UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1100 — Sannsynlighetsregning og statistisk modellering - Løsningsforslag
Eksamensdag:	Tirsdag 31. mai 2016.
Tid for eksamen:	09.00 – 13.00.
Oppgavesettet er på ?? sider.	
Vedlegg:	Ingen
Tillatte hjelpemidler:	Godkjent kalkulator og formelsamling for STK1100/STK1110

Kontroller at oppgavesettet er komplett
før du begynner å besvare spørsmålene.

Oppgave 1.

(a) Siden fordelingen er symmetrisk om μ, er $P(X \leq \mu) = P(X \geq \mu) = 0.5$ som viser at μ er medianen. Vi har at

$$R_X(t) = \ln(M_X(t)) = \mu t - \ln[1 - b^2t^2]$$

$$R'_X(t) = \mu - \frac{-2b^2t}{1 - b^2t^2} = \mu + \frac{2b^2t}{1 - b^2t^2}$$

$$E(X) = R'_X(0) = \mu$$

(b) Vi har

$$R''_X(t) = \frac{2b^2[1 - b^2t^2] - 2b^2t[-2b^2t]}{[1 - b^2t^2]^2}$$

$$V(X) = R''_X(0) = 2b^2$$

(c) Vi har fra formelsamling at $M_{Z_i}(t) = \lambda/(\lambda - t)$. Dermed er

$$M_Y(t) = E[e^{(\mu + Z_1 - Z_2)t}] = e^{\mu t}E[e^{Z_1 t}]E[e^{-Z_2 t}]$$

$$= e^{\mu t}\frac{\lambda}{\lambda - t}\frac{\lambda}{\lambda + t} = e^{\mu t}\frac{\lambda^2}{\lambda^2 - t^2} = \frac{e^{\mu t}}{1 - t^2/\lambda^2}$$

(Fortsettes side 2.)

som svarer til den momentgenererende funksjon til Laplace fordelingen med $b = 1/\lambda$.

(d) For den eksponensielle fordeling har vi at $F_Z(z) = 1 - e^{-\lambda z}$ som gir at man kan generere $Z_i = -\ln(1 - U_i)/\lambda$ der $U_i \sim \text{Uniform}[0,1]$. Ved å velge $\lambda = 1/b$ kan vi så sette $X = \mu + Z_1 - Z_2$.

Alternativt kan en beregne

$$F_X(t) = \begin{cases} \frac{1}{2}e^{(x-\mu)/b} & x < \mu; \\ 1 - \frac{1}{2}e^{-(x-\mu)/b} & x \geq \mu \end{cases}$$

som gir ved inversjonsmetoden

$$X = \begin{cases} \mu + b\ln(2 \cdot U) & U < 0.5 \\ \mu - b\ln(2(1-U)) & U \geq 0.5 \end{cases}$$

(e) Prosedyre for ikke-parametrisk Bootstrapping (for $\hat{\mu}$):

(a) Trekk $x_1^*, ..., x_n^*$ med tilbakelegging fra $\{x_1, ...x_n\}$

(b) Beregn $\hat{\mu}^*$ fra $x_1^*, ..., x_n^*$

Etter å ha gjentatt B ganger har man så Bootstrap sampler $\hat{\mu}_1^*, ..., \hat{\mu}_B^*$ og standard feil estimeres ved

$$\hat{\sigma}_{\hat{\mu}} = \sqrt{\tfrac{1}{B-1}\sum_{b=1}^{B}(\hat{\mu}_b^* - \overline{\hat{\mu}}^*)^2}$$

(f) Med $z_{0.025}$ lik den øvre 2.5% kvantilen i standard normal fordelingen har vi at

$$\begin{aligned} P(-z_{0.025} < \frac{\hat{\mu} - \mu}{\hat{\sigma}_{\hat{\mu}}} < z_{0.025}) &\approx 0.95 \\ P(-z_{0.025}\hat{\sigma}_{\hat{\mu}} < \hat{\mu} - \mu < z_{0.025})\hat{\sigma}_{\hat{\mu}} &\approx 0.95 \\ P(-\hat{\mu} - z_{0.025}\hat{\sigma}_{\hat{\mu}} < -\mu < -\hat{\mu} + z_{0.025})\hat{\sigma}_{\hat{\mu}} &\approx 0.95 \\ P(\hat{\mu} - z_{0.025}\hat{\sigma}_{\hat{\mu}} < \mu < \hat{\mu} - z_{0.025}\hat{\sigma}_{\hat{\mu}}) &\approx 0.95 \end{aligned}$$

som gir at $[\hat{\mu} - z_{0.025}\hat{\sigma}_{\hat{\mu}}, \hat{\mu} + z_{0.025}\hat{\sigma}_{\hat{\mu}}]$ er et tilnærmet 95% konfidensintervall for μ.

