

En modell for Ca^{2+} hendelser i astrocytter

Eivind Heggelund Tørstad

Masteroppgave, våren 2023



Denne masteroppgaven er levert inn under masterprogrammet *Stokastisk modellering, statistikk og risikoanalyse*, studieretning *Statistikk*, ved Matematisk institutt, Universitetet i Oslo. Oppgaven er normert til 60 studiepoeng.

Forsiden viser et utsnitt av rotsystemet til den eksepsjonelle liegruppen E_8 , projisert ned i planet. Liegrupper ble oppfunnet av den norske matematikeren Sophus Lie (1842–1899) for å uttrykke symmetriene til differensiallikninger og spiller i dag en sentral rolle i flere deler av matematikken.

Abstract

Astrocytes play an important part in the central nervous system, but their role and structure are not yet fully understood. Previous studies of cultured astrocytes and astrocytes in brain slices have concluded that the sizes of Ca^{2+} events, the main signalling pathway of astrocytes, are power-law distributed. This thesis argues that a more comprehensive modeling approach is necessary and develops a variant of a gamma process model fitted on *in-vivo* measurements of Ca^{2+} . Though the model type fits well, it is difficult to estimate the importance of this modeling approach, both due to the idiosyncratic nature of Ca^{2+} measurement methods and the lack of any comparable efforts in statistical modeling of Ca^{2+} events. Doing inference within our framework we find no evidence of what we call a relative refractory period after Ca^{2+} events in astrocytes, though we do find a robust positive relation between time since the last Ca^{2+} event and the size of the next. We also find weak evidence that supports a negative relation between the size of a Ca^{2+} event and the size of the following Ca^{2+} event.

Acknowledgements

I løpet av denne oppgaven har jeg fått stor hjelp av mine veiledere Celine Cunen og Gudmund Hermansen, som har hjulpet meg kontinuerlig gjennom hele prosessen og gitt meg mer faglig og emosjonell støtte enn hva man kan forvente av vanlige veiledere. Jeg vil også takk veileder Rune Enger og miljøet på GliaLab for tilgang til data og en introduuksjon til astrocyttenes fantastiske verden.

Innhold

Abstract	i
Acknowledgements	ii
Innhold	iii
Figurer	iv
Tabeller	v
1 Introduksjon	1
1.1 Astrocytter	1
1.2 Refraktærperiode	3
1.3 Størrelse av Ca ²⁺ hendelser	4
1.4 Gammaprocess	7
2 Data	8
2.1 Måling av Ca ²⁺ aktivitet og Ca ²⁺ hendelser	8
2.2 ROA-algoritmen	9
2.3 Hva er en Ca ²⁺ hendelse	10
3 Gamma-prosess	14
3.1 intro til gamma prosess	14
3.2 Transformert fordeling	17
4 Modellevaluering	22
4.1 OLS-modell	23
4.2 Enkel gammaprocess	26
4.3 Trunkert gammaprocess	29
5 Diskusjon	41
Appendices	44
A The First Appendix	45
B The Second Appendix	46

Figurer

1.1 Bilde av astrocytt, fra Gerry Shaw , CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=29369565	2
1.2 Estimert parameter og observert data for en enkel regresjonsmodell: $Y_i = \beta_0 + \beta X_i$. Y_i er størrelse på den i -te Ca2+ hendelsen og X_i er tid til neste Ca2+ hendelse som er større enn Y_i . For nærmere definisjon av variabler se kap 2.	4
1.3 Figur fra Jung et al. (1998). S er størrelsen av en Ca2+ hendelse og $p(s)$ er frekvens av hendelse. Fylte trekant er observert data og tomme kvadrater er simulert data.	5
1.4 Noe om figuren her	6
2.1 Grafisk fremstilling av prosessering av ROA-algoritmen [1]	10
2.2 Hvordan størrelse og tid blir målt. Den blå striplede linja er terskelverdien, y og t er de observerte størrelsene og tidsverdiene.	11
2.3 To forskjellige ROA tracer med forskjellige maksverdier, men lik oppførsel. ROA tracer fra to ulike -mus.	12
3.1 Sannsynlighetstetthetene til en vanlig gammafordeling ($f_Z(z)$) og den transformerte fordelingen ($f_Y(y)$) evaluert for $y \in [0.01, 0.99]$ og $z = -\log(1 - y)$	18
3.2 Simulerte gammaprosesser og transformasjon av de simulerte gammaprosessene. Tidssteget er $dt = 1$ og $t = 1, 2, 3, \dots, 100$. De transformerte gammaprosessene er $Y(t) = 1 - \exp(Z(t))$	19
3.3	21
3.4	21
4.1 Plot av observert størrelse til hendelse opp mot tid siden forrige hendelse i sort. Tilpassede verdier i blå.	23
4.2 Observerte størrelser av Ca2+ hendelser i sort og tilpasset størrelse av hendelsene, kontrollert for hjulfart i blått	24
4.3 a) Histogram av OLS residualer og b) sammenhengen mellom residualer og størrelse på Ca2+ hendelse	25
4.4 Observerte størrelser av Ca2+ hendelser i sort og tilpassede størrelser fra OLS modell med logtransformerte kovariater i blått	26
4.5 Observerte størrelser av Ca2+ hendelser (sort) og tilpassede størrelser (blått)	27

4.6	Observerte størrelser av Ca ²⁺ hendelser (sort) og tilpassede størrelser fra modell 1.1 (blått)	28
4.7	Hvert plott viser observerte data i sort, og simulert størrelse gitt kovariater i rødt	29
4.8	Tilpassede størrelsesverdier fra modell 2.1.1 i blått	31
4.9	Tilpassede størrelsesverdier fra modell 2.1.2 i blått	32
4.10	Tilpassede størrelsesverdier fra modell 2.2.1 i blått	34
4.11	Tilpassede størrelsesverdier fra modell 2.2.2 i blått	35
4.12	Tilpassede verdier for størrelse fra modell 2.3.1 i blått	36
4.13	Tilpassede størrelsesverdier fra modell 2.3.2 i blått	37
4.14	Tilpassede verdier for størrelse fra modell 2.3.3 i blått	38
4.15	Tilpassede verdier for størrelse fra modell 3.0 i blått	39
4.16	Tilpassede verdier for størrelse fra modell 3.1 i blått	40

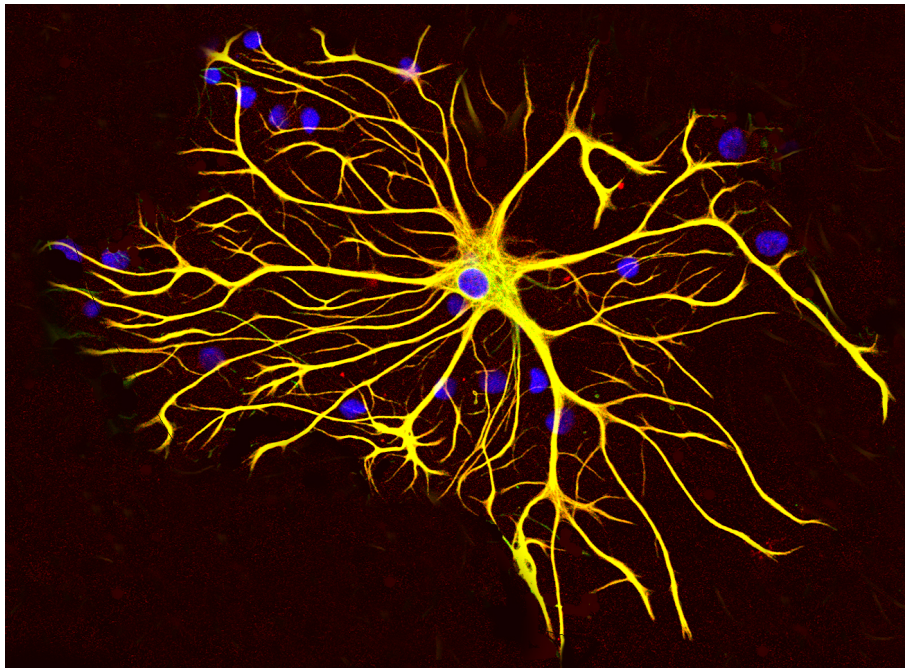
Tabeller

KAPITTEL 1

Introduksjon

1.1 Astrocytter

Nervesystemet består hovedsaklig av to typer celler: nevroner og gliaceller. Gliaceller finnes over hele nervesystemet, det er et stort antall av dem og det finnes noen forskjellige subtyper. Astrocytter er en subtype av gliaceller, som man finner i sentralnervesystemet og astrocytter finner man flest av i hjernen. Astrocytter spiller en rolle i mange funksjoner i hjernen og sentralnervesystemet generelt. Fra kjernen av astrocytten (soma) sprer det seg ut lange armer (endeføtter). Endeføttene til astrocyttene ligger rundt viktige strukturer i hjernen, spesielt blodvener/årer som går gjennom hjernen og synapser, koblinger mellom nevroner.



Figur 1.1: Bilde av astrocytt, fra Gerry Shaw , CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=29369565>

Den viktigste signalmåten for celler i hjernen er gjennom Ca^{2+} signaler, slik er det også i astrocytter. Innad i astrocytter heves og senkes nivået av Ca^{2+} ioner. Heving av Ca^{2+} nivået innad i cellene, som kalles en Ca^{2+} hendelse, korrelerer med andre viktige funksjonene til astrocytter. Det er gjennom analyse av Ca^{2+} hendelser i astrocytter man har påvist hvordan astrocytter påvirker andre funksjoner i hjernen, samt hvordan astrocytter blir påvirket av annen stimuli.

Ca^{2+} nivået i en astrocytt måles ved at Ca^{2+} flurosensindikatorer binder seg til Ca^{2+} molekyler i cellen og at man gjennom to-foton mikroskopi måler flurosensnivået i en spesifikk celle, en del av en celle eller over en samling av celler. Ca^{2+} flurosensindikatorer påvirker hvordan Ca^{2+} molekyler oppfører seg og hindrer til en viss grad Ca^{2+} molekyler å binde seg til Ca^{2+} reseptorene i astrocytten. Det er altså ikke sikkert at Ca^{2+} nivået man måler i cellen gjenspeiler det Ca^{2+} nivået som hadde vært hvis ikke flurosensindikatorene hadde vært der. I tillegg er det et åpent spørsmål hvorvidt stimuli utenfor astrocytten påvirker sannsynligheten for at det oppstår en heving av Ca^{2+} nivå, eller om stimuli påvirker dynamikken til Ca^{2+} hevingen, f.eks. hvor raskt eller i hvilken retning Ca^{2+} aktiviteten sprer seg gjennom cellen. [10]

Det eksisterer per i dag flere ulike måter å måle Ca^{2+} nivået i astrocytter, ulike måter å modellere den romlige og temporale dynamikken i hevinger av Ca^{2+} nivået og heller ingen robust måte å definere hva en Ca^{2+} hendelse er. I denne oppgaven brukes det data som er innhentet fra en metode som måler andelen

av astrocytten som er "aktiv", der det observeres Ca^{2+} aktivitet, over et 2D plan. En Ca^{2+} hendelse blir slik vi definerer det en observert, sammenhengende heving av Ca^{2+} nivået over en gitt terskel. Dette diskuteres i kapittel 2.

1.2 Refraktærperiode

Knytte modellering av aksjonspotensiale i nevroner til Ca^{2+} hendelser i astrocytter. vise at det finnes noe analogt til en relativ refraktærperiode for astrocytter.

Noe mer om refraktærperiode her: Estimering, interspike intervalls, problemer med astrocytt-data osv...

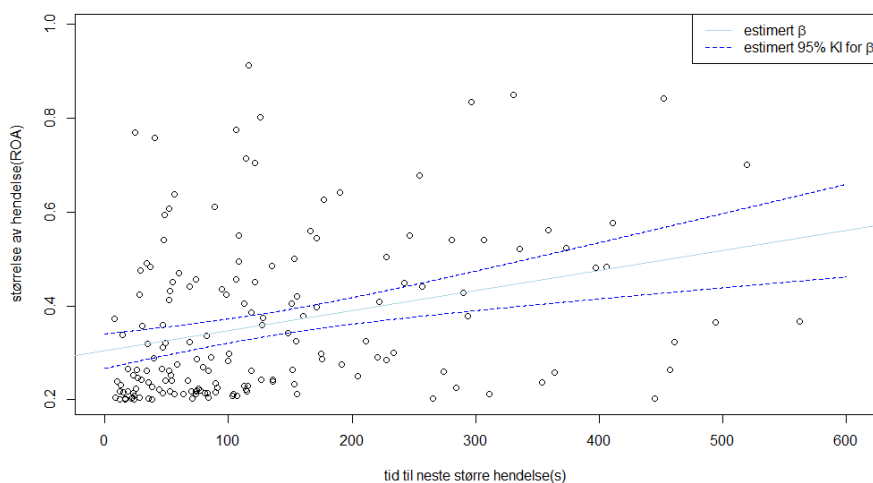
Utspringet for denne oppgaven startet med spørsmålet om det finnes en refraktærperiode for store Ca^{2+} hendelser i astrocytter. En refraktærperiode er et tidsintervall der cellen ikke er i stand til å gjenta en "action". I vår kontekst betyr dette at det finnes en periode etter en Ca^{2+} hendelse i en astrocytt, der astrocytten ikke er i stand til å produsere en like stor Ca^{2+} hendelse.

Refraktærperiode er viktig i modellering av tiden mellom aksjonspotensialer i nevroner. Tiden mellom kalles en interspike-interval. Etter en aksjonspotensiale tar det tid før nevronen er i stand til å produsere en ny aksjonspotensiale. Refraktærperioden til en nevron kan deles inn i absolutt og relativ refraktærperiode. I den absolutte refraktærperioden kan ingen kan ingen injisert strøm skape en ny aksjonspotensiale. I den den relative refraktærperioden er terskelen for hvor mye strøm som trengs for å produsere en aksjonspotensiale betydelig forhøyet. Den relative refraktærperioden er generelt lengre enn den absolutte refraktærperioden [5]. Det finnes ulike metoder for å estimere refraktærperioden i et nevron, avhengig av obsererte verdier av interspike intervaller, hva slags type nevroner og hvilken modell for interspike intervaller man bruker. Se Hampel(2008)[6] for en gjennomgang og evaluering av ulike metoder for å estimere refraktærperiode i nevroner. Sannsynligheten for at det oppstår et aksjonspotensiale, for en gitt tid, i en nevron er avhengig av mange prosesser i sentralnervesystemet (i tillegg til ekstern stimuli), men også obsererte tidligere interspike intervaller i samme nevron. Song et al.(2018) [11] finner at korte interspike intervaller følges opp av lengre intervaller.

Det finnes ingen kjent fysiologisk forklaring for hvorfor astrocytter kan ha en refraktærperiode for Ca^{2+} hendelser, og det er generelt vanskelig å måle om de har det. Som tidligere nevnt er det ingen enighet om en presis definisjon for Ca^{2+} hendelser, som f.eks. når de starter og når de slutter, i tillegg er det knyttet betydelig usikkerhet til *in-vivo* målinger av Ca^{2+} nivå i astrocytter, samt at det ikke er utbredt å måle Ca^{2+} signal i 3D s.a. det ikke er gitt at man nøyaktig vet når Ca^{2+} hendelsen starter og slutter (selv om man hadde hatt en enighet om presis definisjon av Ca^{2+} hendelser). Det vi vet er at det er en tendens til at store Ca^{2+} er spredt i tid, i.e. ikke kommer rett etter hverandre. En parallell til den *relative refraktærperioden* for nevroner, i.e. tidsintervallet der nevronet kan produsere et aksjonspotensiale, men det krever betydelig forhøyet elektrisk strøm, er for astrocytter tidsintervallet der astrocytten ikke er i stand til å produsere en like stor, eller større, Ca^{2+} hendelse som den foregående. En enkel regresjonsanalyse bekrefter at det finnes en tendens til at store Ca^{2+} hendelser er spredt i tid og at jo større Ca^{2+} hendelsene, jo mer sannsynlig er

1.3. Størrelse av Ca²⁺ hendelser

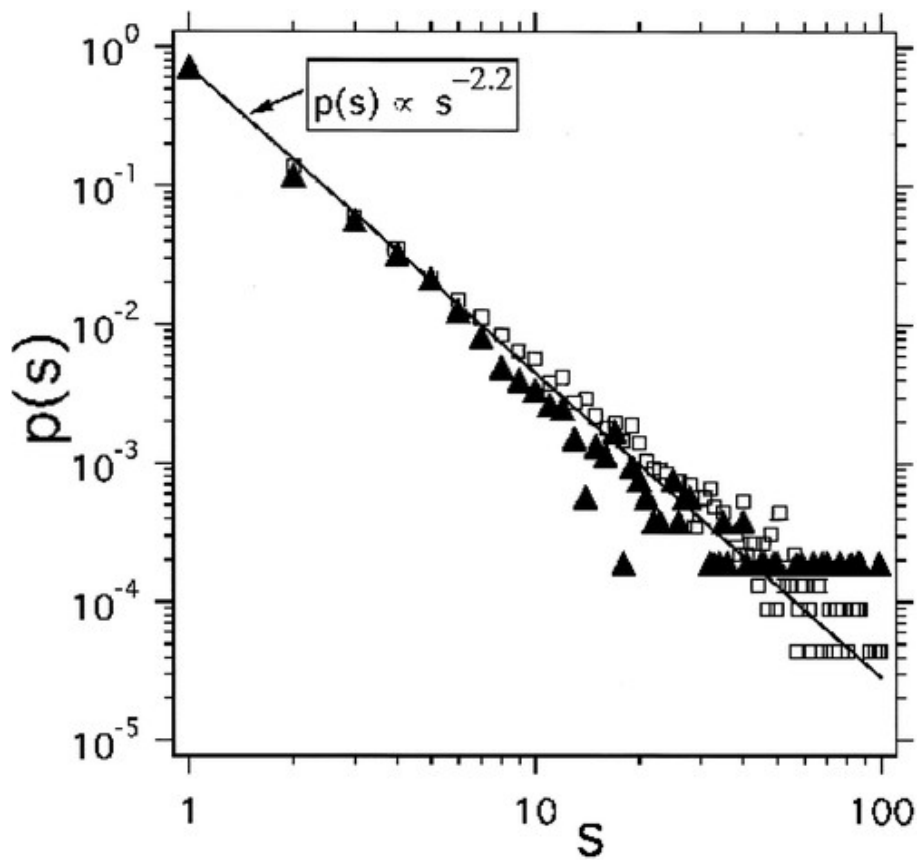
det at tidsintervallet mellom dem er stort. Diskusjon rundt data og definisjoner i kap. 2.



