

Prior distributions expressing ignorance about convex increasing failure rates

Jørund Gåsemyr¹ | Aliaksandr Hubin¹

Department of mathematics, University of Oslo, Oslo, Norway

Correspondence

Jørund Gåsemyr, Department of Mathematics, University of Oslo, Oslo, POBox 1053 Blindern 0316 Oslo, Norway.
Email: gaasemyr@math.uio.no

Abstract

This paper deals with the specification of probability distributions expressing ignorance concerning annual or otherwise discretized failure or mortality rates, when these rates can safely be assumed to be increasing and convex, but are completely unknown otherwise. Such distributions can be used as noninformative priors for Bayesian analysis of failure data. We demonstrate why a uniform distribution used in earlier work is unsatisfactory, especially from the point of view of insensitivity with respect to the time scale that is chosen for the problem at hand. We suggest alternative distributions based on Dirichlet distributed weights for the extreme points of relevant convex sets, and discuss which consequences a requirement for scale neutrality has for the choice of Dirichlet parameters.

KEYWORDS

Bayesian estimation, convex sets, mortality rates, noninformative prior distributions, scaling

1 | INTRODUCTION

Determination of annual mortality rates is an important task for actuarial science. Models for death or failure rates per time unit also arise in biological and technological contexts as well, with relevant time units changing from application to application. Since the main motivating example for this paper is taken from life insurance, we will mostly use the term “mortality

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Scandinavian Journal of Statistics* published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.

rate.” Various parametric distributions have been used in different applications, such as the Gompertz–Makeham distribution (Bowers et al., 1986) customarily used in the insurance industry. A common flexible framework for the various applications can be built on considering each of the mortality or failure rates as a parameter, and to introduce dependence between these parameters by restrictions on the parameter space. Such a model is introduced in Gelman et al. (1996), where in accordance with the increasing convexity of the hazard rate function of the Gompertz–Makeham distribution, the constraint is that the mortality rate curve should be increasing and convex. This would seem to be a natural restriction in many cases. Considering a data set from life insurance, Gelman et al. (1996) perform both a maximum likelihood and a Bayesian analysis based on this model. The Bayesian analysis requires a prior distribution, which is updated by the likelihood according to Bayes’ theorem to form the posterior, which is the basis for the inference. But the analysis can be made “objective” by choosing a noninformative prior, expressing ignorance about the mortality rates. How to do this is not at all obvious, however. In the present paper, we examine how a distribution expressing ignorance for a vector of increasing, convex mortality rates can be specified.

Since the mortality rate is the probability of dying within a given time interval for a randomly chosen individual from a given age class, the choice of a noninformative prior distribution for the parameter p of a binomial distribution is particularly relevant in our context. The conjugate prior is the beta distribution $\text{beta}(p; \alpha, \beta)$ with density proportional to $p^{\alpha-1}(1-p)^{\beta-1}$. With $\alpha = \beta = \epsilon$ close to 0, the posterior expectation of p based on a data set with y successes among N trials is $(\alpha + y)/(\alpha + \beta + N)$, which approximates the maximum likelihood estimate y/N . Note that, if we change the time scale by, for example, considering monthly rather than annual mortality rates, the posterior estimate based on this noninformative prior is on average scaled appropriately down in accordance with the smaller number of deaths over the shorter intervals. This is in accordance with the desired goal expressed in Gelman (1996) “that as the scale of discretization changes with fixed data the strength of the prior distribution, the Bayesian analysis would remain roughly constant.” In other words, we seek a prior that is approximately neutral with respect to the time scale that is chosen for the given problem. However, even though we may lack useful prior information about the level of each mortality rate individually, it would be unreasonable to use the noninformative beta prior to estimating the mortality rate for each age class separately. The constraint introduced in Gelman et al. (1996) of an increasing and convex mortality rate curve is a natural way of formalizing prior beliefs about the shape of this curve. With such a model constraint the specification of a noninformative and scale neutral prior distribution is not so straightforward. Also, it may be difficult to verify for a particular distribution that “the Bayesian analysis would remain roughly constant” under changes of scale. We cannot expect to find a conjugate prior for which the posterior expectations can be analytically computed, such as in the beta-binomial case, nor is there necessarily an obvious target such as y/N to aim for. If, for a particular prior, we are nevertheless able to analyze the effect a change in scale would have on the inference. Such an analytic examination can save us from performing numerical computations with an unreasonable prior.

It is natural to consider the uniform distribution as a candidate for a noninformative and scale neutral prior for the convex and increasing mortality rate model, since in general, the uniform distribution is a common choice for a prior distribution intended to be noninformative. A major disadvantage of this choice is that the prior changes under reparameterization, but at least it has the virtue that maximization of the posterior density is equivalent to maximizing the likelihood. Accordingly, the maximum likelihood analysis in Gelman et al. (1996) is followed up by a Bayesian analysis based on a uniform prior on the restricted parameter space. Comparison of

draws from the posterior distribution of the mortality rates with the data, especially concerning the higher age classes, suggests that this may not be a good choice. In Theorem 1 of the present paper, we provide a mathematical explanation why the mortality rate of the highest age class, in particular, seems to be overestimated by using the uniform distribution as a prior. An explanation involving the whole vector of mortality rates is given in Gelman (1996), Section 3.3, which attributes the apparent lack of fit to the prior expectation of the mortality rate curve climbing too steeply for higher age classes, as well as to the marginal variances of the prior distribution converging to 0 as the scale of discretization becomes finer, implies that a fixed set of data is eventually unable to move the posterior distribution away from the prior expectation. Such behavior of the marginal prior variances is certainly unsatisfactory. Since we find the argument to be somewhat inaccurate, we provide a more detailed derivation of the limiting behavior under the uniform prior distribution of the expected shape of the mortality rate curve and the marginal mortality rate variances under scale refinement, see Proposition 2, respectively, Corollary 1 in Section 5 of the present paper. A new framework developed in Section 4 of this paper provides the basis for the analysis, and gives rise to a class of prior distributions that according to Theorem 5, with a suitable choice of parameters, contains distributions for which the marginal variances stay bounded away from 0 regardless of the scale of discretization. Hence, these distributions should have a potential for overcoming the difficulties reported in Gelman (1996) and therefore accomplishing the desired goal of approximate insensitivity to change in the scale of discretization. The good properties of these distributions are demonstrated numerically on the insurance dataset in Section 6.

The rest of the paper is organized as follows: Section 2 provides the insurance data and presents formally the model for convex, increasing mortality rates and the uniform prior on this set. In Section 3, we provide new explanations why the uniform prior is problematic, based on establishing the distributions of the initial and final mortality rates, that is, the mortality rates for the lowest and highest age class, respectively, which essentially determine the location of the mortality rate curve. In order to examine whether uniformity can nevertheless be part of a sensible noninformative modeling for this problem, we introduce distributions involving such uniform modeling in Section 4. Motivated by alternative characterizations of these distributions given in Theorems 3 and 4, we suggest in Definition 1 and Definition 2 new classes of prior distributions for convex and increasing mortality rate functions. In Section 5, we analyze the limiting behavior under scale refinement of the prior expectation and variance of the mortality rates under some of these priors. With the uniform prior being a special case, this provides an additional explanation why the uniform prior is unsatisfactory, not only for the location, but also for the shape of the mortality rate curve. This analysis is also the basis for suggesting distributions that have a potential for dealing more satisfactorily with the requirement for scale neutrality. Numerical experiments presented in Section 6 confirm the good behavior of these distributions. Some conclusions are given in Section 7.

2 | THE MODEL AND THE UNIFORM PRIOR FOR CONVEX, INCREASING MORTALITY RATES

In a Bayesian model evaluation context, Gelman et al. (1996) reanalyzes a dataset from life insurance originally studied in Broffitt (1988), containing the number N_t of insured and the number y_t of deaths ($t = 1, 2, \dots, 30$) in 30 consecutive age classes, ranging from 35 to 64, under a certain life insurance policy. The data are displayed in Table 1 and are henceforth referred to as

TABLE 1 Mortality rate data (LIN) from Broffitt (1988)

Age, t	N_t	y_t	A age, t	N_t	y_t
35	1771.5	3	50	1516.0	4
36	2126.5	1	51	1371.5	7
37	2743.5	3	52	1343.0	4
38	2766.0	2	53	1304.0	4
39	2463.0	2	54	1232.5	11
40	2368.0	4	55	1204.5	11
41	2310.0	4	56	1113.5	13
42	2306.5	7	57	1048.0	12
43	2059.5	5	58	1155.0	12
44	1917.0	2	59	1018.5	19
45	1931.0	8	60	945.0	12
46	1746.5	13	61	853.0	16
47	1580.0	8	62	750.0	12
48	1580.0	2	63	693.0	6
49	1467.5	7	64	594.0	10

the LIN dataset. Bayesian models for analysing these data were also discussed in Carlin (1993). For this dataset, the observed mortality rates y_t/N_t show a generally increasing trend, but jump quite irregularly with occasional downward jumps from one age class to the next. Attributing this irregular pattern to random fluctuations, Gelman et al. (1996) want to build into their prior for the underlying mortality rates θ_t the constraint that these rates are increasing and convex. Having also performed a maximum likelihood analysis, it is stated in Gelman (1996) that “since we were willing to use the maximum likelihood estimate, it seems reasonable to use a uniform prior distribution on θ .” Being uniform, this prior distribution might naively be perceived as noninformative, although it is not portrayed as such in Gelman et al. (1996) or Gelman (1996). In the next section, we investigate the potential for the uniform distribution on the set of an arbitrary number n increasing and convex mortality rates for being noninformative, by analyzing the consequences of this choice for the marginal prior distribution of θ_1 and θ_n , that is, the initial (first) and final (last) mortality rates, respectively, and hence for the location of the mortality rate curve under the uniform prior. Some possible consequences for the posterior distribution of these mortality rates in general and in particular with the LIN data are discussed. In Section 6, a numerical investigation of the posterior distribution for these data under the uniform and other priors is provided.

We phrase the problem in somewhat more general terms. The data are assumed to arise from recording failures of objects or deaths of individuals belonging to a certain homogeneous population, occurring at ages contained in an interval $(T_0, T_n]$. In a laboratory experiment studying for example, the effect of exposure to some chemical substance, “age” may refer to the time since onset of exposure rather than physical age. Also, age could be replaced by some relevant measure of accumulated load such as, for example, total mileage of a car. The data are discretized by counting the number of deaths occurring at ages in intervals $(T_{t-1}, T_t]$, $t = 1, 2, \dots, n$, with an arbitrary number n (to be chosen by researchers or decision-makers) of equidistant measuring

points T_1, \dots, T_n in the interval $(T_0, T_n]$, the common distance $T_t - T_{t-1} = (T_n - T_0)/N$ between consecutive measuring points hence also being arbitrary. We denote by N_t the number of individuals at risk at age T_{t-1} and by Y_t the number of deaths at ages in $(T_{t-1}, T_t]$. Some adjustments of N_t may be made due to individuals joining the population or leaving the population without dying during the age interval $(T_{t-1}, T_t]$. In the LIN dataset (Table 1), such individuals are counted as 0.5. These adjustments are assumed not to depend on the underlying mortality rates. An arbitrary degree of right censoring may take place also due to individuals surviving the age T_n or not reaching this age before the collection of data is completed.

