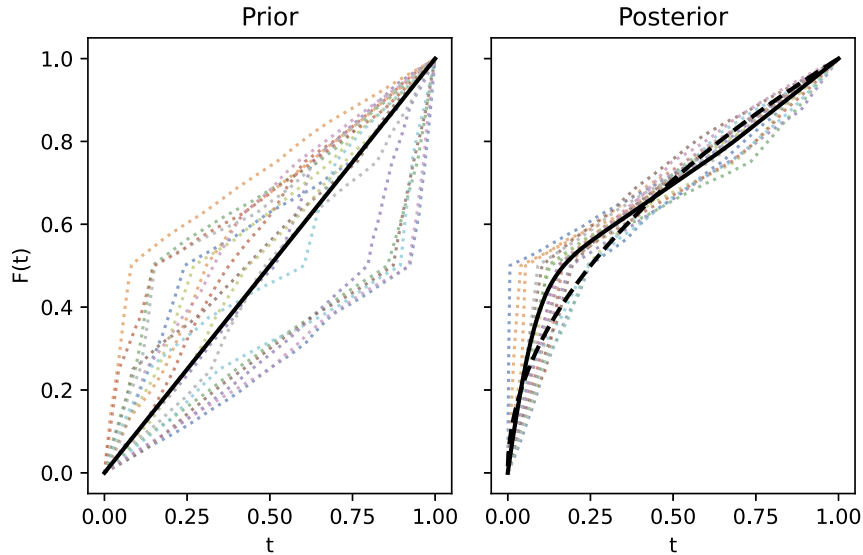
Figure 6: Prior and posterior estimates of the Beta quantile pyramid model given $n = 100$ binary response data points. On the left, the solid curve is the prior mean and the dotted curves are realisations of the prior process. On the right, the solid curve is the posterior mean, the dotted curves are realisations of the posterior process and the dashed curve is the true $Beta(1/2, 1)$ cdf.

4.2   Real current status data

We also applied the new importance sampling algorithm to a real rubella seroprevalence data set, originally studied by Keiding et al. (1996), provided by the Institute of Virology, Vienna. In this data set, the immunisation status of $n = 230$ Austrian males older than three months was tested during the period 1–25 March 1988. On a log scale, the data were scaled linearly so that the standard probit model $\mathbb{P}(y = 1) = \Phi(x)$ gave the best probit model fit. To model the time to infection, we used a Pitman-Yor multinomial (PYM) process (Lijoi et al., 2020) as prior for P. For the sake of simplicity, we used the realised probability distributions of the process to model the survival distribution directly, rather than imposing a PYM mixture model. Lijoi et al. (2020) showed that given exact observations, posterior simulation of PYM processes are possible without the use of MCMC methods, via the empirical marginalisation of a latent variable. Unfortunately, this algorithm does not apply to censored data. Although it would theoretically be possible to apply the SSS method (Doss, 1994) or similar algorithms to the PYM process model given censored data, each iteration of the sampling algorithm would require the aforementioned marginalisation. As a result, this approach would be computationally expensive and of questionable accuracy. On the other hand, since it is straightforward to generate samples from the PYM process model (Ridout, 2009; Lijoi et al., 2020), the new importance sampling algorithm can be directly applied
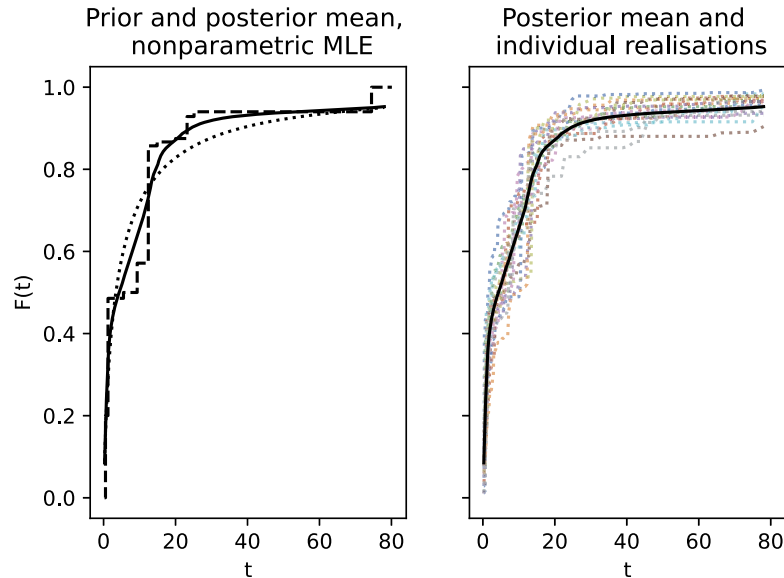
Figure 7: Posterior estimates of the Pitman-Yor multinomial process model given the Rubella data set ($n = 230$). On the left, the dotted curve is the prior mean, the solid curve is the posterior mean and the dashed curve is the nonparametric MLE. On the right, the solid curve is the posterior mean and the dotted curves are realisations of the posterior process.

for posterior inference. In the notation of Lijoi et al. (2020), we set the hyperparameters $\alpha = 2.0, \sigma = 0.7, H = 100$, along with the standard normal distribution $\mathcal{N}(0, 1)$ as base measure. These choices serve only as an illustration, and with more work, one could optimise this choice further, or impose hyperparameter priors in order to achieve a fully Bayesian approach. The log marginal likelihood was calculated to be -92.864. Reaching an ESS of 2000 required $T = 393,011$ iterations. Out of these, $219,492$ yielded a vanishing permanent and could be discarded. Calculating the permanents took 3 hours, 55 minutes and 27 seconds of computation time. In the left plot of Figure 7, we see the prior mean, the posterior mean and the nonparametric maximum likelihood estimator (MLE) (Ayer et al., 1955) of the cdf F. The nonparametric MLE is a frequentist estimator, analogous to the that by Kaplan and Meier (1958) for right censored data, as shown by Turnbull (1974). In the right plot, we see individual realisations from the posterior process, along with the posterior mean. The Kolmogorov-Smirnov distance from the prior mean and the posterior mean to the nonparametric MLE are 0.204 and 0.158, respectively, indicating an improved fit. Indeed, we see the same qualitative behaviour in the posterior mean as the penalty term models studied by Keiding et al. (1996). For the PYM process model, possible improvements may be achieved by optimising parameter choices, or indeed by introducing a similar penalty mechanism into the model.

5 Binary classification with multidimensional data

So far, we have assumed that the x_i are all one-dimensional. In this section, we show how the theory developed in the previous sections can be extended to problems where[2] $x_i \in \mathbb{R}^p$ for $p \geq 1$. This will both enable the addition of covariates, as well as provide an inference framework for binary classification models with multidimensional data.

Let $g \sim \pi(\cdot)$ be a (possibly random) function from $\mathbb{R}^p$ to $\mathbb{R}$ and let $F \sim \pi(\cdot)$ be a (possibly random) cdf on $\mathbb{R}$. We model the binary responses y_i as $y_i \mid F, g \sim$ Bernoulli($F(g(x_i))$) independently for $i = 1, \ldots, n$.

Example. *We now consider three key examples which are covered by the above setup.*

- *Let g almost surely be a neural network and let F almost surely be the sigmoid activation function $F(a) = 1/(1 + \exp(-a))$. Then the above model is the standard neural network model for binary classification (Bishop, 2006).*

- *Let g be a Gaussian process and let F almost surely be the sigmoid activation function. Then the above model is the standard Gaussian process model for binary classification (Rasmussen and Williams, 2006).*

- *Let g almost surely be a linear function with coefficients β and let F be the cdf of a random probability distribution P. Then the above model is semiparametric, and corresponds to the addition of covariates in the basic model introduced in Section 2.1.*

Again, we first provide an estimator for the marginal likelihood $\pi(y)$ of our n observations $y = (y_1, \ldots, y_n) \in \{0, 1\}^n$, which now takes the form

$$\pi(y) = \mathbb{E}\left[\prod_{i=1}^n F(g(x_i))^{y_i} \left\{1 - F(g(x_i))\right\}^{1-y_i}\right].$$

We need to introduce some notation before we can write down our estimator. Given F, let $z = (z_1, \ldots, z_n) \sim \pi(\cdot \mid F)$ be distributed such that $\mathbb{P}(z_i \leq t \mid F) = F(t)$ independently for all $i = 1, \ldots, n$. Also, given g, write $\mathcal{B}_g = \mathcal{B}_{g,1} \times \cdots \times \mathcal{B}_{g,n}$, where

$$\mathcal{B}_{g,i} = \begin{cases} (-\infty, g(x_i)] & \text{if } y_i = 1, \\ (g(x_i), \infty) & \text{if } y_i = 0. \end{cases}$$

Then,

$$\mathbb{P}(z \in \mathcal{B}_g \mid g, F) = \int_{\mathcal{B}_g} \mathrm{d}\pi(z \mid F) = \prod_{i=1}^{n_0} F(g(x_i))^{y_i}\{1 - F(g(x_i))\}^{1-y_i}. \tag{5.1}$$

---

[2]In generative models for binary classification, there is also a distribution for the inputs x_i. However, for evaluation of the marginal likelihood and posterior inference of the hyperparameters, we condition on these input values, effectively treating them as constant. Hence, everything in this section also applied to generative models, but we omit conditioning on the value of x_i for the sake of clarity.

Now consider the following estimator.

$$\hat{\pi}_T^{\text{IS}}(y) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{n!} w(z^{(t)}; \mathcal{B}_{g^{(t)}}), \tag{5.2}$$

where $z^{(t)} \sim \pi(\cdot \mid F^{(t)})$, $F^{(t)} \sim \pi(\cdot)$ and $g^{(t)} \sim \pi(\cdot)$.

Proposition 5.1. *The statistic $\hat{\pi}_T^{\text{IS}}(y)$ is an unbiased and consistent estimator for the marginal likelihood $\pi(y)$.*

Proof. Using double expectation, we have that

$$\mathbb{E}\left[\frac{1}{n!} w(z; \mathcal{B}_g)\right] = \frac{1}{n!}\mathbb{E}\left[\mathbb{E}[w(z; \mathcal{B}_g) \mid g, F]\right] = \frac{1}{n!}\mathbb{E}\left[\int_{\text{Orb}(\mathcal{B}_g)} w(z; \mathcal{B}_g)\pi(\mathrm{d}z \mid F)\right]$$

$$= \frac{1}{n!}\mathbb{E}\left[\sum_{\sigma \in S_n} \int_{\sigma(\mathcal{B}_g)} \pi(\mathrm{d}z \mid F)\right] = \mathbb{E}\left[\int_{\mathcal{B}_g} \pi(\mathrm{d}z \mid F)\right]$$

$$= \mathbb{E}\left[\prod_{i=1}^{n} F(g(x_i))^{y_i}\{1 - F(g(x_i))\}^{1-y_i}\right] = \pi(y),$$

where the penultimate equality follows from (5.1). This proves that the estimator is unbiased. Applying the law of large numbers to (5.2) establishes consistency. □

As in Section 2.3, it is possible to extend this result to a normalised importance sampling estimator for general posterior inference for g and P.

6 Discussion

We conclude the paper with a few points of discussion which shed light on directions for future work and further improvements.

In this paper, we have only considered binary responses, which for one-dimensional inputs corresponds to left and right censored observations. This is because such data yield block rectangular matching matrices, whose permanents are computable in polynomial time. However, for more complicated observations, such as interval censored data or polychotous responses (as opposed to binary), it is easy to construct examples of matching matrices which are not block rectangular. Hence, in order to apply the methods developed in this paper to such problems, it is necessary to develop an efficient and accurate estimation procedure for the permanents of the corresponding matching matrices. Further work in this direction is encouraged.

We have proved the consistency of the new estimator (2.4), in the sense that as $T \to \infty$, this converges to the posterior mean $\mathbb{E}[\theta \mid x \in \mathcal{B}]$. However, a separate question is whether this posterior mean itself is consistent. For parametric models, this is guaranteed via the Bernstein–von Mises theorem, which asserts consistency of the posterior mean and links Bayesian credibility sets with frequentist confidence intervals.

In contrast, Doss (1985a,b) and Diaconis and Freedman (1986) showed that for nonparametric models, there exist examples of reasonable choices of priors which lead to inconsistent posteriors. Hence, consistency does not automatically apply in Bayesian nonparametrics. Multiple positive consistency results have since been established for specific choices of nonparametric priors (Brunner and Lo, 1996; Ghosal et al., 1999), some of which also allow for censored data (Kim and Lee, 2004; De Blasi et al., 2009; Camerlenghi et al., 2021; Jongbloed et al., 2022). In general, the issue of consistency in Bayesian nonparametrics should be considered only a partially resolved question, especially for problems involving censored data. Further research in this area is needed to answer to which extent the asymptotic theory of the frequentist nonparametric MLE (Ayer et al., 1955; Groeneboom and Jongbloed, 2014) transfers to the Bayesian nonparametric setting.

Supplementary Material

Supplementary Material for "Inference for Bayesian nonparametric models with binary response data via permutation counting" (DOI: 10.1214/22-BA1353SUPP; .pdf). Proofs of Proposition 2.2, Proposition 3.1 and Theorem 3.1.

References

Albert, J. H. and Chib, S. (1993). "Bayesian analysis of binary and polychotomous response data." *Journal of the American Statistical Association*, 88(422): 669–679. MR1224394. 293

Antoniak, C. E. (1974). "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems." *The Annals of Statistics*, 2(6): 1152–1174. MR0365969. 294, 306

Arbel, J., De Blasi, P., and Prünster, I. (2019). "Stochastic approximations to the Pitman–Yor process." *Bayesian Analysis*, 14(4): 1201–1219. MR4136558. doi: https://doi.org/10.1214/18-BA1127. 295

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955). "An empirical distribution function for sampling with incomplete information." *The Annals of Mathematical Statistics*, 26(4): 641–647. MR0073895. doi: https://doi.org/10.1214/aoms/1177728423. 310, 313

Bhattacharya, P. K. (1981). "Posterior distribution of a Dirichlet process from quantal response data." *The Annals of Statistics*, 9(4): 803–811. MR0619283. 294

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Cambridge: Springer. MR2247587. doi: https://doi.org/10.1007/978-0-387-45528-0. 311

Bollobás, B. (1979). *Graph Theory: An Introductory Course*, volume 63 of *Graduate Texts in Mathematics*. New York: Springer. MR0536131. 302

Brunner, L. J. and Lo, A. Y. (1996). "Limiting posterior distributions under mixture of conjugate priors." *Statistica Sinica*, 6(1): 187–197. MR1379056. 313

Cai, B., Lin, X., and Wang, L. (2011). "Bayesian proportional hazards model for current status data with monotone splines." *Computational Statistics & Data Analysis*, 55(9): 2644–2651. MR2802342. doi: https://doi.org/10.1016/j.csda.2011.03.013. 294

Camerlenghi, F., Lijoi, A., and Prünster, I. (2021). "Survival analysis via hierarchically dependent mixture hazards." *The Annals of Statistics*, 49(2): 863–884. MR4255111. doi: https://doi.org/10.1214/20-aos1982. 313

Campbell, T., Huggins, J. H., How, J. P., and Broderick, T. (2019). "Truncated random measures." *Bernoulli*, 25(2): 1256–1288. MR3920372. doi: https://doi.org/10.3150/18-bej1020. 295

Cappé, O., Godsill, S. J., and Moulines, E. (2007). "An overview of existing methods and recent advances in sequential Monte Carlo." In *Proceedings of the IEEE*, volume 95, 899–924. 300

Cappé, O., Guillin, A., Marin, J.-M., and Robert, C. P. (2004). "Population Monte Carlo." *Journal of Computational and Graphical Statistics*, 13(4): 907–929. MR2109057. doi: https://doi.org/10.1198/106186004X12803. 300

Christensen, D. (2022). "Nonparametric Bayesian sensitivity testing with optimal design." In *Proceedings of the 51st International Annual Conference of the Fraunhofer ICT*. Fraunhofer ICT. 293

Christensen, D. (2023). Supplementary Material for "Inference for Bayesian Nonparametric Models with Binary Response Data via Permutation Counting". *Bayesian Analysis*. doi: https://doi.org/10.1214/22-BA1353SUPP. 298, 304, 305

Cox, D. R. (1972). "Regression models and life-tables." *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2): 187–202. MR0341758. 294

De Blasi, P., Peccati, G., and Prünster, I. (2009). "Asymptotics for posterior hazards." *The Annals of Statistics*, 37(4): 1906–1945. MR2533475. doi: https://doi.org/10.1214/08-AOS631. 313

Diaconis, P. and Freedman, D. (1986). "On the consistency of Bayes estimates." *The Annals of Statistics*, 14(1): 1–26. MR0829555. doi: https://doi.org/10.1214/aos/1176349830. 313

Dixon, W. J. and Mood, A. M. (1948). "A Method for obtaining and analyzing sensitivity data." *Journal of the American Statistical Association*, 43(241): 109–126. 293

Doss, H. (1985a). "Bayesian nonparametric estimation of the median; part I: Computation of the estimates." *The Annals of Statistics*, 13(4): 1432–1444. MR0811501. doi: https://doi.org/10.1214/aos/1176349746. 313

Doss, H. (1985b). "Bayesian nonparametric estimation of the median; part II: Asymptotic properties of the estimates." *The Annals of Statistics*, 13(4): 1445–1464. MR0811502. doi: https://doi.org/10.1214/aos/1176349747. 313

Doss, H. (1994). "Bayesian nonparametric estimation for incomplete data via succes-

sive substitution sampling." *The Annals of Statistics*, 22(4): 1763–1786. MR1329167. doi: https://doi.org/10.1214/aos/1176325756.   294, 306, 309

Escobar, M. D. and West, M. (1995). "Bayesian density estimation and inference using mixtures." *Journal of the American Statistical Association*, 90(430): 577–588. MR1340510.   294

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 1(2): 209–230. MR0350949.   294

Ferguson, T. S. (1983). "Bayesian density estimation by mixtures of normal distributions." In *Recent advances in statistics. Papers in honor of Herman Chernoff on his sixtieth birthday*, 287–302. Bibliohound. MR0736538.   294

Gelfand, A. E. and Kuo, L. (1991). "Nonparametric Bayesian bioassay including ordered polytomous response." *Biometrika*, 78(3): 657–666. MR1130934. doi: https://doi.org/10.1093/biomet/78.3.657.   294

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (1999). "Posterior consistency of Dirichlet mixtures in density estimation." *The Annals of Statistics*, 27(1): 143–158. MR1701105. doi: https://doi.org/10.1214/aos/1018031105.   313

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation under Shape Constraints*. New York: Cambridge University Press. MR3445293. doi: https://doi.org/10.1017/CBO9781139020893.   293, 313

Hesterberg, T. (1995). "Weighted average importance sampling and defensive mixture distributions." *Technometrics*, 37(2): 185–194.   300

Hjort, N. L. (1990). "Nonparametric Bayes estimators based on Beta processes in models for life history data." *The Annals of Statistics*, 18(3): 1259–1294. MR1062708. doi: https://doi.org/10.1214/aos/1176347749.   294

Hjort, N. L. and Walker, S. G. (2009). "Quantile pyramids for Bayesian nonparametrics." *The Annals of Statistics*, 37(1): 105–131. MR2488346. doi: https://doi.org/10.1214/07-AOS553.   294, 306, 307, 308

Huh, J. (2022). "A fast quantum algorithm for computing matrix permanent." *ArXiv preprint. Available at arXiv:*2205.01328.   302

Ishwaran, H. and James, L. F. (2001). "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96(453): 161–173. MR1952729. doi: https://doi.org/10.1198/016214501750332758.   294

Jongbloed, G., van der Meulen, F. H., and Pang, L. (2022). "Bayesian nonparametric estimation in the current status continuous mark model." *Scandinavian Journal of Statistics*, 49(3): 1329–1352. MR4471288. doi: https://doi.org/10.1111/sjos.12562.   313

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). "Slice sampling mixture models." *Statistics and Computing*, 21: 93–105. MR2746606. doi: https://doi.org/10.1007/s11222-009-9150-y.   294

Kaplan, E. L. and Meier, P. (1958). "Nonparametric estimation from incomplete observations." *Journal of the American Statistical Association*, 53(282): 457–481. MR0093867. 310

Keiding, N., Begtrup, K., Scheike, T. H., and Hasibeder, G. (1996). "Estimation from current-status data in continuous time." *Lifetime Data Analysis*, 2: 119–129. 293, 306, 309, 310

Kim, Y. and Lee, J. (2004). "A Bernstein–von Mises theorem in the nonparametric right-censoring model." *The Annals of Statistics*, 32(4): 1492–1512. MR2089131. doi: https://doi.org/10.1214/009053604000000526. 313

Kuo, L. (1988). "Linear Bayes estimators of the potency curve in bioassay." *Biometrika*, 75(1): 91–96. MR0932821. doi: https://doi.org/10.1093/biomet/75.1.91. 294

Lenk, P. J. (1988). "The logistic normal distribution for Bayesian, nonparametric, predictive densities." *Journal of the American Statistical Association*, 83(402): 509–516. MR0971380. 294

Lenk, P. J. (1991). "Towards a practicable Bayesian nonparametric density estimator." *Biometrika*, 78(3): 531–543. MR1130921. doi: https://doi.org/10.1093/biomet/78.3.531. 294

Leonard, T. (1978). "Density estimation, stochastic processes and prior information." *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(2): 113–132. MR0517434. 294

Lijoi, A., Mena, R. H., and Prünster, I. (2005). "Hierarchical mixture modeling with normalized inverse–Gaussian priors." *Journal of the American Statistical Association*, 100(472): 1278–1291. MR2236441. doi: https://doi.org/10.1198/016214505000000132. 294

Lijoi, A., Mena, R. H., and Prünster, I. (2007). "Controlling the reinforcement in Bayesian non-parametric mixture models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4): 715–740. MR2370077. doi: https://doi.org/10.1111/j.1467-9868.2007.00609.x. 294

Lijoi, A., Prünster, I., and Rigon, T. (2019). "Finite-dimensional discrete random structures and Bayesian clustering." *Collegio Carlo Alberto Working Paper*, No. 600. 294, 295

Lijoi, A., Prünster, I., and Rigon, T. (2020). "The Pitman–Yor multinomial process for mixture modelling." *Biometrika*, 107(4): 891–906. MR4186494. doi: https://doi.org/10.1093/biomet/asaa030. 294, 295, 306, 309, 310

Little, C. H. C. (1975). "A characterization of convertible (0, 1)-matrices." *Journal of Combinatorial Theory, Series B*, 18(3): 187–208. MR0424583. doi: https://doi.org/10.1016/0095-8956(75)90048-9. 302

Lo, A. Y. (1984). "On a class of Bayesian nonparametric estimates: I. Density estimates." *The Annals of Statistics*, 12(1): 351–357. MR0733519. doi: https://doi.org/10.1214/aos/1176346412. 294

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). "Approximate Bayesian computational methods." *Statistics and Computing*, 22: 1167–1180. MR2992292. doi: `https://doi.org/10.1007/s11222-011-9288-2`. 294

Muliere, P. and Secchi, P. (1995). "A note on a proper Bayesian bootstrap." Technical report, Dipartimento di economia politica e metodi quantitativi, Universita degli studi di Pavia. 295

Murray, I., MacKay, D., and Adams, R. P. (2008). "The Gaussian process density sampler." In *Advances in Neural Information Processing Systems*, volume 21. 294

Neal, R. M. (2000). "Markov chain sampling methods for Dirichlet process mixture models." *Journal of Computational and Graphical Statistics*, 9(2): 249–265. MR1823804. doi: `https://doi.org/10.2307/1390653`. 294

Newton, M. A. and Zhang, Y. (1999). "A recursive algorithm for nonparametric analysis with missing data." *Biometrika*, 86(1): 15–26. MR1688068. doi: `https://doi.org/10.1093/biomet/86.1.15`. 294

Neyer, B. T. (1994). "A D-optimality-based sensitivity test." *Technometrics*, 36(1): 61–70. 293

Orbanz, P. and Teh, Y. W. (2011). "Bayesian nonparametric models." In Sammut, C. and Webb, G. I. (eds.), *Encyclopedia of Machine Learning*. Boston, Massachusetts: Springer. 294

Paisley, J. W., Zaas, A. K., Woods, C. W., Ginsburg, G. S., and Carin, L. (2010). "A stick-breaking construction of the Beta process." In *International Conference on Machine Learning*, 847–854. PMLR. 294

Papaspiliopoulos, O. and Roberts, G. O. (2008). "Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models." *Biometrika*, 95(1): 169–186. MR2409721. doi: `https://doi.org/10.1093/biomet/asm086`. 294

Paulon, G., Müller, P., and Rosas, V. G. S. Y. (2020). "Bayesian nonparametric bivariate survival regression for current status data." *ArXiv preprint. Available at arXiv:2009.06460*. 294

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, Massachusetts: MIT Press. MR2514435. 311

Regazzini, E., Lijoi, A., and Prünster, I. (2003). "Distributional results for means of normalized random measures with independent increments." *The Annals of Statistics*, 31(2): 560–585. MR1983542. doi: `https://doi.org/10.1214/aos/1051027881`. 294

Ridout, M. S. (2009). "Generating random numbers from a distribution specified by its Laplace transform." *Statistics and Computing*, 19: 439–450. MR2565316. doi: `https://doi.org/10.1007/s11222-008-9103-x`. 309

Ryser, H. J. (1963). *Combinatorial Mathematics*, volume 14 of *Carus Mathematical Monographs*. American Mathematical Society. MR0150048. 302

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 4(2): 639–650. MR1309433.   294

Turnbull, B. W. (1974). "Nonparametric estimation of a survivorship function with doubly censored data." *Journal of the American Statistical Association*, 69(345): 169–173. MR0381120.   310

Valiant, L. G. (1979). "The complexity of computing the permanent." *Theoretical Computer Science*, 8(2): 189–201. MR0526203. doi: https://doi.org/10.1016/0304-3975(79)90044-6.   295, 302

Walker, S. G. (2007). "Sampling the Dirichlet mixture model with slices." *Communications in Statistics - Simulation and Computation*, 36(1): 45–54. MR2370888. doi: https://doi.org/10.1080/03610910601096262.   294

Wang, N., Wang, L., and McMahan, C. S. (2015). "Regression analysis of bivariate current status data under the Gamma-frailty proportional hazards model using the EM algorithm." *Computational Statistics & Data Analysis*, 83: 140–150. MR3281802. doi: https://doi.org/10.1016/j.csda.2014.10.013.   294

Acknowledgments

Paper II

perms: Likelihood-free estimation of marginal likelihoods for binary response data in Python and R

Paper III

Models for predicting impact sensitivity of energetic materials based on the trigger linkage hypothesis and Arrhenius kinetics

ORIGINAL PAPER

Models for predicting impact sensitivity of energetic materials based on the trigger linkage hypothesis and Arrhenius kinetics

Tomas L. Jensen[1] · John F. Moxnes[1] · Erik Unneberg[1] · Dennis Christensen[1]

Abstract

In order to predict the impact sensitivity of high explosives, we designed and evaluated several models based on the trigger linkage hypothesis and the Arrhenius equation. To this effect, we calculated the heat of detonation, temperature of detonation, and bond dissociation energy for 70 energetic molecules. The bond dissociation energy divided by the temperature of detonation proved to be a good predictor of the impact sensitivity of nitroaromatics, with a coefficient of determination (R^2) of 0.81. A separate Bayesian analysis gave similar results, taking model complexity into account. For nitramines, there was no relationship between the impact sensitivity and the bond dissociation energy. None of the models studied gave good predictions for the impact sensitivity of liquid nitrate esters. For solid nitrate esters, the bond dissociation energy divided by the temperature of detonation showed promising results ($R^2 = 0.85$), but since this regression was based on only a few data points, it was discredited when model complexity was accounted for by our Bayesian analysis. Since the temperature of detonation correlated with the impact sensitivity for nitroaromatics, nitramines, and nitrate esters, we consider it to be one of the leading predictive factors of impact sensitivity for energetic materials.

Introduction

By using quantum chemical and thermodynamic calculations, new energetic molecules can be designed and characterized in terms of their geometry, density, and performance as explosives and propellants [1–5]. It is important to predict the sensitivity of energetic materials, but despite considerable efforts made during the last decades, developing a reliable and general method is still challenging [6–11]. By the sensitivity of an energetic material, we refer to its susceptibility to initiate due to external thermal, mechanical, or electrostatic stimuli. The study of the underlying causes that govern sensitivity is not only important for understanding liquid and solid-state phenomena in general but above all for ensuring safe handling, transport, and storage of energetic materials.

One of the most well-known measures of the sensitivity of an explosive is its impact sensitivity, which is determined by dropping a mass upon the sample, measuring the critical point at which a pre-decided fixed percentage of the drops will lead to an explosion. We refer to this critical point as the *critical impact level* of the material, which may either be given as the critical height (cm) or potential energy (J). Hence, the impact sensitivity and the critical impact level are inversely correlated.

The impact sensitivity is related to macroscopic parameters such as particle size, crystals defects, polymorphism, and crystal orientation. Defects play a particularly prominent role since they form hot-spots under fast compression of the material. The initiation process can be divided into two steps. First, the material is compressed and deformed, leading to heating of the hot-spots. In the second step, the material inside and surrounding the hot-spots self-ignites and propagates into an explosion, provided that the hot-spot temperatures are sufficiently high. The critical temperature

✉ Dennis Christensen
dennis.christensen@ffi.no

John F. Moxnes
john-f.moxnes@ffi.no

[1] Defence Systems Division, Norwegian Defence Research Establishment, P.O. Box 25, N-2027, Kjeller, Norway

at which the explosive self-ignites during impact has been measured to be between 390 and 1060° C [12, 13].

