

Baysiansk Inferens for Markov-Arrival-Processer



AALBORG UNIVERSITET

Institut for Matematiske fag \bullet Torkild Enge \bullet 10.
semester \bullet MAT 6 \bullet 1.februar - 3.maj 2010



TITEL:

Baysiansk inferens for markov-arrival-processer

PROJEKT PERIODE:

Fra 1. februar 2010 til 3. maj 2010

PROJEKTGRUPPE:

Torkild Enge

VEJLEDER:

Jakob Gulddahl Rasmussen

OPLAGSTAL: 4

ANTAL SIDER: 57

SYNOPSIS:

I rapporten defineres og beskrives Markov Arrival processer og der benyttes baysiansk statistik til at undersøge hvordan punktmønstre repræsenterende biler kan beskrives som MAP. Efter en kort beskrivelse af den baysianske fremgangsmåde, defineres punktmønstre og punktprocesser på den reelle akse. Derefter defineres og beskrives markov-arrival-processen, hvorefter en række summary statistics defineres og nogle analytiske resultater for disse udledes. De data som bliver brugt til $inferens \ bliver \ præcenteret \ sammen \ med$ Metropolis-Hastings-algoritmen, inden to modeller tilpasses data ved baysiansk fremgangsmåde, og disse modellers egnethed til at beskrive data kontrolleres.

I appendiks findes en række definitioner og resultater for markovkæder og en samling af funktioner i programmeringssproget R, som er brugt under udarbejdningen af rapporten.

© Torkild Enge 2010

AALBORG UNIVERSITET

ABSTRACT

Point processes on the real line have many useful applications, both when describing events in time, and the distributions of points along a line in space. In this report a class of point processes named Markov Arrival Processes are defined, described and used for baysian inference upon tree sets of data.

The data upon witch the inference is made, come from the FleetFet project at the University of Mannheim. This data contains the placement of cars on a freeway, and it is the aim of the report to introduce models that can be used to describe these as point processes on the real line.

The first to chapters give a basic presentation of the baysian approach to inference and point processes on the real line. The third chapter describes the Markov Arrival processes and some of its properties. Thereafter several summary statistics are defined and some results fore these are obtained. The sixth chapter describe the data and the Metropolis-Hastings algorithm. The last two chapters describe two direct models used to describe the data as point processes, the poisson process and the Markov modulated renewal process. The distributions of the parameters of these models are calculated, and the abilites of the models to describe the data is assessed.

The appendixes contain various results for markov chains used in the report, as well as several functions in R.

FORORD

Denne rapport er skrevet som speciale på 10. semester ved Aalborg Universitet.

Rapporten omhandler markov-arrival-processer som er en klasse af punktprocesser på den reelle akse. Rapporten forudsætter kendskab til grundlæggende sandsynlighedsregning og baysiansk statistik.

Læseren gøres opmærksom på, at der bruges to slags henvisninger. Når nummeret er i parentes, som f.eks. (1.2) henvises der til en ligning eller ulighed med dette nummer, mens 'sætning 1.2' henviser til en hel sætning med dette nummer. Sidstnævnte henvisning findes ligeledes til lemma, eksempel, figur, algoritme og korollar.

Kilder anføres i starten af afsnit og kapitler, mens henvisninger til specifikke sætninger og beviser anføres i en parentes efter sætnings-, lemma-, eksempel- eller korollarnummer. Notationen [X] henviser til den X'te kilde i kildelisten bagerst.

Decimaltal skrives med punktum. For eksempel vil en fjerdedel noteres som 0.25 og ikke 0,25.

Jeg vil gerne give en tak for tålmodighed og støtte til min vejleder, instituttets sekretærer og min familie.

Aalborg den 3/5 2010

Torkild Enge

INDHOLD

1	Ind	ledning	1
2	Bay	zesiansk statistik	2
	2.1	Likelihood, prior og posterior	3
	2.2	Konjugeret prior	4
	2.3	Central posterior interval	4
3	Pur	nktprocesser på den Reelle Akse	6
	3.1	Grundlæggende begreber	7
	3.2	Poissonproces	9
	3.3	Regulære og tiltrækkende punktprocesser	9
	3.4	Coxprocesser	10
	3.5	Renewal punktprocesser	11
4	Ma	rkov-Arriavel-Proceser	12
	4.1	Egenskaber for MAP	13
	4.2	Markov Arrival Palm Proces	14
	4.3	Eksempler på MAP	15
	4.4	Simulation af MAP	16
5	Sun	nmary Staticstics	17
	5.1	Forbindelsestal	17
	5.2	Forbindelsesafstand	19
	5.3	Forbindelseshop	21
6	Dat	abehandling	25
	6.1	Presentation af Data	25
		6.1.1 Modelkontrol og Summary Statistics	25
	6.2	Metropolis-Hastings-algoritmen	25
7	Pois	ssonprocessen	29
	7.1	Possionproces	29
	7.2	Estimation af parametrene.	29
	7.3	Modelkontrol	29
		7.3.1 Forbindelsestal	30
		7.3.2 Forbindelsafstand	31
		7.3.3 Forbindelseshoptal	32
8	Ma	rkovmoduleret Renewalproces	34
	8.1	Markovmoduleret renewalproces	35

	8.2	Estimation af parametrene.	35
	8.3	Modelkontrol	37
		8.3.1 Forbindelstal	37
		8.3.2 Forbindelsafstand	38
		8.3.3 Forbindelseshoptal	38
	8.4	forslag til ny model	38
9	Kon	nklution	44
Α	Mar	rkovkæder	46
	A.1	Konvergens for Markovkæder	47
	A.2	Diskrete Phasetypefordelinger	49
	A.3	Endelige Markovkæder i Kontinuert Tid	51
в	Pro	grammer i R	52
	B .1	Simulation af MAP	53
	B.2	Metropolis-Hastings	53
	B.3	middelværdi for Y_{cd_+}	55
	B .4	middelværdi for Y_{cd_+}	56
Li	ttera	atur	57

INDLEDNING

Der findes mange typer af fænomener der kan beskrives ved hjælp af punktprocesser. I rummet kan for eksempel placeringen af byer, planter, kræftknuder og epicentre for jordskælv modelleres ved hjælp af punktprocesser. Punktprocesser på linjen finder blandt andet anvendelse til modellering af tidspunkter for begivenheder som jorskælv eller telefonopkald. Også placeringen af objekter på en linje kan naturligt nok modelleres ved hjælp af punktprocesser på linjen. Derfor er det i mange tilfælde nyttigt at kunne lave inferens for punktprocesser.

Denne rapport beskriver metoder til inferens for markov-arrival-processer, som er en klasse af punktprocesser på den reelle akse. Markov-arrival-processer er en meget fleksibel klasse af punktprocesser, idet de kan udgøre både tiltrækkende og regulære punktprocesser. Poissonprocessen er et specialtilfælde af markov-arrival-processer, ligesom nogen former for coxprocesse og renewalprocesser.

Rapporten anvender baysiansk inferens, eftersom det ikke altid er muligt at opstille likelihood for en markov-arrival-proces i en form der lader sig maksimere analytisk.

De data som rapporten inferer over beskriver placeringen af biler på en vej, og er lavet i forbindelse med projektet Fleetnet på universitetet i Mannheim. Fleetnet arbejder med trådløse netværk imellem biler, og disse data er genereret med det formål for øje, at beskrive hvordan et sådant netværk opfører sig imellem biler i bevægende.

I denne rapport betraktes data dog som et øjebliksbillede. Det vil sige at rapporten søger at beskrive hvordan bilerne kan forventes at være fordelt på vejen i et givet tidspunkt.

Som det første gives en beskrivelse af baysiansk statistik. I tredie kapitel beskrives grundlæggende egenskaber for punktprocesser på den reelle akse, og nogle simple familier af punktprocesser beskrives. Herefter defineres markov-arrival-processer og nogle egenskaber for disse beskrives. Femte kapitel beskriver en række summary statistics for punktprocesser på den reelle akse. Disse er ligesom data beregnet på at beskrive hvordan et trådløst netværk imellem biler vil opføre sig. Herefter beskrives metropolis-hastings algoritmen samt de data, som i de to efterfølgende kapiteler bliver forsøgt modelleret med henholdsvis en poissonproces og en markovmoduleret renewalproces.

Bagerst i rapporten findes appendiks omhandlende markovkæder og en række funktioner i programmeringssproget R.

BAYESIANSK STATISTIK

2.1 LIKELIHOOD, PRIOR OG POSTERIOR

Dette afsnit er baseret på [6, side 31-41].

Et ofte forekommende problem ved behandling af data er, at der haves en mængde af observationer, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, ud fra hvilke, det ønskes at sige noget om en ukendt parameter, $\boldsymbol{\theta}$. I Bayesiansk statistik opfattes både \boldsymbol{x} og $\boldsymbol{\theta}$ som stokastiske variable til forskel fra klassisk statistik, hvor kun \boldsymbol{x} betragtes som en stokastisk variabel, og $\boldsymbol{\theta}$ betragtes som en fast parameter. Det antages, at den simultane fordeling, $P(\boldsymbol{x}, \boldsymbol{\theta})$, er ukendt, og at den betingede fordeling $P(\boldsymbol{x}|\boldsymbol{\theta})$ er kendt. Den sidstnævnte fordeling kaldes en observationsmodel.

Likelihoodfunktionen defineres ud fra observationsmodellen på samme måde som i klassisk statistik.

Definition 2.1 (Likelihoodfunktion) Ved likelihoodfunktionen af $\boldsymbol{\theta}$ givet \boldsymbol{x} , $l(\boldsymbol{\theta}|\boldsymbol{x})$, forstås $p(\boldsymbol{x}|\boldsymbol{\theta})$ betragtet som funktion af $\boldsymbol{\theta}$.

Det bør noteres, at likelihoodfunktionen ikke er en sandsynlighedstæthed, da den ikke nødvendigvis integrerer op til én.

Når der gøres brug af Bayesiansk statistik, er man interesseret i værdierne af kukendte størrelser, disse noteres som

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$$

Før der udføres et eksperiment, som har til formål at bestemme værdien af θ , kan der være en viden om værdien af θ . Denne viden kan udtrykkes som en tæthed kaldet priortætheden.

Definition 2.2 (Prior)

Ved en priortæthed, $p(\boldsymbol{\theta})$, forstås den marginale sandsynlighedstæthed for $\boldsymbol{\theta}$.

I nogle tilfælde er det nyttigt at gøre brug af en funktion som ikke integrerer til noget endeligt. I tilfælde hvor man ikke har nogen viden om fordelingen af $\boldsymbol{\theta}$, kan prioren for eksempel antages at $p(\boldsymbol{\theta}) \propto 1$ for alle værdier af $\boldsymbol{\theta}$, som er tilladelige i observationsmodellen. Når en uegentlig prior benyttes bør det altid undersøges om posteriorfunktionen er egentlig.

Bayes' sætning er givet ved

Sætning 2.3 Bayes' sætning [6, side 7] Antag, at $p(\theta) > 0$ og p(x) > 0, da gælder

 $p(\boldsymbol{\theta}, \boldsymbol{x}) = p(\boldsymbol{x})p(\boldsymbol{\theta}|\boldsymbol{x}) = p(\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta}).$

Ved hjælp af Bayes' sætning kan et udtryk for sandsynlighedstætheden for den ubekendte θ givet datasættet x således bestemmes som

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{\theta})}{p(\boldsymbol{x})} p(\boldsymbol{x}|\boldsymbol{\theta})$$

$$\propto p(\boldsymbol{\theta}) l(\boldsymbol{\theta}|\boldsymbol{x}), \qquad (2.1)$$

hvor $p(\boldsymbol{x})^{-1}$ blot er en proportionalitetsfaktor.

Definition 2.4 (Posterior)

Ved en posteriortæthed, $p(\boldsymbol{\theta}|\boldsymbol{x})$, forstås en sandsynlighedstæthed, der er proportional med produktet af priortætheden og likelihoodfunktionen.

Posteriortætheden skal ses som et udtryk for vores viden om værdien af $\boldsymbol{\theta}$, efter der er taget højde for værdien af observationerne. Det er posteriorfordelingen, der i Bayesiansk statistik anvendes til at inferere ud fra. Hvis der senere udføres ydderligere eksperimenter for at bestemme værdien af $\boldsymbol{\Theta}$ kan posteriortætheden for det forige forsøg benyttes som priortæthed ved det næste forsøg.

Ud fra Definition 2.4 ses det, at posteriortætheden forbliver uændret, hvis likelihoodfunktionen multipliceres med en faktor, der er uafhængig af $\boldsymbol{\theta}$.

2.2 KONJUGERET PRIOR

Dette afsnit er baseret på [6, side 56-60] og beskriver en bestemt type prior kaldet en konjugeret prior.

Definition 2.5 (Konjugeret prior) Hvis $l(\boldsymbol{\theta}|\boldsymbol{x})$ er likelihoodfunktionen, da siges en klasse af priortætheder, Π , at danne en konjugeret familie, hvis posteriortætheden,

 $p(\boldsymbol{\theta}|\boldsymbol{x}) \propto p(\boldsymbol{\theta})l(\boldsymbol{\theta}|\boldsymbol{x}),$

også ligger i klassen Π for alle \boldsymbol{x} , når priortætheden ligger i Π .

En konjogeret prior kan med fordel benyttes i tilfælde hvor beregningerne, når posteriortætheden skal bestemmes, er mindre komplicerede, end hvis der ikke vælges en konjugeret prior. Som regel vil den priortæthed, som man bedst mener afspejler den forhåndsviden man har, ikke være en konjugeret prior. I stedet kan man så vælge en konjugeret priortæthed, som ligner den ikke-konjugerede prior mest muligt. I mange tilfælde vil posteriortætheden, bestemt med den konjugerede prior, ikke afvige ret meget fra posteriortætheden bestemt med den ikke-konjugerede prior, men beregningerne vil være væsentligt lettere.

2.3 CENTRAL POSTERIOR INTERVAL

Det er ofte ønskeligt at kunne angive et interval, hvor θ ligger med en vis sadnsynlighed – for eksempel 95%– derfor indføres følgende definition:

Definition 2.6 (Centralt posterior interval) Givet en posteriorfordeling $p(\theta|\mathbf{x})$ for en éndimensional parameter θ og et tal $k \in [0, 1[$, kaldes intervallet [a, b] et k-centralt posterior interval (CPI) for θ , hvis

$$\int_{-\infty}^{a} p(\theta|\mathbf{X}) \mathrm{d}\theta = \int_{b}^{\infty} p(\theta|\mathbf{X}) \mathrm{d}\theta = \frac{1-k}{2}.$$

[Eksempel 2.7 *Et* 95%-*CPI* for $\theta | \boldsymbol{X} \sim \mathcal{N}(0, 1)$ fås ved

$$\int_{-\infty}^{a} p(\theta|\mathbf{X}) \mathrm{d}\theta = \int_{b}^{\infty} p(\theta|\mathbf{X}) \mathrm{d}\theta = \frac{1 - 0.95}{2}.$$

Dette giver a = -1.96 og b = 1.96. Intervallet ses afbildet på Figur 2.1.

En fordel ved CPI er, at dette er invariant under transformation af den stokastiske variabel.

Sætning 2.8

Lad]a, b[være et k-CPI for $p(\theta|\mathbf{x})$, og lad $\varphi = f(\theta)$, hvor $f : \mathbb{R} \to \mathbb{R}$ er en differentiabel, invertibel funktion med $f(-\infty) = -\infty$ og $f(\infty) = \infty$, så er]f(a), f(b)[et k-CPI for φ .

Bevis:

For transformationen af en kontinuert éndimensional stokastisk variabel gælder, at

$$p_y(y) = p_x(f^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} f^{-1}(y) \right|.$$



 $Figur \ 2.1 \it Viser \ 95\% - \it CPI \ for \ standard normal for deling$

Derfor bliver

$$\int_{-\infty}^{f(a)} p_{\varphi}(\varphi | \boldsymbol{x}) d\varphi = \int_{-\infty}^{f(a)} p_{\theta}(f^{-1}(\varphi) | \boldsymbol{x}) \left| \frac{d}{d\varphi} f^{-1}(\varphi) \right| d\varphi$$
$$= \int_{-\infty}^{a} p_{\theta}(\theta | \boldsymbol{x}) d\theta.$$

Tilsvarende

$$\int_{f(b)}^{\infty} p_{\varphi}(\varphi | \boldsymbol{x}) \mathrm{d}\varphi = \int_{b}^{\infty} p_{\theta}(\theta | \boldsymbol{x}) \mathrm{d}\theta.$$

De i dette kapittel presenterede begreber benyttes i kapittelerne 7 og 8 til inferans.

