Eksamen i STK1100 våren 2021 - løsningsforslag

Oppgave 1

a

For hver test, registrerer vi A (positiv) eller A' (negativ). $P(A) = p$ hver gang. Da har vi også $P(A') = 1 - p$. Fra definisjonen av X, er 1 minste mulige verdi (positiv på første test). Rimelig å anta at testene er uavhengige, slik at vi kan multiplisere sannsynligheter. Da kan vi skrive

$$\begin{aligned} P(X = k) =& P(A' \ k-1 \text{ ganger } \cap \ A \text{ i siste}) = (1-p)(1-p)\cdots(1-p)p \\ =& (1-p)^{k-1}p \quad \text{for } k = 1, 2, ... \end{aligned}$$

b

Her får vi bruk for hintet om konvergens av en geometrisk rekke,

$$\sum_{m=0}^{\infty} r^m = 1/(1-r) \text{hvis} \mid r \mid < 1.$$

Bruker definisjonen for momentgenererende funksjon, og setter inn punkt-sannsynligheten fra a:

$$\begin{aligned} M_X(t) =& \sum_{k=1}^{\infty} e^{tk}(1-p)^{k-1}p \\ =& e^t p \sum_{k=1}^{\infty} (e^t(1-p))^{k-1} \\ =& e^t p \sum_{k=0}^{\infty} (e^t(1-p))^{k} \\ =& \frac{e^t p}{1-(1-p)e^t} \end{aligned}$$

For siste overgang (hintet) må vi ha $(1-p)e^t < 1$. Da må vi ha

$$t < \ln(\frac{1}{1-p}).$$

c

Deriverer uttrykket for den momentgenererende funksjon med hensyn på t, og får

$$\begin{aligned} M_X'(t) =& \frac{pe^t}{1-(1-p)e^t} + \frac{(1-p)e^t pe^t}{(1-(1-p)e^t)^2} \\ =& \frac{pe^t}{1-(1-p)e^t}\left(1 + \frac{(1-p)e^t}{1-(1-p)e^t}\right) \\ =& \frac{pe^t}{(1-(1-p)e^t)^2}. \end{aligned}$$

Da finner vi forventningsverdien ved

$$\mathrm{E}(X) = M_X'(0) = \frac{p}{p^2} = \frac{1}{p}.$$

Tilsvarende finner vi

$$\begin{aligned} M_X''(t) =& \frac{pe^t}{(1-(1-p)e^t)^2} + \frac{2(1-p)e^t pe^t}{(1-(1-p)e^t)^3} \\ =& \frac{pe^t}{(1-(1-p)e^t)^2}\left(1 + \frac{2(1-p)e^t}{1-(1-p)e^t}\right) \\ =& \frac{pe^t}{(1-(1-p)e^t)^2}\left(\frac{1+(1-p)e^t}{1-(1-p)e^t}\right). \end{aligned}$$

Videre har vi at

$$\mathrm{E}(X^2) = M_X''(0) = \frac{p}{p^2} \cdot \frac{2-p}{p} = \frac{2-p}{p^2},$$

og endelig

$$\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Med $p = 0.001$ får vi en forventning $\mathrm{E}(X) = 1000$ tester til og med første positive.

d

Her kan vi bruke fra c at $\mathrm{E}(X_i) = 1/p$ og $\mathrm{V}(X_i) = (1-p)/p^2$. Da finner vi, for $Y = \sum_{i=1}^{r} X_i$, at

$$\mathrm{E}(Y) = \mathrm{E}(\sum_{i=1}^{r} X_i) = \sum_{i=1}^{r} \mathrm{E}(X_i) = \frac{r}{p}$$

og

$$\mathrm{V}(Y) = \mathrm{V}(\sum_{i=1}^{r} X_i) = \sum_{i=1}^{r} \mathrm{V}(X_i) = \frac{r(1-p)}{p^2}.$$

For variansen har vi brukt at X_i-ene er uavhengige. Med $r = 10$ og $p = 0.001$, får vi $\mathrm{E}(Y) = 10000$ tester inntil vi har 10 positive.

e

Merk at dette ikke er negativ binomisk fordeling slik den er definert i boken og formelsamlingen, fordi boken bruker antall negative tester frem til r-te positive, mens vi i denne oppgaven har definert Y =antall tester (både negative og positive) til og med r-te positive. Minste mulige verdi for Y blir da r (de r første testene er alle positive).