Figur 1.2: Estimert parameter og observert data for en enkel regresjonsmodell: $Y_i = \beta_0 + \beta X_i$. Y_i er størrelse på den i -te Ca²⁺ hendelsen og X_i er tid til neste Ca²⁺ hendelse som er større enn Y_i . For nærmere definisjon av variabler se kap 2.

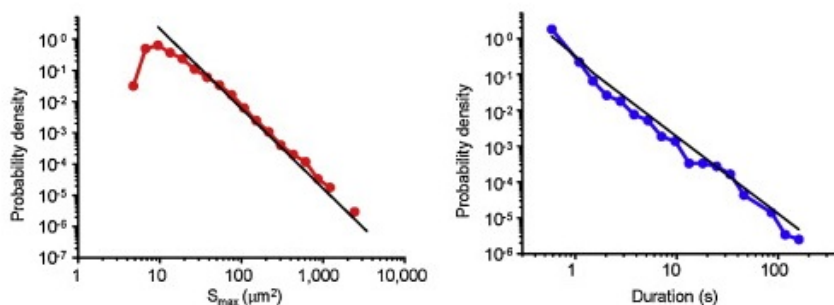
1.3 Størrelse av Ca²⁺ hendelser

I stedet for å se på modeller som forsøker å predikere *når* en Ca²⁺ hendelse vil skje, eller modellere interspike intervallet for astrocytter, vil vi starte med å se på størrelsen av Ca²⁺ hendelser. Det er to studier som har modellert fordelingen til størrelsen av Ca²⁺ hendelser i astrocytter. Jung et al (1998) [7] undersøkte fordelingen av størrelse til Ca²⁺ hendelser i kultiverte astrocytter (cellekulturer utviklet fra hjernen døde dyr) og fant at størrelsene til hendelsene er potenslovsfordelt. Større hendelser er eksponensielt mindre sannsynlig enn mindre hendelser.



Figur 1.3: Figur fra Jung et al. (1998). s er størrelsen av en Ca²⁺ hendelse og $p(s)$ er frekvens av hendelse. Fylte trekanter er observert data og tomme kvadrater er simulert data.

Wu et al. (2014) [13] fant det samme for Ca²⁺ hendelser i astrocytter i oppskårte skiver av musehjerne i 2014. I tillegg fant de at varigheten til Ca²⁺ hendelser også er potenslovsfordelt i de samme skivene.



Figur 1.4: Noe om figuren her

At Ca²⁺ hendelser er potenslovforderlt betyr at størrelsen av Ca²⁺ hendelser, $Y \in (0, \infty)$, kommer fra en fordeling der $f(y) \propto y^{-\alpha}$. Dette er ikke en ekte sannsynlighetsfordeling, siden $\lim_{y \rightarrow 0} = \infty$, så det er vanlig å modellere kontinuerlige variabler som er potenslovforderlt som paretofordelte, slik at $y \in [y_{min}, \infty)$ har sannsynlighetstetthet $f(y) \propto y^{-\alpha}$ og $y < y_{min}$ har tetthet $f(y) = 0$. Se Clauset et al. (2009)[4] for en gjennomgang av potenslovforderlinger. Vi kan legge merke til tre aspekter ved spesielt Wu et al. (2014) og til dels Jung et al. (1998) sine artikler. 1) De motiverer ikke hvorfor størrelsen til Ca²⁺ hendelser bør være potenslovforderlt. Hvis man skal bruke en statistisk modell for å modellere et fysiologisk fenomen bør man tenke grundig gjennom antakelsene som ligger til grunn for den statistiske modellen og argumentere for hvorfor disse antakelsene gjelder. 2) Wu et al. (2014) tester kun om Ca²⁺ hendelser, i de cellene med Ca²⁺ hendelser som passer en potenslovforderling best, passer en potenslovforderling bedre enn kun to andre fordelinger, en lognormalfordeling og en eksponensiellfordeling. Dette illustrerer problemet med å teste for om data passer en viss fordeling. På Wikipedia-siden List of probability distributions er det listet opp rundt 40 ofte brukte kontinuerlige fordelinger med støtte i $(0, \infty)$ og selv om en viss fordeling passer bedre enn alle andre vanlige fordelinger så er det ingen garanti for at en annen mindre kjent fordeling ikke ville passet bedre. 3) Verken Wu et al. (2014) eller Jung et al. (2014) diskuterer ulike varianter av potenslov-liknende fordelinger. Gitt at størrelser av Ca²⁺ hendelser ligner en potenslovforderling, så bør man ha i mente at andre fordelinger kan generere lignende data. Uansett kan man legge til grunn at størrelsen av Ca²⁺ hendelser har en tendens til å ligne et potenslovforderlt variabler. I denne oppgaven vil vi i all hovedsak diskutere ulike varianter av gammafordelingen for å modellere størrelsene til Ca²⁺ hendelser. Vi diskuterer ikke potenslovforderlinger noe mer, men vi kan merke oss at gammafordelingen ligner på en potenslovforderling. Vi har at hvis Y er gammafordelt vil tettheten være $f(y) \propto y^{-\alpha} \exp(-\lambda y)$, som ligner på en potenslovforderling fram til y blir stor.

Analogt til nevroner er det gjort noen forsøk på å modellere hele dynamikken i Ca²⁺ nivå for astrocytter over tid (ved hjelp av partielle differensialligninger og estimater for gjennomstrømming av signalstoffer gjennom kanaler inn til astrocytten) *ex-vivo*, i cellekulturer eller hjerneskiver. Se f.eks. igjen Wu et al.

(2014) [13]. Men det er ikke, så vidt bekjent, gjort noen forsøk på å modellere denne dynamikken *in-vivo*.

1.4 Gammalprosess

Vi er interessert i å utvikle en statistisk modell for å modellere størrelsen av Ca^{2+} hendelser. Modellen bør være fleksibel nok til at den kan integrere ulike kovariater i analysen og reflektere våre grunnleggende antakelser om den fysiologiske prosessen som ligger bak Ca^{2+} hendelser.

Vår fremste antakelse er at jo større tidsintervall det er mellom Ca^{2+} hendelsene, jo større vil påfølgende hendelse sannsynligvis være. En modell type modell som passer disse spesifikasjonene er en gammalprosess-modell. Gamma prosess er en monotont økende "jumpprosess der inkrementene er uavhengig og gamma-fordelte.

Gammalprosess ble først anvendt for å modellere vannvolumet i en damm av Moran [8], der tilsiget av vann er antatt gamma-fordelt. Senere er det blitt en ubredt modell innenfor ingeniørvitenskap for å modellere degradering av fysiske strukturer. Her antar man at forskjellen i degradering ved to forskjellige tidspunkter er gammafordelt slik at den totale summen av degradering opp til et tidpunkt er realisasjoner av en gammalprosess. Se van Noortwijk (2009) [9] for en grundig gjennomgang av anvendelser av gammalprosess-modellen innen ingeniørvitenskap.

Vi tror først og fremst at sannsynligheten for at Ca^{2+} hendelser er større enn en viss verdi øker jo lenger tid som er gått. Samtidig tror vi at denne sannsynligheten øker gradvis, så det lønner seg å bruke en kontinuerlig fordeling.

KAPITTEL 2

Data

2.1 Måling av Ca²⁺ aktivitet og Ca²⁺ hendelser

Som nevnt i forrige kapittel er det ingen konsensus om hva en Ca²⁺ hendelse er eller hvordan man bør måle Ca²⁺ aktivitet i astrocytter *in-vivo*. Samtidig er det heftet noe usikkerhet ved spørsmålet om Ca²⁺ aktiviteten som måles *in-vivo* reflekterer det faktiske Ca²⁺ nivået i astrocytten. I en ideell verden hadde vi hatt tilgang til presise målinger av all Ca²⁺ aktiviteten i astrocytter over en lang tidsperiode, men det er ikke gjennomførbart per nå. Vi har brukt 60 10-minutters opptak som måler Ca²⁺ aktiviteten over et 2D plan som inkluderer hele astrocyttceller. At opptakene er 10 minutter lange gjør oss i stand til å modellere sammenhengen mellom tid mellom Ca²⁺ hendelser og størrelsen til Ca²⁺ hendelser, spesielt når det er stor avstand mellom hendelsene. At vi ikke har brukt målinger av all Ca²⁺ aktiviteten i astrocyttcellene, i tre dimensjoner, gjør at det er noe usikkerhet mellom sammenhengene vi estimerer der det er kort tid mellom Ca²⁺ hendelser er usikre. Siden vi har målinger av Ca²⁺ aktiviteten over hele astrocyttceller er det mindre usikkerhet rundt de målte størrelsene av Ca²⁺ hendelser enn hvis vi hadde fokusert på mindre deler av astrocyttene.

Et fortrinn for *in-vitro* modeller for Ca²⁺ aktivitet i astrocytter, i forhold til *in-vivo*, er at man i større grad kan undersøke presist hvilke mekanismer som forårsaker variasjon i Ca²⁺ aktivitet. For modellering av Ca²⁺ aktivitet *in-vivo* kan man kun kontrollere for det man kan kalle målinger av makro aktivitet i hjernen, som målinger av musens fysiske aktivitet, øyepupilldilatasjon, værhåraktivitet og søvnstadie. Det er vanskelig å måle mikro aktivitet rundt hver celle *in-vivo* og dette har konsekvenser for hva slags kausale slutninger man kan trekke modellene vi presenterer senere. Aalen et al. (2008) fremhever i kapittel 9 av *Survival and event history analysis: a process point of view*[14] en motsetning mellom hvordan statistikere presenterer sine funn og hvordan andre forskere tolker funnene. Aalen et al. peker spesifikt på at særlig innenfor medisinsk forskning blir statistiske studier brukt til å trekke kausale slutninger uten at det finnes noe grunnlag for det. Siden vi i denne oppgaven fokuserer på modellering av *in-vivo* målinger av Ca²⁺ hendelser og ikke har tilgang på målinger av mekanismene som styrer Ca²⁺ nivået i hver individuelle celle er det ingen grunnlag for å trekke kausale slutninger fra denne oppgaven, det må bli overlatt til *in-vitro* studier på cellenivå, men ved å forsøke å modellere Ca²⁺ hendelser *in-vivo* kan vi også få en dypere forståelse for dynamikken i Ca²⁺ aktivitet i astrocytter.

Som nevnt i introduksjonen er det ingen konsensus om en definisjon av en Ca²⁺ hendelse, bortsett fra at det er en rask heving av Ca²⁺ nivået i astrocytten, men forskning på astrocytter er et felt i rask utvikling og det er grunn til å håpe på fremgang også i spørsmålet om hvordan definere Ca²⁺ hendelser. Et av de største framskrittene innen Ca²⁺ modellering i astrocytter er AQUA algoritmen, publisert i 2019 [12]. AQUA algoritmen bruker de samme transformasjonene på de råe gråverdi-tidsrekkene for hver piksel som ROA-algoritmen som ble brukt i vårt datasett, men bruker også ulike clustering-teknikker til å skille ulike Ca²⁺ hendelser fra hverandre. Dette kan gi en mer robust måte å måle størrelsen og tiden til hver enkelt Ca²⁺ hendelse, samt andre viktige aspekter som på hvilken måte hendelsen sprer seg. Foreløpig er det ikke mulig å prosessere lange tidsrekker fra høykvalitetsopptak med denne metoden, men det bør være mulig å utvide modelleringsrammeverket vårt hvis den type data blir tilgjengelig.

Etter en rask gjennomgang av ulemper og fordeler med datamaterialet vårt går vi nå kort gjennom ROA-algoritmen, som ble brukt til å måle Ca²⁺ aktivitet i datasettet som blir brukt i denne oppgaven.

2.2 ROA-algoritmen

Over et 2D plan i astrocyttcellen blir det gjort opptak som måler flurosensverdien av hver piksel 30 ganger i sekundet. Vi har $F_i(t)$ som er gråverdien til piksel i på et tidspunkt t . En gaussisk utjevning over flere nærliggende piksler og over et tidsintervall blir brukt på alle $F_i(t)$ slik at man får $F'_i(t)$. En variansstabiliserende transformasjon blir brukt $F_i^*(t) = \sqrt{F'_i(t)}$ slik at variansen til pikselverdiene ikke skal avhenge i nevneverdig grad av forventingen til pikselverdien. For hver piksel blir det estimert en terskel slik $\tau_i = \hat{\mu}_i + \kappa \cdot \hat{\sigma}_i$, der $\hat{\mu}_i$ er typetallet til gråverdien av pikselen, $\text{mode}(F_i^*)$, og $\hat{\sigma}_i$ er det estimerte standardavviket, $\hat{\sigma}_i^2 = \frac{1}{0.9099} \text{Median}([F_i^*(t+1) - F_i^*(t)]^2)$, og κ er selvvalgt tall som bestemmer. Da får man en dikotomisert tidsrekke for hver piksel som indikerer om det er et aktivt Ca²⁺ signal der eller ikke. For hvert tidspunkt blir andelen aktive piksler utregnet og satt sammen i en ROA trace.

En kort oppsummering er at den råe flurosensverdien for hver piksel blir transformert og utjevna.

$$F_i(t) \longrightarrow F'_i(t) \longrightarrow \sqrt{F'_i(t)} = F_i^*(t)$$

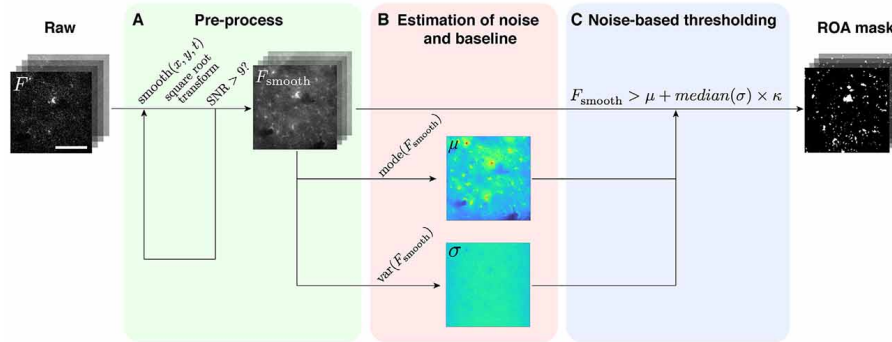
Så estimerer man en terskel for hver piksel og setter sammen en tidsrekke som indikerer om det er Ca²⁺ aktivitet for hver piksel for ulike tidspunktene.

$$ROA_i(t) = \begin{cases} 1, & \text{hvis } F_i^*(t) \geq \hat{\mu}_i + k \cdot \hat{\sigma}_i \\ 0, & \text{ellers} \end{cases}$$

Så beregner man den totale Ca²⁺ aktiviteten for hele planet til ethvert tidspunkt i opptaket ved å utregne den totale andelen piksler med Ca²⁺ aktivitet.

$$ROA(t) = \frac{1}{n} \sum_{i=1}^n ROA_i(t), \quad n \text{ er antall piksler}$$

En grafisk framstilling av denne prosessen er hentet fra artikkelen til Bjornstad et al. (2021) [1].



