Løsningsforslag Eksamen i Statistikk Nov 2001

Oppgave 1

a)

Det fins 8 mulige kombinasjoner. Disse finnes ved å utelate ett og ett tall. Antall utvalg av størrelse 7 blant m er $\binom{m}{7}$.

$$\begin{aligned} \text{Pris} &= \text{Antall Rekker} \cdot 3\,\text{kr.} \\ &= \binom{12}{7} \cdot 3\,\text{kr.} \\ &= \frac{12 \cdot \ldots \cdot 6}{1 \cdot \ldots \cdot 7} \cdot 3\,\text{kr.} \\ &= 2376\,\text{kr.} \end{aligned}$$

b)

Tilnærmelsen, når $np = \mu$, kommer fra resultatet;

$$\lim_{n\to\infty} b(x; n, p) = p(x; \mu).$$

$b(x; n, p)$ er binomisk med parametre n og p. $p(x; \mu)$ er poisson med parametre μ.

Dersom p er liten og n stor, vil poissonfordelingen være en god tilnærmelse til den binomiske fordelingen. Tilnærmelsen vil være gitt ved

$$\begin{aligned} P(X = x) &= b(x; n, p) \\ &\approx p(x; \mu) \\ &= \frac{\mu^x}{x!} \exp(-\mu), \end{aligned}$$

hvor

$$\begin{aligned} \mu &= n \cdot p \\ &= \frac{21\,481\,335}{5\,379\,616} \\ &\approx 4.0. \end{aligned}$$

Poissontilnærmelse:

$$
\begin{aligned}
P(X=0) &= \exp(-4) \\
&\approx 0.0183, \\
P(X=1) &= 4\exp(-4) \\
&\approx 0.0733.
\end{aligned}
$$

Binomisk tilnærmelse:

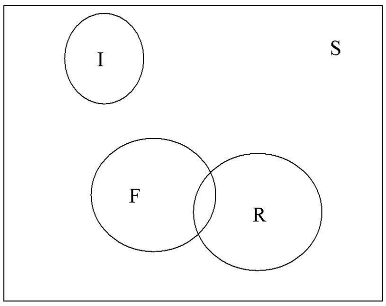
$$
\begin{aligned}
P(X=0) &= (1-p)^n \\
&\approx 0.0184, \\
P(X=1) &= 4\cdot(1-p)^n \\
&\approx 0.0736.
\end{aligned}
$$

Vi ser at tilnærmelsen er meget god.

Oppgave 2

a)

Merk fra Venn diagram at I ikke overlapper F eller R.



$$P(R|F) = \frac{P(R\cap F)}{P(F)} = \frac{0.3}{0.5} = 0.6$$

$$P(R|I') = \frac{P(R\cap I')}{P(I')} = \frac{P(R)}{1-P(I)} = \frac{0.4}{1-0.05} = 0.421$$

b)

Generelle forutsetninger for binomisk fordeling

i) Forsøksrekken består av n enkeltforsøk.

ii) Det registreres kun suksess eller ikke suksess.
iii) Sannsynligheten for suksess er lik i alle forsøk.
iv) Enkeltforsøkene er uavhengige.

For X har vi
i) Det er valgt ut n kamper.
ii) Vi registrerer kun om den som får første målet vinner(suksess) eller ikke.
iii) Sannsynligheten for suksess er p og er antatt å være konstant.
iv) Vi antar at kampene er uavhengige.
Dette er rimelige antakelser.

Sentralgrenseteoremet sier:
Dersom $Z_1, Z_2, \ldots, Z_n$ er uavhengig identisk fordelte fra sannsynlighetsfordelingen $f_Z(z)$, hvor $E(Z) = \mu$ og $Var(Z) = \sigma^2$, så vil $\sqrt{n}\frac{\overline{Z}-\mu}{\sigma}$ konvergere mot en normalfordeling med forventning 0 og varians 1. Der $\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$.
For en binomisk forsøksrekke, definer Z_i slik at: $Z_i = 1$ hvis suksess, og $Z_i = 0$ ellers. Med andre ord:

$$P(Z_i = z) = \begin{cases} p & \text{hvis } z = 1 \\ 1-p & \text{hvis } z = 0 \end{cases}$$

Slik at $E(Z_i) = p$ og $Var(Z_i) = p(1-p)$.

