



Oblig 3b

Gruppe 98

av

Christopher Sanden

i

MA-223

Statistikk

Fakultet for teknologi og realfag

Universitetet i Agder

April 2026

# Innhold

<b>1</b>	<b>Innledning</b>	<b>1</b>
<b>2</b>	<b>Begreper</b>	<b>2</b>
2.1	Poisson-prosess . . . . .	2
2.2	Prediktiv fordeling . . . . .	2
2.3	Gamma-fordeling og gamma-gamma-fordeling . . . . .	2
<b>3</b>	<b>Oppgaver fra punkt 3</b>	<b>3</b>
3.1	Kapittel 13, oppgave 13d . . . . .	3
3.2	Kapittel 13, oppgave 14c . . . . .	3
3.3	Kapittel 14, oppgave 12 . . . . .	3
3.4	Kapittel 14, oppgave 14 . . . . .	4
3.5	Kapittel 14, oppgave 16 . . . . .	5

# 1 Innledning

Denne rapporten besvarer Oblig 3b i MA-223 og omhandler bayesianske metoder for modeller med ukjente parametere. Besvarelsen tar utgangspunkt i sentrale begreper som Poisson-prosess, prediktiv fordeling, gamma-fordeling og gamma-gamma-fordeling, og viser hvordan disse brukes i konkrete oppgaver fra pensum.

Rapporten er bygd opp i to deler. Først gis en kort forklaring av de viktigste begrepene som brukes videre i besvarelsen. Deretter løses oppgaver knyttet til posterior- og prediktive fordelinger, samt sammenligning av ulike parametere ved hjelp av sannsynlighetsberegninger og beslutningsregler.

I oppgavene kombineres teoretiske utledninger med numeriske beregninger i R. Målet er både å finne de relevante sannsynlighetene og å vise hvordan bayesiansk analyse kan brukes til å trekke konklusjoner om framtidige observasjoner og sammenligninger mellom ukjente størrelser.

## 2 Begreper

### 2.1 Poisson-prosess

En Poisson-prosess er en stokastisk prosess som teller antall hendelser over tid, der

- $N(0) = 0$ ,
- prosessen har uavhengige inkremitter,
- antall hendelser i et tidsintervall av lengde  $t$  er Poisson-fordelt med parameter  $\lambda t$ .

Hvis  $N(t)$  er antall hendelser fram til tid  $t$ , sa har vi altsa

$$N(t) \sim \text{Poisson}(\lambda t).$$

Parameteren  $\lambda$  er intensiteten, det vil si forventet antall hendelser per tidsenhet.

### 2.2 Prediktiv fordeling

En prediktiv fordeling er fordelingen til en framtidig observasjon nar vi tar hensyn til usikkerheten i parameteren. Hvis  $\theta$  er ukjent parameter og  $X^+$  er en framtidig observasjon, er den prediktive fordelingen gitt ved

$$f(x^+ | \text{data}) = \int f(x^+ | \theta) \pi(\theta | \text{data}) d\theta.$$

Vi integrerer altsa observasjonsmodellen over posteriorfordelingen til parameteren.

### 2.3 Gamma-fordeling og gamma-gamma-fordeling

Gamma-fordelingen brukes ofte som prior eller posterior for en positiv parameter, for eksempel intensiteten  $\lambda$  i en Poisson-prosess:

$$\lambda \sim \Gamma(k, t).$$

Hvis vi sa ser pa en framtidig ventetid eller en framtidig observasjon og integrerer ut  $\lambda$ , far vi en ny fordeling. I denne settingen far ventetiden en gamma-gamma-fordeling (også kalt beta-prime i en alternativ parametrisering).

Sammenhengen er derfor at gamma-gamma-fordelingen oppstar nar en gammafordelt parameter integreres ut av en gammafordelt betinget modell. Forskjellen er at gamma-fordelingen er en fordeling for selve parameteren, mens gamma-gamma-fordelingen er en prediktiv fordeling for en framtidig storrelse.