The probability of a random individual having survived age T_{t-1} of dying before T_t is denoted by $\theta_t, t = 1, 2, \dots, n$. Hence, Y_t is binomially distributed with parameters N_t , and θ_t . These mortality rates are collected in the vector θ . The mortality rates θ_1 and θ_n are referred to as the initial and the final mortality rates, respectively. Since under the assumption of increasing mortality rates all the rates must lie between θ_1 and θ_n , the value of these parameters essentially determine the location of the mortality rate curve, and they are the focus of the next section. The remaining rates describe the shape of the curve. This aspect of the model will be discussed in Section 5. Assuming that Y_1, \dots, Y_n are independent and binomially distributed given N_1, \dots, N_n and θ , the likelihood for a dataset with observed number of deaths given by $\mathbf{y} = (y_1, \dots, y_n)$ is

$$p(\mathbf{y}|\theta, N_1, \dots, N_n) = \prod_{t=1}^n \theta_t^{y_t} (1 - \theta_t)^{N_t - y_t}.$$

For any integer p we denote by m_p the Lebesgue measure on R^p . Let J_n be the set of convex, increasing θ , that is,

$$J_n = \{\theta \in R^n : 0 \leq \theta_1 \leq \dots \leq \theta_n \leq 1 \text{ and } \theta_{t+1} + \theta_{t-1} \geq 2\theta_t \text{ for } t = 2, \dots, n-1\}. \quad (1)$$

Denoting by I the indicator function, the uniform prior distribution used in Gelman et al. (1996) is hence

$$\pi_{\text{GMS}}(\theta) = I_{J_n}(\theta)/m_n(J_n). \quad (2)$$

3 | THE MARGINAL DISTRIBUTIONS OF θ_1 AND θ_n UNDER THE UNIFORM PRIOR

3.1 | The marginal prior for the final mortality rate θ_n

Using $\pi_{\text{GMS}}(\theta)$ as a prior, based on simulation experiments it was observed in Gelman et al. (1996) that “posterior predictive data sets were mostly higher than the observed data for the later ages.” One possible explanation mentioned by the authors is that this could be a selection effect destroying the convexity, due to the insurance company having “screened out some high risk older people.” As also indicated by Gelman (1996), it is reasonable to ask if this could instead be related to the behavior of the prior for higher age classes. It is therefore natural to examine, in particular, the marginal distribution for the final mortality rate θ_n under the uniform prior on J_n . The following theorem suggests that the problem is indeed due to the inherent mathematical properties of this distribution linked to the geometry of the set J_n . Essentially, as reflected in the

proof, the problem is that the $(n - 1)$ -dimensional volume of the set of $\theta \in J_n$ such that $\theta_n = x$ is proportional to x^{n-1} , resulting in increased probability weight being put on higher values of θ_n as n increases.

Theorem 1. Under π_{GMS} the rate θ_n has the marginal density $\text{beta}(\theta_n; n, 1)$, that is, is beta-distributed with parameters n and 1 .

Proof. For any vector $\theta = (\theta_1, \dots, \theta_n)^T \in R^n$, denote by $\theta_{1:n-1}$ the vector $(\theta_1, \dots, \theta_{n-1})^T \in R^{n-1}$. Define

$$K_n = \{\theta_{1:n-1} \in R^{n-1} : (\theta_{1:n-1}^T, 1)^T \in J_n\}. \tag{3}$$

Then $\theta \in J_n$ if and only if $\theta = 0_n = (0, \dots, 0)^t$ or $0 < \theta_n \leq 1$ and $(1/\theta_n)(\theta_{1:n-1}) \in K_n$, or equivalently, $0 \leq \theta_n \leq 1$ and $\theta_{1:n-1} \in \theta_n K_n$. Hence, calculating the volume of the set of $\theta \in J_n$ such that $\theta_n \leq x$ by integrating along the θ_n -axis, we obtain

$$\begin{aligned} m_n(\{\theta \in J_n : \theta_n \leq x\}) &= \int_0^x m_{n-1}(\{\theta_{1:n-1} : (\theta_{1:n-1}^T, \theta_n)^T \in J_n\}) d\theta_n \\ &= \int_0^x m_{n-1}(\theta_n K_n) d\theta_n = m_{n-1}(K_n) \int_0^x \theta_n^{n-1} d\theta_n \\ &= (m_{n-1}(K_n)/n)x^n. \end{aligned}$$

By the uniformity assumption, the probability under (2) that $\theta_n \leq x$ is proportional to this volume, it follows that the density for θ_n under (2) is proportional to θ_n^{n-1} , that is, equals $(\Gamma(n + 1)/\Gamma(n)\Gamma(1))\theta_n^{n-1}(1 - \theta_n)^{1-1} = \text{beta}(\theta_n; n, 1)$. ■

Remark: According to Gelman (1996), Section 3.1, the beta distribution also characterizes the marginal distributions for the parameters of a discrete approximation to an increasing but not necessarily convex function under a uniform prior.

An immediate consequence of Theorem 1 is that the expected mortality rate for the highest age class under π_{GMS} is $n/(n + 1)$, hence increases with n . This is quite unsatisfactory, since all the mortality rates ought to decrease rather than increase if the interval length $T_t - T_{t-1}$ decreases, and as a result n increases. Moreover, since we want a prior distribution suitable also with an arbitrary degree of right censoring at the final measuring time T_n , the highest mortality rate θ_n may be small and hence the prior expectation $n/(n + 1)$ unreasonably large even for small values of n . This is relevant for the special case analyzed in Gelman et al. (1996), with $n = 30$ representing age 64. According to $\pi_{\text{GMS}}(\theta)$, a person being lucky enough to have survived to the age of 64 will have only a 3% chance of surviving another year. We think this is a rather disturbing consequence, considering that we believe the data to come from a 20th century western human population which is not known to have been exposed to any kind of exceptionally high-risk factors. This is in accordance with Hjort et al. (2006), section 9, which mentions some other gloomy facts facing persons whose survival distributions follow the uniform prior. However, despite its peculiar properties when taken literally as a joint distribution for the vector of mortality rates, Hjort et al. (2006) consider it as a reasonable choice of a vague prior leading to an “objective analysis of the data,” and discards it only in the context of prior predictive model evaluation. We note, however, that the marginal variance of θ_n is $n/((n + 1)^2(n + 2))$. In the context of the LIN data, with $n = 30$, the resulting small SD of 0.03 could hardly be considered as an indication of noninformativity. Since the marginal variance is of order $1/n^2$, a change in scale and a corresponding increase of n , for instance, considering monthly mortality rates, would aggravate this problem.

At the same time, due to the reduced size of the basic time intervals, the true value of θ_n would decrease with such a change in scale, and hence become more and more out of reach for a posterior distribution based on a fixed dataset. This behavior under a change of time scale indicates that the desired scale neutrality is not achieved by the prior (2).

It is worth noting that with the binomial likelihood $\theta_n^{y_n}(1 - \theta_n)^{N_n - y_n}$ for θ_n , the marginal prior for θ_n would be updated to a beta distribution with parameters $n + Y_n$ and $N_n - y_n + 1$, if we were to use only data from age class n . For the LIN data with $n = 30, y_n = 10$ and $N_n = 594$, this would yield the posterior expectation $E(\theta_n|y_n) = (y_n + n)/(N_n - y_n + 1) = (10 + 30)/625$. This is almost four times higher than the posterior expectation when using a standard noninformative beta prior, which is approximately $10/594$. The numerical results given in Section 6 indicate that the consequences may not be quite that bad when taking data for all age groups into consideration, but confirm the predicted overestimation of the mortality rate for the highest age class. Hence it is reasonable to conclude from Theorem 1 that prior (2) is not as neutral as might naively be thought, and is a bad choice of a non-informative prior.

3.2 | Can the problem be remedied by restricting the support of θ_n ?

The mortality rate model of Gelman et al. (1996) is also considered in Johnson (2007) in the context of Model evaluation. In order to obtain a sensible prior predictive sample, Johnson (2007) suggests limiting the support of the mortality rates to the interval $[0, 0.05]$. A similar suggestion appears in Hjort et al. (2006). Forcing the posterior distribution of the mortality rates downward by restricting the support in this way could conceivably remedy the problems with the uniform prior demonstrated above. We can, however, show that the pressure toward higher values for θ_n remains. Updating the marginal prior density, which under the restricted support is proportional to $I_{[0,0.05]}(\theta_n)\theta_n^{n-1}$, only with data y_n from the n th age group, we can show that the posterior expectation of θ_n given y_n is close to the support limit of 0.05. To see this, define integers k, r such that $k - 1 + r = n - 1 + y_n$ and r is minimal with respect to satisfying the inequality $r \geq 0.05(N_n - y_n + r)$. Then the posterior density for θ_n can be written as $I_{[0,0.05]}(\theta_n)\theta_n^{k-1}g(\theta_n)$, where $g(\theta_n)$ is proportional to $\theta_n^r(1 - \theta_n)^{N_n - y_n}$. The function $g(\theta_n)$ is increasing on $[0, 0.05]$ and by the proposition given below it follows that the posterior expectation is at least $0.05k/(k + 1)$. For the LIN data $n = 30, y_n = 10$ and $N_n = 594$, so we can take $r = 31$, giving $k = 9$. Consequently, the posterior expectation is at least 0.045, which is well above the observed mortality rates for the highest age classes.

Proposition 1. *Suppose $\theta \in [0, a]$ has density of the form $f(\theta) = \theta^{k-1}g(\theta)$, where $g(0) = 0$ and $g(\theta)$ is increasing on $[0, a]$. Then $E(\theta) \geq ka/(k + 1)$.*

Proof. Using first integration by parts, then the inequality $\theta^{k+1} \leq a\theta^k$, and finally integration by parts again, we obtain

$$\begin{aligned} E(\theta) &= \int_0^a \theta^k g(\theta) d\theta = (a^{k+1}/(k+1))g(a) - \int_0^a (\theta^{k+1}/(k+1))g'(\theta) d\theta \\ &\geq (ka/(k+1))[(a^k/k)g(a) - \int_0^a (\theta^k/k)g'(\theta) d\theta] \\ &= (ka/(k+1)) \int_0^a \theta^{k-1} g(\theta) d\theta = ka/(k+1). \end{aligned}$$

■

The support $[0, 1]$ for θ_n implied by the prior distribution $\pi_{\text{GMS}}(\theta)$ is needed to cover any possible application of the convex and increasing mortality rate model, although it is unreasonably wide in the insurance application. But even if the support is drastically shortened, this proposition indicates that the uniform distribution with the restricted support $[0, a]$ for θ_n is still not a satisfactory choice of a noninformative prior distribution. An increase in the number n of measuring points would aggravate the effect, provided that the number of deaths y_n in the highest age class decreases more slowly than the increase in n . Indeed, writing $n - 1 + y_n$ as $k - 1 + r$ and $g(\theta_n) \propto \theta_n^r (1 - \theta_n)^{N_n - y_n}$ as above, we see that with a nonincreasing population size N_n , the conditions of the proposition are satisfied with an increasing value of k when n increases. This pushes the lower bound $ak/(k + 1)$ for the posterior expectation of θ_n toward the support limit a when n increases, whereas in order to be scale neutral it ought to decrease when n increases.