PUNKTPROCESSER PÅ DEN REELLE AKSE

I dette kapitel defineres punktprocesser på den reelle akse og en række relaterede begreber. Herefter defineres og beskrives tre klasser af punktprocesser. Kapitlet er skrevet på baggrund af [7], med untagelse af definition 3.13 som kommer fra [2, kappitel V].

3.1 GRUNDLÆGGENDE BEGREBER

For at definere punktprocesser er det først nødvendigt at definere punktmønstre på den reelle akse.

Definition 3.1 (Lokalt endeligt punktmønster)

En punktmængde $x \subseteq \mathbb{R}$ siges at være et punktmønster på \mathbb{R} , hvis x er en tællelig delmængde af \mathbb{R} . Hvis det for alle $a, b \subseteq \mathbb{R}$ gælder at $x \cap [a, b]$ indeholder et endeligt antal punkter, kaldes x for lokalt endelig.

I resten af rapporten vil X betegne en lokalt endelig punktproces på \mathbb{R} , som defineret i følgende definition

Definition 3.2 (Lokalt endelig punktproces)

En stokastisk variabel, hvis udfaldsrum er en mængde af lokalt endelige punktmøndtre på \mathbb{R} kaldes en lokalt endelig punktproces på \mathbb{R}

Med mindre andet er anført vil alle punktprocesser i resten af rapporten antages at være lokalt endelige. En af de ting punktprocesser på \mathbb{R} ofte bruges til, er at beskrive tidspunkter hvor bestemte hendelser finder sted. Et eksempel herpå kan være de tidspunkter, hvor der ankommer kunder til en forretning. Punkterne i en punktproces på \mathbb{R} kaldes derfor også for ankomster. Tiderne imellem to ankomster kaldes ventetider.

Denne rapport omhandler fortrinsvis stationære punktprocesser, som beskrevet i følgende definition.

Definition 3.3 (Tidsstationaritet)

En punktproces X på \mathbb{R} er tidsstationær, hvis dens fordeling er invariant under translation, dvs. fordelingen af $X + s = \{\xi + s | \xi \in X\}$ er den samme som fordelingen for X for ethvert $s \in \mathbb{R}$.

For en tidsstationær punktproces kan tiden sættes til 0 på et hvilket som helst tidspunkt uden at fordelingen for X ændres.

I visse tilfælde vil en punktproces betragtes betinget med at den indeholder et punkt til tiden t_0 . I så fald kan den ikke være tidsstationær, idet den betingede tæthed for X = x er 0 hvis ikke $t_0 \in x$. I dette tilfælde kan begrebet begivenhedsstationær i stedet benyttes.

Definition 3.4 (Begivenhedsstationaritet)

En punktproces X på \mathbb{R} er begivenhedsstationær, hvis det for alle $t_0 \in \mathbb{R}$, gælder at fordelingen af $X + s = \{\xi + s | \xi \in X\}$ betinget med $t_0 \in X + s$ er den samme som fordelingen for X betinget med $t_0 \in X$ for alle s hvorom gælder at $s + t_0 \in X$.

For en tidsstationær punktproces kan tiden sættes til t_0 i et hvilket som helst punkt i X uden at fordelingen for X ændres.

Idet punktprocessen X er en delmængde af den reelle akse har punkterne i X en naturlig rækkefølge. Således vil X₁ betegne min $(\psi \in X | \psi \ge 0)$ og X_i vil betegne min $(\psi \in X | \psi > X_{i-1})$ for i = 2, 3, ...

Et vigtigt begreb i beskrivelsen af punktprocesser er deres intensitet.

Definition 3.5 (Intensitetsmål og intensitetsfunktion) Lad X være en punktproces på \mathbb{R} . Intensitetsmålet μ på \mathbb{R} er da givet ved

$$\mu(B) = \mathbb{E}N(B), \quad B \subseteq \mathbb{R}.$$

Hvis intensitetsmålet μ kan skrives som

$$\mu(B) = \int_B \rho(\xi) d\xi, \quad B \subseteq \mathbb{R}$$

med $\rho \geq 0$ for hele S, kaldes ρ for intensitetsfunktionen for X.

Intensitetsmålet beskriver hvor mange ankomster der kan forventes til forskellige tidspunkter. I det tidligere nevnte eksempel med forretningen kan man for eksempel forestille sig at intensiteten vil være højere om eftermidagen når folk har fri, end om formidagen hvor de fledste er på arbejde. For en tidsstationær punktproces er $\rho(a) = \rho$ for alle $a \in \mathbb{R}$.

Ud over at være beskrevet ved sin intensitet kan en punktproces også beskrives ved sit andenordens faktorielle momentmål og sin andenordens produkttæthed.

Definition 3.6 (Andenordens faktorielt momentmål og andenordens produkttæthed) Lad X være en punktproces på \mathbb{R} . Det andenordens faktorielle momentmål $\alpha^{(2)}$ på \mathbb{R}^2 er da givet ved

$$\alpha^{(2)}(C) = \mathbb{E}\left[\sum_{\xi,\eta\in X}^{\neq} \mathbf{1}[(\xi,\eta)\in C]\right], \quad C\subseteq \mathbb{R}^2.$$

Hvis det andenordens faktorielle momentmå
l $\alpha^{(2)}$ kan skrives som

$$\alpha^{(2)}(C) = \int \int \mathbf{1}[(\xi,\eta) \in C] \rho^{(2)}(\xi,\eta) d\xi d\eta, \quad C \subseteq \mathbb{R}^d \times \mathbb{R}^d,$$

med $\rho^{(2)} \geq 0$, da kaldes $\rho^{(2)}$ for den andenordens produkttæthed.

For en tids stationær punktproces er $\rho^{(2)}(a, a + c) = \rho^{(2)}(b, b + c) = \rho^{(2)*}(c)$ for all e $a, b, c \in \mathbb{R}$. Ud fra definition 3.5 og definition 3.6 kan parkor relaitionsfunktionen defineres som følger:

Definition 3.7 (Parkorrelationsfunktion) Hvis både ρ og $\rho^{(2)}$ eksisterer, er parkorrelationsfunktionen defineret ved

$$g(\xi,\eta) = \begin{cases} \frac{\rho^{(2)}(\xi,\eta)}{\rho(\xi)\rho(\eta)} & \text{for } \rho(\xi)\rho(\eta) \neq 0\\ 0 & \text{for } \rho(\xi)\rho(\eta) = 0 \end{cases}$$

Hvis parkorrelationsfunktionen for punkterne $a, b \in \mathbb{R}$ er større end 1 vil tætheden for at X indeholder punktet b, være større betinget med $a \in X$ end den ubetingede tæthed.

For en tidsstationær punktproces gælder

$$g(\xi,\eta) = \frac{\rho^{(2)}(\xi,\eta)}{\rho^2} = \frac{\rho^{(2)*}(\xi-\eta)}{\rho^2} = g(\xi-\eta).$$
(3.1)

3.2 POISSONPROCES

Poissonprocessen er en simpel punktproces, som kan anvendes til at beskrive begivenheder der er uafhængige af hinanden.

For at definere poissonprocessen defineres først den binomiale punktproces.

Definition 3.8 (Binomial punktproces) Lad f være en tæthedsfunktion på en mængde $B \subseteq \mathbb{R}$, og lad $n \in \mathbb{N}$. En punktproces X bestående af n uafhængige og ensfordelte punkter fordelt med tæthed f kaldes en binomial punktproces med npunkter i B med tæthed f. Dette noteres $X \sim \text{Binomial}(B, n, f)$

Ses der på en delmængde af B, fås der af denne definition, at antallet af punkter i delmængden er binomialfordelt.

Poissonprocessen kan nu defineres.

Definition 3.9 (poissonproces)

En punktproces X på \mathbb{R} kaldes en poissonproces med intensitetsfunktion $\rho(t)$, hvis følgende egenskaber er opfyldt:

- (i) For enhver mængde $B \subseteq \mathbb{R} \mod \mu(B) < \infty$ er antallet af punkter i B Poissonfordelt med middelværdi $\mu(B)$.
- (ii) For ethvert $n \in \mathbb{N}$ og $B \subseteq S$ med $0 < \mu(B) < \infty$ gælder, at betinget med N(B) = n er $X \cap B \sim \text{Binomial}(B, n, f)$, hvor $f(\xi) = \frac{\rho(\xi)}{\mu(B)}$.

En poissonproces med $\rho(t) = k$, hvor k er en konstant, kaldes homogen.

I en homogen poissonproces på \mathbb{R} er alle ventetiderne ifølge [7, Proposition 3.5] uafhængige og eksponentialfordelte med middelværdi ρ^{-1} . Eftersom antallet af punkter i en mængde C er poissonfordelt med middelværdi $\mu(C)$, gælder der for en poissonproces at

$$\begin{aligned} \alpha^{(2)}(C) &= \mathbb{E}\left[\sum_{\xi,\eta\in X}^{\neq} \mathbf{1}[(\xi,\eta)\in C]\right] \\ &= \sum_{n=2}^{\infty} \frac{\mu(C)^n e^{\mu(C)}}{n!} n(n-1) \\ &= \mu(C)^2 \sum_{m=0}^{\infty} \frac{\mu(C)^m e^{\mu(C)}}{m!} \\ &= \mu(C)^2 \\ &= \int_C \int_C \rho(\xi) \rho(\eta) \mathrm{d}\xi \mathrm{d}\eta, \end{aligned}$$

hvor m = n - 2. Således er $\rho^{(2)}(\xi, \eta) = \rho(\xi)\rho(\eta)$ for en poissonproces. Derfor er parkorelationsfunktionen $g(\xi, \eta) = \frac{\rho^{(2)}(\xi, \eta)}{\rho(\xi)\rho(\eta)} = 1$ for $\rho(\xi)\rho(\eta) \neq 0$.

3.3 Regulære og tiltrækkende punktprocesser

Nogen punktprocesser vil have en tendens til at punkterne ligger i klynger, hvor der ligger mange punkter tæt på hinanden, med større afstand mellem klyngerne. I andre punktprocesser vil der være tendens til at der er en jævn afstand imellem punkterne. Eksempelvis kunne man forestille sig at den før omtalte forretning er en resturation, hvor folk ofte kommer i grupper for at spise sammen, eller at det er en barber, hvor den næste kunde først kan komme til når den forige kunde er barberet færdig. For at kunne beskrive sådanne tendenser mere precist indføres K og L-funktionerne.

Definition 3.10 (K og L-funktionerne) For en tidsstationær punktproces på \mathbb{R} defineres K og L-funktionerne for r > 0 ved

$$K(r) = \int_{-r}^{r} g(t) \mathrm{d}, \quad L(r) = (K(r)/2),$$

For at se hvad K og L-funktionerne betyder sættes (3.1) ind i $K(r) = \int_{-r}^{r} g(t) dt$

$$K(r) = \int_{-r}^{r} \frac{\rho^{(2)}(\xi, \xi+t)}{\rho^2} dt = \rho^{-1} \int_{\xi-r}^{\xi+r} \frac{\rho^{(2)}(\xi, \eta)}{\rho(\eta)} d\eta.$$

Eftersom $\rho(\xi)$ kan betrgates som tætheden for at X både indeholder et punkt i en lille omegn af ξ , og $\rho^{(2)}(\xi,\eta)$ kan betrkates som tætheden for at X både indeholder et punkt i en lille omegn af ξ og i en lille omegn af η , kan $\frac{\rho^{(2)}(\xi,\eta)}{\rho(\xi)}$ betraktes som den betingede tæthed for at X indeholder et punkt i en lille omegn af η , givet at X indeholder et punkt i en lille omegn af ξ . Derfor opnås

$$K(r) = \rho^{-1} \mathbb{E}(N(X \cap [\xi - r, \xi + r]) | \xi \in X)$$
(3.2)

Det vil sige at K(r) angiver middelværdien af antallet af nabopunkter indenfor afstanden r.

Hvis L(r) > r siges punktprocessen at være tiltrækkende i afstanden r. For L(r) < r siges punktprocessen at være frastødende i afstanden r. For små værdier af r gælder for poissonprocessen er L(r) = r. Poissonprocessen er således hverken frastødende eller tiltrækkende. Ved små værdier af r er en punktproces mere regulær end poissonprocessen hvis L(r) < r og har større tendens til klyngedanelser hvis L(r) > r.

3.4 COXPROCESSER

Coxprocessen er en generalisering af poissonprocessen og benyttes til at definere markov-arrival-processen i kapitel 4.

Definition 3.11 (Coxproces)

Lad $S \subseteq \mathbb{R}$, og lad $Z = \{Z(\xi) | \xi \in S\}$, hvor $Z(\xi)$ er en positiv stokastisk variabel for alle $\xi \in S$, således at funktionen $\xi \to Z(\xi)$ med sandsynlighed en er lokalt integrabel. Hvis X|Z er en poissonproces på S med intensitetsfunktion Z, kaldes X for en coxproces drevet af Z.

En coxproces er tidsstationær hvis fordelingen af Z(a) er den samme som fordelingen af Z(a+b) for alle $a, b \in \mathbb{R}$.

Sætning 3.12 (Intensitetsfunktionen for en coxproces) [7, Side 60] Lad $X \subseteq S$ og $\xi, \eta \in S$. Hvis X er en coxproces drevet af Z, så er intensitetsfunktionen givet ved

$$o(\xi) = \mathbb{E}[Z(\xi)],$$

og andenordens produkttætheden er givet ved

$$\rho^{(2)}(\xi,\eta) = \mathbb{E}[Z(\xi)Z(\eta)].$$

Bevis:

Lad X være en coxproces på \mathbb{R} og $A, B \subseteq \mathbb{R}$. Dermed fås det ved benyttelse af definition ??, at intensitetsmålet er

$$\mu(B) = \mathbb{E}[N(B \cap X)] = \mathbb{E}[\mathbb{E}[N(B \cap X)|Z]] = \mathbb{E}\left[\int_{B} Z(\xi)d\xi\right] = \int_{B} \mathbb{E}[Z(\xi)]d\xi,$$

og dermed er $\rho(\xi) = \mathbb{E}[Z(\xi)].$

Ved at anvende andenordens momentmålet fra definition 3.5 ses det, at

_

$$\begin{aligned} \alpha^{(2)}(A \times B) &= \mathbb{E}\left[\sum_{\xi,\eta \in X}^{\neq} \mathbf{1}[\xi \in A, \eta \in B]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_{\xi,\eta \in X}^{\neq} \mathbf{1}[\xi \in A, \eta \in B]|Z\right]\right] \\ &= \mathbb{E}\left[\int_{A} \int_{B} [Z(\xi)Z(\eta)]d\xi d\eta\right] \\ &= \int_{A} \int_{B} \mathbb{E}[Z(\xi)Z(\eta)]d\xi d\eta, \end{aligned}$$
(3.3)

3.5 RENEWAL PUNKTPROCESSER

Renewal punktprocesser er en gruppe af punktprocesser, der blandt andet kan beskrive hendelser som sker efter en proces der gentager sig efter når hendelsen er indtruffet. Hvis den tid en frisør bruger på at frisere en kunde er ensfordelt for alle kunder, og frisøren begynder at frisere en ny kunde straks efter at hun er færdig med den forrige, vil de tidspunkter hvor kunderne forlader frisørstolen udgøre en renewal punktproces.

Definition 3.13 (Renewalpunktproces)

Lad X være en punktproces på \mathbb{R} . Hvis alle ventetider i X er uafhængige og ensfordelte, kaldes X en Renewal punktproces på \mathbb{R} .

Eftersom ventetiderne i en homogen poissonproces er uafhængie og ensfordelte, udgør denne en renewalprocess.

MARKOV-ARRIAVEL-PROCESER

Dette kapitel er baseret på [5, side 4] og [2, side 302-304]. I dette kapitel defineres og beskrives Markovaraival-processer (herefter kaldet MAP) der gives eksempler på MAP, og angives en algoritme for simulation af MAP.

4.1 Egenskaber for MAP

Markov-arrival-processer kan defineres på følgende måde.

Definition 4.1 (MAP)

Lad $J_t, t \in \mathbb{R}$ være en markovkæde i kontinuert tid på udfaldsrummet $\{1, \ldots, m\}$. Lad så X_a være Coxprocessen dannet ved en poissonpunktproces med intensitet β_i i de punkter hvor J_t er i tilstand i, og lad X_b være punktprocessen der med sandsynlighed p_{ij} har et punkt i de punkter hvor J_t skifter fra tilstand i til tilstand j. Da kaldes punktprocessen $X = X_a \bigcup X_b$ for en Markov Arrival Process.