### 3 Oppgaver fra punkt 3

#### 3.1 Kapittel 13, oppgave 13d

Prioren er

$$\lambda \sim \Gamma(k_0, t_0), \quad k_0 = 3, \quad t_0 = 73,$$

og vi observerer  $n = 6$  hendelser i løpet av  $t = 119$  tidsenheter.

For en Poisson-prosess med gamma-prior er posterioren

$$\lambda \mid \text{data} \sim \Gamma(k_0 + n, t_0 + t).$$

Dermed far vi

$$\lambda \mid \text{data} \sim \Gamma(3 + 6, 73 + 119) = \Gamma(9, 192).$$

#### 3.2 Kapittel 13, oppgave 14c

Vi er gitt posteriorhyperparametrene

$$k_1 = 29, \quad \tau_1 = 8,$$

og skal finne

$$P(T_{+3} \leq 1) \quad \text{og} \quad P(N_{+1} \leq 5).$$

Når  $\lambda \mid \text{data} \sim \Gamma(k_1, \tau_1)$ , far vi den prediktive fordelingen for framtidig antall hendelser som

$$N_{+l} \sim \text{NegBin}\left(k_1, \frac{\tau_1}{\tau_1 + l}\right).$$

Her blir det

$$N_{+1} \sim \text{NegBin}\left(29, \frac{8}{9}\right).$$

Dermed er

$$P(N_{+1} \leq 5) = \sum_{n=0}^5 \binom{n+28}{n} \left(\frac{8}{9}\right)^{29} \left(\frac{1}{9}\right)^n \approx 0.8298.$$

Videre bruker vi sammenhengen

$$T_{+3} \leq 1 \iff N_{+1} \geq 3.$$

Derfor far vi

$$P(T_{+3} \leq 1) = P(N_{+1} \geq 3) = 1 - P(N_{+1} \leq 2) \approx 0.6849.$$

#### 3.3 Kapittel 14, oppgave 12

Vi bruker neutral priorer. Da far hver middelerdi en marginal posteriorfordeling av Student- $t$ -type:

$$\mu_O \mid \text{data} \sim t_7\left(\bar{x}_O, \frac{s_O}{\sqrt{8}}\right), \quad \mu_K \mid \text{data} \sim t_6\left(\bar{x}_K, \frac{s_K}{\sqrt{7}}\right).$$

Fra dataene far vi

$$\bar{x}_O = 20.625, \quad s_O = 0.9161, \quad \bar{x}_K = 15.5714, \quad s_K = 7.3679.$$

Hypotesen er

$$H_1 : \mu_O > \mu_K,$$

med signifikans  $\alpha = 0.2$ . I denne typen oppgave velger vi  $H_1$  dersom

$$P(\mu_O > \mu_K \mid \text{data}) > 1 - \alpha = 0.8.$$

Ved numerisk integrasjon far vi

$$P(\mu_O > \mu_K \mid \text{data}) \approx 0.9392.$$

Siden  $0.9392 > 0.8$ , velger vi  $H_1$ . Dataene gir derfor tilstrekkelig støtte for at Odds hosteanfall i gjennomsnitt varer lenger enn Kjells.

### 3.4 Kapittel 14, oppgave 14

Vi bruker samme beslutningsregel gjennom hele oppgaven: velg  $H_1$  nar den posterior sannsynligheten for utsagnet i  $H_1$  er større enn  $1 - \alpha$ .

**14a** Vi har

$$\Theta \sim \Gamma(7, 70), \quad \Psi \sim \Gamma(4, 80),$$

og hypotesen

$$H_1 : \Theta > \Psi, \quad \alpha = 0.05.$$

Den relevante posterior sannsynligheten er

$$P(\Theta > \Psi) = \int_0^\infty f_\Theta(x) F_\Psi(x) dx \approx 0.8774.$$

Siden  $0.8774 < 0.95$ , velger vi  $H_0$ .