Figur 2.1: Grafisk fremstilling av prosessering av ROA-algoritmen [1]

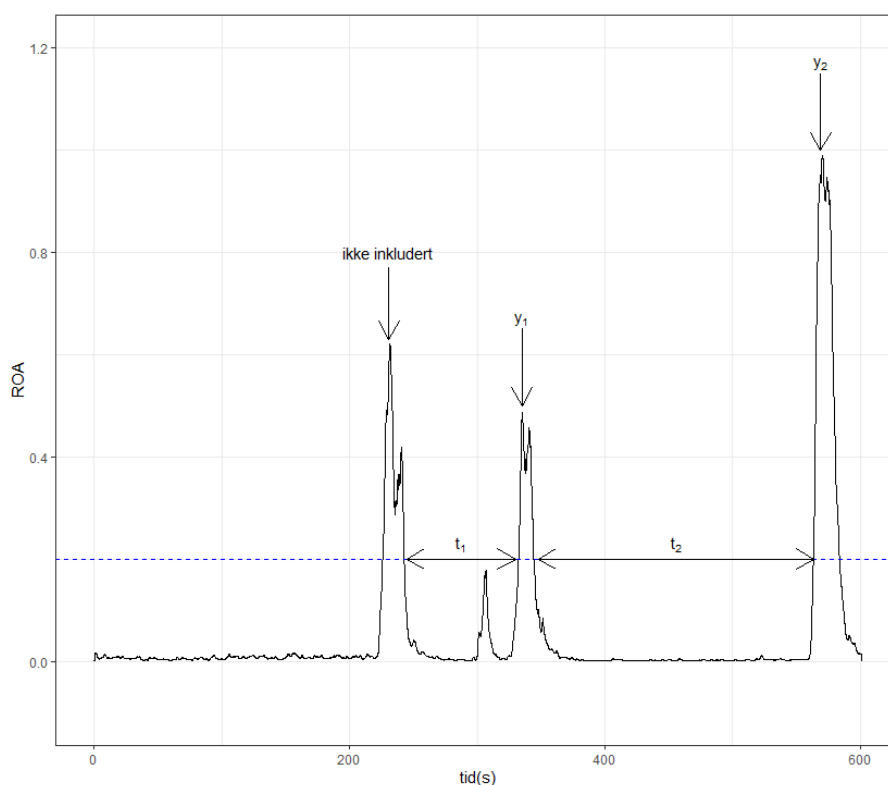
2.3 Hva er en Ca²⁺ hendelse

En naturlig definisjon av Ca²⁺ hendelser fra ROA-data er tidsintervaller der ROA-verdien er sammenhengende positiv, men vi vil være så konservative som mulig i vår definisjon av Ca²⁺ hendelser, siden det fortsatt er usikkerhet rundt målinger av Ca²⁺ aktivitet, så vi setter inn en høyere terskel for hva som er en Ca²⁺ hendelse. Dermed konsentrerer vi oss hovedsaklig om større Ca²⁺ hendelser, som det er mindre usikkerhet rundt, men siden vi kun ser på Ca²⁺ aktivitet over en høyere terskel må vi også ta høyde for at Ca²⁺ aktiviteten kan gå under terskelen noen øyeblikk selv om den er del av samme hendelse. En Ca²⁺ hendelse, i vår definisjon, starter altså når ROA-tracen på et tidspunkt går over en terskel $ROA(t) > \tau$, slutter når ROA-verdien har passert under terskelen, såfremt den ikke går over terskelen igjen innen 3 sekunder. Vi har valgt en overlappverdi på 3 sekunder også med et konservativt utgangspunkt. Jo lenger overlappen er, jo større er sannsynligheten for å måle flere Ca²⁺ hendelser som en, men siden vi vet at Ca²⁺ hendelser kan vare rundt 1-2 sekunder vil en overlappverdi på 3 sekunder garantere at vi ikke måler en hendelse som flere hendelser.

Størrelsen til en Ca²⁺ hendelser kan måles som maksverdien av Ca²⁺ aktivitet under hendelsen, integralet av ROA-kurven under hendelsen eller tidslengden av hendelsen. Vi velger å måle størrelse av Ca²⁺ hendelser som maksverdien siden dette er den vanligste måten å måle Ca²⁺ hendelser på i tidligere litteratur, samt at maksverdien ga sterkest sammenheng med de andre variablene vi bruker til å modellere i enkle analyser. Hjularten som musen løper i under hendelsen måler vi på samme tidspunkt som ROA-verdien når sitt maksimum under hendelsen og tid mellom hendelsene måler vi fra hendelsen er ferdig, der ROA-verdien går under vår terskel, og til ROA-verdien går over terskelen igjen for den påfølgende hendelsen. Siden Ca²⁺ hendelser ikke varer særlig lengde og har en tendens til å spre seg ut over tid er det ikke sannsynlig at når man starter og stopper måling av tid mellom hendelser vil påvirke resultatene av analysen, men måling av hjulart kan med fordel utforskes mer. Den høyere terskelen for Ca²⁺ hendelser

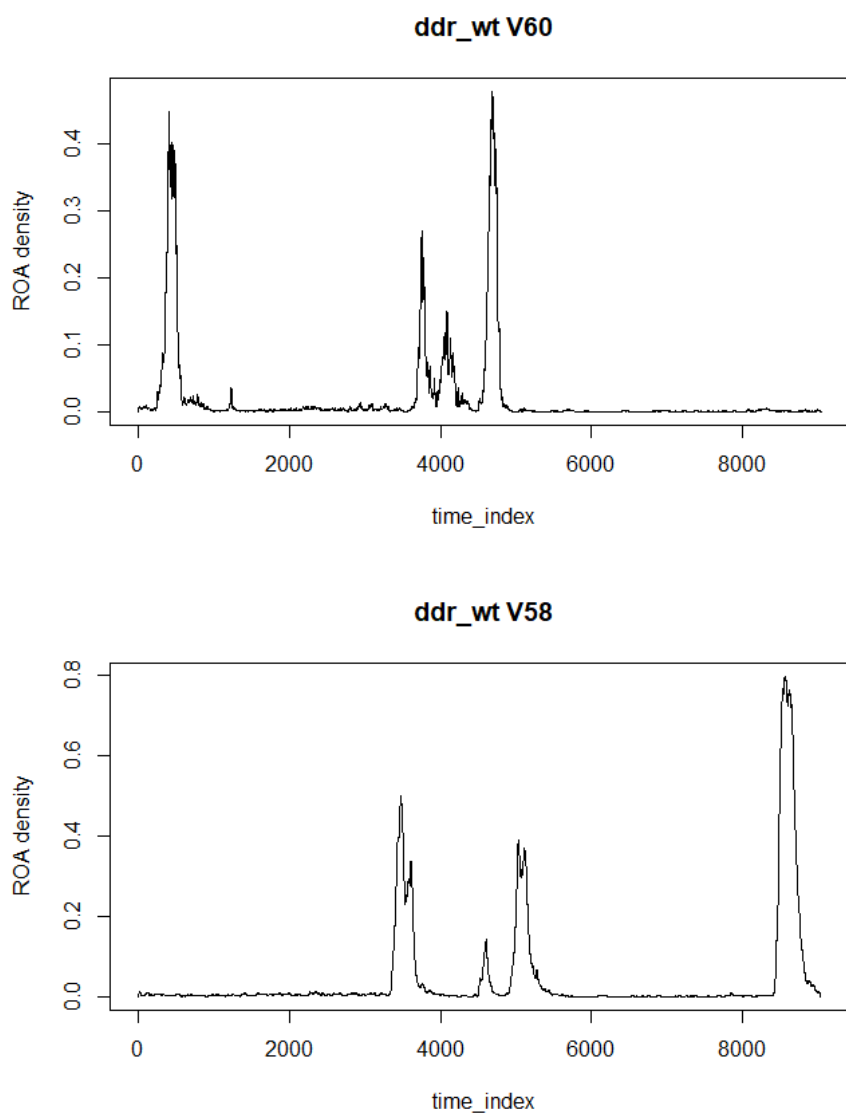
2.3. Hva er en Ca²⁺ hendelse

som vi har valgt er 0.2, kun basert på en visuell inspeksjon av ROA-dataene. 0.2 virker som en naturlig terskel basert på Ca²⁺ aktiviteten vi er interessert i å modellere, de høye store piggene i ROA-dataen, men denne terskelen kan absolutt diskuteres og effekten av å justere terskelverdien kan være stor. Siden vi både er interessert i å måle tid mellom hendelsene og sammenhengen mellom størrelse av to påfølgende hendelser måler vi ikke størrelsen av den første hendelsen i opptaket. At vi ikke måler den første hendelsen vil trolig ikke påvirke resultatene fra vår analyse, da datasettet vårt er stort nok uten dem til å tilpasse modellene godt nok. Nedenfor er det en grafisk framstilling av hvordan vi måler tid mellom hendelser og størrelsen av hendelser.



Figur 2.2: Hvordan størrelse og tid blir målt. Den blå striplede linja er terskelverdien, y og t er de observerte størrelsene og tidsverdiene.

En problemstilling for oss er at hver ROA-trace blir estimert basert på individuelle parametere og det er ikke opplagt at størrelsene for Ca²⁺ hendelser på tvers av opptak er sammenlignbare. Hver mus, hver celle og hvert opptak er unikt og resultatet blir at størrelsene til Ca²⁺ hendelser kan variere vilt. Samtidig er det ingen tvil om at oppførselen til Ca²⁺ aktiviteten på tvers av opptak er sammenfallene, det er generelt lange perioder med liten til ingen aktivitet før det kommer en eller flere pigger som indikerer en massiv økning i Ca²⁺ aktivitet. Dette fenomenet er illustrert med to ulike opptak nedenfor.



Figur 2.3: To forskjellige ROA tracer med forskjellige maksverdier, men lik oppførsel. ROA tracer fra to ulike -mus.

For å gjøre størrelsene av Ca²⁺ hendelser mer sammenliknbare på tvers av opptak har vi valgt å normalisere dem slik at hvert opptak får en maksverdi rett under 1, ved å dele alle ROA-verdiene på summen av den observerte maksverdien i opptaket og en korrigering på 0.01 for at ROA-verdien ikke skal bli nøyaktig 1. Etter denne normaliseringen har opptakene fortsatt samme oppførsel. Selv om vi gjør Ca²⁺ hendelsene, forhåpentligvis, mer sammenliknbare vil det føre til at fordelingen av størrelsene til Ca²⁺ hendelser får en større hale rundt verdien

2.3. Hva er en Ca²⁺ hendelse

1. Hvis vi tenker tilbake på diskusjonen i forrige kapittel om potenslovfordeling vil denne normaliseringen rettferdiggjøre en modifisert potenslovfordeling, der tettheten til størrelsen av Ca²⁺ hendelser er $f(y) \propto y^{-\alpha} \exp(-\lambda y)$, siden den større halen passer godt med det eksponensielle leddet på slutten som vil dominere potenslovleddet for store hendelser.

Til slutt er det en del opptak der det ikke er noe Ca²⁺ aktivitet eller en betydelig drift i baselinenivået på Ca²⁺ aktivitet, dette er typisk en følge av forstyrrelser i mikroskoputsryret. Opptakene der det åpenbart har skjedd noe feil med utstyret blir fjernet før analysen etter en visuell inspeksjon.

KAPITTEL 3

Gamma-prosess

3.1 intro til gamma prosess

gammafordelingen

Vi har at et gamma-fordelt variabel Z med parametere (a, b) $Z \sim \text{Gamma}(a, b)$ har sannsynlighetstetthet $g(a, b)$ og kumulativ sannsynlighetstetthet $G(c, a, b)$

$$g(z, a, b) = \frac{b^a}{\Gamma(a)} z^{a-1} e^{-bz} \quad (3.1)$$

$$G(c, a, b) = \int_0^c g(z, a, b) dz \quad (3.2)$$

Med forventning og varians

$$E(Z) = \frac{a}{b} \quad (3.3)$$

$$\text{Var}(Z) = \frac{a}{b^2} \quad (3.4)$$

En viktig egenskap ved gammafordelte variabler er at en summen av uavhengige gammafordelte variabler, med lik b -parameter, også er gammafordelt.

Lemma 3.1.1. *Gitt uavhengige variabler dZ_1, dZ_2, \dots, dZ_n som er gammafordelte med parametere $(a_1, b), (a_2, b), \dots, (a_n, b)$ har vi at $\sum_{i=1}^n dZ_i \sim \text{Gamma}(\sum_{i=1}^n a_i, b)$.*

Vi kan vise dette ved hjelp av den karakteristiske funksjonen til et gammafordelt variabel. Den karakteristiske funksjonen, $\phi_{dZ}(t) = E[e^{itdZ}]$ til et stokastisk variabel dZ definerer fordelingen til variabelen fullstendig. For en sum av uavhengige variabler, $Z = \sum_{i=1}^n dZ_i$, er den karakteristiske funksjonen til summen gitt ved produktet av de karakteristiske funksjonene til de summerte variablene $\phi_Z(t) = \phi_{dZ_1}(t) \phi_{dZ_2}(t) \cdots \phi_{dZ_n}(t)$.

Bevis. Den karakteristiske funksjonen for en gammafordelt variabel, $dZ_i \sim \text{Gamma}(a_i, b)$ er $\phi_{dZ_i} = (1 - it/b)^{-a_i}$ og for en sum av uavhengige variabler $Z = \sum_{i=1}^n dZ_i$ er den karakteristiske funksjonen $\phi_Z = \phi_{dZ_1} \phi_{dZ_2} \cdots \phi_{dZ_n}$. Vi har

$$\phi_Z = \phi_{dZ_1} \phi_{dZ_2} \cdots \phi_{dZ_n}$$

$$\begin{aligned}
&= (1 - it/b)^{-a_1} (1 - it/b)^{-a_2} \dots (1 - it/b)^{-a_n} \\
&= (1 - it/b)^{-(a_1+a_2)} (1 - it/b)^{-(a_3)} \dots (1 - it/b)^{-a_n} \\
&= (1 - it/b)^{-\left(\sum_{i=1}^n a_i\right)}
\end{aligned}$$

som er den karakteristiske funksjonen for et gammafordelt variabel med parametre $(\sum_{i=1}^n a_i, b)$. Altså har vi at $Z = X_1 + X_2 + \dots + X_n$ er gamma-fordelt $Z \sim \text{Gamma}(\sum_{i=1}^n a_i, b)$. ■

Definition 3.1.2. En gamma-prosess $\{Z(t) : t > 0\}$ er en sum av uavhengige, gammafordelte inkremer $dZ(t) \sim \text{Gamma}(adt, b)$, der $dZ(t) = Z(t+dt) - Z(t)$ og $Z(t) = \sum_{s=1}^{t/dt} dZ(s)$ slik at $Z(t) \sim \text{Gamma}(at, b)$. Mer generelt kan man også si at $Z(t) \sim \text{Gamma}(aM(t), b)$ er en gammaprosess, hvis $M(t)$ er en positiv, monotont økende funksjon for t .

Denne definisjonen av en gammaprosess gir oss en fleksibel måte å modellere fordelingen til Ca2+ hendelser. Vi vil la $M(t)$ være en funksjon av ulike kontrollvariabler X_1, X_2, \dots, X_p , samt t , slik $M(t) = \exp(\beta_t t + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$. Siden vi med denne typen $M(t)$ alltid har $M(t) > 0$ vil $Z(t) \sim \text{Gamma}(aM(t), b)$ være en gammaprosess så lenge $\beta_t > 0$.

Generelt for gammafordelingen er det ulike måter å estimere parametrene og det finnes forskjellige estimatorer med en lukket form, men for maximum likelihood estimatorer må man bruke ulike iterative algoritmer for å maksimere likelihoodfunksjonen

Trunkert fordeling

En enkel gammaprosess modell kan fange opp det vi mener er de sentrale aspektene ved fordelingen til Ca2+ hendelser, nemlig at størrelsen til en gitt hendelse sannsynligvis blir større jo lenger tid som er gått siden forrige hendelse og jo mindre forrige hendelse var. Samtidig er det en svakhet, fra definisjonen vår av en Ca2+ hendelse, at vi ikke observerer hendelser som er mindre enn vår egendefinerte terskel, som vi kaller τ , men en enkel gammaprosess modell vil uansett gi en positiv sannsynlighet for at størrelsen er mindre enn τ , men for. Siden vi bare kan måle størrelsen til Ca2+ hendelser som er større enn vår terskel, kan vi modellere fordelingen som en *trunkert* fordeling.

Vi kan merke oss her at det ikke er snakk om en form for *sensurering*. Sensurering betyr vi ikke kjenner til den eksakte verdien av en observasjon, men at verdien ligger i et intervall. I vår setting ville dette betydd at vi hadde observasjoner av Ca2+ hendelser som vi vet er under vår terskel, men vi definerer en Ca2+ hendelse som en sammenhengende aktivering over terskelen, derfor kan vi ikke observere en Ca2+ hendelse som er mindre enn terskelen, ut ifra vår definisjon.

En trunkert fordeling innskrenker støtten til den opprinnelige fordelingen. I vårt tilfelle vil det bety at sannsynligheten for å observere en Ca2+ hendelse som er mindre enn vår terskel, τ , vil være 0. Innenfor den innskrenka støtten vil den trunkerte fordelingen være definert den opprinnelige fordelingen (gitt de samme parameterverdiene), delt på en normaliseringsfaktor. Slik at innenfor den innskrenka støtten vil den trunkerte fordelingen være proporsjonal til den opprinnelige fordelingen.

