UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1100 — Sannsynlighetsregning og statistisk modellering -
Eksamensdag:	Friday 6. juni 2014.
Tid for eksamen:	14.30 – 18.30.
Oppgavesettet er på 6 sider.	
Vedlegg:	Ingen
Tillatte hjelpemidler:	Godkjent lommeregner og Formelsamling for STK1100 og STK1110

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

Oppgave 1

(a) Den kalles den Geometriske fordeling. Vi har at

$$\begin{aligned} E(X) =& \sum_{x=1}^{\infty} x(1-p)^{x-1}p \\ =& p\sum_{x=1}^{\infty} x(1-p)^{x-1} \\ =& p\tfrac{1}{(1-(1-p))^2} = \tfrac{1}{p} \end{aligned}$$

(b) Hvis vi ser på begivenheten $X > x$ så er det ekvivalent med å få feil i de x første forsøk. Sannsynligheten for dette er $(1-p)^x$ og dermed blir

$$F(x) = P(X \leq x) = 1 - P(X > x) = 1 - (1-p)^x$$

Alternativt har vi at

$$\begin{aligned} P(X \leq x) =& 1 - P(X > x) \\ =& 1 - \sum_{y=x+1}^{\infty} (1-p)^{y-1}p \\ =& 1 - (1-p)^x p \sum_{y=x+1}^{\infty} (1-p)^{y-1-x} \\ =& 1 - (1-p)^x p \sum_{y=0}^{\infty}(1-p)^y \\ =& 1 - \tfrac{(1-p)^x p}{p} \\ =& 1 - (1-p)^x \end{aligned}$$

(Fortsettes på side 2.)

(c) Vi har at y kan anta enhver verdi $1, 2, 3, ...$ Vi må ha at $x > y$.

Man kan her ta utgangspunkt i at den geometriske fordeling er basert på en serie av uavhengige forsøk. Hvis vi vet at $X > y$ så vet vi at det allerede har vært y feil. Dermed kan vi se på $X - y$ som da gitt $X > y$ svarer til antall forsøk etter de y første før suksess som også blir Geometrisk fordelt med forventing $1/p$ og varians $(1 - p)/p^2$. Videre svarer $X = x$ til at $X - y = x - y$ som gir at

$$P(X = x|X > y) = P(X - x = x - y|X > y) = \quad (1 - p)^{x-y-1}p$$

Alternativt:

$$\begin{aligned}
P(X = x|Y > y) =& \frac{P(X = x \cap X > y)}{P(X > y)} \\
=& \tfrac{(1-p)^{x-1}p}{(1-p)^y} \\
=& (1 - p)^{x-y-1}p
\end{aligned}$$

(d) Fra argumentasjonen i forrige oppgave får vi direkte at forventingen blir $y + 1/p$ mens variansen blir $(1 - p)/p^2$.

Mer direkte beregning:

$$\begin{aligned}
E(X|X > y) =& \sum_{x=y+1}^{\infty} x(1 - p)^{x-1-y}p \\
=& \sum_{x=1}^{\infty}(y + x)(1 - p)^{x-1}p \\
=& y\sum_{x=1}^{\infty}(1 - p)^{x-1}p + \sum_{x=1}^{\infty} x(1 - p)^{x-1}p \\
=& y + 1/p
\end{aligned}$$

Videre er

$$\begin{aligned}
E(X^2|X > y) =& \sum_{x=y+1}^{\infty} x^2(1 - p)^{x-1-y}p \\
=& \sum_{x=1}^{\infty}(y + x)^2(1 - p)^{x-1}p \\
=& y^2\sum_{x=1}^{\infty}(1 - p)^{x-1}p + 2y\sum_{x=1}^{\infty} x(1 - p)^{x-1}p + \sum_{x=1}^{\infty} x^2(1 - p)^{x-1}p \\
=& y^2 + 2y/p + E(X^2) \\
=& y^2 + y/p + (1 - p)/p^2 + 1/p^2
\end{aligned}$$

(Fortsettes på side 3.)

og

$$\begin{aligned}V(X|X>y) =&y^2+2y/p+(1-p)/p^2+1/p^2-(y+1/p)^2\\ =&y^2+2y/p+(1-p)/p^2+1/p^2-y^2-2y/p-1/p^2\\ =&(1-p)/p^2\end{aligned}$$