The suggestion to restrict the support of θ_n is also not entirely satisfactory from a puristic Bayesian point of view, since it seems to be based on an initial inspection of the data revealing that the maximum observed mortality rate is less than 0.02. In fact, Carlin (1993) uses the extremely data-based prior support $[0, 0.02]$. On the other hand, restricting the support of the prior without information of this sort, there is a risk of the support not containing the true value. Ironically, while Hjort et al. (2006) warn against such a risk, this is the case for its rather casual suggestion of $[0.03, 0.07]$ as support.

3.3 | The marginal prior for the initial mortality rate θ_1

Before turning to the introduction of alternative priors, it is natural to consider the possibility that the prior (2) could lead to biased estimation also for the initial mortality rate θ_1 . The following counterpart to Theorem 1 indicates that this could happen in some cases.

Theorem 2. *Under π_{GMS} the rate θ_1 has the marginal density $\text{beta}(\theta_1; 1, n)$, that is, is beta-distributed with parameters 1 and n .*

The proof is similar to the proof of Theorem 1. Parallel to the set K_n of (3) it builds on the set

$$L_n = \{ \theta_{2:n} \in R^{n-1} : (0, \theta_{2:n}^T)^T \in J_n \}. \quad (4)$$

The proof is a little more complex and is given in the Appendix.

It follows from Theorem 2 that the marginal expectation of the uniform distribution on J_n is $1/N$, and hence decreases at a pace that matches the decrease in the mortality rates that we would expect as the basic time intervals become shorter. This should indicate a less problematic prior modeling under (2) for the initial than for the final mortality rate. However, if we were to update the prior with data only from age group 1, the posterior expectation would be $E(\theta_1 | y_1) = (y_1 + 1)/(N_1 + n + 1)$. If θ_1 is very close to 0 and as a result y_1 is very small, and if not only absolute but also relative error is important, the extra 1 added to y_1 could lead to an undesirable relative overestimation of θ_1 when combining $\pi_{\text{GMS}}(\theta)$ with the likelihood. When n increases, relatively more weight is put on this extra 1 in this posterior conditional expectation, since, for a fixed data set, the value of y_1 decreases with decreasing the length of the basic time interval. Depending on the true shape of the mortality rate curve, this could to some extent be compensated for by data from the next age classes. Nevertheless, it seems reasonable to conclude that also for the initial mortality rate the desired scale neutrality is not achieved in general.

4 | ALTERNATIVE NONINFORMATIVE PRIORS

Our desired goal is to create a scale neutral noninformative prior or family of priors, equally well suited for arbitrary populations including nonhuman ones, as well as arbitrary time intervals between measuring points. Consider, for example, a laboratory experiment testing the toxicity of a certain chemical substance to a population of some animal species, measured on a suitable time scale (e.g., hours, days, or weeks). In some of such cases, one might very well be in a state of complete ignorance prior to the experiment, and a scale neutral noninformative prior distribution is called for. In view of the problems demonstrated in the previous section, we propose in general to first specify a noninformative marginal prior distribution for the final and/or initial mortality rates, enabling the data to locate the curve in the correct region. Thereafter a joint conditional distribution is specified for the remaining mortality rates, aimed at enabling the data to determine the correct shape of the curve. In particular, we will examine to what extent the uniform prior distribution, found inappropriate when applied to the entire θ -vector, can be more successfully applied for this more restricted task. This examination is performed both theoretically and numerically, and is reported in Sections 5 and 6, respectively. Both the theoretical analysis and the Monte Carlo simulation needed for the numerical investigation built on alternative characterizations of the relevant distributions given in Theorems 3 and 4. More flexible alternative suggestions presented in Definitions 1 and 2 are also motivated by these characterizations.

In the following subsection, we focus specifically on the problems discussed in Section 3.1 concerning the final mortality rate. By choosing a noninformative prior for θ_n the magnitude of this parameter is determined within a Bayesian framework, and we avoid the non-Bayesian, data-based specification of an upper bound for the support used by Carlin (1993).

4.1 | Modifying the marginal prior for the final mortality rate

As a marginal prior for θ_n , it is natural to choose a beta distribution, which is a conjugate prior for the factor $\theta_n^{y_n}(1 - \theta_n)^{N_n - y_n}$ of the binomial likelihood. As a start, we combine it with a uniform conditional distribution for the rest of the parameters. Recalling the definition of K_n in (3), We thus introduce the prior (the superscript F indicating “final”):

$$\pi_{\alpha, \beta}^F(\theta) = \text{beta}(\theta_n; \alpha, \beta) I_{\theta_n K_n}(\theta_{1:n-1}) / m_{n-1}(\theta_n K_n). \quad (5)$$

Various degrees of weaker or stronger informativity based on subjective knowledge can be accommodated for by different choices of the parameters α and β . The uniform distribution on J_n is obtained as the special case $\alpha = n$, $\beta = 1$, since then (5) is proportional to

$$\theta_n^{n-1} I(\theta_n \leq 1) I_{\theta_n K_n}(\theta_{1:n-1}) / m_{n-1}(\theta_n K_n) = I_{J_n}(\theta) / m_{n-1}(K_n) \propto \pi_{\text{GMS}}(\theta).$$

From the point of view of scale neutrality and noninformativity, as indicated in the introduction, the choice $\alpha = \beta = \epsilon$ close to 0 is preferable. If the sample size N_n is not too small, and the time scale is sufficiently coarse, data y_n from the n th age group alone should then suffice to locate the marginal posterior distribution of θ_n adequately. If the number n of measuring points is drastically increased, and as a result, the true value of θ_n and the observed y_n become very small, data from lower age classes should still contribute to a sensible location of this distribution.

However, we also want the prior to be noninformative with respect to the shape of the mortality rate curve, which is determined by the other parameters. In Section 5, we will discuss whether the conditional uniform distribution on these parameters is a satisfactory prior choice from this point of view. The basis for this discussion is Theorem 3 below, expressing the prior distribution (5) in terms of familiar random variables. This builds on the following lemma:

Lemma 1. Denote by \bar{K}_n the set of convex and increasing functions f on $\{1, 2, \dots, n\}$ satisfying $0 \leq f(1) \leq f(n) = 1$. Then \bar{K}_n is a convex set spanned by n extreme points. Identifying f with the vector $(f(1), \dots, f(n))^T$, the extreme points are

$$\begin{aligned} \mathbf{w}^1 = \mathbf{1}_n &= (1, 1, \dots, 1)^T, & \mathbf{w}^2 &= \frac{1}{n-1}(0, 1, 2, \dots, n-1)^T, \\ \mathbf{w}^3 &= \frac{1}{n-2}(0, 0, 1, 2, \dots, n-2)^T, \dots, \\ \mathbf{w}^i &= \frac{1}{n-i+1}(0, \dots, 0, 1, 2, \dots, n-i+1)^T, \dots, & \mathbf{w}^n &= (0, 0, \dots, 0, 1)^T. \end{aligned}$$

Proof. It is straightforward to verify that the set \bar{K}_n is convex. Clearly $\mathbf{w}^1 = \mathbf{1}_n$ is an extreme point, since, for all t , $u_t \leq 1$ for every $\mathbf{u} \in \bar{K}_n$. Consider \mathbf{w}^i for $i \geq 2$. If $\mathbf{w}^i = a\mathbf{u}^1 + (1-a)\mathbf{u}^2$ with \mathbf{u}^1 and $\mathbf{u}^2 \in \bar{K}_n$ and $0 < a < 1$, then clearly $u_t^j = 0$ for $1 \leq t \leq i-1, j = 1, 2$. Since the graph of the function represented by \mathbf{w}^i is a straight line from $(i-1, 0)$ to $(n, 1)$, the convexity assumption implies that \mathbf{u}^1 and \mathbf{u}^2 have to follow this straight line as well. Hence, $\mathbf{u}^1 = \mathbf{u}^2 = \mathbf{w}^i$, and it follows that \mathbf{w}^i is an extreme point. Now let $\mathbf{u} \in \bar{K}_n$ be arbitrary. Note that the $n \times n$ matrix W^n formed by the column vectors $\mathbf{w}^1, \dots, \mathbf{w}^n$ is lower triangular with non-zero values on the diagonal. Hence, its determinant is positive, and it follows that $\mathbf{w}^1, \dots, \mathbf{w}^n$ are linearly independent. We may therefore write \mathbf{u} uniquely as a linear combination $\mathbf{u} = a_1\mathbf{w}^1 + \dots + a_n\mathbf{w}^n$. Since $u_n = 1$ and $w_n^i = 1$ for $i = 1, \dots, n$, it follows that $a_1 + \dots + a_n = 1$. It remains to prove that $a_i \geq 0$ for $i = 1, \dots, n$. Clearly, $a_1 = u_1 \geq 0$. Since $u_2 - u_1 = a_2/(n-1)$ and \mathbf{u} is increasing, it follows that $a_2 \geq 0$. It is not too hard to see more generally that for $i \geq 2$ we have

$$u_i - u_{i-1} = a_2/(n-1) + \dots + a_i/(n-i+1).$$

Consequently, for $i \geq 3$ we have $a_i/(n-i+1) = (u_i - u_{i-1}) - (u_{i-1} - u_{i-2})$, which is nonnegative by the convexity assumption. Hence, \mathbf{u} is a convex combination of $\mathbf{w}^1, \dots, \mathbf{w}^n$. ■

We denote by $\text{Dir}(n; \gamma_1, \gamma_2, \dots, \gamma_n)$ the Dirichlet distribution for the vector $(X_1, X_2, \dots, X_{n-1})^T$ whose density is given by

$$(\Gamma(\gamma_1 + \dots + \gamma_n) / (\Gamma(\gamma_1) \dots \Gamma(\gamma_n))) X_1^{\gamma_1-1} \dots X_{n-1}^{\gamma_{n-1}-1} (1 - X_1 - \dots - X_{n-1})^{\gamma_n-1}.$$

Using Lemma 1, the distribution $\pi_{\alpha, \beta}^F$ can now be characterized as follows:

Theorem 3. A variable θ distributed according to $\pi_{\alpha, \beta}^F$ has the distribution of $\theta_n(X_1\mathbf{w}^1 + \dots + X_n\mathbf{w}^n)$, where $\theta_n \sim \text{beta}(\cdot; \alpha, \beta)$, $X_n = 1 - X_1 - \dots - X_{n-1}$ and $(X_1, \dots, X_{n-1})^T$ is Dirichlet distributed $\text{Dir}(n; 1, 1, \dots, 1)$, independently of θ_n .