Definition 3.1.3. Gitt en stokastisk variabel X med sannsynlighetstetthet $f(x)$ og kumulativ sannsynlighetsfunksjon $F(x)$, der begge har støtte $(a^*, b^*]$, da har vi for en innskrenka støtte $(a, b]$, der $a, b \in (a^*, b^*]$ og $a < b$, at tettheten til X , gitt $a < X \leq b$ er

$$f(x|a < X \leq b) = \frac{f(x) \cdot \mathbb{I}(a < x \leq b)}{F(b) - F(a)}$$

Fra vår definisjon av en Ca2+ hendelse kan vi ikke observere hendelser under en viss størrelse, derfor er vi interessert i fordelingen til et trunkert gammafordelt variabel $Z > \tau$, $Z \sim \text{Gamma}(a, b)$, $\tau \in (0, \infty)$. Sannsynlighetstettheten til $Z|Z > \tau$, $f_{Z|Z > \tau}(z)$ er gitt ved

$$f_{Z|Z > \tau}(z) = \frac{h(z, a, b)}{1 - G(\tau, a, b)}$$

der $h(z, a, b) = g(z, a, b)\mathbb{I}(z > \tau)$

og $\mathbb{I}(z > \tau) = \begin{cases} 1, & \text{hvis } z > \tau \\ 0, & \text{ellers} \end{cases}$

Vi kan finne forventning til $Z|Z > \tau$ ved

$$E(Z|Z > \tau) = \int_0^\infty \frac{z h(z, a, b)}{1 - G(\tau, a, b)} dz$$

og varians til $Z|Z > \tau$

$$\text{Var}(Z|Z > \tau) = \int_0^\infty \frac{z^2 h(z, a, b)}{1 - G(\tau, a, b)} dz - [E(Z|Z > \tau)]^2$$

og får da at

$$E(Z|Z > \tau) = \frac{\Gamma(a + 1, b\tau)}{b\Gamma(a, b\tau)}$$

$$\text{Var}(Z|Z > \tau) = \frac{\Gamma(a + 2, b\tau)}{b^2\Gamma(a, b\tau)} - \frac{\Gamma(a + 1, b\tau)^2}{b^2\Gamma(a, b\tau)^2}$$

der $\Gamma(a, \tau)$ er den øvre ukomplette gamma funksjonen, $\Gamma(a, \tau) = \int_\tau^\infty x^{a-1} e^{-x} dx$. Vi bruker at $G(\tau, a, b) = \frac{\gamma(a, b\tau)}{\Gamma(a)}$, der $\gamma(a, b\tau)$ er den nedre ukomplette gamma funksjonen $\gamma(a, \tau) = \int_0^\tau x^{a-1} e^{-x} dx$, og dermed er $1 - G(\tau, a, b) = \frac{\Gamma(a, b\tau)}{\Gamma(a)}$.

Gitt u.i.f. observasjoner $\mathbf{z}' = (z'_1, z'_2, \dots, z'_n)$, der $Z'_i \sim \text{TrunkertGamma}(a, b, \tau)$ For å estimere parameterne (a, b) maksimerer vi log likelihood-funksjonen $l(a, b|\mathbf{z}', \tau)$ mhp. (a, b) ¹

$$\max_{\hat{a}_{ML}, \hat{b}_{ML}} \left\{ \sum_{i=1}^n \log \left(\frac{h(z_i, \hat{a}_{ML}, \hat{b}_{ML})}{1 - G(\tau, \hat{a}_{ML}, \hat{b}_{ML})} \right) \right\}.$$

¹se evt denne <https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-27/issue-2/Estimating-the-Parameters-of-a-Truncated-Gamma-Distribution/10.1214/aoms/1177728272.full>

Det er ingen enkel måte å å utlede et lukket uttrykk for estimatorene(eller kanskje) $\hat{a}_{ML}, \hat{b}_{ML}$, men man kan enkelt maksimere $l(a, b|\mathbf{z}', \tau)$ numerisk.

3.2 Transformert fordeling

Vi skal observere spredningen til n Ca2+ hendelser, $Y_i \in (0, 1)$, for $i = 1, 2, \dots, n$. Siden en gammafordeling sin support er i $(0, \infty)$ må vi bruke en transformasjon fra $(0, \infty)$ til $(0, 1)$. Vi bruker transformasjonen

$$Y_i = c(Z_i) = 1 - e^{-Z_i} \quad (3.5)$$

der $Z_i \sim \text{Gamma}(a, b)$. Da vil

$$Z_i = c^{-1}(Y_i) = -\log(1 - Y_i) \quad (3.6)$$

Siden $c(Z)$ er økenende for $Z \in (0, \infty)$ finner vi sannsynlighetstettheten til det transformerte variabelen med formelen

$$f_Y(y) = g(c^{-1}(y), a, b) \frac{d}{dy} c^{-1}(y)$$

Vi får da at sannsynlighetstettheten til Y er

$$f_Y(y) = \frac{b^a}{\Gamma(a)} (1 - y)^{b-1} (-\log(1 - y))^{a-1} \quad (3.7)$$

kan finne forventning og varians

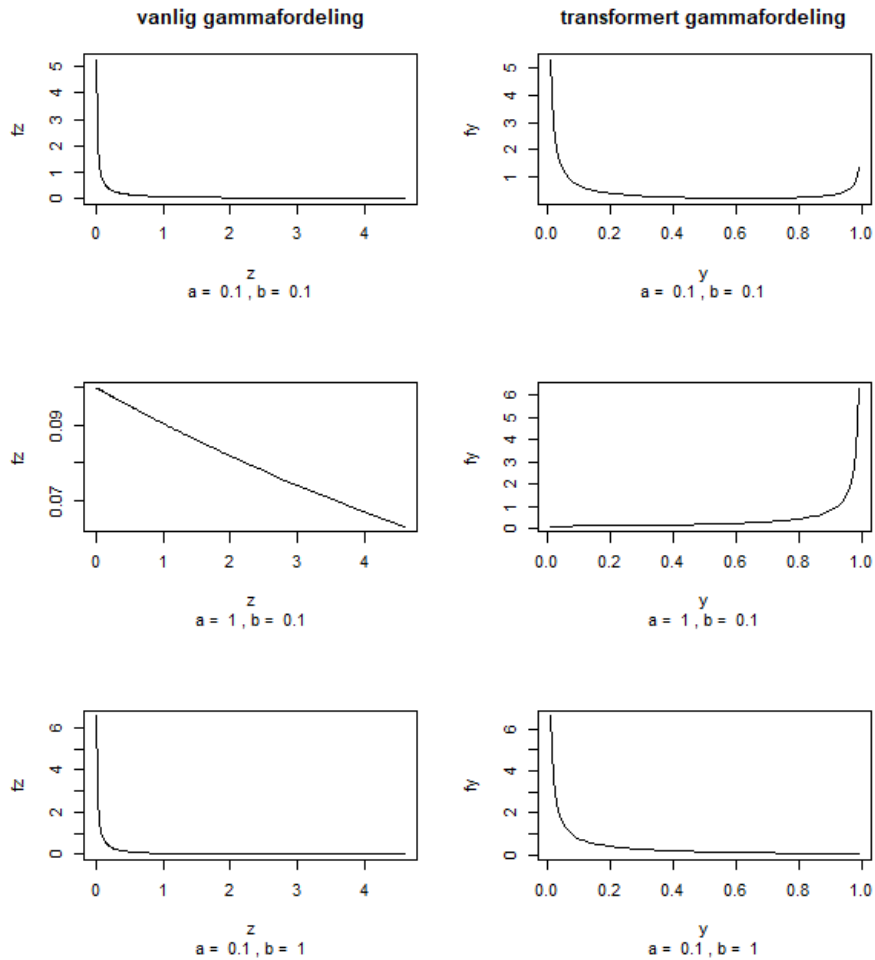
$$E(Y) = \int_0^{\infty} c(z)g(z, a, b), \quad \text{Var}(Y) = \int_0^{\infty} [c(z)]^2 g(z, a, b)$$

$$E(Y) = 1 - \left(\frac{b}{b+1}\right)^a \quad (3.8)$$

$$\text{Var}(Y) = \left(\frac{b}{b+2}\right)^a - \left(\frac{b}{b+1}\right)^{2a} \quad (3.9)$$

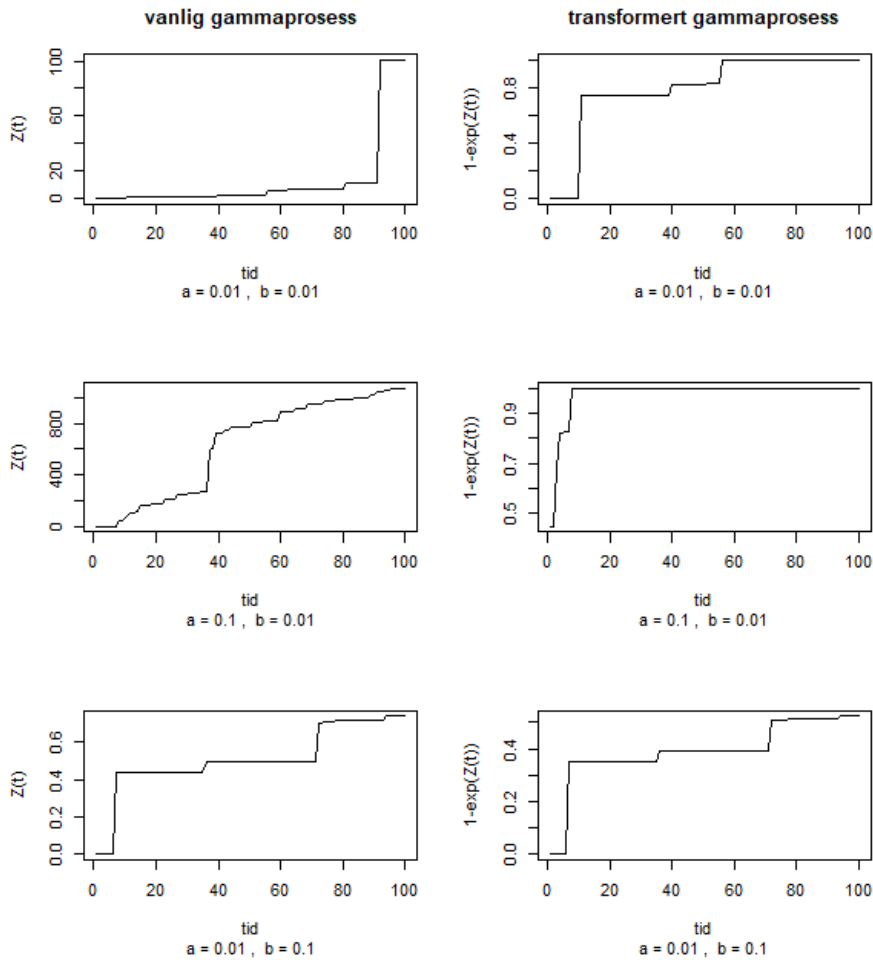
Vi legger merke til at siden $\frac{b}{b+1}$ går mot 1 når b vokser vil forventet verdi av Y bli mindre jo større b er. Siden $\frac{b}{b+1} < 1$ vil forventet verdi av Y vokse når a vokser. Variansen er ikke monotont økende i hverken b eller a .

Vi kan sammenligne sannsynlighetstettheten til Y og Z for ulike verdier av a, b .



Figur 3.1: Sannsynlighetstetthetene til en vanlig gammafordeling ($f_Z(z)$) og den transformerte fordelingen ($f_Y(y)$) evaluert for $y \in [0.01, 0.99]$ og $z = -\log(1-y)$

Vi ser at tettheten til det transformerte variabelet har en tendens til å ha større hale og at den er noe flatere enn tettheten til en vanlig gammafordelt variabel, gitt de samme parameterverdiene.



Figur 3.2: Simulerte gammaprosesser og transformasjon av de simulerte gammaprosessene. Tidssteget er $dt = 1$ og $t = 1, 2, 3, \dots, 100$. De transformerte gammaprosessene er $Y(t) = 1 - \exp(Z(t))$.

Vi ser at den transformerte gammaprosessen går raskere mot maksimum av støtten og så flater ut mer enn en vanlig gammaprosess, for de samme parameterverdiene.

Trunkert og transformert fordeling

Vi har altså en fordeling med støtte i $(0, 1)$, men som nevnt har vi definert at alle Ca^{2+} hendelser må være over en viss terskel $\tau \in (0, 1)$. Vi vet altså at alle observerte hendelser må være større enn vår terskel så vi må se på fordelingen til det trunkerte variabelen $Y|Y > \tau$. Tettheten til det trunkerte variabelen kan vi finne enkelt siden vi vet at den kumulative sannsynlighetsfordelingen

3.2. Transformert fordeling

til Y er $F_Y(y, a, b) = P(Y < y|a, b) = P(c(Z) < y|a, b) = P(Z < c^{-1}(y)|a, b) = G(c^{-1}(y), a, b)$, der $Z \sim \text{Gamma}(a, b)$. Så fra definisjon 3.1.3 har vi

$$f_{Y|Y>\tau}(y) = \frac{f_Y(y)\mathbb{I}(y > \tau)}{1 - F_Y(\tau, a, b)}$$

Da får vi

$$f_{Y|Y>\tau}(y|y > \tau) = \frac{b^a}{\Gamma(a, bc^{-1}(\tau))} (c^{-1}(y))^{a-1} (1-y)^{b-1} \mathbb{I}(y > \tau) \quad (3.10)$$

Vi kan finne forventning og varians til for det trunkerte variabelet ved

$$\begin{aligned} E(Y|Y > \tau) &= \int_0^\infty \frac{c(z)g(z, a, b)\mathbb{I}(z > c^{-1}(\tau))}{1 - G(c^{-1}(\tau), a, b)} \\ \text{Var}(Y|Y > \tau) &= \int_0^\infty \frac{[c(z)]^2 g(z, a, b)\mathbb{I}(z > c^{-1}(\tau))}{1 - G(c^{-1}(\tau), a, b)} \\ &\quad - [E(Y|Y > \tau)]^2 \end{aligned}$$

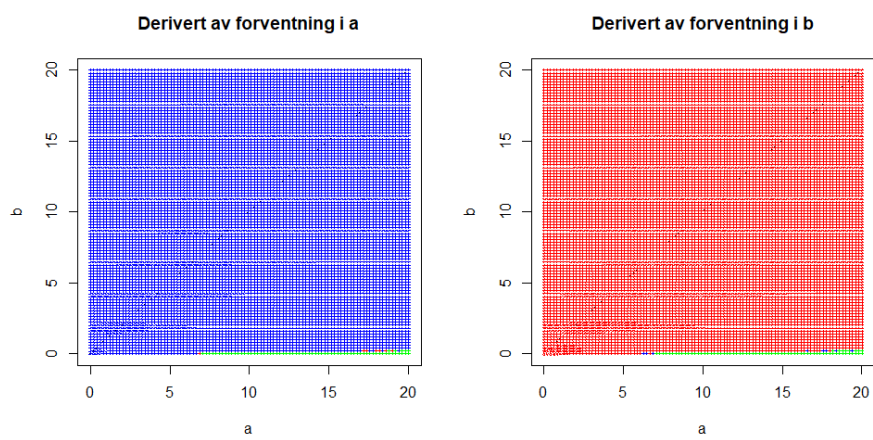
Vi får da at

$$E(Y|Y > \tau) = 1 - \left(\frac{b}{b+1}\right)^a \frac{\Gamma(a, (b+1)c^{-1}(\tau))}{\Gamma(a, bc^{-1}(\tau))} \quad (3.11)$$

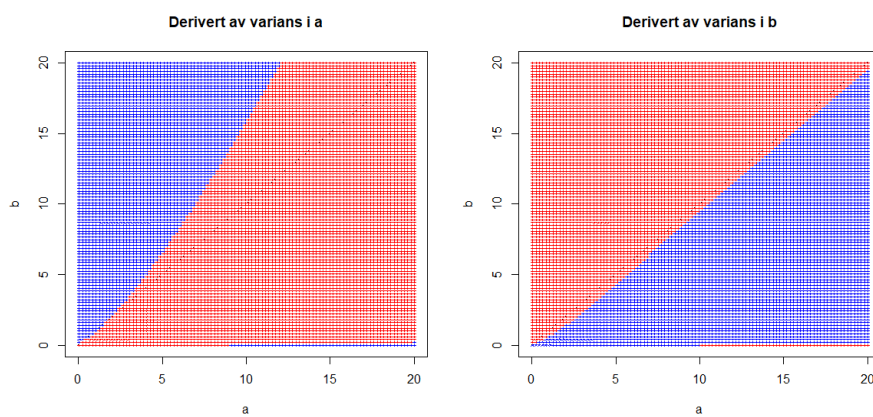
$$\begin{aligned} \text{Var}(Y|Y > \tau) &= \left(\frac{b}{b+2}\right)^a \frac{\Gamma(a, (b+2)c^{-1}(\tau))}{\Gamma(a, bc^{-1}(\tau))} \\ &\quad - \left(\frac{b}{b+1}\right)^{2a} \frac{\Gamma(a, (b+1)c^{-1}(\tau))^2}{\Gamma(a, (b)c^{-1}(\tau))^2} \end{aligned} \quad (3.12)$$

Det er ikke enkelt å finne den deriverte av forventningen på en analytisk form, men ved å fikserer terskelen, τ , kan vi se på ulike verdier av b at forventningen generelt øker i a , og vice-versa for ulike verdier av a vil forventningen generelt synke i b .