(e) Vi har at

$$\begin{aligned}E(e^{tX}) =&\sum_{x=1}^{\infty} e^{tx}(1-p)^{x-1}p\\ =&pe^t\sum_{x=1}^{\infty} e^{t(x-1)}(1-p)^{x-1}\\ =&pe^t\sum_{x=0}^{\infty}[e^t(1-p)]^x\\ =&\frac{pe^t}{1-e^t(1-p)}\end{aligned}$$

der den siste ulikheten er riktig hvis $e^t(1-p)<1$ som gir at $e^t<\frac{1}{1-p}$ og dermed $t<-\log(1-p)$. Siden $-\log(1-p)>0$ har vi at $E(e^{tX})$ eksisterer i en omegn av 0 og dermed vil den momentgenererende funksjon eksistere.

(f) Vi har at

$$\begin{aligned}M_{\overline{X}}(t) =&\prod_{i=1}^{n} M_X(\tfrac{1}{n}t)=[M_X(\tfrac{1}{n}t)]^n\\ =&\left[\frac{pe^{t/n}}{1-(1-p)e^{t/n}}\right]^n=\frac{p^ne^t}{[1-(1-p)e^{t/n}]^n}\end{aligned}$$

(g) Her kan en bli lurt til å tro en igjen må bruke den momentgenererende funksjon, men man kan bruke direkte at

$$E(\overline{X}) =E(\tfrac{1}{n}\sum_{i=1}^{n}X_i)=\tfrac{1}{n}\sum_{i=1}^{n}E(X_i)=\tfrac{1}{n}\sum_{i=1}^{n}\tfrac{1}{p}=\tfrac{1}{p}$$

der vi har brukt at forventningen til en lineærkombinasjon er lineærkombinasjonen av forventningene. Videre har vi at

$$V(\overline{X}) =V(\tfrac{1}{n}\sum_{i=1}^{n}X_i)=\tfrac{1}{n^2}V(\sum_{i=1}^{n}X_i)=\tfrac{1}{n^2}\sum_{i=1}^{n}V(X_i)=\tfrac{1}{n^2}\sum_{i=1}^{n}\tfrac{1-p}{p^2}=\tfrac{1-p}{np^2}$$

der vi først har brukt at variansen til en konstant multiplisert med en tilfeldig variabel er variansen til den tilfeldige variabelen multiplisert med konstanten kvadrert og deretter brukt at variansen til en sum av uavhengige variable er summen av variansene.

(Fortsettes på side 4.)

(h) Vi kan simulere $x_1, ..., x_n$ fra $p_X(x)$ og deretter beregne $\bar{x}^3$. Dette kan så gjentas M ganger. Vi kan så lage oss et MC estimat ved

$$\widehat{\theta} = \frac{1}{M}\sum_{m=1}^{M} \bar{x}_m^3$$

Dette estimatet vil være forventningsrett og ha en varians omvendt propsjonal med antall simuleringer M.

Oppgave 2

(a) For at $f_X(x)$ skal være en lovlig sannsynlighetstetthetsfunksjon så må

$$f_X(x) \geq 0 \text{ for alle } x, \text{ og } \int_{-\infty}^{\infty} f_X(x) = 1$$

Vi har at

$$\begin{aligned}
\int_{-\infty}^{\infty} f_X(x)dx =& \int_{-1}^{1} f_X(x)dx \\
=& 2\int_0^1 f_X(x)dx \qquad\qquad \text{pga symmetri} \\
=& 2\int_0^1 [a-x]dx \\
=& 2\int_0^1 [a-x]dx \\
=& 2[ax - 0.5x^2]_0^1 \\
=& 2[a - 0.5]
\end{aligned}$$

som gir at $a = 1$. Merk at da vil også $f_X(x) \geq 0$ for alle x.

(b) Vi har at

$$\begin{aligned}
E(X) =& \int_{-\infty}^{\infty} x f_X(x)dx \\
=& \int_{-1}^{0} x[1+x]dx + \int_0^1 x[1-x]dx \\
=& \int_{-1}^{0} [x+x^2]dx + \int_0^1 [x-x^2]dx \\
=& [\tfrac{1}{2}x^2 + \tfrac{1}{3}x^3]_{-1}^0 + [\tfrac{1}{2}x^2 - \tfrac{1}{3}x^3]_0^1 \\
=& -\tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{2} - \tfrac{1}{3} = 0
\end{aligned}$$

Alternativt kan en argumentere med at siden $f_X(x)$ er symmetrisk om 0, så har den også forventning 0.