There are many operational factors which also affect the critical impact level measured, such as the type of fallhammer used, the test procedure itself, and operator-related judgment of explosion/no-explosion. In addition, the impact sensitivity will depend on the thickness and size of the sample [14, 15]. In general, the measured critical impact level depends on a variety of experimental factors in addition to molecular-related properties. Consequently, predicting impact sensitivities with reasonable accuracy appears to be too subtle a problem to be explained by a model based on fallhammer measurements alone.

Additional key factors responsible for the sensitivity of an energetic material include the molecular properties related to the kinetics and the thermodynamics of the decomposition reactions. Numerous studies have been carried out in order to correlate the impact sensitivity with properties like heat of detonation [16–18], detonation velocity [19], bond dissociation energy [22, 23], oxygen balance [24], electrostatic potential of the molecular surface [25–27], band gap [28], ^{15}N NMR chemical shift [29], "doorway modes" in the region 200–1000 cm^{-1} [30], and free space in the crystal lattice [31]. More recent studies focusing on physical factors report that there are only weak correlations or trends between the impact sensitivity and heat of detonation, electrostatic potential, and free space in the crystal lattice [7, 8, 17].

Due to the complexity of initiation of the decomposition pointed out by Dlott, care must be taken before drawing mechanistic conclusions based on simple correlation studies [32]. Moreover, if a study is based on too few compounds to make conclusive judgments, we risk asserting accidental correlations [33]. Using a large set of molecules, Keshavarz et al. derived models based on the CHNO ratios and different molecular moieties [34, 35]. Quantitative structure-property relationship (QSPR) models have also been developed for large sets of molecules [36, 37], and seem to be able to predict the impact sensitivity with reasonable accuracy. However, unlike models based on physical factors, these QSPR models do not reveal much information about the intrinsic factors that govern the impact sensitivity. Since they generally contain a surfeit of adjustable parameters, they are also prone to over-fitting.

Models for predicting the impact sensitivity based on the Arrhenius equations and the thermodynamics of the decomposition were introduced in the 1940s and 1950s [12]. In a study of 15 molecules, Wu et al. showed that the ratio between the dissociation energy of the weakest –NO$_2$ bond and the heat of decomposition correlated with the impact sensitivity [1, 44]. This approach was refined by Mathieu et al., investigating models based on larger sets of molecules [38–40]. For various families of energetic materials, they found correlations between the impact sensitivity and the bond dissociation energy divided by the decomposition energy. In these models, the enthalpies of formation were either neglected or calculated with a simplified method, and the decomposition energies were computed assuming that the energetic molecule decomposed to H_2O–CO_2 arbitrarily [38–40]. Instead of calculating the bond dissociation energies for each molecule separately, the –NO$_2$ bond dissociation energy was assigned to a constant value by considering into which family of energetic materials the molecule belonged, along with the functional groups in the neighboring position of the nitro group [38–40].

In this work, we report density functional theory (DFT) calculations of the –NO$_2$ bond dissociation energies and thermodynamic calculations of the heat and temperature of detonation for 70 energetic molecules. We then apply our results to investigate how these properties can predict impact sensitivity. All –NO$_2$ bond dissociation energies in the molecules are calculated separately. When overlap between the data sets is accounted for, we reach a total of 91 data points on which our regression models are based. Impact sensitivity models based on bond dissociation energies, heats of detonation, detonation temperatures, and total energies are evaluated for these molecules. 1,3,5-Triamino-2,4,6-trinitrobenzene (TATB) is known for its low sensitivity, and has been frequently used in models for predicting impact sensitivities [18, 19, 34, 37, 38]. However, its critical impact level (490 cm; 2.5 kg drop weight) is not a measured value but an estimate based on extrapolating the measured critical heights and oxygen balances of only three energetic molecules. The measured critical height of TATB is only reported to be higher than 320 cm [41]. We will therefore use our most promising model to make a more accurate prediction of this value.

Theory and methods

### Modeling the critical impact level

We model the critical impact level of an energetic molecule as a continuous random variable I with the property that its (natural) logarithm $\log I$ is governed by a normal distribution with mean μ and variance σ^2, so that

$$\log I \sim \mathcal{N}\left(\mu, \sigma^2\right). \tag{1}$$

In general, μ will depend on the individual choice of molecule, whereas σ^2 is assumed to be constant across families of molecules. The variance will depend on the level of statistical noise in the data set under consideration, which is largely due to experimental inaccuracies. However, when modeling quantum mechanical phenomena, genuine

randomness in nature may also have an effect on the measurements. The Bruceton method and UN test procedure are among the most common schemes for measuring critical impact levels. The former gives the impact energy level I_{50} (J) or height H_{50} (cm) at which 50% of the test samples are expected to explode, whereas the latter gives the impact energy level $I_{1:6}$ (J) or height $H_{1:6}$ (cm) that results in an explosion for at least one in six test drops [42, 43]. For dimension reasons, we introduce a reference value I^0 of 1 J. We let the critical impact level I denote either I_{50}/I^0 or $I_{1:6}/I^0$, depending on whether the data set upon which the regression is based is in accordance with the Bruceton method or the UN test procedure.

We now motivate our choice of models. According to the hot-spot theory, when an energetic material is subjected to a mechanical impact, material deformation will increase the hot-spot temperatures. If this temperature is above a critical level, the molecule will decompose. The trigger linkage hypothesis states that the first step in the initiation of an energetic molecule is a bond cleavage. The decomposition is triggered by the homolytic fission of an A–NO$_2$ bond, and so the reaction is given by

$$A\text{–}NO_2 \rightarrow (A\text{–}NO_2)^\star \rightarrow A\cdot + \cdot NO_2, \tag{2}$$
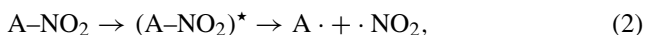
where the star $\star$ denotes the transition state and A is either C, N, or O.

At high temperatures, similar to what the material is exposed to by impact or shock, the C–NO$_2$ bond dissociation is the dominant reaction in the initial decomposition phase for nitroaromatic molecules [33]. However, at lower temperatures, reactions involving the other functional groups on the aromatic ring may occur. Furthermore, auto-catalyzed reactions and self-heating of the material because of exothermal reactions determine the rate of reaction r in the next decomposition phase. The kinetic theory attributed to Arrhenius dictates that r is inversely exponentially dependent on the activation energy E_a (the energy required to transform the reactant into the transition state), giving us

$$-\frac{d\,[A\text{–}NO_2]}{dt} = r = c\,[A\text{–}NO_2]^n \exp\left(-\frac{E_a}{RT}\right), \tag{3}$$

where $[A\text{–}NO_2]$ is the molar concentration of A–NO$_2$, t is time, c is a constant (the pre-exponential factor), n is the reaction order, R is the molar gas constant, and T is the absolute temperature. Since it has been assumed that the rate of reaction given by the Arrhenius equation correlates negatively with the impact sensitivity of an energetic material [1, 12, 38, 44], we make the assumption that these quantities are inversely proportional. That is, the

sum $\mu + \log r$ is constant. Combining this with taking logarithms on both sides of Eq. 3, we get that

$$-\mu = c + n \log\,[A\text{–}NO_2] - \frac{E_a}{RT}. \tag{4}$$

When $[A\text{–}NO_2]$ and n are assumed to be constant, Eq. 4 takes the form

$$\mu = c_1 + \frac{E_a}{RT}, \tag{5}$$

where c_1 is a constant.

Equation 5 forms the basis for our models. In order to calculate μ, we first need to calculate E_a and T. Unfortunately, these parameters are difficult to determine. Even though E_a can be calculated by quantum mechanical methods, several transition states and different decomposition routes need to be considered, making the calculations very time-consuming. For this reason, we have evaluated various approximation schemes for E_a and T. In our first three models, we assume that the activation energy in Eq. 5 is constant.

The Arrhenius law requires a particular temperature. During the complex sequence of events leading to an explosion, the temperature in the surroundings of a decomposed molecule will deviate from the ambient temperature. A small number of neighboring molecules are envisioned to decompose and release energy, increasing the local temperature. If the decomposition reaction produces more heat than is lost in conjunction with the heating of the nearby species, heat convection, and conduction to the surroundings, the temperature will rise rapidly. If this occurs, the reaction may propagate into an explosion. Therefore, T is a local temperature, varying in space and time during these events. In this picture, the more energy released during decomposition, the higher the local temperature. Consequently, the heat of detonation Q (kJ dm^{-3}) is assumed to be proportional to T. This gives our first model, which was studied in References [1, 7, 8, 16–18, 44], namely

$$\mu = c_1 + \frac{c_2}{Q}. \tag{6}$$

Here, the constants c_1, c_2 are fitted to critical impact level data by linear regression and Q is calculated in EXPLO5 [45], which uses the chemical formula, the enthalpy of formation, and density to calculate the detonation properties. The enthalpies of formation and densities are taken from References [46–49]. These references do not contain enthalpy of formation or density for all the energetic materials we shall consider, so for the remaining molecules, these parameters are estimated using the method described in References [50–54].

Song et al. assumed that the total energy E_{total} (Hartree) of an energetic molecule correlates with the energy release of the decomposition reaction [20, 21]. This motivates the

assumption that E_{total} and T are proportional, giving us the model

$$\mu = c_1 + \frac{c_2}{E_{total}}, \tag{7}$$

where E_{total} is calculated by Gaussian09 [55]. The zero-point energy is not included in this calculation.

A new model, to our knowledge not discussed in the literature, results from assuming that T in Eq. 5 is proportional to the detonation temperature T_{ex} (K) of the material, leading us to

$$\mu = c_1 + \frac{c_2}{T_{ex}}, \tag{8}$$

where T_{ex} is calculated in EXPLO5.

The models above require a constant E_a for the homolytic cleavage of the A–NO_2 bond for each class of energetic material. This is clearly a rough approximation. It has been proposed that E_a is proportional to the bond dissociation energy BDE (kJ mol^{-1}), that is, the energy required to break the trigger-linkage [1, 38–40, 44]. However, a proportional relationship between BDE and E_a is likely to hold only for compounds where the resonance stabilization and the structure of the transition states are relatively similar. Khrapkovskii et al. reported a significant correlation between the measured value of E_a for the C–NO_2 homolysis in nitroaromatic substances with different substituents and the values of BDE calculated by the hybrid DFT functional B3LYP and a small basis set 6-31G(d,p), with a coefficient of determination R^2 of 0.72 [56]. This motivates the assumption that E_a and BDE are proportional.

Our fourth model in question was studied in References [22, 23]. It is based on BDE alone, and reads

$$\mu = c_1 + c_2 (BDE), \tag{9}$$

where we have assumed that T in Eq. 5 is constant. By including BDE and the approximations used in Eqs. 6, 7, and 8 for the local temperature into Eq. 5, we arrive at our final three models of consideration, which take the form

$$\mu = c_1 + c_2 \left(\frac{BDE}{Q} \right), \tag{10}$$

$$\mu = c_1 + c_2 \left(\frac{BDE}{E_{total}} \right), \tag{11}$$

$$\mu = c_1 + c_2 \left(\frac{BDE}{T_{ex}} \right). \tag{12}$$

Equations 10 and 11 were studied in References [1, 19, 38–40, 44] and [20, 21], respectively.

We study three families of energetic molecules: nitroaromatics, nitramines, and nitrate esters. For each family, we make a choice of which bond rupture we believe to be the key step in the initiation process. We choose C–NO_2,

N–NO_2, and O–NO_2 for nitroaromatics, nitramines, and nitrate esters, respectively.

Density functional theory calculations

Our original intention was to optimize the geometry of molecules and radicals with the M06 functional and the 6-311+G(2d,p) basis set, since M06 is reported to calculate homolytic dissociation of C–NO_2 bonds accurately [57]. However, we were not able to calculate several hundred values of BDE with this choice of functional and basis set due to limited computer power. In order to avoid this difficulty, we instead chose the B3LYP functional, which is widely used in optimizing the geometry of energetic materials. This functional is known to systematically undershoot the value of BDE for C–NO_2 [57], but Khrapkovskii et al. have shown that with a small basis set (6-31G(d,p)), it calculates BDE for substituted nitroaromatics with similar accuracy as wB97xd/6-31+G(2df,p), G2, G3, G3B3 and CBS-QB3 [56]. In Table 1, we show how the calculated value of the C–NO_2 BDE for nitrobenzene converges by increasing the size of the basis set. This value has been measured to be 298.7 kJ mol^{-1} [58] and 314.5 kJ mol^{-1} [59].

The BDE values in Table 1 are calculated according to the method in Reference [60]. Table 1 also illustrates how the B3LYP functional undershoots the BDE values, but it should be borne in mind that in our models, differences in BDE are more important than the particular values they take. The calculation is defined by

$$BDE = E(A\cdot) + E(NO_2\cdot) - E(A-NO_2), \tag{13}$$

where $E(A\cdot)$, $E(NO_2\cdot)$, and $E(A-NO_2)$ denote the ground state electronic energies (open shell model) of the species A·, NO_2· and A–NO_2, respectively. In Eq. 13, the zero-point energy is neglected since Song et al. have shown that it bears no important role for the correlation between I_{50} and BDE/E_{total} [20].

Table 1 The C–NO_2 bond dissociation energy BDE (kJ mol^{-1}) for nitrobenzene calculated with the B3LYP and M06 functional at 298 K by using different basis sets

Basis set	B3LYP	M06
6-31G	301.3	329.9
6-31G(d)	290.1	312.6
6-31G(d,p)	290.2	312.6
6-311G(d,p)	278.7	299.7
6-31+G(d,p)	282.6	305.4
6-311+G(d,p)	276.3	297.5
6-311+G(2df,2p)	277.3	299.6

Experimental measurements

Since variation in the measurements of critical impact level causes difficulties in parametrizing and validating models, it is important to keep observational uncertainties to a minimum when conducting experiments. The critical impact level data for the nitroaromatic materials is obtained from the Wilson et al. data set, where the tests were performed in the same laboratory with the same equipment and according to the same test procedure [48]. Wilson et al. also ensured that the molecules considered had a similar particle distribution, and so we find this data set to be the most useful one for our purposes. It should be noted that the critical height of 1,3-diamino-2,4,6-trinitrobenzene (DATB) and TATB are only given as a lower limit below which they did not explode, namely $H_{50} > 200$ cm for both molecules. Therefore, these values are not included in the training set from Wilson et al. In our analysis, we also consider two other data sets for nitroaromatics in order to reduce the risk of asserting any accidental correlations. These are taken from Storm et al. [41] and Meyer et al. [47]. The critical impact levels for the nitramines and nitrate esters are obtained from Storm et al. and Meyer et al., respectively. For the Wilson et al. and Storm et al. data sets, the Bruceton procedure was used, while the Meyer et al. data set is based on the UN test procedure. Hence, there are systematic differences between the measured critical impact levels in these data sets, which emphasizes that we cannot easily combine them in order to parametrize and validate the models collectively [8].

Statistical analysis

In order to evaluate the predictive ability of our models, we calculate the coefficient of determination (R^2), root-mean-square error (RMSE), absolute mean, and maximum deviation between our predictions and the measurements from our data sets. The most promising model is also developed in parallel via Bayesian regression, taking model complexity into account. We evaluate the predictive power of this Bayesian model via the model evidence function. In addition, we also perform a simple sensitivity analysis in order to evaluate the consequences of inaccurate calculations or measurements.

In the frequentist framework, the mean μ and variance σ^2 in the distribution (1) are estimated using the sample mean and correctly scaled sample variance, respectively, which are unbiased. We also note that our assumption (1) may be rephrased as

$$I \sim \text{Lognormal}\left(\mu, \sigma^2\right), \qquad (14)$$

and so in the frequentist framework, we may predict the mean and variance of a new critical impact level I_{new} as

$$\mathbb{E}\left[I_{\text{new}}\right] = \exp\left(\mu + \frac{\sigma^2}{2}\right), \qquad (15)$$

$$\text{var}(I_{\text{new}}) = \left[\exp\left(\sigma^2\right) - 1\right]\exp\left(2\mu + \sigma^2\right). \qquad (16)$$

In the Bayesian framework, we introduce a conjugate prior distribution for the model coefficients and estimate the variance by maximizing the evidence function. Then any new critical impact level $I_{\text{new}}|\mathcal{D}$ given observed data $\mathcal{D}$ will be governed by the predictive distribution, which will also be lognormal due to the functional form of the prior distribution. That is,

$$I_{\text{new}}|\mathcal{D} \sim \text{Lognormal}\left(\mu_{\mathcal{D}}, \sigma_{\mathcal{D}}^2\right), \qquad (17)$$

where $\mu_{\mathcal{D}}$ and $\sigma_{\mathcal{D}}^2$ are the mean and variance of the predictive distribution, respectively. Recall that unlike in the frequentist framework, $\sigma_{\mathcal{D}}^2$ will depend on the individual choice of molecule considered. The expectation and variance of the new critical impact level $I_{\text{new}}|\mathcal{D}$ may then be calculated as

$$\mathbb{E}\left[I_{\text{new}}|\mathcal{D}\right] = \exp\left(\mu_{\mathcal{D}} + \frac{\sigma_{\mathcal{D}}^2}{2}\right), \qquad (18)$$

$$\text{var}\left(I_{\text{new}}|\mathcal{D}\right) = \left[\exp\left(\sigma_{\mathcal{D}}^2\right) - 1\right]\exp\left(2\mu_{\mathcal{D}} + \sigma_{\mathcal{D}}^2\right). \qquad (19)$$

Later, we predict the critical impact level for DATB and TATB, both in the frequentist and Bayesian framework.

Results and discussion

The molecular structure of these energetic materials together with the computed data are available in the Electronic Supplementary Material.

Nitroaromatics

We perform the linear regression for nitroaromatic materials in two separate ways, both in a frequentist framework and via Bayesian regression. In the former framework, the model coefficients are determined by minimizing a non-regularized sum-of-squares error function, and the predictive ability of the model is evaluated via cross-validation. Note that with this approach, the issue of model complexity is not addressed. However, when performing regression with relatively few data points, the issue of model complexity becomes a key point for avoiding over-fitting. In the Bayesian framework, analysis of model complexity is built in by design, which leads us to somewhat different conclusions than those relying on the frequentist approach, in particular for the Storm et al. data set. In the

Table 2 The coefficient of determination between the log critical impact level and the reciprocal of the heat of detonation, the reciprocal of the total energy, the reciprocal of the temperature of detonation, the bond dissociation energy, the bond dissociation energy divided by the total energy, the bond dissociation energy divided by the heat of detonation, and the bond dissociation divided by the temperature of detonation. The regression is based on the Wilson et al., Storm et al., and Meyer et al. data sets, respectively

Data set	$1/Q$ (dm^3 kJ^{-1})	$1/E_{\text{total}}$ (E_{h}^{-1})	$1/T_{\text{ex}}$ (K^{-1})	BDE (kJ mol^{-1})	BDE/E_{total} (kJ mol^{-1} E_{h}^{-1})	BDE/Q (dm^3 mol^{-1})	BDE/T_{ex} (kJ mol^{-1} K^{-1})
Wilson et al.	0.20	0.24	0.41	0.56	0.48	0.76	0.81
Storm et al.	0.41	0.26	0.54	0.56	0.42	0.64	0.67
Meyer et al.	0.64	0.21	0.75	0.41	0.40	0.70	0.69

Bayesian framework, the predictive quality of the model is addressed by evaluating the model evidence rather than cross-validation.

Frequentist framework

The results for the nitroaromatic materials based on the Wilson et al., Storm et al., and Meyer et al. data sets are shown in Table 2.

We note that $1/T_{\text{ex}}$ correlates better with $\log I$ than $1/Q$ for all three data sets. The most promising predictor of critical impact levels overall is BDE/T_{ex}, with $R^2 = 0.81, 0.67$, and 0.69. Figure 1 illustrates the merit of this model for the Wilson et al. data set.

Table 2 shows only a weak correlation between BDE/E_{total} and $\log I$. These results are not in line with those of Song et al., which indicate merit for this model [20, 21]. However, they derived this correlation by using a small data set. For molecules of similar structure, E_{total} is likely to correlate with Q, but this is unlikely to hold in general; we get virtually no correlation ($R^2 = 0.03$) when plotting these parameters against each other.

As there are only 24, 17, and 16 molecules in the Wilson et al., Storm et al., and Meyer et al. data sets, respectively, our regression is a priori prone to over-fitting. In order to evaluate its predictive ability, we use leave-one-out cross-validation; the results of which are summarized in Table 3.
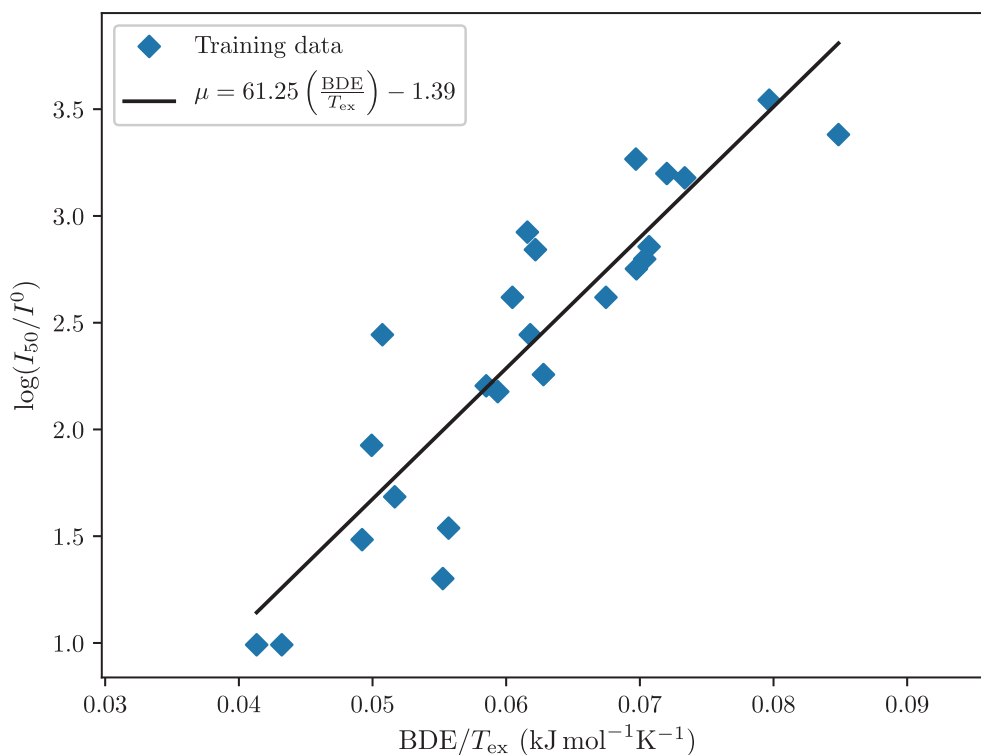

Fig. 1 The log critical impact level of the nitroaromatics in the Wilson et al. data set plotted against the bond dissociation energy divided by the detonation temperature, along with the best-fitting regression line ($R^2 = 0.81$)

Table 3 Leave-one-out cross-validation of the model based on the bond dissociation energy divided by the temperature of detonation. The RMSE, absolute average deviation, maximum deviation, and the compounds with largest deviation are given. See the "Statistical analysis" section for how the predicted critical impact level is calculated from the model

Data set	RMSE (J)	Average abs. deviation (J)	Maximum deviation (J)	Maximum deviation compound
Wilson et al.	6.1	3.7	25	CL-14
Storm et al.	15	10	40	Styphnic acid
Meyer et al.	9.1	6.6	21	Picramic acid

Our cross-validation does not show any particularly convincing results for either data set, revealing that more parameters than just BDE/T_{ex} are needed to predict the critical impact level of nitroaromatic compounds. The most promising numbers are for the Wilson et al. data set, with an RMSE of 6.1 J and an average absolute deviation of 3.7 J between the predicted and measured I_{50}. 5,7-Diamino-4,6-dinitrobenzofuroxan (CL-14) has the largest deviation (25 J). The (Wilson et al.) model predicts the critical impact level I_{50} of CL-14 to be 48 J, while the measured I_{50} is 29 J. We note that the calculated C–NO$_2$ BDE in CL-14 is quite large, at 318 kJ mol^{-1}. However, this NO$_2$ group is surrounded by an amino group on each of the neighboring carbons, as is also the case for TATB. Hence, this value for CL-14 is not particularly surprising, as the BDE of TATB is calculated to be 310 kJ mol^{-1}. The deviation between the measured and predicted critical impact level of CL-14 may indicate that this molecule follows another decomposition route.

CL-14 contains a furoxan ring, as is also the case for three other molecules in the Wilson et al. data set, namely 7-amino-4,6-dinitrobenzofuroxan (ADNBF), 4,6-dinitrobenzofuroxan (DNBF), and 8-amino-7-nitrobenzobisfuroxan (CL-18). These have deviations of -3 J, -8 J, and 2 J (respectively) between the predicted and measured I_{50}. If the decomposition is initiated in the furoxan ring, a larger deviation may be expected.

From Table 3, we see that in the Storm et al. data set, 1,3-dihydroxy-2,4,6-trinitrobenzene (styphnic acid) has the largest deviation between the predicted and measured I_{50}. Our model predicts this value to be 45 J, while it is measured to be 11 J. The BDE is calculated to be 287 kJ mol^{-1} by using the M06 functional and the 6-311G+(2d,p) basis

set. This is similar to the value calculated when using B3LYP/6-31G(d), which is 274 kJ mol^{-1}.

Bayesian regression

We now perform Bayesian linear regression separately on the three data sets for the most promising model, namely (12). In order to adapt the Bayesian framework, we introduce a zero-mean, isotropic bivariate normal prior distribution over the model coefficients $\mathbf{c} = (c_1, c_2)^T$ with covariance matrix $\tau^2 I$, so that

$$\mathbf{c} \sim \mathcal{N}\left(\mathbf{0}, \tau^2 I\right). \tag{20}$$

We follow the process outlined in Reference [61], in which the first step is to maximize the evidence function in order to obtain estimates for the parameters σ^2 and τ^2, a technique also known as emperical Bayes. We then compute the posterior distributions for the model coefficients by updating our prior distributions separately over the the data sets. Using the mean of this posterior distribution as our estimate for the model coefficients $\mathbf{c}$, we obtain our linear models. Note that this process is equivalent to minimizing a regularized sum-of-squares error function with regularization term $\lambda = \sigma^2/\tau^2$, and so model complexity is intrinsically accounted for. A summary of the model coefficients computed, along with the relevant parameters, is provided in Table 4.

We see that the penalizing regularization term plays a substantial role for the regression on the Storm et al. data set. Although values of R^2 in the frequentist analysis for Storm et al. and Meyer et al. were relatively close in value (0.67

Table 4 Bayesian regression for the three nitroaromatic data sets, based on Eq. 12, with a prior distribution given by Eq. 20. The variances σ^2 and τ^2, the regularization coefficient $\lambda = \sigma^2/\tau^2$ and the effective number of parameters $\gamma = \mathbf{c}^T \mathbf{c}/\tau^2$ are also included

Data set	c_1	c_2	σ^2	τ^2	$\lambda = \sigma^2/\tau^2$	$\gamma = \mathbf{c}^T\mathbf{c}/\tau^2$
Wilson et al.	− 1.31	59.92	0.11	1.7×10^3	6.2×10^{-5}	1.98
Storm et al.	2.80	2.18	0.82	12	6.7×10^{-3}	1.03
Meyer et al.	− 0.81	44.40	0.23	1.0×10^3	2.2×10^{-4}	1.93

Table 5 Evaluation of the (log) model evidence for the proposed linear model (12), along with the (log) model evidence for an alternative constant model (21). The Bayes factor and its preference are also listed for the three data sets

Data set	Linear hypothesis model evidence (log)	Constant hypothesis model evidence (log)	Bayes factor (linear against constant)	Preferred model
Wilson et al.	-14.46	-26.13	1.2×10^5	Linear
Storm et al.	-23.40	-21.12	0.10	Constant
Meyer et al.	-17.78	-18.96	3.25	Linear

and 0.69, respectively), the Bayesian analysis suggests that there is little evidence supporting a linear term in the model for the Storm et al. data set. Considering the interpretation of the parameter $\gamma = \mathbf{c}^T \mathbf{c}/\tau^2$ provided in Reference [62], namely as the *effective number of parameters* for the model, we see that for the Storm et al. data set, there is effectively only a single parameter (the constant term) governing the distribution of data points. This is in contrast to the corresponding results for the Wilson et al. and Meyer et al. data sets, where the effective number of parameters is calculated to be approximately equal to 2, supporting the claim that $\log I$ indeed depends linearly on $\mathrm{BDE}/T_{\mathrm{ex}}$.

We evaluate the predictive ability of our models in the Bayesian framework by calculating the (log) model evidence function. Since the Bayesian regression penalizes the model complexity for the Storm et al. data set, we also compare the evidence for our proposed model to that of a separate constant model which asserts no correlation between $\log I$ and $\mathrm{BDE}/T_{\mathrm{ex}}$. That is, our other model claims that I is governed by a lognormal distribution of the form

$$I \sim \mathrm{Lognormal}\left(\mu_0, \sigma_0^2\right) \tag{21}$$

with constant mean μ_0 and constant variance σ_0^2. We perform the corresponding Bayesian regression for this alternative model and compare the results to our original model (12) by calculating the Bayes factor. The results of this model comparison are presented in Table 5. A Bayesian factor larger than unity indicates preference towards the linear model, while a factor smaller than unity indicates preference towards the constant model.

