

Kandidatnummer: 250712



Universitetet
i Stavanger

«Lokal trendestimering av Covid-19-smitte ved
bayesiansk analyse»

LMABAC

Bacheloroppgave 2021

Det teknisk-naturvitenskapelige fakultet

Institutt for matematikk og fysikk

Adrian Serigstad

Sammendrag

Dette studiet tar for seg trendmodellen for Covid-19-smitte publisert i Stavanger Aftenblad. Hensikten er å videreformidle trendmodellens bayesianske bakteppe og fremheve bayesiansk analyse som metode for estimering under usikkerhet.

For å kunne beskrive smittetrender i Rogaland ble det gjennomført en bayesiansk statistisk analyse. Kombinasjonen av tidligere kjent informasjon og smittedata fra de to foregående ukene førte med seg en fordeling av interesse. Videre ble denne fordelingen sammen med smittedata fra siste uke brukt til å lage en ny fordeling. Dette resulterte i trendgrenser som kunne beskrive om smittetrenden i områdene var «stabil», «økende» eller «synkende» gitt stabil tilstand.

I analysen ble det tatt utgangspunkt i fiktive smittetall for Rogaland. Tallet på smittede de to foregående ukene og siste uke ble henholdsvis satt til 700 og 450. Dette resulterte i en økende smittetrend. Deretter ble det gjort tilsvarende med Time kommune i mer konsise trekk. Denne gangen med reelle smittedata fra de tre første ukene i april 2021. Med 19 smittede de to første ukene og seks siste uke, indikerte modellen stabil trend.

Vi kan konkludere med at bayesiansk metode er velegnet til lokal trendestimering av Covid-19-smitte. Denne metoden gir en fin balanse mellom skjønn og logisk matematikk i en situasjon der det er fornuftig. I en klassisk analyse ville vi blant annet risikert meningsløse resultater i kommuner uten smitte.

Innholdsliste

1. INTRODUKSJON	4
2. BAYESIANSK STATISTIKK	4
3. TRENDMODELLEN	5
4. BAYESIANSK ANALYSE	7
4.1 POISSONFORDELING	7
4.2 PRIORFORDELING	8
4.3 RIMELIGHETSFUNKSJON	10
4.4 POSTERIORFORDELING	11
4.5 BAYESESTIMAT	13
4.6. PREDIKTIV FORDELING	13
5. SMITTETREND I TIME KOMMUNE	16
6. KONKLUSJON	18
7. LITTERATURLISTE	19

1. Introduksjon

Statistikk er en gren innen matematikken jeg synes er ekstra spennende. Det å beregne sannsynligheter og analysere statistiske modeller er noe vi alle gjør i større eller mindre grad i det daglige. Statistikk er noe vi omfavnes av hele tiden uten å tenke noe særlig over det. Politiske meningsmålinger, oddsen for Man United-seier og gårsdagens koronasmittetall inngår alle i statistikkfagfeltet. Slike hverdagslige temaer som baserer seg på matematiske konsepter er noe som fatter interesse.

Da veileder introduserte meg for trendmodellen for Covid-19-smitte i Aftenbladet, ble jeg raskt nysgjerrig. Noe så dagsaktuelt kombinert med en ukjent del av statistikkverden måtte studeres nærmere. Derfor falt valget på å utforske den bayesianske metoden bak denne trendmodellen og videreformidle dette på en ryddig og enkel måte.

Studiets hovedvekt ligger på den bayesianske analysen bak trendmodellen for Covid-19-smitte. I tillegg vil studiet gå inn på bayesiansk statistikk som helhet og fremheve denne tankegangen under estimering ved usikkerhet. Klassisk statistisk analyse blir nevnt til sammenligning, men vil ikke være et fokusert område i dette studiet.

2. Bayesiansk statistikk

Statistikk handler om å samle inn, organisere, analysere, tolke og presentere ulike data til ulike formål. Det kan være til eksperimenter, undersøkelser eller annen forskning. Vi skiller gjerne mellom klassisk og bayesiansk statistikk. I klassisk statistikk er vi ofte interessert i estimering av faste ukjente parametere (Løvås, 2018, s. 235). Dette ved å utføre gjentatte delforsøk slik vi får et datasett $x = (x_1, \dots, x_n)$ (Løvås, 2018, s. 235). Målet er i bunn og grunn å finne det beste estimatet til den sanne parameteren og bruke konfidensintervall til å anslå området som parameteren trolig ligger i (Løvås, 2018, s. 234).