I resten af rapporten forkortes Markov Arrival Process til MAP

Markovprocessen J_t er beskrevet ved matricen Λ . Lad $\Lambda = C + D$, hvor indgangene i C og D er givet ved hhv

$$d_{ij} = \begin{cases} \beta_i & i=j\\ \lambda_{ij}p_{ij} & i\neq j \end{cases} \text{ og } c_{ij} = \begin{cases} \lambda_{ii} & i=j\\ \lambda_{ij}(1-p_{ij}) & i\neq j \end{cases}.$$

$$(4.1)$$

Bemærk at λ_{ii} ikke angiver en af indgangene i matricen Λ . I stedet beskriver λ_{ii} raten for punkter og skift i J. Lad J være i tilstand i til tiden t_0 , og lad t_1 betegne den mindste værdi af $t > t_0$, hvor J skifter tilstand, eller hvor X indeholder et punkt. Da er $t_1 - t_0$ eksponentialfordelt med rateparameter $-\lambda_{ii}$. Det ses at en MAP er entydigt bestemt ud fra matricerne C og D samt α , som er fordelingen for J til tiden t = 0. Da det i denne sammenhæng er stationære og ergodiske punktprocesser som betragtes, må den stationære fordeling $\alpha_0 = \pi$ beyttes, for hvilken det gælder at $\pi \Lambda = 0$.

Sætning 4.2 [5, Side 4] Intensiteten for en tidsstationær MAP er givet ved $\rho = \pi D \mathbf{1} \top$

Bevis:

Intensiteten for en ankomst til tiden t, betinget med at $J_t = i$ er $\sum_{j=1}^m d_{ij}$. Dermed bliver intensiteten $\rho = \sum_{i=1}^m \pi_i \sum_{j=1}^m d_{ij} = \pi D \mathbf{1}^\top$.

Betragt nu den omvendte proces $\tilde{X} = \{x \in \mathbb{R} | -x \in X\}.$

Sætning 4.3 [5, Side 4]

Den omvendte proces $\tilde{X} = \{x \in \mathbb{R} | -x \in X\}$ er en MAP, beskrevet ved matricerne \tilde{C} og \tilde{D} med indgangene $\tilde{c}_{ij} = \frac{\alpha_j}{\alpha_i} c_{ji}$ og $\tilde{d}_{ij} = \frac{\alpha_j}{\alpha_i} d_{ji}$. Den stationære fordeling α er den samme for \tilde{X} og X.

Bevis:

Betragt \tilde{J} , defineret ved $\tilde{J}_i = J_{-i}$. Markovegenskaben for J gør at også \tilde{J} er en makov proces. Således udgør \tilde{X} betinget med \tilde{J} en superposition af en posionpunktproces med intensitet β_i i de punkter hvor \tilde{J} er i tilstand i og punktprocessen der med sandsynlighed p_{ij} har et punkt i de punkter hvor \tilde{J} skifter fra tilstand j til tilstand i. Det er derfor klart at \tilde{X} er en MAP.

Betragt nu matricen $\tilde{\Lambda}$, der beskriver \tilde{J} . Eftersom de tider \tilde{J} og J tilbringer i tilstand i har samme fordelinger, er $\lambda_{ii} = \tilde{\lambda}_{ii}$. For J gælder endvidre at når processen skifter til tilstand j har den forrige tilstand været i med sandsynlighed $\tilde{p}_{ij} = p(J=i)p_{ji}/p(J=j)$. Da J antages at være i sin stationære fordeling er $p(J=i) = \alpha_i$ og $p(J=j) = \alpha_j$. Eftersom $\lambda_{ij} = -\lambda_{ii}p_{ij}$ er

$$\tilde{\lambda}_{ij} = -\tilde{\lambda}_{ii}\tilde{p}_{ij} = -\lambda_{ii}\frac{\alpha_i p_{ji}}{\alpha_j} = \lambda_{ji}\frac{\alpha_i}{\alpha_j},$$

hvilket betyder at

$$\begin{split} \tilde{d}_{ij} &= \begin{cases} \beta_i & i = j \\ \tilde{\lambda}_{ij} \tilde{p}_{ij} & i \neq j \end{cases} \\ &= \begin{cases} \beta_i & i = j \\ \lambda_{ji} \alpha_j / \alpha_i p_{ji} & i \neq j \end{cases} \\ &= \frac{\alpha_j}{\alpha_i} d_{ji}. \end{split}$$

Tilsvarende fås

$$\begin{split} \tilde{c}_{ij} &= \begin{cases} \tilde{\lambda}_{ii} & i=j\\ \tilde{\lambda}_{ij}(1-\tilde{p}_{ij}) & i\neq j \end{cases} \\ &= \begin{cases} \lambda_{ii} & i=j\\ \lambda_{ji}p(j)/p(i)(1-p_{ij}) & i\neq j \end{cases} \\ &= \frac{\alpha_j}{\alpha_i} d_{ji}. \end{split}$$

I forbindelse med inferens for MAP er det nyttigt at have et udtryk for tætheden, der kan benyttes som likelihood. Dette gives i følgende sætning.

Sætning 4.4 [2, Proposition 1.5] For en MAP med parametre (α, C, D) er tætheden for de første n ventetidertider T_1, \ldots, T_n givet ved $P(T_1 = t_1, \ldots, T_n = t_n) = \boldsymbol{\alpha} e^{Ct_1} D e^{Ct_2} D \ldots e^{Ct_n} D \mathbf{1}^\top$

Bevis:

Betragt den hændelse at den første ventetid er t_1 , den anden ventetid er t_2 , $J_0 = i$, $J_{x_1-} = j$, $J_{x_1} = k$, $J_{x_2-} = l$ og $J_{x_2} = m$. Tætheden for denne hendelse er givet ved

$$\alpha_i \boldsymbol{e}_i e^{Ct_1} \boldsymbol{e}_j^\top d_{jk} \boldsymbol{e}_k e^{Ct_1} \boldsymbol{e}_l^\top d_{lm},$$

hvor e_i er enhedsvektoren med 1 i den *i*'te indgang og 0 i de resterende indgange. Ved at summere over i, j, k, l og m opnåes resultatet for n = 2. Ved tilsvarende udregninger kan resultatet opnås for n > 2. \Box

4.2 MARKOV ARRIVAL PALM PROCES

Ved palmfordelingen, Y, forstås fordelingen af X betinget med at X indeholder et punkt til tiden t_0 . Hvis palmprocessen er begivenhedsstationær, kan t_0 sættes til 0 uden tab af generalitet. Derfor vil palmfordelingen være betinget med at der ligger et punkt til tiden t = 0.

Sætning 4.5 [2, Proposition 1.4] Markov Arrival palm processen er begivenhedsstationær, hvis $\alpha = \alpha_0 = \frac{\pi D}{\pi D \mathbf{1}^{\top}}$ hvor π er den stationære fordeling for J.

Bevis:

Lad Z_n være værdien af J_t umiddelbart efter den *n*'te ankomst. Sætningen kan bevises ved at vise at α_0 er den stationære fordeling for markovkæden Z. Tætheden for at $J_t = j$ og at der ingen ankomster er i tidsrummet (0,t) under betingelse af $J_0 = i$ er $e_i e^{Cx_1} e_j^{\top}$, hvilket betyder at sandsynligheden for at $Z_{n+1} = j$ givet $Z_n = i$ er givet ved

$$p_{ij} = \sum_{k=0}^{m} \int_{0}^{\infty} \boldsymbol{e}_{i} e^{Ct} \boldsymbol{e}_{k}^{\top} d_{kj} \mathrm{d}t.$$

Dette betyder at overgangsmatricen for Y bliver

$$P = \int_0^\infty e^{Ct} D \mathrm{d}t = -C^{-1} D.$$



Figur 4.1 På grafen ses en MAP genereret som specifiseret i eksempel 4.6. Cirklerne viser punkterne i processen, mens de vandrette linjestykker angiver hvilken tilstand J er i til pågældende tidspunkt.

Lad nu π betegne den stationære fordeling for J. Da er $\pi \Lambda = \pi (C+D) = \mathbf{0}$ hvilket medfører $\pi D = -\pi C$. Herfra fås

$$\boldsymbol{\alpha}_{\mathbf{0}} P = \frac{-\pi C}{\pi D \mathbf{1}^{\top}} (-C^{-1}D) = \frac{\pi D}{\pi D \mathbf{1}^{\top}} = \boldsymbol{\alpha}_{\mathbf{0}},$$
onære fordeling for Z.

hvilket viser at α_0 er den stationære fordeling for Z.

Palm fordelingen for de første n mellemankomst
tider fås ved

$$f_0(t_1,\ldots,t_n) = \boldsymbol{\alpha}_{\mathbf{0}} e^{Ct_1} D \ldots e^{Ct_n} D \mathbf{1}^\top$$
(4.2)

hvor $\alpha_0 = \frac{\pi D}{\pi D \mathbf{1}^{\top}}$

4.3 Eksempler på MAP

Der gives nu to eksempler på MAP'er.

[Eksempel 4.6 Et simpelt eksempel på en MAP er on/of poissonprocessen beskrevet ved

$$D = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \text{og } C = \begin{bmatrix} -6 & 1 \\ 1 & -1 \end{bmatrix} \text{og } \boldsymbol{\alpha} = (0.5, 0.5)$$

Denne MAP udgør en tiltrækkende punktproces eftersom punkterne alle vil ligge i de områder hvor den underliggende markovproces J er i tilstand 1, og der imellem disse områder vil ligge områder hvor J er i tilstand 2 og der derfor ingen punkter er. En realisation af denne proces kan ses i figur 4.1.

[Eksempel 4.7 Et andet simpelt eksempel på en MAP er renevalprocessen beskrevet ved

$$D = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{og } C = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \text{og } \alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$$



Figur 4.2 På grafen ses en MAP genereret som specifiseret i eksempel 4.6. Cirklerne viser punkterne i processen, mens de vandrette linjestykker angiver hvilken tilstand J er i til pågældende tidspunkt.

Punkterne i denne MAP ligger de steder hvor underliggende markovproces J skifter fra tilstand 1 til tilstand 2. Således vil afstanden imellem to nabopunkter udgøre summen af tre eksponentialfordelte variable med middelværdi 1. Derfor vil den være gammafordelt med formparameter 3 og skalaparameter 1. Denne MAP udgør en regulær punktoroces. En realisation af denne proces kan ses i figur 4.2.

4.4 SIMULATION AF MAP

Givet $m \times m$ matricerne C og D, opfyldende 4.1 og begyndelsesfordelingen α kan følgende algoritme benyttes til at simulere en MAP.

Algoritme 4.8 Start med at generere starttilstanden, j_0 , fra fordelingen α , og sæt $t_0 = 0$. For $n = 0, 1, 2, \ldots$ foretages nu følgende:

- Generer Δt_n fra en eksponentialfordeling med rateparameter $-\lambda_{j_n j_n}$. Lad $t_{n+1} = t_n + \Delta t_n$.
- Generer $j_n + 1$ fra sandsynlighedsfordelingen hvor værdien j_n har sandsynligheden $P(j_n) = \frac{\beta_{j_n}}{-\lambda_{j_n j_n}}$ og de øvrige mulige værdier $i \in \{1, \ldots, m\} \setminus j_n$ har sandsynlighederne $P(i) = \frac{\lambda_{j_n i}}{-\lambda_{j_n j_n}}$.
- Hvis $j_n = j_{n+1}$ tilføjes processen et punkt beliggende i t_{n+1} . Hvis $j_n \neq j_{n+1}$ tilføjes punktet med sandsynligheden $p_{j_n j_{n+1}}$

I afsnit B.1 findes algoritmen som en funktion i R.

SUMMARY STATICSTICS

I dette kapitel indføres en række summary statistics for palmprocessen Y. Disse funktioner vil i kapitel 7 og kapitel 8 blive brugt til modelkontrol. Kapitlet er baseret på [5, side 4-8].

I resten af kapitlet benyttes følgende definition af at to punkter i X står i forbindelse med hinanden.

Definition 5.1 (Enkelt og multi-hopsforbindelse)

For ethvært positivt reelt tal R siges to punkter x_1 og x_2 i \mathbb{R} at stå i enkelthopsforbindelse, hvis $|x_1 - x_2| \leq R$. Dette noteres ved $x_1 \sim x_2$.

Hvis der findes punkter $x_1, x_2, \ldots, x_n \in X$ sådan at $x_1 \sim x_2 \sim \ldots \sim x_n$ siges x_1 og x_n at være flerhopforbundne, og det noteres $x_1 \sim x_n$.

Hvis x_i og x_j står i flerhopforbindelse og forbindelse kan gøres med n hop, men ikke med færre hop, noteres dette med $x_i \dot{\sim}_n x_j$.

5.1 FORBINDELSESTAL

Forbindelsestallene viser hvor mange punkter et givet punkt står i forbindelse med. Det højresidede enkelthops forbindelsestal, kaldes højresidet eftersom det kun medregner punkter til højre for det punkt for hvilket det udregnes.

Definition 5.2 (Højresidet enkelthop forbindelsestal)

Enkelthops forbindelsestallet $Y_{\rm scn}$ er defineret ved antallet af punkter der kan nås ved et enkelt hop fra x_0 .

$$\operatorname{scn}(x_0, X) = N(\{x_i \in X \setminus \{x_0\} | x_i \sim x_0\})$$

Det højreside
de enkelthop forbildelsestal Y_{scn_+} er defineret ved antallet af punkter til højre for x_0 der kan nås ved et enkelt hop fra x_0 .

$$\operatorname{scn}_+(x_0, X) = N(\{x_i \in X | x_i \sim x_0, x_i > x_0\})$$

I resten af rapporten bruges notationen $Y_{\text{scn}} = \text{scn}(x_0, Y)$ og $Y_{\text{scn}_+} = \text{scn}_+(x_0, Y)$ idet vi betrakter palmprocessen betinget $x_0 = 0$. Således bliver Y_{scn_+} antallet af punkter i intevallet (0, R). Det bør også bemærkes at ifølge(3.2) er $K(r) = \rho \mathbb{E}(Y_{\text{scn}})$ for R = r.

På tilsvarende vis defineres flerhops forbindelsestallet

Definition 5.3 (Højresidet flerhop forbindelsestal)

flerhop forbildelsestal Y_{mcn} er defineret ved antallet af punkter der kan nås ved et eller flere hop.

$$\operatorname{mcn}(x_0, X) = N(\{x_i \in X | x_i \dot{\sim} x_0\})$$

Det højresidede flerhop forbildelsestal Y_{mcn_+} er defineret ved antallet af punkter til højre for x_0 der kan nås ved et eller flere hop.

$$mcn_{+}(x_{0}, X) = N(\{x_{i} \in X | x_{i} \dot{\sim} x_{0}, x_{i} > x_{0}\})$$

Det venstresidede flerhop forbildelsestal $Y_{mcn_{-}}$ er defineret ved antallet af punkter til venstre for x_0 der kan nås ved et eller flere hop.

$$mcn_{-}(x_{0}, X) = N(\{x_{i} \in X | x_{i} \dot{\sim} x_{0}, x_{i} < x_{0}\})$$

Notationen Y_{mcn} , Y_{mcn_+} og Y_{mcn_-} benyttes på tilsvarende vis som for enkelthopforbindelsestallet. Følgende sætning beskriver fordelingen for Y_{mcn_+} og Y_{mcn_-} .

Sætning 5.4 [5, Proposition 1]

Hvis X er en stationær, ergodisk og ikke-tom MAP, med representation (C, D), så er $Y_{mcn_+} + 1$ fordelt med en phasetypefordeling med representation $(\boldsymbol{\alpha}_0, A)$, hvor $A = C^{-1}(e^{CR} - I)D$. Tilsvarende er $Y_{mcn_-} + 1$ fordelt med en phasetypefordeling med reprecentation $(\boldsymbol{\alpha}_0, \tilde{A})$, hvor $\tilde{A} = \tilde{C}^{-1}(e^{\tilde{C}R} - I)\tilde{D}$.

Bevis:

Betragt markovprocesen Z, hvor $Z_n = J_{X_n}$ for $X_n \sim X_0$, og $Z_n = 0$ for $X_n \sim X_0$. Hvis X_n ikke står i forbindelse med X_0 , kan heller ingen efterfølgende punkter stå i forbindelse med X_0 . Tilstanden 0 er derfor en absorberende tilstand. Det bemærkes at absorbtionstiden for Z er lig med $Y_{mcn_+} + 1$. Overgangsmatricen for Z kan skrives som

$$Z = \left[\begin{array}{cc} T & \mathbf{T}^0 \\ \mathbf{0} & 0 \end{array} \right]$$

Hvor T er en $m \times m$ matrix hvor den ij'te indgang angiver sandsynligheden for at gå fra j til i, medens \mathbf{T}^0 er en vektor hvor den i'te indgang er sandsynligheden for at gå fra i til 0. Fordelingen for absorbtionstiden for Z er dermed phasetypefordeling med representation ($\boldsymbol{\alpha}_0, T$).

