Løsningsforslag øving 11

11.30

Vi antar at X er en Bernoulli-variabel. Da kan X enten ta verdien 0 eller 1. Vi vet at $0^2 = 0$ og $1^2 = 1$. Det betyr at $(\omega : X(\omega) = 0) = (\omega : X^2(\omega) = 0)$ og $(\omega : X(\omega) = 1) = (\omega : X^2(\omega) = 1)$. Da har vi vist at $X(\omega) = X^2(\omega)$ for alle verdier av ω i utfallsrommet. Altså er de to tilfeldige variablene like.

11.3

Tapsfunksjonen vår er gitt som $l(\Gamma, \hat{\gamma}) = 1 - \delta(\Gamma - \hat{\gamma})$. Da er risikoen lik

$$r(\hat{\gamma}) = E[l(\Gamma, \hat{\gamma})] = \int l(\gamma, \hat{\gamma}) g(\gamma) d\gamma = \int (1 - \delta(\gamma - \hat{\gamma})) g(\gamma) d\gamma.$$

Likning (11.3) sier at

$$\int \delta(\gamma - \hat{\gamma}) g(\gamma) d\gamma = g(\hat{\gamma})).$$

Dette kan vi bruke for å finne forventningsverdien til tapsfunksjonen. Vi får

$$r(\hat{\gamma}) = \int (1 - \delta(\gamma - \hat{\gamma})) g(\gamma) d\gamma = \int g(\gamma) d\gamma - \int \delta(\gamma - \hat{\gamma}) g(\gamma) d\gamma = 1 - g(\hat{\gamma}).$$

Vi ønsker å minimere risikoen. Da er det klart at vi må maksimere g. Derfor blir den optimale estimatoren lik

$$\hat{\gamma} = \arg\min(1 - g) = \arg\max(g).$$

11.4

Vi får

$$\begin{aligned} E\,|(\Gamma - \mu) + (\mu - \hat{\gamma})|^2 &= E\left[(\Gamma - \mu)^2 + 2(\Gamma - \mu)(\mu - \hat{\gamma}) + (\mu - \hat{\gamma})^2\right] \\ &= E\left[(\Gamma - \mu)^2\right] + 2(\mu - \hat{\gamma}) E\left[(\Gamma - \mu)\right] + E\left[(\mu - \hat{\gamma})^2\right], \end{aligned}$$

fordi forventningsverdi er lineær. Det midterste leddet ser vi at blir lik null, fordi $E(\Gamma - \mu) = \mu - \mu = 0$. Det venstre leddet kjenner vi igjen som definisjonen av variansen. Det høyre leddet er forventningsverdien til en konstant, som betyr at vi kan fjerne forventningsverdien. Da kan vi se at vi får

$$E\,|(\Gamma - \mu) + (\mu - \hat{\gamma})|^2 = \mathrm{Var}(\Gamma) + (\mu - \hat{\gamma})^2 = \mathrm{Var}(\Gamma) + |\mu - \hat{\gamma}|^2.$$

11.7

Hvis det ikke er slik at $E|\Gamma| < \infty$, så vil $\arg\min E|\Gamma - \gamma| = \arg\min \infty$ være udefinert, og vi kan ikke finne medianen fra dette uttrykket, som er grunnen til at vi må bruke en alternativ definisjon,

$$m = \arg\min_{\gamma} E(|\Gamma - \gamma| - |\Gamma - \gamma_0|).$$

Hvis det derimot er slik at $E|\Gamma| < \infty$, så kan vi forenkle definisjonen. Da får vi

$$m = \arg\min_{\gamma} E(|\Gamma - \gamma| - |\Gamma - \gamma_0|) = \arg\min_{\gamma} E|\Gamma - \gamma| - E|\Gamma - \gamma_0|.$$

Her er det klart at det andre leddet kun er en konstant, som ikke inneholder γ. Derfor er

$$\arg\min_{\gamma} E|\Gamma - \gamma| - E|\Gamma - \gamma_0| = \arg\min_{\gamma} E|\Gamma - \gamma|,$$

og vi får definisjonen for medianen:

$$m = \arg\min_{\gamma} E|\Gamma - \gamma|.$$

11.12

α-fraktilen til en tilfeldig variabel Γ er definert slik at $P(\Gamma \leq \gamma_\alpha) = \alpha$, hvor γ_α er α-fraktilen. Hvis γ_α er α-fraktilen til Γ, så er $P(\Gamma \leq \gamma_\alpha) = \alpha$. Vi trekker fra γ fra begge sider og får $P(\Gamma - \gamma \leq \gamma_\alpha - \gamma) = \alpha$. Derfor er $\gamma_\alpha - \gamma$ lik α-fraktilen til $\Gamma - \gamma$.