Proof. Note that the $\text{Dir}(n; 1, 1, \dots, 1)$ distributed vector $(X_1, \dots, X_{n-1})^T$ is uniformly distributed on the set

$$S_n = \{(x_1, x_2, \dots, x_{n-1})^T : 0 \leq x_i \leq 1 \text{ for } i = 1, \dots, n-1; x_1 + \dots + x_{n-1} \leq 1\}.$$

Recall that W^n is the matrix formed by the column vectors $\mathbf{w}^1, \dots, \mathbf{w}^n$ of Lemma 1. Let $W_{-n;-n}^n$ be the matrix obtained from W^n by deleting the last row, consisting of 1s, and the last column, consisting of 0s except for the last entry. Define the $n - 1$ -dimensional vector variable

$$\boldsymbol{\phi} = W_{-n;-n}^n(X_1, \dots, X_{n-1})^T = X_1\mathbf{w}_{1:n-1}^1 + \dots + X_{n-1}\mathbf{w}_{1:n-1}^{n-1}. \tag{6}$$

Since the first $n - 1$ entries of the last column of W^n are 0, it follows that $\boldsymbol{\phi}$ consists of the first $n - 1$ entries of $W^n(X_1, \dots, X_n)^T = X_1\mathbf{w}^1 + \dots + X_n\mathbf{w}^n$. The last entry of this vector is 1, and consequently $X_1\mathbf{w}^1 + \dots + X_n\mathbf{w}^n = (\boldsymbol{\phi}^T, 1)^T$. By Lemma 1, the set \bar{K}_n is spanned by such vectors, and accordingly the set K_n of (3) is spanned by the vectors $\boldsymbol{\phi}$. By (6), the variable $\boldsymbol{\phi}$ is the image of $(X_1, \dots, X_{n-1})^T$ under the linear transformation defined by the matrix $W_{-n;-n}^n$. Since this transformation has a constant Jacobian determinant, and since $(X_1, \dots, X_{n-1})^T$ is uniformly distributed on S_n , it follows that the distribution of $\boldsymbol{\phi}$ on K_n is uniform as well. Scaling $\boldsymbol{\phi}$ down by multiplying with θ_n distributed according to $\text{beta}(\cdot; \alpha, \beta)$ results in a uniform distribution for $\theta_{1:n-1}$ given θ_n on $\theta_n K_n$, in accordance with (5). ■

In view of Theorem 3, it seems natural to introduce the following generalization of (5):

Definition 1. Let $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)^T$ be an n -dimensional vector of positive, real numbers. We define $\pi_{\alpha,\beta,\boldsymbol{\gamma}}^F(\boldsymbol{\theta})$ to be the density for a variable of the form $\theta_n(X_1\mathbf{w}^1 + \dots + X_n\mathbf{w}^n)$, where $\theta_n \sim \text{beta}(\cdot; \alpha, \beta)$, $X_n = 1 - X_1 - \dots - X_{n-1}$ and $(X_1, \dots, X_{n-1})^T$ is Dirichlet distributed $\text{Dir}(n; \gamma_1, \gamma_2, \dots, \gamma_n)$, independently of θ_n .

The posterior distribution resulting from updating this prior with the likelihood is proportional to

$$\pi_{\alpha,\beta,\boldsymbol{\gamma}}^F(\boldsymbol{\theta}) \prod_{t=1}^n \theta_t^{\gamma_t} (1 - \theta_t)^{N_t - \gamma_t}.$$

the resulting posterior does not belong to the same class of distributions as the prior, but can be obtained using MCMC. In Section 6, we use Stan (Carpenter et al., 2017) for the analysis of the LIN data. The results show that using Definition 1 allows significant improvements over the uniform distribution on J_n and even over (5) with $\alpha = \beta$ close to 0. The choice of parameters is motivated by the theoretical analysis of Section 5, where it is demonstrated that it is possible to choose Dirichlet parameters for $\pi_{\alpha,\beta,\boldsymbol{\gamma}}^F(\boldsymbol{\theta})$ in such a way that the marginal prior variances for the mortality rate parameters are bounded away from 0, regardless of the number n of measuring points. This should allow data to adjust the prior, which indicates that this generalisation of (5) is useful from the point of view of scale neutrality. However, for completeness, we first consider parallels to (5) where either θ_1 or both θ_1 and θ_n follow noninformative beta priors, and define variants of these distributions corresponding to the prior distributions of Definition 1.

4.2 | Modified priors involving the initial mortality rate

The discussion at the end of Section 3.3 suggests that the possibility of overestimation of a very small θ_1 could be a matter of concern. This could be handled by an approach parallel to that used in the previous subsection. Recalling Equation (4), which defines the set L_n , the idea leading to (5) suggests the prior (the superscript I indicating “initial”):

$$\pi_{\alpha,\beta}^I(\boldsymbol{\theta}) = \text{beta}(\theta_1; \alpha, \beta) I_{\theta_1 \mathbf{1}_{n-1} + (1-\theta_1)L_n}(\boldsymbol{\theta}_{2:n}) / m_{n-1}((1 - \theta_1)L_n). \tag{7}$$

To avoid the possible disadvantage of the $\pi_{\text{GMS}}(\theta)$ prior discussed at the end of Section 3.3, we suggest to choose α, β close to 0.

The concerns raised with respect to the marginal distributions for θ_1 and θ_n under the $\pi_{\text{GMS}}(\theta)$ prior can be handled simultaneously by combining (5) and (7) by specifying a joint prior for these parameters, and a uniform conditional distribution for $\theta_{2:n-1}$ given θ_1 and θ_n . Define $M_n = \{\theta_{2:n-1} \in R^{n-2} : (0, \theta_{2:n-1}^T, 1)^T \in J_n\}$. Similarly to the proofs of Theorems 1 and 2, it can be seen that $\theta \in J_n$ if and only if $0 \leq \theta_1 \leq \theta_n \leq 1$ and $\theta_{2:n-1} \in \theta_1 \mathbf{1}_{n-2} + (\theta_n - \theta_1)M_n$. This leads to the following definition, (the superscript IF indicating “initial and final”):

$$\pi_{\alpha_1, \beta_1, \alpha_n, \beta_n}^{\text{I,F}}(\theta) \propto I_{\theta_1 \leq \theta_n}((\theta_1, \theta_n)^T) \text{beta}(\theta_1; \alpha_1, \beta_1) \text{beta}(\theta_n; \alpha_n, \beta_n) \\ I_{\theta_1 \mathbf{1}_{n-2} + (\theta_n - \theta_1)M_n}(\theta_{2:n-1}) / m_{n-2}((\theta_n - \theta_1)M_n). \quad (8)$$

The following lemma describes the sets of extreme points needed to characterize the distributions (7) and (8) analogously to the characterization of (5) in Theorem 3.

Lemma 2.

- (i) The set \bar{L}_n of convex, increasing functions on $\{1, 2, \dots, n\}$ satisfying $0 = f(1) \leq f(n) \leq 1$ is a convex set spanned by n extreme points. the extreme points are $\mathbf{w}^0 = \mathbf{0}_n = (0, 0, \dots, 0)^T$ and the $n - 1$ vectors $\mathbf{w}^2, \dots, \mathbf{w}^n$ given in Lemma 1.
- (ii) The set \bar{M}_n of convex, increasing functions on $\{1, 2, \dots, n\}$ satisfying $0 = f(1) \leq f(n) \leq 1$ is a convex set spanned by the $n - 1$ extreme points $\mathbf{w}^2, \mathbf{w}^3, \dots, \mathbf{w}^n$.

The proof is given in the Appendix.

We can now give the following characterization of the distributions (7) and (8):

Theorem 4.

- (i) a variable θ distributed according to $\pi_{\alpha, \beta}^{\text{I}}$ has the distribution of $\theta_1 \mathbf{1}_n + (1 - \theta_1)(X_2 \mathbf{w}^2 + \dots + X_n \mathbf{w}^n)$, where $\theta_1 \sim \text{beta}(\cdot; \alpha, \beta)$, and $(X_2, \dots, X_n)^T$ is Dirichlet distributed $\text{Dir}(n; 1, 1, \dots, 1)$.
- (ii) a variable θ distributed according to $\pi_{\alpha_1, \beta_1, \alpha_2, \beta_2}^{\text{I,F}}$ has the distribution of $\theta_1 \mathbf{1}_n + (\theta_n - \theta_1)(X_2 \mathbf{w}^2 + \dots + X_n \mathbf{w}^n)$, where $(\theta_1, \theta_n)^T$ has marginal density proportional to $I_{\theta_1 \leq \theta_n}((\theta_1, \theta_n)^T) \text{beta}(\theta_1; \alpha_1, \beta_1) \text{beta}(\theta_n; \alpha_n, \beta_n)$, $X_n = 1 - X_2 - \dots - X_{n-1}$ and $(X_2, \dots, X_{n-1})^T$ is Dirichlet distributed $\text{Dir}(n - 1; 1, 1, \dots, 1)$.

The proof is given in the Appendix.

Based on Theorem 4, and in analogy with Definition 1, it is natural to introduce the following generalizations of (7) and (8):

Definition 2.

- (i) Let $\gamma = (\gamma_2, \gamma_3, \dots, \gamma_n, \gamma_0)^T$ be an n -dimensional vector of positive, real numbers. We define $\pi_{\alpha, \beta, \gamma}^{\text{I}}(\theta)$ to be the density for a variable of the form $\theta_1 \mathbf{1}_n + (1 - \theta_1)(X_2 \mathbf{w}^2 + \dots + X_n \mathbf{w}^n)$, where $\theta_1 \sim \text{beta}(\cdot; \alpha, \beta)$, and vector $(X_2, \dots, X_n)^T$ is Dirichlet distributed $\text{Dir}(n; \gamma_2, \gamma_3, \dots, \gamma_n, \gamma_0)$, independently of θ_1 .

(ii) Let $\boldsymbol{\gamma} = (\gamma_2, \gamma_3, \dots, \gamma_n)^T$ be an $n - 1$ -dimensional vector of positive, real numbers. We define $\pi_{\alpha_1, \beta_1, \alpha_n, \beta_n, \boldsymbol{\gamma}}^{1, F}(\boldsymbol{\theta})$ to be the density for a variable of the form $\theta_1 \mathbf{1}_n + (\theta_n - \theta_1)(X_2 \mathbf{w}^2 + \dots + X_n \mathbf{w}^n)$, where the joint density for $(\theta_1, \theta_n)^T$ is proportional to $I_{\theta_1 \leq \theta_n}((\theta_1, \theta_n)^T) \text{beta}(\theta_1; \alpha_1, \beta_1) \text{beta}(\theta_n; \alpha_n, \beta_n)$, $X_n = 1 - X_2 - \dots - X_{n-1}$ and $(X_2, \dots, X_{n-1})^T$ is Dirichlet distributed $\text{Dir}(n - 1; \gamma_2, \gamma_3, \dots, \gamma_n)$, independently of θ_1 and θ_n .

5 | MARGINAL PRIOR EXPECTATIONS AND VARIANCES FOR MORTALITY RATE PARAMETERS

We think it is fair to claim that the wish for scale neutrality for the initial and final mortality rates are met by choosing the respective α and β -parameters of Definitions 1 and 2 close to 0. In this section, we will discuss how the parameters γ_i of these definitions, describing the conditional distributions for the remaining mortality rates, should be chosen in order to obtain a noninformative and approximately scale neutral modeling also for the shape of the mortality rate curve. For simplicity, we will limit attention to the priors discussed in Section 4.1. As a byproduct, we will also derive the limiting behavior of the uniform distribution on J_n , see Proposition 2 and Corollary 1. These results confirm that this distribution is far from scale neutral. We start by considering the expected shape of the mortality rate curve.

5.1 | Prior expectation of the mortality rate curve

For convenience, we first assume that the prior distribution (5) is chosen. By Theorem 3, this means that each γ_i of Definition 1 equals 1. We are interested in the effect on the expected shape of the mortality rate curve of changing the scale by increasing n . A standardized representation of the mortality rate curve is given by the vector $\boldsymbol{\phi}$ defined in (6), shown in Theorem 3 to follow the required uniform distribution on K_n when the vector $(X_1, \dots, X_{n-1})^T$ follows the $\text{Dir}(n; 1, \dots, 1)$ distribution. This standardization simplifies the comparison of the curves for different values of n . To emphasize the dependence on the number n of measuring points, we denote the components ϕ_t of $\boldsymbol{\phi}$ defined in (6) by $\phi_t^n, t = 1, \dots, n - 1$. In addition, we define $\phi_n^n = 1$.