Betinget med at $J_{X_n} = i$ er tætheden for at $J_{X_{n+1}} = j$ og $X_{n+1} = X_n + t$ givet ved $\mathbf{e}_i e^{Ct} D \mathbf{e}_j^T$ således bliver

$$T_{ji} = \int_0^R \mathbf{e}_i e^{Ct} D \mathbf{e}_j^\top dt = \left[\mathbf{e}_i C^{-1} e^{Ct} D \mathbf{e}_j^\top \right]_0^R = \mathbf{e}_i C^{-1} (e^{CR} - I) D \mathbf{e}_j^\top.$$

Hvilket betyder at T = A, hvilket beviser at $Y_{mcn_+} + 1$ er fordelt med en phasetypefordeling med reprecentation (α_0, A). Fordelingen for $Y_{mcn_-} + 1$ kan vises på tilsvarene måde.

På tilsvarende vis kan følgende sætning udledes

Sætning 5.5 [5, Proposition 2]

Hvis X er en stationær, ergodisk og ikke-tom MAP, med repesentation (C, D), så er $Y_{mcn} + 2$ fordelt med en phasetypefordeling med reprecentation $((\boldsymbol{\alpha}_{0,1}\mathbf{e}_1), \mathbf{0}, \ldots, \boldsymbol{\alpha}_{0,m}\mathbf{e}_m, \mathbf{0}), B)$, hvor B er en $m^2 \times m^2$ blokdiagonal matrix, hvor den j'te blok er givet ved

$$B_j = \begin{pmatrix} A & (I-A)\mathbf{1}^\top \mathbf{e}_j \\ O & \tilde{A} \end{pmatrix},$$

Bevis:

På samme måde som i beviset for 5.4 kan det vises at Y_{mcn_+} betinget med $J_0 = j$ er phasetypefordelt med reprecentation (\mathbf{e}_j, A) , hvor $A = C^{-1}(e^{CR} - I)D$, og at Y_{mcn_-} betinget med $J_0 = j$ en phasetypefordeling med reprecentation $(\mathbf{e}_j, \tilde{A})$, hvor $\tilde{A} = C^{-1}(e^{\tilde{C}R} - I)\tilde{D}$.

Fordelingen for $Y_{mcn+} = Y_{mcn_+} + 1 + Y_{mcn_-} + 1$ betinget med $J_0 = j$ bliver så foldningen af fordelinger ne for $Y_{mcn_+} + 1$ og $Y_{mcn_-} + 1$, eftersom disse er uafhængige betinget med $J_0 = j$. Foldningen af to phasetypefordelinger med representationerne (α_1, T_1) og (α_2, T_2) er ifølge [8, side 381] er en phasetypefordeling med representation

$$\left((\boldsymbol{\alpha_1}, \boldsymbol{\alpha_0 \alpha_2}), \left[\begin{array}{cc} T_1 & \mathbf{T}_1^0 \boldsymbol{\alpha_2} \\ 0 & T_2 \end{array} \right] \right),$$

hvor $\alpha_0 = 1 - \boldsymbol{\alpha}_1 \mathbf{1}^T$, og $\mathbf{T}_1^0 = (I - T)\mathbf{1}$. Således bliver fordelingen for $Y_{mcn} + 2$ betinget med $J_0 = j$ en phasetypefordeling med representation $((\mathbf{e}_j, 0), B_j)$

Den ubetingede fordeling for $Y_{mcn} + 2$ er en mixturfordeling af fordelingerne for $Y_{mcn} + 2$ betinget med $J_0 = j$ for j = 1, ..., m, hvor hver af disse fordelinger er givet vægten α_j . Heraf følger sætningen umiddelbart, eftersom en mixturfordeling af phasetypefordelinger med representationerne (α_1, T_1) og (α_2, T_2) og vægtene p og p - 1 netop er en phasetypefordeling med representation

$$\left((p\boldsymbol{\alpha_1},(1-p)\boldsymbol{\alpha_2}),\left[\begin{array}{cc}T_1&0\\0&T_2\end{array}\right]\right)$$

I kapitel 7 og kapitel 8 benyttes \bar{Y}_{mcn_+} til modelkontrol.

5.2 FORBINDELSESAFSTAND

I denne sektion defineres forbindelsafstanden og laplacetransformationen af denne udledes. Forbindelsesafstanden defineres ved den afstand et signal kan rejse ved hjælp af hop imellem punkterne i x.

Definition 5.6 (Forbindelsesafstand) Forbindelsesafstanden og dens højre- og venstre-sidede versioner defineres ved

$$cd_{+}(x, X) = \sup(x_{j} + R; x_{j} \in X, x_{j} \sim x) - x$$
$$cd_{-}(x, X) = x - \inf(x_{j} - R; x_{j} \in X, x_{j} \sim x)$$
$$cd(x, X) = cd_{+}(x, X) + cd_{-}(x, X)$$

Ud fra denne definition bliver $Y_{cd} = cd(x_0, Y_0)$, $Y_{cd+} = cd_+(x_0, Y_0)$ og $Y_{cd-} = cd_-(x_0, Y_0)$. Inden laplacetransformationen for denne udledes, precenteres og bevises et lemma der er nødvendigt for beviset af de efterfølgende sætninger.

Lemma 5.7 [5, Lemma 1]

Antag at funktionen f kan faktoriseres på følgende måde

$$f(t_1,\ldots,t_n) = \prod_{i=1}^n f_i(t_i)$$

og at laplacetransformationen af den aflede af

$$F_1(t) = \int_0^t \dots \int_0^{t-t_1-\dots-t_{n-1}} f(t_1,\dots,t_n) \mathrm{d}t_n \dots \mathrm{d}t_1$$

eksisterer. Da gælder der at

$$\int_0^\infty e^{-st} \frac{\mathrm{d}}{\mathrm{d}t} F_1(t) \mathrm{d}t = \prod_{i=1}^n \int_0^\infty e^{-st_i} f_i(t_i) \mathrm{d}t_i$$

Bevis:

$$F_{1}(t) = \int_{0}^{t} \dots \int_{0}^{t-t_{1}-\dots-t_{n-1}} \prod_{i=1}^{n} f_{i}(t_{i}) dt_{n} \dots dt_{1}$$
$$= \int_{0}^{t} f_{1}(t_{1}) F_{2}(t-t_{1}) dt$$
$$= (f_{1} * F_{2})(t).$$

hvor F_2 er defineret ved

$$F_2(t) = \int_0^\infty \dots \int_0^{t-t_1-\dots-t_{n-2}} \prod_{i=2}^n f_i(t_i) dt_n \dots dt_2.$$

Da F_2 er på samme form som F_1 kan samme fremgangsmåde benyttes rekursivt, således at

$$F_1(t) = (f_1 * f_2 * \dots f_{n-1} * F_n)(t),$$

hvor

$$F_n(t) = \int_0^t f_n(t_n) \mathrm{d}t_n.$$

Idet $\frac{d}{dt}(f * g)(t) = \left(f * \frac{d}{dt}g\right)(t)$, kan nu ved brug af infinitesimalregningens hovedsætning opnås at

$$\frac{\mathrm{d}}{\mathrm{d}t}F_1(t) = (f_1 * \ldots * f_n)(t)$$

Ved at tage laplace transformationen af $\frac{\mathrm{d}}{\mathrm{d}t}F_1(t)$ opnås lemma et.

Ved brug af dette lemma kan følgende sætning bevises.

Sætning 5.8 [5, Proposition 3]

Hvis X er en stationær, ergodisk og ikke-tom MAP, med repesentation (C, D), så har Y_{cd_+} og Y_{cd_-} laplace-Stieltjens-transformationerne henholdsvis

$$\mathcal{L}_{cd_+}(s) = -e^{-sR}g(s,\alpha_0,C,D)$$

og

$$\mathcal{L}_{cd_{-}}(s) = -e^{-sR}g(s,\alpha_0,\tilde{C},\tilde{D})$$

hvor

$$g(s, \alpha, C, D) = \alpha (I - (C - sI)^{-1} (e^{(C - sI)R} - I)D)^{-1} C^{-1} e^{CR} D \mathbf{1}^{\top}$$

Bevis:

Betragt for delingsfunktionen for $Y_{cd_{\pm}}-R$

$$\begin{split} F_{cd_{+}}^{0}(t) &= \sum_{n=1}^{\infty} P\left(\sum_{j=1}^{n-1} \leq t, t_{i} \leq R(i=1,\ldots,n-1)\right) \\ &= \sum_{n=1}^{\infty} \int_{0}^{t} \ldots \int_{0}^{t-t_{1}-\ldots-t_{n-2}} \prod_{i=1}^{n-1} \mathbf{1}(t_{i} \leq R) \int_{0}^{\infty} f_{0}(t_{1},\ldots,t_{n}) dt_{n} \ldots dt_{1} \\ &= -\alpha_{\mathbf{0}} \sum_{n=1}^{\infty} \int_{0}^{t} \ldots \int_{0}^{t-t_{1}-\ldots-t_{n-2}} \prod_{i=1}^{n-1} \mathbf{1}(t_{i} \leq R) e^{Ct_{1}} D \ldots e^{Ct_{n}} D dt_{n-1} \ldots dt_{1} C^{-1} e^{CR} D \mathbf{1}^{\top} \end{split}$$

Ved at tage Laplace-Stieltes-transformationen af $Y_{\rm cd_+}-R$ opnås

$$\begin{aligned} \mathcal{L}_{cd_{+}}^{0}(s) &= \int_{0}^{\infty} e^{-st} dF_{cd_{+}}^{0}(t) \\ &= -\alpha_{0} \sum_{n=1}^{\infty} \left(\prod_{i=1}^{n-1} \int_{0}^{\infty} \mathbf{1}(t_{i} \leq R) e^{(C-sI)t_{i}} D dt_{i} \right) C^{-1} e^{CR} \mathbf{1}^{\top} \\ &= -g(s, \alpha_{0}, C, D). \end{aligned}$$

For Y_{cd+} bliver Laplace-Stieltes-transformationen så

$$\mathcal{L}_{\mathrm{cd}_{+}}(s) = \mathbb{E}\left[e^{-sY_{\mathrm{cd}_{+}}}\right]$$
$$= e^{-sR}\mathbb{E}\left[e^{-s(Y_{\mathrm{cd}_{+}}-R)}\right]$$
$$= -e^{-sR}g(s,\alpha_{0},C,D)$$

Ved at følge samme fremgangsmåde kan resultatet for Y_{cd} vises.

Sætning 5.9 [5, Proposition 4]

Hvis X er en stationær, ergodisk og ikke-tom MAP, med repesentation (C, D), så har Y_{cd} laplace-Stieltjens-transformationen

$$\mathcal{L}_{\rm cd}(s) = \sum_{j=1}^{m} \alpha_{0,j} e^{-s2R} g(s, \mathbf{e}_j, C, D) g(s, \mathbf{e}_j, \tilde{C}, \tilde{D})$$

hvor

$$g(s, \alpha, C, D) = \alpha (I - (C - sI)^{-1} (e^{(C - sI)R} - I)D)^{-1} C^{-1} e^{CR} D \mathbf{1}^{\top}$$

Bevis:

Lad $\mathcal{L}_{cd_{+}}^{j}$ og $\mathcal{L}_{cd_{-}}^{j}$ være Laplace-Stieltjens-transformationen af henholdsvis $Y_{cd_{+}}$ og $Y_{cd_{-}}$ betinget med $J_{0} = j$. Ved samme fremgangsmåde som i beviset for sætning 5.8 fås

$$\mathcal{L}^{j}_{\mathrm{cd}_{+}} = -e^{-sR}g(s, \mathbf{e}_{j}, C, D)$$

og

$$\mathcal{L}_{\mathrm{cd}_{-}}^{j} = -e^{-sR}g(s, \mathbf{e}_{j}, \tilde{C}, \tilde{D}).$$

Eftersom Y_{cd_+} og Y_{cd_-} er uafhængige betinget med $J_0 = j$ bliver Laplace-Stieltjens-transformationen af Y_{cd} betinget med $J_0 = j$

$$\begin{aligned} \mathcal{L}_{cd}^{j}(s) &= & \mathbb{E}\left[e^{-sY_{cd}}|J_{0}=j\right] \\ &= & \mathbb{E}\left[e^{-sY_{cd_{-}}}e^{-sY_{cd_{+}}}|J_{0}=j\right] \\ &= & \mathcal{L}_{cd_{+}}^{j}\mathcal{L}_{cd_{-}}^{j} \\ &= & e^{-s2R}g(s,\mathbf{e}_{i},C,D)g(s,\mathbf{e}_{i},\tilde{C},\tilde{D}). \end{aligned}$$

For \mathcal{L}_{cd} fås så

$$\mathcal{L}_{cd}(s) = \sum_{j=1}^{m} \alpha_{0,j} \mathcal{L}_{cd}^{j}(s)$$
$$= \sum_{j=1}^{m} \alpha_{0,j} e^{-s2R} g(s, \mathbf{e}_{j}, C, D) g(s, \mathbf{e}_{j}, \tilde{C}, \tilde{D})$$

_	_
_	_

Når Laplace-Stieltestransformationen af Y_{cd_+} er interessant skyldes det at den udgør den momentfrembringende funktion for Y_{cd_+} . Laplace-Stieltestransformationen af den stokastiske variabel Z med tæthedsfunktion p(Z = z) = f(z) er givet ved $\mathcal{L}_Z(s) = \int_0^\infty e^{-sz} df(z)$, hvilket per definition er middelværdien af e^{-sz} . Eftersom $e^{-sz} = 1 + \frac{-sz}{1} + \frac{(-sz)^2}{2} + \dots$ bliver $\mathcal{L}_Z(s) = 1 + \frac{E(-sz)}{1} + \frac{E((-sz)^2)}{2} + \dots$ Således er $\frac{d^n}{ds^n}\mathcal{L}_Z(0) = E((-z)^n)$, hvilket betyder at sætning 5.8 og sætning 5.9 kan bruges til at beregne momenterne for Y_{cd_+} og Y_{cd} .

Det betyder at

$$\mathbb{E}(Y_{\mathrm{cd}_{+}}) = \left. \frac{\mathrm{d}}{\mathrm{d}s} \mathcal{L}_{cd_{+}}(s) \right|_{s=0} = \left. e^{-sR} \frac{\partial}{\partial s} g(s, \alpha_0, C, D) - Re^{-sR} g(s, \alpha_0, C, D) \right|_{s=0}$$

I kapitel 7 og kapitel 8 benyttes $\bar{Y}_{cd_{+}}$ til modelkontrol.

5.3 FORBINDELSESHOP

Forbindelseshoptallet defineres på følgende måde

Definition 5.10 (Forbindelseshoptal)

Ved forbindelseshoptallet forstås det mindste antal hop nødvendigt for at nå alle punkter, som kan nås igennem flerhopforbindelser.

$$ch(x_0, X) = \sup \left(\inf(n \in \mathbb{N}; x_0 \dot{\sim}_n x_j); x_j \in X \right)$$

Det højresidede Forbindelseshoptal er det mindste antal hop nødvendigt for at nå alle punkter til højre for x_0 , som kan nås igennem flerhopforbindelser.

$$ch_+(x_0, X) = \sup\left(\inf\{n \in \mathbb{N}; x_0 \dot{\sim}_n x_j\}; x_j \in X, x_j > x_0\right)$$

Igen benyttes notationen $Y_{\rm ch}$ og $Y_{\rm ch_{+}}$.



Figur 5.1 En illustration of s_1 , s_2 og s_3

Der vil i denne rapport ikke blive givet nogen gennerelle resultater for forbindelseshoptallet. I stedet vil følgende sætning beskrive middelværdien for forbindelseshoptallet, for poissonprocessen.

Sætning 5.11[5, Proposition 5]Hvis X er en stationer poisson
proces med intensitet $\lambda > 0$ da er

$$\mathbb{E}[Y_{\mathrm{ch}_+}] = \mathbb{E}[Y_{\mathrm{ch}_-}] = h(0) - 1$$

hvor funktionen h(t) er løsningen til integralligningen

$$h(t) = 1 + \int_0^{R-t} \lambda e^{-\lambda s} h(s) \mathrm{d}s.$$
(5.1)

Bevis:

Lad x_i , i = 1, 2, ... være defineret ved det fjerneste punkt i X, som kan nås ved i hop. Det vil sige $x_i = \sup\{x \in X | x \sim x_0\}$, og lad $x_0 = 0$. Definer envidere s_i , i = 1, 2, ... som $s_i = x_{i-1} - x_i + R$, samt $s_0 = 0$. Se figur 5.1.