11.13

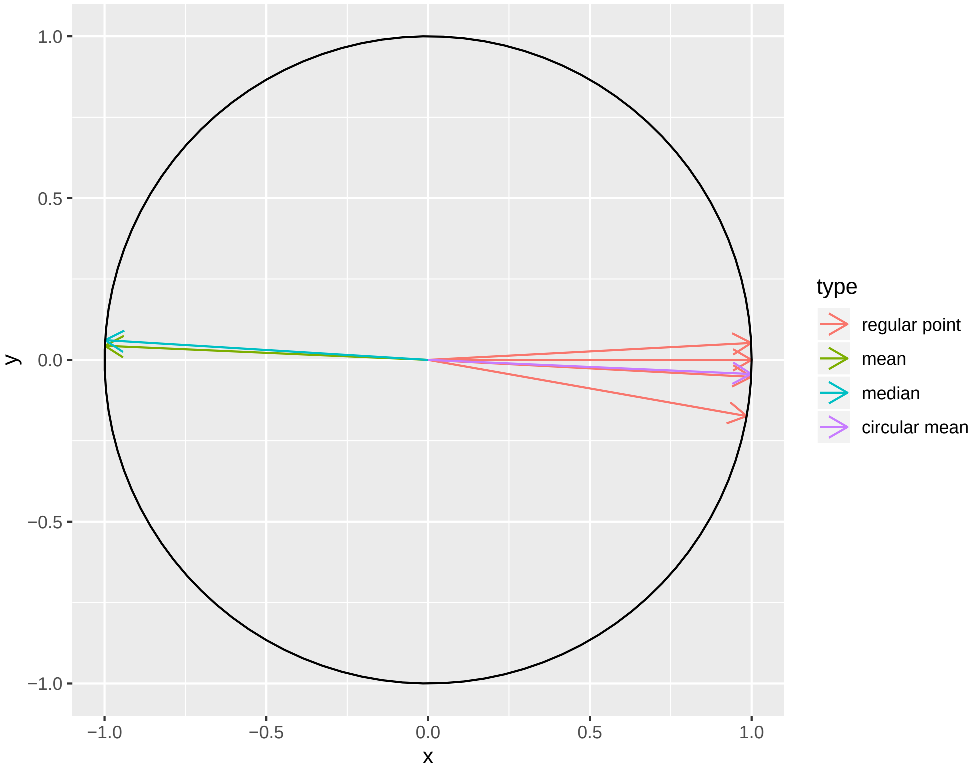
Her er $\hat{\gamma}$ et estimat for hvor mange pølser som kommer til å bli kjøpt på 17. mai, mens γ er den sanne verdien. Når $\hat{\gamma} \leq \gamma$ så har Bør Børson bestilt for få pølser, og da har han en tapt inntekt på $a = a_1 N$ pga. pølser som kunne blitt solgt. Når $\hat{\gamma} > \gamma$ har det blitt bestilt for mange pølser, som Bør Børson ikke får solgt. Da taper han $b = b_1 N$ pga. innkjøpte pølser som aldri blir solgt. Da representerer b_1 innkjøpsprisen per pølse. Hvis $a > b$ betyr det at tapet er større om det kjøpes inn en pølse for lite enn om det kjøpes inn en pølse for mye. Da er det mer optimalt å kjøpe inn litt for mange pølser, siden dette vil føre til et mindre tap enn om det ble kjøpt inn for få pølser. Da er det klart at innkjøpsestimatet $\hat{\gamma}$ bør være større enn medianen til γ, slik at det er mindre enn 50% sannsynlighet for at det kjøpes inn for få pølser. Caspar modellerer γ ved en betafordeling, og kommer frem til at denne har parametrene $s \approx 153.2$ og $f \approx 229.8$. Dette gir $E\Gamma \approx 0.4$ og $\mathrm{Var}(\Gamma) \approx 0.0006$. Fordelingen har altså svært lav varians, som vil si at "de fleste" fraktilene ligger tett inntil hverandre. Derfor er det liten avstand mellom medianen og estimatet for γ.