From Table 5, we see that Bayesian regression on the Wilson et al. and Meyer et al. data sets prefers the linear

model (12), whereas it prefers the constant model (21) for the Storm et al. data set. Hence, our prediction of the critical impact level of TATB will be based on the Wilson et al. data set rather than that of Storm et al. Note that the model comparison for the Wilson et al. data set is several orders of magnitude more decisive than for that of Meyer et al.

We make two predictions of the critical impact level I_{50} of DATB and TATB, one based on the frequentist framework and another from the Bayesian predictive distribution. The predictions and their uncertainties are calculated using Eqs. 15 and 16. The results are given in Table 6.

Storm et al. measured the critical impact level (H_{50}, 2.5 kg drop weight) of TATB to be larger than 320 cm ($I_{50} > 78$ J). Additionally, they predicted H_{50} to be 490 cm ($I_{50} = 120$ J) based on the measured impact sensitivity and the oxygen balance for trinitrobenzene, 2,4,6-trinitroaniline and DATB [41]. Using the Wilson et al. data set, our frequentist and Bayesian prediction of I_{50} for TATB are 97 J and 95 J, respectively. We note that the Bayesian prediction is less certain since the variance of the predictive distribution accounts for uncertainty related to the model coefficients, as well as noise in the data. This difference in certainty for the two models is illustrated in Fig. 2.

The frequentist and Bayesian predictions of I_{50} for DATB are 61 J and 60 J, respectively. In the Storm et al. data set, its measured value is 78 J [41], which is consistent with the difference between the Wilson et al. data set and the Storm et al. data set. Indeed, the former data set is more sensitive than the latter for materials with a critical impact level larger than 15 J. For more sensitive materials ($I_{50} < 15$ J), the differences in critical impact level are smaller [8]. In the Storm et al. data set, a sample mass of 40 mg was used, while in the Wilson et al. data set, it was 35 mg. It has been reported that nitroaromatics such

Table 6 Frequentist and Bayesian predictions of the critical impact level I_{50} (J) of DATB and TATB based on Eq. 12. The standard deviation of the respective predictions are included in parentheses

	μ	$\mu_{\mathcal{D}}$	σ^2	$\sigma_{\mathcal{D}}^2$	Frequentist predicted I_{50} (J)	Bayesian predicted I_{50} (J)
DATB	4.06	4.03	0.11	0.15	61 (21)	60 (24)
TATB	4.52	4.48	0.11	0.16	97 (34)	95 (40)

120

Fig. 2 Frequentist and Bayesian predictions of the critical impact level of TATB, based on the lognormal model fitted to the Wilson et al. data set. The red-shaded region comprises one standard deviation on either side of the exponential curve, as calculated by Eqs. 16 and 19 for the frequentist and Bayesian prediction, respectively
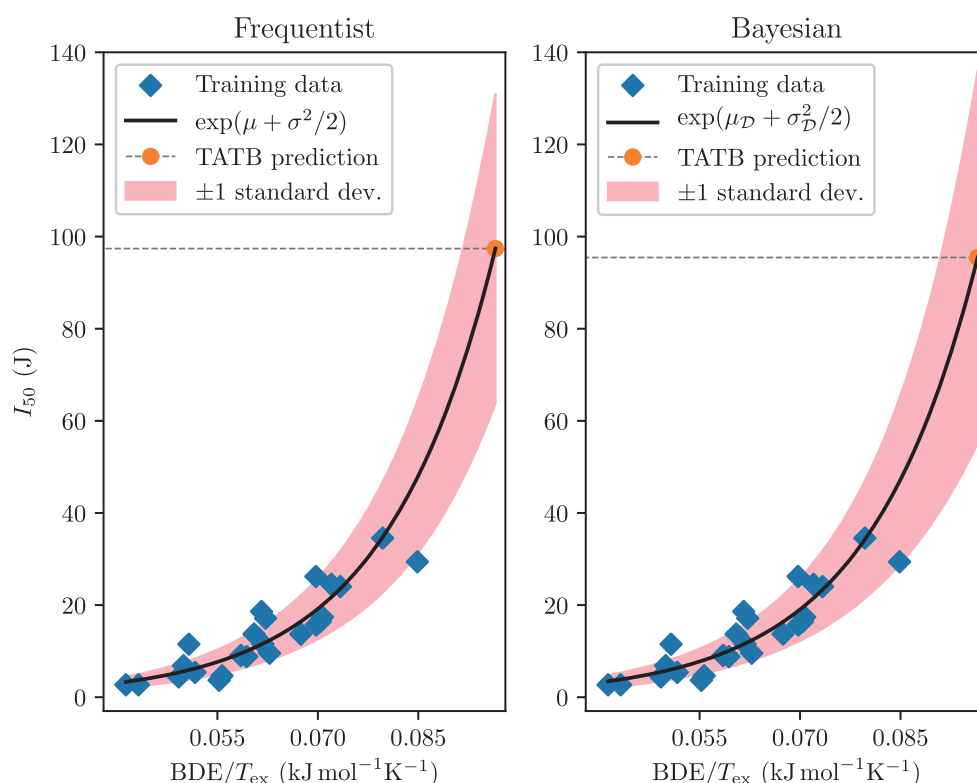
as tetryl, 2,4,6-trinitrotoluene (TNT) and o-trinitrophenol become more sensitive when the thickness or the amount of sample is reduced [14, 15]. The smaller the sample, the more energy is released per unit volume of explosive. The Wilson et al. data set is based on a smaller sample mass than that used in the Storm et al. data set, which may be one of the main reasons for the difference in measured critical impact levels.

Model sensitivity analysis

We now perform a simple sensitivity analysis of our frequentist and Bayesian models, investigating how sensitive they are to slight changes in the input variables upon which the regression is based. From the "Bayesian regression"

section, we expect the Bayesian model to be more sensitive to such alterations than the frequentist model, since the variance of the Bayesian predictive distribution also accounts for uncertainty in the model coefficients. This can be seen explicitly by observing how the predicted value of I_{50} for DATB changes if we over- and underestimate the values of BDE and T_{ex} in turn in the Wilson et al. data set. When multiplying BDE and T_{ex} by $(1 + \varepsilon)$, where ε = -0.05, -0.04, ..., 0.04, 0.05, the frequentist prediction of I_{50} for DATB changes by no more than $10^{-13}\%$ of its original value. On the other hand, the corresponding difference in the Bayesian prediction ranges from $10^{-8}\%$ to $10^{-6}\%$.

Inaccuracies in the measured critical impact level may also affect how the model makes new predictions. We measure the sensitivity of our frequentist and Bayesian models

Table 7 The coefficient of determination (R^2) between the log critical impact level of nitramine compounds and the reciprocal of the heat of detonation, the reciprocal of the total energy, the reciprocal of the temperature of detonation, the bond dissociation energy, the bond dissociation energy divided by the total energy, the bond dissociation energy divided by the heat of detonation, and the bond dissociation divided by the temperature of detonation. The calculations are first based on using the weakest (N-N) bond, and then the weakest C–NO$_2$ or N–NO$_2$ bond. The regression is based on the Storm et al. data set

Data set	$1/Q$ (dm^3 kJ^{-1})	$1/E_{total}$ (E_h^{-1})	$1/T_{ex}$ (K^{-1})	BDE (kJ mol^{-1})	BDE/E_{total} (kJ mol^{-1} E_h^{-1})	BDE/Q (dm^3 mol^{-1})	BDE/T_{ex} (kJ mol^{-1} K^{-1})
N–NO$_2$	0.34	0.12	0.44	0.04	0.10	0.41	0.34
C–NO$_2$ or N–NO$_2$	0.34	0.12	0.44	0.11	0.12	0.49	0.41

by multiplying the values of I in the data set by $(1 + \varepsilon)$, where ε is drawn (separately for each molecule) from a uniform probability distribution of range $[-0.05, 0.05]$. We then calculate the absolute difference in the predicted I_{50} of DATB before and after perturbing the input, given as a percentage of the original predicted value of I_{50}. Averaging over 500 such random simulations, we find the predicted I_{50} of DATB to deviate from the original prediction by a factor of 1.3% for both the frequentist and the Bayesian model. This result indicates that inexact experimental measurements may sully the accuracy of the model.

Nitramines

The results for the nitramines based on the Storm et al. data set are shown in Table 7. We see that neither $1/Q$ nor $1/T_{ex}$ is strongly correlated with log I ($R^2 = 0.34$ and 0.44, respectively), and for the N–NO$_2$ BDE, we get virtually no correlation with log I ($R^2 = 0.04$).

In addition to N–NO$_2$ bond, some of the nitramines contain one or more C–NO$_2$ bonds. For such molecules, the C–NO$_2$ bond may be weaker than the N–NO$_2$ bond when three NO$_2$ groups are attached to the same carbon atom. This is due to repulsive forces between the NO$_2$ groups and also their attraction to electrons in the C–NO$_2$ bond.

If the weakest bond in the model is taken to be either N–NO$_2$ or C–NO$_2$, there is still low correlation between log I and BDE, indicating that molecular rearrangements and auto-catalyzed reactions play a key role in the initial nitramine decomposition. Contrary to the promising result for nitroaromatics, we only see a weak correlation between log I and BDE/T_{ex} ($R^2 = 0.41$). This observation is illustrated in Fig. 3, where log I is plotted against BDE/T_{ex}.

We now comment on the outliers marked in Fig. 3. First, bis-(2,2,2-trinitroethyl)-nitramine has a positive oxygen balance which results in a relatively low value of T_{ex} since the molecule does not contain enough carbon or hydrogen to utilize all the oxygen. This leads to a particularly high value of BDE/T_{ex}, which illustrates how our model predicts compounds with a positive oxygen balance to be less sensitive to impact than what they actually are. Next, N,N'-dinitromethanediamine and N,N'-dinitro-1,2-ethanediamine contain the –NH–NO$_2$ functional group. The values of the N–NO$_2$ BDE for these molecules (220 kJ mol^{-1} and 214 kJ mol^{-1}, respectively) are relatively high compared with those of nitramines, which usually range from 150 to 170 kJ mol^{-1}. We calculated BDE of these compounds with the M06 functional and the 6-311+G(2d,p) basis set in order to exclude the possibility for any erroneous geometry optimization caused by the B3LYP functional. The M06 functional with the 6-311+G(2d,p) basis set predicts
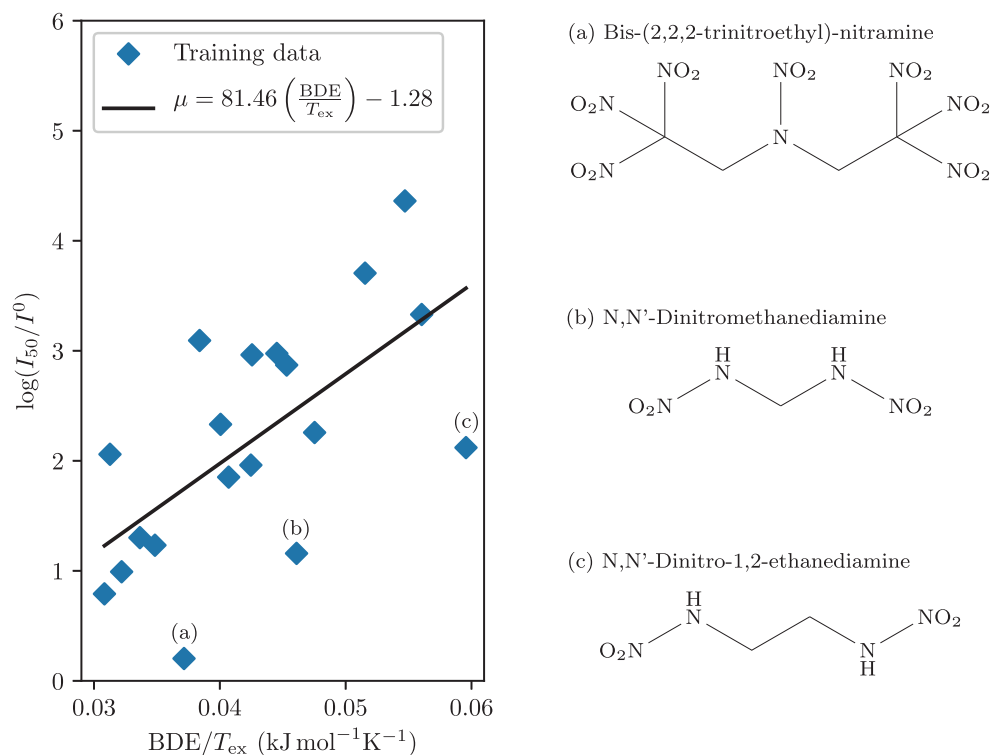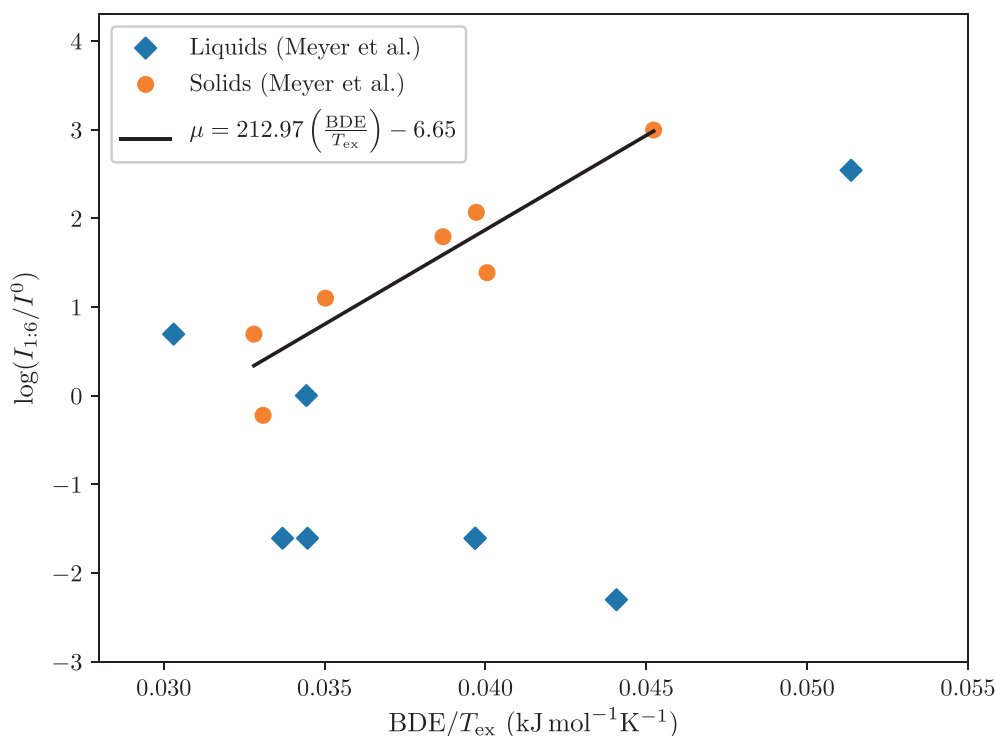

Fig. 3 The log critical impact level of the nitramines in the Storm et al. data set plotted against the bond dissociation energy divided by the detonation temperature, along with the best-fitting linear regression line ($R^2 = 0.41$). The weakest N–NO$_2$ or C–NO$_2$ bond is used as the weakest bond for the model

Table 8 The coefficient of determination (R^2) between the log critical impact level of nitrate ester compounds and heat of detonation, total energy, detonation temperature, and bond dissociation energy and ratios between the bond dissociation energy and heat of detonation, total energy, and detonation temperature. The regression is based on the Meyer et al. data set

Data set	$1/Q$ (dm^3 kJ^{-1})	$1/E_{\text{total}}$ (E_{h}^{-1})	$1/T_{\text{ex}}$ (K^{-1})	BDE (kJ mol^{-1})	BDE/E_{total} (kJ mol^{-1} E_{h}^{-1})	BDE/Q (dm^3 mol^{-1})	BDE/T_{ex} (kJ mol^{-1} K^{-1})
Solids + liquids	0.00	0.19	0.17	0.00	0.19	0.00	0.11
Solids	0.29	0.83	0.83	0.68	0.49	0.49	0.85
Liquids	0.05	0.16	0.16	-0.02	0.11	0.03	0.10

BDE for N,N'-dinitromethanediamine and N,N'-dinitro-1,2-ethanediamine to be 244 kJ mol^{-1} and 241 kJ mol^{-1}, respectively. Thus, the divergence of these two molecules from the regression line cannot be explained by the high BDE calculated by the B3LYP functional alone.

The initial decomposition of nitramines can take place through several mechanistic routes. For example, at least four initial mechanisms for the decomposition of 1,3,5-trinitroperhydro-1,3,5-triazine (RDX) have been suggested: N–NO$_2$ homolysis, HONO elimination, the "tripple whammy" mechanism, and NONO isomerization. Using the couple cluster theory, the E_a value of the HONO elimination has been calculated to be lower than that of the N–NO$_2$ homolytic reaction for RDX [63]. A recent study of the initial decomposition process of liquid RDX has revealed that HONO elimination is likely to be the major decomposition pathway [64].

In order to investigate whether HONO elimination is an alternative decomposition route for N,N'-dinitromethanediamine, we calculate the energy required to break the N–NOOH bond. The HONO BDE is calculated to be 420 kJ mol^{-1}. When the hydrogen atom is moved from the nitrogen atom to the oxygen atom, the length of the N–N bond decreased from 1.379 to 1.257 Å (M06/6-311+G(2d,p)). The bond becomes more like a double bond, making it unlikely to break without more molecular rearrangements.

Nitrate esters

The nitrate ester data set (from Meyer et al.) consists of both liquids and solids. Table 8 shows that when these are considered simultaneously, none of our models seems to give any satisfactory predictions. When solids and

Fig. 4 The log critical impact level of the liquid and solid nitrate esters in the Meyer et al. data set plotted against the bond dissociation energy divided by the detonation temperature, along with the best-fitting linear regression line for the solids ($R^2 = 0.85$)

liquids are treated separately, no model shows any notable improvement for the liquids. Denisaev et al. have found a correlation between $\sqrt{H_{50}}$ and ρQ, where ρ is the density, for liquid nitrate esters [16]. The critical impact levels of liquid nitrate esters are sensitive to experimental factors such as the presence of bubbles in the liquid which can significantly alter the impact sensitivity of a liquid explosive [12].

There seems to be a strong correlation between $\log I$ and BDE/T_{ex} for solid nitrate esters, but as can be seen in Fig. 4, this assertion is based on very few data points. In order to investigate whether this correlation is genuine or just a result of over-fitting, we perform a Bayesian model comparison with a constant model (which asserts no correlation between BDE/T_{ex} and $\log I$) as in the "Bayesian regression" section. When model complexity is accounted for, the Bayes factor comes out to be 0.554, in favor of the constant model. Hence, more data points are needed for further investigation of the the merit of the linear model. Similar results hold for the other seemingly promising model, namely (8), with a Bayes factor of 0.552, again favoring the constant model.

Conclusion

By investigating 70 energetic nitroaromatics, nitramines, and nitrate esters, we have evaluated seven models for predicting critical impact level, the quantity from which impact sensitivity is determined. Input parameters were the molecules' temperature of detonation, heat of detonation, and bond dissociation energy. Our regression was based on three separate data sets comprising 91 data points in total.

For the largest nitroaromatics data set, the bond dissociation energy divided by the temperature of detonation was the best predictor of critical impact level, with a coefficient of determination (R^2) of 0.81. Leave-one-out cross-validation gave a root-mean-square error (RMSE) of 6.1 J and the absolute average deviation was 3.7 J between the predicted and the measured values. A separate Bayesian regression also assigned similar merit to the predictive power of this model, also when accounting for model complexity. The frequentist and Bayesian models predicted the critical impact level of TATB to be 97 J and 95 J, respectively. Our sensitivity analysis showed this prediction to be more robust under changes in calculations of molecular parameters than measurements of critical impact levels.

For nitramines, our analysis showed that the temperature of detonation is a moderately useful predictor of critical impact level, unlike the N–NO$_2$ bond dissociation energy, which we found to have virtually no such predictive ability. Hence, N–NO$_2$ homolysis is unlikely to be the only reaction taking place in the initial phase of the decomposition.

None of the models was able to predict the critical impact level of liquid nitrate esters, but for a small data set of solid nitrate esters, there seemed to be promising results for bond dissociation energy divided by the temperature of detonation. However, our Bayesian model comparison revealed that more data points are necessary to validate this correlation.

Our results regarding nitroaromatics, nitramines, and solid nitrate esters allude to the temperature of detonation being a better predictor of critical impact level than the heat of detonation. Moreover, a high ratio of bond dissociation energy to temperature of detonation indicates low impact sensitivity, whereas a small ratio suggests that the material is highly sensitive to impact. Having evaluated the predictive power of our models, we conclude that predicting impact sensitivity of energetic materials with acceptable accuracy may require inclusion of additional parameters such as hardness, crystal defects, particle size, amount of sample, heat conductivity, and heat capacity.

References

1. Fried LE, Manaa MR, Pagoria PF, Simpson RL (2001) Design and synthesis of energetic materials. Annu Rev Mater Res 31(1):291–321
2. Politzer P, Murray JS, Seminario JM, Lane P, Grice ME, Concha MC (2001) Computational characterization of energetic materials. J Mol Struct Theochem 573(1–3):1–10
3. Rice BM, Byrd EF (2006) Theoretical chemical characterization of energetic materials. J Mater Res 21(10):2444–2452
4. Klapötke TM (2012) Chemistry of high-energy materials, 2nd edn. Walter de Gruyter GmbH & Co KG, Berlin/Boston
5. Brinck T, Rahm M (2014) Theoretical design of green energetic materials: predicting stability, detection, synthesis and performance. In: Brinck T, Green energetic materials (eds). Wiley, Chichester, pp 15–43
6. Politzer P, Murray JS (2014) Detonation performance and sensitivity: a quest for balance. In: Advances in quantum chemistry, vol 69. Elsevier, pp 1–30
7. Politzer P, Murray JS (2016) High performance, low sensitivity: conflicting or compatible? Propellants, Explosives and Pyrotechnics 41(3):414–425
8. Politzer P, Murray JS (2015) Some molecular/crystalline factors that affect the sensitivities of energetic materials: molecular

surface electrostatic potentials, lattice free space and maximum heat of detonation per unit volume. J Mol Model 21(2):25

9. Yan QL, Zeman S (2013) Theoretical evaluation of sensitivity and thermal stability for high explosives based on quantum chemistry methods: a brief review. Int J Quantum Chem 113(8):1049–1061

10. Zeman S, Jungová M (2016) Sensitivity and performance of energetic materials. Propellants, Explosives and Pyrotechnics 41(3):426–451

11. Chen ZX, Xiao HM (2014) Quantum chemistry derived criteria for impact sensitivity. Propellants, Explosives and Pyrotechnics 39(4):487–495

12. Bowden FP, Yoffe AD (1952) Initiation and growth of explosion in liquids and solids. The Syndics of the Cambridge University Press, Cambridge

13. Wenograd J (1961) The behaviour of explosives at very high temperatures. Trans Faraday Soc 57:1612–1620

14. Afanasyev GT, Bobolev VK (1968) Initiation of solid explosives by impact. Nauka Press, Moscow

15. Wen Y, Long X, Xiang Y, Yu S, Dai X, Li M (2015) Mass dependent of explosion probability of RDX, tetryl, and a PBX powder in drop hammer test. Propellants, Explosives and Pyrotechnics 40(3):433–438

16. Denisaev A, Korsunskii B, Pepekin V, Matyushin YN (2010) Impact sensitivity of liquid explosives. Combustion, Explosion, and Shock Waves 46(1):74–80

17. Politzer P, Murray JS (2015) Impact sensitivity and the maximum heat of detonation. J Mol Model 21(10):262

18. Rice BM, Hare JJ (2002) A quantum mechanical investigation of the relation between impact sensitivity and the charge distribution in energetic molecules. J Phys Chem A 106(9):1770–1783

19. Mathieu D (2017) Sensitivity of energetic materials: theoretical relationships to detonation performance and molecular structure. Industr Eng Chem Res 56(29):8191–8201

20. Song X, Cheng X, Yang X, Li D, Linghu R (2008) Correlation between the bond dissociation energies and impact sensitivities in nitramine and polynitro benzoate molecules with polynitro alkyl groupings. J Hazard Mater 150(2):317–321

21. Song XS, Cheng XL, Yang XD, He B (2006) Relationship between the bond dissociation energies and impact sensitivities of some nitro-explosives. Propellants, Explosives and Pyrotechnics 31(4):306–310

22. Owens F (1996) Calculation of energy barriers for bond rupture in some energetic molecules. J Mol Struct Theochem 370(1):11–16

23. Rice BM, Sahu S, Owens FJ (2002) Density functional calculations of bond dissociation energies for NO_2 scission in some nitroaromatic molecules. J Mol Struct Theochem 583(1–3):69–72

24. Kamlet M, Adolph H (1979) The relationship of impact sensitivity with structure of organic high explosives. ii. polynitroaromatic explosives. Propellants, Explosives and Pyrotechnics 4(2):30–34

25. Murray JS, Concha MC, Politzer P (2009) Links between surface electrostatic potentials of energetic molecules, impact sensitivities and C–NO_2/N–NO_2 bond dissociation energies. Mol Phys 107(1):89–97

26. Politzer P, Murray JS (1996) Relationships between dissociation energies and electrostatic potentials of C–NO_2 bonds: applications to impact sensitivities. J Mol Struct 376(1–3):419–424

27. Murray JS, Lane P, Politzer P (1998) Effects of strongly electron-attracting components on molecular surface electrostatic potentials: application to predicting impact sensitivities of energetic molecules. Mol Phys 93(2):187–194

28. Zhang H, Cheung F, Zhao F, Cheng XL (2009) Band gaps and the possible effect on impact sensitivity for some nitro aromatic explosive materials. Int J Quantum Chem 109(7):1547–1552

29. Zeman S (2006) New aspects of initiation reactivities of energetic materials demonstrated on nitramines. J Hazard Mater 132(2–3):155–164

30. Ye S, Tonokura K, Koshi M (2003) Energy transfer rates and impact sensitivities of crystalline explosives. Combust Flame 132(1–2):240–246

31. Pospíšil M, Vávra P, Concha MC, Murray JS, Politzer P (2010) A possible crystal volume factor in the impact sensitivities of some energetic compounds. J Mol Model 16(5):895–901

32. Dlott DD (2003) Fast molecular processes in energetic materials. In: Theoretical and computational chemistry, vol 13. Elsevier, pp 125–191

33. Brill TB, James KJ (1993) Kinetics and mechanisms of thermal decomposition of nitroaromatic explosives. Chem Rev 93(8):2667–2692

34. Keshavarz MH (2013) A new general correlation for predicting impact sensitivity of energetic compounds. Propellants, Explosives and Pyrotechnics 38(6):754–760

35. Keshavarz MH (2010) Simple relationship for predicting impact sensitivity of nitroaromatics, nitramines, and nitroaliphatics. Propellants, Explosives and Pyrotechnics 35(2):175–181

36. Prana V, Fayet G, Rotureau P, Adamo C (2012) Development of validated QSPR models for impact sensitivity of nitroaliphatic compounds. J Hazard Mater 235:169–177

37. Wang R, Jiang J, Pan Y, Cao H, Cui Y (2009) Prediction of impact sensitivity of nitro energetic compounds by neural network based on electrotopological-state indices. J Hazard Mater 166(1):155–186

38. Mathieu D (2013) Toward a physically based quantitative modeling of impact sensitivities. J Phys Chem A 117(10):2253–2259

39. Mathieu D, Alaime T (2015) Impact sensitivities of energetic materials: exploring the limitations of a model based only on structural formulas. J Molec Graph Modell 62:81–86

40. Mathieu D, Alaime T (2014) Predicting impact sensitivities of nitro compounds on the basis of a semi-empirical rate constant. J Phys Chem A 118(41):9720–9726

41. Storm C, Stine J, Kramer J (1990) Sensitivity relationships in energetic materials. In: Chemistry and physics of energetic materials. Springer, pp 605–639

42. Dixon WJ, Mood AM (1948) A method for obtaining and analyzing sensitivity data. J Am Stat Assoc 43(241):109–126

43. (2015) Recommendations on the transport of dangerous goods: manual of tests and criteria, 6th revised edn. UN, New York. https://doi.org/10.18356/c2b83494-en. Accessed 8 August 2019

44. Wu CJ, Fried LE (1998) First-principles study of high explosive decomposition energetics. In: Conference: eleventh international detonation (1998) Symposium, Snowmass, CO. August 31-September 4, 1998

45. Sućeska M (2010) EXPLO5, version 5.04. Zagreb

46. (1997) Data base of thermochemical data, 4th edn. Fraunhofer Institut für Chemische Technologie

47. Meyer R, Köhler J, Homburg A (2007) Explosives, 6th edn. Wiley, Weinheim

48. Wilson WS, Bliss DE, Christian SL, Knight DJ (1990) Explosive properties of polynitroaromatics, NXC TP 7073 Naval Weapons Center. China Lake

49. Adolph HG, Holden JR, Cichra DA (1981) Relationships between the impact sensitivity of high energy compounds and some molecular properties which determine their performance; N, M, and ρ_0, NSWC TR 80-495. Naval Surface Weapons Center, Silver spring

50. Keshavarz MH (2009) Predicting condensed phase heat of formation of nitroaromatic compounds. J Hazard Mater 169(1–3):890–900

51. Keshavarz MH (2011) Prediction of the condensed phase heat of formation of energetic compounds. J Hazard Mater 190(1–3):330–344

52. Rice BM, Hare JJ, Byrd EF (2007) Accurate predictions of crystal densities using quantum mechanical molecular volumes. J Phys Chem A 111(42):10,874–10,879

53. Lu T, Chen F (2012) Multiwfn: a multifunctional wavefunction analyzer. J Comput Chem 33(5):580–592

54. Lu T, Chen F (2012) Quantitative analysis of molecular surface based on improved marching tetrahedra algorithm. J Molecul Graph Modell 38:314–323

55. Frisch M et al (2009) Gaussian 09, version A.02. Gaussian Inc., Wallingford

56. Khrapkovskii G, Sharipov D, Shamov A, Egorov D, Chachkov D, Van BN, Tsyshevsky R (2013) Theoretical study of substituents effect on C–NO$_2$ bond strength in mono substituted nitrobenzenes. Comput Theor Chem 1017:7–13

57. Qi C, Lin QH, Li YY, Pang SP, Zhang RB (2010) C–N bond dissociation energies: an assessment of contemporary DFT methodologies. J Mol Struct THEOCHEM 961(1–3):97–100

58. Gonzalez AC, Larson CW, McMillen DF, Golden DM (1985) Mechanism of decomposition of nitroaromatics. laser-powered homogeneous pyrolysis of substituted nitrobenzenes. J Phys Chem 89(22):4809–4814

59. Pruitt CJM, Goebbert DJ (2013) The C–N dissociation energies of nitrobenzene and nitrotoluene radical anions and neutrals. Chem Phys Lett 580:21–27

60. Foresman JB, Frisch Æ (1996) Exploring chemistry with electronic structure methods, 2nd edn. Gaussian, Inc., Pittsburgh

61. Bishop CM (2006) Pattern recognition and machine learning. Springer, Cambridge

62. MacKay DJ (1992) Bayesian interpolation. Neural Comput 4(3):415–447

63. Molt RW, Watson T, Bazanté AP, Bartlett RJ, Richards NG (2016) Gas phase RDX decomposition pathways using coupled cluster theory. Phys Chem Chem Phys 18(37):26,069–26,077

64. Khichar M, Patidar L, Thynell ST (2018) Improvement and validation of a detailed reaction mechanism for thermal decomposition of RDX in liquid phase. Combust Flame 198:455–465

Paper IV

Improved measurements of impact sensitivities of energetic materials

Improved measurements of impact sensitivities of energetic materials

Dennis Christensen*[†]   Erik Unneberg[†]   Eirik Høyheim[‡]
Tomas Lunde Jensen[†]   Nils Lid Hjort*

* Department of Mathematics, University of Oslo, Oslo, Norway

[†] Norwegian Defence Research Establishment (FFI), Kjeller, Norway

[‡] Norwegian University of Life Sciences, Ås, Norway

dennis.christensen@ffi.no

Abstract

Accurate estimation of impact sensitivity is crucial for safe production, handling, storage and transport of energetic materials. Indeed, molecular characteristics will affect sensitivity, and for solid materials, factors like particle size, lattice defects and morphology also play a role and make reproducibility difficult. As various synthesis and recrystallisation methods may lead to differences in crystal properties, it is important to determine the impact sensitivity whenever an energetic material is prepared. Of particular interest is the median h_{50}, namely the impact energy level at which there is a probability of 50% of an explosion occurring. This value has been shown to correlate with quantum chemical properties of the energetic material in question, providing insight into the underlying causes which govern sensitivity. However, in practical applications, it may be more important to estimate extreme values like the 99% quantile h_{99}. In addition to providing point estimates, we would like to derive confidence intervals to address their uncertainty. In this work, we cover the most common methods for constructing such confidence intervals (the delta method, Fieller's theorem and the likelihood-ratio test) and compare their performance on sensitivity data via simulations. Our experiments indicate that Fieller's theorem is the superior method, and we therefore use it to construct confidence intervals for h_{50} and h_{99} for cyclotetramethylene-tetranitramine (HMX), using new data. Based on our results, we formulate recommendations for researchers measuring sensitivities of synthesised molecules.