I en bayesiansk statistisk analyse har vi ofte kjent informasjon før datainnsamlingen forekommer basert på tidligere fordeling på parameteren (Klugman, Panjer & Willmot, 2004, s. 361). Denne allerede kjente informasjonen kalles prior og uttrykkes ved en priorfordeling (Klugman et.al., 2004, s. 361). Det å ha prior informasjon står i motsetning til en klassisk statistisk analyse der man kun forholder seg til de konkrete dataene som er samlet inn (van de Schoot et.al., 2013). Ved å ha denne kjente informasjonen i tillegg til data, blir det enklere å estimere parameteren da antall mulige verdier av parameteren begrenses (van de Schoot et.al., 2013). Kombinasjonen av en priorfordeling og innsamlet data gir en posteriorfordeling ved bruk av Bayes formel (Klugman et.al., 2004, s. 362). Posteriorfordelingen er det vi ønsker å sitte igjen med til slutt da dette gir den beste informasjonen om parameteren. (Klugman et.al., 2004, s. 362).

En annen viktig forskjell ifølge tidligere professor ved UIO, Bent Natvig, er hvilken grad skjønn spiller en rolle i beslutningen. I bayesiansk metode tar man utgangspunkt i de konkrete dataene og lar de subjektive tankene påvirke. Natvig poengterer at en lege for eksempel ikke kun kan legge generelle data til grunn i vurderingen av en hjerteoperasjon. Hvis pasienten har dårlig fysisk helse med underliggende sykdommer av betydning, må legen ta en avgjørelse basert på skjønn i tillegg til konkrete data. Det er ikke nok å vite estimatet for hvor stor sannsynlighet det er for at operasjonen lykkes basert på foregående hendelser. Legen må i tillegg gjøre en subjektiv vurdering av pasientens tilstand, understreker Natvig. Denne bayesianske måten å angripe et problem på er laget for å kunne ta hensyn til beslutterens subjektivitet samtidig som den er i tråd med reglene innenfor sannsynlighetsregning. Dette gjør bayesiansk metode velegnet til å ta beslutninger under usikkerhet (Natvig, 1997).

3. Trendmodellen

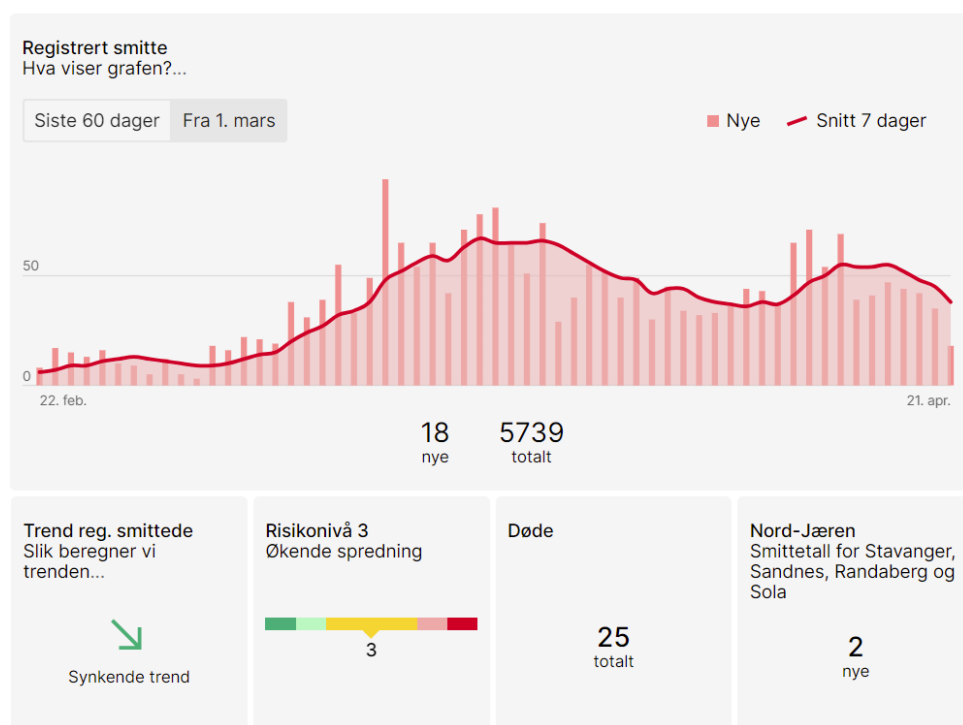
Trendmodellen publisert i Aftenbladet kan vi si har et bayesiansk fundament. Dette skal vi ta for oss i mer detalj i neste kapittel. Modellen viser antall registrerte koronasmittede i Rogaland de siste ukene og siden pandemiens utbrudd. Den tar også for seg kommunene hver for seg. Antall registrerte smittede illustreres med søyler og en graf

Kandidatnummer: 250712

med hensyn på disse. Grafen formes av gjennomsnittet av nye koronasmittede de siste syv dagene. Det vil si at grafen ikke legger seg på hver enkel søyletopp, men får en jevnere kurve basert på siste ukens snitt. Nye smittede er definert som summen av dagens og gårsdagens registrerte smittetilfeller (Aftenbladet, 2021).