Betragt fordelingen af s_{i+1} givet s_i og x_i . Da X ikke indeholder punkter intervallet $(x_i, x_i + s_i]$, lever fordelingen på mængden $[0, R - s_i) \bigcup \{R\}$, hvor $s_i = R$ svarer til at ingen nye punkter nås i det i'te hop. For $t \ge R$ bliver $P(s_{i+1} \le t | s_i, x_i) = 1$, for $t \in [R - s_i, R)$ fås $P(s_{i+1} \le t | s_i, x_i) = P(s_{i+1} \le R - t_i | s_i, x_i)$,

og for $t \in [0, R-s_i]$ fås

$$\begin{split} P(s_{i+1} \leq t | s_i, x_i) &= \sum_{k=0}^{\infty} P\left(s_{i+1} \leq t | N((x_i + s_i, x_i + R]) = k\right) P\left(N((x_i + s_i, x_i + R])) = k\right) \\ &= \sum_{k=0}^{\infty} P\left(N((x_i + R - t, x_i + R]) \neq 0 | N((x_i + s_i, x_{i-1} + R]) = k\right) \frac{(\lambda(R - s_i))^k}{k!} e^{-\lambda(R - s_i)} \\ &= \sum_{k=1}^{\infty} \left(1 - \left(\frac{R - t - s_i}{R - s_i}\right)^k\right) \frac{(\lambda(R - s_i))^k}{k!} e^{-\lambda(R - s_i)} \\ &= (e^{\lambda(R - s_i)} - 1)e^{-\lambda(R - s_i)} - \sum_{k=1}^{\infty} \left(\frac{(\lambda(R - t - s_i))^k}{k!}\right) e^{-\lambda(R - s_i)} \\ &= 1 - e^{-\lambda(R - s_i)} - \left(\sum_{k=0}^{\infty} \frac{(\lambda(R - t - s_i)^k}{k!} - 1\right) e^{-\lambda(R - s_i)} \\ &= 1 - e^{-t}. \end{split}$$

Dette vil sige at der er tale om en trunkeret eksponentialfordeling i intervallet $(0, R - s_i]$ mikset med en diskret sandsynlighed for $s_{i+1} = R$.

Eftersom fordelingen af s_i kun afhænger af s_{i-1} , kan vi betragte x_i , $i = 0, 1, \ldots$ som en markovkæde. Eftersom forbindelseshoptallet ikke medregner det hop, som ikke når nogen nye punkter, vil denne markovkæde have en absorbsionstid som nettop svarer til forbindelseshoptallet med 1 lagt til. Derfor må (5.1) være opfyldt ifølge sætning 5.12.

I beviset for sætning 5.11 benyttes følgende resultat, som er skrevet på baggrund af [9]

Sætning 5.12 Antag at Y er en stationær markovkæde med udfaldsrum Ω , da gælder for $A \subset \Omega$ at

$$k_A(t) = \mathbb{E}[K_A|Y_0 = t] = \begin{cases} 0 & \text{for } t \in A\\ 1 + \int_{\Omega} k_A(s) d\mu_t(s) & \text{for } t \in \Omega \setminus A \end{cases}$$
(5.2)

Hvor $K_A = \inf \{i \in \{0, 1, \ldots\} | y_s \in A\}$ og μ_t er sandsynlighedsmålet for Y_n givet $Y_{n-1} = t$

Bevis:

Først vises at k_A opfylder (5.2). At $k_A(t) = 0$ for $t \in A$ er trivielt. For $t \notin A$ gælder ifølge markovegenskaben og stationæriteten at

$$\mathbb{E}[K_A|Y_0 = t, Y_1 = s] = 1 + \mathbb{E}[K_A|Y_0 = s].$$

Envidere gælder at

$$\mathbb{E}[K_A|Y_0 = t] = \int_{\Omega} \mathbb{E}[K_A|Y_0 = t, Y_1 = s] d\mu_t(s)$$
$$= \int_{\Omega} (1 + \mathbb{E}[K_A|Y_0 = s]) d\mu_t(s)$$
$$= \int_{\Omega} (1 + k_A(s)) d\mu_t(s)$$

Г		

Sætning 5.11 giver et udtryk for middelværdien for Y_{ch_+} , men den indeholder en integralligning som ikke kan løses analytisk. I stedet kan løsningen approksimeres ved hjælp af sætning 5.13.

Lad $\mathcal H$ betege mængden af målelige funktioner $H:[0,R]\to [0,\infty)$ og definer φ ved

$$\varphi(H) = 1 + \int_0^{R-t} \lambda e^{-\lambda s} h(s) \mathrm{d}s.$$

Bemærk at for $H \in \mathcal{H}$ gælder at $\varphi(H) \in \mathcal{H}$. Envidere udstyres \mathcal{H} med supremumnormen

$$||H||_{\infty} = \sup_{t \in [0,R]} |H(t)|$$

Sætning 5.13 [5, Proposition 6] Lad $h_0 \in \mathcal{H}$ og definer funktionerne h_1, h_2, \dots ved $h_i = \varphi(h_{i-1})$. Da gælder $\lim_{i \to \infty} h_i = h$ (5.3) og $||h - h_i||_{\infty} \le (1 - e^{-\lambda R})^i e^{\lambda R} ||h - h_0||_{\infty}$ (5.4)

Bevis:

Lad H_1 og H_2 tilhøre \mathcal{H} . Da gælder

$$\begin{aligned} ||\varphi(H_1) - \varphi(H_2)||_{\infty} &= ||\int_0^{R-t} \lambda e^{-\lambda s} (H_1(s) - H_2(s)) \mathrm{d}s||_{\infty} \\ &\geq \int_0^{R-t} \lambda e^{-\lambda s} ||H_1 - H_2||_{\infty} \mathrm{d}s \\ &= ||H_1 - H_2||_{\infty} (1 - e^{-\lambda R}). \end{aligned}$$

Eftesom $\varphi(h) = h$ fås for $||h - h_i||_{\infty}$ at

$$\begin{aligned} ||h - h_i||_{\infty} &\geq (1 - e^{\lambda R}) ||h - h_{i-1}||_{\infty} \\ &\vdots \\ &\geq (1 - e^{-\lambda R})^i ||h - h_0||_{\infty}, \end{aligned}$$

Hvilket viser (5.3). På tilsvaende vis fås

$$|h_{i-1} - h_i||_{\infty} \ge (1 - e^{-\lambda R})^i ||h_0 - h_1||_{\infty}.$$
(5.5)

envidere ses det ved brug af trekantsuligheden at

$$||h - h_{0}||_{\infty} = \lim_{i \to \infty} ||h_{i} - h_{0}||_{\infty}$$

$$\leq \lim_{i \to \infty} \sum_{j=0}^{i-1} ||h_{j+1} - h_{j}||_{\infty}$$

$$\leq \lim_{i \to \infty} ||h_{1} - h_{0}||_{\infty} \sum_{j=0}^{i-1} (1 - e^{-\lambda R})^{i}$$

$$= e^{\lambda R} ||h_{1} - h_{0}||_{\infty}.$$
(5.6)

Ved at sætte (5.6) ind i (5.5) fås (5.4).

I kapitel 7 og kapitel 8 benyttes \bar{Y}_{ch_+} til modelkontrol.

DATABEHANDLING

I dette kapitel beskrives de data som i de næste kapitler bliver brugt til at inferere over. Herefter beskrives Metropolis-Hastings algoritmen. Denne algoritme er brugt til at approksimere posteriorfordelingerne for parametrene i modellerne, i de tilfælde hvor likelihooden for modellen ikke er på en form som er let overskuelig og til at beregne maksimum eller konfidensintervaller for.

6.1 PRESENTATION AF DATA

De data som er brugt til inferancen i denne rapport kommer fra institut for informatik ved universitetet i Mannheim, og er genereret i forbindelse med projektet Fleetnet. Dette projekt omhandler brugen af mobile netværksssystemer ved hjælp af trådløse forbindelser imellem biler. For mere information se [3], [4] og [1]. Hele datasættet inkluderer 40 filer med data, som beskriver bilers bevægelse på et stykke tysk motorvej. Ud af disse er valgt 3 sæt af data. Disse set kaldes her i rapporten for datasæt A, B og C. Datasæt A består af linje nr. 50 fra filen combined-108 nodes-120 tsteps-2 2 npkm-2 lpd-0. Datasæt B består af linje nr. 50 fra filen combined-459 nodes-120 tsteps-6 11 npkm-2 lpd-0.txt. Dataset C består, ligesom datasæt B, af linje nr. 50 fra filen combined-459 nodes-120 tsteps-6 11 npkm-2 lpd- θ .txt, dog indeholder dette sæt kun punkter som representer biler, der kører imod højre, det vil sige biler hvis position til tidspunkt 51 ligger til højre for deres possition til tidspunkt 50. At disse tre er udvalgt skyldes at de er rimeligt forskellige med hensyn til intensitet og autokorrelation af ventetiderne. På figur 6.2 ses autokorrelation og spredning af ventetiderne samenlignet med histogrammer for poissonprocesser med tilsvarende intensitet. Filerne indeholder data som for hver bil repræsenterer position på vejen, den fart hvormed de kører og deres akseleration. I hvert sæt er alle værdier sorteret fra som ikke representer positionen af en bil på vejen, således at hvert datasæt udgør et punktmønster på den reelle akse. Hvert datasæt repræsenterer et øjebliksbillede af hvordan bilerne er placeret på vejen. Man kan forestille sig at lignende data kan fås ved at tage et luftfoto af en lige vejstrækning, og notere positionerne for bilerne på vejen.

Det ses at datasæt A minder om poissonprocessen, mens datasæt B og især datasæt C har større autocorealition end forventet for possionprocessen. Alle tre datasæt har varians af mellemankomstiderne, som minder om poissonprocessen. I kapitel 7 findes flere sammenligninger mellem data og poisonprocessen.

6.1.1 Modelkontrol og Summary Statistics

Idet data er genereret med henblik på at undersøge egenskaberne for trådløse forbindelser imellem biler, er det naturligt at benytte summary statistics der beskriver disse forbindelser. Derfor benyttes de summary statistics som er defineret i forrige kapitel. Hvis hver bil er udstyret med en radiosender der har en rækkevide R, vil $Y_{\rm scn}$, angive antallet af biler, som ligger indenfor senderens rækkevide og $Y_{\rm mcn}$ angiver det samlede antal biler som står i kontakt med t_0 . Ligeledes vil $Y_{\rm cd}$ angive hvor lang en strækning forbindelsen strækker sig over og $Y_{\rm ch}$ vil vise hvor mange gange en besked skal sændes fra en bil til en anden for at nå til den fjernest beliggende bil indenfor netværket.

6.2 Metropolis-Hastings-algoritmen

Dette afsnit er baseret på [10, side 10-13], og beskriver Metropolis-Hastings-algoritmen. I afsnit A.1 findes en række definitioner og resultater som benyttes i beviset for at algoritmen virkelig konvergerer imod den ønskede fordeling.

Metropolis-Hastings-algoritmen er en Markovkæde-Montecarlo-metode til at kunne approksimere en sandsynlighedstæthed, π , som er kendt op til proportionalitet. Lad Z være en markovkæde i diskret tid, for hvilken sandsynligheden for at gå fra punktet z til punktet $z^* \neq z$ er givet ved

$$p(z, z^*) = q(z, z^*)a(z, z^*) = q(z, z^*)\min\{1, H(z, z^*)\},\$$





t

Figur 6.1 Plot af datasættene



Figur 6.2 På graferne ses histogrammer af autokorrelation og varians for 500 poisonprocesser, med intensitet svarende til 1 over gennemsnittet af ventetiderne for hvert af datasættene. De stiplede linjer viser autokorrelation og spredning for datasættene.

hvor $q(z, z^*)$ angiver sandsynlighedstætheden for at z^* bliver foreslået og $a(z, z^*) = \min\{1, H(z, z^*)\}$ er sandsynligheden for at forslaget bliver accepteret. Hastings-forholdet $H(z, z^*)$ er defineret ved

$$H(z, z^*) = \frac{\pi(z^*)q(z^*, z)}{\pi(z)q(z, z^*)}.$$

Som det ses forekommer π både over og under brøkstregen, hvilket er grunden til at det er tilstrækkeligt at kende π op til proportionalitet.

Således er Metropolis-Hastings algoritmen som følger

Algoritme 6.1 Lad begyndelsestilstanden $Z_0 = z$ være således, at $\pi(z) > 0$. For n = 0, 1, ... givet Z_n udføres følgende trin,

- generer $U_{n+1} \sim \mathcal{U}(0,1)$
- generer $Z_{n+1}^* \sim y \mapsto q(X_n, y)$
- Sxt

$$Z_{n+1} = \begin{cases} Z_{n+1}^* & \text{hvis } U_{n+1} \le H(Z_n, Z_{n+1}^*) \\ Z_n & \text{ellers} \end{cases}$$

At algoritmen kan bruges til simulation af π vises af følgende sætning

Sætning 6.2

Hvis der for alle $z, z^* \mod \pi(z) > 0$ og $\pi(z^*) > 0$ gælder at $q(z, z^*) > 0$, vil der for markovkæden Z, dannet ved brug af Metropolis-Hastings algoritmen gælle at

$$P(Z_n \in A) \to \pi(A) \text{for} n \to \infty,$$

forudsat at $Z_0 = z_0 \mod \pi(z_0) > 0$.

Bevis:

Ifølge sætning A.11 vil sætningen gælde hvis Z har π som stationær tæthed, er π -ireducibel, aperiodisk og Harris-rekurrent.

Antag at $H(z,z^*) \leq 1$. Da gælder at $H(z,z^*) \geq 1$, således at $a(z^*,z) = 1$. Dette betyder at for $H(z,z^*) \leq 1$ gælder

$$\begin{aligned} \pi(z)a(z,z^*) &= \pi(z)q(z,z^*)H(z,z^*) \\ &= \pi(z)q(z,z^*)\frac{\pi(z^*)q(z^*,z)}{\pi(z)q(z,z^*)} \\ &= \pi(z^*)q(z^*,z) \\ &= \pi(z^*)a(z^*,z). \end{aligned}$$

På tilsvarende vis gælder det samme for $H(z, z^*) \ge 1$. Derfor opfylder markovkæden Z detailed balance condition, hvilket medfører at den er reversibel, med stationær tæthed π . Yddermere medfører $q(z, z^*) > 0$ at $p(z, z^*) > 0$ for alle $z, z^* \mod \pi(z) > 0$ og $\pi(z^*) > 0$, hvilket viser at Z er π -ireducibel, samtidig følger Harris-rekurrens også af disse betingelser ifølge [10, Theorem 4]. Endelig er Z aperiodisk eftersom den har positiv sandsynlighed for at $Z_n = Z_{n+1}$.

I afsnit B.2 findes en Metropolis-Hastings-algoritme som funktion i R.

POISSONPROCESSEN

I dette kapitel benyttes poissonprocessen som model for data. For hvert sæt af data bliver posteriorfordelingen for poissonprocessens intensitet beregnet ved brug af en konjugeret prior. Derefter kontrolleres det hvor godt modellen beskriver data.

7.1 **POSSIONPROCES**

Den simpleste form for MAP er possion
procesen. Det er derfor naturligt at starte med at samenligne data med denne proces. Possion
processen kan udtrykkes som en MAP, hvori der indgår en enketl
t parameter $\beta > 0.$

$$D = [\beta] \text{ og } C = [-\beta] \text{ og } \boldsymbol{\alpha} = (1).$$

Poissobprocessen kan tænkes at repræsentere en situation hvor alle biler på vejen kører uafhængigt af hinanden.

7.2 ESTIMATION AF PARAMETRENE.

For poissonprocessen reduceres sætning 4.4 til

$$\alpha e^{Ct_1} D e^{Ct_2} D \dots e^{Ct_n} D \mathbf{1}^{\top} = \beta^n e^{-\beta \sum_{i=1}^n t_i}$$

eftersom mellemankomsttiderne er fordelt med uafhængige, ensfordelte eksponentialfordelinger.

Som prior vælges en gammafordeling idet den udgør en konjugeret prior for eksponentialfordelingen. Under antagelse af at bilisterne kører med et minimum af ansvarlighed må det forventes at $\beta \leq 1$ idet $\beta = 1$ svarer til en middelafstand på en meter imellem bilerne. Ligeledes kan det næppe heller forventes at de tyske motorveje ligger så øde hen at der er under en bil per kilometer, hviket svarer til $\beta \geq 0.001$. Ved at lade formparameteren være a = 0.2307 og skalaparameteren b = 1.639 fås en middelværdi for β på 0.1408 og en median på 0.0201. Dette svarer til en middelafstand på henholdsvis 7.10 og 48.3 meter imellem bilerne, hvilket virker plausibelt. Samtidig ligger 95% af sandsynlighedsmassen imellem 0.001 og 1.00.