Keywords: Energetic materials, Impact sensitivity; Bruceton method; Confidence intervals; HMX

1 Introduction

Energetic materials are molecules or compounds that release chemical energy quickly when they decompose and are necessary ingredients in explosives, propellants and pyrotechnics. They have both civilian and military applications, e.g. in mining, oil and gas well technology, construction, demolition, ammunition, and rocket motors. The production and usage of several energetic compounds and compositions are expected to increase the next years [1], and the search for new materials is vigorously pursued internationally today.

The development of energetic material has a long history, and traditional workhorses such as 2,4,6-trinitrotoluene (TNT), cyclotrimethylene-trinitramine (RDX) and cyclotetramethylene-tetranitramine (HMX) have been around for decades [2–4]. However, a downside with many high-performance compounds is that they are sensitive to external stimuli, e.g. to heating, friction, shock, spark or impact. As safety is an essential issue for production, handling, storage, transport and disposal of energetic materials, synthesis of less sensitive—but still well performing—materials is encouraged. One example is 1,1-diamino-2,2-dinitroethylene (FOX-7)

[5], which has the same empirical formula (and consequently the same oxygen balance) as HMX and RDX, but is far less sensitive to impact [6].

It is thus utterly important to carry out good sensitivity measurements of new energetic materials. Possible intermediate products should also be considered in this respect. Moreover, the sensitivity of some molecules may be altered upon crystal modification, and a good knowledge of the sensitivity of recrystallised products is therefore also advantageous. In addition to the safety aspect, insight into the sensitivity may make us better able to understand the decomposition mechanisms that external stimuli initiate.

When measuring impact sensitivity, a fallhammer apparatus is used. In this device, a sample of the explosive is positioned between two cylinders covered by a ring. A weight of known mass is then repeatedly dropped from various heights onto the top cylinder. The operator registers whether the drop caused the sample to explode via visual and auditory clues, as well as with a decomposition gas detector. The explosive's sensitivity is then estimated from these observations [7, 8].

The log energy levels $x = (x_1, \ldots, x_n)$ used in the fallhammer tests (equivalently, the log drop heights of the weight) are commonly decided with the Bruceton "up-and-down" method, originally introduced in 1948 by Dixon and Mood [9]. This method is commonly applied in the energetic materials industry and recognised by both NATO and the U.S. Department of Defense [7, 8]. The Bruceton method depends on two parameters, namely an initial impact energy level x_1 and a step size d. The first drop test is performed at energy level x_1, at which point the researcher observes the outcome $y_1 \in \{0,1\}$, where 1 and 0 denote explosion and non-explosion, respectively. For the remaining drops, the impact energy level x_i is obtained from x_{i-1} by adding or subtracting d, depending on whether the drop at x_{i-1} caused an explosion or not. More precisely, for i = 2, ..., n, we have

$$x_i = \begin{cases} x_{i-1} + d & \text{if} \quad y_{i-1} = 0 \\ x_{i-1} - d & \text{if} \quad y_{i-1} = 1. \end{cases} \tag{1.1}$$

We shall write $y = (y_1, \ldots, y_n)$ and refer to the tuple $\{x, y\}$ as a *Bruceton data* set.

For a large number of explosives, the impact sensitivity can be accurately modelled using probit regression. That is,

$$\mathbb{P}(y_i = 1) = \Phi(\alpha + \beta x_i), \tag{1.2}$$

where α, β are parameters to be estimated and $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-t^2/2} \, dt$ is the standard normal cumulative density function (CDF). The parameters α, β of the model are estimated by maximizing the log-likelihood, which, up to a constant, is given by

$$\ell(\alpha, \beta) = \sum_{i=1}^{n} \{y_i \log \Phi(\alpha + \beta x_i) + (1 - y_i) \log[1 - \Phi(\alpha + \beta x_i)]\}, \tag{1.3}$$

where log denotes the natural logarithm. We let $\hat{\alpha}, \hat{\beta}$ denote the resulting maximum likelihood estimators (MLEs). Since (1.3) is globally concave, $\hat{\alpha}, \hat{\beta}$ may be easily computed numerically with the Newton-Raphson method.

A subtle remark needs to be made about how the generative process (1.1) for the x_i affects the model. Normally, in problems dealing with regression, the inputs x_i are interpreted as fixed, given constants. However, in our case, they are generated by the process (1.1), which depends on the observations y_i. Although the log-likelihood (1.3) looks the same in both paradigms, the underlying models are different, depending on whether or not we condition on the x_i, viewing

them as fixed. This affects, for example, the value of the Fisher information matrix, to be introduced in Section 2.1. However, all results in this article are based on large-sample theory, and so for our purposes, the two paradigms are asymptotically equivalent [10, 11, 12]. We may therefore view the x_i as given without further issues.

It should also be noted at this point that the original article by Dixon and Mood also introduces a way of approximating the MLEs $\hat{\alpha}, \hat{\beta}$ with pen and paper, avoiding any numerical optimisation [9]. However, as numerical optimisation tools are built into virtually any programming language or spreadsheet software nowadays, this approximation has become obsolete and should be avoided. Crucially, there are many instances where the MLEs exist and converge, but where the approximation of them is deemed invalid because a certain numerical criterion is violated [9]. It is particularly unfortunate when researchers apply said approximation and are forced to discard their experimental data on the grounds that they are invalid (where "invalid" here means that the *approximation* is invalid), rather than calculating the MLEs directly and getting a valid answer. Throughout this paper, the term *Bruceton method* refers to the "up-and-down" procedure only, and not to the accompanying approximation of the MLEs.

In the study of energetic materials, we are particularly interested in estimating the median $h_{50} = -\alpha/\beta$, which corresponds to the impact energy level at which there is a probability of $1/2$ of an explosion occurring. Indeed, h_{50} has been shown to correlate with a variety of quantum chemical properties, studied either via statistical modelling or machine learning [13, 14]. However, for practical applications, other quantiles may be more important, such as h_{01} or h_{99}, namely the impact energy levels at which there is a probability of 0.01 and 0.99 of an explosion occurring, respectively. Note that in general, for a probability $p \in [0,1]$, we have, by definition of the probit model, that

$$h_{100p} = \frac{1}{\beta}(z_p - \alpha), \tag{1.4}$$

where $z_p = \Phi^{-1}(p)$. In particular, we recover $h_{50} = -\alpha/\beta$, since $z_{1/2} = 0$. Thus, estimates of any quantile h_{100p} can be obtained from estimates of α, β via $\hat{h}_{100p} = (z_p - \hat{\alpha})/\hat{\beta}$.

When estimating h_{50} (or generally, h_{100p}) for a particular material, it is crucial to address the uncertainty of the estimate by means of a confidence interval (CI). Indeed, a single point estimate $\hat{h}_{50} = -\hat{\alpha}/\hat{\beta}$ is of little value if we do not know how uncertain it is. Thus, in addition to the MLE $\hat{h}_{50}$, we would like to derive a CI $[u(x,y), v(x,y)]$ such that

$$\mathbb{P}\big(u(x,y) \leq h_{50} \leq v(x,y)\big) = 1 - \gamma, \tag{1.5}$$

where $1 - \gamma$ is the desired confidence level, with the most standard choice being $\gamma = 0.05$. Three of the most common ways of constructing such a CI are via the delta method, via the likelihood-ratio test and via Fieller's theorem [15, 16]. In this paper, we shall explore the properties of these methods and how they apply to the case of estimating impact sensitivities from Bruceton data sets.

The remainder of the paper is structured as follows. In the next section, we present the methods for constructing CIs. Then, in Section 3, we determine which method yields the most satisfactory results via simulations. In Section 4, we introduce the notion of confidence curves, which provide an intuitive geometric illustration addressing the confidence in our estimates. We then look at real measurements for cyclotetramethylene-tetranitramine (HMX) in Section 5 and show how confidence curves allow us to extract more information about the explosive from the data. Finally, we conclude and provide a couple of points of discussion in Section 6.

2 Confidence intervals

In this section, we go through the main methods for constructing a CI for h_{50} (and more generally for h_{100p}), starting with the delta method.

2.1. Via the delta method

The first approach relies on the large-sample properties of the MLEs. Let $\theta = (\alpha, \beta)^\top$ and $\hat\theta = (\hat\alpha, \hat\beta)^\top$. The Fisher information matrix $\mathcal{J}(\theta)$ is given by

$$\mathcal{J}(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta^\top}\middle|\theta\right] = \sum_{i=1}^n \frac{\phi^2(\eta_i)}{\Phi(\eta_i)[1-\Phi(\eta_i)]}\begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}, \tag{2.1}$$

where $\eta_i = \alpha + \beta x_i$ and $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ is the standard normal probability density function (PDF). Writing $V = \mathcal{J}(\theta)^{-1}$, the asymptotic normality of the probit model yields the large-sample approximation

$$\hat\theta \approx N(\theta, V), \tag{2.2}$$

where $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. Now, applying the delta method (see for example [17]) in this setting yields the approximation

$$\hat h_{50} \approx N\left(h_{50}, \frac{1}{\beta^2}(V_{11} + 2h_{50}V_{12} + h_{50}^2 V_{22})\right). \tag{2.3}$$

The $1-\gamma$ CI derived via the delta method is thus

$$\hat h_{50} \pm \frac{z_{\gamma/2}}{\hat\beta}\sqrt{V_{11} + 2\hat h_{50}V_{12} + \hat h_{50}^2 V_{22}}. \tag{2.4}$$

In order to derive the corresponding $1-\gamma$ CI for h_{100p}, we just replace $\hat h_{50}$ with $\hat h_{100p} = (z_p - \hat\alpha)/\hat\beta$ in (2.4).

2.2. Via the likelihood-ratio test

The second approach for constructing a CI is via the likelihood-ratio (LR) test. Rather than using the approximation (2.2), the LR test relies on Wilks' theorem [18], which is another convergence result involving the log-likelihood (1.3). The test is used for answering whether we have sufficient evidence to reject the null hypothesis $H_0 : -\alpha/\beta \in \mathbb{R}$ in favour of an alternative hypothesis of the form $H_1 : -\alpha/\beta = \rho$, for some value ρ. The $1-\gamma$ CI will then consist of precisely those values ρ for which the null hypothesis is rejected when the significance level is set at γ. Unlike that derived via the delta method, this CI is in general not symmetric about the MLE $\hat h_{50}$, and can sometimes be unbounded.

In order to perform the test, we first define the *profile log-likelihood* ℓ_{prof} by

$$\ell_{\text{prof}}(\rho) = \sup\{\ell(\alpha, \beta) : -\alpha/\beta = \rho\}. \tag{2.5}$$

That is, the maximum value the log-likelihood function takes given that that H_1 holds. Similarly, we let $\ell_{H_0} = \ell(\hat\alpha, \hat\beta)$ denote the maximum value the log-likelihood function takes overall. In

order to determine whether H_0 should be rejected for a specific value ρ, we use the *deviance function* $D(\rho)$ as our test statistic, which is defined as twice the difference of these two quantities. Namely,

$$D(\rho) = 2\{\ell_{H_0} - \ell_{\mathrm{prof}}(\rho)\}. \tag{2.6}$$

Inferring the probability distribution of $D(\rho)$ is in general an infeasible problem. However, Wilks' theorem asserts that as $n \to \infty$, $D(\rho)$ converges in distribution to a chi-squared distributed random variable with one degree of freedom. Thus, we obtain the approximation $D(\rho) \approx \chi_1^2$, and the resulting CI will comprise all those values ρ which satisfy that $D(\rho) < \Gamma^{-1}(1 - \gamma)$, where Γ is the CDF of the χ_1^2 distribution. By the relation between the chi-squared and normal distributions, one can show that $\Gamma^{-1}(1 - \gamma) = z_{\gamma/2}^2$. Thus, the resulting CI contains those values ρ which satisfy that $D(\rho) < z_{\gamma/2}^2$. One can analogously derive the CI for h_{100p} by replacing $H_1: -\alpha/\beta = \rho$ with $H_1: (z_p - \alpha)/\beta = \rho$ in the above derivation. This means redefining the profile log-likelihood as $\ell_{\mathrm{prof}}^p(\rho) = \sup\{\ell(\alpha, \beta) : (z_p - \alpha)/\beta = \rho\}$ and the deviance function as $D^p(\rho) = 2\{\ell_{H_0} - \ell_{\mathrm{prof}}^p(\rho)\}$.

2.3. Via Fieller's theorem

Finally, the third method for deriving CIs is via Fieller's theorem [19], which in this context can be viewed as a hybrid between the delta method and the LR test. We start the same way as the former, using the asymptotic approximation (2.2). We then write down the log-likelihood for this approximation, which, up to a constant, takes the form

$$\lambda(\theta) = -\frac{1}{2}(\hat{\theta} - \theta)^{\mathsf{T}} V^{-1}(\hat{\theta} - \theta). \tag{2.7}$$

As a function of α and β, this equals

$$\lambda(\alpha, \beta) = -\frac{1}{2}\frac{1}{V_{11}V_{22} - V_{12}^2}\left\{V_{22}(\hat{\alpha} - \alpha)^2 - 2V_{12}(\hat{\alpha} - \alpha)(\hat{\beta} - \beta) + V_{11}(\hat{\beta} - \beta)^2\right\}. \tag{2.8}$$

The CI derived via Fieller's theorem is then obtained from applying the LR test to (2.8) rather than (1.3). In this way, Fieller's theorem can be seen as a special case of Wilks' theorem. A nice property of this approach is that the profile likelihood $\lambda_{\mathrm{prof}}(\rho) = \sup\{\lambda(\alpha, \beta) : -\alpha/\beta = \rho\}$ can be derived analytically. In fact,

$$\lambda_{\mathrm{prof}}(\rho) = -\frac{1}{2}\frac{(\hat{\alpha} + \hat{\beta}\rho)^2}{V_{11} + 2V_{12}\rho + V_{22}\rho^2}. \tag{2.9}$$

This results in the deviance function

$$\Delta(\rho) = 2\{\lambda(\hat{\alpha}, \hat{\beta}) - \lambda_{\mathrm{prof}}(\rho)\} = -2\lambda_{\mathrm{prof}}(\rho) = \frac{(\hat{\alpha} + \hat{\beta}\rho)^2}{V_{11} + 2V_{12}\rho + V_{22}\rho^2}, \tag{2.10}$$

and thus the $1 - \gamma$ CI obtained via Fieller's theorem consists of the values ρ which satisfy that

$$\frac{(\hat{\alpha} + \hat{\beta}\rho)^2}{V_{11} + 2V_{12}\rho + V_{22}\rho^2} < z_{\gamma/2}^2. \tag{2.11}$$

Rearranging, we obtain

$$\left(\hat{\beta}^2 - z_{\gamma/2}^2 V_{22}\right)\rho^2 + 2\left(\hat{\alpha}\hat{\beta} - z_{\gamma/2}^2 V_{12}\right)\rho + \hat{\alpha}^2 - z_{\gamma/2}^2 V_{11} < 0, \tag{2.12}$$

a quadratic inequality in ρ. We now see that if $\hat{\beta}^2 - z_{\gamma/2}^2 V_{22} < 0$, it is possible for the resulting CI to be the entire real line, or the disjoint union of two semi-infinite intervals. On the other hand, when $\hat{\beta}^2 - z_{\gamma/2}^2 V_{22} > 0$, we get a (possibly empty) bounded interval, whose endpoints are given by

$$\hat{h}_{50} + \frac{g}{1-g}\left(\hat{h}_{50} + \frac{V_{12}}{V_{22}}\right)$$
$$\pm \frac{z_{\gamma/2}}{\hat{\beta}(1-g)}\left\{V_{11} + 2\hat{h}_{50}V_{12} + \hat{h}_{50}^2 V_{22} - g\left(V_{11} - \frac{V_{12}^2}{V_{22}}\right)\right\}^{1/2}, \tag{2.13}$$

where $g = z_{\gamma/2}^2 V_{22}/\hat{\beta}^2 < 1$. Note that as expected, this interval is not symmetric about $\hat{h}_{50}$. If we instead wish to derive a CI for h_{100p}, we replace $\hat{h}_{50}$ by $\hat{h}_{100p} = (z_p - \hat{\alpha})/\hat{\beta}$ in (2.13).

A particularly nice property of Fieller's theorem is that if $\hat{\theta} \sim N(\theta, V)$ were to be true, then $\Delta(\rho)$ actually follows a χ_1^2 distribution *exactly*. Hence, unlike the delta method, any inaccuracies in the CIs derived from Fieller's theorem will stem from the approximation $\hat{\theta} \approx N(\theta, V)$ alone.

3   Simulations

To see in practice which of the methods outlined in the previous section yields the most satisfactory results, we conducted two simulation studies in which we calculated CIs for h_{50} and h_{99}. For both quantiles, we looked at data sets of size $n = 30$ and $n = 100$. Various similar studies have been conducted for h_{50} in order to compare the performance of the approaches described above [20–23], and have generally found that the delta method is inferior for small data sets. Our contribution is to report such simulations for Bruceton data sets specifically. In particular, in our simulations, the inputs x_i were not seen as fixed, but were rather generated by the process (1.1) separately for each iteration.

3.1.   Results for h_{50}

In our first simulation study, we generated $T = 10,000$ Bruceton data sets $\{x, y\}_{t=1}^T$ with underlying true parameters $\alpha = 0$ and $\beta = 1$. Note that if we had chosen other values for these two parameters, we could nevertheless rescale the experiment to obtain $\alpha = 0$ and $\beta = 1$. The initial impact energy level x_1 was set to be $x_1 = 0$, whereas the step size d ranged over the values $d \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. We also ran simulations for $d = 0.4$, but for $n = 30$, both Fieller's theorem and the LR test produced unbounded CIs more than 10% of the time. Similarly, for $d = 1.1$, the MLEs failed to converge in more than 10% of the runs when $n = 30$. Consequently, these runs were discarded.

For each of the data sets, we calculated CIs for h_{50} with significance level 0.05 via the delta method, the LR test and Fieller's theorem. All unbounded CIs were discarded. We then checked whether each CI contained the true value $h_{50} = -\alpha/\beta = -0/1 = 0$. For a theoretically exact CI, this will happen 95% of the time. The best performing method is thus that which yielded a coverage probability closest to 0.95. In Table 1, we see the resulting observed coverage

probabilities. As we can see, the delta method consistently underperforms compared to the other methods. For all step sizes, Fieller's theorem provides the best performing CIs, both for $n = 30$ and $n = 100$. However, Fieller's theorem produced unbounded CIs more often than the LR test. For example, for $n = 30$ and step sizes $d = 0.5, \dots, 1.0$, there were 1187, 562, 300, 92, 35 and 11 unbounded CIs by the former method, compared with 563, 12, 2, 0, 0 and 0 for the latter. One way of resolving this is by increasing the step size. Indeed, for $n = 100$, Fieller's theorem yielded unbounded CIs only twice (for $d = 1.0$). To researchers performing sensitivity tests, we recommend employing Fieller's theorem when deriving confidence intervals, due to a consistently better performance.

Table 1. Results from the first simulation study, comparing the coverage probabilities for h_{50} for the delta method (Delta), the likelihood-ratio test (LR) and Fieller's theorem (Fieller). The best coverage probability is typed in bold. *More than 10% of the CIs were unbounded.

	$n = 30$			$n = 100$		
d	Delta	LR	Fieller	Delta	LR	Fieller
0.5	84.33	91.18	**92.86***	91.28	93.45	**93.96**
0.6	85.44	92.43	**92.84**	91.78	93.66	**94.18**
0.7	85.98	92.79	**93.20**	92.05	93.78	**94.19**
0.8	85.96	93.04	**93.52**	92.41	93.95	**94.52**
0.9	86.10	93.48	**93.86**	92.60	93.84	**94.38**
1.0	86.21	93.58	**93.99**	92.80	93.90	**94.52**

3.2. Results for h_{99}

We also repeated the simulation study explained in the previous section for h_{99}. The results are given in Table 2. As we can see, the delta method is still underperforming for all step sizes, both for $n = 30$ and $n = 100$. Unlike the simulation study for h_{50}, we do not really observe a consistent satsifactory performance from any particular method for $n = 30$. This indicates that the sample size $n = 30$ is too small for estimating an extreme quantile like h_{99} [24]. For the increased sample size $n = 100$, we again see that Fieller's theorem is the best performing method overall, and that it produces accurate results for all step sizes. We thus recommend using a sample size of $n = 100$ and Fieller's theorem when estimating an extreme quantile, such as h_{99}, of an impact sensitivity.

Table 2. Results from the first simulation study, comparing coverage probabilities for h_{99}. *More than 10% of the CIs were unbounded.

	$n = 30$			$n = 100$		
d	Delta	LR	Fieller	Delta	LR	Fieller
0.5	72.40	92.33	**94.62***	86.43	93.96	**94.72**
0.6	76.33	92.48	**97.47**	87.94	94.24	**95.24**
0.7	80.96	**93.82**	98.17	89.58	94.08	**95.11**
0.8	81.95	**96.21**	98.17	89.77	94.25	**94.73**
0.9	83.85	**96.78**	98.44	90.10	94.34	**95.50**
1.0	90.17	**97.18**	98.10	90.05	94.09	**95.45**

4 Confidence curves

In this section, we introduce the notion of confidence curves (CCs), which provide a visual interpretation of the confidence in our estimates. Based on the results from the previous section, we shall focus on creating CCs from Fieller's theorem. Recall from Section 2.3 the approximation $\Delta(\rho) \approx \chi_1^2$, where $\Delta(\rho)$ is defined in (2.10). Now, the *confidence curve (CC)* for h_{50} is defined by

$$cc(\rho) = \Gamma(\Delta(\rho)) = \Gamma\left(\frac{(\hat{\alpha} + \hat{\beta}\rho)^2}{V_{11} + 2V_{12}\rho + V_{22}\rho^2}\right), \tag{4.1}$$

where we recall that Γ denotes the CDF of the χ_1^2 distribution. More generally, the CC for h_{100p} is defined by

$$cc^p(\rho) = \Gamma\left(\frac{(\hat{\alpha} - z_p + \hat{\beta}\rho)^2}{V_{11} + 2V_{12}\rho + V_{22}\rho^2}\right). \tag{4.2}$$

Now, it is not difficult to show that as $n \to \infty$, we have

$$\mathbb{P}(\{\rho : cc(\rho) \le 1 - \gamma\}) \to 1 - \gamma \tag{4.3}$$

for all $\gamma \in (0,1)$. Hence, we can read off CIs of any desired confidence level $1 - \gamma$ directly from the curve.

How to read information from a CC is best seen with an example. In the following illustration, we generated a Bruceton data set with $n = 30$, $\alpha = 0$, $\beta = 1$, $x_1 = 0$, $d = 0.7$, and used (4.1) to obtain the CC drawn in Figure 1. Here, we can read off the CI for any wanted significance level γ. For example, the points at which the curve intersects the line $y = 0.95$ gives the endpoints of the 95% CI for h_{50}, namely $[-0.53, \ 0.33]$. Note also that the curve is zero at the MLE $\hat{h}_{50} = -0.06$.
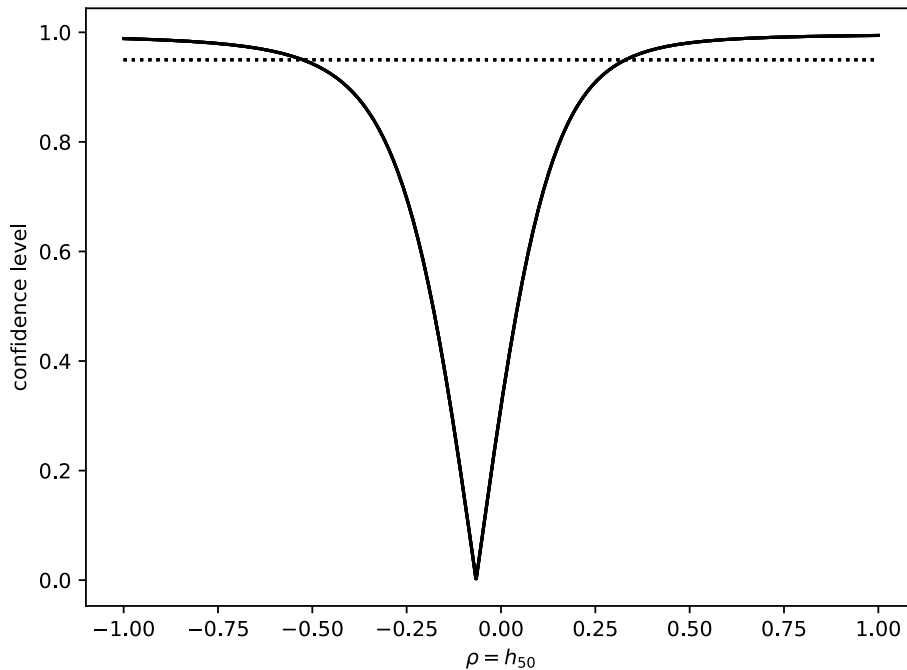
Figure 1. A CC for h_{50} after $n = 30$ observations. The solid curve is the actual confidence curve, and the dotted line is given by $y = 0.95$. The points at which this line intersects the curve is the 95% CI for h_{50}.