I tillegg til å vise registrerte tilfeller gjennom grafer og tall, viser modellen om smittetrenden er stigende, stabil eller synkende. Dette illustreres med ei pil som henholdsvis peker skrått oppover, bortover eller skrått nedover. Basert på siste ukens smittetall sammenlignet med de to foregående ukens tall, vil modellen kunne si noe om smittetrenden i fylket og kommunene. Er smitten signifikant høyere den siste uken sammenlignet med de to foregående ukene gitt stabil tilstand, vil trenden være stigende. Hvis smittetallet er signifikant lavere den siste uken sammenlignet med de to foregående ukene gitt stabil tilstand, vil trenden være synkende. Dersom ingen av disse tilfellene forekommer, vil modellen indikere stabil trend (Aftenbladet, 2021).

Rogaland



Figur 1: Oversikt over koronasmittede i Rogaland. Utklipp fra Aftenbladet, 2021.

4. Bayesiansk analyse

Bak modellens trendpiler ligger det et helt stykke matematisk arbeid. Vi skal nå ta for oss i grunnleggende trekk hvordan en bayesiansk analyse gjennomføres med blikket rettet mot trendmodellen for registrerte koronasmittede. Dette vil underveis eksemplifiseres med relevante verdier.

4.1 Poissonfordeling

Trendmodellen er bygget basert på bayesiansk tanke sett der antall nye smittede hver uke blir presentert uavhengig gjennom en poissonfordeling (Aftenbladet, 2021). Årsaken til bruk av denne fordelingen er at vi ønsker å studere antall forekomster av en hendelse innenfor et visst tidsintervall (Løvås, 2018, s. 186). Vi er nemlig interessert i antall tilfeller av nye koronasmittede i fylket i løpet av de tre siste ukene (Aftenbladet, 2021). I tillegg antar vi at kravene en poissonfordeling har oppfylles (Løvås, 2018, s. 186).

Et krav er at forventet antall forekomster av hendelsen skal være konstant lik λ per tidsenhet (Løvås, 2018, s. 186). Det oppfylles da vi antar at antall forventede tilfeller av koronasmitte vil være konstant per uke. I tillegg kan ikke to forekomster av hendelsen være fullstendig sammenfallende på tidsaksen (Løvås, 2018, s. 186). I og med to personer ikke blir koronasmittet nøyaktig samtidig, blir dette kravet oppfylt også. Et siste krav er at forekomstene av hendelsen er uavhengige av hverandre (Løvås, 2018, s. 186). Dette strider i og for seg med gjeldende praksis fordi en koronasmittet vil kunne smitte andre personer. Likevel antar vi at hendelsene er uavhengige som en forenkling til modellen.

I og med en poissonfordeling er passende i denne sammenhengen, vil λ være den ukjente parameteren som skal estimeres (Løvås, 2018, s. 186). Da er parameteren bestemt og første steg i den bayesianske analysen er gjennomført.

4.2 Priorfordeling

Neste steg i analysen vil være å velge en prior for parameteren $p(\lambda)$ (van de Schoot et.al., 2013). Denne prioren skal representere den subjektive kunnskapen med hensyn på λ før data er samlet inn (van de Schoot et.al., 2013). I vårt tilfelle vil dette illustreres gjennom en gammafordeling betegnet som gamma (α, β). En slik fordeling utfolder seg avhengig av en formparameter α og en skalaparameter β (Natvig, 1997). Formparameteren bestemmer formen på fordelingen (Wikipedia, 2021). Høy formparameter, gjør at gammafordelingen tilnærmer seg en normalfordeling. (Wikipedia, 2021).

Formparameter lik 1 vil gi en eksponentialfordeling (Walpole, Myers, Myers & Ye, 2012, s. 195). Skalaparameteren angir spredningen til gammafordelingen (Wikipedia, 2021). Jo større skalaparameteren er, jo større spredning vil fordelingen få (Wikipedia, 2021).

Prioren vår vil være gamma ($1, \lambda_0$) – fordelt der $\lambda_0 = 7 \cdot \left(\frac{n}{10\,000}\right)$ er forventningen. Det vil si vi antar at 0,01% av befolkningen (n) tester positivt en gitt dag. Hvis vi ganger det med 7, får vi forventet antall positive tester en gitt uke. Siden formparameteren er lik 1 vil vi få en eksponentialfordeling (Walpole, et.al., 2012, s. 195). Eksponentialfordeling er med andre ord et spesialtilfelle av gammafordelingen (Walpole, et.al., 2012, s. 194).

Den generelle formelen for tettheten til en gamma (α, β) – fordelt prior er gitt ved (Klugman, et.al., 2004, s. 369):

$$p(\lambda) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda/\beta}$$

Siden skalaparameteren β er avhengig av befolkningstallet n , må vi først bestemme en verdi for n . Ut fra SSB sine tall, har Rogaland ca. 480 000 innbyggere (Statistisk sentralbyrå, 2021).