Posteriorfordelingen for β bliver så

$$p(\beta) \propto \beta^{n} e^{-\beta \sum_{i=1}^{n} t_{i}} \frac{b^{a}}{\Gamma(a)} \beta^{a-1} e^{-b\beta}$$
$$= \frac{b^{a}}{\Gamma(a)} \beta^{a+n-1} e^{-(b+\sum_{i=1}^{n} t_{i})\beta}$$

hvilket betyder at posteriorfordelingen bliver $\text{Gamma}(a+n, b+\sum_{i=1}^{n} t_i)$. Posteriorfordelingen for datasæt A bliver Gamma(91.23, 11654.04), hvilket giver en middelværdi for β på 0.00782, og et 95%-CPI på [0.007262, 0.00951].

For datasæt B bliver posteriorfordelingen Gamma(393.23, 12061.44), hvilket giver en middelværdi for β på 0.0326, og et 95%-CPI på [0.0315, 0.0359].

Endeligt bliver posteriorfordelingen for datasæt C Gamma(246.23, 11985.84), hvilket giver en middelværdi for β på 0.0205, og et 95%-CPI på [0.0181, 0.0231].

7.3 MODELKONTROL

I dette afsnit kontroleres hvor godt poissonprocessen beskriver data ved at sammenligne de målte værdier af summary statistics med punktmønstre trukket fra den posteriorprediktive fordeling. Ved den posteriorprediktive fordeling forstås fordelingen af punktmønstre genereret ved først at trække β fra posteriorfordelingen, for derefter at trække et punktmønster fra poissonprocessen med intensitet α . Ved



Figur 7.1 De fuldt optrukne linjer viser \bar{Y}_{scn_+} for de tre datasæt. De stiplede linjer angiver et 95%-konfidensinterval.

estimeringen af de forskællige summary statistics for forskellige punktmønstre er de sidste tyve punkter i hvert punktmønster ikke medtaget. Det sker for at undgå at medregne værdier af $Y_{mcn_+} + 1$ som er begrænset, af det ydderste punkt i datasættet af naturlige årsager ikke har forbindelse til nogen efterfølgende punkter. I de tilfælde hvor en teoretisk værdi er beregnet, $\mathbb{E}(Y_{mcn_+})$, $\mathbb{E}(Y_{cd_+})$ og $\mathbb{E}(Y_{ch_+})$, er middelværdien for β benyttet.

7.3.1 Forbindelsestal

På figur 7.1 ses grafer for \bar{Y}_{scn_+} for de tre datasæt. Det ses at der er god overensstemmelse imellem \bar{Y}_{scn_+} for datasæt A og de genererede punktmønstre. For R < 50 har datasæt A færre par af punkter indenfor en afstand af R end det typisk kan forventes for poissonprocessen, men værdien af \bar{Y}_{mcn} ligger stadig indenfor 95%-konfidensintervallet.

Punkterne i datasæt har B vesentligt færre naboer indenfor en afstand af R < 30 end det kan forventes for poissonprocessen. Det samme kan siges om datasæt C, hvor tendensen er mere udtalt.

Ved brug af sætning 5.4 ses det at fordelingen $Y_{mcn_+} + 1$ er en phasetypefordeling med representation $(1, 1 - e^{-\beta R})$. Det betyder at fordelingsfunktionen for $Y_{mcn_+} + 1$ bliver

$$F(k) = 1 - \alpha_0 A^k \mathbf{1}^\top = 1 - (1 - e^{-\beta R})^k.$$

Herudfra er middelværdierne for Y_{mcn_+} beregnet, og disse er plottet som funktion af R figur 7.2, sammen med grafer for gennemsnittet af $Y_{mcn_+} + 1$. Derudover er der for hvert datasæt genereret 500 poisson-





Figur 7.2 plots for $\bar{Y}_{mcn_{+}}$ for de tre datasæt. De stiplede linier viser de i datasættene målte værdier for $\bar{Y}_{mcn_{+}}$. De prikkede viser middelværdien for $\bar{Y}_{mcn_{+}}$, beregnet ved hjælp af sætning 5.4. De fuldt optrukne linjer angiver et 95%-konfidensinterval approksimeret ved hjælp af 500 datasæt gennereret fra den posteriorpredektive fordeling.

proceser med $\beta = \mathbb{E}(\beta)$. Ved hjælp af disse er 95%-konfidensintervaller approksimeret og tegnet ind i grafen.

På figur 7.2 ses grafer der sammenligner de i datasæt A opserverede værdier for \overline{Y}_{mcn_+} , med $\mathbb{E}(Y_{mcn_+})$. Det ses at der er god overensstemmelse imellem data og model for både R > 60 men værdierne ligger under hvad modellen forudsiger for R < 60.

Tilsvarende grafer ses også for datasæt B. Det ses tydeligt at poisson
procesen ikke er en god beskrivelse af datasætet for R > 40, i
det der er vesentligt flere større værdier end for poisson
procesen. Der er således for store værdier af R samlet set flere punkter der står i for
bindelse med et givet x_0 , end for poisson
processen, uden at der et tilsvarende større antal som står i enkelthopforbindelse. Dette kan forklares ved at der forekommer færre mellemankomst
tider over 40 end for poison
processen. En alternativ forklaring kan være at de længere mellemankomst
tider ligger samlet så der er områder hvor der lokalt forekommer færre lange mellemankomst
tider. Samtidig ligger værdien af \bar{Y}_{mcn+} under konfiden
sintervallet, hvilket tyder på at der i datasættet er færre mellemankomst
tider under 30 end det kan forventes ud fra modellen.

For datasæt C gælder de samme opservationer som for datasæt B. Bare i højere grad.

7.3.2 Forbindelsafstand

Fra sætning 5.8 kendes den momentfrembringende funktion. Ved hjælp fra den kan middelværdien for $Y_{cd_{\perp}}$ beregnes. I afsnit B.3 ses en funktion der beregner denne middelværdi.



Figur 7.3 Plots for $\bar{Y}_{cd_{+}}$ for de tre datasæt. De stiplede linier viser de i datasættene målte værdier for $\bar{Y}_{cd_{+}}$. De prikkede viser middelværdien for $\bar{Y}_{cd_{+}}$, beregnet ved hjælp af sætning 5.4. De fuldt optrukne linjer angiver et 95%-konfidensinterval approksimeret ved hjælp af 500 datasæt gennereret fra den posteriorpredektive fordeling.

Ligesom for Y_{mcn_+} , er den beregnede middelværdi og det målte gennemsnit for Y_{cd_+} plottet på figur 7.3 sammen med et estimationen af et 95%-konfidensinterval lavet ud fra 500 punktmønstre trukket fra den posteriorprediktive model.

Det ses for datasæt A at der med henhold til \overline{Y}_{cd_+} ingen modstrid er imellem modellen og data. For datasæt B ses på figur 7.3 at for R = 10 til R = 15 ligger \overline{Y}_{ch_+} under hvad man kan forvente for poissonprocessen, medens den ligger over for R = 30 60 R = 60. Dette tyder igen på at datasættet er mere regulært end poissonprocessen, og derfor har færre meget korte og meget lange ventetider.

For datasæt C ses at for værdier af R omkring R = 15 ligger \overline{Y}_{ch_+} under hvad man kan forvente for poissonprocessen, medens den ligger over for R > 30 i endu mere udtalt grad end for datasæt B. Dette tyder som sagt på at datasættet er mere regulært end poissonprocessen, og derfor har færre meget korte og meget lange ventetider.

7.3.3 Forbindelseshoptal

Fra sætning 5.11 kendes middelværdien for Y_{ch_+} , og fra sætning 5.13 kendes metoden til at approksimere den. I afsnit B.3 findes funktioner i R til at approksimere middelværdien for Y_{ch_+} . I kapittel B ses en funktion i R til at approksimere $E(Y_{ch_+})$.

I 7.4 ses grafer for \overline{Y}_{ch_+} og $E(Y_{ch_+})$. Derudover er der ligesom tidligere for hvert af datasættene genereret 500 poissonprocesser, som er brugt til estimation af et 95%-konfidensinterval.



Figur 7.4 Grafer for $\bar{Y}_{ch_{+}}$ for de tre datasæt. De stiplede linier viser de i datasættene målte værdier for $\bar{Y}_{ch_{+}}$. De prikkede viser middelværdien for $\bar{Y}_{ch_{+}}$, beregnet ved hjælp af sætning 5.4. De fuldt optrukne linjer angiver et 95%-konfidensinterval approksimeret ved hjælp af 500 datasæt gennereret fra den posteriorpredektive fordeling.

Det ses for datasæt A, at der med henhold til \bar{Y}_{ch_+} er overensstemmelse imellem modellen og data for R < 60. For mindre værdier af R ligger \bar{Y}_{ch_+} under konfidensintervallet. Igen kan dette forklares ved at der er færre korte mellemankomsttider.

For datasæt B ses på figur 7.4 at for R = 20 ligger \overline{Y}_{ch_+} under hvad man kan forvente for poissonprocessen. Dette tyder ligesom for datasæt A på at der er færre punkter der ligger tæt på hinanden end for poissonprocessen. For værdier af R imellem 30 og R = 50 kan man i datasættet foretage vesentligt flere hop imellem punkterne, før man støder på en ventetid der er større end R, end man ville kunne ved possionprocessen. Dette kan skyldes at der er færre lange mellemankomsttider, eller at de kortere mellemankomsttider ligger samlet, sådan at man i en samling af kortere mellemankomsttider kan foretage et større antal hop inden man kommer til et område med længere mellemankomsttider.

For datasæt C ses ligesom for datasæt B, at der er færre mellemankomsttider under 25 end for poissonprocessen. For R > 30 kan man dog foretage vesentligt flere hop end forudsagt af modellen. Igen kan dette tyde på at der findes færre lange mellemankomsttider for datasætet end for poissonprocessen.

Overordnet set kan det siges at datasæt A med en vis rimelighed kan modeleres ved en poissonprocess, selvom der er færre mellemankomsttider under 30 end man kan forvente for poissonprocessen. Både datasæt B og datasæt C kan dårligt beskrives af poissonprocessen, da der er vesentlig modstrid imellem model og data, for både større og mindre værdier af R. Det kan forventes at datasættene bedre kan beskrives af en model der er mere regulær. En sådan model præsenteres i det følgende kapittel.

MARKOVMODULERET RENEWALPROCES

I dette kapitel beskrives en model for data. Herefter benyttes Metropolis-Hastings algoritmen til at approksimere posteriorfordelinger for parametrene, hvorefter modellens egnethed til at beskrive data kontrolleres.

8.1 MARKOVMODULERET RENEWALPROCES

Ofte vil man på længere vejstrækninger se at der nogle steder er relativt få biler, der kører uafhængigt af hinanden, og andre steder vil bilerne køre tæt eller endda i kø. Ud fra denne simple opfattelse af hvordan bilerne fordeler sig på vejen, kan man opstille følgende MAP som model for data:

$$D = \begin{bmatrix} 0 & \lambda_1(1-q) & \lambda_1 q \\ 0 & 0 & 0 \\ 0 & \lambda_2 & \beta \end{bmatrix} \text{ og } C = \begin{bmatrix} -\lambda_1 & 0 & 0 \\ \lambda_1 & -\lambda_1 & 0 \\ 0 & 0 & -\beta - \lambda_2 \end{bmatrix} \text{ og } \alpha = \left(\frac{\lambda_2}{p}, \frac{\lambda_2}{p}, \frac{\lambda_1 q}{p}\right),$$

hvor $p = 2\lambda_2 + \lambda_1 q$. De områder hvor bilerne kører tæt repræsenteres af tilstand 1 og 2, medens de områder hvor de kører med større afstand repræsenteres af tilstand 3. På den måde kommer afstandene imellem bilerne til at være mere regulære for de biler som kører tæt på hinanden, medens bilerne på de mindre trafikerede strækninger kører uafhængigt af hinanden. Da mellemankomstiderne er eksponintialfordelte i de områder hvor J = 3 og afstanden imellem bilerne er gammafordelt når J er i tilstand 1 og 2, kalles denne model for en Markovmodelleret renewalproces.

Eftersom det ønskes at tilstand 3 repræsentere strækninger med større afstand imellem bilerne bør middelværdien for den tid markovprocessen J tilbringer i tilstand 3 være større end middelværdien for afstanden imellem bilerne på de tættere trafikerede strækninger. Ligeledes bør også den gennemsnitlige afstand imellem bilerne være større på de mindre trafikerede strækninger. Derfor tilføjes følgende begrænsninger på de variable

$$\frac{3}{\lambda_1} < \frac{1}{\lambda_2} \mathrm{og} \ \frac{3}{\lambda_1} < \frac{1}{\beta} \mathrm{og} \ q \in (0,1)$$

8.2 ESTIMATION AF PARAMETRENE.

Det er ingen let sag at beregne maximum likelihood estimatet for parametrene i denne model, da udtrykket for liklyhooden indeholder pruduktet af en lang række matricer. I stedet benyttes derfor en baysiansk fremgansmåde, hvor posteriorfordelingen af parametrene approksimeres ved hjælp af en Metropolis hastings algoritme. En beskrivelse af algoritmen findes i sætning 6.2. Som prior for parametrene er brugt en uegentlig prior der for alle parametre er proportional med 1.

Parameter	1. Kvartil	Median	Middelværdi	3. Kvartil	95%-CPI
λ_1	0.0201	0.0249	0.0263	0.0302	[0.0155, 0.0451]
λ_2	0.0017	0.0027	0.0028	0.0037	[0.0004, 0.0057]
β	0.0026	0.0039	0.0039	0.0051	[0.0007, 0.0074]
q	0.5308	0.6994	0.6785	0.8456	[0.2657, 0.9818]

Tabel 8.1 Beskrivelse af fordelingerne for parametrene i modellen for datasæt A.

I figur 8.1, figur 8.2 og figur 8.3 ses trace og autokorrelation for parametrene for datasættene, og i tabel 8.1,8.2 og 8.3 ses en opsumering af posteriorfordelingerne.

Det ser ud til at modellen giver en egentlig posteriorfordeling, selvom en uegentlig prior er benyttet.



Figur 8.1 Trace, autokorrelation og histogram for parametrene for datasæt A

Parameter	1. Kvartil	Median	Middelværdi	3. Kvartil	95%-CPI
λ_1	0.06368	0.07001	0.07428	0.08262	[0.05630880, 0.09068483]
λ_2	0.007047	0.013700	0.012430	0.017660	[0.001266049, 0.018989510]
β	0.01545	0.01867	0.01920	0.02296	[0.0100895, 0.0254438]
q	0.5447	0.7664	0.6975	0.8983	[0.0100895, 0.0254438]

Tabel 8.2 Beskrivelse af fordelingerne for parametrene i modellen for datasæt B.

Parameter	1. Kvartil	Median	Middelværdi	3. Kvartil	95%-CPI
λ_1	0.0533	0.0558	0.0560	0.0587	[0.0486, 0.0649]
λ_2	0.0014	0.0020	0.0023	0.0028	[0.0006, 0.0054]
β	0.0091	0.0105	0.0105	0.0119	[0.0067, 0.0145]
q	0.0444	0.0654	0.0748	0.0926	[0.0200, 0.1904]

Tabel 8.3 Beskrivelse af fordelingerne for parametrene i modellen for datasæt C.



Figur 8.2 Trace og autokorrelation for parametrene for datasæt B

8.3 MODELKONTROL

I dette afsnit kontrolleres hvor godt denne model beskriver data ved at sammenligne de målte værdier af summaristatistic med de forvendete ved brug af den posteriorpredektive fordeling for modellen. Ved den posteriorprediktive fordeling, forstås fordelingen af punktmønstre dannet ved først at trække et sæt af parametre $(\lambda_1, \lambda_2, \beta, q)$ fra posteriorfordelingen, og derefter trække et punktmønster fra en MAP beskrevet af (8.1), ved brug af disse parametre. Ligessom i forige kapitel unlades de sidste punkter i hvert punktmønster ved approktimation af summary statistics, og middelværdierne for Y_{cd_+} og Y_{scn_+} beregnes ved brug af middelværdierne for parametrene.