5 Application: HMX

In this section, we demonstrate how to calculate CIs in practice with a real data example. We performed $n = 100$ drops in a Bundesanstalt für Materialprüfung (BAM) fallhammer apparatus on 40 mm³ samples of Eurenco HMX Class 1. The molecular structure and a scanning electron micrograph of the crystals are shown in Figure 2. We used a 2 kg drop weight with $x_1 = 1.30$ and $d = 0.05$ on a $\log_{10}$ scale. That is, the height of the first drop was $10^{1.30} = 19.95$ cm, and so on. The data from the experiment is given in Table 3.
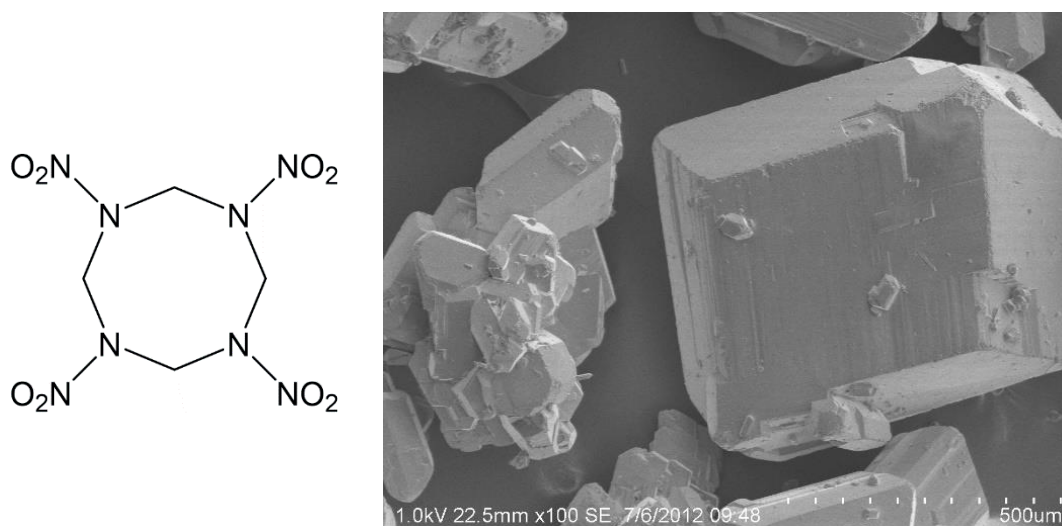
Figure 2. Molecular structure of HMX (left) and scanning electron micrograph of HMX Class 1 crystals. The distance between each mark on the bar scale is 50 μm.

Table 3. Bruceton data from the fallhammer drops on HMX with $n = 100$.

Height ($\log_{10}$)	1.15	1.20	1.25	1.30	1.35	1.40
# Explosions	0	3	18	15	12	2
# Trials	3	21	33	27	14	2

Based on these data, we obtain $\hat{\alpha} = -15.53$ and $\hat{\beta} = 12.26$, which further yields (still on a $\log_{10}$ scale) $\hat{h}_{50} = 1.27$ and $\hat{h}_{99} = 1.46$, with 95% CIs $[1.24, 1.29]$ and $[1.40, 1.61]$, respectively. For the sake of completeness, we also computed $\hat{h}_{01} = 1.08$, with the 95% CI $[0.93, 1.14]$. Converting back to cm, this corresponds to $10^{\hat{h}_{01}} = 11.94$ cm, $10^{\hat{h}_{50}} = 18.49$ cm and $10^{\hat{h}_{99}} = 28.62$ cm, with 95% confidence intervals $[8.43, 13.73]$, $[17.50, 19.55]$ and $[24.87, 40.68]$, respectively. In Figure 2, we see CCs for h_{50} and h_{99} on the $\log_{10}$ scale. We see that the CC for h_{50} is much narrower than that for h_{99}, reflecting that the former is easier to estimate than the latter.
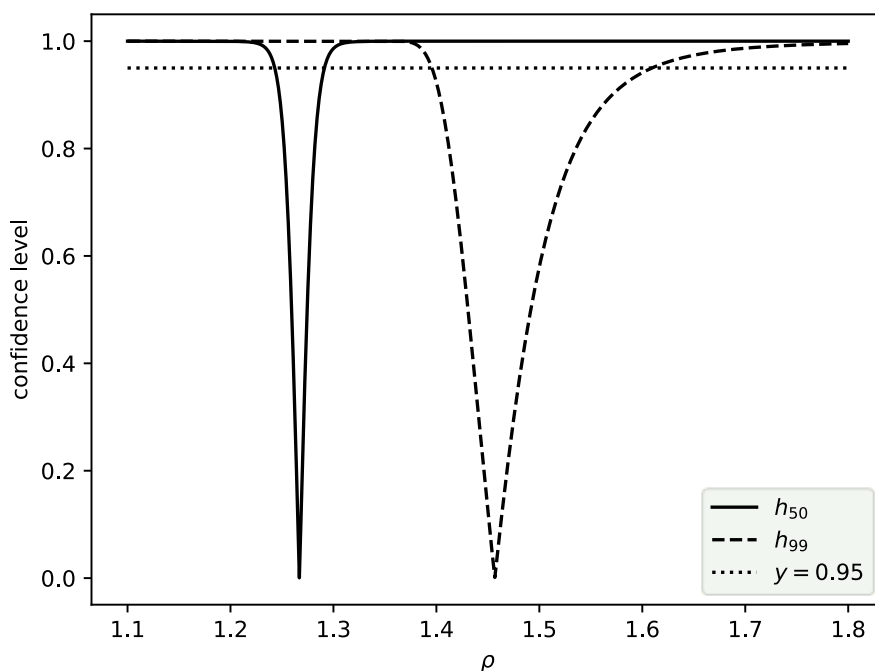
Figure 3. Confidence curves for h_{50} and h_{99} for HMX based on Bruceton data with $n = 100$.

6 Conclusion

In this paper, we have covered the main methods for computing confidence intervals for h_{50} (and more generally for h_{100p} where $p \in [0,1]$), which is crucial to ensure safe handling of energetic materials, including during synthesis of molecules. We found in our simulation study that Fieller's theorem yielded satisfactory results for a standard Bruceton data set with $n = 30$ when estimating h_{50}, and thus recommend that researchers employ this method. For h_{99}, we found that a larger sample size, such as $n = 100$, was necessary in order to obtain similar results, again via Fieller's theorem. We then calculated confidence intervals for h_{01}, h_{50} and h_{99} for HMX from a newly obtained Bruceton data set.

Through the calculation of confidence intervals and confidence curves, we have seen how to extract more information from sensitivity data, allowing us to assess the uncertainty of our estimates. We therefore strongly recommend to all researchers working with impact sensitivity that they include their raw data in their publications, rather than simply reporting point estimates of α and β.

In future work, it would be useful to consider more refined techniques for constructing confidence intervals. In particular, one may improve the likelihood-ratio test, for instance by applying the Bartlett correction [25–27]. We refer the reader to [28–30] for further reading on this point. Also, in this article, we have only considered the construction of confidence intervals for h_{100p} for a single explosive. It would be interesting to conduct a similar simulation study to investigate approaches to hypothesis testing for comparing the sensitivities of two or more explosives.

References

[1] Transparency Market Research report. RDX and HMX Market – Global industry analysis, size, share, growth, trends, and forecast, 2021–2031. Report ID TMRGL82506, 2021.
[2] J. Wildbrand. Notiz über trinitrotoluol. *Annalen der Chemie und Pharmacie*, 128(2):178–179, 1863.

[3]   G.F. Henning. Verfahren zur Darstellung eines Nitrokörpers aus Hexamethylentetramin. Patentschrift Nr. 104280, Berlin, 1898.

[4]   W.E. Bachmann and J.C. Sheehan. A new method of perparing the high explosive RDX. *Journal of the American Chemical Society*, 71(5):1842–1845, 1949.

[5]   N.V. Latypov, J. Bergman, A. Langlet, U. Wellmar and U. Bremm. Synthesis and reactions of 1,1-diamino-2,2-dinitroethylene. *Tetrahedron*, 54(38):11525–11536, 1998.

[6]   W. Trzciński and A. Belaada. 1,1-Diamino-2,2-dinitroethene (DADNE, FOX-7) – Properties and formulations (a review). *Central European Journal of Energetic Materials*, 13(2):527–533, 2016.

[7]   North Atlantic Treaty Organization (NATO). STANAG 4489: Explosives, impact sensitivity tests, 1999.

[8]   U.S. Department of Defense. MIL-STD-1751A: Safety and performance tests for the qualification of explosives (high explosives, propellants, and pyrotechnics), 2001.

[9]   W.J. Dixon and A.M. Mood. A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association*, 43(241):109–126, 1948.

[10]   M.J. Crowder. Maximum likelihood estimation for dependent observations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(1):45–53, 1976.

[11]   P.J. Bickel, Y. Ritov and T. Rydén. Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *The Annals of Statistics*, 26(4):1614–1635, 1998.

[12]   P. Giudici, T. Rydén and P. Vandekerkhove. Likelihood-ratio tests for hidden Markov models. *Biometrics*, 56(3):742–747, 2000.

[13]   T.L. Jensen, E. Unneberg, J.F. Moxnes and D. Christensen. Models for predicting impact sensitivity of energetic materials based on the trigger linkage hypothesis and Arrhenius kinetics. *Journal of Molecular Modeling*, 26(4):1–14, 2020.

[14]   J.L. Lansford, B.C. Barnes, B.M. Rice and K.F. Jensen. Building chemical property models for energetic materials from small datasets using a transfer learning approach. *Journal of Chemical Information and Modeling*, 62(22):5397–5410, 2022.

[15]   D. Finney. *Probit Analysis*. Cambridge University Press, London, third edition, 1971.

[16]   B.J.T. Morgan. *Analysis of Quantal Response Data*. Chapman & Hall, London, 1992.

[17]   T.S. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall, London, 1996.

[18]   S.S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

[19]   E.C. Fieller. Some problems in interval estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 16(2):175–185, 1954.

[20]   K.M. Abdelbasit and R.L. Plackett. Experimental design for binary data. *Journal of the American Statistical Association*, 78(381):90–98, 1983.

[21]   C. Cox. Fieller's theorem, the likelihood and the delta method. *Biometrics*, 46(3):709–817, 1990.

[22]   R.R. Sitter and C.F.J. Wu. On the accuracy of Fieller intervals for binary response data. *Journal of the American Statistical Association*, 8(423):1021–1025, 1993.

[23]   D. Faraggi, P. Izikson and B. Reiser. Confidence intervals for the 50 per cent response dose. *Statistics in Medicine*, 22(12):1977–1988, 2003.

[24]   L.J. Young and R.G. Easterling. Estimation of extreme quantiles based on sensitivity tests: A comparative study. *Technometrics*, 36(1)48–60, 1994.

[25]   M.S. Bartlett. Properties of sufficiency and statistical tests. In *Proceedings of the Royal Society of London, Series A*, 160(901)268–282, 1937.

[26]   D.N. Lawley. A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, 43(3):295–303, 1956.

[27]   G.M. Corderio. Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(3):404–413, 1983.

[28]   P. Harris, M. Hann, S.P.J. Kirby and J.C. Dearden. Interval estimation of the median effective dose for a logistic dose-response curve. *Journal of Applied Statistics*, 26(6):715–722, 1999.

[29]   R.L. Paige, P.L. Chapman and R.W. Butler. Small sample LD50 confidence intervals using saddlepoint approximations. *Journal of the American Statistical Association*, 106(493):334-344, 2011.

[30]   T. Schweder and N.L. Hjort. *Confidence, Likelihood, Probability: Statistical Inference with Confidence Distributions*. Cambridge University Press, Cambridge, 2016.

Paper V

Sequential experimental designs in regression: Theory for the Bruceton and Langlie designs

Paper VI

Estimating sensitivity with the Bruceton method: Setting the record straight

Paper VII

Increased impact sensitivity in ageing high explosives; analysis of Amatol extracted from explosive remnants of war

Increased impact sensitivity in ageing high explosives; analysis of Amatol extracted from explosive remnants of war

Geir P. Novik[ac*] & Dennis Christensen[bc]

a. Department of Safety, Economics and Planning, Faculty of Science and Technology, University of Stavanger, P.O. Box 8600, 4036 Stavanger, Norway
b. Department of Mathematics
   University of Oslo, P.O. Box 1053 Blindern, 0316 Oslo
c. Norwegian Defence Research Establishment (FFI), P.O. Box 25, 2027 Kjeller, Norway

* Corresponding author: Geir P. Novik, e-mail: geir-petter.novik@ffi.no

Abstract

Millions of tonnes of explosive remnants of war remain in nature, and the volume is continuously growing. The explosive legacy of wars represents an increasing threat to the environment and to societal safety and security. As munitions continue to deteriorate, harmful constituents will eventually leak into the environment, poisoning ecological receptors and contaminating the surrounding soil and groundwater. Moreover, deterioration due to exposure to various environmental factors may ultimately cause the munitions to become increasingly sensitive to external stimuli and susceptible to accidental detonation. To thoroughly assess how to address these ageing munitions, we must first establish certain threshold values for the safe and secure handling and final disposal of the explosive ordnance. One key factor is to establish how the impact sensitivity of the explosives evolves over time. In the present work, we investigated the high explosive substance Amatol extracted from ageing explosive remnants of war. The results obtained in the analysis indicate that the high explosives in the examined specimens generally were much more sensitive to impact than previously assumed. Furthermore, the analysis revealed that the standardised methodology of impact sensitivity testing was insufficient for estimating the sensitivities in question, and a more careful statistical analysis was required.

Keywords

Unexploded ordnance; ERW; the Bruceton method; sensitivity analysis; high explosives

1. Introduction

In the event known as the Shell Crisis of 1915, the stock of UK artillery shells was unexpectedly depleted due to an unanticipated and prolonged period with a high rate of fire on the front lines in World War I (WWI). It soon became evident that the supply of high explosives in use (predominantly 2,4,6-trinitrotoluene [TNT] and 2,4,6-trinitrophenol [picric acid]) was insufficient (Fedoroff, Aaronson, Reese, Sheffield, & Clift, 1960). To eke out the available supply of TNT for shell, grenade and bomb fillings, the Research Department at the Royal Arsenal in Woolwich developed mixtures of ammonium nitrate and TNT. These binary mixtures, known as Amatols, were easy to manufacture

and exhibited several favourable properties, including the effectiveness they exhibited in shell-bursting trials. Ammonium nitrate, which was being manufactured from atmospheric nitrogen for the first time, was a readily available explosive ingredient and was all the more valuable since it leaves no solid residue upon decomposition and ensures a high volume of gaseous explosion products (Urbanski, 1967). Live fire gun trials substantiated the trials at rest, and the adoption of Amatols as high explosive fillings in munitions followed quickly thereafter (Robertson, 1920).

In addition to being an easily available explosive in a time of necessity, the Amatols also enabled a highly economical output of explosive material, as the cost of ammonium nitrate was about one quarter of that of TNT. The Amatols were therefore proposed with the objective of economising the volume of TNT and simultaneously taking advantage of the excess of oxygen contained by ammonium nitrate to compensate partially or completely for TNT's oxygen deficiency (The War Office [UK], 1925). For similar reasons, several governments authorised its use shortly after Great Britain (e.g. War Department [US], 1944).

TNT and Amatol were the preferred high explosive fillings for most high explosive artillery shells at the outset of World War II (WWII), largely due to their availability and their combination of high power and low sensitivity. In particular, they were easy and safe to handle and transport. Towards the end of WWII, the rapid production of an enormous supply of TNT eventually removed the necessity of using ammonium nitrate as a substitute for TNT. Another contributing factor to the disuse of Amatols as high explosives in munitions was the emergence of other explosives during WWII, such as pentaerythritol tetranitrate (PETN) and cyclotrimethylenetrinitramine (RDX) and their binary and ternary mixtures, which are more powerful than TNT (U.S. Army Material Command, 1965).

Although they are now mostly obsolescent, Amatols were universally used for several decades by all nations in all types of ammunition as a substitute for TNT (U.S. Army Material Command, 1965). Consequently, the only time Amatols are normally encountered in explosive ordnance today is in legacy munitions, at ammunition dumping sites and in explosive remnants of war (ERW).

As a considerable percentage of both WWI and WWII munitions contained Amatol filling, its ageing characteristics are a subject of immense importance. Several studies have revealed that the deterioration of explosive fillers can make the munitions increasingly sensitive to external stimuli and susceptible to detonation if subjected to heat, shock or friction (Albright, 2012; Hamer, 2004; Long, 2005; Pfeiffer, 2012). An increasing number of spontaneous detonations have also been reported in ageing munitions, possibly resulting from deteriorating or changing technical or chemical properties (Ford, Ottemöller, & Bapite, 2005; Nordaas, 2019). Earlier studies regarding samples of high explosives extracted from ERW (e.g. TNT and PETN) have indicated that the impact sensitivity of ageing explosives does not appear to have been reduced over the last eight decades and that in some cases, the explosives can even become increasingly sensitive to stress (Geir P. Novik, 2022). Some reports have also indicated that under specific circumstances, Amatols can form dangerous compounds that may increase their sensitivity (i.e. Picatinny Arsenal, 1943; U.S. Army Material Command, 1965; War Department [US], 1944). However, few studies have analysed the properties of ageing Amatols in ERW. Consequently, we do not have sufficient data available to properly assess the risks related to spontaneous detonation or the clearance and handling of ERW with Amatol filling.

ERW at terrestrial and aquatic sites also present an international environmental problem due to the release of explosive material from the corroding ordnance, in addition to the risks associated with the potential for accidental detonations (Sunahara et al., 2009). Similar to most explosive fillings used in munitions, Amatols represent a source of contamination which can be toxic to ecological

receptors, damaging the impacted sites and surrounding areas exposed to the offsite migration of contaminants. As several of the chemicals used in ammunition are highly poisonous and have been proven to contaminate living organisms and the surrounding soil and groundwater, the leakage and bioaccumulation of toxic constituents from corrosive munitions pose a formidable threat to the ecosystem (ATSDR, 1995; Koske, Goldenstein, & Kammann, 2019; Koske et al., 2020; Schuster et al., 2021; Yinon, 1990). Some munitions' constituents have also been proven to enter the food chain and could therefore directly affect human health upon the consumption of contaminated food (Maser & Strehse, 2021).

As munition casings continue to deteriorate, we expect an increase in the release of their harmful constituents in the future (Geir P. Novik, Sommer, & Abrahamsen, 2022). Consequently, there is a time limit pertaining to when ERW and their explosive fillings can be identified and handled safely and appropriately based on an assessment of the viable options. As a result of the potential dangers related to ERW risks, their removal is a highly prioritised task for many countries and international organisations, such as the North Atlantic Treaty Organization (NATO) and the United Nations (UN) (Geir P. Novik, Abrahamsen, & Sommer, 2023). To properly assess and ideally mitigate the risks related to accidental detonations and the uncontrolled release of harmful substances, we rely on accurate data and a proper statistical analysis of the specifics of the applicable ERW, including its sensitivity.

In this study, we have analysed the impact sensitivity of Amatol extracted from ageing ERW via statistical analyses of new fallhammer measurements. In accordance with the recommendations of Christensen et al. (2023), we have employed the Bruceton up-and-down test procedure and compute confidence intervals using Fieller's theorem. Our analysis shows that all collected samples were more sensitive than the standard reported value for Amatol in the literature. This study therefore also serves to illustrate why these standards are insufficient and require an update.

2. Materials and methods
2.1 Sample characteristics

The first experiments using ammonium nitrate ($H_4N_2O_3$) as a component in explosive mixtures began in the second half of the 19th century, although the substance was originally discovered 200 years earlier (Urbanski, 1965). However, it only gained supreme military importance as an ingredient of high explosives during WWI (The War Office [UK], 1925). One of the most commonly used military high explosives at the outbreak of WWI was TNT ($C_7H_5N_3O_6$). This was partially due to its explosive characteristics (i.e. high output and low sensitivity) but also because of its ease of manufacture and its suitability for melt loading, either as a pure explosive or as a binary mixture (Gibbs & Popolato, 1980). Since the colossal demand for high explosives in WWI could not be fulfilled by the output of explosives such as TNT and picric acid, various compositions, such as mixtures of aromatic compounds with ammonium nitrate, were introduced and widely implemented (Urbanski, 1967). These compositions were explosives in which two or more explosive compounds were mixed to produce an explosive substance with more suitable characteristics. Generally, the properties of the composition exhibit an intermediate state between those of the individual explosive ingredients (U.S. Department of the Army, 1984).

Amatols ($C_7H_9N_5O_9$) are binary mixtures of ammonium nitrate and TNT, as illustrated in Figure 1. Compared to TNT, they were cheaper to produce, and they produced greater volumes of gas per unit weight upon explosion (Fedoroff et al., 1960). When TNT detonates, free carbon is present, suggesting that it is deficient in oxygen (The War Office [UK], 1925). For Amatols, on the other hand, the addition of ammonium nitrate, which is rich in oxygen, yields a more complete combustion of the

TNT component. For this reason, the smoke produced by the detonation of Amatol has a light white-yellowish colour, in contrast to the heavy black smoke produced by the detonation of pure TNT (Fedoroff et al., 1960). In general, due to its tendency to increase chemical stability and decrease sensitivity to friction and shock, ammonium nitrate is the most widely used oxygen carrier in explosives (Urbanski, 1965). Although it is technically possible to detonate straight ammonium nitrate given a sufficiently powerful impulse, its chemical properties suggest that it should not be used alone as an explosive (The War Office [UK], 1925).

There are many types of Amatols, which differ only in terms of the proportion at which TNT and ammonium nitrate are present (The War Office [UK], 1925). The constitution of any one of these is reflected in the nomenclature for each mixture. Thus, *Amatol 80/20* denotes a mixture of 80% ammonium nitrate with 20% TNT, by mass. Generally, the first number invariably refers to the percentage of ammonium nitrate, although in German nomenclature, the Amatol compositions (known as various types of Füllpulver, shortened Fp. 60/40, 20/80, etc.), the numerators refer to the percentage of TNT present. The principal Amatols are 40/60 and 80/20. Examples of other proportions that have been used are 45/55, 50/50, 83/17 and 90/10.

a)   **b)**   **c)**
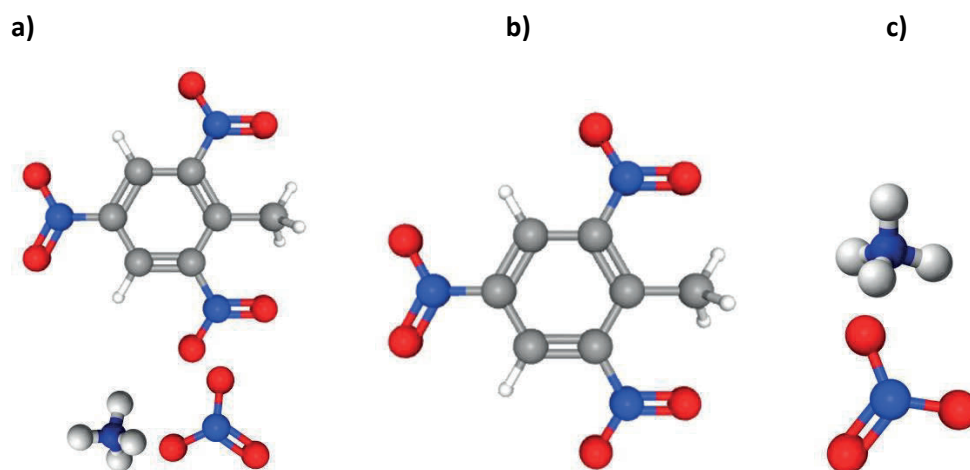
Fig. 1. Chemical formulas of a) Amatol and its constituents; b) TNT; and c) ammonium nitrate (PubChem, 2024).

Amatols were commonly used during WWI and WWII in most countries. Particularly, the mixture consisting of 40% ammonium nitrate and 60% TNT accrued immense importance (Urbanski, 1967). It was known in Germany as Füllpulver No 13 or Fp. 60/40 and in Great Britain as Amatol 40/60, and it was cast loaded in a broad variety of bombs and shells (Fedoroff, Aaronson, Clift, & Reese, 1958). However, as Amatols are generally considered comparatively insensitive, they require a special exploder system to ensure their complete detonation (The War Office [UK], 1925). When efficiently detonated, Amatol 40/60 is slightly less violent than TNT alone. Owing to the hygroscopic nature of ammonium nitrate, Amatols are considered highly unstable in storage, unless it is possible to exclude moisture. For example, at 90% relative humidity (RH) and 30 °C, Amatol 80/20 would contain approximately 61% moisture in two days. This would not only lower the sensitivity and velocity of the detonation to a low order, but it could result in a failure to detonate (Fedoroff et al., 1960). Another effect that has been observed as a result of the exposure to moisture and high temperatures is that Amatol may congeal into a dense, hard mass as a result of changes in the crystalline form of the ammonium nitrate (The War Office [UK], 1925).

In the existing literature, Amatols are generally considered to be equally or less sensitive to impact than TNT (e.g. Fedoroff et al., 1960; Hershkowitz & Akst, 1975; The War Office [UK], 1925; U.S. Army Material Command, 1965; U.S. Department of the Army, 1984; U.S. War Department, 1945; War Department [US], 1944). However, some reports indicate that the introduction of impurities to the production of Amatol can result in slightly increased sensitivity compared to that of pure TNT (The War Office [UK], 1925). According to a study by Hackel (1937, as cited in Urbanski, 1967), the impact sensitivity of mixtures of nitro compounds with ammonium nitrate (Amatols) was found to be higher than that for pure nitro compounds due to the friction produced by the hard crystals of the ammonium nitrate. In this study, Hackel found mixtures containing 30 to 60% of ammonium nitrate to be equally as sensitive as picric acid, an explosive substance that is slightly more sensitive to impact than TNT (U.S. Army Material Command, 1965). However, due to the hygroscopic nature of ammonium nitrate, it will begin to deteriorate if exposed to water, and studies have demonstrated that the explosive compositions containing ammonium nitrate can become progressively less sensitive to impact as the moisture content increases (Montesi & Menichelli, 1964). It has also been proven that the impact sensitivity can be reduced to such a level where the amount of force required for the initiation of the substances renders them impracticable as explosives (e.g. Johansen, 2005), as standard means of initiation would result in failures to detonate (Fedoroff et al., 1960). Moreover, a high moisture content can decrease the detonation velocity, which in many cases would entail that a continuation of an explosive shockwave within the substance by its own means is unachievable.

However, studies have also revealed that the presence of moisture, when combined with other factors, can contribute to an increase of the impact sensitivity in Amatols; it is known that explosive compositions containing ammonium nitrate might become sensitised if contaminated with small amounts of metals or if the composition comes into contact with metal, as these contaminating metals might react chemically with the ammonium nitrate, forming complex salts and sensitising the mixture (Montesi & Menichelli, 1964). A contamination of the Amatols could occur during normal handling and mixing, or it could come into contact with bare metal surfaces when loaded in ordnance or if any preventive lacquers deteriorate over time. An investigation of the stability of various mixtures of ammonium nitrate and TNT conducted after WWII at the Laboratoire Centrale des Poudres in Paris also demonstrated that mixtures of military grade TNT and pure ammonium nitrate had in some cases decomposed with the evolution of ammonia that attacked the TNT to form various unstable coloured compounds (F. M. Lang and J. Boileau, 1952, as cited in Fedoroff et al., 1960). According to Fedoroff et al. (1960), in the presence of iron, the hydrolysis of moist ammonium nitrate may occur with the formation of ammonia solution (NH_4OH), which reacts with TNT to form an exudate of a brown oily material igniting at 67°C . This can be detected by the discolouration of the explosive and the odour of ammonia (NH_3). In addition to being reactive to iron, mixtures of Amatols may, in the presence of moisture, also attack metals such as copper, brass, bronze and lead and can form dangerously sensitive compounds with copper and its alloys (Picatinny Arsenal, 1943; U.S. Army Material Command, 1965). However, as this was a well-known attribute of Amatols, it was considered general practice at the time to coat the insides of munitions with acid-proof paint prior to loading to prevent the corrosion caused by contact between Amatols and metal (Fedoroff et al., 1960).

2.2 Sampling location and methodology

To ensure the reliability of the data, all samples of Amatol in this study have been extracted from live ordnance originating from WWII during national explosive ordnance disposal (EOD) clearance operations in Norway. Consequently, all explosive objects utilised in this analysis originate from explosive ordnance that were produced before May 1945. All the explosive objects were localised and reported to the relevant governmental agencies by members of the public prior to their

exploitation and final disposal. In all cases, the munitions were subject to handling (moving the object) by the discoverer or by the designated EOD team. All the samples of high explosives were extracted from the relevant objects and analysed within the last three years (2021 - 2023). The first author personally executed the physical extraction of the high explosives from the ordnance. In all situations, it was deemed safe to move the explosive objects to a site suitable for the extraction of the high explosive samples as well as the final disposal of the ordnance. All explosive objects included in this study are of German origin and were located in an area heavily contaminated with explosive remnants of WWII, namely Finnmark county in the northernmost region of Norway. The required disassembly of the ordnance to gain access to their high explosive fillings was performed with the use of explosive cutting charges (shaped charges), as indicated in the example in Figure 2.

After the required dismantlement of the explosive ordnance, an initial sample of the high explosives was retrieved at the point of entry, centre mass of the explosives. As some of the studied objects were found to contain several types of high explosive fillings, multiple samples were collected of the various compounds (in the case of Amatol fillings, casted TNT was frequently used as a seal to prevent any moisture from coming into contact with the hygroscopic Amatols).

In total, high explosive samples were collected from over 20 unexploded objects potentially containing Amatol fillings. Of these, five unique samples of Amatol were included in this study. Of these five, three were discovered with their respective fuzes installed and (based on a visual inspection) in a condition that would suggest that they were still fully intact (i.e. no visible cracks or fractures in their outer casings). The remaining two objects were also apparently intact but were found without fuzes installed, increasing their explosive fillings' exposure to environmental factors. Of these two, one was retrieved from an ammunition dumping site (lake) at about a five-metre depth. The remaining four objects were all located on land. The specifics for the particular objects are as follows: one German HE artillery projectile (no fuze, located in water [hereafter designated as substance A]), one German HE aerial bomb (no fuze, located on land [substance B]), one German HE artillery projectile (fuze installed, located on land [substance C]), two German HE mortar projectiles (fuze installed, located on land [substances D and E, respectively]).