**14b** Vi har

$$\Theta \sim \Gamma(9, 20), \quad \Psi \sim \Gamma(11, 20),$$

og hypotesen

$$H_1 : \Theta < \Psi, \quad \alpha = 0.1.$$

Her gir beregningen

$$P(\Theta < \Psi) \approx 0.6762.$$

Siden  $0.6762 < 0.9$ , velger vi  $H_0$ .

**14c** Nar parameterne i 14b ganges med 10, far vi

$$\Theta \sim \Gamma(90, 200), \quad \Psi \sim \Gamma(110, 200),$$

med samme hypotese som i 14b. Da blir

$$P(\Theta < \Psi) \approx 0.9220.$$

Siden  $0.9220 > 0.9$ , velger vi na  $H_1$ .

Det er altså en forskjell: med ti ganger så mye informasjon blir posteriorfordelingene mer konsentrerte, og det blir lettere å skille mellom parameterne.

### 3.5 Kapittel 14, oppgave 16

Her sammenligner vi beta-fordelte variable. Jeg viser både eksakt beregning og normalapprosimasjon.

For normalapprosimasjonen bruker vi at dersom

$$U \sim \beta(a, b),$$

sa er

$$E(U) = \frac{a}{a+b}, \quad \text{Var}(U) = \frac{ab}{(a+b)^2(a+b+1)}.$$

For to uavhengige variable  $\psi$  og  $\pi$  approksimerer vi deretter

$$D = \psi - \pi \approx N(E(\psi) - E(\pi), \text{Var}(\psi) + \text{Var}(\pi)).$$

**16a**

$$\psi \sim \beta(2, 5), \quad \pi \sim \beta(4, 3), \quad H_1 : \psi < \pi, \quad \alpha = 0.1.$$

Eksakt beregning gir

$$P(\psi < \pi) \approx 0.8788,$$

mens normalapprosimasjonen gir

$$P(\psi < \pi) \approx 0.8861.$$

Begge er mindre enn 0.9, sa vi velger  $H_0$ .

**16b**

$$\psi \sim \beta(23, 17), \quad \pi \sim \beta(17, 23), \quad H_1 : \psi > \pi, \quad \alpha = 0.1.$$

Eksakt beregning gir

$$P(\psi > \pi) \approx 0.9131,$$

og normalapprosimasjonen gir

$$P(\psi > \pi) \approx 0.9153.$$

Begge er større enn 0.9, sa vi velger  $H_1$ .

**16c**

$$\psi \sim \beta(20, 20), \quad \pi \sim \beta(17, 23), \quad H_1 : \psi > \pi, \quad \alpha = 0.05.$$

Eksakt beregning gir

$$P(\psi > \pi) \approx 0.7520,$$

og normalapproximasjonen gir

$$P(\psi > \pi) \approx 0.7527.$$

Begge er mindre enn 0.95, så vi velger  $H_0$ .

**16d** Nar parameterne i 16c ganges med 10, far vi

$$\psi \sim \beta(200, 200), \quad \pi \sim \beta(170, 230), \quad H_1 : \psi > \pi, \quad \alpha = 0.05.$$

Eksakt beregning gir

$$P(\psi > \pi) \approx 0.9834,$$

og normalapproximasjonen gir

$$P(\psi > \pi) \approx 0.9837.$$

Begge er større enn 0.95, så vi velger  $H_1$ .

Dette viser at mer data kan gi en tydelig forskjell selv nar forholdet mellom positive og negative observasjoner er det samme som for 16c.