11.14

Vi tegner opp de 4 observasjonene som punkter på enhetssirkelen. Punktene kan sees i bildet nedenfor. Gjennomsnittet av vinklene vil være et dårlig mål på midlere retning. Vi får $\bar{\theta} = 177.5°$, som vil si at en vektor med vinkel $\bar{\theta}$ vil peke nesten rett mot venstre. For å regne ut sirkulær middelverdi summerer vi de fire vektorene, og får vektoren $(3.98, -0.17)$, som har vinkel $\theta = 357.5°$. Empirisk median for de observerte retningene er lik $(350° - 3°)/2 = 173.5°$, og empirisk forventning er lik gjennomsnitt, så den har vi allerede regnet ut.

```
library(ggplot2)

## Create data.frame of the four points
thetas <- c(0, 350, 3, 357) # angle in degrees
thetas <- thetas * pi / 180 # angle in radians
x <- cos(thetas)
y <- sin(thetas)
df <- data.frame(x, y, type = "point")

## Plot the mean, median and circular mean angles
theta_mean <- mean(thetas)
theta_median <- median(thetas)
mean_df <- data.frame(x = cos(theta_mean),
y = sin(theta_mean),
type = "mean")
median_df <- data.frame(x = cos(theta_median),
y = sin(theta_median),
type = "median")
xy_circ <- c(sum(x), sum(y))
xy_circ <- xy_circ / sqrt(sum(xy_circ^2))
circ_df <- data.frame(x = xy_circ[1],
y = xy_circ[2],
type = "circular mean")

df <- rbind(df, mean_df, median_df, circ_df)

## Plot the four points
plot <- ggplot() +
geom_segment(data = df,
aes(xend = x, yend = y, y = 0, x = 0, col = type),
arrow = arrow(length = unit(.15, "inches")))

## Plot the unit circle
unit_circle <- data.frame(theta = seq(0, 2 * pi, length = 100))
unit_circle$x <- cos(unit_circle$theta)
unit_circle$y <- sin(unit_circle$theta)
plot <- plot +
geom_path(data = unit_circle, aes(x = x, y = y))
```

```
## Draw the plot
plot
```



11.54

Caspar har samlet inn data og observert $s = 14$ suksesser og $f = 18$ fiaskoer, i det som kan beskrives ved en Bernoulli forsøksrekke med suksessannsynlighet θ. Med en Beta(α, β) a priori fordeling for suksessen får vi altså a posteriori en Beta$(s + \alpha, f + \beta)$-fordeling. Ved å sette inn tallene $\alpha \approx 153.2$ og $\beta \approx 229.8$ får vi en Beta$(167.2, 247.8)$-fordeling. Dette er ikke en veldig stor relativ endring i parametrene. Vi finner $E[\Gamma|Y] \approx 0.4$ og $\mathrm{Var}(\Gamma|Y) \approx 0.0006$, så lite har endret seg, og konklusjonen er tilnærmet den samme

11.55

Vi antar en kvadratisk tapsfunksjon her, da dette er standard fremgangsmåte. En kvadratisk tapsfunksjon fører til at Bayes-estimatoren er a posteriori forventningsverdi. Vi finner først a posteriori-fordelingen for lengden l gitt observasjonene y. A priori-fordelingen til l er gitt ved

$$\pi(l) = \frac{1}{\sqrt{2\pi\sigma_l^2}} e^{-\frac{(l-\mu_l)^2}{2\sigma_l^2}},$$

og fordelingen til observasjonene gitt lengden er lik

$$\pi(y_1, \ldots, y_n | l) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - l)^2} = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n \left((y_i-\bar{y})^2 + (l-\bar{y})^2\right)}$$

Dette gir a posteriori-fordelingen

$$\begin{aligned}
\pi(l|y_1, \ldots, y_n) &\propto \pi(y_1, \ldots, y_n)\pi(l) \\
&\propto \exp\left\{-\frac{1}{2\sigma^2} n(l - \bar{y}^2)\right\} \exp\left\{-\frac{(l-\mu_l)^2}{2\sigma_l^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(l^2\left(\frac{1}{\sigma_l^2} + \frac{n}{\sigma^2}\right) - 2l\left(\frac{\mu_l}{\sigma_l^2} + \frac{n\bar{y}}{\sigma^2}\right)\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(l^2 \frac{\sigma^2 + n\sigma_l^2}{\sigma^2\sigma_l^2} - 2l\frac{\sigma^2\mu_l + n\sigma_l^2\bar{y}}{\sigma^2\sigma_l^2}\right)\right\}.
\end{aligned}$$