Now, for any positive real number r , let $[r]$ be the smallest integer for which $[r] \geq r$. Then $\phi_{[nr]}^n$ is the standardized mortality rate ϕ_t^n for an index t which approximately satisfies $t/n = \tau$. For any integer n define the step function $f_n(\tau) = E(\phi_{[nr]}^n), \tau \in (0, 1]$. Note that by Lemma 1 the t th row of the matrix W^n is given by

$$\begin{aligned} & (w_{t,1}, w_{t,2}, w_{t,3}, \dots, w_{t,t}, w_{t,t+1}, \dots, w_{t,n}) \\ & = (1, (t - 1)/(n - 1), (t - 2)/(n - 2), \dots, 1/(n - t + 1), 0, \dots, 0). \end{aligned} \tag{9}$$

Hence, it follows from (6), (9) and the fact that $E(X_i) = 1/n$ for every i , that

$$\begin{aligned} f_n(\tau) &= (1/n) \left(1 + \sum_{i=1}^{[nr]-1} ((nr) - i)/(n - i) \right) \\ &= (1/n) + \sum_{i=1}^{[nr]-1} (1/n) (([nr]/n) - (i/n))/(1 - (i/n)). \end{aligned} \tag{10}$$

The sum in (10) is an approximation to the integral

$$\int_0^\tau (\tau - x)/(1 - x) dx = (\tau - 1) \int_0^\tau 1/(1 - x) dx + \int_0^\tau dx = (1 - \tau) \log(1 - \tau) + \tau.$$

Hence, as n increases the step functions $f_n(\tau)$ concentrate around the curve

$$\lim_{n \rightarrow \infty} f_n(\tau) = f(\tau) \stackrel{\text{def}}{=} (1 - \tau) \log(1 - \tau) + \tau. \quad (11)$$

We may define the random vector ϕ^n through (6) with (X_1, \dots, X_{n-1}) Dirichlet distributed with any Dirichlet parameter vector γ used in Definition 1. Since $E(X_i) = 1/n$ also under the $\text{Dir}(n; \gamma, \dots, \gamma)$ -distribution for any value of $\gamma > 0$, the argument leading to (11) also applies to the prior distributions of Definition 1, if all components of γ are identical. Hence, rescaling essentially does not influence this part of the prior modeling significantly. The result is also valid for the original uniform prior on θ of Gelman et al. (1996). This can be proved by a slightly modified argument involving the extra extreme point vector 0, as well as a Dirichlet distribution with marginal expectations $1/(n+1)$ rather than $1/n$. Hence, we have the following proposition:

Proposition 2. *Under the uniform prior distribution on J_n , it holds that for any $\tau \in (0, 1]$ we have $\lim_{n \rightarrow \infty} E(\theta_{[n\tau]}) = f(\tau) = (1 - \tau) \log(1 - \tau) + \tau$.*

Interestingly, Gelman (1996) claims that, in our notation, the limiting curve is instead the quadratic τ^2 . The argument is that under convexity, the differences $\phi_i^n - \phi_{i-1}^n$ form a positive increasing sequence, and under the uniformity assumption such sequences approach linearity as n increases. Hence, the sequence ϕ_i^n itself should approach a quadratic. The reason that this argument is not valid is that the restriction $(\phi_2^n - \phi_1^n) + \dots + (\phi_n^n - \phi_{n-1}^n) = 1 - \phi_1^n \leq 1$ is not taken into consideration. The vector $(\phi_1^n, (\phi_2^n - \phi_1^n), \dots, (\phi_n^n - \phi_{n-1}^n))$ is a one-to-one linear transformation of $(\phi_1^n, \dots, \phi_n^n)$ and is hence uniformly distributed if this latter vector is assumed to have a uniform distribution. But since the sum of the components is bounded by 1, this distribution is restricted to a proper subset of the set of increasing functions. Hence, the models for this vector and a general uniformly distributed increasing vector are not identical, as asserted in Gelman (1996). In particular, the curve does not necessarily become concentrated around a straight line in the limit.

The perceived quadratic form of the mortality rate curve in Gelman (1996) is taken as a partial explanation for the apparent overestimation of the mortality rates for higher age classes; the form of the curve forcing these mortality rates upwards. To the extent that such an effect is important, it is even stronger because of the logarithmic shape of $f(\tau)$, having the derivative $-\log(1 - \tau)$ and the second derivative $1/(1 - \tau)$. However, we doubt that there exists a “neutral” shape, ideal for being updated by data toward a true mortality rate curve irrespective of the true shape of this curve. But if such a shape is believed to exist, one could obtain an a priori expected curve of this shape by changing the parameters of the Dirichlet distribution for the X_i . The parameters should be chosen such that the weights $E(X_i)$ are equal to (or proportional to) the components of the vector obtained by applying the inverse of W^n to a vector of discrete approximations to the desired curve. This strategy could also be followed if it is desirable to build into the prior an informative preference for a certain shape, such as the shape of the failure rate of a Gompertz–Makeham distribution.

The overestimation for higher age classes is also partially attributed by Gelman (1996) to small marginal prior variances for the mortality rates with the rather large ($n = 30$) number of measuring points. Undoubtedly, this is a very significant point, and we think choosing Dirichlet parameters giving large enough variances for each ϕ_i^n to allow the data to determine the shape of the posterior expected mortality rate curve, irrespective of the shape of the prior expectation, is likely to be more important than this shape. The limiting behavior of the marginal variances with increasing n is analyzed in the next subsection. How to choose such parameters is discussed in Section 5.3.

5.2 | Prior variances of the standardized mortality rates

In an example discussed in sections 3.1 and 3.2 of Gelman (1996), it is shown that under a uniform prior the marginal variances of parameters representing a gridded approximation to a continuous increasing curve converges to 0 as the distance between the grid points converges to 0. This is undesirable behavior, since it may imply that a fixed data set is unable to pull the curve which has a linear expectation under the uniform prior, toward the true shape if a too fine grid is chosen. In the words of Gelman (1996): “The strength of the prior distribution thus depends on the discretization, with potentially grave consequences.” Based on considering the increasing and convex mortality rate model “as a slight variant of the example of sections 3.1–3.2,” Gelman (1996) in Section 3.3 argues that this model is subject to the same problem. Part of the argument is based on the incorrect expected linear shape of the curve of mortality rate differences mentioned in the previous subsection, but the result is confirmed in Corollary 1 below. The corollary follows from a special case of Theorem 5 below.

Under the prior distributions $\pi_{\alpha, \beta, \gamma, \dots, \gamma}^F$ of Definition 1, the standardized mortality rate variances $v_n(\tau) \stackrel{\text{def}}{=} \text{var}(\phi_{[n\tau]}^n)$, $\tau \in (0, 1]$, can be computed using (6) and (9) in a similar way as in the computation of the expectations $f_n(\tau)$ in (10) of the previous subsection. After carrying out this calculation and going to the limit as $n \rightarrow \infty$, we arrive at the following theorem, where the main point is the existence of a limiting curve rather than its exact analytic form:

Theorem 5. *Under the prior distributions $\pi_{\alpha, \beta, \gamma, \dots, \gamma}^F$ of Definition 1, the standardized mortality rate variances satisfy*

$$\lim_{n \rightarrow \infty} (n\gamma + 1)v_n(\tau) = 2\tau(1 - \tau) + 2(1 - \tau)^2 \log(1 - \tau) - (1 - \tau)^2(\log(1 - \tau))^2.$$

Proof. Noting that $\text{var}(X_i) = (n - 1)/(n^2(n\gamma + 1))$ and, for $i \neq j$, $\text{cov}(X_i, X_j) = -1/(n^2(n\gamma + 1))$ we obtain from (6)

$$v_n(\tau) = ((n - 1)/(n^2(n\gamma + 1))) \left\{ \sum_{i=1}^n w_{[n\tau],i}^2 \right\} + (-1/(n^2(n\gamma + 1))) \left\{ \sum_{i,j=1, i \neq j}^n w_{[n\tau],i} w_{[n\tau],j} \right\}.$$

For convenience we add the sum $\sum_{i=1}^n w_{[n\tau],i}^2/(n^2(n\gamma + 1))$ to the first summand and subtract it from the second. Using (9) and then (10), this yields

$$\begin{aligned}
v_n(\tau) &= (1/(n(\gamma + 1))) \left\{ 1 + \sum_{i=1}^{\lfloor n\tau \rfloor - 1} ((\lfloor n\tau \rfloor - i)/(n - i))^2 \right\} \\
&\quad + (-1/(n^2(\gamma + 1))) \left\{ 1 + \sum_{i=1}^{\lfloor n\tau \rfloor - 1} (\lfloor n\tau \rfloor - i)/(n - i) \right\} \\
&\quad \cdot \left\{ 1 + \sum_{j=1}^{\lfloor n\tau \rfloor - 1} (\lfloor n\tau \rfloor - j)/(n - j) \right\} \\
&= (1/(n\gamma + 1)) \left\{ (1/n) + \sum_{i=1}^{\lfloor n\tau \rfloor - 1} (1/n)((\lfloor n\tau \rfloor/n) - (i/n))/(1 - (i/n))^2 \right\} \\
&\quad - (1/(n\gamma + 1))f_n(\tau)^2, \tag{12}
\end{aligned}$$

where we have used (10) in the last step. The sum in (12) is an approximation to the integral

$$\int_0^\tau (\tau - x)^2/(1 - x)^2 dx = 2\tau - \tau^2 + 2(1 - \tau) \log(1 - \tau).$$

Since the limit of $f_n(\tau)$ as $n \rightarrow \infty$ is $f(\tau)$ given by (11), it follows that

$$\begin{aligned}
\lim_{n \rightarrow \infty} (n\gamma + 1)v_n(\tau) &= 2\tau - \tau^2 + 2(1 - \tau) \log(1 - \tau) - f(\tau)^2 \\
&= 2\tau - \tau^2 + 2(1 - \tau) \log(1 - \tau) - ((1 - \tau) \log(1 - \tau) + \tau)^2 \\
&= 2\tau(1 - \tau) + 2(1 - \tau)^2 \log(1 - \tau) - (1 - \tau)^2 (\log(1 - \tau))^2. \tag{13}
\end{aligned}$$

■

It follows from Theorem 5 that for a fixed $\gamma > 0$ the standardized mortality rate variance $v_n(\tau)$ converges to 0 as $n \rightarrow \infty$. Hence, using a common fixed γ regardless of n is subject to the above-mentioned concern expressed in Gelman (1996). This applies also to the marginal variances for the mortality rate vector θ when distributed according to the uniform distribution on J_n :