Skalaparameteren kan vi da regne ut:

$$\beta = \lambda_0 = 7 \cdot \left(\frac{n}{10\,000}\right) = 7 \cdot \left(\frac{480\,000}{10\,000}\right) = 336$$

Da har vi verdiene $\alpha = 1$ og $\beta = 336$ og kan sette dette inn i formelen for tettheten til gamma (α, β) - fordelt prior:

$$p(\lambda) = \frac{1}{336^1 \Gamma(1)} \lambda^{1-1} e^{-\lambda/336} = \frac{1}{336} e^{-\lambda/336}$$

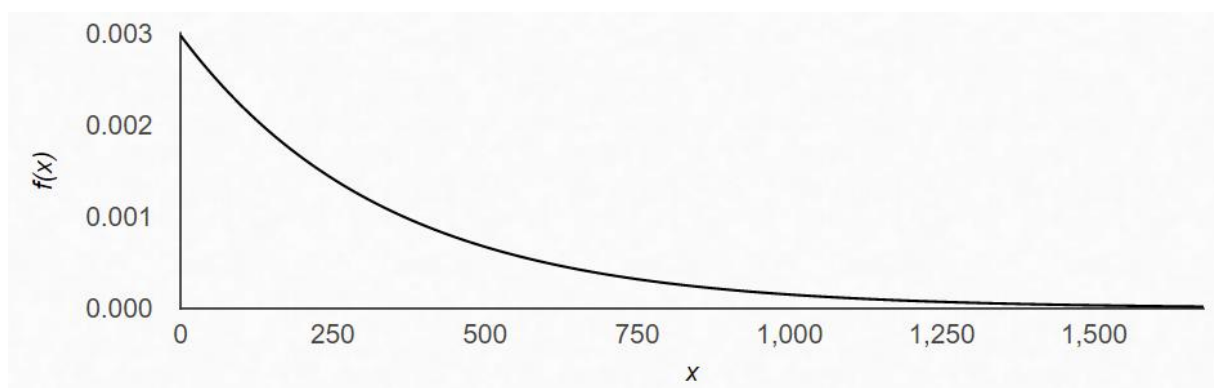
Gammafunksjonen forsvinner fordi:

$$\alpha = 1 \Rightarrow \Gamma(1) = (1 - 1)! = 0! = 1$$

Konstanten foran eksponentialfunksjonen vil ikke spille en rolle i videre beregninger. Årsaken er at vi er uinteressert i faktorer som ikke inneholder parameteren ettersom det er verdien av λ vi ønsker å finne (Walpole, et.al., 2012, s. 712). Dette kalles proporsjonalitet (\propto) (Walpole, et.al., 2012, s. 712). Da sitter vi igjen med følgende uttrykk for prioren som tas med videre i analysen:

$$p(\lambda) = \frac{1}{336} e^{-\lambda/336} \propto e^{-\lambda/336}$$

Priorfordelingen vil med $\alpha = 1$ og $\beta = 336$ se slik ut:



Prior: $\lambda \sim \text{gamma}(1, 336)$

Figur 2: Priorfordeling

Forventning: $E(\lambda) = \alpha \cdot \beta = 1 \cdot 336 = 336$

Dette vil si vi regner med en parameterverdi på rundt 336. Med andre ord forventer vi at det en gitt uke vil bli registrert rundt 336 nye koronasmittede. Dette kun basert på vår prior informasjon. Ut fra priorfordelingen kan vi i tillegg regne ut ulike kvantiler for å sette fordelingen i perspektiv:

```
> qexp(0.95, rate=1/336)
[1] 1006.566
> qexp(0.50, rate=1/336)
[1] 232.8975
> qexp(0.05, rate=1/336)
[1] 17.23455
```

Figur 3: Kvantiler i R

I 95 % av ukene en gitt tidsperiode regner vi med et smittetall på under 1007 per uke.

I 50 % av ukene en gitt tidsperiode regner vi med et smittetall på under 233 per uke.

I 5 % av ukene en gitt tidsperiode regner vi med et smittetall på under 18 per uke.

4.2 Rimelighetsfunksjon

Etter å ha valgt en priorfordeling som representerer priorinformasjonen til λ , ønsker vi å kombinere dette med observerte data (Natvig, 1997). De observerte dataene $x = x_1, \dots, x_n$ kan presenteres gjennom en rimelighetsfunksjon $L(\lambda)$ (Natvig, 1997).

Rimelighetsfunksjonen er en funksjon av den ukjente parameteren som gjenspeiler de mest sannsynlige verdiene for parameteren gitt dataene (van de Schoot et.al., 2013).