Siden enkeltforsøkene er uavhengige så er Z_i ene også uavhengige. Av sentralgrenseteoremet følger at $\sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}}$ konvergerer mot en normalfordeling med forventning 0 og varians 1. Der $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} Z_i$.
c)

H_0: $p \geq 0.8$ mot H_1: $p < 0.8$
Eventuelt: H_0: $p = 0.8$ mot H_1: $p < 0.8$

Vi ønsker å forkaste dersom $\hat{p} < k$, hvor k bestemmes slik at

$$P(\hat{p} < k) = \alpha = 0.05$$

Vi benytter at $Z = \sqrt{n}\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)}}$ er tilnærmet normalfordelt med forventing 0 og varians 1 under H_0. Da har vi fra ligningen over:

$$P(Z < \frac{\sqrt{n}(k-p_0)}{\sqrt{p_0(1-p_0)}}) = 0.05$$

$$\frac{\sqrt{n}(k-p_0)}{\sqrt{p_0(1-p_0)}} = -Z_{0.05}$$

Dette gir $k = p_0 - Z_{0.05}\sqrt{\frac{p_0(1-p_0)}{n}}$. Vi forkaster H_0 dersom:

$$\hat{p} < p_0 - Z_{0.05}\sqrt{\frac{p_0(1-p_0)}{n}} = 0.8 - 0.658\frac{1}{\sqrt{n}}$$

For $n = 24$ og $X = \sum_{i=1}^{n} Z_i = 17$ får vi $\hat{p} = 0.71$, $k = 0.67$. Vi forkaster ikke H_0.

Vi kan ikke pastå at ekspertkommentatoren tar feil på 5 prosent nivå.

d)

Vi ønsker at styrken på testen i alternativet $p = 0.7$ skal være minst 0.9. Dvs

$$P(\hat{p} < 0.8 - 0.658\frac{1}{\sqrt{n}}|p = 0.7) = 0.9$$

Vi benytter at $Z = \sqrt{n}\frac{\hat{p}-0.7}{\sqrt{0.7\cdot 0.3}}$ er tilnærmet normalfordelt med forventing 0 og varians 1 under alternativet med $p = 0.7$. Innsatt i kravet fra ligningen over gir dette:

$$P(Z < \sqrt{n}\frac{0.8 - 0.7}{\sqrt{0.7\cdot 0.3}} - \frac{0.658}{\sqrt{0.7\cdot 0.3}}) = 0.9$$

0.1 percentilen i normalfordelingen er lik $Z_{0.1} = 1.28$. Kravet som n må oppfylle blir dermed:

$$\sqrt{n}\frac{0.1}{\sqrt{0.7\cdot 0.3}} - \frac{0.658}{\sqrt{0.7\cdot 0.3}} = 1.28$$

Løsningen blir $n = 155.1$ kamper. Dvs at vi må se minst 156 kamper for å oppnå den ønskede styrken på testen.

Oppgave 3

a)

$$\begin{aligned}
P(\text{Bot}) &= P(\text{P.vakt ankommer før Katrine}) \\
&= P(T < 2) \\
&= \int_0^2 \frac{1}{5}\exp(-\frac{t}{5})\,dt \\
&= 1 - \exp(-\frac{2}{5}) \\
&\approx 0.33.
\end{aligned}$$

Definerer $K =$ kostnad ved parkeringsopphold. Får da

$$
\begin{aligned}
P(K=300) &= P(\text{Bot i løpet av } t \text{ timer}) \\
&= P(T<t) \\
&= \int_0^t \lambda \exp(-t\lambda) dt \\
&= 1-\exp(-\lambda t).
\end{aligned}
$$

$$
\begin{aligned}
P(K=0) &= P(\text{Ikke bot i løpet at } t \text{ timer}) \\
&= \exp(-\lambda t).
\end{aligned}
$$

$$
\begin{aligned}
E(K) &= 300(1-\exp(-\lambda t))+0 \\
&= 300(1-\exp(-\lambda t)).
\end{aligned}
$$

Forventet kostnad i løpet av 8 timer:

$$
\begin{aligned}
E(K_8) &= 300(1-\exp(-\frac{8}{5})) \\
&= 239.40\text{kr}.
\end{aligned}
$$

Ordinær pris: $8 \cdot 30$ kr $= 240$ kr.
Det lønner seg (økonomisk) å ikke betale.

b) Finner sannsynlighetsmaksimeringsfunksjonen

$$
\begin{aligned}
L(\lambda) &= f(t_1, \ldots, t_n; \lambda) \\
&= \prod_{i=1}^{n} \lambda \exp(-\lambda t_i),
\end{aligned}
$$

den naturlige logaritmen av sannsynlighetsmaksimeringsfunksjonen,

$$
\begin{aligned}
l(\lambda) &= \ln(L(\lambda)) \\
&= n \ln(\lambda) - \lambda \sum_{i=1}^{n} t_i
\end{aligned}
$$

og den deriverte av den naturlige logaritmen av sannsynlighetsmaksimeringsfunksjonen

$$
\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} t_i.
$$

Den dobbeltderiverte er

$$\frac{\partial^2 l}{\partial \lambda^2} = -\frac{n}{\lambda^2}.$$

Setter den deriverte lik null, og får da et uttrykk for sannsynlighetsmaksimeringsestimatoren