3.2. Transformert fordeling



Figur 3.3:



Figur 3.4:

KAPITTEL 4

Modellevaluering

Vi skal se på og evaluere noen varianter av de ulike modellene som vi har presentert. Vi estimerer parameterne i modellen, $\hat{\theta}$ hovedsakelig ved å minimalisere den negative log-likelihood funksjonen, $-\mathcal{L}(\mathbf{y}|\theta)$, med L-BFGS-B metoden [2] i `optim()`-funksjonen i R. Ved å estimere maksimum-likelihood estimatorene vet vi at vi finner parameterverdiene som minimaliserer AIC-verdien, et modellavlueringsverktøy som blir diskutert under. Samtidig kan vi bruke den asymptotiske fordelingen til maksimum-likelihood estimatorer $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow \mathcal{N}(0, \mathcal{I}^{-1})$, der $\hat{\theta}$ er den estimerte parameterverdien, θ_0 er den sanne parameterverdien og \mathcal{I} er Fisher informasjonsmatrisen, til å beregne konfidensintervall for parameterverdien og p-verdien for å kvantifisere hvor sannsynlig det er at vi hadde estimert den gitte parameterverdien, hvis den sanne parameterverdien er 0.

Kullback-Leibner divergens er et standard mål for avstanden mellom to sannsynlighetsfordelinger. Gitt to sannsynlighetsfordelinger P og Q definert over utfallsrommet \mathcal{X} , med sannsynlighetstetthet $p(x)$ og $q(x)$ for $x \in \mathcal{X}$, er Kullback-Leibner divergensen definert som $D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx$. Vi vet at maksimum likelihood estimatet for parameterne i en modell vil minimalisere Kullback-Leibner divergensen mellom modellen og den sanne datagenererende prosessen. Hvis vi hadde visst den sanne datagenererende prosessen kunne vi målt Kullback-Leibner divergensen mellom ulike kandidatmodeller og så valgt den modellen som minimaliserer Kullback-Leibner divergensen [3].

Vi kan ikke vite hva den sanne datagenererende prosessen er, men vi kan estimere hvor mye lengre unna den sanne datagenererende prosessen flere ulike kandidatmodeller er ved å bruke Akaike's information criterion (AIC). AIC er definert som $AIC = 2k - 2\log(\hat{\mathcal{L}})$, der k er antall parametere i modellen og $\hat{\mathcal{L}}$ er den maksimerte verdien av likelihoodfunksjonen. Modellen med lavest AIC-verdi vil da ha estimert lavest Kullback-Leibner divergens til den sanne datagenererende prosessen [3].

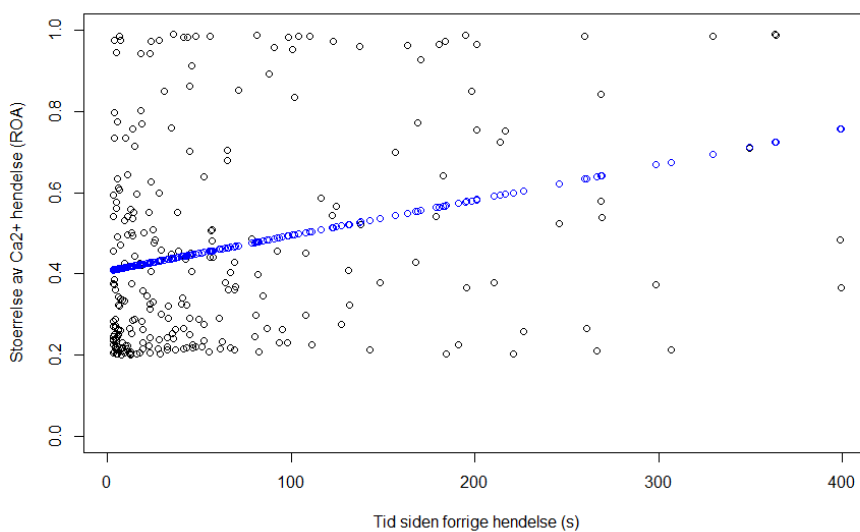
Modellen med lavest AIC-verdi er modellen med høyest verdi av den maksimerte likelihoodfunksjonen, korrigert for antall parametere. En annen vanlig måte å velge modell på ut ifra verdi av den maksimerte likelihoodfunksjonen, korrigert for antall parametere, er Bayesian information criterion (BIC), definert som $BIC = k \log(n) - 2\log(\hat{\mathcal{L}})$, der n er antall observerte datapunkter. Jo lavere BIC-verdien til en modell er, jo større estimert sannsynlighet er det at modellen

er den sanne datagenererende prosessen. BIC ligner på AIC, men for $n > 7$ vil BIC straffe modeller med flere parametere strengere enn AIC[3].

4.1 OLS-modell

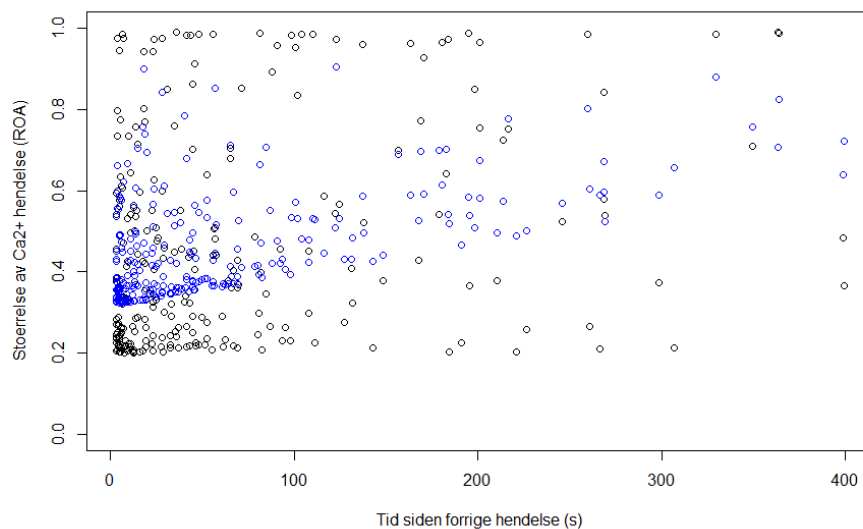
La oss først se på en enkel lineær regresjonsmodell modell som et utgangspunkt. La $Y_i = \beta_0 + \beta_t t_i + \epsilon_i$ for $i = 1, \dots, n$ der Y_i er størrelsen til den i -te Ca2+ hendelsen, $\epsilon_i \sim N(0, \sigma)$ og t_i er tiden mellom forrige hendelse og den i -te hendelsen (forrige hendelse er ikke den $(i-1)$ -te hendelse da den første hendelsen i opptaket ikke er inkludert i dataen).

Vi får at β blir estimert å være positiv så det er estimert positiv sammenheng mellom tid siden forrige hendelse og størrelse til den nåværende hendelse. Denne sammenhengen er robust for ulike definisjoner av terskel- og overlappverdien for hendelser. Estimaten for β er mer signifikant annerledes enn 0 når terskelverdien er mindre, siden det relativt blir flere små hendelser med kort tid siden forrige hendelse enn små hendelser med lang tid siden forrige hendelse.



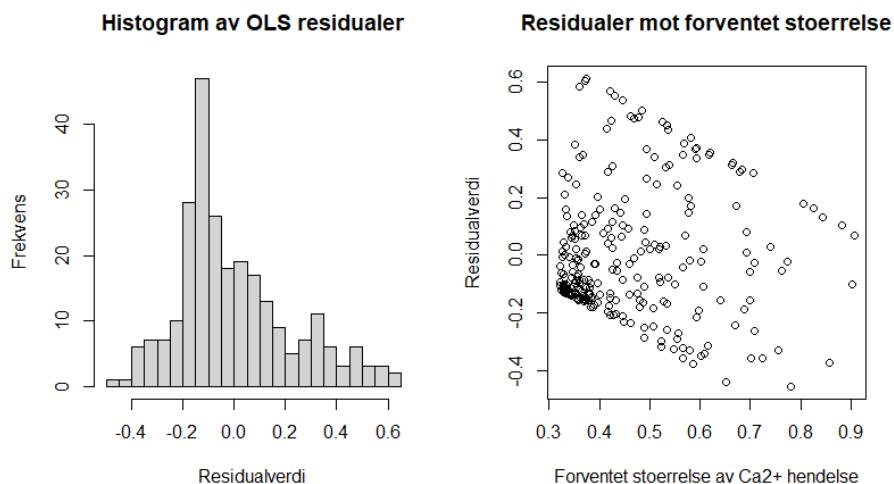
Figur 4.1: Plot av observert størrelse til hendelse opp mot tid siden forrige hendelse i sort. Tilpassede verdier i blå.

Vi kan også kontrollere for løping i hjul hvis vi lar $Y_i = \beta_0 + \beta_t t_i + \beta_w w_i + \epsilon_i$, der w_i er hjulfarten for den i -te hendelsen og $\epsilon_i \sim N(0, \sigma)$.



Figur 4.2: Observerte størrelser av Ca²⁺ hendelser i sort og tilpasset størrelse av hendelsene, kontrollert for hjulfart i blått

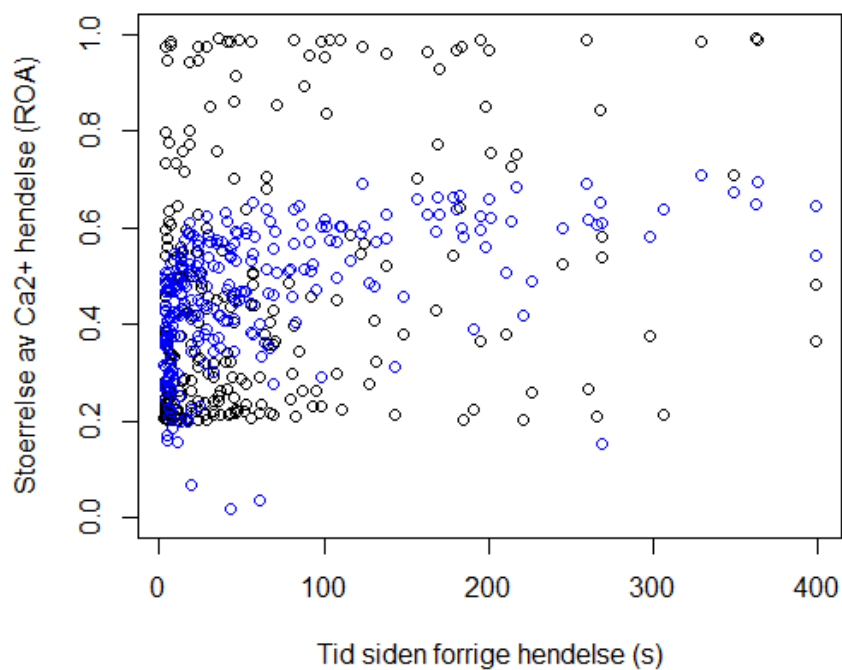
OLS modellen med hjul-fart har $AIC = -35.22$ og $BIC = -17.51$. Vi tror nok ikke at denne enkle OLS modellen er en god måte å modellere sammenhengen mellom størrelse på Ca²⁺ hendelser og tid mellom hendelsene. Vi ser det direkte på et plot av data at sammenhengen ikke er særlig lineær, og vi kan også se på residualene fra OLS modellen at modellen er noe feilspesifisert.



Figur 4.3: a) Histogram av OLS residualer og b) sammenhengen mellom residualer og størrelse på Ca²⁺ hendelse

For en OLS modell antar man at residualene er normalfordelt med forventning 0 og konstant varians $\epsilon \sim N(0, \sigma)$. Vi ser fra histogrammet av residualene at ikke ser ut som de er normalfordelt med forventning 0, og vi ser fra plottet i b) at variansen ikke ser ut til å være konstant heller.

Siden forholdet mellom størrelse av Ca²⁺ hendelsene og kovariatene er åpenbart ikke-lineært er det fristende å gjøre en ikke-lineær transformasjon av kovariatene. Siden det er mange mindre hendelser med kort tid siden forrige hendelse og lav hjulfart er det naturlig å bruke en log-transformasjon. Vi ser på modellen $Y_i = \beta_0 + \beta_t \log(t_i) + \beta_w \log(w_i + 0.001) + \epsilon_i$. Siden det er mange hendelser der hjulfarten er 0 må man legge til en liten korrigering i observert hjulfart for å kunne log-transformere.



Figur 4.4: Observerte størrelser av Ca²⁺ hendelser i sort og tilpassede størrelser fra OLS modell med logtransformerte kovariater i blått

Vi ser at modellen til en viss grad passer godt for små Ca²⁺ hendelser med kort tid siden forrige hendelse, men ikke passer godt for de store hendelsene. I tillegg forventer modellen at noen hendelser er mindre enn terskelen vi har satt for å definere Ca²⁺ hendelser. AIC og BIC verdiene er marginalt bedre enn OLS modellen med ikke-transformerte kovariater med AIC = -38.86 og BIC = -24.70. Resultatet blir omtrent likt hvis man kun logtransformerer tid siden forrige hendelse.

4.2 Enkel gammaprosess

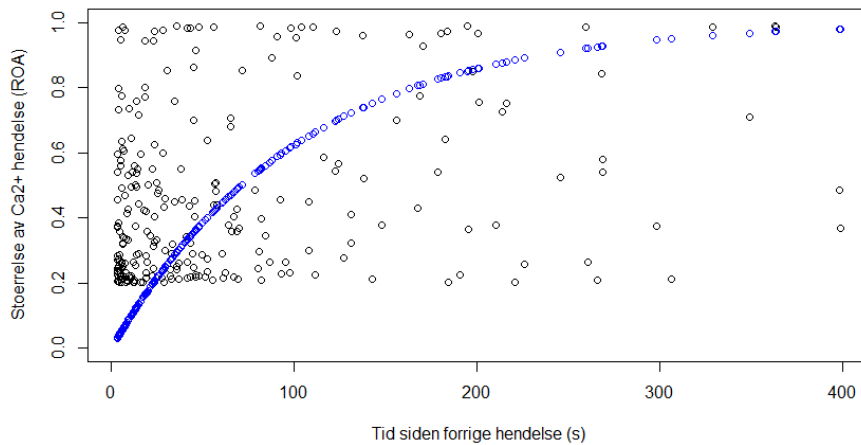
Modell 1.0

La oss først se på den aller enkleste gammaprosess modellen der $Z_i \sim \text{Gamma}(at_i, b)$ og vi lar $Y_i = 1 - \exp(-Z_i)$. Denne modellen har log-likelihoodfunksjon

$$\log \mathcal{L} = \sum_{i=1}^n [at_i \cdot \log(b) - \log(\Gamma(at_i)) + (b-1)\log(1-y_i) + (at_i-1)\log(-\log(1-y_i))]$$

Vi finner $(\hat{a}, \hat{b}, \log \hat{\mathcal{L}})$ numerisk og kan finne forventet verdi for alle hendelser med uttrykket for forventning fra kapittel 3.

$$E(Y_i) = 1 - \left(\frac{\hat{b}}{\hat{b}+1}\right)^{\hat{a}t_i}$$



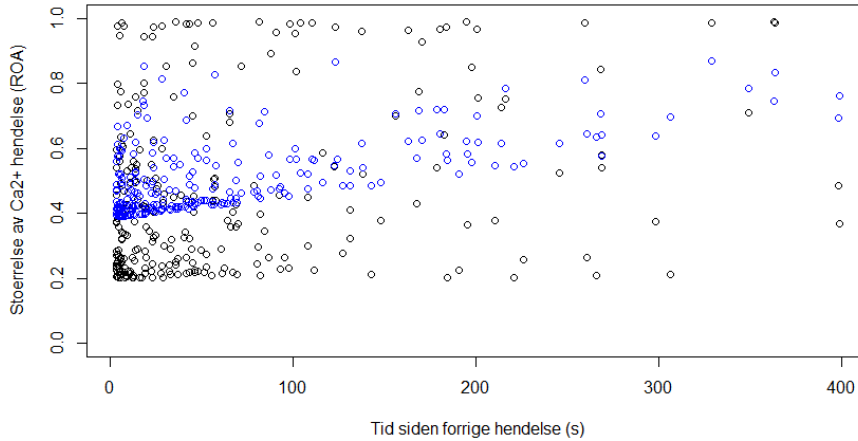
Figur 4.5: Observerte størrelser av Ca²⁺ hendelser (sort) og tilpassede størrelser (blått)

Vi ser den er noe feil ved første øyekast, siden forventet størrelse er mindre enn terskelen vi bruker for å definere Ca²⁺ hendelser når det er kort tid siden forrige hendelse. AIC og BIC verdiene er også betydelig dårligere enn OLS modellene med AIC = 44.49 og BIC = 447.57.