Den er som i klassisk statistikk gitt ved produktet av sannsynlighetsmassefunksjoner med hensyn på parameteren vår (Klugman, et.al., 2004, s. 361):

$$L(\lambda) = p(x | \lambda) = f(x_1 | \lambda)f(x_2 | \lambda) \cdots f(x_n | \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

Som i prioren vil enkelte faktorer falle bort. Vi kvitter oss med konstantene og står igjen med faktorer som inneholder λ :

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \propto \lambda^{x_1+x_2+\dots+x_n} e^{-n\lambda}$$

Estimeringen av λ gjøres på data for de to foregående ukene (Aftenbladet, 2021). Det vil si mellom 8 og 21 dager siden. Vi lar X være summen av alle registrerte koronatilfeller i dette tidsrommet ved n observasjoner:

$$X = x_1 + x_2 + \dots + x_n$$

La oss si det har blitt registrert 400 smittede første uken og 300 smittede andre uken. Da får vi følgende uttrykk for rimelighetsfunksjonen som tas med videre i analysen:

$$L(\lambda) \propto \lambda^X e^{-n\lambda} = \lambda^{400+300} e^{-2\lambda} = \lambda^{700} e^{-2\lambda}$$

4.3 Posteriorfordeling

Når vi nå både har valgt en passende prior og gjort rede for dataene gjennom en rimelighetsfunksjon, kan vi utlede en posteriorfordeling $p(\lambda | x)$ (van de Schoot et.al., 2013). Dette blir vår oppdaterte kunnskap basert på prior informasjon $p(\lambda)$ kombinert med dataene $L(\lambda)$ (van de Schoot et.al., 2013). Denne kombinasjonen kan presenteres gjennom en simultanfordeling $p(x, \lambda)$ (Natvig, 1997). Dette er en fordeling som tar for seg mer enn én variabel i motsetning til en marginalfordeling (Natvig, 1997).

Simultanfordelingen finner vi ved å multiplisere rimelighetsfunksjonen med prioren (Natvig, 1997). Nå kan vi finne uttrykket for posterioren ved Bayes teorem:

$$p(\lambda | x) = \frac{p(x | \lambda) p(\lambda)}{p(x)} = \frac{L(\lambda) p(\lambda)}{p(x)} = \frac{p(x, \lambda)}{p(x)}$$

$p(x)$ er kun en normaliseringskonstant vi kan se bort fra i denne sammenhengen da den ikke inneholder parameteren (Walpole, et.al., 2012, s. 712). Derfor vil posterioren være proporsjonal til simultanfordelingen:

$$p(\lambda | x) \propto p(x, \lambda) = L(\lambda) p(\lambda)$$

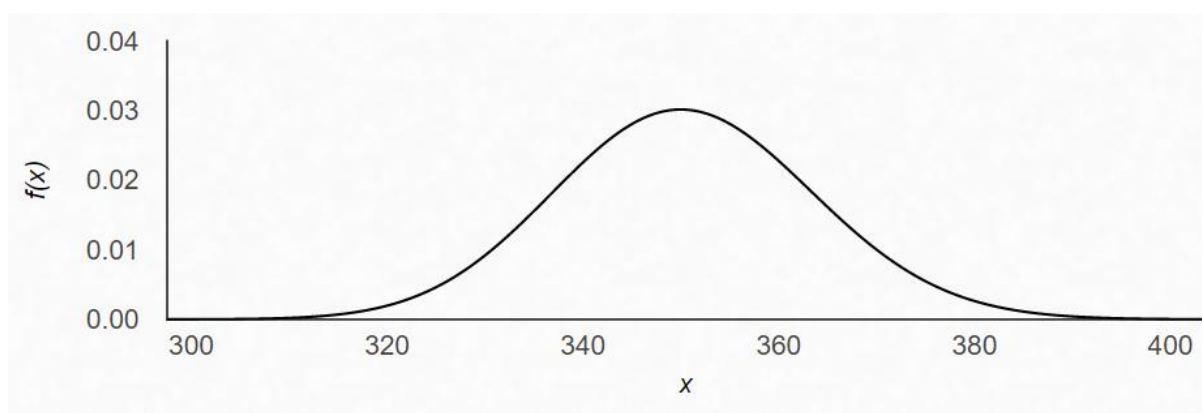
I vårt tilfelle vil vi få følgende uttrykk for posterioren etter å ha gjort om på ønsket form $\lambda^{\alpha-1} e^{-\lambda/\beta}$ (Walpole, et.al., 2012, s. 713):

$$p(\lambda | X) \propto \lambda^X e^{-n\lambda} e^{-\lambda/\lambda_0} = \lambda^X e^{-\lambda/\frac{1}{\frac{1}{\lambda_0}+n}}$$

Nå kan vi nemlig utlede posteriorfordelingen ved å definere formparameteren α_1 og skalaparameteren β_1 . Vi ser fra uttrykket over at $\alpha_1 = X + 1$ og $\beta_1 = (n + \lambda_0^{-1})^{-1}$. Siden vi har med to ukentlige observasjoner å gjøre, får vi $n = 2$. Da får vi at posterioren er gamma $(X + 1, (2 + \lambda_0^{-1})^{-1})$ -fordelt.