Tilsvarende beregninger kan gjøres basert på $\tilde{\mu}$ som gir $[\tilde{\mu} - z_{0.025}\hat{\sigma}_{\tilde{\mu}}, \tilde{\mu} + z_{0.025}\hat{\sigma}_{\tilde{\mu}}]$ er et tilnærmet 95% konfidensintervall for μ.

(g) Gjennomsnittet vil være tilnærmet normalfordelt ifølge sentralgrenseteoremet. Histogrammet av Bootstrap simuleringene indikerer også at

(Fortsettes side 3.)

normalfordelingen er rimelig. Vi vet at forventning er forventningsrett, noe også likheten mellom $\hat{\mu}$ og $\overline{\hat{\mu}}^*$ indikerer.

For medianen, så ser vi at siden $\overline{\tilde{\mu}}^* - \tilde{\mu} = 1.9155 - 1.895 = 0.0206$ er såpass liten, så kan vi anta at medianen er tilnærmet forventingsrett. Vi har ikke noe tilsvarende resultat om normalfordeling for medianen, men vi har generelt snakket om at mange estimatorer er tilnærmet normalfordelt. Fra histogrammet ser det dog ut som normaltilnærmingen kan være noe tvilsom. Vi bør derfor ikke stole for mye på konfidensintervallet basert på $\tilde{\mu}$.

Oppgave 2.

(a) La A være begivenheten at studenten svarer riktig. Da er $P(A) = r + (1-r)\cdot 0.2 = 0.6 + 0.4 \cdot 0.2 = 0.68$.

La B være begivenheten at kandidaten visste svaret. Bayes setning:

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} = \frac{r \cdot 1}{r + (1-r)\cdot 0.2} = \frac{0.6}{0.68} = 0.88$$

(b) Estimat: $\hat{p} = x/n = 2433/3570 = 0.682$

Estimert standard feil $\hat{\sigma}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/n} = 0.0080$

99% konfidensintervall:

$$\hat{p} \pm 2.576\hat{\sigma}_{\hat{p}} = [0.661, 0.701]$$

(c) Binært utfall ok. Samme sannsynlighet: Noe tvilsomt da vanskelighetsgraden på de ulike punkter er forskjellige. Uavhengighet: Tvilsomt da en flink student nok vil svare korrekt på mange oppgaver.

(d) Vi har at $q = 0.8r + 0.2$ slik at hvis $P(L < q < U) = 0.99$ så er $P(L < 0.8r + 0.2 < U) = 0.99$ og $P((L-0.2)/0.8 < r < (U-0.2)/0.8) = 0.99$. Dermed blir $[(0.185-0.2)/0.8, (0.117-0.2)/0.8] = [-0.018, 0.117]$. Dette kan vi eventuelt gjøre om til $[0.000, 0.117]$ da vi vet at r ikke kan bli negativ. Siden konfidensintervaller inkluderer 0 som en mulig verdi er det ikke noen grunn til å tro at studentene kunne denne oppgaven!

Oppgave 3.

(Fortsettes side 4.)

(a) Vi har

$$f_X(x) = \int_y f(x,y)dy = \int_0^1 (x+y)dy = [xy + 0.5y^2]_0^1 = x + 0.5$$

og pga symmetri blir tilsvarende $f_Y(y) = y + 0.5$.

(b) Vi har at

$$\begin{aligned}
P(Y \geq 0.5) =& \int_{0.5}^1 (y+0.5)dy \\
=& [0.5y^2 + 0.5y]_{0.5}^1 \\
=& 0.5 + 0.5 - 0.125 - 0.25 = 0.625 \\
P(X \geq 0.5, Y \geq 0.5) =& \int_{0.5}^1 \int_{0.5}^1 (x+y)dydx \\
=& \int_{0.5}^1 [xy + 0.5y^2]_{0.5}^1 dx \\
=& \int_{0.5}^1 [x + 0.5 - 0.5x - 0.125]dx \\
=& \int_{0.5}^1 [0.5x + 0.375]dx \\
=& [0.25x^2 + 0.375x]_{0.5}^1 \\
=& 0.25 + 0.375 - 0.0625 - 0.1875 = 0.375 \\
P(X \geq 0.5 | Y \geq 0.5) =& \frac{P(X \geq 0.5, Y \geq 0.5)}{P(Y \geq 0.5)}
\end{aligned}$$

$$= \frac{0.3125}{0.625} = 0.6$$

SLUTT