Modell 1.1

Vi kan prøve å utvide modellen for å ta hensyn til hjulfart også. Da lar vi igjen $Y_i = 1 - \exp(-Z_i)$, men denne gangen utvider vi formparameteren slik at $Z_i \sim \text{Gamma}(a \cdot \exp(\beta_i t_i + \beta_w w_i), b)$, der w_i er hjulfarten i den i -te hendelsen. Denne modellen har samme loglikelihoodfunksjon og forventning som den forrige modellen, bortsett fra at at_i er byttet ut med den utvidede

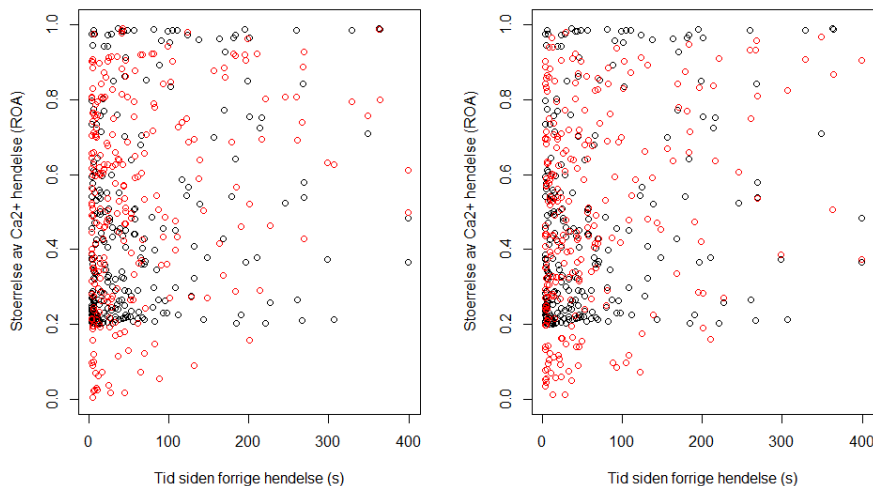
formparameteren $a \cdot \exp(\beta_t t_i + \beta_w w_i)$. Vi kan se på forventede verdier av størrelsen til Ca²⁺ hendelser i denne modellen opp mot faktiske størrelse.



Figur 4.6: Observerte størrelser av Ca²⁺ hendelser (sort) og tilpassede størrelser fra modell 1.1 (blått)

Her ser vi at forventet verdi passer en del bedre med observert data. AIC og BIC er også en del bedre enn OLS modellene med $AIC = -50.73$ og $BIC = -36.57$. Vi estimerer samme antall parametere i denne modellen og OLS-modellen, her har vi $\theta = (a, b, \beta_t, \beta_w)$ og i OLS-modellen hadde vi $\theta = (\beta_0, \beta_t, \beta_w, \sigma)$. Det er altså log-likelihood funksjonen som har en noe høyere verdi enn i OLS-modellene, noe som ikke er overraskende med tanke på hvor ikke-linær sammenhengen mellom tid siden forrige hendelse og størrelse på hendelse er. I motsetning til OLS-modellen er her vi også begrenset støtten til Y_i , fra $(-\infty, \infty)$ i OLS-modellen til $(0, 1)$ i denne gammaprosessmodellen, noe som også hjelper. Vi estimerer at $\hat{\beta}_t = 2.14$. Vi kan konstruere et 95% konfidensintervall for β_t og får $(1.45, 2.83)$. Vi estimerer altså i en vanlig gammaprosess-modell at det er en tydelig positiv sammenheng mellom tid fra forrige Ca²⁺ hendelse og størrelsen på den nåværende.

Selv om forventede verdier for størrelse av Ca²⁺ hendelser er større enn terskelen for Ca²⁺ hendelser, vil det fortsatt i denne modellen være en positiv sannsynlighet for at størrelsen til hendelser er mindre enn terskelen. Vi kan enkelt simulere fra modellen vår ved å bruke maksimum likelihood estimatene for parameterne til å simulere verdier fra en vanlig gammafordeling, $\tilde{Z}_i \sim \text{Gamma}(\hat{a} \cdot \exp(\hat{\beta}_t t_i + \hat{\beta}_w w_i), \hat{b})$, og så transformere de simulerte verdiene med transformasjonen $\tilde{Y}_i = 1 - \exp(-\tilde{Z}_i)$.



Figur 4.7: Hvert plott viser observerte data i sort, og simulert størrelse gitt kovariater i rødt

Vi ser at modellen fortsatt kan simulere at noen Ca2+ hendelser er mindre enn terskelen.

4.3 Trunkert gammaprosess

Den trunkerte gammaprosess-modellen vi presenterte i kap. 3 kan kanskje takle utfordringen med å komme nærmere den ekte datagenerende prosessen noe bedre.

modell 2.0

Da vil $Z'_i \sim \text{TrunkertGamma}(a, 1, c^{-1}(0.2))$, og $Y_i = 1 - \exp(-Z'_i)$. Denne modellen, med konstant forventning og varians har AIC = -200.74 og BIC = -197.20. Hvis vi lar rateparameteren variere får vi AIC = -228.19 og BIC = -221.10.

	AIC	BIC	\hat{a}, \hat{b}
a,b	-228.18	-221.10	<0.001, 0.570
a, 1	-200.74	-197.20	0.167, 1

Man kan utvide formlen i flere retninger. Hvis man fortsetter i et typisk gammaprosessrammeverk vil det være naturlig å la formlen være $\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) t_i$, som vil tilsvare $\exp(\beta_0 + \log(t_i) + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$, der X_i for $i = 1, 2, \dots, p$ er kovariater. Man kan også la tid bli eksponensiert på samme måte som de andre kovariatene og påvirke fordelingen gjennom en regresjonskoeffisient, da vil formlen være

4.3. Trunkert gammaprosess

$\exp(\beta_0 + \beta_i t_i + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$. Det kan også være at det lønner seg å la tiden siden forrige hendelse påvirke fordelingen gjennom en regresjonskoeffisient, men også la tid være på samme skala som i det første alternativet, eksponensielt og logtransformert, slik at formparameteren blir $\exp(\beta_0 + \beta_i \log(t_i) + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$, som vil tilsvare $t_i^{\beta_i} \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p)$. Så lenge alle kovariater har bare positive verdier og effekten av kovariatet på størrelsen av Ca2+ hendelser er positiv kan man også la formparameteren være $\beta_0 + \beta_i t_i + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$. Hvis verdiene eller effekten av et kovariat ikke er positive, kan man transformere kovariatet.

modell 2.1.1

Vi starter med den enkleste utvidelsen av formparameteren $Z'_i \sim \text{TrunkertGamma}(at_i, b, c^{-1}(0.2))$, og $Y_i = 1 - \exp(-Z'_i)$. Merk at at_i tilsvarende $\exp(\log(a) + \log(t_i))$, så $\log(a)$ er det samme som β_0 for det første alternativet for å utvide formparameteren vi beskrev ovenfor.

Loglikelihoodfunksjonen for denne modellen er

$$\log \hat{\mathcal{L}} = \sum_{i=1}^n [at_i \log(b) - \log(\Gamma(at_i, bc^{-1}(0.2)))] \\ + (at_i) \log(c^{-1}(y_i)) + (b-1) \log(1 - y_i)]$$

Der vi husker at $c^{-1}(\tau) = -\log(1 - \tau)$ fra 3.6. Forventet størrelse av hver hendelse under denne modellen kan finnes ved å modifisere uttrykket fra 3.11

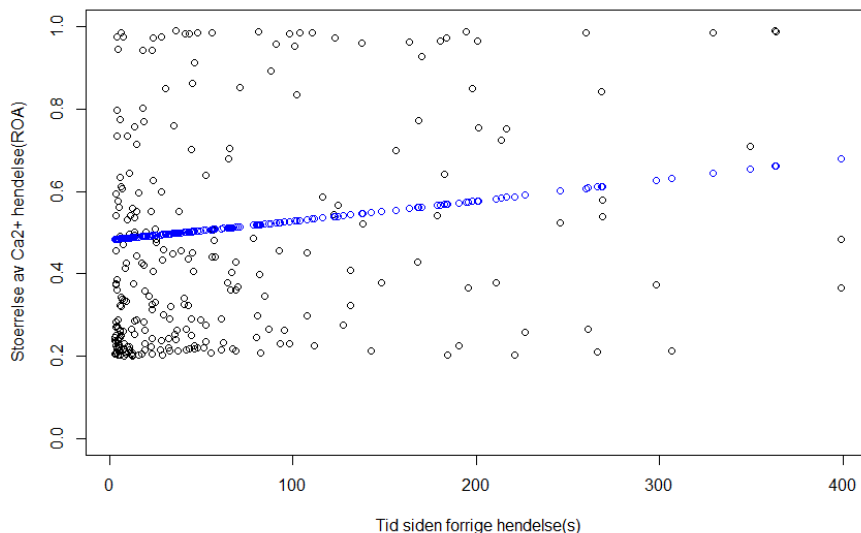
$$E(Y_i) = 1 - \left(\frac{b}{b+1}\right)^{at_i} \frac{\Gamma(at_i, (b+1)c^{-1}(\tau))}{\Gamma(at_i, bc^{-1}(\tau))}$$

I R kan man beregne verdien av den øvre ukomplette gammafunksjonen ved å huske at $\Gamma(a, b\tau) = \Gamma(a)(1 - G(\tau, a, b))$, så $\Gamma(a, b\tau) = \text{gamma}(a) * \text{pgamma}(\tau, a, \text{rate}=b, \text{lower.tail}=\text{FALSE})$.

For denne modellen har vi

	at_i, b
AIC	-236.28
BIC	-229.19
\hat{a}, \hat{b}	2.654, 0.708
95% KI for \hat{a}	(1.173, 4.136)
p-verdi for \hat{a}	<0.0002

Vi finner de tilpassede verdiene for størrelsene til Ca2+ hendelser ved å bruke maksimum-likelihood estimatene for parameterne i formelen for forventningen.



Figur 4.8: Tilpassede størrelsesverdier fra modell 2.1.1 i blått

Denne modellen ser på øyemål ut som den passer observerte data en god del bedre enn de ikke-trunkerte gammaprosessmodellene og AIC og BIC verdiene bekrefter dette. Vi får $AIC = -236.28$ og $BIC = -229.19$. Vi husker at AIC og BIC for modellen der tid ikke påvirker formparameteren og rateparameteren var fri var -228.19 og -221.10 , så å la tid påvirke formparameteren på denne måten gir en noe bedre modell, men ikke dramatisk bedre. Punktestimater for formparameteren er $\hat{a} = 2.65$ og 95% konfidensintervall for formparameteren er $(1.17, 4.14)$. Det er samtidig vanskelig å argumentere for at siden konfidensintervallet ikke dekker null og p-verdien til \hat{a} er lav så er det en sammenheng mellom tid og størrelse, siden vi husker at vi må ha $a > 0$ for at $P(Y > 0) > 0$. Det er viktigere for oss at AIC og BIC gir noe lavere verdier enn grunnmodellen, som tyder på at modellen med tid inkludert passer bedre med data.

modell 2.1.2

La oss utvide denne modellen ved å ta hensyn til hjulfart i formparameteren. Da vil $Z'_i \sim \text{TrunkertGamma}(\exp(\beta_w w_i) a t_i, b, c^{-1}(0.2))$, og $Y_i = 1 - \exp(-Z'_i)$, der formparameteren tilsvarer $\exp(\beta_0 + \beta_w w_i) t_i$ hvor $\beta_0 = \log(a)$.

vi har log-likelihood funksjonen

$$\log \hat{\mathcal{L}} = \sum_{i=1}^n [\exp(\beta_w w_i) a t_i \log(b) - \log(\Gamma(\exp(\beta_w w_i) a t_i, b c^{-1}(0.2)))]$$

4.3. Trunkert gammaprosess

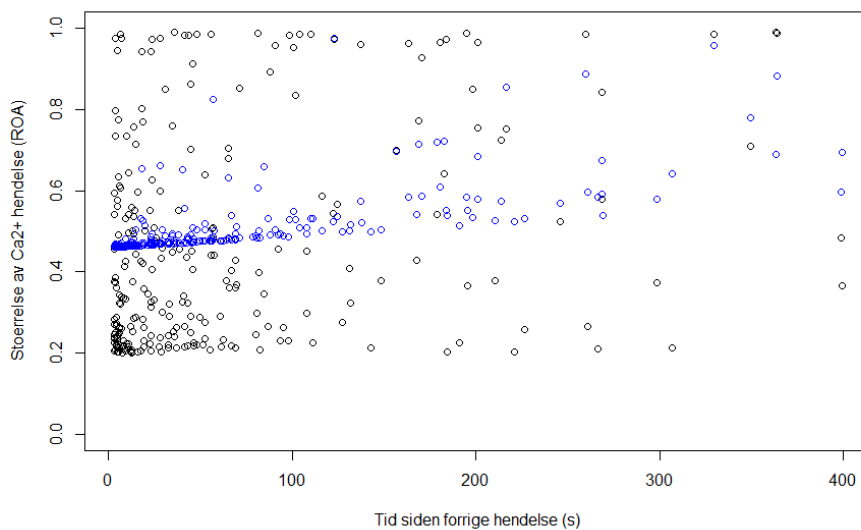
$$+(\exp(\beta_w w_i) a t_i) \log(c^{-1}(y_i)) + (b - 1) \log(1 - y_i)]$$

Forventet størrelse av hendelser vil ligne på den forrige, men med noen avvik for hendelser der hjulfarten var stor.

$$E(Y_i) = 1 - \left(\frac{b}{b+1}\right)^{\exp(\beta_w w_i) a t_i} \frac{\Gamma(\exp(\beta_w w_i) a t_i, (b+1)c^{-1}(\tau))}{\Gamma(\exp(\beta_w w_i) a t_i, bc^{-1}(\tau))}$$

For denne modellen får vi

$\exp(\beta_w w_i) a t_i, b$	
AIC	-264.68
BIC	-254.05
$\hat{a}, \hat{b}, \hat{\beta}_w$	1.840, 0.867, 0.126
95% KI for \hat{a}	(1.306, 2.373)
p-verdi for \hat{a}	<1e-11



Figur 4.9: Tilpassede størrelsesverdier fra modell 2.1.2 i blått

For denne modellen har vi at $AIC = -264.68$ og $BIC = -254.05$, så denne modellen passer betydelig bedre enn uten hjulfart. Særlig er inkluderingen av hjulfart i modellen egnet til å fange opp de store Ca^{2+} hendelsene der hjulfarten er stor, men fortsatt kan vi ikke stole på at konfidensintervallet og signifikanstesten av \hat{a} representerer den faktiske sammenhengen mellom tid og størrelse av hendelse, siden vi fortsatt må ha $a > 0$.

modell 2.2.1

Vi har foreløpig vist at denne typer transformerte og trunkerte gammamodeller kan forbreides ved å ta hensyn til tid og at den enkelt kan utvides for å ta hensyn til andre kovariater som gjør modellen bedre, men vi har ennå ikke formelt testet signifikansen av sammenhengen mellom tid og størrelse av Ca2+ hendelser. Denne sammenhengen er knyttet til et spørsmål vi startet denne oppgaven med, nemlig om astrocytter har en type *relativ* refraktærperiode på samme måte som nevroner. Vi kan ikke teste direkte om det finnes slik relativ refraktærperiode, men hvis sammenhengen mellom tid og størrelse er sterkt positiv vil det i det minste underbygge påstanden at astrocytter også har en relativ refraktærperiode.

En naturlig modell for å kunne teste sammenhengen mellom tid og størrelse til Ca2+ hendelse, hvis vi fortsetter i samme rammeverk som til nå, vil være å la formparameteren være på formen $\exp(\beta_0 + \beta_t t_i)$. Merk at det ikke hjelper å la formparameteren være på formen $\beta_0 + \beta_t t_i$, siden vi fortsatt må ha $\beta_0 + \beta_t t_i > 0$ og β_0 vil neppe være stor nok å la β_t variere fritt. I tillegg har modeller med denne typen formparameter lavere AIC og BIC verdi enn den type modell vi nettopp gikk gjennom.