Posteriorfordelingen kan dermed formuleres slik: $\lambda | X \sim \text{gamma}(1 + X, (2 + \lambda_0^{-1})^{-1})$

Hvis vi nå setter inn $X = 700$ og $\lambda_0 = 336$, får vi formparameter $\alpha_1 = 701$ og skalaparameter $\beta_1 = 0.5$. Dette gir følgende posteriorfordeling:



Posterior: $\lambda | X \sim \text{gamma}(701, 0.5)$

Figur 4: Posteriorfordeling

4.4 Bayesestimat

Fra posteriorfordelingen kan vi nå regne ut et punktestimat (Walpole, et.al., 2012, s. 717). Dette punktestimatet kalles Bayesestimatet $\hat{\lambda}_{Bayes}$ og er forventet verdi av parameteren ut fra posterior (prior og data) (Walpole, et.al., 2012, s. 717). Selv om posteriorfordelingen gir den mest utfyllende informasjonen, kan det være greit å konkretisere med en estimert verdi. Siden vi har med en gamma posterior å gjøre er Bayesestimatet gitt ved produktet av form- og skalaparameteren (Walpole, et.al., 2012, s. 718):

$$\hat{\lambda}_{Bayes} = E(\lambda | X) = \alpha_1 \cdot \beta_1 = 701 \cdot 0.5 = 350.5$$

$$E(\lambda) = 336$$

Vi ser her at forventet verdi av λ er noe høyere ut fra posterior sammenlignet med prior. De observerte dataene har altså bidratt til en økning av parameterens forventningsverdi. I praksis vil dette si at det antatte smittetallet en gitt uke har økt fra 336 til ca. 351 etter å ha inkludert observerte smittedata.

4.5 Prediktiv fordeling

Frem til nå har vi sett på data fra de to forgående ukene gitt ved X . Nå skal vi legge til et nytt datasett Y . Dette representerer antall påvist smittede siste uke. La oss si $Y = 450$. Det vi nemlig er interessert i er å sammenligne dataene fra den siste uken med de to foregående ukene for å se smittetrenden (Aftenbladet, 2021). Vi ønsker med andre ord å finne en prediktiv fordeling $p(Y | X)$ (Klugman et.al., 2004, s. 362).

λ har vi til nå sett på som en fast ukjent størrelse. Videre skal vi se på den som en stokastisk (tilfeldig) variabel (Klugman et.al., 2004, s. 362). For å finne den prediktive fordelingen må vi først se på simultanfordelingen for λ og $Y | X$ gitt ved $p(\lambda, Y | X)$ (Klugman et.al., 2004, s. 362).

Simultanfordelingen finner vi ved å ta produktet av massefunksjonen til Y gitt ved $p(Y | \lambda)$ og posterioren $p(\lambda | X)$ (Klugman et.al., 2004, s. 362):

$$p(\lambda, Y | X) = p(Y | \lambda) p(\lambda | X)$$

Siden vi kun er interessert i Y -delen av simultanfordelingen, integrerer vi bort λ og får følgende marginalfordeling (Klugman et.al., 2004, s. 369):

$$\begin{aligned} p(Y | X) &= \int_0^{\infty} p(Y | \lambda) p(\lambda | X) d\lambda \\ &= \int_0^{\infty} \frac{\lambda^Y}{Y!} e^{-\lambda} \frac{1}{\beta_1^{\alpha_1} \Gamma(\alpha_1)} \lambda^{\alpha_1-1} e^{-\lambda/\beta_1} d\lambda \\ &= \frac{1}{\beta_1^{\alpha_1} Y! \Gamma(\alpha_1)} \int_0^{\infty} \lambda^{\alpha_1+Y-1} e^{-\lambda(1+\beta_1^{-1})} d\lambda \\ &= \frac{1}{\beta_1^{\alpha_1} Y! \Gamma(\alpha_1)} \frac{\Gamma(\alpha_1 + Y)}{(1 + \beta_1^{-1})^{\alpha_1+Y}} \\ &= \frac{(\alpha_1 + Y - 1)!}{(\alpha_1 - 1)! Y!} \frac{\beta_1^{-\alpha_1}}{(1 + \beta_1^{-1})^{\alpha_1+Y}} \\ &= \binom{Y + \alpha_1 - 1}{Y} \left(\frac{\beta_1^{-1}}{1 + \beta_1^{-1}} \right)^{\alpha_1} \left(\frac{1}{1 + \beta_1^{-1}} \right)^Y \\ &= \binom{Y + \alpha_1 - 1}{Y} \left(\frac{1}{1 + \beta_1} \right)^{\alpha_1} \left(\frac{\beta_1}{1 + \beta_1} \right)^Y \quad \blacksquare \end{aligned}$$