According to the relevant literature, all ordnance included in this study were identified as carrying high explosive fillings of the substance Füllpulver 60/40, otherwise known as Amatol 40/60 (Der Reichsminister der Luftfart, 1942; Ordnance Bomb Disposal Center [US]; The War Office [UK], 1944; U.S. War Office, 1953a, 1953b). The identification of Amatols was later confirmed by analysing the samples with an ion chromatograph and an ultra-performance liquid chromatography-mass spectrometer (UPLC-MS/MS).

Fig. 2. A German 88 mm HE projectile (type 8.8 cm Sprgr. Patr. L/4.5 (Kz)) cut with a flexible linear-shaped charge. Its filling is identified as Füllpulver 60/40 (i.e., Amatol 40/60).

2.3 Storage and preparation of samples

Once extracted, the Amatol samples were immediately placed in airtight containers (50 ml sterile polypropylene screw-cap tube) and stored in approved ammunition storage facilities. Aside from humidity control (at most 50% RH), the samples were stored in normal atmospheric conditions, in continuation of the normal temperature fluctuations that would appear in nature, albeit with less violent variations, as the samples were stored under cover and protected from direct sunlight.

The physical appearance of all samples was found to resemble a grainy, brown sugar-like form (as opposed to the white to light buff colour normally associated with Amatol), indicating the presence of impurities in the composition or that the Amatols had been subjected to exposure to light and moisture (U.S. Army Material Command, 1965).

At the time of extraction, all substances, with the exception of substance A, appeared to be dry and powdery. We established the exact moisture content of each sample as follows. First, we introduced a dried Pyrex crystalliser with a ribbed cover, of combined mass W_1 (all masses were accurate up to 1/10 mg). The ribbed cover was used to catch the small amounts of TNT which sublime upon heating (Fedoroff et al., 1960). Each substance was then analysed by adding a sample of mass W_S to the crystalliser. The total mass of the crystalliser, cover and sample was obtained, and the specimen was heated for 2 to 3 hours at a temperature of 75°C and then cooled in a desiccator. Letting W_2 denote the combined mass of the specimen after this process, the original moisture content w (as a percentage) of the sample is yielded by

$$w = 100\{W_S - (W_2 - W_1)\}/W_S.$$

It was found that the moisture content of substance A was 22.2% and that the remaining substances had moisture contents of 0.33 ± 0.25%. These results coincide with the samples' individual physical appearances at the time of extraction. However, their discolouration indicates that all of the tested substances may have undergone some exposure to moisture at one point in time.

Preceding the impact sensitivity analysis, the samples were prepared in accordance with the requirements of NATO STANAG 4489 (NATO, 1999) and the United Nations Manual of Tests and Criteria - Classification Procedures, Test Methods and Criteria Relating to Explosives, Test 3 (a) (ii) (United Nations, 2019). Powdered substances are to be sieved, and only the fraction with a particle size of 0.5-1.0 mm is to be used for testing. For pressed or cast substances, where the powder was excessively coarse to all pass the sieve, the particle sizes are reduced by gently crushing it using a pestle and mortar. Only the fraction passing a 1000 µm sieve and retained on a 500 µm sieve was used for the test.

As one of the substances consisted of a paste-like material (substance A), it was treated as a *paste-like or gel-type substance* as per United Nations (2019, p. 86) test procedures, wherein a cylindrical tube of 40 mm^3 capacity (3.7 mm diameter and 3.7 mm height) is inserted into the substance, and, after levelling off the surplus, the sample is removed from the tube by means of a wooden rod. A sample from this substance was placed in a humidity-controlled environment to reduce the moisture level of the sample to about 0.5% in preparation for further analysis. This particular substance underwent analysis in both its original (22.2% moisture) and prepared (0.33% moisture) state, hereafter denoted respectively as substance A_1 (original) and substance A_2 (prepared).

2.4 Impact sensitivity testing

The impact sensitivity of an explosive substance is its susceptibility to detonation under impact. This parameter characterises the safety of explosives in handling and transportation (Rădeanu, Rus, Jitea, Miron, & Vasilescu, 2020). To determine the impact sensitivity of a substance, a type of device known as a fallhammer apparatus is normally applied. There are several versions of these types of devices, but the United Nations recommends the Bundesanstalt für Materialforschung und -prüfung (BAM) fallhammer, which has also become the most frequently used standard impact sensitivity measuring device (Gruhne et al., 2019). However, the various apparatuses all operate on the same principle: A sample of assorted sizes of the tested explosive substance are subjected to the impact of falling weights, and the researcher estimates the sensitivity of the explosive based on which heights resulted in explosions (Meyer, Köhler, & Homburg, 2005). The main differences between the various fall hammer apparatuses are mainly related to their design and the manner in which the sample is subjected to the drop weight impact via different types of plungers (Suceska, 1995). It is currently an active area of research to better understand how energy is transferred through the explosive sample in the fallhammer test (e.g., Monogarov, Meerov, Fomenkov, & Pivkina, 2023; Samseth, 2022).

The BAM fallhammer test was initially developed to obtain better reproductive data compared to that of existing tests at the time (Meyer et al., 2005), and is generally considered to yield reasonably reproducible results (Suceska, 1995). In this analysis, the OZM BFH 12 BAM Impact Apparatus was applied, and the tests were performed in accordance with the requirements of the test procedure described in NATO STANAG 4489, Annex C; BAM Impact Machine (NATO, 1999). The BAM Impact Machine is presented in Figure 3a. The essential parts of the BAM fallhammer are the steel block with the base, the anvil, the guiding rods, the drop weight with the locking and unlocking device and the impact device. The impact device, as presented in Figure 3b, consists of two coaxially arranged

steel cylinders with polished surfaces and rounded edges, held in place by a cylindrical steel guide ring with an inner diameter of 10 mm.

a)   b)
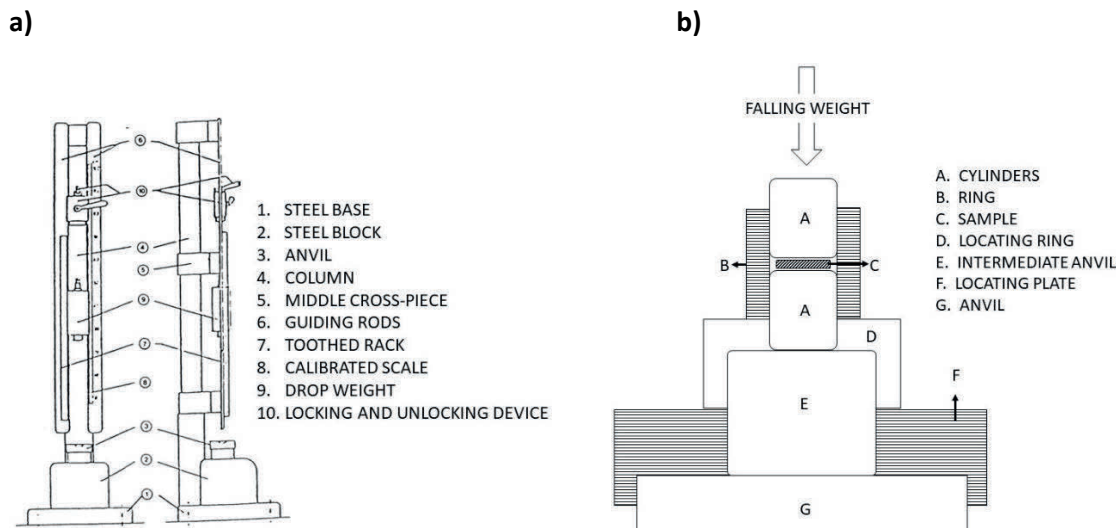
Fig. 3. a) The BAM impact machine (NATO, 1999) and b) The fallhammer impact device (NATO, 1999).

The device is prepared by partially pushing one of the cylinders into a guide ring and positioning it on the intermediate anvil fitted with a locating ring. Using a measuring spoon, 40 mm^3 of the prepared high explosive samples (e.g. crushed and sieved to a particle size of 500-1000 μm) are placed inside the impact device, ensuring that a centre heap is formed. The impact device is then closed with a second steel cylinder by carefully pressing it into the guide ring until it touches the sample. For the impact sensitivity testing, assorted drop weights with masses ranging from 0.25 kg to 10 kg are available. The body of each drop weight has two guide grooves, in which it moves between the guide rails. It is equipped with a suspension spigot that arrests the weight in the release mechanism and is further provided with a cylindrical striker, a height marker and the rebound catch for stopping the weight after rebounding from the anvil. Based on the anticipated results (e.g. on the basis of the specific characteristics of the explosive substance undergoing the test), the drop weight is secured in the release mechanism, and the weight is then positioned to the desired height. When the release mechanism is activated, the drop weight is unlocked, and its striking head impacts the upper cylinder of the impact device.

Depending on the characteristics of the tested explosive substance, the mass of the drop weight and the drop height (the combined product of which is the impact energy), the sample may or may not initiate upon impact. In judging the results, a distinction is made between *no reaction*, *decomposition* (without flame or explosion) and *explosion* (with weak to strong report or inflammation). The verification of *decomposition* and *explosion* can be recognised based on several factors, including sound, gas, flame and smoke or via an inspection of the impact device for sooty deposits after the upper cylinder has been removed. If none of these effects are noticed, an initiation failure (*no reaction*) is registered. Of the three possible types of reactions, both *decomposition* and *explosion* are considered positive test reactions (initiations) according to STANAG test procedures (NATO, 1999). In our experiments, in addition to audio-visual observations, a decomposition gas detector (MultiRAE model PGM6208) was used to classify the reactions.

The tests were performed at ambient temperatures (i.e. 23.6°C +1.4°C), according to the United Nations' (2019, p. 80) recommended test conditions. As the scope of the test method was within the range of -30°C to +80°C, no particular environmental modification was required.

As repeated drops from the same height in a fallhammer will not invariably yield the same result (reaction versus no reaction), the impact sensitivity of an energetic material must be estimated statistically. Hence, the weight is dropped repeatedly from a range of (log) heights $x_1, \ldots, x_n$, and for each x_i, we observe a binary outcome $y_i \in \{0,1\}$, where $y_i = 1$ if a reaction occurred and $y_i = 0$ otherwise. In accordance with STANAG 4489, the heights are determined according to the Bruceton up-and-down procedure (Dixon & Mood, 1948), meaning that an initial height x_1 is chosen for the first drop, and the consecutive heights are chosen inductively by

$$x_i = \begin{cases} x_{i-1} + d & \text{if } y_{i-1} = 0 \\ x_{i-1} - d & \text{if } y_{i-1} = 1, \end{cases} \tag{1}$$

for $i = 2, \ldots, n$, where $d > 0$ is the step size of the test, chosen by the operator. That is, we descend one step if a reaction is observed and ascend one step if not. In our experiments, the step size was set as $d = 0.05$, in accordance with STANAG 4489.

When assessing sensitivity, our primary interest is in quantiles such as h_{50}, which is the height from which there is a 50% probability of a reaction occurring. The median h_{50} is of particular interest, since it is known to correlate with quantum chemical properties of the energetic material (Jensen, Moxnes, Unneberg, & Christensen, 2020).

In addition to point estimates, we also wish to quantify the uncertainty of our results by means of confidence intervals (CIs). The use of large-sample theory to construct CIs for the Bruceton up-and-down method was verified by Christensen, Stoltenberg, Hjort (2023). Christensen et al. (2023) found via simulations that Fieller's theorem yields the most satisfactory CIs for the quantiles when the Bruceton up-and-down-method is employed. As recommended by Christensen et al. (2023), we used the existence of a bounded 95% CI for h_{50} via Fieller's theorem as a necessary criterion for terminating our fallhammer experiments. This resulted in most of the data sets comprising more than 30 drops. The fact that 30 drops were not sufficient could allude to inhomogeneity of the substances tested, or statistical model misspecification. Although it would be possible to simply employ the delta method to construct CIs, as suggested by Dixon & Mood (1948), simulation studies consistently show that CIs constructed via Fieller's theorem are more accurate for sensitivity data (see Christensen et al. (2023) and the references therein). In particular, the use of Fieller's theorem does not impact the qualitative conclusions reached in this paper, but rather increases the accuracy with which they are derived.

3. Results

Prior to all testing, a reference material of recently produced TNT ('Trinitrotoluene Type 1, Flake' with a 0.44% content of hexanitrostilbene (HNS), produced by Zaklady Chemiczne 'NITRO-CHEM' S.A. in Bydgoszcz, Poland, released for sale following the Certification of Compliance and Analysis on 8th September 2017, was tested. The test of the reference sample indicated an impact sensitivity (h_{50}) of 29.8 J, coinciding with the reported value (30 J) as described in STANAG 4489 (NATO, 1999). The full data from the impact sensitivity tests using the BAM Impact Apparatus are available at Novik & Christensen (2023). Here, we go over the main results.

For substance A_1, we initially aimed to obtain a single reaction with a 5 kg weight, but when this was not achieved, we proceeded to drop a 10 kg weight to increase the impact energy. After the first five drops, we still had no reactions, and we therefore decided to execute 10 drops from the maximum height of 100 cm with the 10 kg weight. Out of these, only a single drop caused a reaction. Thus, for this experiment, the maximum likelihood estimators (MLEs) do not exist, and we have highly limited information about the true underlying parameters governing the sensitivity of substance A_1. We can, however, assert with relatively high confidence that h_{50} is above 98.07 J, that is, 100 cm with a 10 kg weight.

For substance A_2, we did not obtain a bounded 95% CI for h_{50} after the first $n = 30$ drops, and we therefore increased the number of drops in increments by 10 at a time until a valid confidence interval was achieved. This happened after $n = 70$ drops. The resulting estimate for h_{50} is 10.99 J, or 22.41 cm with the 5 kg weight. The 95% and 99% confidence intervals for h_{50} is [8.26 J, 13.06 J] and [5.25 J, 14.45 J], respectively. We see that the value of h_{50} is significantly less than 30 J.

For substance B, we decided to stop the experiment after $n = 30$ drops, since this proved to be sufficient for obtaining a bounded 95% CI for h_{50}. From the data, the resulting estimate for h_{50} is 7.52 J, or 15.34 cm with a 5 kg weight. The 95% CI for h_{50} is [3.53 J, 9.25 J]. We did not obtain a 99% CI for h_{50}, since we only did $n = 30$ drops. Anyhow, we see that the value is significantly smaller than 30 J.

For substance C, as with substance A_2, we had not achieved a bounded 95% CI for h_{50} after the first 30 drops, and therefore decided to augment the dataset by increments of 10 drops until this was achieved. After $n = 50$ drops, we had a 95% CI for h_{50}. The resulting estimate of h_{50} is 31.60 J, or 64.42 cm with a 5 kg weight. The 95% and 99% CIs for h_{50} are [28.70 J, 35.90 J] and [27.01 J, 41.06 J], respectively. In particular, we do not have sufficient evidence to reject the hypothesis that $h_{50} = 30$ J.

For substance D, since we had not obtained a bounded CI for h_{50} after 30 drops, we increased the number of drops by increments of 10 until this was achieved, at n=70. The resulting estimate of h_{50} is 13.51 J. The accompanying 95% CI for h_{50} is [4.88 J, 19.14 J]. As with substance B, we did not obtain a bounded 99% CI for h_{50} for substance D. Anyhow, we still see that the value for h_{50} is significantly smaller than 30 J.

For substance E, since 30 drops were not sufficient for obtaining a bounded 95% CI for h_{50}, we increased the number of drops by increments of 10 until a valid CI was obtained, after $n = 70$ drops. The resulting estimate of h_{50} is 15.37 J, with the 95% CI [12.13 J, 19.06 J]. Again, this is significantly smaller than 30 J.

To summarise our results graphically, we draw the confidence curves for h_{50} for the substances A_2, B, C, D and E in Fig. 4. These were drawn using Fiellers theorem, as explained by Christensen et al. (2023). Using these curves, we may obtain all CIs for any confidence level. For example, if we were to calculate where the line $y = 0.95$ intersects these curves, we would recover the 95% CIs reported in the previous sections. As we can see, there is a substantial distance between the confidence curve for substance C and the other substances, whose confidence curves overlap more. This reflects how substance C exhibited an impact sensitivity in accordance with the existing literature on Amatol (30 J), whilst all the other substances were significantly more sensitive to impact. Note also how some of the curves are skewed, which reflects the asymmetric confidence intervals reported in the previous sections.
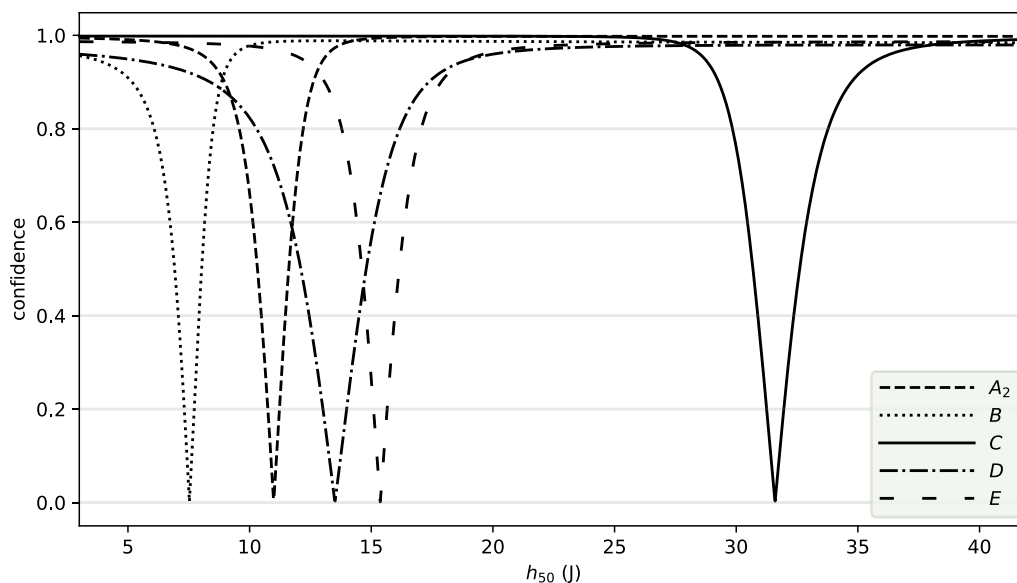
Fig. 4. Confidence curves for h_{50} for the substances A_2, B, C, D and E.

4. Summation and discussion

This study demonstrates that Amatols extracted from ERW, with the expected exception of the sample of high moisture content, are still sensitive to impact. For only one of the samples studied, the impact sensitivity coincided with what is recorded in the literature as expected values for Amatol. All the other samples studied were on the other hand significantly more sensitive to impact. In the most extreme case, namely substance B, the substance was more nearly than four times more sensitive than expected for Amatol (the estimate value of h_{50} being only 7.52 J, which is nearly less than a quarter of the expected value of 30 J). Note also that for this substance, we observed reactions with impacts as low as 6.18 J, as shown in Novik & Christensen (2023). The study therefore shows that the impact sensitivity of Amatol high explosives extracted from ageing ERW are susceptible of becoming increasingly sensitive to impact. Earlier studies have shown that explosive compositions containing ammonium nitrate might become sensitised if contaminated and/or exposed to certain environmental factors (e.g. moisture, heat). Albeit all the explosive objects included in this study, which displayed that a significant increase in impact sensitivity were located in cold climatic areas, and their moisture contents were found to be negligible; it cannot be disregarded that these explosives at one point of time have been subjected to heat and/or moisture, or that the explosives can have been contaminated by impurities. Further research into possible variations, resulting from chemical, technical or environmental differences, will be required, in order to gain further knowledge on ageing munitions containing Amatol.

Although not able to conclusively trace the origin of the increase in impact sensitivity, our study shows unambiguously that Amatols in ageing munitions can be much more sensitive to impact than previously assumed. This is imperative, as most risk assessments concerning ERW regularly seek to ascertain the societal risks related to ageing munitions. In this regard, particularly the risk of an unintentional explosion of the munitions are of great relevance, as the ordnance is prone to detonate given sufficient stimulus. Such an explosion could be the result of a number of causes. For example, it could occur as a result of an intended act of crime or terrorism, chemical or technical deterioration could cause a spontaneous detonation, or it could arise as a result of an accidental or

intentional disturbance (e.g. construction work, fishing, recreational activities). Sometimes negligence towards the risks posted by the ERW can result in unauthorised handling of the ordnance, with false reassurance that the explosives does not pose any significant risk (Geir P. Novik, 2022).

As munitions can remain intact and functional for decades, and even centuries, after the end of hostilities, ERW contamination is generally considered a major threat to societal safety and security. Simultaneously, toxic compounds, including nitroaromatic explosives are released into the environment by deteriorating munitions, representing an acute ecological and health hazard, resulting in serious environmental pollution problems in several countries and regions worldwide (Barreto-Rodrigues, Silva, & Paiva, 2009; Luo et al., 2023). Therefore, the clearance of ERW is a prioritised task in affected areas, and is recognised as a vital risk reduction tool (Geir P. Novik, 2023). However, all munitions subject to EOD clearance are, by nature, prone to be handled in one form or another (e.g. moving, relocating, rendering safe). Consequently, if the impact sensitivity of the explosives is in fact significantly higher than previously assumed, this would influence how ERW related risks are perceived, and form new boundaries for safe and practically feasible disposal techniques.

5. Conclusions

In this study we have analysed the composition high explosive substance Amatol, that were extracted from ageing explosive remnants of war. Our results clearly show that the samples studied were significantly more sensitive to impact than what one would expect based on the existing literature. A proper understanding of the hazard properties of explosive remnants of war is of vital importance, as there exist millions of tonnes of such remnants in nature in the form of unexploded ordnance and munitions disposed of at dumping sites on land, in lakes and at sea. The munitions are continuously deteriorating, resulting in hazardous materials being released into the environment, potentially posing an environmental, as well as a societal, risk. Moreover, as the explosives deteriorate over time, often resulting from inferior storage conditions or the presence of undesired factors such as moisture and certain metals, the munitions may become increasingly sensitive to external stimuli and susceptible to accidental detonation. In this study we analysed the composition high explosive substance Amatol, that were extracted from ageing explosive remnants of war.

All explosive ordnance subject to this analysis were originally deemed safe to move and to transport by the EOD operative in charge. This is a risk decision based on a number of factors, most predominantly the risks related to a detonation in situ and their correlating risk mitigating actions, and the risks associated with moving or transporting the object to a location that is more suited for conducting a controlled detonation. In this risk assessment it is imperative to evaluate the technical condition of the explosive object, including its sensitivity to impact. This is an essential part of assessing whether the object should or could be relocated. However, as this study has shown, these risk assessments were all conducted on the basis of what has proven to be erroneous information, as this study proved that Amatol can potentially have a significantly increased impact sensitivity compared to what is listed in most of the literature. As such, all risk assessments involving Amatols must account for the fact that handling these substances can involve a greater risk of accidental detonation as a result of increased impact sensitivity than what was the original premise.

Furthermore, in addition to Amatol being one of the high explosive compositions most extensively used up until the end of World War II, several seemingly identical explosive objects were produced with alternating fillings in which the same object could contain several explosives or explosive

compositions. Consequently, not only do we need to assume that the filling in explosive ordnance containing Amatol can have developed an increased impact sensitivity, but, as the exact filling of various ordnance cannot always be verified by external features alone, the risk of increased impact sensitivity must be included for all explosive ordnance potentially containing Amatols. We therefore recommend that EOD operators and other risk assessors now must account for the increase in impact sensitivity of Amatol in ageing ordnance and factor in this when encountering all munitions potentially containing Amatols. Moreover, as the required number of drops as described in the relevant standardised methodology of impact sensitivity testing (e.g. as NATO STANAG 4489) was not enough to produce a valid confidence interval in the majority of the experiments, these standards should therefore be revised to include a suitable method for constructing confidence intervals, such as Fieller's theorem. In particular, no fallhammer test should be terminated until a 95% confidence interval for h_{50} has been obtained.

Ethics

Our institution does not require ethics approval for collecting samples included in this study.

Data accessibility

Our data are deposited at Dryad: https://doi.org/10.5061/dryad.sqv9s4n91.

CRediT authorship contribution statement

Geir P. Novik performed conceptualisation, writing - original draft and formal analysis.
Dennis Christensen performed writing - original draft and formal analysis.

Declaration of competing interest

Geir P. Novik reports financial support was provided by Norwegian Defence Research Establishment. Dennis Christensen reports financial support was provided by Norwegian Defence Research Establishment.

Funding

Acknowledgements

We are also grateful for the valuable contributions made by the peer reviewers in shaping this paper into its final form. Their dedication to advancing the field have been instrumental in improving the quality and impact of this research.

References

Albright, R. (2012). *Cleanup of Chemical and Explosive Munitions: Location, Identification and Environmental Remediation* (2nd ed.). Massachusetts, United States: William Andrew.

ATSDR. (1995). *Toxicological Profile for 2,4,6-trinitrotoluene*. Retrieved from U.S. Department of Health and Human Services – Agency for Toxic Substances and Disease Registry https://wwwn.cdc.gov/TSP/ToxProfiles/ToxProfiles.aspx?id=677&tid=125

Barreto-Rodrigues, M., Silva, F. T., & Paiva, T. C. B. (2009). Characterization of wastewater from the Brazilian TNT industry. *Journal of Hazardous Materials, 164*(1). doi:https://doi.org/10.1016/j.jhazmat.2008.07.152

Christensen, D., Stoltenberg, E. A., & Hjort, N. L. (2023). *Sequential experimental designs in regression: Theory for the Bruceton and Langlie designs.* . arXiv preprint arXiv:2312.13387.

Christensen, D., Unneberg, E., Høyheim, E., Jensen, T. L., & Hjort, N. L. (2023). *Improved measurements of impact sesitivities of energetic materials.* Paper presented at the 25th International Seminar on New Trends in Research of Energetic Materials (NTREM).

Der Reichsminister der Luftfart. (1942). *Die Munition der Flakartillerie, Beschreibung, Teil 1, L.Dv.4402/1*. Berlin: Der Reichsminister der Luftfart

Dixon, W. J., & Mood, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association, 43*(241), 109-126.

Fedoroff, B. T., Aaronson, H. A., Clift, G. D., & Reese, E. F. (1958). *Dictionary of explosives, ammunition and weapons (German section), PATR 2510*. Dover, New Jersey: U.S. Army Research and Development Command.

Fedoroff, B. T., Aaronson, H. A., Reese, E. F., Sheffield, O. E., & Clift, G. D. (1960). *Encyclopedia of explosives and related items. PATR 2700. Volume 1* (U. S. A. R. a. D. Command Ed.). Picatinny Arsenal, New Jersey: U.S. Army Research and Development Command.

Ford, G., Ottemöller, L., & Bapite, B. (2005). *Analysis of Explosions in the BGS Seismic Database in the Area of Beaufort's Dyke, 1992-2004*. Retrieved from British Geological Survey https://webarchive.nationalarchives.gov.uk/20121203195642/http://www.mod.uk/NR/rdonlyres/712B6133-E353-4030-9DD0-F677DC3B6F38/0/bgs_beauforts.pdf

Gibbs, T. R., & Popolato, A. (Eds.). (1980). *LASL Explosive Property Data*. Los Angeles: University of Calofornia Press.

Gruhne, M. S., Lommel, M., Wurzenberger, M. H. H., Szimhardt, N., Klapötke, T. M., & Stierstorfer, J. (2019). OZM Ball Drop Impact Tester (BIT-132) vs. BAM Standard Method – a Comparative Investigation. *Propellants, Explosives, Pyrotechnics., 45*(1). doi:https://doi.org/10.1002/prep.201900286

Hamer, M. (2004, 21st August). The doomsday wreck. *New Scientist* Retrieved from https://www.newscientist.com/article/mg18324615-100-the-doomsday-wreck/

Hershkowitz, J., & Akst, I. (1975). *A new approach to improving the performance of non-ideal explosives containing ammonium nitrate*. Retrieved from Dover, New Jersey:

Jensen, T. L., Moxnes, J. F., Unneberg, E., & Christensen, D. (2020). Models for predicting impact sensitivity of energetic materials based on the trigger linkage hypothesis and Arrhenius kinetics. *Journal of Molecular Modeling, 26*(65). doi:https://doi.org/10.1007/s00894-019-4269-z

Johansen, S. R. (2005). *Ammonal fra Sola overlevert av kaptein Novik*. Analytical report. Dyno Nobel.

Koske, D., Goldenstein, N. I., & Kammann, U. (2019). Nitroaromatic compounds damage the DNA of zebrafish embryos (Danio rerio). *Aquatic Toxicology, 217*(105345). doi:https://doi.org/10.1016/j.aquatox.2019.105345

Koske, D., Goldenstein, N. I., Rosenberger, T., Machulik, U., Hanel, R., & Kammann, U. (2020). Dumped munitions: New insights into the metabolization of 2,4,6-trinitrotoluene in Baltic flatfish. *Marine Environmental Research, 160*(104992). doi:https://doi.org/10.1016/j.marenvres.2020.104992

Long, T. P. (2005). *Introduction to Sea Dumped Munitions and Hazardous Wrecks*. https://www.drpaulwcollins.com/wp-content/uploads/2019/04/Introduction-to-Sea-Dumped-Munitions-and-Hazardous-Wrecks-Terrance-P.-Long.pdf

Luo, J., Li, Y., Cao, H., Zhu, Y., Liu, X., Li, H., & Liao, X. (2023). Variations of microbiota in three types of typical military contaminated sites: Diversities, structures, influence factors, and co-occurrence patterns *Journal of Hazardous Materials, 443*. doi:https://doi.org/10.1016/j.jhazmat.2022.130290

Maser, E., & Strehse, J. S. (2021). Can seafood from marine sites of dumped World War relics be eaten? *Archives of Toxicology*. doi:https://doi.org/10.1007/s00204-021-03045-9

Meyer, R., Köhler, J., & Homburg, A. (2005). *Explosives*. Weinheim: Wiley-VCH Verlag GmbH.