Corollary 1. *Let $\tau \in (0, 1)$ be arbitrary. Then under the uniform prior on J_n we have*

$$\lim_{n \rightarrow \infty} \text{var}(\theta_{\lfloor n\tau \rfloor}) = 0.$$

Proof. It is noted just after (5) that $\pi_{\text{GMS}}(\theta) = \pi_{\alpha, \beta}^{\text{F}}(\theta)$ with $\alpha = n, \beta = 1$. By Theorem 3 and Definition 1, this is the same distribution as $\pi_{\alpha, \beta, \gamma}^{\text{F}}(\theta)$ with $\alpha = n, \beta = 1$ and $\gamma = (1, 1, \dots, 1)$. Under this distribution, we have $\theta_{\lfloor n\tau \rfloor} = \theta_n \phi_{\lfloor n\tau \rfloor}^n$, and hence

$$\text{var}(\theta_{\lfloor n\tau \rfloor}) = \text{var}(\theta_n \phi_{\lfloor n\tau \rfloor}^n).$$

Since by Theorem 3 θ_n and $\phi_{\lfloor n\tau \rfloor}^n$ are independent, we obtain by conditioning on θ_n

$$\text{var}(\theta_n \phi_{\lfloor n\tau \rfloor}^n) = E(\theta_n^2) \text{var}(\phi_{\lfloor n\tau \rfloor}^n) + (E(\phi_{\lfloor n\tau \rfloor}^n))^2 \text{var}(\theta_n) \leq \text{var}(\phi_{\lfloor n\tau \rfloor}^n) + \text{var}(\theta_n).$$

The first summand approaches 0 as $n \rightarrow \infty$ by Theorem 5, while the second summand equals $n/(n + 1)^2(n + 2)$ by Theorem 1. This proves the corollary. ■

5.3 | Consequences for the choice of Dirichlet parameters

We knew already from Theorem 1 and the subsequent discussion that the uniform distribution is not satisfactory as a noninformative prior on J_n . The main lesson to learn from Theorem 5 is therefore to guide us in choosing Dirichlet parameters in Definition 1. Consider for instance $\pi_{\alpha,\beta}^F$ defined in (5), which is a special case of Definition 1 with $(\gamma_1, \dots, \gamma_n) = (1, \dots, 1)$. The conclusion of Corollary 1 does not hold when choosing $\alpha = \beta = \epsilon$ close to 0, since by (10), (11) and the above proof we then in the limit have $\text{var}(\theta_{[n\tau]}) \approx (1/4)f(\tau)^2$ for large n . But for the shape of the curve, the distribution (5) becomes increasingly problematic when n increases, regardless of the choice of α and β , due to Theorem 5. Hence, it seems unlikely that uniformity can play a useful role in non-informative prior modeling of convex, increasing mortality rates. By Theorem 5, the only choice of γ for which $v_n(\tau)$ stays bounded away from 0 for every value of n is $\gamma = 0$, corresponding to the improper Dirichlet density proportional to $x_1^{-1}x_2^{-1} \cdots x_n^{-1}$. But for a given value of n we may approximate this arbitrarily well by choosing γ sufficiently small. It would seem natural from the point of view of scale neutrality to choose γ inversely proportional to n . The standardized variance $v_n(\tau)$ then approaches a limit proportional to (13), and is of the same order of magnitude as this limit. Hence, the desired stability of the prior variances of the standardized mortality rates under rescaling can be achieved.

Because the limit of (13) as $\tau \rightarrow 0$ is 0, implying that, under the conditions of Theorem 5 with a fixed value of γ , we have $\lim_{n \rightarrow \infty} \text{var}(\phi_1^n) = 0$, it might be a good idea to choose a parameter vector for the Dirichlet distribution of $(X_1, \dots, X_{n-1})^T$ of the form $(\gamma, \gamma/n, \dots, \gamma/n)$ instead of $(\gamma, \gamma, \dots, \gamma)$. Since by (6) $\phi_1^n = X_1$, this choice would give a beta distribution with parameters γ and $(n-1)\gamma/n$ for ϕ_1^n . This will ensure $\text{var}(\phi_1^n) \approx 1/4$ if γ is small enough, allowing data to adjust the distribution of ϕ_1^n .

If another limiting shape for the prior expectation of the mortality rate curve than (11) is preferred, this can be achieved in our framework by choosing a non-constant vector of Dirichlet parameters, as explained in Section 5.1. Also in this case sensible scaling is important. Fortunately, the argument leading to (13) can be generalized to the situation when the Dirichlet parameters $\gamma_1, \gamma_2, \dots, \gamma_n$ are allowed to be unequal. We conclude this section with a sketch of the proof. Assume that these parameters under different scalings are derived from a common continuous bounded function $g(\tau)$ on the interval $[0, 1]$. For a given value of n , define accordingly scaled parameters $\gamma_i = (1/n)g(i/n), i = 1, \dots, n$. If we require $\int_0^1 g(\tau)d\tau = 1$, we obtain that $\Gamma \stackrel{\text{def}}{=} \sum_{i=1}^n \gamma_i \approx 1$. The case treated in Theorem 5, combined with the requirement that the Dirichlet parameters add up to 1, independently of n , corresponds to $g(\tau) = 1$. We now have $\text{var}(X_i) = \gamma_i(\Gamma - \gamma_i)/(\Gamma^2(\Gamma + 1))$ and $\text{cov}(X_i, X_j) = -\gamma_i\gamma_j/(\Gamma^2(\Gamma + 1))$. These quantities can be incorporated in the expansion of $\text{var}(\phi_{[n\tau]}^n)$ as in (12). When adapting the proof to this situation, the sums corresponding to the sums in (12) converge to integrals involving $g(\tau)$, and the sum Γ corresponds to the quantity $n\gamma$ appearing in Theorem 5. Since requiring g to have unit integral implies that γ is close to 1, the required stability under rescaling of the marginal prior variances of the standardized mortality rates can be obtained also in this case.

6 | APPLICATION TO THE LIFE INSURANCE DATASET

In this paper, we first address a LIN dataset (Broffitt, 1988) on the graduation of mortalities from Table 1. Additionally, this dataset was extended to monthly observed data (LINM) through a

uniform upsampling procedure to allow for monthly data over ages from 35 to 64 resulting in 360 observed time points in LINM instead of 30 in LIN. LINM dataset is available in Table 2 in the Appendix. For LIN and LINM data, we follow Gelman (1996) in assuming that the observed deaths y_t at each age t follow independent binomial distributions with rates equal to the unknown mortality rates θ_t and known population sizes N_t . Further, consistently following Gelman (1996) due to the large population sizes for each age group and due to low empirical death rates, we use the Poisson approximation for mathematical convenience. Thus, the following model is assumed:

$$p(\mathbf{y}|\theta, N_1, \dots, N_n) \propto \prod_{t=1}^n \theta_t^{y_t} \exp(-N_t \theta_t). \quad (14)$$

Then, we performed inference with the uniform $\pi_{\text{GMS}}(\theta)$ prior (Gelman, 1996) specified in Equation (2), as well as the suggested in this paper $\pi_{\alpha, \beta}^{\text{F}}(\theta)$, $\pi_{\alpha, \beta}^{\text{I}}(\theta)$, and $\pi_{\alpha_1, \beta_1, \alpha_n, \beta_n}^{\text{I,F}}(\theta)$ priors from Equations (5), (7), and (8), respectively.

We used Stan (Carpenter et al., 2017) for Bayesian inference. In all of the experiments, 20 parallel chains were run for 100,000 iterations each with a warm-up of 10,000 iterations. Otherwise, the default tuning parameters from Stan (Carpenter et al., 2017) were used.

In Figure 1, we present extensions of figure 2 from Gelman et al. (1996) and figures 2 and 3 from Gelman (1996). Figure 1 presents posterior means under the studied priors as well as 95% credible intervals obtained within the same settings. These figures show that replacing the marginal prior $\text{beta}(\theta_{30}; 30, 1)$ with a noninformative beta prior while retaining the uniform prior for the rest of the parameters results in a drastic improvement of the posterior distribution. On the other hand, a corresponding change in the marginal prior for θ_1 seems to have a negligible effect. The same conclusions with an even stronger improvement are valid for the LINM data as shown in Figure 2.

Further, for both of the addressed datasets, we performed inference with a $\pi_{\alpha_1, \beta_1, \alpha_n, \beta_n}^{\text{I,F}}(\theta)$ prior and different values of $\gamma_2, \dots, \gamma_n$ for the Dirichlet distribution of $(X_2, \dots, X_{n-1})^T$. More

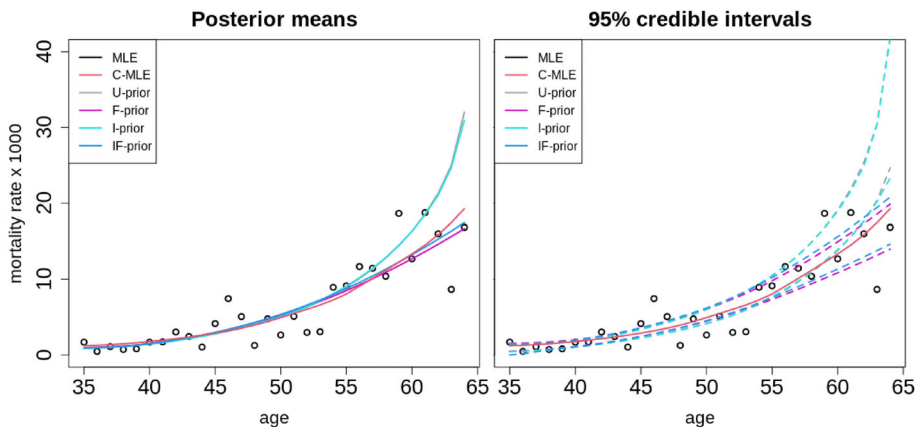


FIGURE 1 Posterior means (left) and credible intervals (right) for the mortality graduation data (LIN) from Broffitt (1988) under $\pi_{\text{GMS}}(\theta)$ (U-prior), $\pi_{\alpha, \beta}^{\text{F}}(\theta)$ (F-prior), $\pi_{\alpha, \beta}^{\text{I}}(\theta)$ (I-prior), and $\pi_{\alpha_1, \beta_1, \alpha_n, \beta_n}^{\text{I,F}}(\theta)$ (IF-prior) priors as well as MLE (dots) and constrained MLE (C-MLE) estimates of mortality rates from Gelman et al. (1996) [Colour figure can be viewed at wileyonlinelibrary.com]

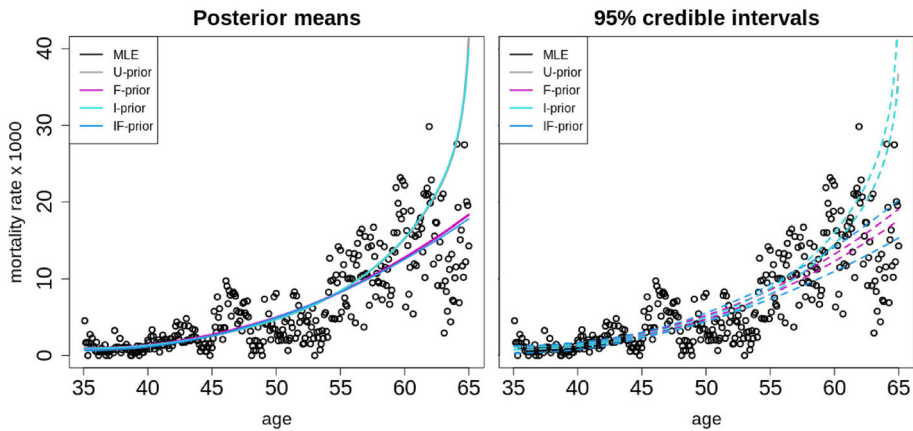


FIGURE 2 Posterior means (left) and credible intervals (right) for the simulated mortality graduation data (LINM) based on Broffitt (1988) under $\pi_{\text{GMS}}(\theta)$ (U-prior), $\pi_{\alpha,\beta}^{\text{F}}(\theta)$ (F-prior), $\pi_{\alpha,\beta}^{\text{I}}(\theta)$ (I-prior), and $\pi_{\alpha_1,\beta_1,\alpha_n,\beta_n}^{\text{I,F}}(\theta)$ (IF-prior) priors as well as MLE (dots) estimates of mortality rates [Colour figure can be viewed at wileyonlinelibrary.com]