La oss derfor se på modellen der

$$Z'_i \sim \text{TrunkertGamma}(\exp(\beta_0 + \beta_t t_i), b, c^{-1}(0.2)),$$

og $Y_i = 1 - \exp(-Z'_i)$. Denne modellen har loglikelihood-funksjonen

$$\begin{aligned} \log \hat{\mathcal{L}} = \sum_{i=1}^n & [\exp(\beta_0 + \beta_t t_i) \log(b) - \log(\Gamma(\exp(\beta_0 + \beta_t t_i), bc^{-1}(0.2)))] \\ & + (\exp(\beta_0 + \beta_t t_i)) \log(c^{-1}(y_i)) + (b-1) \log(1-y_i) \end{aligned}$$

Forventning til hver enkelt Ca2+ hendelse vil være

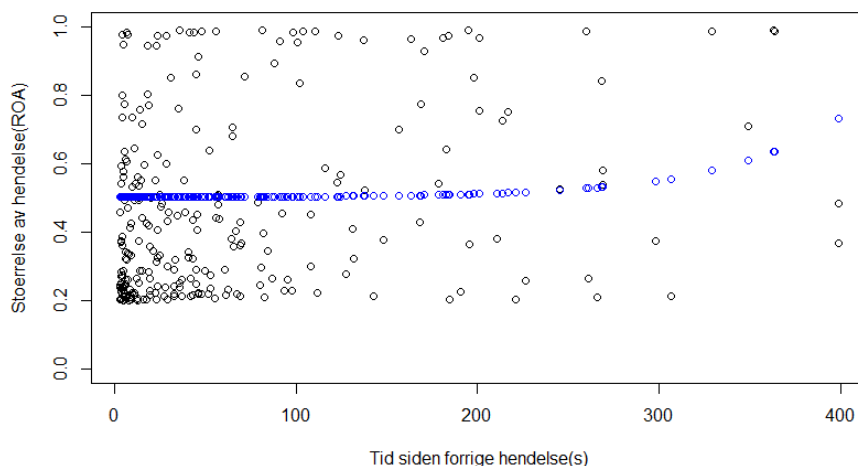
$$E(Y_i) = 1 - \left(\frac{b}{b+1}\right)^{\exp(\beta_0 + \beta_t t_i)} \frac{\Gamma(\exp(\beta_0 + \beta_t t_i), (b+1)c^{-1}(\tau))}{\Gamma(\exp(\beta_0 + \beta_t t_i), bc^{-1}(\tau))}$$

Vi får for denne modellen

	$\exp(\beta_0 + \beta_t t_i), b$
AIC	-229.72
BIC	-219.09
$\hat{\beta}_0, \hat{\beta}_t, \hat{b}$	-5.809, 14.93, 0.599
95% KI for $\hat{\beta}_t$	(12.12, 17.75)
p-verdi for $\hat{\beta}_t$	<1e-25

I denne modellen tyder signifikanstesten og konfidensintervall for $\hat{\beta}_t$ på at det er en sammenheng mellom tid og størrelse for Ca2+ hendelser, men AIC og BIC verdiene er noe høyere enn modellen der formparameteren var at_i . Vi legger også merke til at verdien for $\hat{\beta}_0$ er veldig lav, så selv om den sammenhengen mellom

tid og størrelse er signifikant positiv så er det ikke en særlig sterk sammenheng. At sammenhengen ikke er særlig sterk underbygges ved å se på tilpassede verdier av størrelse for denne modellen. I tillegg er loglikelihood-funksjonen til denne modellen ustabil rundt maksimum og verdien til funksjonen, samt den estimerte hessematrisen, bryter sammen hvis grenseverdiene for parameterene ikke justeres. Justeringen av grenseverdiene for parameterne påvirker da også AIC og BIC verdiene.

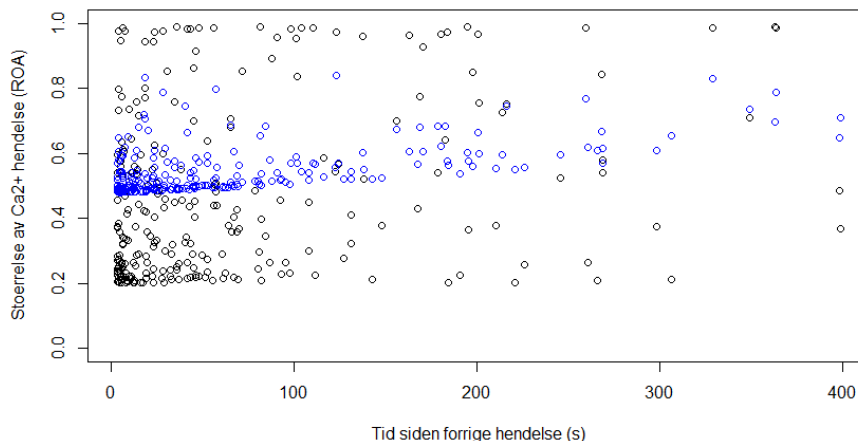


Figur 4.10: Tilpassede størrelsesverdier fra modell 2.2.1 i blått

modell 2.2.2

Vi kan også utvide formparameteren for å kontrollere for hjulfart slik at $Z'_i \sim \text{TrunkertGamma}(\exp(\beta_0 + \beta_t t_i + \beta_w w_i), b, c^{-1}(0.2))$ og $Y_i = 1 - \exp(-Z'_i)$. For denne modellen får vi

$\exp(\beta_0 + \beta_t t_i + \beta_w w_i), b$	
AIC	-258.21
BIC	-244.04
$\hat{\beta}_0, \hat{\beta}_t, \hat{\beta}_w, \hat{b}$	-3.176, 7.556, 0.836, 0.143
95% KI for $\hat{\beta}_t$	(6.313, 8.800)
p-verdi for $\hat{\beta}_t$	<1e-33



Figur 4.11: Tilpassede størrelsesverdier fra modell 2.2.2 i blått

Vi ser at, som tidligere når vi har inkludert hjulfart, vil tilpasset verdi for en del hendelser være større. AIC og BIC verdiene er litt større enn for den tidligere modellen med hjulfart, med $AIC = -258.21$ og $BIC = -244.04$. Sammenhengen mellom tid siden forrige hendelse og størrelsen til nåværende hendelse estimerer vi at er signifikant positiv, men som den forrige modellen sørger den lave verdien til $\hat{\beta}_0$ for at effekten ikke er særlig stor.

Vi legger merke til at modellene med formparameter av typen $\exp(\beta_0 + \dots)t_i$ har lavere AIC og BIC enn modeller med tilsvarende kovariater og formparameter av typen $\exp(\beta_0 + \beta_t t_i \dots)$, som kan tyde på to ting. Enten gjør ikke den ekstra parameteren, β_t , ikke nok for å heve loglikelihoodet i forhold til straffen AIC og BIC gir for en ekstra parameter, eller så lønner det seg ikke å eksponensiere tidsvariabelen. Vi kan kompensere for det siste ved å ta logaritmen av tid innenfor eksponensieringen, dvs at formparameteren blir av typen $\exp(\beta_0 + \beta_t \log(t_i) \dots)$. Vi ser at $\exp(\beta_0)t_i = \exp(\beta_0 + \log(t_i) \dots)$, så $\exp(\beta_0)t_i$ tilsvarer $\exp(\beta_0 + \beta_t \log(t_i) \dots)$, der vi fikserer $\beta_t = 1$. Hvis vi altså lar β_t variere kan vi både teste om sammenhengen mellom tid og størrelse til Ca^{2+} hendelser er signifikant, samt at vi (muligens) kan få bedre modeller siden tidsvariabelen vil være på en mer naturlig skala.

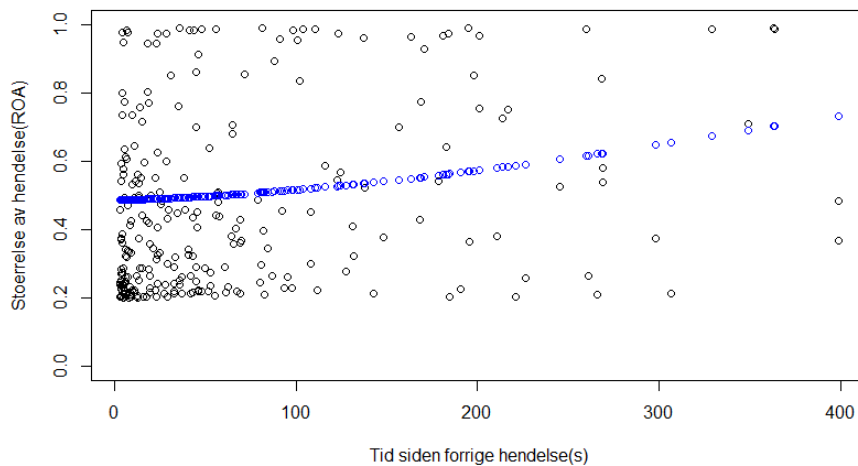
modell 2.3.1

Vi ser først på modellen der $Z'_i \sim \text{TrunkertGamma}(\exp(\beta_0 + \beta_t \log(t_i)), b, c^{-1}(0.2))$, og $Y_i = 1 - \exp(-Z'_i)$.

4.3. Trunkert gammaprosess

$\exp(\beta_0 + \beta_t \log(t_i)), b$	
AIC	-235.63
BIC	-225.01
$\hat{\beta}_0, \hat{\beta}_t, \hat{b}$	1.590, 1.443, 0.684
95% KI for $\hat{\beta}_t$	(1.058, 1.827)
p-verdi for $\hat{\beta}_t$	<1e-13

Så modellen med en egen parameter for tid gjør at AIC og BIC verdiene blir marginalt høyere enn i modell 2.1.1 og modellen er derfor noe dårligere tilpasset data, samtidig er ikke AIC og BIC verdiene markant større og det kan hende at det i modell 2.1.1 tilfeldigvis passet rimelig godt at $\hat{\beta}_t$ er 1. Viktigere er det at vi estimerer at sammenhengen mellom tid og størrelse er positiv og p-verdien til denne parameteren er lav. De tilpassede verdiene for størrelse av Ca2+ hendelser ligner på det vi fikk fra modell 2.1.1



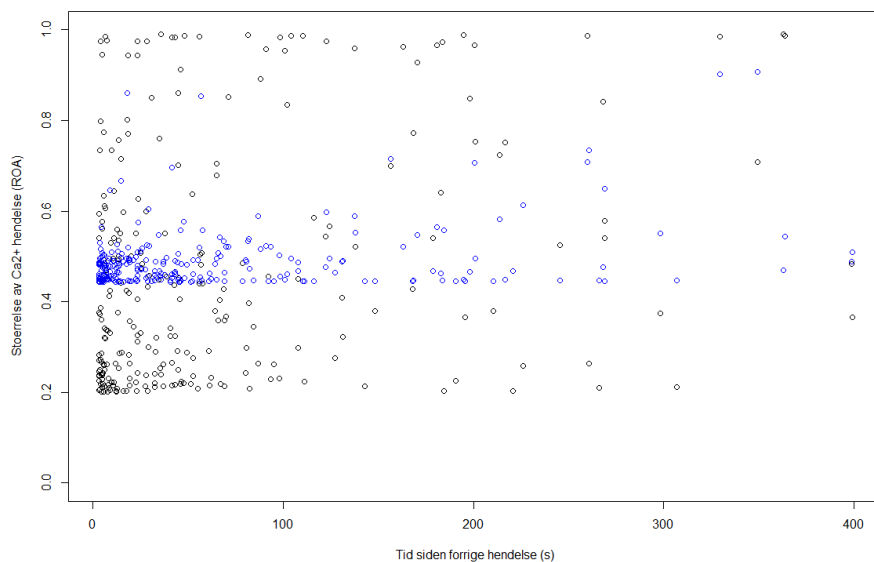
Figur 4.12: Tilpassede verdier for størrelse fra modell 2.3.1 i blått

modell 2.3.2

Vi kun utvide modellen til å kontrollere for hjulfart, slik at formparameteren blir $Z'_i \sim \text{TrunkertGamma}(\exp(\beta_0 + \beta_t \log(t_i) + \beta_w w_i), b, c^{-1}(0.2))$, og $Y_i = 1 - \exp(-Z'_i)$.

$\exp(\beta_0 + \beta_t \log(t_i) + \beta_w w_i), b$	
AIC	-262.73
BIC	-248.56
$\hat{\beta}_0, \hat{\beta}_t, \hat{\beta}_w, \hat{b}$	0.562, 0.949, 0.124, 0.876
95% KI for $\hat{\beta}_t$	(0.804, 1.094)
p-verdi for $\hat{\beta}_t$	<1e-38

Igen får vi analoge resultater for denne modellen i forhold til modell 2.1.2, som for 2.3.1 i forhold til 2.1.1. Modelltilpassningen er noe dårligere, men sammenhengen mellom tid og størrelse er estimert til å være positiv. Tilpassede verdier av størrelsene ligner på det vi har sett tidligere.



Figur 4.13: Tilpassede størrelsesverdier fra modell 2.3.2 i blått

Som nevnt i kapittel 1 er det vanlig innenfor modellering av interspikeintervaller for nevroner å ta hensyn til størrelsen av forrige intervall. Jo mindre forrige intervall var, jo større forventer man at neste intervall blir. I vår kontekst kan man tenke seg at jo større forrige Ca²⁺ hendelse var, jo mindre vil neste Ca²⁺ hendelse være. En enkel måte å estimere effekten av størrelsen av forrige Ca²⁺ hendelse i vårt rammeverk vil være å la observert størrelse av forrige Ca²⁺ hendelse være en del av en utvidet formparameter. Merk at på grunn av måten datasettet vårt er strukturert på, den første hendelsen i et ROA-opptak er ikke med, vil alle Ca²⁺ hendelser i vårt datasett kunne bruke størrelsen av forrige Ca²⁺ hendelse som kovariat.

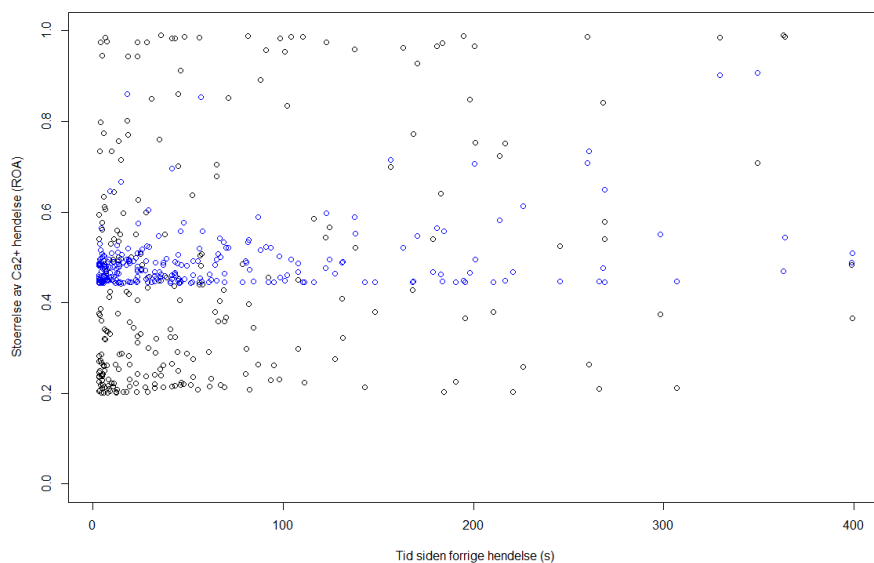
modell 2.3.3

Vi lar da $Z'_i \sim \text{TrunkertGamma}(\exp(\beta_0 + \beta_t \log(t_i) + \beta_w w_i + \beta_s s_i), b, c^{-1}(0.2))$, der s_i er den observerte størrelsen av Ca²⁺ hendelsen før den i -te hendelsen, og $Y_i = 1 - \exp(-Z'_i)$.

4.3. Trunkert gammaprosess

$\exp(\beta_0 + \beta_t \log(t_i) + \beta_w w_i + \beta_s s_i), b$	
AIC	-261.99
BIC	-244.29
$\hat{\beta}_0, \hat{\beta}_t, \hat{\beta}_w, \hat{\beta}_s, \hat{b}$	0.905, 0.950, 0.121, -0.568, 0.886,
95% KI for $\hat{\beta}_t$	(0.807, 1.092)
p-verdi for $\hat{\beta}_t$	<1e-40
95% KI for $\hat{\beta}_s$	(-1.094,-0.041)
p-verdi for $\hat{\beta}_s$	0.015

Å inkludere størrelsen av forrige hendelse gir noe dårligere modelltilpassning, men sammenhengen mellom størrelsen av forrige hendelse og størrelse av neste går i den retningen vi forventet. Vi estimerer at jo større forrige hendelse var, jo mindre vil forventningen til neste hendelse være, men denne sammenhengen er ikke særlig sterk og det vi blir ikke overbevist om at den er veldig forskjellig fra 0. Tilpassede verdier av størrelse fra denne modellen ligner også på det vi tidligere har sett.



Figur 4.14: Tilpassede verdier for størrelse fra modell 2.3.3 i blått

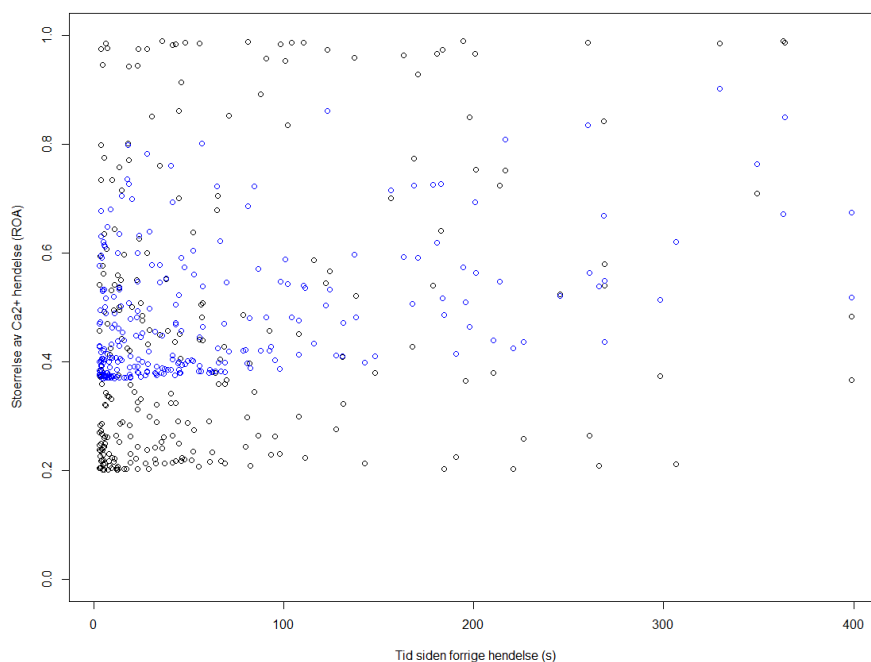
modell 3.0

Siden vi vet at i en vanlig gammaprosessmodell vil form- og rateparameteren påvirke varians og forventning ulikt er det interessant å se hvordan modellen blir hvis vi lar hjulfart og størrelsen til forrige hendelse påvirke rateparameteren.