La A = testen er positiv. Da får vi:

$$\begin{aligned}
P(Y = k) =& P(A\ r-1 \text{ ganger i } k-1 \text{ tester } \cap\ A \text{ i } k\text{-te test})\\
=& P(A\ r-1 \text{ ganger i } k-1 \text{ tester}) \cdot P(A \text{ i } k\text{-te test})\\
=& \binom{k-1}{r-1} p^{r-1}(1-p)^{k-1-(r-1)} \cdot p\\
=& \binom{k-1}{r-1} p^{r}(1-p)^{k-r} \quad \text{for } k = r, r+1, ...
\end{aligned}$$

f

Har fra oppgaveteksten at

$$P(A|\text{ smittet}) = P(A|S) = 0.75$$

og

$$P(A|\text{ ikke smittet}) = P(A|S') = 0.0001,$$

der S = studenten er smittet. Har også at $P(A) = p = 0.001$. Skal finne $P(S)$. Bruker setningen om total sannsynlighet for A,

$$
\begin{aligned}
P(A) =& P(A|S)P(S) + P(A|S')P(S') \\
=& P(A|S)P(S) + P(A|S')(1 - P(S)),
\end{aligned}
$$

og løser for $P(S)$:

$$
P(S) = \frac{P(A) - P(A|S')}{P(A|S) - P(A|S')} = \frac{0.001 - 0.0001}{0.75 - 0.0001} = 0.0012.
$$

g

$$
P(S|A) = \frac{P(A|S)P(S)}{P(A)} = \frac{0.75 \cdot 0.0012}{0.001} = 0.90.
$$

Oppgave 2

a

For $x \leq 0$ er $F_X(x) = 0$ da $f_X(x) = 0$ for $x < 0$, og for $x > 0$ har vi

$$
F_X(x) = \int_{-\infty}^{x} f_X(z)dz = \int_0^x \frac{\alpha}{\beta^\alpha} z^{\alpha-1} e^{-\left(\frac{z}{\beta}\right)^\alpha} dz.
$$

La $w = \left(\frac{z}{\beta}\right)^\alpha$, slik at $z = \beta w^{1/\alpha}$ og $dz = \beta/\alpha w^{1/\alpha - 1} dw$. Vi får

$$
\begin{aligned}
F_X(x) &= \int_0^{\left(\frac{x}{\beta}\right)^\alpha} \frac{\alpha}{\beta^\alpha} (\beta w^{1/\alpha})^{\alpha-1} e^{-w} \frac{\beta}{\alpha} w^{1/\alpha - 1} dw \\
&= \int_0^{\left(\frac{x}{\beta}\right)^\alpha} e^{-w} dw = \left[-e^{-w}\right]_0^{\left(\frac{x}{\beta}\right)^\alpha} = 1 - e^{-\left(\frac{x}{\beta}\right)^\alpha}.
\end{aligned}
$$

Medianen $\tilde{\mu}$ finner vi ved

$$
\begin{aligned}
F_X(\tilde{\mu}) = 1 - e^{-\left(\frac{\tilde{\mu}}{\beta}\right)^\alpha} &= \frac{1}{2} \\
\longrightarrow \left(\frac{\tilde{\mu}}{\beta}\right)^\alpha &= -\ln\left(\frac{1}{2}\right) = \ln(2) \\
\longrightarrow \tilde{\mu} &= \beta \left(\ln(2)\right)^{1/\alpha}.
\end{aligned}
$$

b
Vi har:

$$\begin{aligned} F_Y(y) = \mathrm{P}(Y \leq y) =& \mathrm{P}\left(\left(\frac{X}{\beta}\right)^{\alpha} \leq y\right) \\ &= \mathrm{P}\left(X \leq \beta y^{\frac{1}{\alpha}}\right) \\ &= F_X\left(\beta y^{\frac{1}{\alpha}}\right) = 1 - e^{-\left(\frac{\beta y^{\frac{1}{\alpha}}}{\beta}\right)^{\alpha}} = 1 - e^{-y}. \end{aligned}$$

