Oblig 3b

Gormery K. Wanjiru

10. april 2024

Innhold

1 Begreper

1.1 Hva er en Poisson-prosess?

En poissonprosess er en prosess der hendelser av en bestemt type skjer ifølge spesifiserte sannsynlighetslover, blant annet at antall hendelser som skje i ikke-overlappende intervaller er uavhengige av hverandre. dise egenskaper er:

1. $N(0) = 0$. Initialtilstand: Antall hendelser ved starten er null

2. For enhver $0 \leq t_1 < t_2 \leq t_3 < t_4$ er $N(t_2) - N(t_1)$ og $N(t_4) - N(t_3)$ uavhengige stokastiske variabler. Uavhengighet: Antallet hendelser som skjer i ikke-overlappende intervaller er uavhengige av hverandre.

3. For enhver $t \geq 0$ og $\Delta t > 0$ har vi at

$$P(N(t + \Delta t) - N(t) = 1) = \lambda \Delta t + o(\Delta t),$$

$$P(N(t + \Delta t) - N(t) \geq 2) = o(\Delta t),$$

der hver $o(\Delta t)$ angir en funksjon som oppfyller

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

Sannsynligheten for at en hendelse skjer i et lite tidsintervall Δt er proporsjonal med lengden av intervallet, og sannsynligheten for at mer enn én hendelse skjer i dette intervallet er neglisjerbar.

1.2 Hva er prediktiv fordeling?

En prediktiv fordeling beskriver sannsynligheten for fremtidige observasjoner gitt eksisterende data. for eksempel i bayesian statistikk representerer den distribusjonen av mulige utfall for en fremtidig observasjon basert på en nåværende postulert modell og de observerte dataene.

1.3 Hva er sammenhengen mellom gamma-fordeling og gamma-gamma-fordeling, og hva er forskjellen på de to?

Gamma-fordelingen er en kontinuerlig sannsynlighetsfordeling som ofte brukes til å modellere ventetiden til den k -te hendelsen i en Poisson-prosess.

Den er karakterisert ved to parametere, formparameteren k og raten λ , og den har en bred anvendelse i queuing teori, finans, og forsikringsmatematikk.

Gamma-gamma-fordelingen refererer vanligvis til en sammensatt fordeling hvor formparameteren i en gamma-fordeling selv følger en gamma-fordeling, ofte brukt i bayesianske hierarkiske modeller. Denne fordelingen brukes for å modellere over-dispersjon i data hvor variansen er større enn middelverdien, som er typisk i visse typer forsikringskrav eller kjøpsfrekvensdata.

Sammenhengen mellom dem ligger i at gamma-gamma-fordelingen utvider gamma-fordelingen ved å tillate at dens parametere selv er tilfeldige variabler med sine egne fordelinger, noe som fører til en mer fleksibel modell. Forskjellen er i denne fleksibiliteten og i den ytterligere nivå av usikkerhet som innføres ved å ha parametere som selv er stokastiske.

2 Kapittel 13: oppgave 13d

Gitt en prior normalfordeling for μ med kjent σ, hvor hyperparameterne for prioren er $\kappa_0 = 0$, $\mu_0 = 0$, og $\sigma_0 = 200$. Det er observert $n = 12$ med gjennomsnitt $\bar{x} = 35723.7$ og vi har at $\sigma = 3125$.

Posteriorfordelingen for μ etter å ha observert dataene er også en normalfordeling. Posteriorens parametre, μ_n og σ_n^2, kan beregnes som følger:

$$\mu_n = \frac{\kappa_0\mu_0 + n\bar{x}}{\kappa_0 + n}$$

$$\sigma_n^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$$

Setter vi inn verdiene får vi:

$$\mu_n = \frac{0 \cdot 0 + 12 \cdot 35723.7}{0 + 12} = 35723.7$$

$$\sigma_n^2 = \frac{1}{\frac{1}{200^2} + \frac{12}{3125^2}}$$

Den oppdaterte verdien av σ_n^2 blir den inverse av summen av de inverse variansene.