## R-kode

```
1  # Posterior for a Poisson process with Gamma(shape = k0, rate = t0) prior
2  # and data n events observed during time t.
3  poisson_posterior <- function(k0, t0, n, t) {
4    list(shape = k0 + n, rate = t0 + t)
5  }
6
7  # Predictive probability for future counts in a Poisson process when
8  # lambda | data ~ Gamma(shape, rate).
9  # R's pnbinom uses the same negative-binomial form we need here.
10 predictive_count_cdf <- function(m, shape, rate, horizon) {
11   pnbinom(m, size = shape, prob = rate / (rate + horizon))
12 }
13
14 # Predictive probability for the k-th future waiting time.
15 # We use  $P(T_{\{+k\}} \leq l) = P(N_{\{+l\}} \geq k) = 1 - P(N_{\{+l\}} \leq k - 1)$ .
16 predictive_waiting_time_cdf <- function(k, l, shape, rate) {
17   1 - predictive_count_cdf(k - 1, shape, rate, l)
18 }
19
20 # Posterior probability  $P(X > Y)$  for independent gamma variables.
21 gamma_gt_prob <- function(shape_x, rate_x, shape_y, rate_y) {
22   integrate(
23     function(x) {
24       dgamma(x, shape = shape_x, rate = rate_x) *
25       pgamma(x, shape = shape_y, rate = rate_y)
26     },
27     lower = 0,
28     upper = Inf,
29     subdivisions = 2000L,
30     rel.tol = 1e-10
31   )$value
32 }
33
34 # Exact probability  $P(U > V)$  for independent beta variables with integer
35 # first shape parameter in U, from the finite-sum identity used in Rule A.3.3.
36 beta_gt_prob_exact <- function(a, b, c, d) {
37   total <- 0
38   for (i in 0:(a - 1)) {
39     total <- total + beta(c + i, b + d) /
40       ((b + i) * beta(1 + i, b) * beta(c, d))
41   }
42   total
43 }
44
45 # Normal approximation for pairwise beta comparison.
46 # If U and V are independent beta variables, then  $D = U - V$  is approximated
47 # by a normal variable with mean  $E(U) - E(V)$  and variance  $Var(U) + Var(V)$ .
48 beta_compare_prob_normal <- function(a, b, c, d, direction = c("gt", "lt")) {
49   direction <- match.arg(direction)
50 }
```

```

51 mean_u <- a / (a + b)
52 mean_v <- c / (c + d)
53 var_u <- a * b / ((a + b)^2 * (a + b + 1))
54 var_v <- c * d / ((c + d)^2 * (c + d + 1))
55
56 mean_d <- mean_u - mean_v
57 sd_d <- sqrt(var_u + var_v)
58
59 if (direction == "gt") {
60   1 - pnorm(0, mean = mean_d, sd = sd_d)
61 } else {
62   pnorm(0, mean = mean_d, sd = sd_d)
63 }
64 }
65
66 # Posterior probability  $P(\mu_x > \mu_y)$  for Gaussian samples under neutral priors.
67 # Marginally, each mean has a Student t posterior:
68 #  $\mu \mid \text{data} \sim t_{\{n-1\}}(\bar{x}, s / \sqrt{n})$ .
69 gaussian_mean_gt_prob <- function(x, y) {
70   nx <- length(x)
71   ny <- length(y)
72   mx <- mean(x)
73   my <- mean(y)
74   sx <- sd(x)
75   sy <- sd(y)
76
77   integrate(
78     function(z) {
79       dt((z - mx) / (sx / sqrt(nx)), df = nx - 1) / (sx / sqrt(nx)) *
80       pt((z - my) / (sy / sqrt(ny)), df = ny - 1)
81     },
82     lower = -Inf,
83     upper = Inf,
84     subdivisions = 5000L,
85     rel.tol = 1e-10
86   )$value
87 }
88
89 # -----
90 # Point 3a: Chapter 13, task 13d
91 # -----
92
93 post_13d <- poisson_posterior(k0 = 3, t0 = 73, n = 6, t = 119)
94
95 cat("Chapter 13, task 13d\n")
96 cat("Posterior shape =", post_13d$shape, "\n")
97 cat("Posterior rate =", post_13d$rate, "\n\n")
98
99 # -----
100 # Point 3b: Chapter 13, task 14c
101 # -----
102
103 shape_14c <- 29

