

# Rapport: Oblig 3a

Gormery K. Wanjiru

Mars 2024

## 1 Oppgave 1 (5%)

**Kapittel 11: oppgave 2 utelatt for nå**

## 2 Oppgave 2 (10%)

### (a) Hva er en parameter?

En parameter er en ukjent tallverdi som beskriver en befolkning. Parametere er sentrale i statistiske modeller fordi de hjelper til med å definere fordelingen av en datamengde.

### (b) Hva er en observator?

En observator er en målbar funksjon av dataene, som brukes til å estimere verdien av en parameter.

### (c) Hva er en hyperparameter?

En hyperparameter er en parameter i en bayesiansk statistisk modell som ikke er direkte relatert til datamengden, men som påvirker fordelingen av modellparametere.

### (d) Hva er en Poisson-prosess?

En Poisson-prosess er en type statistisk prosess som beskriver antall hendelser som skjer i et fast tidsintervall, under forutsetning av at disse hendelsene skjer med en konstant rate og uavhengig av hverandre. Enkel sagt den beskriver hendelser som skjer tilfeldig over tid, hvor antall hendelser i et gitt tidsintervall følger en Poisson-fordeling

**(e) Hva er en Bernoulli-prosess?**

En Bernoulli-prosess er en sekvens av uavhengige og identisk fordelt binære tilfeldige variabler, hvor hver variabel kan anta verdien 1 med sannsynlighet  $p$  og 0 med sannsynlighet  $1 - p$ .

**(f) Hva er en posterior fordeling?**

En posterior fordeling er sannsynlighetsfordelingen for en parameter gitt observasjonsdata, basert på en kombinasjon av priorinformasjon og den likelihood som dataene gir.

**(g) Hva er en prediktiv fordeling?**

En prediktiv fordeling er en fordeling av fremtidige observasjoner gitt eksisterende data, som tar hensyn til usikkerheten rundt parameterestimaterne.

## **3 Gaussisk**

### **3.1 Oppgave 3 til 4**

utelatt for nå

## **4 Bernoulli**

### **4.1 Oppgave 5 til 7**

utelatt for nå

## **5 Poisson**

### **5.1 Oppgave 8 til 9**

utelatt for nå

## **6 Oppgave 10: Tre inferenser**

### **a) Bernoulli (25%)**

#### **i. Grunnleggende**

For å generere 30 Bernoulli-forsøk med parameter  $p = 0.349$ , kan vi bruke følgende R-kode:

```
# Genererer 30 uavhengige Bernoulli-forsøk
rbinom(30, 1, 0.349)
# Genererer antall suksesser i 30 Bernoulli-forsøk
rbinom(1, 30, 0.349)
```

Den første linjen utfører 30 uavhengige forsøk, som returnerer et vektor med 30 elementer, hvor hvert element representerer utfallet av hvert forsøk (suksess eller fiasko). Den andre linjen utfører et sett av 30 forsøk og returnerer totalt antall suksesser. Den første metoden gir mer detaljert informasjon om hvert enkelt forsøk, mens den andre gir en summarisk oversikt over antall suksesser.

## ii. Observasjons-versjon

Denne koden simulerer 50 grupper av 10 Bernoulli-forsøk hver, beregner suksessraten for hver gruppe, og visualiserer distribusjonen av disse ratene gjennom et histogram. Koden markerer også gjennomsnittet, standardavviket, og den sanne verdien av  $p$ .

```
p = 0.349
m = 50
n = 10
z = rbinom(m, n, p) / n
hist(z, breaks = seq(-0.5, n + .5, 1) / n)
m = mean(z)
s = sd(z)
abline(v = m, col = "green", lwd = 1)
abline(v = m + c(-s, s), col = "pink", lwd = 1)
abline(v = p, col = "blue", lwd = 1)
```

Dette illustrerer hvordan empiriske suksessrater distribuerer seg rundt den sanne parameteren  $p$ , med gjennomsnitt og spredning visualisert.

## iii. Effekt av større $m$ og $n$

Økningen i  $m$  (antall grupper av forsøk) og  $n$  (antall forsøk per gruppe) påvirker resultatene på følgende måter:

- Økning i  $m$  gir en jevnere og mer stabil estimasjon av fordelingen av suksessraten siden vi har flere datapunkter å basere estimatene på.
- Økning i  $n$  vil typisk føre til en smalere distribusjon rundt den sanne verdien av  $p$ , ettersom loven om store tall er at gjennomsnittet av en stor prøve vil nærme seg den forventet verdien.

## iv. Inferens-versjon

For inferens bruker vi Jeffreys' nøytrale prior for en Bernoulli-prosess, som er en Beta-distribusjon med  $a_0 = 0.5$  og  $b_0 = 0.5$ . Dette reflekterer en usikkerhet før vi ser data, og er spesielt valgt fordi den er ikke-informativ.

utfører deretter forsøkene for  $n = 10, 100, 1000, 10000$  og oppdaterer vår tro basert på observasjonene:

```
# Jeffreys' nøytrale prior
a0 = 0.5
b0 = 0.5

# Simulerer forsøk og oppdaterer troen for forskjellige n
n_values = c(10, 100, 1000, 10000)
for (n in n_values) {
  k = sum(rbinom(n, 1, p)) # Antall suksesser
  l = n - k # Antall fiaskoer
  a1 = a0 + k
  b1 = b0 + l
  # Tegn posterior sannsynlighetsfordeling for p
  curve(dbeta(x, a1, b1), 0, 1, main = paste("Posterior for n=", n))
}
```

Den demonstrerer hvordan vår kunnskap om  $p$  forbedres med flere data, og hvordan posterior distribusjonen blir mer konsentrert rundt den sanne verdien av  $p$  som  $n$  øker.

### **b) Poisson (25%)**

utelatt for nå

### **c) Gaussisk (25%)**

utelatt for nå