Sannsynlighetene for μ og prediksjonene for fremtidig observasjon X_+ kan beregnes ved å bruke den oppdaterte posteriore normalfordelingen for μ.

$$P(\mu < 3125) = 99.48\%$$
$$P(X_{+} < 3125) = 76.15\%$$

Oppgave 14(c)

Gitt at stokastiske variabelen X er normalfordelt, $X \sim \mathcal{N}(5, 3.1^2)$, og en nyttefunksjon $u(x)$ som er definert som:

$$u(x) = \begin{cases} -1 & \text{for } x < 3 \\ 1.5 & \text{for } 3 \leq x < 6 \\ 4 & \text{for } x \geq 6 \end{cases}$$

Forventet nytte U kan beregnes ved å ta forventningen av $u(X)$ med hensyn på fordelingen til X:

$$U = E[u(X)] = -1 \cdot P(X < 3) + 1.5 \cdot P(3 \leq X < 6) + 4 \cdot P(X \geq 6)$$

Ved å sette inn de relevante sannsynlighetene fra den normalfordelte stokastiske variabelen får vi:

$$U = -1 \cdot \Phi\left(\frac{3-5}{3.1}\right) + 1.5 \cdot \left(\Phi\left(\frac{6-5}{3.1}\right) - \Phi\left(\frac{3-5}{3.1}\right)\right) + 4 \cdot \left(1 - \Phi\left(\frac{6-5}{3.1}\right)\right)$$

$$U = -1 \cdot \Phi(-0.64516) + 1.5 \cdot (\Phi(0.32258) - \Phi(-0.64516)) + 4 \cdot (1 - \Phi(0.32258))$$

$$U \approx 2.1081$$

3 De tre eksperimentene:

3.1 (a) Bernoulli

3.1.1 i. Intervallestimater

For å regne ut 90% kredibilitetsintervall for parameteren π for alle 4 n-verdier, bruker vi Betafordelingen Beta(a, b), hvor a er antall suksesser pluss 1, og

b er antall feil pluss 1. For $n = 10, 100, 1000, 10000$, anta vi har observerte suksesser gitt i oppgaven.

Kredibilitetsintervallet for en Betafordeling er gitt ved:

$$\text{CI} = \left[\text{Beta}^{-1}\left(\frac{\alpha}{2}; a, b\right), \text{Beta}^{-1}\left(1 - \frac{\alpha}{2}; a, b\right)\right]$$

hvor $\alpha = 0.1$ for et 90% intervall.

3.1.2 ii. Punktestimater

Vi antar at for $n = 100$ og $n = 10000$, antallet suksesser (a) og feil (b) er gitt. Vi beregner punktestimatene for disse to n-verdiene.

For å gjøre dette, bruker vi R. R-kode for beregning av måleparameter (forventningsverdi) og varians for Betafordelingen for $n = 100$ og $n = 10000$.

3.2 R Kode for Punktestimater

```
# For n=100
a100 <- 70 # Antall suksesser + 1
b100 <- 31 # Antall feil + 1
mu100 <- a100 / (a100 + b100)
var100 <- (a100 * b100) / ((a100 + b100)^2 * (a100 + b100 + 1))

# For n=10000
a10000 <- 7000 # Antall suksesser + 1
b10000 <- 3001 # Antall feil + 1
mu10000 <- a10000 / (a10000 + b10000)
var10000 <- (a10000 * b10000) / ((a10000 + b10000)^2 * (a10000 + b10000 + 1))

# resultatene
print(c(mu100, var100))
print(c(mu10000, var10000))
```

3.2.1 iii. Hypotesetesting

For å teste hypotesen $H_1 : \pi > 0.5$ med signifikansnivå $\alpha = 0.1$, beregner vi $P(H_0)$ for alle de fire posterior-fordelingene. Dette kan gjøres ved å bruke den kumulative fordelingsfunksjonen for Betafordelingen i R.

3.3 R Kode for Hypotesetesting

```
#for å beregne P(H0) for de fire posterior-fordelingene
library(stats)

# Anta at a og b er definert som tidligere for hvert n
# Beregning for n=10
a10 <- 7; b10 <- 4
PH0_n10 <- 1 - pbeta(0.5, a10, b10)

# Beregning for n=100, a100 og b100 definert tidligere
PH0_n100 <- 1 - pbeta(0.5, a100, b100)

# Beregning for n=1000
a1000 <- 700; b1000 <- 301
PH0_n1000 <- 1 - pbeta(0.5, a1000, b1000)

# Beregning for n=10000, a10000 og b10000 definert tidligere
PH0_n10000 <- 1 - pbeta(0.5, a10000, b10000)

# resultatene
print(c(PH0_n10, PH0_n100, PH0_n1000, PH0_n10000))
```

Vedlegg

Vedlegg A

Referanser

[1] `https://tma4245.math.ntnu.no/viktige-diskrete-fordelinger/poissonprosess-og-poissonfordeling` *NTNU*