Vi kan allerede kjenne igjen $F_Y(y)$ som den kumulative fordelingene til eksponensialfordelingen med parameter $\lambda = 1$. Eventuelt kan vi finne tettheten til Y ved

$$f_Y(y) = \frac{d}{dy} F_Y(y) = e^{-y},$$

som er tettheten til eksponensialfordelingen med parameter $\lambda = 1$. Alternativt kan en løse oppgaven ved å bruke at $X = \beta Y^{\frac{1}{\alpha}} = v(Y)$ og formelen

$$\begin{aligned} f_Y(y) =& f_X(\beta y^{\frac{1}{\alpha}}) \cdot \left|\frac{d}{dy} v(y)\right| \\ =& \frac{\alpha}{\beta^{\alpha}} (\beta y^{\frac{1}{\alpha}})^{\alpha-1} e^{-\left(\frac{\beta y^{\frac{1}{\alpha}}}{\beta}\right)^{\alpha}} \cdot \frac{\beta}{\alpha} y^{\frac{1}{\alpha}-1} = e^{-y}. \end{aligned}$$

Oppgave 3

a
Vi har:

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^{\alpha}\right)^{\frac{1}{\alpha}} = \frac{1}{n^{1/\alpha}} \left(\sum_{i=1}^{n} X_i^{\alpha}\right)^{\frac{1}{\alpha}} = \frac{1}{n^{1/\alpha}} \left(\beta^{\alpha} \sum_{i=1}^{n} \frac{X_i^{\alpha}}{\beta^{\alpha}}\right)^{\frac{1}{\alpha}} = \frac{\beta}{n^{1/\alpha}} \left(\sum_{i=1}^{n} Y_i\right)^{\frac{1}{\alpha}}.$$

Da blir

$$n\left(\frac{\hat{\beta}}{\beta}\right)^{\alpha} = \sum_{i=1}^{n} Y_i.$$

Videre er $Y_1, \ldots, Y_n$ uavhengige da $X_1, \ldots, X_n$ er det, og fra Oppgave 2 b vet vi at $Y_i \sim \text{gamma}(1, 1)$. Da må $\sum_{i=1}^{n} Y_i = n(\hat{\beta}/\beta)^{\alpha} \sim \text{gamma}(n, 1)$.

b
Fra a vet vi at $n(\hat{\beta}/\beta)^{\alpha} \sim \text{gamma}(n, 1)$, slik at

$$\begin{aligned} &\mathrm{P}\left(a \leq n\left(\frac{\hat{\beta}}{\beta}\right)^{\alpha} \leq b\right) = 0.95 \\ \longrightarrow &\mathrm{P}\left(\left(\frac{a}{n}\right)^{1/\alpha} \leq \frac{\hat{\beta}}{\beta} \leq \left(\frac{b}{n}\right)^{1/\alpha}\right) = 0.95 \\ \longrightarrow &\mathrm{P}\left(\hat{\beta}\left(\frac{n}{b}\right)^{1/\alpha} \geq \beta \geq \hat{\beta}\left(\frac{n}{a}\right)^{1/\alpha}\right) = 0.95. \end{aligned}$$

Da er et 95% konfidensintervall for β gitt ved

$$\left(\hat{\beta}\left(\frac{n}{b}\right)^{1/\alpha}, \hat{\beta}\left(\frac{n}{a}\right)^{1/\alpha}\right),$$

der $\hat{\beta}$ er estimatet beregnet basert på observerte data.

c
Vi setter inn og får estimatet

$$\hat{\beta} = \left(\frac{1}{30}\sum_{i=1}^{30} x_i^{\alpha}\right)^{\frac{1}{\alpha}} = (4.08)^{\frac{1}{1.50}} \approx 2.55.$$

Konfidensintervallet blir da

$$\left(2.55\left(\frac{30}{41.65}\right)^{1/1.50}, 2.55\left(\frac{30}{20.24}\right)^{1/1.50}\right) \approx (2.05, 3.31).$$

d
Vi har:

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}\left(\frac{\beta}{n^{1/\alpha}}\left(\sum_{i=1}^{n} Y_i\right)^{1/\alpha}\right) = \frac{\beta}{n^{1/\alpha}}\mathrm{E}\left(\left(\sum_{i=1}^{n} Y_i\right)^{1/\alpha}\right).$$