```

```

104 rate_14c <- 8
105 k_14c <- 3
106 l_14c <- 1
107 m_14c <- 5
108
109 p_n_le_m <- predictive_count_cdf(m = m_14c, shape = shape_14c, rate = rate_14c, horizon =
  ↪ l_14c)
110 p_t_le_l <- predictive_waiting_time_cdf(k = k_14c, l = l_14c, shape = shape_14c, rate =
  ↪ rate_14c)
111
112 cat("Chapter 13, task 14c\n")
113 cat("P(N_{+1} <= 5) =", p_n_le_m, "\n")
114 cat("P(T_{+3} <= 1) =", p_t_le_l, "\n\n")
115
116 # -----
117 # Point 3c: Chapter 14, task 12
118 # -----
119
120 odd <- c(22, 20, 21, 20, 21, 21, 19, 21)
121 kjell <- c(15, 12, 32, 12, 11, 13, 14)
122
123 p_mu_odd_gt_kjell <- gaussian_mean_gt_prob(odds, kjell)
124
125 cat("Chapter 14, task 12\n")
126 cat("Odd: n =", length(odds), "mean =", mean(odds), "sd =", sd(odds), "\n")
127 cat("Kjell: n =", length(kjell), "mean =", mean(kjell), "sd =", sd(kjell), "\n")
128 cat("P(mu_Odd > mu_Kjell | data) =", p_mu_odd_gt_kjell, "\n\n")
129
130 # -----
131 # Point 3d: Chapter 14, task 14
132 # -----
133
134 p_14a <- gamma_gt_prob(7, 70, 4, 80)
135 p_14b <- 1 - gamma_gt_prob(9, 20, 11, 20)
136 p_14c <- 1 - gamma_gt_prob(90, 200, 110, 200)
137
138 cat("Chapter 14, task 14\n")
139 cat("P(Theta > Psi) in 14a =", p_14a, "\n")
140 cat("P(Theta < Psi) in 14b =", p_14b, "\n")
141 cat("P(Theta < Psi) in 14c =", p_14c, "\n\n")
142
143 # -----
144 # Point 3e: Chapter 14, task 16
145 # -----
146
147 # 16a: H1 is psi < pi, so exact probability is 1 - P(psi > pi).
148 p_16a_exact <- 1 - beta_gt_prob_exact(2, 5, 4, 3)
149 p_16a_norm <- beta_compare_prob_normal(2, 5, 4, 3, direction = "lt")
150
151 # 16b: H1 is psi > pi.
152 p_16b_exact <- beta_gt_prob_exact(23, 17, 17, 23)
153 p_16b_norm <- beta_compare_prob_normal(23, 17, 17, 23, direction = "gt")

```

```

154
155 # 16c: H1 is psi > pi.
156 p_16c_exact <- beta_gt_prob_exact(20, 20, 17, 23)
157 p_16c_norm <- beta_compare_prob_normal(20, 20, 17, 23, direction = "gt")
158
159 # 16d: same as 16c, but all parameters multiplied by 10.
160 p_16d_exact <- beta_gt_prob_exact(200, 200, 170, 230)
161 p_16d_norm <- beta_compare_prob_normal(200, 200, 170, 230, direction = "gt")
162
163 cat("Chapter 14, task 16\n")
164 cat("16a exact =", p_16a_exact, "\n")
165 cat("16a normal =", p_16a_norm, "\n")
166 cat("16b exact =", p_16b_exact, "\n")
167 cat("16b normal =", p_16b_norm, "\n")
168 cat("16c exact =", p_16c_exact, "\n")
169 cat("16c normal =", p_16c_norm, "\n")
170 cat("16d exact =", p_16d_exact, "\n")
171 cat("16d normal =", p_16d_norm, "\n")

```

---

**Listing 1:** R-kode for hele rapporten