4.3. Trunkert gammaprosess

$\exp(\beta_0 + \beta_t \log(t_i)), (\exp(\beta_1 + \beta_w w_i + \beta_s s_i))^{-1}$	
AIC	-290.03
BIC	-272.32
$\hat{\beta}_0, \hat{\beta}_t, \hat{\beta}_1, \hat{\beta}_w, \hat{\beta}_s$	1.299, 1.200, -0.731, 0.181, 0.012,
95% KI for $\hat{\beta}_t$	(0.858, 1.541)
p-verdi for $\hat{\beta}_t$	<1e-12
95% KI for $\hat{\beta}_s$	(-0.330, 0.354)
p-verdi for $\hat{\beta}_s$	0.472

Vi får da en betydelig bedre AIC og BIC score, så denne fremgangsmåten kan lønne seg videre. Sammenhengen mellom tid og størrelse er omtrent den samme som de tidligere modellene, men sammenhengen mellom størrelsen av forrige hendelse og størrelsen til neste er tilnærmet 0. Vi ser også for denne typen modell at det gir bedre tilpassning å ikke inkludere størrelse av forrige hendelse, uten β_s er AIC = -292.03, BIC = -277.86. Tilpassede verdier for denne modellen ligner på de tidligere modellene, men det betydelig større spredning.



Figur 4.15: Tilpassede verdier for størrelse fra modell 3.0 i blått

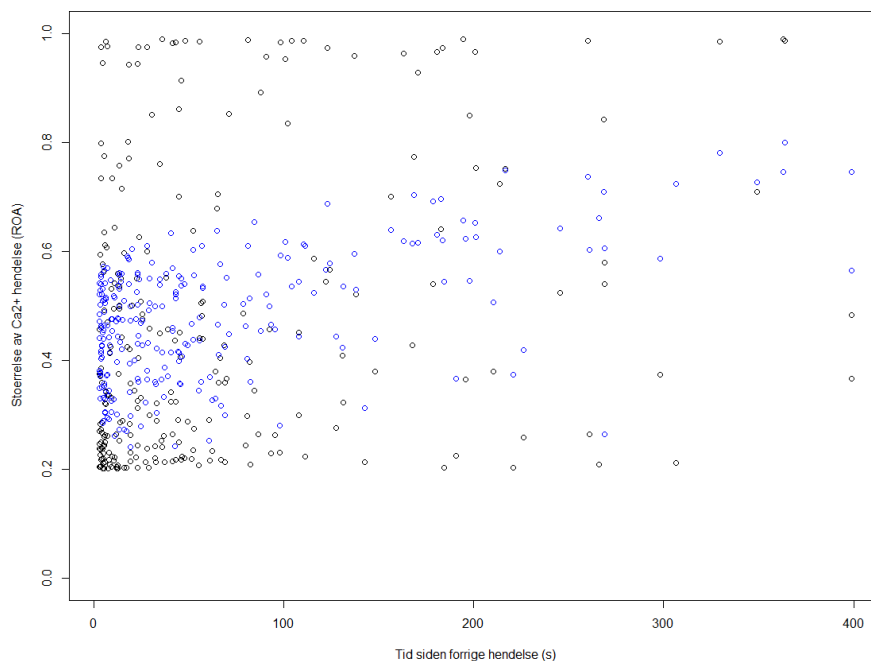
modell 3.1

Siden vi så at det lønner seg å logtransformere tiden etter å ha eksponensiert den kan vi undersøke om det også gjelder for de andre kovariatene.

4.3. Trunkert gammaprosess

$\exp(\beta_0 + \beta_t \log(t_i)), (\exp(\beta_1 + \beta_w \log(w_i + 0.01) + \beta_s \log(s_i)))^{-1}$	
AIC	-318.85
BIC	-301.14
$\hat{\beta}_0, \hat{\beta}_t, \hat{\beta}_1, \hat{\beta}_w, \hat{\beta}_s$	1.048, 0.899, -0.160, 0.499, 0.262,
95% KI for $\hat{\beta}_t$	(0.641, 1.156)
p-verdi for $\hat{\beta}_t$	<1e-11
95% KI for $\hat{\beta}_s$	(0.087, 0.438)
p-verdi for $\hat{\beta}_s$	0.0014

Vi ser at å logge alle kovariatene gir en del bedre modelltilpassning, samtidig som sammenhengen mellom tid og størrelse blir estimert til å være ca. den samme som tidligere. I denne modellen er også sammenhengen mellom størrelsen av forrige og størrelsen til neste igjen negativ, men denne gangen er sammenhengen estimert til å være en del større. Denne modellen er også den eneste der AIC verdien blir lavere av å inkludere størrelsen av forrige hendelse, men BIC-verdien blir fortsatt større. Uten β_s får vi AIC = -318.76, BIC = -304.59. De tilpassede størrelsene av denne modellen har en ganske annen form enn det vi har sett tidligere og er bedre egnet til å fange opp de små hendelsene med kort tid siden forrige.



Figur 4.16: Tilpassede verdier for størrelse fra modell 3.1 i blått

KAPITTEL 5

Diskusjon

I denne oppgaven har vi fokusert på et hovedspørsmål: Hvordan kan vi best mulig modellere størrelsene til Ca^{2+} hendelser i astrocytter som gammaprosesser. Vi startet med en enkel gjennomgang astrocytters oppbygging og egenskaper, en rask oppsummering av hva tidligere litteratur kan si oss om fordelingen av størrelsen til Ca^{2+} hendelser *in-vitro*, hvorfor vi tror gammaprosessliknende modeller kan være en god modell for fordelingen av størrelsene til Ca^{2+} hendelser og en kort gjennomgang av hva slags typer problemstillinger gammaprosessmodeller er blitt brukt på tidligere. I korte trekk kan man si at motivasjonen vår for å bruke gammaprosessliknende modeller handler om at vi tror tid siden forrige hendelse har en positiv effekt på størrelsen av neste Ca^{2+} hendelse, samt at effekten er monotont og kontinuerlig økende. Deretter redegjorde vi for datasettet som er brukt i denne oppgaven, egenskapene til en gammaprosessmodell og hvilke modifikasjoner av gammaprosessmodellen vi kan gjøre for å tilpasse den bedre til vår data.

Den viktigste og største delen av denne oppgaven er de ulike forsøkene på å tilpasse en gammaprosessliknende modell til vårt datasett. Vi finner at vår modell passer datasettet rimelig godt og kan enkelt utvides til å inkludere flere kontrollvariabler eller modifiseres for å tilpasse en bedre modell. Det er vanskelig å si noe om hva slags virkninger vår modell vil ha for videre modellering av størrelsene av Ca^{2+} hendelser, først og fremst fordi den er presentert i en middelmådig, norskspråklig masteroppgave, men også fordi dette forsøket på å modellere størrelsene til Ca^{2+} hendelser mangler sammenligning. Tidligere er det kun publisert to studier som foreslår en statistisk modell for størrelsen av Ca^{2+} hendelser i astrocytter, disse studiene bruker ikke data som er samlet inn *in-vivo*, de tar ikke hensyn til andre kontrollvariabler og de gjør få forsøk på å evaluere modellene. Samlet sett har litteraturen vært for raske til å bastant konkludere med at størrelsene til Ca^{2+} hendelser er potenslovsfordelt uten å diskutere antakelsene for eller konsekvenser av denne konklusjonen. Da det ikke finnes sammenlignbare forsøk på å modellere Ca^{2+} hendelser er det vanskelig å konkludere med at gammaprosessmodellene som blir presentert i denne oppgaven vil være av noe nytte. I tillegg blir nytten av modellene som blir presentert i denne oppgaven ytterligere dratt i tvil av at det per nå er flere ulike måter å måle størrelsen av Ca^{2+} hendelser på, alle med sine fordeler og ulemper, samt at det ikke finnes noen operasjonaliserbar definisjon av hva en Ca^{2+} hendelse er. Selv om evalueringene av de ulike modellene som blir presentert i denne oppgaven neppe gir oss noen større forståelse av de ekte dynamikkene som

påvirker Ca²⁺ signaler i astrocytter, og selv om forskningen på dette feltet er så variert at modellene neppe vil bli brukt uten store modifikasjoner, så viser denne oppgaven at det er mulig å tilpasse en modell for ulike datasett av Ca²⁺ hendelser *in-vivo* der man har en gjennomsliktig og klar diskusjon om hvilke antakelser som bør ligge til grunn.

En del av oppgaven gikk med til å diskutere spørsmålet om det finnes noe analogt til relative refraktærperioder, som man observerer i nevroner, for Ca²⁺ hendelser i astrocytter. Vi gikk gjennom hva en relativ refraktærperiode i en nevron er, et tidsintervall etter et nevronevner signal der nevronen krever betydelig større stimuli for å sende et like stort signal, samt hvordan man kan tenke på en relativ refraktærperiode i astrocytter, et tidsintervall etter en Ca²⁺ hendelse der det er lav sannsynlighet for at en ny Ca²⁺ hendelse vil være stor. Grunnet at vår definisjon av Ca²⁺ hendelser innebærer at det er vanskelig å skille to ulike Ca²⁺ hendelser fra hverandre hvis de kommer tett på hverandre, samt at vi ikke kan presist måle når en Ca²⁺ hendelse starter og slutter så lenge vi kun måler Ca²⁺ aktivitet over et 2D plan, vil det være vanskelig eller fånyttet å forsøke å estimere lengden på en slik relativ refraktærperiode. Istedenfor forsøkte vi, innenfor vårt gammaprosessrammeverk, å inkludere tid siden forrige hendelse på ulike måter i modellene, for å se hvilken måte det påvirker fordelingen til størrelsene av Ca²⁺ hendelser. Vi forsøkte også å estimere effekten av tiden som er gått siden forrige hendelse på fordelingen til størrelsen av neste hendelse, samt tallfeste sannsynligheten for at effekten er positiv. Vi kom fram til en måte å inkludere tid i formparameteren, $\exp(\beta_0 + \beta_t \log(t_i) + \dots)$, på som ga konsistent bedre tilpassede modeller enn de andre alternativene, samt at i modelltypen som var nærmest den sanne datagenererende prosessen så ble modellen bedre tilpasset av å ta hensyn til tid siden forrige hendelse. Estimaten for effekten av tid siden forrige hendelse varierte stort fra modell til modell, men i de modellene som var best tilpasset data kan vi si at effekten var betydelig. Formen på effekten vi estimerte var noe overraskende, vi tenkte at effekten av tid ville være størst når det er gått kort tid siden forrige hendelse, men i den best tilpassede modellen fant vi at effekten av tid ble større jo lengre det er gått siden forrige hendelse. At effekten av tid blir større jo lenger tid som er gått bør møtes med en viss skepsis, det kan tenkes at det vil lønne seg å sjekke om modellen blir bedre tilpasset hvis tid blir inkludert på en annen måte enn de som blir presentert i denne oppgaven eller at opptak over lenger tid enn hva vi hadde tilgjengelig vil gi data som gir et annet bilde. Siden en hovedantakelse for gammaprosessmodeller er at effekten av tid er positiv måtte vi utvide formparameteren for å finne sannsynligheten for at effekten av tid ikke er positiv, i de modellene fant vi at det sannsynligheten for at effekten av tid er positiv er svært stor. Alt i alt fant vi ikke overbevisende bevis for at det finnes en slags relativ refraktærperiode for astrocytter etter Ca²⁺ hendelser, særlig siden den estimerte effekten av tid i de best tilpassede modellene ikke er større når det er gått kort tid siden forrige hendelse, men vi finner at det er en robust positiv sammenheng mellom tid som er gått siden forrige hendelse og forventet størrelse av neste hendelse.

En liten del av oppgaven ble brukt til å diskutere sammenhengen mellom størrelse av forrige Ca²⁺ hendelse og størrelsen til neste Ca²⁺ hendelse. Vi viste til analyser av interspike-intervaller i nevroner der det er påvist at jo større forrige interspike-intervall var, jo mindre vil forventet lengde av neste intervall

være. Analogt til fordelingen av størrelsene av Ca^{2+} hendelser vil det være naturlig å tenke at jo større forrige hendelse var, jo mindre vil neste hendelse bli. For de fleste modellene som ble presentert er parameteren som påvirket fordelingen til neste hendelse gitt en størrelsen til forrige hendelse estimert til å være tilnærmet 0. Samtidig er det verdt å merke seg at i modellen som er best tilpasset datasettet er også den modellen der sammenhengen er estimert til å være sterkst og der sannsynligheten for at effekten er negativ er størst. Når man beregner p-verdien til en parameter flere ganger øker sannsynligheten for at p-verdien er lav, så man bør ikke konkludere at det er en sammenheng basert på beregnet p-verdi fra en av flere modeller, men det tyder på at denne sammenhengen bør undersøkes nærmere.

Appendices

TILLEGG A

The First Appendix

a,b	AIC
$t_i \exp(\beta_0), 1$	-229.19
$t_i \exp(\beta_0), b$	-236.28
$t_i \exp(\beta_0 + \beta_w w_i), b$	-264.68
$t_i \exp(\beta_0 + \beta_w w_i + \beta_s s_i), b$	-263.94

a,b	AIC
$\exp(\beta_0), 1$	-200.74
$\exp(\beta_0), b$	-216.25
$\exp(\beta_0 + \beta_t t_i), b$	-226.19
$\exp(\beta_0 + \beta_t t_i + \beta_w w_i), b$	-258.21
$\exp(\beta_0 + \beta_t t_i + \beta_w w_i + \beta_s s_i), b$	-256.24

a,b	AIC
$\beta_0, 1$	-200.74
$\beta_0 + \beta_t t_i, 1$	-227.15
$\beta_0 + \beta_t t_i, b$	-234.20
$\beta_0 + \beta_t t_i + \beta_w w_i, b$	-259.49
$\beta_0 + \beta_t t_i + \beta_w w_i + \beta_s s_i, b$??

TILLEGG B

The Second Appendix

Bibliografi

- [1] Bjørnstad, D. M. mfl. «Begonia—a two-photon imaging analysis pipeline for astrocytic ca^{2+} signals». I: *Frontiers in Cellular Neuroscience* årg. 15 (2021), s. 681066.
- [2] Byrd, R. H. mfl. «A limited memory algorithm for bound constrained optimization». I: *SIAM Journal on scientific computing* årg. 16, nr. 5 (1995), s. 1190–1208.
- [3] Claeskens, G., Hjort, N. L. mfl. «Model selection and model averaging». I: *Cambridge Books* (2008).
- [4] Clauset, A., Shalizi, C. R. og Newman, M. E. «Power-law distributions in empirical data». I: *SIAM review* årg. 51, nr. 4 (2009), s. 661–703.
- [5] Gabbiani, F. og Cox, S. J. *Mathematics for neuroscientists*. Academic Press, 2017.
- [6] Hampel, D. og Lansky, P. «On the estimation of refractory period». I: *Journal of neuroscience methods* årg. 171, nr. 2 (2008), s. 288–295.
- [7] Jung, P. mfl. «Noise-induced spiral waves in astrocyte syncytia show evidence of self-organized criticality». I: *Journal of neurophysiology* årg. 79, nr. 2 (1998), s. 1098–1101.
- [8] Moran, P. *The Theory of Storage*. Methuen's monographs on applied probability and statistics. Methuen, 1961.
- [9] Noortwijk, J. M. van. «A survey of the application of gamma processes in maintenance». I: *Reliability Engineering & System Safety* årg. 94, nr. 1 (2009), s. 2–21.
- [10] Semyanov, A., Henneberger, C. og Agarwal, A. «Making sense of astrocytic calcium signals—from acquisition to interpretation». I: *Nature Reviews Neuroscience* årg. 21, nr. 10 (2020), s. 551–564.
- [11] Song, S. mfl. «Mathematical modeling and analyses of interspike-intervals of spontaneous activity in afferent neurons of the zebrafish lateral line». I: *Scientific Reports* årg. 8, nr. 1 (2018), s. 1–14.
- [12] Wang, Y. mfl. «Accurate quantification of astrocyte and neurotransmitter fluorescence dynamics for single-cell and population-level physiology». I: *Nature neuroscience* årg. 22, nr. 11 (2019), s. 1936–1944.

- [13] Wu, Y.-W. mfl. «Spatiotemporal calcium dynamics in single astrocytes and its modulation by neuronal activity». I: *Cell calcium* årg. 55, nr. 2 (2014), s. 119–129.
- [14] Aalen, O., Borgan, O. og Gjessing, H. *Survival and event history analysis: a process point of view*. Springer Science & Business Media, 2008.