8.3.1 Forbindelstal

På figur 8.4 ses grafer der viser viser \bar{Y}_{scn_+} for de tre datasæt, sammenlignet med estimater for et 95%konfidensinterval. Disse er lavet ud fra 500 punktmønstre trukket fra den posteriorprediktive fordeling. Det ses at \bar{Y}_{scn_+} for datasæt C er en smule mindre end modellen forudsiger for R. Ved større værdier er der overensstemmelse immelem model og data. Ligeledes er der ingen modstrid imellem modellen og datasættene A og B.

På ses at der igen ingen modstrid er imellem modellen og datasæt A. For datasæt B og C er \bar{Y}_{mcn_+} ved værdier af R omkring 20 en smule lavere en det ud fra modellen bør forventes. For R omkring 40 står punkterne i kontakt med flere andre punkter end modellen beskriver, for disse to datasæt. Dette kan ligesom nævnt i forrige kapittel skyldes at datasættene er mere rægulære end modellen forudsiger.



Figur 8.3 Trace og autokorrelation for parametrene for datasæt BC

8.3.2 Forbindelsafstand

Figur 8.6, viser middelværdier for Y_{cd_+} , sammenlignet med gennemsnit for \overline{Y}_{cd_+} for datasættene og konfidensintervaller, lavet på samme måde som tidligere beskrevet.

Graferne for \overline{Y}_{cd_+} minder om dem i for \overline{Y}_{mcn_+} . Igen ligger det målte gennemsnit for datasæt B og datasæt C udenfor konfidensintervallet i et lille område omkring R = 20. Der er betydelig større overensstemmelse imellem model og data, for disse datasæt end ved brug af poissonprocessen. For datasæt A er der igen ingen synlig modstrid imellem data og model.

8.3.3 Forbindelseshoptal

Figur 8.7 viser \overline{Y}_{ch_+} for datasættene sammenlignet med et 95%-konfidensinterval, som er fremstillet på samme måde som tidligere beskrevet. For datasæt A ses at \overline{Y}_{ch_+} holder sig indenfor 95%-konfidensintervalet. For datasæt B og C ligger \overline{Y}_{ch_+} 0ver det forventede for R omkring de 40. Dette er langt mere tydeligt for datasæt B end for datasæt C. Ligelædes ligger \overline{Y}_{ch_+} for datasæt C også udenfor 95%-konfidensintervalet ved R = 20. Modellen kan derfor med større rimelighed siges at beskrive datasæt B end datasæt C. For alle tre datasæt gælder det at denne model passer bedre til data med hensyn til \overline{Y}_{ch_+} , end poissonprocessen.

8.4 FORSLAG TIL NY MODEL

Selvom den i dette kapittel beskrevne model passer vesentligt bedre på data end poissonprocessen, er der stadig modstrid imellem modellen og datasættene B og C. Denne modstrid skyldes tilsyneladende at datasættene indeholder færre mellemankomsttider omkring 20 og flere omkring 40 end modellen



4

ო

2

~

0

0

 $\overline{\gamma}_{\rm scn+}$

Datasæt A

150

R

2.0

1.5

0.5

0.0

0 50

 $\overline{Y}_{\rm scn+}$



80

R

120



Figur 8.4 De stiplede linjer viser \bar{Y}_{scn_+} for de tre datasæt. De fuldt optrukne linjer angiver et 95%-konfidensinterval.

40



Figur 8.5 plots for $\overline{Y}_{mcn_{+}}$ for de tre datasæt. De stiplede linier viser de i datasættene målte værdier for $\overline{Y}_{mcn_{+}}$. De prikkede viser middelværdien $\mathbb{E}(Y_{mcn_{+}})$, beregnet ved hjælp af sætning 5.4. De fuldt optrukne linjer angiver et 95%-konfidensinterval approksimeret ved hjælp af 500 datasæt gennereret fra den posteriorpredektive fordeling.



Figur 8.6 Plots for $\bar{Y}_{cd_{+}}$ for de tre datasæt. De stiplede linier viser de i datasættene målte værdier for $\bar{Y}_{cd_{+}}$. De prikkede viser middelværdien for $\bar{Y}_{cd_{+}}$, beregnet ved hjælp af sætning 5.4. De fuldt optrukne linjer angiver et 95%-konfidensinterval approksimeret ved hjælp af 500 datasæt gennereret fra den posteriorpredektive fordeling.



Figur 8.7 Grafer for \bar{Y}_{ch_+} for de tre datasæt. De stiplede linier viser de i datasættene målte værdier for \bar{Y}_{ch_+} . De fuldt optrukne linjer angiver et 95%-konfidensinterval approksimeret ved hjælp af 500 datasæt gennereret fra den posteriorpredektive fordeling.

forudsiger. Det er derfor tænkeligt at en mere rægulær model kan beskrive disse datasæt bedre. Et forslag til en sådan model er

hvor $p = 3\lambda_2 + \lambda_1 q$. Denne model minder meget om den der allerede er beskrevet i dette kapittel, men i områderne med tæt trafik vil afstandene imellem bilerne variere mindre. Dette vil muligvis føre til færre helt korte mellemankomsttider, og flere omkring 40.

KONKLUTION

I løbet af rapporten er det blevet beskrevet hvordan stokastiske punktmønstre på den reelle akse kan beskrives ved hjælp af punktprocesser, herunder markov-arrival-processer. Der er givet eksempler på markov-arrival-processer, der er både regulære og tiltrækkende.

Tre punktmønstre fra projektet Fleetnet er forsøgt beskrevet ved hjælp af to forskellige modeller, en poissonproces og en mere kompleks model kaldet en markovmoduleret renewalproces. Parametrene for disse modeller er tilpasset data ved hjælp af baysiansk statistik, og deres egnethed er kontrolleret ved hjælp af en række summary statistics, som beskriver antallet af punkter der kan nås ved hjælp af spring på en afstand R.

Datasæt A kan med en vis rimelighed beskrives af både poissonprocessen, men beskrives vesentligt bedre af den markovmodulerede renewalproces. Datasæt B og datasæt C kan bedst beskrives af den markovmodulerede renewalproces, men begge disse datasæt har vesentligt flere mellemankomsttider omkring 40 end modellen forudsiger. Det er formodentligt muligt at opstille modeller som beskriver data bedre end de to som er anvendt i denne rapport. et forslag til en sådan model kunne være en ny markovmoduleret renewalproces beskrevet ved (8.1).

MARKOVKÆDER

I dette kapitel defineres markovkæder i diskret og kontinuert tid. En række af egenskaber for markovkæder defineres, og nogle resultater for markovkæder presenteres. Kapitlet er skrevet på baggrund af [10, side 4-8]

Definition A.1 (Markovkæde)

En Markovkæde er en stokastisk proces (X_0, X_1, \ldots) med tilstandsrum Ω , for hvilken det gælder, at

$$P(X_{n+1} \in A | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = P(X_{n+1} \in A | X_n = x_n)$$

= $P(x_n, A)$

for alle $n \in \mathbb{Z}_+$, alle $A \subseteq \Omega$ og alle $x_1, x_2, \ldots, x_n \in \Omega$. Sandsynlighedsfunktionen P kaldes overgangskernen. Fordelingen for X_0 kaldes begyndelsesfordelingen.

Overgangskernen kan i nogle tilfælde beskrives ved hjælp af tæthedsfunktionen p(x, y) for et givet x, hvorom det for alle $A \subseteq \Omega$ gælder, at

$$P(x_n, A) = \int_A p(x, y) \mathrm{d}y.$$

Således er p(x, y) sandsynlighedstætheden for X_{n+1} under betingelsen, at $X_n = x$. Bemærk at det er muligt at beregne fordelingen for X_n for et vilkårligt n > 0, hvis overgangskernen og begyndelsesfordelingen er kendt.

Hvis $\Omega = \{s_1, s_2, \ldots, s_m\}$ kan overgangskærnen repræsenteres ved en overgangsmatricen

$$P = \begin{pmatrix} P(s_1, s_1) & P(s_1, s_2) & \cdots & P(s_1, s_m) \\ P(s_2, s_1) & P(s_2, s_2) & \cdots & P(s_2, s_m) \\ \vdots & \vdots & \ddots & \vdots \\ P(s_m, s_1) & P(s_m, s_2) & \cdots & P(s_n, s_m) \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}$$

Per definition gælder at $\sum_{j=1}^{m} p_{ij} = 1$ for alle $i = 1, \dots, m$.

A.1 KONVERGENS FOR MARKOVKÆDER

De Markovkæder, der er interessante i forbindelse med dette projekt, har en række egenskaber, som her vil blive beskrevet.

Definition A.2 (Stationær fordeling og tæthed) En fordeling Π kaldes en stationær fordeling med stationær tæthed π for Markovkæden med tilstandsrum Ω og overgangskerne P, hvis der for alle $A \subseteq \Omega$ gælder at

$$\int_{\Omega} \pi(x) P(x, A) \mathbf{x} = \int_{A} \pi(x) \mathbf{x} = \Pi(A).$$

En stærkere egenskab for Markovkæder er reversibilitet.

Definition A.3 (Reversibilitet) En Markovkæde kaldes reversibel, hvis der findes en fordeling Π , for hvilken det gælder, at (X_0, X_1) og (X_1, X_0) har samme fordeling, når $X_0 \sim \Pi$.

For reversible Markovkæder gælder følgende sætning.

Sætning A.4

En reversibel Markovkæde har en stationær tæthed.

Bevis:

Lad X være en reversibel Markovkæde, og lad π være tætheden, for hvilken X opfylder Definition A.3. Antag nu, at π ikke er en stationær tæthed. Da vil X_0 og X_1 ikke være ensfordelte, når $X_0 \sim \pi$, hvilket betyder, at (X_0, X_1) og (X_1, X_0) heller ikke kan have samme fordeling, hvorved modstrid opstår. Til gengæld er en Markovkæde med en stationær tæthed ikke nødvendigvis reversibel.

Definition A.5 (detailed balance condition) Hvis der for en Markovkæde gælder, at der findes en fordeling Π , således at

 $\pi(x)p(x,y) = \pi(y)p(y,x)$

for all $x, y \in \Omega$, siges Markovkæden at opfylde detailed balance condition (DBC).

DBC er en stærkere egenskab end reversibilitet, som det fremgår af følgende sætning.

Sætning A.6

En Markovkæde, der opfylder DBC, er reversibel og har en stationær fordeling.

Bevis:

Lad X være en Markovkæde, som opfylder DBC, og lad $X_0 \sim \pi$. Betragt nu fordelingen for (X_0, X_1) , som er givet ved

$$p(X_0 = x, X_1 = y) = p(X_0 = x)p(X_1 = y | X_0 = x)$$

= $\pi(x)p(x, y)$
= $\pi(y)p(y, x)$
= $p(X_0 = y, X_1 = x),$

hvorved reversibiliteten er opfyldt. Ifølge sætning A.4 har X derfor en stationer tæthed.

En anden og meget vigtig egenskab for Markovkæder er irreducibilitet.

Definition A.7 (irreducibilitet)

Lad Π være en stationær fordeling for en Markovkæde. Markovkæden siges da at være Π -irreducibel, hvis der for alle $A \subseteq \Omega \mod \Pi(A) > 0$ og alle $x \in \Omega$ findes et $n \in \mathbb{Z}_+$, således at $P^n(x, A) > 0$.

Irreducibilitet medfører ligeledes, at Π er en entydig stationær fordeling.

En Markovkæde er Π -irreducibel, hvis det er muligt at nå fra et vilkårligt punkt $x \mod \pi(x) > 0$ til enhver mængde $A \mod \Pi(A) > 0$ i løbet af endelig tid. En Markovkæde kan endvidere være Harris-rekurrent.

Definition A.8 (Harris-rekurrens)

En markovkæde med stationær fordeling Π kaldes Harris-rekurrent, hvis der for alle $A \subseteq \Omega$, med $\Pi(A) > 0$ og alle $x \in \Omega$ gælder, at

 $P(X_n \in A \text{ for uendeligt mange } n | X_0 = x) = 1.$

En vigtig egenskab ved irreducible Markovkæder er Store tals stærke lov for Markovkæder, som beskrevet i følgende sætning.

Sætning A.9 [10, Theorem 1]

Lad (X_0, X_1, \ldots) være en Π -irreducibel Markovkæde, hvor π er en tilhørende stationær tæthed, og lad $h: \Omega \mapsto \mathbb{R}$ være en funktion, således at middelværdien $\theta = \int_{\Omega} h(x)\pi(x)x$ har en endelig værdi. Lad endvidere det empiriske gennemsnit, $\hat{\theta}_n$, være givet ved

$$\widehat{\theta}_n = \frac{1}{n+1} \sum_{i=m}^{m+n} h(X_i),$$

hvor $m, n \in \mathbb{Z}_+$. Da vil der findes en mængde $C \subseteq \Omega$, hvorom det gælder $\Pi(C) = 1$, og at

$$P(\hat{\theta}_n \to \theta | X_m = x) = 1 \quad \text{for } n \to \infty$$

for alle $x \in C$. Hvis Markovkæden er Harris-rekurrent, kan C vælges således, at $C = \Omega$.

For bevis se [11, kapitel 17]

Nu defineres periodisitet og aperiodicitet for markovkæder.

Definition A.10 (Periodiositet og aperiodiositet)

En Π -irreducibel Markovkæde siges at være periodisk, hvis der for n > 1 findes n+1 disjunkte mængder U_0, U_1, \ldots, U_n , således at $\Omega = U_0 \cup U_1 \cup \ldots \cup U_n$ for hvilke, det gælder, at $\Pi(U_0) = 0$ og

$$\begin{array}{rcl} x \in U_1 & \Rightarrow & P(x,U_2) = 1, \\ x \in U_2 & \Rightarrow & P(x,U_3) = 1, \\ & & \vdots \\ x \in U_{n-1} & \Rightarrow & P(x,U_n) = 1, \\ x \in U_n & \Rightarrow & P(x,U_1) = 1. \end{array}$$

Ellers kaldes Markovkæden aperiodisk.

Sætning A.11 [10, Theorem 1] Lad Π være en stationær fordeling for en Π -irreducibel, aperiodisk Markovkæde, da findes en mængde $C \subseteq \Omega$, således at $\Pi(C) = 1$, og der for alle $x \in C$ og alle $A \subseteq \Omega$ gælder, at

$$P(X_n \in A | X_0 = x) \to \Pi(A) \quad \text{for} n \to \infty.$$

Hvis Markovkæden er Harris-rekurrent, kan C vælges således, at $C = \Omega$.

For bevis se [11, kapitel 17]

A.2 DISKRETE PHASETYPEFORDELINGER

I dette afsnit presenteres phasetype fordelinges, som benyttes i afsnit 5.1 og et par egenskaber for dem udledes. Afsnittet bygger på [8, Appendix 2]

Først defineres absorberende tilstande for en markovkæde.

Definition A.12 (Absorberende tilstand) Hvis en markovkæde X_1, X_2, \ldots med udfaldsrum Ω har en tilstand $s_0 \in \Omega$, hvorom gælder at

$$P(X_{n+1} = s_0 | P(X_n) = s_0) = 1$$

kaldes en s_0 en absorberende tilstand.

Definition A.13 (Phasetypefordeling)

Betrakt en markovkæde X_0, X_1, \ldots med udfaldsrummet $\Omega = \{s_1, s_2, \ldots, s_m, s_{m+1}\}$ og begyndelsesfordeling $(\boldsymbol{\alpha}, \alpha_0)$, hvor α_0 angiver sandsynligheden for $X_0 = s_{m+1}$, og indgangene i $\boldsymbol{\alpha}$ betegner sandsynligederne for at X er i hver af de andre tilstande til tiden 0.

Lad tilstanden m + 1 er absorberende, og $\lim_{n\to\infty} P(X_n = s_{m+1}|X_0 = s) = 1$ for alle $s \in \Omega$. For denne markovkæde kan overgangsmatricen skrives som

$$P = \begin{bmatrix} T & \mathbf{T}^{0^{\top}} \\ \mathbf{0} & 1 \end{bmatrix}, \text{ hvor } \begin{bmatrix} p_{11} & \cdots & p_{1_m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{bmatrix} \text{ og } \mathbf{T}^0 = (p_{1m+1}, \dots, p_{mm+1}).$$

Fordelingen for $\min(i|X_i = s_{m+1})$ kalles da en diskret phasetypefordeling med representation $(\boldsymbol{\alpha}, T)$.