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}.$$

Denne vet vi er verdien som maksimerer λ, fordi den dobbeltderiverte alltid er mindre enn null her.

$$\begin{aligned} E(\hat{\lambda}) &= E(\frac{n}{\sum_{i=1}^{n} t_i}) \\ &= nE(\frac{1}{\sum_{i=1}^{n} t_i}) \\ &= \frac{n}{n-1}\lambda \end{aligned}$$

Vi ser da at estimatoren ikke er forventningsrett, og foreslår da

$$\tilde{\lambda} = \frac{n-1}{n}\hat{\lambda}.$$

$$\begin{aligned} E(\tilde{\lambda}) &= \frac{n-1}{n}E(\hat{\lambda}) \\ &= \frac{n-1}{n}\frac{n}{n-1}\lambda \\ &= \lambda \end{aligned}$$

$\tilde{\lambda}$ er nå forventningsrett, og kan estimeres:

$$\begin{aligned} \tilde{\lambda} &= \frac{19}{42.51} \\ &= 0.447. \end{aligned}$$

c)

Fra tabellen kan man lese: For T eksponensielt fordelt,

$$M_T(t) = \frac{1}{1-\frac{t}{\lambda}}.$$

For U χ^2-fordelt med *2n* frihetsgrader,

$$M_U(t) = \frac{1}{(1-2t)^n}.$$

Benytter nå standard regneregler for MGF:

$$\begin{aligned}
M_{\sum_{i=1}^n T_i}(t) &= \prod_{i=1}^n M_{T_i}(t) \\
&= \frac{1}{(1-\frac{t}{\lambda})^n}.
\end{aligned}$$

$$\begin{aligned}
M_{2\lambda\sum_{i=1}^n T_i}(t) &= M_{\sum_{i=1}^n T_i}(2\lambda t) \\
&= \frac{1}{(1-2t)^n}
\end{aligned}$$

Har nå vist at $V = 2\lambda\sum_{i=1}^n T_i$ har samme MGF som en χ^2-fordelt variabel med *2n* frihetsgrader og dermed er V χ^2-fordelt med *2n* frihetsgrader.

d)

Fra *c)* har vi følgende

$$P(\chi^2_{1-\frac{\alpha}{2}} < 2\lambda\sum_{i=1}^n T_i < \chi^2_{\frac{\alpha}{2}}) = 1-\alpha,$$

hvor $\chi^2_{1-\frac{\alpha}{2}}$ og $\chi^2_{\frac{\alpha}{2}}$ er kvantiler i en χ^2-fordelt med *2n* frihetsgrader. Har da

$$P(\frac{\chi^2_{1-\frac{\alpha}{2}}}{2\sum_{i=1}^n T_i} < \lambda < \frac{\chi^2_{\frac{\alpha}{2}}}{2\sum_{i=1}^n T_i}),$$

og vi får et $(1-\alpha)\cdot 100\%$ konfidensintervall for λ

$$[\frac{\chi^2_{1-\frac{\alpha}{2}}}{2\sum_{i=1}^n T_i}, \frac{\chi^2_{\frac{\alpha}{2}}}{2\sum_{i=1}^n T_i}] = [\lambda_L, \lambda_R].$$

For den aktuelle situasjonen

$$\begin{aligned}
\chi^2_{1-\frac{\alpha}{2},2n} &= \chi^2_{0.975,40} \\
&= 24.43, \\
\chi^2_{\frac{\alpha}{2},2n} &= \chi^2_{0.025,40} \\
&= 59.34.
\end{aligned}$$

Får da

$$\begin{aligned}\lambda_L &= 0.2874,\\ \lambda_R &= 0.6980.\end{aligned}$$

(studentene ble ikke bedt om å beregne disse).
Siden $300(1-\exp(-\lambda t))$ er monotont økende i λ, så blir $(1-\alpha)\cdot 100\%$ konfidensintervall for λ lik

$$[300(1-\exp(-\lambda_L t)), 300(1-\exp(-\lambda_R t))]$$

, hvor λ_L og λ_R er som tidligere. For $t = 8$ får vi intervallet

$$[269.90, 298.87].$$