Dette uttrykket ser vi kan kobles opp mot massefunksjonen til en negativ binomisk fordeling $nbin(r, p)$ med $r = \alpha_1$ og $p = (1 + \beta_1)^{-1}$ (Klugman et.al., 2004, s. 76):

$$\binom{k + r - 1}{k} \left(\frac{1}{1 + \theta} \right)^r \left(\frac{\theta}{1 + \theta} \right)^k = \binom{k + r - 1}{k} p^r (1 - p)^k$$

Kandidatnummer: 250712

Vi får altså en prediktiv fordeling som er $nbin(\alpha_1, (1 + \beta_1)^{-1})$ – fordelt.

Det å blande en Poissonfordeling med en gammafordeling ser vi gir en negativ binomisk fordeling. Derfor kan negativ binomisk fordeling ses på som en gamma-Poisson-blanding (Klugman et.al., 2004, s. 77).

Hvis vi nå spinner videre på verdiene brukt tidligere, får vi:

$$\alpha_1 = X + 1 = 700 + 1 = 701$$

$$(1 + \beta_1)^{-1} = (1 + (2 + \lambda_0^{-1})^{-1})^{-1} = (1 + 0.5)^{-1} = \frac{2}{3}$$

$$\Rightarrow Y | X \sim nbin(701, 2/3)$$

Dette gir følgende trendgrenser:

```
> qnbinom(p=0.95, size=701, prob=2/3)
[1] 389
> qnbinom(p=0.05, size=701, prob=2/3)
[1] 313
```

Figur 5: Trendgrenser for Rogaland i R

Hvis antall registrerte smittede siste uke Y er større enn 95-prosentilen til $Y | X$, vil smittetrenden være «økende» (Aftenbladet, 2021). Siden modellen baserer seg på smittede i en stabil tilstand, regner vi det som lite sannsynlig at Y overstiger dette. Tilsvarende er det hvis antall registrerte smittede siste uke Y er mindre enn 5-prosentilen til $Y | X$ (Aftenbladet, 2021). Da vil smittetrenden regnes som «synkende». Igjen er sannsynligheten for dette utfallet liten gitt stabil tilstand. Hvis ingen av disse tilfellene forekommer, vil smittetrenden være «stabil» (Aftenbladet, 2021).

Ut fra trendgrensene ser vi at hvis antall registrerte smittede i fylket overstiger 389 siste uke, vil smittetrenden være økende. Hvis antall registrerte smittede siste uke er mindre enn 313, vil smittetrenden være synkende. Med et smittetall mellom disse grensene, vil smittetrenden være stabil. Disse trendgrenseneverdiene er altså gitt at totalt 700 er registrert smittet de to foregående ukene (fra X). Siden vi tidligere definerte $Y = 450$, ville altså modellen indikert økende smittetrend.

5. Smittetrend i Time kommune

I dette kapittelet skal vi se på smittetrenden i hjemkommunen min fra 1. april til og med 21. april 2021. Vi skal se om det var økende, stabil eller synkende smittetrend blant innbyggerne i Time i denne perioden ved hjelp av bayesiansk analyse. Siden selve analysen allerede er gjennomført, endrer jeg bare de nødvendige verdiene.

Priorfordeling:

gamma (α, β) - fordelt med $\alpha = 1$ og $\beta = \lambda_0 = 7 \cdot \left(\frac{n}{10\,000}\right)$

Time kommune har et folketall på ca. 19 100 ved inngangen av 2021 (Statistisk sentralbyrå, 2021).

$$\Rightarrow \beta = 7 \cdot \left(\frac{19\,100}{10\,000}\right) = 13.37$$

$$p(\lambda) \propto e^{-\lambda/\beta} = e^{-\lambda/13.37}$$

Rimelighetsfunksjon:

Poissonfordelt med $X = x_1 + x_2 + \dots + x_n$ og $n = 2$

Smittetallene i Time kommune de to foregående ukene:

Dato	01.04	02.04	03.04	04.04	05.04	06.04	07.04
Smittetall	1	0	0	3	1	1	1

Dato	08.04	09.04	10.04	11.04	12.04	13.04	14.04
Smittetall	0	5	3	0	2	1	1

Tabell 1: Smittedata i Time kommune 01.04.21 – 14.04.21. Hentet fra Aftenbladet, 2021.

$$\Rightarrow X = 1 + 3 + 1 + 1 + 1 + 5 + 3 + 2 + 1 + 1 = 19$$

$$L(\lambda) \propto \lambda^X e^{-n\lambda} = \lambda^{19} e^{-2\lambda}$$

Posteriorfordeling:

gamma (α_1, β_1) – fordelt med $\alpha_1 = X + 1$ og $\beta_1 = (2 + \lambda_0^{-1})^{-1}$

$$\alpha_1 = 19 + 1 = 20$$

$$\beta_1 = (2 + 13.37^{-1})^{-1} \approx 0.482$$

$$p(\lambda | X) \propto \lambda^X e^{-\lambda / \frac{1}{\lambda_0 + 2}} = \lambda^{19} e^{-\lambda / 0.482}$$