Vi kjenner igjen den kvadratiske formen inne i en eksponentialfunksjon som den algebraiske kjernen til en normalfordeling. Vi prøver derfor å skrive om uttrykket til formen $\exp\left\{-\frac{(l-\mu_{l|y})^2}{2\sigma_{l|y}^2}\right\}$. Da får vi

$$\begin{aligned}
\pi(l|y_1, \ldots, y_n) &\propto \exp\left\{-\frac{1}{2}\left(l^2 \frac{\sigma^2 + n\sigma_l^2}{\sigma^2\sigma_l^2} - 2l\frac{\sigma^2\mu_l + n\sigma_l^2\bar{y}}{\sigma^2\sigma_l^2}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2\frac{\sigma^2+n\sigma_l^2}{\sigma^2\sigma_l^2}}\left(l - \frac{\sigma^2\mu_l + n\sigma_l^2\bar{y}}{\sigma^2 + \sigma_l^2 n}\right)\right\}.
\end{aligned}$$

Dette betyr at Bayes-estimatoren er lik

$$E[l|y_1, \ldots, y_n] = \frac{\sigma^2\mu_l + n\sigma_l^2\bar{y}}{\sigma^2 + n\sigma_l^2} = \frac{10^{-6}\cdot 2 + 11\cdot 100^2 \cdot 2.5}{10^{-6} + 11 \cdot 100^2}\text{m} \approx 2.5\text{m}.$$

Dette er et rimelig svar, siden det er det samme som den observerte middelverdien. Vanligvis ville a priori-fordelingen påvirket oss slik at a posteriori-fordelingen ble endret, men i dette tilfellet er usikkerheten a priori såpass stor sammenliknet med den observerte middelverdien og den kjente målefeilen, at vår a priori kunnskap blir fullstendig unyttig, og ikke påvirker resultatet vårt.

Obs: her ville alle andre "standard" valg av tapsfunksjon også gitt det samme resultatet, dette fordi i en normalfordeling så er snittet lik medianen lik toppunktet til sannsynlighetstettheten.

11.56

Dataene $y_1, \ldots, y_n$ er gammafordelte med parametre α og $\lambda = 1/\beta$, som gir

$$\pi(y_1, \ldots, y_n|\alpha, \lambda) = \prod_{i=1}^{n} \pi(y_i|\alpha, \lambda) = \prod_{i=1}^{n} \frac{\lambda^\alpha}{\Gamma(\alpha)} y_i^{\alpha-1} e^{-y_i\lambda} = \frac{\lambda^{n\alpha}}{\Gamma(\alpha)^n} \left(\prod_{i=1}^{n} y_i\right)^{\alpha-1} \exp\left\{-\lambda \sum_{i=1}^{n} y_i\right\}.$$

Verdien til α er kjent, mens λ har a priori-fordelingen

$$\pi(\lambda) = \frac{1}{b^a\Gamma(a)} \lambda^{a-1} e^{-\lambda/b}.$$

Da er a posteriori-fordelingen til λ lik

$$\begin{aligned}
\pi(\lambda|y_1, \ldots, y_n) &\propto \pi(y_1, \ldots, y_n|\alpha, \lambda)\pi(\lambda) \\
&= \frac{\lambda^{n\alpha}}{\Gamma(\alpha)^n} \left(\prod_{i=1}^{n} y_i\right)^{\alpha-1} \exp\left\{-\lambda \sum_{i=1}^{n} y_i\right\} \frac{1}{b^a\Gamma(a)} \lambda^{a-1} e^{-\lambda/b} \\
&\propto \lambda^{n\alpha} \exp\{-\lambda n\bar{y}\} \lambda^{a-1} \exp\{-\lambda/b\} \\
&= \lambda^{(n\alpha+a)-1} \exp\left\{-\lambda(n\bar{y} + b^{-1})\right\}.
\end{aligned}$$

Dette kjenner vi igjen som den algebraiske kjernen til en Gammafordeling med parametre $n\alpha + a$ og $(n\bar{y} + b^{-1})^{-1}$. Vi har altså

$$[\lambda|y_1, \ldots, y_n] \sim \text{Gamma}\left(n\alpha + a, (n\bar{y} + b^{-1})^{-1}\right)$$