Monogarov, K. A., Meerov, D. B., Fomenkov, I. V., & Pivkina, A. N. (2023). Energy transferred to energetic materials during impact test at reaction threshold: Look back to go forward. *FirePhysChem, 3*(3), 255-262. doi:https://doi.org/10.1016/j.fpc.2022.11.003

Montesi, L. J., & Menichelli, V. J. (1964). *Evaluation of ammonium nitrate, aluminium mixture (80/20)*. Retrieved from U.S. Naval Ordnance Laboratory, White Oak, Maryland:

NATO. (1999). *NATO STANAG 4489 Explosives, Impact Sensitivity Tests*. North Atlantic Treaty Organization (NATO)

Nordaas, M. T. K. (2019, 13th October). Ammunisjonen kan bli brukt til bomber i tilsiktede handlinger [The ammunition can be used for bombs in intentional acts]. *Nærnett [Norwegian]*. Retrieved from https://www.nernett.no/artikler/nyhende/sjodumpet-ammunisjonukjente-konsekvenser

Novik, G. P. (2022). Analysis of samples of high explosives extracted from explosive remnants of war. *Science of The Total Environment, 842*(156864). doi:https://doi.org/10.1016/j.scitotenv.2022.156864

Novik, G. P. (2023). When a Safety Measure Becomes a Risk Accelerant: Removing the Option to Blast-in-Place When Clearing Explosive Remnants of War. *The Journal of Conventional Weapons Destruction, 27*(1). Retrieved from https://commons.lib.jmu.edu/cisr-journal/vol27/iss1/5

Novik, G. P., Abrahamsen, E. B., & Sommer, M. (2023). Improving the decision-making basis by strengthening the risk assessments of unexploded ordnance and explosive remnants of war. *Safety Science, 160*(106065). doi:https://doi.org/10.1016/j.ssci.2023.106065

Novik, G. P., & Christensen, D. (2023). *Increased impact sensitivity in ageing high explosives; analysis of Amatol extracted from explosive remnants of war [Dataset]*.

Novik, G. P., Sommer, M., & Abrahamsen, E. B. (2022). A risk-increasing safety strategy? Evaluating the traditional risk mitigating strategy in dealing with dumped ammunition and explosive remnants of war. *Journal of Military and Strategic Studies, 22*(1). Retrieved from https://jmss.org/article/view/72969/56178

Ordnance Bomb Disposal Center [US]. *German artillery projectiles and fuzes*. Washington: Ordnance Bomb Disposal Center

Pfeiffer, F. (2012). Changes in properties of explosives due to prolonged seawater exposure. *Marine Technology Society Journal, 46*, 102-110. doi:http://dx.doi.org/10.4031/MTSJ.46.1.5

Picatinny Arsenal. (1943). *Study Properties of Tetrammino Cupric Nitrate*. Dover, N. J.: Technical Group, Chemical Department, Research Division, Picatinny Arsenal.

PubChem. (2024). Coumpound summary - Amatol. Retrieved from
https://pubchem.ncbi.nlm.nih.gov/compound/Amatol. from National Library of Medicine
https://pubchem.ncbi.nlm.nih.gov/compound/Amatol

Rădeanu, C., Rus, D. C., Jitea, I. C., Miron, C., & Vasilescu, G. (2020). *Assessing the impact sensitivity of explosives using the BHF-12A equipment.* Paper presented at the 9th International Symposium on Occupational Health and Safety (SESAM 2019).

Robertson, R. (1920). The Research Department, Woolwich. *Nature, 105*, 710-712.
doi:https://doi.org/10.1038/105710a0

Samseth, J. Ø. (2022). *Which bond cleavage is the explosive's sensitivity most sensitive to?* Paper presented at the 51st International Annual Conference of the Fraunhofer ICT.

Schuster, R., Strehse, J. S., Ahvo, A., Turja, R., Maser, E., Bickmeyer, U., . . . Brenner, M. (2021). Exposure to dissolved TNT causes multilevel biological effects in Baltic mussels (Mytilus spp.). *Marine Environmental Research, 167*(105264).
doi:https://doi.org/10.1016/j.marenvres.2021.105264

Suceska, M. (1995). *Test Methods for Explosives*. New York: Springer.

Sunahara, G. I., Kuperman, R. G., Lotufo, G. R., Hawari, J., Thiboutot, S., & Ampleman, G. (2009). Introduction. In G. I. Sunahara, G. R. Lotufo, R. G. Kuperman, & J. Hawari (Eds.), *Ecotoxicology of explosives* (pp. 1-3). Boca Raton, FL: CRC Press.

The War Office [UK]. (1925). *Text book of explosives used in the service*: The War Office.

The War Office [UK]. (1944). *Handbook of enemy ammunition, Pamphlet No. 11*. London: His Majesty's Stationery Office

U.S. Army Material Command. (1965). *Handbook of foreign explosives*. Washington, D.C.: US Army Foreign Science and Technology Center.

U.S. Department of the Army. (1984). *Military Explosives TM 9-1300-214*. Department of the Army

U.S. War Department. (1945). *War Department Technical Manual - Ammunition, General, TM 9-1900*. Washington: War Department

U.S. War Office. (1953a). *German explosive ordnance (bombs, fuzes, rockets, land mines, grenades and igniters) TM 9-1985-2*. Washington: United Stated Government Printing Office

U.S. War Office. (1953b). *German explosive ordnance (projectiles and projectile fuzes) TM 9-1985-3*. Washington: United States Government Printing Office

United Nations. (2019). *Manual of Tests and Criteria* (7th rev. ed.). New York: United Nations.

Urbanski, T. (1965). *Chemistry and technology of explosives, Volume II*. Oxford: Pergamon Press.

Urbanski, T. (1967). *Chemistry and technology of explosives, Volume III*. Oxford: Pergamon Press.

War Department [US]. (1944). *Ammunition Inspection Guide, TM 9-1904*: War Department.

Yinon, J. (1990). *Toxicity and Metabolism of Explosives*. Florida, United States: CRC Press.

Paper VIII

Unlevel playing field: Evidence of ethnic discrimination in the access to children's football from a field experiment in Norway

Appendices

Appendix A

Supplementary material of Paper I

Appendix A. Supplementary material of Paper I

Supplementary Material for "Inference for Bayesian nonparametric models with binary response data via permutation counting"

Dennis Christensen[*,†]

Appendix A: Proof of Proposition 2.2

In this section, we prove Proposition 2.2 from the main text, which we restate here as Proposition A.1.

Proposition A.1. *Let $x \in [0, \infty)^n$. Then $w(x; \mathcal{B}) > 0$ if and only if x is $\mathcal{B}$-admissible.*

Proof of Proposition A.1. Let $\sigma \in S_n$ be as in Definition 2.1 from the main text. Clearly, if x is $\mathcal{B}$-admissible, then by definition, there exists a permutation in S_n, namely σ, such that $\sigma(x) \in \mathcal{B}$.

We prove the converse by induction on n. The base case $n = 1$ is trivial. For the inductive step, assume the proposition holds for $\dim(\mathcal{B}) < n$ and suppose $w(x; \mathcal{B}) > 0$. Then there exists a permutation τ such that $\tau(x) \in \mathcal{B}$. Let j be the index such that $\sigma(1) = \tau(j)$. We claim that we may assume that $j = 1$ without loss of generality.

To prove the claim, it suffices to show that $x_{\tau(1)} \in \mathcal{B}_j$. Indeed, then we can redefine τ to map 1 to $\tau(j)$ and j to $\tau(1)$. To prove that $x_{\tau(1)} \in \mathcal{B}_j$, first assume that $\mathcal{B}_j$ extends to the left, so $\mathcal{B}_j = [0, t_j)$. In this case, we have $x_{\tau(1)} \in \mathcal{B}_1 \subseteq \mathcal{B}_j$, so $x_{\tau(1)} \in \mathcal{B}_j$. On the other hand, if $\mathcal{B}_j$ extends to the right, then $\mathcal{B}_j = [t_j, \infty)$ and $x_{\tau(1)} \in \mathcal{B}_j$ as $x_{\tau(j)} \leq x_{\tau(1)}$. In either case, we have that $x_{\tau(1)} \in \mathcal{B}_j$, so we may assume $j = 1$.

Now, let $x_{-1} = (x_2, \ldots, x_n)$ and $\mathcal{B}_{-1} = \mathcal{B}_2 \times \cdots \times \mathcal{B}_n$. As $\tau(x) \in \mathcal{B}$, it must also be the case that $\tau'(x_{-1}) \in \mathcal{B}_{-1}$, where τ' is the restriction of τ to the domain $\{2, \ldots, n\}$. Hence, $w(x_{-1}; \mathcal{B}_{-1}) > 0$, and so by induction hypothesis, x_{-1} is $\mathcal{B}_{-1}$-admissible. That is, $\sigma'(x_{-1}) \in \mathcal{B}_{-1}$, where σ' is the restriction of σ to the domain $\{2, \ldots, n\}$. Since we also know that $x_{\sigma(1)} = x_{\tau(1)} \in \mathcal{B}_1$, we see that x is $\mathcal{B}$-admissible. $\square$

Appendix B: Proof of Proposition 3.1

In this section, we prove Proposition 3.1 from the main text, which we restate here as Proposition B.1

Proposition B.1. *Let x be $\mathcal{B}$-admissible and let A be the matching matrix of x. Then, after permuting its columns if necessary, A is a complete block rectangular matrix.*

Proof of Proposition B.1. If A contained a row of zeros, then this would mean that for some $j \in \{1, \ldots, n\}$, we must have $x_i \notin \mathcal{B}_i$ for all $i = 1, \ldots, n$. But then x is not $\mathcal{B}$-admissible, a contradiction.

Without loss of generality, we can assume that $x_1 \leq \cdots \leq x_n$. This corresponds to permuting the columns of A, or equivalently, to hitting x with the permutation σ from Definition 2.1 from the main text. By the ordering of the $\mathcal{B}_i$, each column a_j of A takes the form

$$a_j = (\underbrace{0, \ldots, 0}_{k_j^{(1)}}, \underbrace{1, \ldots, 1}_{k_j^{(2)}}, \underbrace{0, \ldots, 0}_{k_j^{(3)}})^T,$$

where

$$k_j^{(1)} = \# \{i \mid x_j \geq t_i, i \leq n_0\}$$
$$k_j^{(2)} = \# \{i \mid x_j < t_i, i \leq n_0\} + \# \{i \mid x_j \geq t_i, i > n_0\}$$
$$k_j^{(3)} = \# \{i \mid x_j < t_i, i > n_0\}.$$

Furthermore, since the x_i are in ascending order, we see that if $j \leq j'$, then $k_j^{(1)} \leq k_{j'}^{(1)}$ and $k_j^{(3)} \geq k_{j'}^{(3)}$. This means that the matrix A comprises a sequence of descending rectangular blocks of ones, and can thus be described with parameters α, β, γ, m, as in Definition 3.2 from the main text. Clearly, α, β, γ, m satisfy condition (3.2) from the main text. It remains to show that they also satisfy condition (3.3) from the main text.

Assume for a contradiction that $\sum_{r=1}^t \beta_r > \sum_{r=1}^t \alpha_r$ for some $t \in \{1, \ldots, k\}$. Then $a_{ij} = 0$ for all $i \leq \sum_{s=1}^t \beta_r$ and $j > \sum_{s=1}^t \alpha_s$. That is, $x_i \notin \mathcal{B}_j$ for all $i \leq \sum_{s=1}^t \beta_r$ and $j > \sum_{s=1}^t \alpha_s$. But then x is not $\mathcal{B}$-admissible, since there exists no surjection from $\{1, \ldots, \sum_{r=1}^t \alpha_r\}$ to $\{1, \ldots, \sum_{r=1}^t \beta_r\}$. The corresponding condition for γ can be proved similarly. $\qquad\square$

Appendix C: Proof of Theorem 3.1

In this section, we prove Theorem 3.1 from the main text, which we restate here as Theorem C.1.

Theorem C.1. *Let A be an $m \times n$ complete block rectangular matrix. Then there exists an implementable algorithm for computing* $\mathrm{perm}(A)$, *whose computational complexity grows polynomially with n.*

The first step is to introduce the notion of *subpermanents*.

C.1 Subpermanents

To motivate the definition of subpermanents, we need a lemma which links permanents to counting permutations.

Lemma C.1. *Let A be an $m \times n$ $(0,1)$-matrix, where $m \leq n$. Then $\operatorname{perm}(A) = \#\wp(A)$, where*

$$\wp(A) = \left\{ \tau \in S_{n,m} \mid a_{i,\tau(i)} = 1 \text{ for } i = 1, \ldots, m \right\}. \tag{C.1}$$

Proof. This follows from Definition 3.1 from the main text and the fact that A is a $(0,1)$-matrix. $\qquad\square$

The notion of subpermanents now arises by sorting the number of permutations in $\wp(A)$ by how many entries from the *right-most* and *left-most* rectangular blocks are in each permutation.

Definition C.1. *Let A be an $m \times n$ block rectangular matrix minimally parametrised by α, β, γ, m, where $m \leq n$ and $k = \dim(\alpha) \geq 2$. For $0 \leq r \leq \alpha_1$ and $0 \leq s \leq \alpha_k$, define the set $\wp_{rs}(A)$ by*

$$\wp_{rs}(A) = \left\{ \tau \in \wp(A) \;\middle|\; \#\{i \mid \tau(i) \leq \alpha_1\} = r \text{ and } \#\{i \mid \tau(i) > \sum_{l=1}^{k-1} \alpha_l\} = s \right\} \tag{C.2}$$

We define $\operatorname{perm}_{rs}(A) = \#\wp_{rs}(A)$ to be the r,s-subpermanent of A.

Note that

$$\operatorname{perm}(A) = \sum_{r=0}^{\alpha_1} \sum_{s=0}^{\alpha_k} \operatorname{perm}_{rs}(A). \tag{C.3}$$

Hence, if we can calculate all subpermanents of A, we obtain the permanent as well. It turns out that the subpermanents of A can be explicitly related to those of another block rectangular matrix which has fewer rows than A. This removal of rows will happen in a systematic fashion via series of moves, called *trimming, splitting* and *merging*, to be defined momentarily. By applying these moves repeatedly to A, we end up with a block rectangular matrix with a very simple parametrisation, whose subpermanents can be calculated explicitly. Having done so, we may then reverse the moves we performed and update the subpermanents accordingly. This will allow us to obtain the subpermanents of A. We now proceed by defining trimming, splitting and merging, and establishing the subpermanent relations.

C.2 Trimming, splitting and merging

Throughout this section, let A be an $m \times n$ complete block rectangular matrix minimally parametrised by α, β, γ, m, where $m \leq n$ and $k = \dim(\alpha) \geq 2$.

Definition C.2. *The operations* trimming, splitting *and* merging *are defined as follows.*

- *If $m > \beta_1 > 0$, a* top trim *of A is the process of replacing A with the block rectangular matrix A^{tt} parametrised by $\alpha, \beta^{\mathrm{tt}}, \gamma, m^{\mathrm{tt}}$, where $\beta^{\mathrm{tt}} = (0, \beta_2, \ldots, \beta_{k-1})$ and $m^{\mathrm{tt}} = m - \beta_1$. This is equivalent to deleting the top β_1 rows of A.*

- If $m > \gamma_{k-1} > 0$, a bottom trim *of A is the process of replacing A with the block rectangular matrix A^{bt} parametrised by $\alpha, \beta, \gamma^{\mathrm{bt}}, m^{\mathrm{bt}}$, where $\gamma^{\mathrm{bt}} = (\gamma_1, \ldots, \gamma_{k-2}, 0)$ and $m^{\mathrm{bt}} = m - \gamma_{k-1}$. This is equivalent to deleting the bottom γ_{k-1} rows of A.*

- If $m > \beta_{k-1} > 0$, a top split *of A is the process of replacing A with the block rectangular matrix A^{ts} parametrised by $\alpha, \beta^{\mathrm{ts}}, \gamma, m^{\mathrm{ts}}$, where $\beta^{\mathrm{ts}} = (\beta_1, \ldots, \beta_{k-2}, 0)$ and $m^{\mathrm{ts}} = m - \beta_{k-1}$. This is equivalent to deleting rows $\sum_{r=1}^{k-2} \beta_r + 1, \ldots, \sum_{r=1}^{k-1} \beta_r$ from A.*

- If $m > \gamma_1 > 0$, a bottom split *of A is the process of replacing A with the block rectangular matrix A^{bs} parametrised by $\alpha, \beta, \gamma^{\mathrm{bs}}, m^{\mathrm{bs}}$, where $\gamma^{\mathrm{bs}} = (0, \gamma_2, \ldots, \gamma_{k-1})$ and $m^{\mathrm{bs}} = m - \gamma_1$. This is equivalent to deleting rows $m - \sum_{r=1}^{k-1} \gamma_r + 1, \ldots, m - \sum_{r=1}^{k-2} \gamma_r$ from A.*

- If $\beta_1 = \gamma_1 = 0$, *then a* left merge *of A is the process of replacing α, β, γ by*

$$\alpha^{\mathrm{lm}} = (\alpha_1 + \alpha_2, \alpha_3, \ldots, \alpha_k), \quad \beta^{\mathrm{lm}} = (\beta_2, \ldots, \beta_{k-1}), \quad \gamma^{\mathrm{lm}} = (\gamma_2, \ldots, \gamma_{k-1}).$$

  This is equivalent to merging the two left-most rectangular blocks when their vertical positions and heights match.

- If $\beta_{k-1} = \gamma_{k-1} = 0$, *then a* right merge *of A is the process of replacing α, β, γ by*

$$\alpha^{\mathrm{rm}} = (\alpha_1, \ldots, \alpha_{k-1}, \alpha_{k-1} + \alpha_k), \quad \beta^{\mathrm{rm}} = (\beta_1, \ldots, \beta_{k-2}), \quad \gamma^{\mathrm{rm}} = (\gamma_1, \ldots, \gamma_{k-2}).$$

  This is equivalent to merging the two right-most rectangular blocks when their vertical positions and heights match. Note that merging does not actually change the matrix A, only its parametrisation.

Combining trimming, splitting and merging in the following way, we obtain the *reduction algorithm.*

Lemma C.2. *After applying the reduction algorithm, A is still a complete block rectangular matrix with a minimal parametrisation.*

Proof. It is easy to check that if conditions (3.2) and (3.3) from the main text were satisfied to begin with, then they are still satisfied after applying the algorithm. Hence, A is still block rectangular. Also, since trimming and splitting only remove rows, and merging only changes the parametrisation of A, it is not possible that either move will introduce a new row of zeros. Hence, A is still complete.

Assume for a contradiction that A is no longer minimally parametrised. Then for some $1 \le r \le k - 1$, we have $\beta_r = \gamma_r = 0$. As A was minimally parametrised before applying the algorithm, we cannot have $1 < r < k - 1$. Hence, $r = 1$ or $r = k - 1$. If $r = k - 1$, then the right merge step in line 17 was not performed, a contradiction. Similarly, if $r = 1$, then the left merge step in line 8 was not performed, since none of the steps from line 10 to line 18 change the values of β_1, γ_1. Hence there exists no such r, and so the parametrisation of A is still minimal. $\qquad\square$

Algorithm 1 Reduction

Require: Matrix A
 1: **if** $m > \beta_1 > 0$ **then**
 2:     $A \leftarrow A^{\text{tt}}$
 3: **end if**
 4: **if** $m > \gamma_1 > 0$ **then**
 5:     $A \leftarrow A^{\text{bs}}$
 6: **end if**
 7: **if** $\beta_1 = \gamma_1 = 0$ **then**
 8:     $A \leftarrow A^{\text{lm}}$
 9: **end if**
10: **if** $m > \gamma_{k-1} > 0$ **then**
11:     $A \leftarrow A^{\text{bt}}$
12: **end if**
13: **if** $m > \beta_{k-1} > 0$ **then**
14:     $A \leftarrow A^{\text{ts}}$
15: **end if**
16: **if** $\beta_{k-1} = \gamma_{k-1} = 0$ **then**
17:     $A \leftarrow A^{\text{rm}}$
18: **end if**

Note that in each `if` statement of the reduction algorithm, the value of $k + m$ is strictly decreased, so that if we repeatedly apply the algorithm, there will come a time where no move is possible. We say that the matrix A is *reduced* if this is the case.

Lemma C.3. *The matrix A is reduced if and only if exactly one of the following holds.*

1. $k = 1$,

2. $k = 2, \beta_1 = 0$ *and* $\gamma_1 = m$,

3. $k = 2, \beta_1 = m$ *and* $\gamma_1 = 0$,

4. $k = 3, \beta = (0, m)$ *and* $\gamma = (m, 0)$.

Proof. It is easy to check that if either condition holds, then no further reduction of A is possible. For the converse, we consider case by case. If $k = 1$, then we are done. Suppose that $k = 2$. If $0 < \beta_1 < m$ or $0 < \gamma_1 < m$, then it would be possible to perform a trim or a split. Also, if $\beta_1 = \gamma_1 = 0$, then it would be possible to perform a left (or right) merge. Hence one of case 2 or 3 must hold.

Finally, assume that $k \geq 3$. By the same argument as above, one of the following must be true.

- $\beta = (0, \ldots, 0, m)$ and $\gamma = (m, 0, \ldots, 0)$,

- $\beta = (m, 0, \ldots, 0)$ and $\gamma = (0, \ldots, 0, m)$.

Assume that the former is true. By Lemma C.2, we know that the parametrisation is minimal. This forces $k = 3$ and so case 4 holds. Finally, if the latter is true, then again minimality of the parametrisation forces $\beta = (m, 0)$ and $\gamma = (0, m)$. However, this parametrisation will yield a matrix with a row of zeros, which contradicts Lemma C.2. □

In Figure 1, we see how the matrix from Figure 3a from the main text is affected by the reduction algorithm. Note that the final parametrisation belongs to case 3 in Lemma C.3.

We will now derive formulae for the subpermanents of a reduced matrix A. In Definition C.1, we only defined subpermanents for matrices with $k = \dim(\alpha) \geq 2$. In condition 1 in Lemma C.3, A is an $m \times n$ matrix of ones, parametrised with $k = 1$. We therefore modify the reduction algorithm slightly by declaring that if $k = 2$ and $\beta_1 = \gamma_1 = 0$, then we ignore the left and right merge moves in lines 7 and 16. This allows us to change condition 1 in Lemma C.3 to the following.

1′. $k = 2$ and $\beta_1 = \gamma_1 = 0$.

Note that this parametrisation is no longer minimal.

Notation. *For whole numbers i, j, if $j \leq i$, we let*

$$\binom{i}{j} = \frac{i!}{j!(i-j)!}, \quad \begin{Bmatrix} i \\ j \end{Bmatrix} = \frac{i!}{(i-j)!},$$

and declare that $\binom{i}{j} = \begin{Bmatrix} i \\ j \end{Bmatrix} = 0$ if $j > i$.

Lemma C.4. *The subpermanents of a reduced matrix A take the following form.*

1′. *If $k = 2$ and $\beta_1 = \gamma_1 = 0$, then*

$$\mathrm{perm}_{rs}(A) = \begin{cases} \binom{m}{r} \begin{Bmatrix} \alpha_1 \\ r \end{Bmatrix} \begin{Bmatrix} \alpha_2 \\ s \end{Bmatrix} & \text{if } r + s = m, \\ 0 & \text{otherwise.} \end{cases}$$

2. *If $k = 2, \beta_1 = 0$ and $\gamma_1 = m$, then*

$$\mathrm{perm}_{rs}(A) = \begin{cases} \begin{Bmatrix} \alpha_2 \\ m \end{Bmatrix} & \text{if } r = 0, s = m, \\ 0 & \text{otherwise.} \end{cases}$$

3. *If $k = 2, \beta_1 = m$ and $\gamma_1 = 0$, then*

$$\mathrm{perm}_{rs}(A) = \begin{cases} \begin{Bmatrix} \alpha_1 \\ m \end{Bmatrix} & \text{if } r = m, s = 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$\left(\begin{array}{cccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right) \xrightarrow[\text{split}]{\text{bottom}} \left(\begin{array}{cccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right)$$

$$\xrightarrow[\text{merge}]{\text{left}} \left(\begin{array}{cccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right) \xrightarrow[\text{split}]{\text{top}} \left(\begin{array}{cccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right)$$

$$\xrightarrow[\text{merge}]{\text{right}} \left(\begin{array}{cccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right) \xrightarrow[\text{trim}]{\text{top}} \left(\begin{array}{cccc|ccc} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right)$$

$$\xrightarrow[\text{split}]{\text{bottom}} \left(\begin{array}{ccccc|cc} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right) \xrightarrow[\text{merge}]{\text{left}} \left(\begin{array}{ccccc|cc} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right)$$

$$\xrightarrow[\text{trim}]{\text{bottom}} \left(\begin{array}{ccccc|cc} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{array}\right)$$

Figure 1: Reducing the matrix in Figure 3a.

4. *If $k = 3$, $\beta = (0, m)$ and $\gamma = (m, 0)$, then*

$$\operatorname{perm}_{rs}(A) = \begin{cases} \left\{ {\alpha_2 \atop m} \right\} & \text{if } r = s = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. We prove the first case for illustration. First note that $\wp_{rs}(A) = \emptyset$ when $r + s \neq m$. When $r + s = m$, every permutation $\tau \in \wp_{rs}(A)$ can be constructed as follows.

(i) Choose a subset I of size r from $\{1, \ldots, m\}$.

(ii) Assign values $\tau(i)$ for all $i \in I$ such that $\tau(i) \leq \alpha_1$.

(iii) Assign values $\tau(i)$ for all $i \notin I$ such that $\tau(i) > \alpha_1$.

There are $\binom{m}{r}$ ways to perform step (i), $\left\{ {\alpha_1 \atop r} \right\}$ ways to perform step (ii), and $\left\{ {n - \alpha_1 \atop s} \right\} = \left\{ {\alpha_2 \atop s} \right\}$ ways to perform step (iii). $\qquad \square$

We now show how the subpermanents of A are affected by trimming, splitting and merging.

Lemma C.5. *The subpermanents of A are related to those of A^{tt}, A^{bt}, A^{ts}, A^{bs}, A^{lm} and A^{rm} via the following relationships.*

$$
\begin{aligned}
\operatorname{perm}_{rs}(A) &= \left\{ {\alpha_k + \gamma_{k-1} - s \atop \gamma_{k-1}} \right\} \operatorname{perm}_{r, s - \gamma_{k-1}}(A^{\mathrm{bt}}) \\
&= \left\{ {\alpha_1 + \beta_1 - r \atop \beta_1} \right\} \operatorname{perm}_{r - \beta_1, s}(A^{\mathrm{tt}}) \\
&= \sum_{l=0}^{\alpha_k} \left\{ {\sum_{q=2}^{k-1} \alpha_q - (m - \gamma_1 - r - l) \atop \gamma_1 + l - s} \right\} \left\{ {\alpha_k - l \atop s - l} \right\} \binom{\gamma_1}{s - l} \operatorname{perm}_{rl}(A^{\mathrm{bs}}) \\
&= \sum_{l=0}^{\alpha_1} \left\{ {\sum_{q=2}^{k-1} \alpha_q - (m - \beta_{k-1} - l - s) \atop \beta_{k-1} + l - r} \right\} \left\{ {\alpha_1 - l \atop r - l} \right\} \binom{\beta_{k-1}}{r - l} \operatorname{perm}_{ls}(A^{\mathrm{ts}}) \\
&= \sum_{l=r}^{\alpha_1 + \alpha_2} \frac{\left\{ {\alpha_2 \atop l - r} \right\} \left\{ {\alpha_1 \atop r} \right\}}{\left\{ {\alpha_1 + \alpha_2 \atop l} \right\}} \binom{l}{r} \operatorname{perm}_{ls}(A^{\mathrm{lm}}) \\
&= \sum_{l=s}^{\alpha_{k-1} + \alpha_k} \frac{\left\{ {\alpha_{k-1} \atop l - s} \right\} \left\{ {\alpha_k \atop s} \right\}}{\left\{ {\alpha_{k-1} + \alpha_k \atop l} \right\}} \binom{l}{s} \operatorname{perm}_{rl}(A^{\mathrm{rm}}).
\end{aligned}
$$

Proof. For the bottom trim, observe that every permutation $\tau \in \wp_{r, s - \gamma_{k-1}}(A^{\mathrm{bt}})$ may be extended to a permutation $\bar{\tau} \in \wp_{rs}(A)$ by defining $\bar{\tau}(i) = \tau(i)$ for $1 \leq i \leq \sum_{l=1}^{k-2} \gamma_l$, and choosing values $\bar{\tau}(\sum_{l=1}^{k-2} \gamma_l + 1), \ldots, \bar{\tau}(\sum_{l=1}^{k-1} \gamma_l)$ distinctly amongst the indices $\{\sum_{l=1}^{k-1} \alpha_l + 1, \ldots, n\}$. There are $\left\{ {\alpha_k + \gamma_{k-1} - s \atop \gamma_{k-1}} \right\}$ ways to to this, and every permutation in $\wp_{rs}(A)$ can be recovered this way, thus yielding the result. The corresponding relation for the top trim is proven similarly.