Da $\sum_{i=1}^{n} Y_i \sim \text{gamma}(n, 1)$ og $r = 1/\alpha > -\alpha$, får vi fra resultatet gitt i oppgaven (om $\mathrm{E}(Z^r)$) at

$$\mathrm{E}(\hat{\beta}) = \frac{\beta}{n^{1/\alpha}} 1^{1/\alpha} \frac{\Gamma(n + 1/\alpha)}{\Gamma(n)} = \beta \frac{\Gamma(n + 1/\alpha)}{n^{1/\alpha}\Gamma(n)}.$$

For $\alpha = 1$ får vi

$$\mathrm{E}(\hat{\beta}) = \beta \frac{\Gamma(n+1)}{n\Gamma(n)} = \beta \frac{n\Gamma(n)}{n\Gamma(n)} = \beta.$$

Altså er $\hat{\beta}$ forventningsrett for β når $\alpha = 1$. Imidlertid er $\frac{\Gamma(n+1/\alpha)}{n^{1/\alpha}\Gamma(n)} \neq 1$ for $\alpha \neq 1$, og da er $\hat{\beta}$ ikke forventningsrett for β. Det er dog ikke så stor forskjell for store n.

e

En forventningsrett estimator for β får vi ved

$$\tilde{\beta} = \hat{\beta} \frac{n^{1/\alpha}\Gamma(n)}{\Gamma(n+1/\alpha)}$$

for da er

$$\mathrm{E}(\tilde{\beta}) = \mathrm{E}\left(\hat{\beta} \frac{n^{1/\alpha}\Gamma(n)}{\Gamma(n+1/\alpha)}\right) = \frac{n^{1/\alpha}\Gamma(n)}{\Gamma(n+1/\alpha)} \mathrm{E}(\hat{\beta}) = \frac{n^{1/\alpha}\Gamma(n)}{\Gamma(n+1/\alpha)} \beta \frac{\Gamma(n+1/\alpha)}{n^{1/\alpha}\Gamma(n)} = \beta.$$

Når vi setter inn observerte verdier og bruker opplysningen i oppgaven, får vi

$$\tilde{\beta} = 2.55 \frac{30^{1/1.50}\Gamma(30)}{\Gamma(30+1/1.50)} = 2.55 \frac{30^{1/1.50}}{9.62} \approx 2.56.$$

Vi ser at det er svært liten forskjell mellom $\hat{\beta}$ og $\tilde{\beta}$ her, og det skyldes at n er såpass stor at $\frac{\Gamma(n+1/\alpha)}{n^{1/\alpha}\Gamma(n)} \approx 1$.

f

Framgangsmåten for å estimere standardfeilene til $\hat{\beta}$ og β^* ved hjelp av parametrisk bootstrap er som følger.

For $b = 1, \ldots, B$

1. Trekk $x_1^*, \ldots, x_{30}^*$ fra Weibull$(\alpha, \hat{\beta})$
2. Beregn $\hat{\beta}_b^* = \left(\frac{1}{30}\sum_{i=1}^{30}(x_i^*)^\alpha\right)^{\frac{1}{\alpha}}$
3. Trekk $x_1^{**}, \ldots, x_{30}^{**}$ fra Weibull$(\alpha, \tilde{\beta})$
4. Beregn $\tilde{\beta}_b^* = \left(\frac{1}{30}\sum_{i=1}^{30}(x_i^{**})^\alpha\right)^{\frac{1}{\alpha}} \frac{\Gamma(30+1/\alpha)}{30^{1/\alpha}\Gamma(30)}$.

Estimér standardfeilene $\sigma_{\hat{\beta}}$ og $\sigma_{\tilde{\beta}}$ med

$$s_{\hat{\beta}} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\beta}_b^* - \bar{\hat{\beta}}^*)^2}$$

og

$$s_{\tilde{\beta}} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\tilde{\beta}_b^* - \bar{\tilde{\beta}}^*)^2},$$

der $\bar{\hat{\beta}}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_b^*$ og $\bar{\tilde{\beta}}^* = \frac{1}{B}\sum_{b=1}^{B}\tilde{\beta}_b^*$.

Vi ser at de to estimatorene har svært like standardfeil, og estimatene er også svært like. I dette tilfellet ar det altså svært liten effekt å korrigere for forventningsskjevheten, trolig fordi n er såpass stor.