Sætning A.14

Fordelingsfunktionen for en diskret phasetypefordeling med representation (α, T) er

 $F(n) = 1 - \boldsymbol{\alpha} T^n \mathbf{1}^\top$, for $n = 1, 2, \dots$

Bevis:

Betrakt markovkæden X hvis absorbtionstid er fordelt med en diskret phasetypefordeling med representation $(\boldsymbol{\alpha}, T)$. Sandsynligheden for $X_n = s_i$ er $(\boldsymbol{\alpha}, \alpha_0) P^n \boldsymbol{e}_i^\top$, hvilket for $i = 1, \ldots, m$ er lig med $\boldsymbol{\alpha} T^n \boldsymbol{e}_i^\top$, således er sandsynligheden for at $X_n \neq s_{m+1}$ givet ved $\boldsymbol{\alpha} T^n \mathbf{1}^\top$. Da denne sandsynlighed netop er sandsynligheden for at X endnu ikke er den absorberende tilstand er sætningen bevist.

Sætning A.15 [8, side 381]

Lad $F_1(\cdot)$ og $F_2(\cdot)$ være uafhængige phasetypefordelte stokastiske variable med representationer (α_1, T_1) og (α_2, T_2) . Mixrurfordelingen $pF_1(\cdot) + (p-1)F_2(\cdot)$ med $p \in [0,1]$ er da en phasetypefordeling med representationen

 $\left((p\boldsymbol{\alpha}_1, (1-p)\boldsymbol{\alpha}_1), \left[\begin{array}{cc} T_1 & 0\\ 0 & T_2 \end{array} \right] \right)$ (A.1)

Bevis:

Lad Z^1 og Z^2 betegne markovkæder hvis absorbtionstid er fordelt med henholdsvis $F_1(\cdot)$ og $F_2(\cdot)$. Lad envidere $\Omega_1 = \{s_1^1, s_2^1, \dots, s_{m+1}^1\}$ og $\Omega_1 = \{s_1^2, s_2^2, \dots, s_{n+1}^2\}$ betegne udfaldsrumne for Z^1 og Z^2 .

Betrakt markovkæden Z dannet ved med $Z_i = Z_i^j$ hvor P(j = 1) = p og P(j = 1) = 1 - p. Hvis g(Z) betegner det mindste *i* for hvilket $Z_i \in \{s_{m+1}^1, s_{n+1}^2\}$. Fordelingen for g(Z) er da netop $pF_1(\cdot) + (p-1)F_2(\cdot)$. Lad envidere markovkæden Z* med udfaldsrummet udfaldsummet $\Omega_1 = \{s_1^1, \ldots, s_m^1, s_1^2, \ldots, s_n^2, s^*\}$ være givet ved $Z_i^* = Z_i$ for $Z_i \notin \{s_{m+1}^1, s_{n+1}^2\}$ og $Z_i^* = Z_i$ for $Z_i \in \{s_{m+1}^1, s_{n+1}^2\}$. Absorbtionstiden for Z^* er således lig med g(Z). Per konstruktion har Z* begyndelsesfordelingen $(p\alpha_1, (1-p)\alpha_1)$ og overgangsmatricen

$$\left[\begin{array}{ccc} T_1 & 0 & {\boldsymbol{T}_1^0}^\top \\ 0 & T_2 & {\boldsymbol{T}_2^0}^\top \\ \boldsymbol{0} & \boldsymbol{0} & 1 \end{array} \right]$$

derfor har absorbtionstiden for Z^* en fasetypefordeling med representation (A.1), og dermed er sætningen bevist.

Sætning A.16 [8, side 381]

Lad $F_1(\cdot)$ og $F_2(\cdot)$ være uafhængige phasetypefordelte stokastiske variable med representationer (α_1, T_1) og (α_2, T_2) . Foldningen $F_1(\cdot) * F_2(\cdot)$ er da en phasetypefordeling med representationen

$$\left((p\boldsymbol{\alpha}_1,(1-p)\boldsymbol{\alpha}_1), \left[\begin{array}{cc}T_1 & \boldsymbol{T}_1^0\boldsymbol{\alpha}_2^\top\\ 0 & T_2\end{array}\right]\right)$$

Bevis:

Lad X og Y betegne markovkæder hvis absorbtionstid er fordelt med henholdsvis $F_1(\cdot)$ og $F_2(\cdot)$. Lad envidere $\Omega_1 = \{s_{X1}, s_{X2}, \ldots, s_{Xm+1}\}$ og $\Omega_1 = \{s_1^2, s_2^2, \ldots, s_{n+1}^2\}$ betegne udfaldsrumne for X og Z^2 . Betrakt markovkæden Z defineret ved

$$Z_i = \begin{cases} Z_i^1 & i < k \\ Z_{i-k}^2 & i \ge k \end{cases},$$

hvor k det mindste i for hvilket $X_i =_{Xm+1}$. Da vil s_{n+1}^2 være en absorberende tilstand for Z og absorbtionstiden for Z vil være summen af absorbstiderne for X og Y, og således vil fordelingen for absorbtionstiden for Z være $F_1(\cdot) * F_2(\cdot)$.

A.3 ENDELIGE MARKOVKÆDER I KONTINUERT TID

I dette afsnit defineres markovkæder i kontinuert tid.

Definition A.17 (Endelig Markovkæde i kontinuert tid)

Lad $X_t \mod t \in \mathbb{R}$ være en familie af stokastiske variable med en endeligt udfaldsrum $\{1, \ldots, m\}$ og lad der eksistere $\lambda_1, \ldots, \lambda_m > 0$ og sandsynlighedsfordelinger $(p_{i,1}, \ldots, p_{i,i-1}, p_{i,i+1}, \ldots, p_{i,1})$ med $i \in \{1, \ldots, m\}$ hvorom gælder:

- For alle $i \in \{1, \ldots, m\}$ med $\lambda_i \neq 0$ vil $\min(s > 0 | X_{t+s} \neq i)$ givet $X_t = i$ være eksponentialfordelt med middelværdi λ_i^{-1} .
- For alle $i \in \{1, \ldots, m\}$ med $\lambda_i \neq 0$ og alle s > 0 gælder at $X_t = i \Rightarrow X_{t+s} = i$
- For all $i \in \{1, \dots, m\}$ med er $P(X_{t+s} = j) | X_t = i) = p_{i,j} \mod s = \min(s > 0 | X_{t+s} \neq i).$

Da kaldes $X = \{X_t\}$ en markovkæde i kontinuert tid.

Den tid Markovkæden forbliver i hver tilstand er således eksponentialfordelt, og når kæden skifter tilstand afhænger kun af den tilstand kæden var i umiddelbart inden den skiftede tilstand. Derfor gælder markov

$$p(X_{t_3}|X_{t_2}, X_{t_1}) = p(X_{t_3}|X_{t_2}) \text{ foralle } t_1 < t_2 < t_3.$$
(A.2)

Ofte beskrives kæden ved hjælp af matricen Q, hvis indgange er givet ved

$$q_{ij} = \begin{cases} -\lambda_i & i = j\\ \lambda_i p_{ij} & i \neq j \end{cases}$$

og ved dens begyndelsesfordeling α .

Programmer i R

Dette bilag indeholder programmer i R til

B.1 SIMULATION AF MAP

Denne funktion gennererer en MAP. Derudover gæmmes den bagvedliggende markovkæde J.

```
MAP <- function(alpha, C,D,tmax){</pre>
#alpha, C og D er parametrene for MAP'en. tmax angiver hvor langt MAP'en kører.
C2<-C
C2[row(C2) == co1(C2)] <- 0
A<-C2+D
j<-sample(1:length(alpha), 1, replace=TRUE, prob = alpha)
#Den tilstand markovprocessen J er i.
t<-0 #Det tidspunkt som vi er nået til.
T<-0 #Vektor med de tidspunkter hvor markovprocessen J skifter tilstand.
J<-j #Vektor med de tilstande markovprocessen J skifter til.
X<-0 #Vektor indeholdende punkterne i punktprocessen X.
while (t<tmax) {
   dt<-rexp(1,sum(A[j,])) #Tiden inden der sker en hendelse.
case<-sample(1:length(A[j,]),1,prob=A[j,]) #Afgør hvilken hendelse der sker.</pre>
t < -(t+dt)
if (case==j){ #Der tilføjes et punkt uden skift i J.
  X < -c(X,t)
else{ #Skift i J med eller uden punkt.
 p<-sample(1:2,1,prob=c(C[j,case],D[j,case]))</pre>
            #Bestemmer om der skal tilføges et punkt eller ej.
  j<-case
  J <-c(J,j)
 T < -c(T,t)
  if (p==2) X<-c(X,t)
  }
}
X < -X[-1]
if (X[length(X)]>tmax) \{X<-X[1:(length(X)-1)]\}
list(tmax=tmax,punkter=X,J=array(c(J,T),dim=c(length(T),2)))
}
```

B.2 Metropolis-Hastings

Denne funktion er delt i tre, hvoraf den første vurderer likelihooden ved brug af sætning 4.4

```
MAPlikelihood<-function(alpha,C,D,X,skala){
n<-dim(D)[1]
D<-Matrix(D)
C<-Matrix(C)
Out<-Matrix(array(alpha,dim=c(1,n)))
for (i in 1:length(X)){
    Out<-Out%*% expm(C*X[i])%*%D*skala
    }
V1<-Matrix(array(1,dim=c(n,1)))
Out<-Out%*%V1
Out[1,1]
}</pre>
```

Den anden funktion tjekker at parametrene ligger indenfor det tillatte område

Den tredie funktion udgør selve MH algoritmen:

```
MAPMH3d<-function(X,n,parm,skala,spr=c(0.002,0.001,0.001,0.1)){#parm=(11,13,b3,p23)
X<-sort(X)
X < -X - c(0, X[-length(X)])
X<-X[-1]
parmf<-parm
OUT<-array(-1,dim=c(5,n))</pre>
plot(c(1,n),c(1,1))
C<-array(c(-parm[1],parm[1],0,
0,-parm[1],0,
0,0,-parm[2]-parm[3])
,dim=c(3,3))
D<-array(c(0,0,0,
parm[1]*(1-parm[4]),0,parm[2],
parm[1]*parm[4],0,parm[3])
,dim=c(3,3))
   alpha<-c(parm[2],parm[2],parm[1]*parm[3])/(parm[2]*2+parm[1]*parm[3])
  alpha0<-alpha%*%D
  alpha0<-alpha0/sum(alpha0)
PI<-MAPlikelyhood(alpha0,C,D,X,skala)*prior(parmf)
for (i in 1:n){
 C<-array(c(-parm[1],parm[1],0,
0,-parm[1],0,
0,0,-parm[2]-parm[3])
,dim=c(3,3))
  D<-array(c(0,0,0,
parm[1]*(1-parm[4]),0,parm[2],
parm[1]*parm[4],0,parm[3])
,dim=c(3,3))
  alpha<-c(parm[2],parm[2],parm[1]*parm[3])/(parm[2]*2+parm[1]*parm[3])
  alpha0<-alpha%*%D
  alpha0<-alpha0/sum(alpha0)
 if (PI>10^300){
points(i,1.2)
skala<-skala*0.5
PI<-MAPlikelyhood(alpha0,C,D,X,skala)
}
```

```
parmf[1] <-rnorm(1,parm[1],spr[1])</pre>
  parmf[2] <-rnorm(1,parm[2],spr[2])</pre>
  parmf[3]<-rnorm(1,parm[3],spr[3])</pre>
  parmf[4]<-rnorm(1,parm[4],spr[4])</pre>
  Cf<-array(c(-parmf[1],parmf[1],0,
0,-parmf[1],0,
0,0,-parmf[2]-parmf[3])
, dim = c(3,3))
  Df<-array(c(0,0,0,
parmf[1]*(1-parmf[4]),0,parmf[2],
parmf[1]*parmf[4],0,parmf[3])
,dim=c(3,3))
  alphaf<-c(parmf[2],parmf[2],parmf[1]*parmf[3])/(parmf[2]*2+parmf[1]*parmf[3])</pre>
  alphaOf<-alphaf%*%D
  alpha0f<-alpha0f/sum(alpha0f)</pre>
PIf<-MAPlikelyhood(alphaOf,Cf,Df,X,skala)*prior(parmf)</pre>
 U<-runif(1,0,1)
 H1<-PIf
 H2<-PI
  OUT[5,i]<-PI
if (!(is.na(H1))&!(is.na(H2))&!(H1==0&H2==0)){
H < -H1/H2
if (!(is.na(H))&(U<=H)){
   points(i,0.8)
           parm<-parmf
      PI<-PIf
   alpha0<-alpha0f}
    OUT[1:4,i]<-c(parm)</pre>
}
else{points(i,1.1,col=2,pch="*")}
 if (floor(i/10)==i/10 ){
points(i,1)
}
}
OUT
}
```

B.3 MIDDELVÆRDI FOR Y_{cd_+}

Denne funktion beregner den momentgenererende funktion for $Y_{\mathrm{cd}_+}.$

```
MGF<-function(s,alpha0,C,D,R){
    D<-Matrix(D)
    C<-Matrix(C)
    alpha<-Matrix(array(alpha0,dim=c(1,dim(D)[1])))
    I<-Matrix(diag(dim(D)[1]))
    A<-C-s*I
    B<-I-solve(A)%*%(expm(A*R)-I)%*%D
    V1<-Matrix(array(1,dim=c(dim(D)[1],1)))
    g<-alpha%*%solve(B)%*%solve(C)%*%expm(C*R)%*%D%*%V1
    out<-g*exp(-s*R)
    out[1,1]
}</pre>
```

Denne funktion be
regner ved hjælp af den forrige funktion middelværdien af
 $Y_{\rm cd_+}.$

```
YcdPlusTeori<-function(alpha0,C,D,R){
    s<-0
    A<-function(t){MGF(t,alpha0,C,D,R)}
    numericDeriv(quote(A(s)), "s")
}</pre>
```

B.4 MIDDELVÆRDI FOR Y_{cd_+}

Denne funktion approksimerer middelværdien for Y_{cd_+} .

```
YchPlusTeori<-function(lambda,R){</pre>
```

```
funk<-function(s,Y){# Denne delfunktion tilnærmer en difrantiabel funktion ud
                      # en vektor med dens værdier i forskellige punkter.
   n<-length(Y)</pre>
   i < -s*(n-1)/R+1
   if (s==R) \{ Out < -Y[n] \}
    else{
l<-Y[floor(i)]</pre>
h<-Y[floor(i)+1]</pre>
Out<-l+(i-floor(i))*(h-l)</pre>
}
   Out
}
h0<-function(t){
1
H<-function(t,X){</pre>
1+integrate(Vectorize(function(s){lambda*exp(-lambda*s)*funk(s,X)}),0,R-t)$value
}
torsk<-Vectorize(h0)(0:1000*R/1000)
repeat{
laks<-Vectorize(function(t){H(t,torsk)})(0:1000*R/1000)</pre>
torsk<-Vectorize(function(t){H(t,laks)})(0:1000*R/1000)</pre>
if (max(c(laks-torsk,torsk-laks))<0.0000001){break}</pre>
}
torsk[1]-1
}
```

LITTERATUR

- [1] Beskrivelse af Fleetnet ved Mannheim Universitets hjemmeside. "http://pi4.informatik.unimannheim.de/pi4.data/content/projects/fleetnet/".
- [2] Søren Asmussen. Applied Probability and Queues. Springer, 2 edition, 2003.
- [3] Holger Füßler, Marc Torrent-Moreno, Roland Krüger, Matthias Transier, and Hannes Hartensteinog Wolfgang Effelsberg. Studying Vehicle Movements on Highways and their Impact on Ad-Hoc Connectivity, 2005. "urlhttp://dsn.tm.uni-karlsruhe.de/publications_959.php".
- [4] Holger Füßler, Marc Torrent-Moreno, Roland Krüger, Matthias Transier, Hannes Hartenstein, and Wolfgang Effelsberg. Statistical Analysis of the FleetNet Highway Movement Patterns, 2005. "http: //madoc.bib.uni-mannheim.de/madoc/volltexte/2005/1091/".
- [5] Martin Bøgsted Hansen, Jakob Gulddahl Rasmussen, and Hans Peter Schwefel. Connectivity analysis of one-dimensional ad-hoc networks. 2008.
- [6] Peter M. Lee. Bayesian Statistics, an introduction. Hodder Arnold, third edition, 2004.
- [7] Jesper Møller and Rasmus Plenge Waagepetersen. Statistical Inference and Simulation for Spatial Point Processes. Chapman & Hall/CRC, first edition, 2004.
- [8] Marcel F. Neuts. Algorithmic probability: A collection of problems. Chapman & Hall, 1995.
- [9] J. R. Norris. Markov Chains. Cambridge University Press, 1997.
- [10] Kasper K. Berthelsen og Jesper Møller. A short diversion into the theory of Markov chains, with a view to Markov chain Monte Carlo methods.
- [11] S.P. Meyn & R.L. Tweedie. Markov chains and stochastic stability. Springer-Verlag, 1993.