Prediktiv fordeling:

nbin(r, p) – fordelt med $r = \alpha_1$ og $p = (1 + \beta_1)^{-1}$

$$\alpha_1 = 20$$

$$(1 + \beta_1)^{-1} = (1 + 0.482)^{-1} \approx 0.675$$

$$\Rightarrow Y | X \sim \text{nbin}(20, 0.675)$$

```
> qnbinom(p=0.95,size=20,prob=0.675)
[1] 16
> qnbinom(p=0.05,size=20,prob=0.675)
[1] 4
```

Figur 6: Trendgrenser for Time kommune i R.

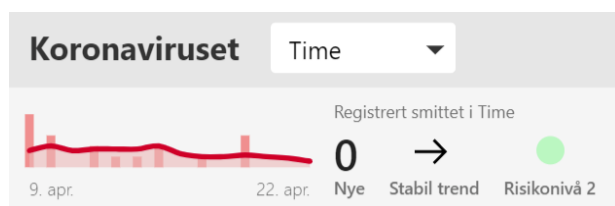
Smittetallene i Time kommune siste uke:

Dato	15.04	16.04	17.04	18.04	19.04	20.04	21.04
Smittetall	2	0	1	0	3	0	0

Tabell 2: Smittedata i Time kommune 15.04.21 – 21.04.21. Hentet fra Aftenbladet, 2021.

$$\Rightarrow Y = 2 + 1 + 3 = 6$$

Med seks observerte smittetilfeller siste uke, betyr det en nedgang fra de 19 smittede de to foregående ukene. Ut fra trendgrenseverdiene i R, ser vi at nedre grense for stabil tilstand er på fire smittede siste uke gitt 19 tilfeller de to foregående. Det vil si selv med en klar nedgang i smitte blant Time-beboerne den siste uken, indikerer modellen at smittetrenden er stabil i perioden 01.04.21 – 22.04.21.



Figur 7: Smittetrend Time kommune 01.04.21 – 22.04.21. Utklipp fra Aftenbladet, 2021.

Hadde vi tatt utgangspunkt i en liten kommune uten smitte, ville vi fått problemer uten en bayesiansk modell. En klassisk sannsynlighetsmaksimeringsestimering ville gitt en estimert parameterverdi lik 0 siden dette kun baseres på data. Samtidig er det ofte vanskelig å oppnå meningsfulle resultater når utvalgsstørrelsen er liten under en slik estimering (van de Schoot et.al., 2013).

6. Konklusjon

Bayesiansk metode ser vi er velfungerende i forbindelse med trendestimering under usikkerhet. Det har vært omdiskutert blant statistikere hvorvidt denne metoden bør ta plass i statistiske analyser. Bruken av skjønn som inngår i denne analyseformen er noe enkelte konservative matematikere er skeptiske til. Likevel har den bayesianske metoden utviklet seg til å bli en viktig del av den moderne matematikken.

Trendmodellen for Covid-19-smitte viser nytten av å ta i bruk den bayesianske tankegangen. Selv i små kommuner uten smitte ville bayesiansk metode gitt meningsfulle resultater. Dette hadde nødvendigvis ikke vært tilfelle i en klassisk analyse.

7. Litteraturliste:

- Klugman, S. A., Panjer & H. H, Willmot, G. E. (2004). *Loss Models From Data to Decisions* (2.utg.). New Jersey: John Wiley & Sons, Inc.
- Natvig, B. (1997). *En introduksjon til bayesiansk statistikk og beslutningsteori* (2.utg.). Oslo: Universitetsforlaget.
- Statistisk sentralbyrå. (2021, 22. april). Kommunefakta Time. Hentet fra <https://www.ssb.no/kommunefakta/time>
- Stavanger Aftenblad. (2021, 22. april). Oversikt over koronasmittede i Rogaland. Hentet fra <https://www.aftenbladet.no/lokalt/i/4q8WxR/oversikt-over-koronasmittede-i-rogaland#/>
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J. & van Aken, M. A. G. (2013) A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. *Child Development*, 85(3), 842-860. <https://doi-org.ezproxy.uis.no/10.1111/cdev.12169>
- Walpole, R. E., Myers R. H., Myers S. L & Ye K. (2012). *Probability & Statistics for Engineers & Scientists* (9.utg.). Boston: Pearson Education, Inc.
- Wikipedia. (2021, 10. mai). Gamma distribution. Hentet fra https://en.wikipedia.org/wiki/Gamma_distribution