Next, we prove the result for the top split. Given $1 \leq l \leq \alpha_1$, we may extend every permutation $\tau \in \wp_{ls}(A^{\mathrm{ts}})$ to a permutation $\bar{\tau} \in \wp_{rs}(A)$ as follows. First, we define $\bar{\tau}(i) = \tau(i)$ for $1 \leq i \leq \sum_{q=1}^{k-2} \beta_q$ and $\bar{\tau}(i) = \tau(i - \beta_{k-1})$ for $\sum_{q=1}^{k-1} \beta_q < i \leq m$. To define $\bar{\tau}(i)$ for $\sum_{q=1}^{k-2} \beta_q < i \leq \sum_{q=1}^{k-1} \beta_q$, we do the following.

(i) Choose a subset $I \subseteq \{\sum_{q=1}^{k-2} \beta_q + 1, \ldots, \sum_{q=1}^{k-1} \beta_q\}$ of size $r - l$.

(ii) For all $i \in I$, assign values $\bar{\tau}(i)$ such that $\bar{\tau}(i) \leq \alpha_1$. Note that l such values out of the possible α_1 are already taken.

(iii) For all $i \notin I$, assign values $\bar{\tau}(i)$ such that $\alpha_1 < \bar{\tau}(i) \leq \sum_{q=1}^{k-1} \alpha_q$. Note that $m - \beta_{k-1} - l - s$ such values out of the possible $\sum_{q=2}^{k-1} \alpha_q$ are already taken.

There are $\binom{\beta_{k-1}}{r-l}$ ways to perform step (i), $\left\{ \substack{\alpha_1 - l \\ r - l} \right\}$ ways to perform step (ii) and $\left\{ \substack{\sum_{q=2}^{k-1} \alpha_q - (m - \beta_{k-1} - l - s) \\ \beta_{k-1} + l - r} \right\}$ ways to perform step (iii). Any permutation in $\wp_{rs}(A)$ can be recovered this way. Thus, summing over $l = 0, \ldots, \alpha_1$, we get the result. The corresponding relation for the bottom split is proven similarly.

For the left merge, fix l and let $\tau \in \wp_{ls}(A^{\mathrm{lm}})$. Let $I = \tau^{-1}(\{1, \ldots, \alpha_1 + \alpha_2\})$ and let τ' be the restriction of τ to $\{1, \ldots, m\} \setminus I$. By assumption, $\#I = l$. Note that there are exactly $\mathrm{perm}_{ls}(A) / \left\{ \substack{\alpha_1 + \alpha_2 \\ l} \right\}$ permutations in $\wp_{ls}(A^{\mathrm{lm}})$ which yield the same τ'. Given τ', we may then extend it to a permutation $\bar{\tau} \in \wp_{rs}(A)$ as follows.

(i) Choose a subset $J \subseteq I$ of size r.

(ii) Assign values $\bar{\tau}(i)$ such that $\bar{\tau}(i) \leq \alpha_1$ for all $i \in J$.

(iii) Assign values $\bar{\tau}(i)$ such that $\alpha_1 < \bar{\tau}(i) \leq \alpha_1 + \alpha_2$ for all $i \in I \setminus J$.

There are $\binom{l}{r}$ ways to perform step (i), $\left\{ \substack{\alpha_1 \\ r} \right\}$ ways to perform step (ii), and $\left\{ \substack{\alpha_2 \\ l - r} \right\}$ ways to perform step (iii). For all permutations τ in $\wp_{rs}(A)$, there exists an l such that τ can be recovered this way. Hence, summing over $l = r, \ldots, \alpha_1 + \alpha_2$, we get the result. The corresponding relation for the right merge is proven similarly. $\square$

Proof of Theorem C.1. Consider the following procedure. First, reduce A by repeatedly applying the reduction algorithm until termination, and keep track of all the parameters involved in memory. Next, calculate the subpermanents of the resulting reduced matrix using Lemma C.4. Then for every move performed in reverse order, update the subpermanents using the relations in Lemma C.5. Once completed, we end up with all subpermanents of the original matrix A. The permanent of A is then the sum of the subpermanents, as shown in (C.3).

Now, reducing A requires at most $k + m \leq 2n$ applications of the reduction algorithm. Further, each calculation involved in updating the subpermanents can be computed in polynomial time. Thus, the complexity of the procedure above grows polynomially with n, as required. $\square$

In practice, we compute all subpermanents on a log scale, and use the LogSumExp function to compute the relations in Lemma C.5.

Appendix B

Supplementary material of Paper III

Appendix B. Supplementary material of Paper III

Supplementary information

Models for predicting impact sensitivity of energetic materials based on the trigger linkage hypothesis and Arrhenius kinetics
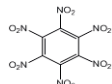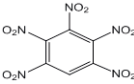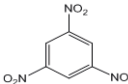
Journal of Molecular Modeling
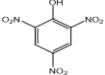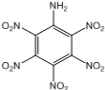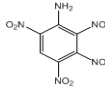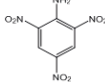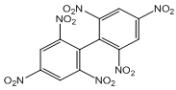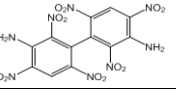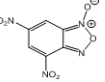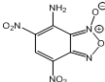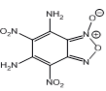
Tomas L. Jensen, John F. Moxnes, Dennis Christensen, Erik Unneberg
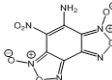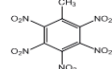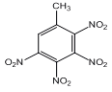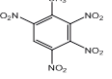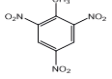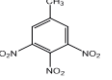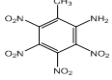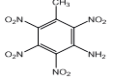
Defence Systems Division, Norwegian Defence Research Establishment, P.O. Box 25, N-2027 Kjeller, Norway
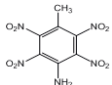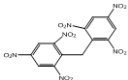
Corresponding author: john-f.moxnes@ffi.no

Table 1 The nitroaromatic compounds in the Wilson et al. data set [1]. The sum formula, enthalpy of formation (ΔH_f), density, bond dissociation energy (BDE), heat of detonation (Q), temperature of detonation (T_{ex}), impact energy (I_{50}) and molecular structure
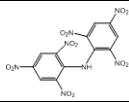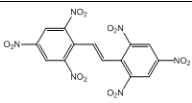
Compound	Formula	ΔH_f(s)[a] (kJ/mol)	Density[d] (g/cm^3)	BDE (kJ/mol)	Q (kJ/kg)	T_{ex} (K)	I_{50}[d] (J)	Molecular structure
Hexanitrobenzene (HNB)	C6N6O12	125.9	1.99	225.5	6816	5456	2.7	
1,2,3,5,6-Pentanitrobenzene (PNB)	C6HN5O10	49.4	1.91	223.1	6578	5163	2.7	
1,2,3,5-Tetranitrobenzene (TetNB)	C6H2N4O8	-43.5	1.82	226.1	5904	4528	6.9	
1,3,5-Trinitrobenzene (TNB)	C6H3N3O6	-47.7	1.68	284.6	5401	4026	17.4	
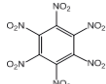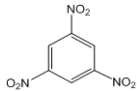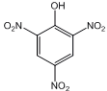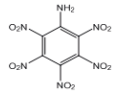
Name	Formula							Structure
2,4,6-Trinitrophenol (Picric Acid)	C6H3N3O7	-231.2	1.77	267.6	5029	3837	15.7	
2,3,4,5,6-Pentanitroaniline (PNA)	C6H2N6O10	-388.3	1.86	213.0	5027	4123	5.4	
2,3,4,6-Tetranitroaniline (TetNA)	C6H3N5O8	-49.0	1.87	216.6	5656	4269	11.5	
2,4,6-Trinitroaniline (TNA)	C6H4N4O6	-74.5	1.76	296.8	5077	3725	34.5	
2,2',4,4',6,6'Hexanitrobiphenyl (HNBP)	C12H4N6O12	68.2	1.69	267.4	5657	4300	17.2	
3,3'-Diamino-2,2',4,4',6,6'-hexanitrobiphenyl (DIPAM)	C12H6N8O12	-28.5	1.79	276.1	5250	3924	16.4	
4,6-Dinitrobenzofuroxan (DNBF)	C6H2N4O6	190.0	1.76	277.5	6012	4506	18.6	
7-Amino-4,6-dinitrobenzofuroxan (ADNBF)	C6H3N5O6	153.9[b]	1.88	296.5	5640	4117	24.5	
5,7-Diamino-4,6-dinitrobenzofuroxan (CL-14)	C6H4N6O6	86.0[c]	1.95	317.8	5163	3745	29.4	

238
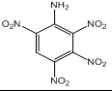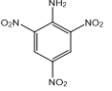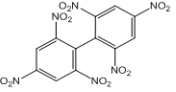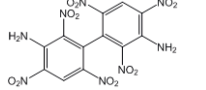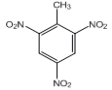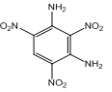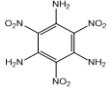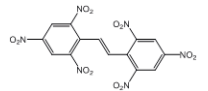
Name	Formula							Structure
8-Amino-7-nitrobenzobisfuroxan (CL-18)	C6H2N6O6	428.9[c]	1.93	313.1	6337	4641	13.7	
2,3,4,5,6-Pentanitrotoluene (PNT)	C7H3N5O10	-63.7[c]	1.76	226.8	6002	4608	4.4	
2,3,4,5-Tetranitrotoluene (2,3,4,5TetNT)	C7H4N4O8	-69.8[c]	1.71	230.1	5616	4164	3.7	
2,3,4,6-Tetranitrotoluene (2,3,4,6-TetNT)	C7H4N4O8	-69.8[c]	1.71	231.9	5616	4164	4.7	
2,4,6-Trinitrotoluene (2,4,6-TNT)	C7H5N3O6	-67.1	1.67	263.5	5097	3592	24.0	
2,3,4-Trinitrotoluene (2,3,4-TNT)	C7H5N3O6	15.1	1.63	231.4	5514	3827	13.7	
3,4,5-Trinitrotoluene (3,4,5-TNT)	C7H5N3O6	-5.0	1.63	263.9	5432	3786	26.2	
2-Amino-3,4,5,6-tetranitrotoluene (TetN-o-Tol)	C7H5N5O8	-234.7[c]	1.72	219.6	4897	3699	8.8	
3-Amino-2,4,5,6-tetranitrotoluene (TetN-m-Tol)	C7H5N5O8	-234.7[c]	1.73	216.0	4899	3692	9.1	
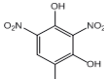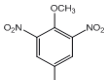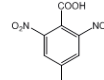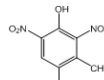
Compound	Formula	ΔH$_f$[a] (kJ/mol)	Density[c] (g/cm^3)	BDE (kJ/mol)	Q (kJ/kg)	T$_{ex}$ (K)	I$_{50}$[d] (J)	Molecular structure
4-Amino-2,3,5,6-tetranitrotoluene (TetN-p-Tol)	C7H5N5O8	-234.7[c]	1.72	228.6	4897	3699	11.5	
2,2',4,4',6,6'- Hexanitrodiphenylmethane (HNDPM)	C13H6N6O12	43.5[c]	1.71	254.6	5537	4055	9.6	

[a]Ref.[2] [b]Ref. [3] [c]Ref. [4] [d]Ref. [1]

Table 2 The nitroaromatic compounds in the Storm et al. data set [5]. The sum formula, enthalpy of formation (ΔH$_f$), density, bond dissociation energy (BDE), heat of detonation (Q), temperature of detonation (T$_{ex}$), impact energy (I$_{50}$) and molecular structure

Compound	Formula	ΔH$_f$[a] (kJ/mol)	Density[c] (g/cm^3)	BDE (kJ/mol)	Q (kJ/kg)	T$_{ex}$ (K)	I$_{50}$[d] (J)	Molecular structure
Hexanitrobenzene (HNB)	C6N6O12	125.9	1.99	225.5	6816	5456	2.9	
1,3,5-Trinitrobenzene (TNB)	C6H3N3O6	-47.7	1.68	284.6	5401	4026	25	
2,4,6-Trinitrophenol (Picric Acid)	C6H3N3O7	-231.2	1.77	267.6	5029	3837	21	
2,3,4,5,6-Pentanitroaniline (PNA)	C6H2N6O10	-388.3	1.86	213.0	5027	4123	3.7	
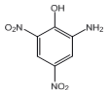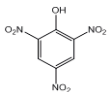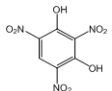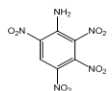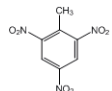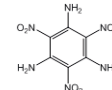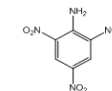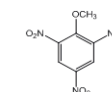
Compound	Formula							Structure
2,3,4,6-Tetranitroaniline (TetNA)	C6H3N5O8	-49.0	1.87	216.6	5656	4269	10	
2,4,6-Trinitroaniline (TNA)	C6H4N4O6	-74.5	1.76	296.8	5077	3725	43	
2,2',4,4',6,6'Hexanitrobiphenyl (HNBP)	C12H4N6O12	68.2	1.69	267.4	5657	4300	21	
3,3'-Diamino-2,2',4,4',6,6'-hexanitrobiphenyl (DIPAM)	C12H6N8O12	-28.5	1.79	276.1	5250	3924	32	
2,4,6-Trinitrotoluene (2,4,6-TNT)	C7H5N3O6	-67.1	1.67	263.5	5097	3592	39	
1,3-Diamino-2,4,6-trinitrobenzene (DATB)	C6H5N5O6	-98.7	1.84	310.2	4805	3486	78	
1,3,5-Triamino-2,4,6-trinitrobenzene (TATB)	C6H6N6O6	-139.7	1.94	310.2	4440	3214	>78	
2,4,6,2',4',6'-Hexnitrodiphenylamine (HNDP)	C12H5N7O12	41.4[b]	1.64[b]	245.9	5451	4158	12	
Hexanitrostilbene (HNS)	C14H6N6O12	78.2[b]	1.74[b]	243.9	5474	3981	9.6	
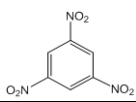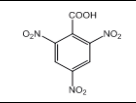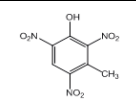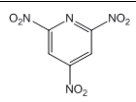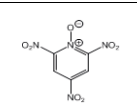
Compound	Formula	ΔH_f (kJ/mol)	Density (g/cm^3)	Q (kJ/kg)	BDE (kJ/mol)	T_{ex} (K)	I_{50} (J)	Molecular structure
1,3-Dihydroxy-2,4,6-trinitrobenzene (Styphnic acid)	C6H3N3O8	-523.0[b]	1.83[b]	274.3	4430	3516	11	
2,4,6-Trinitroanisole	C7H5N3O7	-153.2[b]	1.61[b]	244.9	5254	3780	47	
2,4,6-Trinitrobenzoic acid	C7H3N3O8	-402.9[b]	1.75[a]	269.3	4643	3639	27	
2,4,6-Trinitrocresol	C7H5N3O7	-252.3[b]	1.68[b]	255	4891	3568	47	

[a]Ref. [2] [b]Ref.[6] [c]Ref. [1] [d]Ref. [5]

Table 3 The nitroaromatic compounds in the Meyer et al. data set [6]. The sum formula, enthalpy of formation (ΔH_f), density, bond dissociation energy (BDE), heat of detonation (Q), temperature of detonation (T_{ex}), impact energy (I_{50}) and molecular structure
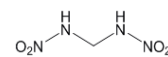
Compound	Formula	ΔH_f[a] (kJ/mol)	BDE (kJ/mol)	Density[a] (g/cm^3)	Q (kJ/kg)	T_{ex} (K)	I_{50}[a] (J)	Molecular structure
2,4,6,2',4',6'-Hexnitrodiphenylamine (HNDP)	C12H5N7O12	41.4	245.9	1.64	5451	4158	7.5	
Hexanitrostilbene (HNS)	C14H6N6O12	78.2	243.9	1.74	5474	3981	5	
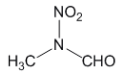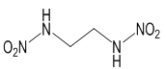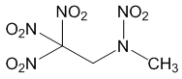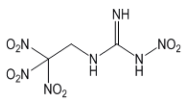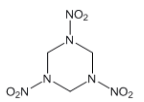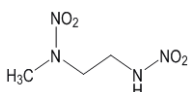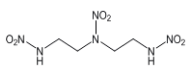
Name	Formula							Structure
1,3-Dinitrobenzene	C6H4N2O4	-27.2	295.0	1.50	4807	3388	39	
2-Amino-4,6-dinitrophenol (Picramic acid)	C6H5N3O5	-248.5	306.4	1.76[b]	4189	3026	34	
2,4,6,-Trinitrophenol (Picric acid)	C6H3N3O7	-241.6	267.6	1.77	5029	3837	7.4	
1,3-Dihydroxy-2,4,6-trinitrobenzene (Styphnic acid)	C6H3N3O8	-523.0	274.3	1.83	4430	3516	7.4	
2,3,4,6-Tetranitroaniline	C6H3N5O8	-49.0	216.6	1.87	5656	4269	6	
2,4,6-Trinitrotoluene (TNT)	C7H5N3O6	-67.1	263.5	1.65	5097	3592	15	
1,3,5-Triamino-2,4,6-trinitrobenzene (TATB)	C6H6N6O6	-139.7	310.2	1.93	4440	3214	50	
2,4,6-Trinitroaniline (TNA)	C6H4N4O6	-84.0	296.8	1.76	4978	3673	15	
2,4,6-Trinitroanisole	C7H5N3O7	-153.2	244.9	1.61	5254	3780	20	
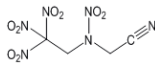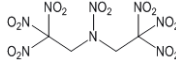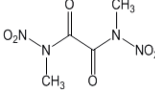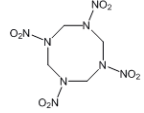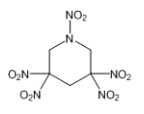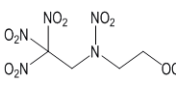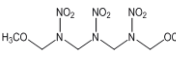
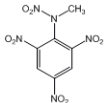| Compound	Chemical Formula	ΔHf (kJ/mol)	BDE (kJ/mol)	Density (g/cm³)	Q (kJ/kg)	Tex (K)	I50 (J)	Molecular structure
1,3,5-Trinitrobenzene	C6H3N3O6	-43.5	284.6	1.76	5630	3780	7.4	
2,4,6-Trinitrobenzoic acid	C7H3N3O8	-402.9	269.3	1.75[c]	4643	3639	10	
2,4,6-Trinitrocresol	C7H5N3O7	-252.3	254.9	1.68	4891	3568	12	
2,4,6-Trinitropyridine	C5H2N4O6	78.8	255.0	1.77	5879	4467	5.5	
2,4,6-Trinitropyridine N-oxide	C5H2N4O7	98.7	260.4	1.86	6407	4786	2.3	
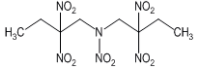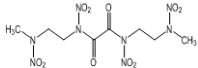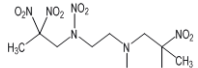
[a]Ref. [6] [b]Ref. [7] [c]Ref. [2]

Table 4 The nitramine data set [5]. The sum formula, enthalpy of formation (ΔH$_f$), density, bond dissociation energy (BDE), heat of detonation (Q), temperature of detonation (T$_{ex}$), impact energy (I$_{50}$) and molecular structure

Compound	Chemical Formula	ΔH$_f$[a] (kJ/mol)	Density (g/cm³)	BDE (N-NO2) (kJ/mol)	BDE (C-NO2) (kJ/mol)	Q (kJ/kg)	T$_{ex}$ (K)	I$_{50}$[g] (J)	Molecular structure
N,N'-Dinitro-methanediamine	C1H4N4O4	113.3	1.74[c]	220.3	-	7029	4779	3.2	
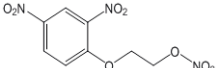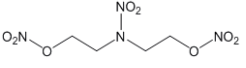
Name	Formula								Structure
N-Nitro-N-methyl-formamide	C2H4N2O3	-80.3	1.52[d]	206.7	-	5653	3778	78.4	
N,N'-Dinitro-1,2-ethanediamine	C2H6N4O4	-103.8[b]	1.71[b]	214.4	-	5432	3600	8.3	
Methyl-2,2,2-trinitro-ethylnitramine	C3H5N5O8	-319.8	1.80[e]	165.1	143.3	6497	4648	2.2	
Trinitroethylnitro-guanidine	C3H5N7O8	-30.0	1.77[f]	156.5	152.6	6113	4535	3.7	
Cyclo-1,3,5-trimethylene-2,4,6-trinitramine (RDX)	C3H6N6O6	66.9[b]	1.82[b]	171.9	-	6141	4224	6.4	
N-Methyl-N,N'-dinitro-1,2-ethanediamine	C3H8N4O4	-85.0	1.53[d]	192.5	-	5385	3435	27.9	
N,N'-Dinitro-N-[2-(nitramino)ethyl]-1,2-ethanediamine	C4H10N6O6	30.5	1.63[f]	182.0	-	5996	3829	9.6	

Name	Formula								Structure
Trinitroethylcyano-methylnitramine	C4H4N6O8	-155.5	1.75[f]	154.3	138.7	5606	4308	2.7	
Bis-(2,2,2-trinitroethyl)-nitramine	C4H4N8O14	-353.8	1.97[f]	147.2	137.1	4373	3690	1.2	
N,N'-dimethyl-N,N'-dinitrooxamide	C4H6N4O6	-305.4[b]	1.52[b]	150.4	-	4772	3534	19.4	
Cyclo-1,3,5,7-tetramethylene-2,4,6,8-tetranitramine (HMX)	C4H8N8O8	75.0[b]	1.96[b]	173.3	-	6036	4081	7.1	
1,3,3,5,5-Pentanitropiperdine	C5H6N6O10	-210.0	1.82[f]	172.2	150.7	5949	4325	3.4	
Trinitroethyl-2-methoxy-ethylnitramine	C5H9N5O9	-424.2	1.62[f]	166.7	150.7	5343	3762	10.3	
1,7-Dimethoxy-2,4,6-trinitro-2,4,6-triazaheptane	C6H14N6O8	-400.5	1.55[f]	170.5	-	5077	3308	40.7	
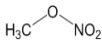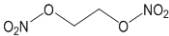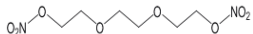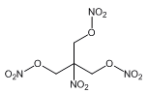
2,4,6-Trinitrophenyl-methyl-nitramine (Tetryl)	C7H5N5O8	20.0[b]	1.73[b]	130.8	258.3	5761	4183	7.8	
N-(2,2-Dinitrobutyl)-N-2,2-trinitro-1-butanamine	C8H14N6O10	-495.5	1.67[d]	150.6	148.5	4991	3335	19.6	
N,N'-Dinitro-N,N'-bis-(3-nitrazabutyl)-oxamide	C8H14N8O10	162.3	1.66[d]	152.8	-	6277	3980	22.1	
2,2,4,7,9,9-Hexanitro-4,7-diazadecane	C8H14N8O12	-438.4	1.63[d]	164.8	168	5297	3637	17.6	

[a]Ref. [3] [b]Ref. [6] [c]Ref. [8] [d]Ref. [9] [e]Ref. [10] [f]Ref. [11] [g]Ref. [5]

Table 5 The nitrate ester data set [6]. The sum formula, enthalpy of formation (ΔHf), density, bond dissociation energy (BDE), heat of detonation (Q), temperature of detonation (Tex), impact energy (I50) and molecular structure

Compound	Formula	ΔH_f[a] (kJ/mol)	Density[a] (g/cm^3)	BDE (kJ/mol)	Q (kJ/kg)	T_{ex} (K)	I_{50}[a] (J)	Molecular structure
Dinitrophenoxy-ethylnitrate	C8H7N3O8	-292.8	1.60	161.6	5058	3573	20	
Dioxyethylnitramine dinitrate (DINA)	C4H8N4O8	-275.7	1.49	160.1	5968	4139	6	

Dipentaerythritol hexanitrate (DIPEHN)	C10H16N6O19	-978.6	1.63	159.6	5754	3985	4	
Erythritol tetranitrate	C4H6N4O12	-502.5[b]	1.60	144.1	5709	4393	2	
Mannitol hexanitrate (MHN)	C6H8N6O18	-675.6	1.60	147.9	5801	4470	0.8	
Pentaerythriol tetranitrate (PETN)	C5H8N4O12	-538.8	1.76	151.9	6203	4337	3	
Trinitrophenoxyethylnitrate	C8H6N4O10	-260.3	1.68	158.7	5495	3996	7.9	
Nitroglycerine (NG)	C3H5N3O9	-370.6	1.59	153.0	6087	4541	0.2	
Methyltrimethylol-methanetrinitrate (TMETN)	C5H9N3O9	-425.0	1.46	154.4	5840	3891	0.2	
Butanetriol-trinitrate (BTTN)	C4H7N3O9	-405.9	1.52	149.1	6119	4330	1	
Diethyleneglycol dinitrate (DEGDN)	C4H8N2O7	-436.7	1.38	164.3	5495	3728	0.1	

248

Methylnitrate	C1H3N1O3	-155.9	1.22	175.3	6310	4418	0.2	
Nitroglycol (EGDN)	C2H4N2O6	-242.7	1.48	160.9	6424	4668	0.2	
Triethyleneglycol dinitrate (TEGDN)	C6H12N2O8	-628.8	1.34	165.3	4918	3218	12.7	
Trimethylolnitromethane trinitrate (NIBTN)	C4H6N4O11	-228.1	1.68	149.9	6815	4946	2	

[a]Ref. [6] [b]Ref. [2]

References

1. Wilson WS, Bliss DE, Christian SL (1990) Explosive Properties of Polynitroaromatics. Naval Weapons Center: China Lake, California
2. (1997) Data Base of Thermochemical Data. Fraunhofer Institut für Chemische Technologie
3. Keshavarz MH (2011) Prediction of the condensed phase heat of formation of energetic compounds. Journal of Hazardous Materials **190**(1):330-344
4. Keshavarz MH (2009) Predicting condensed phase heat of formation of nitroaromatic compounds. Journal of Hazardous Materials **169**(1):890-900
5. Storm CB, Stine JR, Kramer JF (1990) Sensitivity Relationships in Energetic Materials. In: Chemistry and Physics of Energetic Materials, Springer, pp 605-639
6. Meyer R, Köhler J, Homburg A (2007) Explosives, 6th edn. Wiley-VCH Verlag GmbH, Weinheim
7. (2007) ICT Database for Thermochemical values (Demo Version). Fraunhofer Institut fur Chemische Technologie, Pfinztal
8. Pristera F et al (1960) Analysis of explosives by using infrared spectroscopy. Anal. Chem. **32**:495-508.
9. Rice BM, Hare JJ, Byrd EF (2007) Accurate Predictions of Crystal Densities Using Quantum Mechanical Molecular Volumes. The Journal of Physical Chemistry A **111**(42):10,874-10,879

10. Chen SP et al (2003) Kinetics of the exothermic decomposition reaction of N-methyl-N-nitro-2,2,2-trinitroethanamine. *Chin. J. Chem.* **21**(11):1,419-1,421
11. Adolph HG, Holden JR, Chicra DA (1981) Relationships Between the Impact Sensitivity of High Energy Compounds and Some Molecular Properties Which Determine Their Performance. Naval Surface Weapons Center: Silver Spring, Maryland

Appendix C

Supplementary material of Paper V

Appendix D

Supplementary material of Paper VII

Appendix D. Supplementary material of Paper VII

Data from impact sensitivity tests of ageing Amatol using the BAM Impact Apparatus

Table 1

The results from the fallhammer drops on substance A_1. The asterisks denote that a 5 kg weight was used. For the rest, a 10 kg weight was used.

Height (log$_{10}$ cm)	1.75*	1.90*	1.70	1.80	1.85	2.00
# Reactions	0	0	0	0	0	1
# Trials	1	1	1	1	1	10

Table 2

The results from the fallhammer drops on substance A_2, using a 5 kg weight and the Bruceton up-and-down method with n=70.

Height (log$_{10}$ cm)	1.20	1.25	1.30	1.35	1.40	1.45	1.50	1.55	1.60	1.65	1.70	1.75	1.80
# Reactions	0	4	2	8	6	8	5	1	1	1	2	1	1
# Trials	4	6	9	13	13	12	5	1	1	2	2	1	1

Table 3

The results from the fallhammer drops on substance B, using a 5 kg weight and the Bruceton up-and-down method with n=30 drops.

Height (log$_{10}$ cm)	1.05	1.10	1.15	1.20	1.25	1.30	1.35	1.40	1.45	1.50	1.55
# Reactions	0	1	2	2	5	2	2	2	1	1	1
# Trials	1	3	3	6	6	3	3	2	1	1	1

Table 4

The results from the fallhammer drops on substance C, using a 5 kg weight and the Bruceton up-and-down method with n=50 drops.

Height (log$_{10}$ cm)	1.65	1.70	1.75	1.80	1.85	1.90	1.95
# Reactions	0	0	4	9	4	5	1
# Trials	1	5	14	14	9	6	1

Table 5

The results from the fallhammer drops on substance D, using a 5 kg weight and the Bruceton up-and-down method with n=70 drops.

Height (log$_{10}$ cm)	1.20	1.25	1.30	1.35	1.40	1.45	1.50	1.55	1.60	1.65	1.70	1.75
# Reactions	0	1	1	6	2	4	5	6	5	4	1	3
# Trials	1	2	7	8	6	8	10	10	8	4	3	3

Table 6

The results from the fallhammer drops on substance E, using a 5 kg weight and the Bruceton up-and-

down method with n=70 drops. Since we used a different template in this experiment than in the others, the unit for the experimental design was $\log_{10}$ J rather than $\log_{10}$ cm.

Energy ($\log_{10}$ J)	0.95	1.00	1.05	1.10	1.15	1.20	1.25	1.30	1.35
# Reactions	0	1	2	2	3	10	10	6	1
# Trials	1	3	4	5	13	20	16	7	1

Appendix E

Supplementary material of Paper VIII