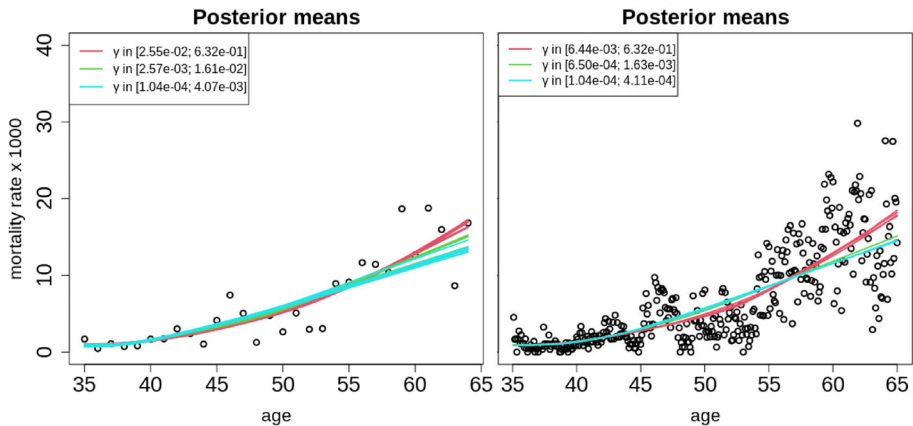


FIGURE 3 Posterior means for the mortality graduation data (LIN) from Broffitt (1988) (left) and the simulated mortality graduation data (LINM) based on Broffitt (1988) (right) under $\pi_{\alpha_1,\beta_1,\alpha_n,\beta_n}^{\text{I,F}}(\theta)$ (IF-prior) priors for different values of $\gamma^{(i)}$ [Colour figure can be viewed at wileyonlinelibrary.com]

specifically, we varied the values of all γ 's within $\pi_{\alpha_1,\beta_1,\alpha_n,\beta_n}^{\text{I,F}}(\theta)$ using the following sequence $\gamma^{(i)} = (1 - \exp(-1))^i$. For LIN, we allowed $i \in \{1, \dots, 20\}$. For LINM, for computational reasons we decided to only use $i \in \{1, 4, 7, 11, 14, 17, 20\}$. The results for different values of $\gamma^{(i)}$ are shown in the left and right panels of Figure 3 for LIN and LINM datasets, respectively. For the LIN dataset from Broffitt (1988), we clearly see three clusters for three ranges of $\gamma^{(i)}$. The general tendency as expected is that for the smaller values of $\gamma^{(i)}$ we are shifting the mortality rates closer to zero. The same conclusions are overall valid for LINM data. Thus, as expected, for both datasets the overall tendency is that the curve becomes flatter and seems to fit the observed data better with smaller values of $\gamma^{(i)}$.

7 | CONCLUSIONS

Results from the analysis of a specific life insurance dataset in Gelman et al. (1996) indicate that a uniform prior on the set of increasing, convex mortality rates is not a satisfactory choice of a noninformative prior distribution on this set. Theorems 1 and 2 of the present paper provide theoretical support for this empirical evidence, and together with Proposition 2 and Corollary 1 indicate an underlying mathematical explanation. By establishing the marginal distributions for the final and the initial mortality rates, and the marginal expectations and variances for parameters characterizing the shape of the mortality rate curve, these results also show that the uniform prior falls short of a requirement for scale neutrality, which is an attractive goal to pursue noninformative priors. Our results are in line with the explanation given in section 3.3 in Gelman (1996), which is that the “prior distribution becomes ever stronger as the scale of the time intervals becomes smaller.” Defining noninformative marginal distributions for the initial and/or final mortality rates directly, uniformity could still potentially be used in the conditional distribution for the rest of the mortality rates. Alternative characterizations of such distributions given in Theorems 3 and 4 motivate a whole new class of priors described in Definitions 1 and 2 through the choice of parameters for Dirichlet distributions. The uniform distribution corresponds to all the Dirichlet parameters being 1 and as a consequence their sum equals the number n of measuring points. Theorem 5, as well as the numerical results from Section 6, indicate that such a uniform conditional distribution on $\theta_{1:n-1}$ given θ_n is still not satisfactory from the point of view of scale neutrality. The subsequent discussion suggests that the desired scale neutrality may be obtained by instead requiring the sum of the Dirichlet parameters to stay constant, regardless of the number of measuring points. Hence, the prior distribution has to be changed according to the number of measuring points. This is in accordance with the goal expressed in Gelman (1996), Section 5: “Instead of requiring that a single probability model be invariant under scaling ... we demand a family of models, indexed by scale, that are mutually consistent.” Alternatively, the improper Dirichlet distribution with all parameters equal to 0 might serve as a single default option suitable for any number of measuring points.

ACKNOWLEDGMENTS

We are very grateful to professor Ida Scheel and two referees for very valuable advice and many helpful suggestions.

ORCID

Jørund Gåsemyr  <https://orcid.org/0000-0002-0561-9406>

Aliaksandr Hubin  <https://orcid.org/0000-0002-3244-6571>

REFERENCES

- Bowers, N. L., Gerber, H. U., & Hickman, J. C. (1986). *Actuarial mathematics* (1st ed.). *The Society of Actuaries*.
- Broffitt, J. D. (1988). Increasing and increasing convex Bayesian graduation (with discussion). *Transactions of the Society of Actuaries*, 40, 115–148.
- Carlin, B. P. (1993). A simple Monte Carlo approach to Bayesian graduation. *Transactions of the Society of Actuaries*, 44, 55–76.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32.
- Gelman, A. (1996). Bayesian model building by pure thought. Some principles and examples. *Statistica Sinica*, 6, 215–232.
- Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion and rejoinder). *Statistica Sinica*, 6, 733–807.

Hjort, N. L., Dahl, F. A., & Steinbakk, G. H. (2006). Post-processing posterior predictive p -values. *Journal of the American Statistical Association*, 101, 1157–1174.

Johnson, V. (2007). Bayesian model assessment using pivotal quantities. *Bayesian Analysis*, 2, 719–734.

How to cite this article: Gåsemyr, J., & Hubin, A. (2022). Prior distributions expressing ignorance about convex increasing failure rates. *Scandinavian Journal of Statistics*, 1–25. <https://doi.org/10.1111/sjos.12588>

APPENDIX

Proof of Theorem 2. Similarly to (3), define

$$L_n = \left\{ \boldsymbol{\theta}_{2:n} \in R^{n-1} : (0, \boldsymbol{\theta}_{2:n}^T)^T \in J_n \right\}.$$

Using that $\theta_1 \leq \theta_t \leq 1$ implies that $0 \leq \theta_t - \theta_1 \leq 1 - \theta_1, t = 2, \dots, n$, it is not too hard to see that $\boldsymbol{\theta} \in J_n$ if and only if $\boldsymbol{\theta} = \mathbf{1}_n$ or $0 \leq \theta_1 < 1$ and $(1/(1 - \theta_1))(\boldsymbol{\theta}_{2:n} - \theta_1 \mathbf{1}_{n-1}) \in L_n$, or equivalently, $0 \leq \theta_1 \leq 1$ and $\boldsymbol{\theta}_{2:n} \in (\theta_1 \mathbf{1}_{n-1}) + (1 - \theta_1)L_n$. Hence, calculating the volume of the set of $\boldsymbol{\theta} \in J_n$ such that $\theta_1 \leq x$ by integrating along the θ_1 -axis, we obtain

$$\begin{aligned} m_n(\{\boldsymbol{\theta} \in J_n : \theta_1 \leq x\}) &= \int_0^x m_{n-1}(\{\boldsymbol{\theta}_{2:n} : (\theta_1, \boldsymbol{\theta}_{2:n}^T)^T \in J_n\})d\theta_1 \\ &= \int_0^x m_{n-1}((1 - \theta_1)L_n)d\theta_1 \\ &= m_{n-1}(L_n) \int_0^x (1 - \theta_1)^{n-1}d\theta_1 \\ &= (m_{n-1}(L_n)/n)(1 - (1 - x)^n). \end{aligned}$$

By the uniformity assumption, the probability under (2) that $\theta_1 \leq x$ is proportional to this volume, and it follows that the density for θ_1 under (2) is proportional to $(1 - \theta_1)^{n-1}$, proving the theorem. ■

Proof of Lemma 2. For both sets, the fact that the given vectors are extreme points are proved by the same argument as in the proof of Lemma 1. Let $\mathbf{u} = (u_1, \dots, u_n)^T$ be an arbitrary element of \bar{L}_n . By adding $1 - u_n$ to each coordinate of \mathbf{u} we obtain a vector belonging to the set \bar{K}_n of Lemma 1. Formally, $\mathbf{u} + (1 - u_n)\mathbf{1}_n \in \bar{K}_n$. This vector can be written uniquely as a convex combination of the extreme points given in Lemma 1, that is,

$$\mathbf{u} + (1 - u_n)\mathbf{1}_n = a_1 \mathbf{w}^1 + \dots + a_n \mathbf{w}^n.$$

Since $\mathbf{w}^1 = \mathbf{1}_n$, this implies $\mathbf{u} = (a_1 - 1 + u_n)\mathbf{1}_n + a_2 \mathbf{w}^2 + \dots + a_n \mathbf{w}^n$. Since $u_1 = 0$, we must have $a_1 = 1 - u_n$, and hence $\mathbf{u} = a_2 \mathbf{w}^2 + \dots + a_n \mathbf{w}^n = a_0 \mathbf{w}^0 + a_2 \mathbf{w}^2 + \dots + a_n \mathbf{w}^n$, where $a_0 = 1 - (a_2 + \dots + a_n) \geq 0$. This proves the first part of Lemma 2. Now suppose instead that $\mathbf{u} \in \bar{M}_n$. Then \mathbf{u} is also a member of \bar{K}_n , and has a unique expression as a convex combination $\mathbf{u} = a_1 \mathbf{w}^1 + \dots + a_n \mathbf{w}^n$ by Lemma 1. Since \mathbf{u} is also a member of \bar{L}_n , and since $\mathbf{w}^1 = \mathbf{1}_n$, we must have $a_1 = 0$, and the assertion follows. ■

TABLE 2 Simulated mortality graduation data (LINM) based on Broffitt (1988)

Age, t	N_t	y_t	Age, t	N_t	y_t	Age, t	N_t	y_t	Age, t	N_t	y_t	Age, t	N_t	y_t	Age, t	N_t	y_t
35.00	1764	8	40.00	2368	4	45.00	1806	4	50.00	1528	4	55.00	1237	7	60.00	949	6
35.08	1800	3	40.08	2383	2	45.08	1912	11	50.08	1555	3	55.08	1230	13	60.08	920	15
35.17	1768	3	40.17	2415	6	45.17	1890	7	50.17	1494	4	55.17	1181	14	60.17	976	14
35.25	1798	0	40.25	2390	8	45.25	1988	6	50.25	1481	4	55.25	1198	9	60.25	1002	8
35.33	1759	3	40.33	2389	6	45.33	1962	15	50.33	1511	3	55.33	1199	8	60.33	957	11
35.42	1817	5	40.42	2350	2	45.42	2012	10	50.42	1490	5	55.42	1242	8	60.42	955	18
35.50	1820	2	40.50	2364	3	45.50	1966	12	50.50	1485	4	55.50	1197	12	60.50	951	15
35.58	1778	4	40.58	2387	4	45.58	1942	7	50.58	1594	7	55.58	1196	16	60.58	923	5
35.67	1752	2	40.67	2356	3	45.67	1905	6	50.67	1495	3	55.67	1193	12	60.67	920	11
35.75	1721	2	40.75	2444	5	45.75	1915	4	50.75	1468	7	55.75	1189	6	60.75	900	14
35.83	1729	1	40.83	2278	1	45.83	1929	7	50.83	1554	1	55.83	1200	12	60.83	937	11
35.92	1752	3	40.92	2292	4	45.92	1945	7	50.92	1537	3	55.92	1192	15	60.92	950	16
36.00	2171	1	41.00	2272	3	46.00	1745	17	51.00	1327	3	56.00	1128	15	61.00	887	14
36.08	2071	0	41.08	2257	6	46.08	1769	16	51.08	1402	6	56.08	1175	5	61.08	804	14
36.17	2122	1	41.17	2258	3	46.17	1778	12	51.17	1373	5	56.17	1103	13	61.17	836	15
36.25	2064	1	41.25	2338	5	46.25	1770	15	51.25	1401	11	56.25	1093	16	61.25	845	14
36.33	2099	2	41.33	2350	4	46.33	1780	14	51.33	1326	11	56.33	1071	14	61.33	888	12
36.42	2080	1	41.42	2418	5	46.42	1756	9	51.42	1340	3	56.42	1125	18	61.42	855	18
36.50	2127	1	41.50	2334	3	46.50	1723	14	51.50	1416	11	56.50	1069	11	61.50	885	9
36.58	2152	3	41.58	2265	2	46.58	1811	15	51.58	1370	4	56.58	1092	18	61.58	863	18
36.67	2062	0	41.67	2297	5	46.67	1739	14	51.67	1380	9	56.67	1146	12	61.67	872	19
36.75	2171	1	41.75	2299	5	46.75	1678	13	51.75	1393	6	56.75	1079	4	61.75	857	18
36.83	2211	1	41.83	2359	3	46.83	1712	10	51.83	1363	7	56.83	1159	18	61.83	838	25

(Continues)

TABLE 2 (Continued)

Age, t	N _t	y _t	Age, t	N _t	y _t	Age, t	N _t	y _t	Age, t	N _t	y _t	Age, t	N _t	y _t	Age, t	N _t	y _t
36.92	2188	0	41.92	2273	4	46.92	1697	7	51.92	1367	8	56.92	1122	12	61.92	806	16
37.00	2757	2	42.00	2338	6	47.00	1594	9	52.00	1357	2	57.00	1045	11	62.00	743	17
37.08	2726	2	42.08	2345	9	47.08	1553	7	52.08	1282	2	57.08	1063	15	62.08	678	14
37.17	2678	2	42.17	2288	10	47.17	1599	9	52.17	1327	3	57.17	1068	7	62.17	755	9
37.25	2722	1	42.25	2269	6	47.25	1558	5	52.25	1329	6	57.25	1037	16	62.25	770	12
37.33	2763	9	42.33	2322	5	47.33	1614	11	52.33	1376	6	57.33	1049	15	62.33	749	13
37.42	2777	3	42.42	2239	9	47.42	1573	11	52.42	1340	7	57.42	1054	10	62.42	752	13
37.50	2694	2	42.50	2313	7	47.50	1625	9	52.50	1462	4	57.50	1017	17	62.50	767	7
37.58	2779	3	42.58	2363	4	47.58	1537	8	52.58	1326	2	57.58	1084	14	62.58	783	10
37.67	2813	3	42.67	2281	4	47.67	1549	11	52.67	1352	5	57.67	1044	8	62.67	742	11
37.75	2702	6	42.75	2316	9	47.75	1594	8	52.75	1298	8	57.75	1027	11	62.75	780	16
37.83	2812	1	42.83	2320	4	47.83	1585	3	52.83	1354	3	57.83	1034	10	62.83	768	7
37.92	2699	2	42.92	2284	11	47.92	1579	5	52.92	1313	0	57.92	1054	10	62.92	713	15
38.00	2802	3	43.00	1971	4	48.00	1536	0	53.00	1318	6	58.00	1160	17	63.00	683	2
38.08	2815	3	43.08	2079	8	48.08	1633	2	53.08	1316	4	58.08	1198	14	63.08	720	7
38.17	2802	3	43.17	2015	4	48.17	1539	1	53.17	1307	3	58.17	1100	8	63.17	719	6
38.25	2752	4	43.25	2054	4	48.25	1535	3	53.25	1286	4	58.25	1176	11	63.25	694	4
38.33	2664	2	43.33	2105	5	48.33	1582	1	53.33	1385	1	58.33	1127	16	63.33	635	8
38.42	2784	3	43.42	2035	7	48.42	1591	3	53.42	1339	2	58.42	1130	13	63.42	713	8
38.50	2722	2	43.50	2036	5	48.50	1622	6	53.50	1238	4	58.50	1095	10	63.50	684	3
38.58	2725	0	43.58	2102	7	48.58	1606	2	53.58	1349	3	58.58	1156	13	63.58	681	9
38.67	2774	0	43.67	2118	5	48.67	1522	2	53.67	1316	7	58.67	1194	15	63.67	693	5
38.75	2802	1	43.75	2042	4	48.75	1543	1	53.75	1252	7	58.75	1199	8	63.75	711	5

(Continues)

TABLE 2 (Continued)

Age, t	N_t	y_t	Age, t	N_t	y_t	Age, t	N_t	y_t	Age, t	N_t	y_t	Age, t	N_t	y_t	Age, t	N_t	y_t
38.83	2721	2	43.83	2065	4	48.83	1589	0	53.83	1275	4	58.83	1152	9	63.83	691	7
38.92	2829	1	43.92	2092	3	48.92	1662	3	53.92	1267	3	58.92	1173	10	63.92	692	8
39.00	2556	1	44.00	1834	1	49.00	1498	11	54.00	1211	10	59.00	1022	14	64.00	617	17
39.08	2454	3	44.08	1933	2	49.08	1469	6	54.08	1182	15	59.08	998	16	64.08	622	12
39.17	2358	1	44.17	1922	4	49.17	1467	6	54.17	1232	12	59.17	1038	14	64.17	581	4
39.25	2388	0	44.25	1971	2	49.25	1454	6	54.25	1245	6	59.25	1006	22	64.25	596	9
39.33	2509	3	44.33	1966	3	49.33	1472	8	54.33	1216	12	59.33	998	16	64.33	611	10
39.42	2546	2	44.42	1906	1	49.42	1532	5	54.42	1251	6	59.42	1009	20	64.42	599	7
39.50	2450	1	44.50	1916	1	49.50	1458	3	54.50	1267	6	59.50	1033	19	64.50	588	6
39.58	2500	1	44.58	1866	2	49.58	1442	6	54.58	1236	13	59.58	1036	24	64.58	619	17
39.67	2451	2	44.67	1948	3	49.67	1442	5	54.67	1264	12	59.67	1012	18	64.67	573	7
39.75	2457	3	44.75	1873	0	49.75	1438	11	54.75	1268	18	59.75	967	22	64.75	599	12
39.83	2373	2	44.83	1884	3	49.83	1457	5	54.83	1254	6	59.83	1066	20	64.83	563	11
39.92	2514	5	44.92	1985	2	49.92	1481	12	54.92	1164	16	59.92	1037	23	64.92	560	8

Proof of Theorem 4. By Lemma 2, part (i), vectors of the form $X_2\mathbf{w}^2 + \dots + X_n\mathbf{w}^n$ span the set \overline{L}_n . Let $W_{-1;-1}^n$ be the matrix formed by deleting the first row and first column of the matrix W^n . Define the $n - 1$ -dimensional vector variable

$$\boldsymbol{\phi} = W_{-1;-1}^n(X_2, \dots, x_n)^T = X_2\mathbf{w}_{2:n}^2 + \dots + X_n\mathbf{w}_{2:n}^n. \tag{A1}$$

Since all vectors $\mathbf{w}^2, \dots, \mathbf{w}^n$ have 0 as their first entry, it follows that $X_2\mathbf{w}^2 + \dots + X_n\mathbf{w}^n = (0, \boldsymbol{\phi}^T)^T$. Consequently, by Lemma 2, part (i), vectors of the form $(0, \boldsymbol{\phi}^T)^T$ span the set \overline{L}_n , and hence the vectors $\boldsymbol{\phi}$ of (14) span L_n . It follows as in the proof of Theorem 3 that $\boldsymbol{\phi}$ is uniformly distributed on L_n . Hence, scaling and translating $\boldsymbol{\phi}$, part (i) follows from (7). To prove part (ii), we first note that by Lemma 2, part (ii), the set \overline{M}_n is spanned by vectors of the form

$$X_2\mathbf{w}^2 + \dots + X_n\mathbf{w}^n = (X_2(\mathbf{w}^2 - \mathbf{w}^n) + \dots + X_{n-1}(\mathbf{w}^{n-1} - \mathbf{w}^n)) + \mathbf{w}^n.$$

Recall that $\mathbf{w}^n = (0, \dots, 0, 1)^T$ and that all the vectors $\mathbf{w}^2, \dots, \mathbf{w}^{n-1}$ have 0 as their first and 1 as their last entry. Hence, for every $i = 2, \dots, n - 1$ we have that $\mathbf{w}^i - \mathbf{w}^n = (0, (\mathbf{w}_{2:n-1}^i)^T, 0)^T$. Let $W_{-1,-n;-1,-n}^n$ be the matrix formed by deleting the first and last row and the first and last column from W^n , that is, the matrix formed by the vectors $\mathbf{w}_{2:n-1}^i, i = 2, \dots, n - 1$. Define the $n - 2$ -dimensional vector variable

$$\boldsymbol{\phi} = W_{-1,-n;-1,-n}^n(X_2, \dots, X_{n-1})^T = X_2\mathbf{w}_{2:n-1}^2 + \dots + X_{n-1}\mathbf{w}_{2:n-1}^{n-1}. \tag{A2}$$

From the above, it follows that $X_2\mathbf{w}^2 + \dots + X_n\mathbf{w}^n = ((0, \boldsymbol{\phi}^T, 0)^T + \mathbf{w}^n) = ((0, \boldsymbol{\phi}^T, 1)^T)$. Since vectors of this form span \overline{M}_n , it follows that vectors $\boldsymbol{\phi}$ of the form (A1) span the set M_n . Uniformity of $\boldsymbol{\phi}$ on this set is proved as before, and the proof is completed by scaling and translation as in the proof of part (i). ■