(Fortsettes på side 5.)

(c) Ved å bruke transformasjonssetningen med $Y = g(X) = \mu + bX$ og dermed $X = h(Y) = \frac{Y-\mu}{b}$, har vi at

$$\begin{aligned} f_Y(y) =& f_X(\tfrac{y-\mu}{b})\tfrac{1}{b} \\ =& \begin{cases} \frac{1}{b}[1 + \frac{y-\mu}{b}] & \text{for } \mu - b \leq x \leq \mu \\ \frac{1}{b}[1 - \frac{y-\mu}{b}] & \text{for } \mu \leq x \leq \mu + b \end{cases} \end{aligned}$$

Alternativt har vi at

$$\begin{aligned} F_Y(y) =& P(Y \leq y) \\ =& P(\mu + bX \leq y) \\ =& P\left(X \leq b^{-1}(y-\mu)\right) = F_X(b^{-1}(y-\mu)) \end{aligned}$$

Vi har videre at

$$F_X(x) = \begin{cases} x + \frac{1}{2}x^2 & \text{for } -1 \leq x \leq 0 \\ \frac{1}{2} + x - \frac{1}{2}x^2 & \text{for } 0 \leq x \leq 1 \end{cases}$$

som gir

$$F_Y(y) = \begin{cases} b^{-1}(y-\mu) + \frac{1}{2}(b^{-1}(y-\mu)^2 & \text{for } \mu - b \leq y \leq \mu \\ \frac{1}{2} + b^{-1}(y-\mu) - \frac{1}{2}(b^{-1}(y-\mu)^2 & \text{for } \mu \leq x \leq \mu + b \end{cases}$$

som videre gir

$$f_Y(y) = F'_Y(y) = \begin{cases} b^{-1} + (b^{-1}(y-\mu) & \text{for } \mu - b \leq y \leq \mu \\ b^{-1} - (b^{-1}(y-\mu)^2 & \text{for } \mu \leq x \leq \mu + b \end{cases}$$

(d) Vi har at $E(Y) = \mu + bE(X) = \mu$.
Vi har at

$$V(Y) = V(\mu + bX) = b^2 V(X)$$

som gir at $k = V(X)$. Videre er

$$\begin{aligned} E(X^2) =& \int_{-\infty}^{\infty} x^2 f_X(x)dx \\ =& \int_{-1}^{0} x^2[1+x]dx + \int_{0}^{1} x^2[1-x]dx \\ =& \int_{-1}^{0} [x^2 + x^3]dx + \int_{0}^{1} [x^2 - x^3]dx \\ =& [\tfrac{1}{3}x^3 + \tfrac{1}{4}x^4]_{-1}^{0} + [\tfrac{1}{3}x^3 - \tfrac{1}{4}x^4]_{0}^{1} \\ =& \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{3} - \tfrac{1}{4} = \tfrac{8-6}{12} = \tfrac{1}{6} \end{aligned}$$

Dermed er $V(Y) = b^2 V(X) = \frac{b^2}{6}$.

(Fortsettes på side 6.)

(e) Vi har at

$$E(Y^2) = V(Y) + (E(Y))^2 = \tfrac{b^2}{6} + \mu^2$$

Vi finner da momentestimatorene ved å løse likningssystemet

$$\begin{aligned} \mu =& \tfrac{1}{n}\sum_{i=1}^{n} y_i \\ \tfrac{b^2}{6} + \mu^2 =& \tfrac{1}{n}\sum_{i=1}^{n} y_i^2 \end{aligned}$$

som gir

$$\begin{aligned} \hat{\mu} =& \tfrac{1}{n}\sum_{i=1}^{n} y_i \\ \hat{b}^2 =& 6[\tfrac{1}{n}\sum_{i=1}^{n} y_i^2 - \hat{\mu}^2] \end{aligned}$$

(f) Vi har at et tilnærmet 95% konfidensintervall for μ er gitt ved $[\hat{\mu} - z_{0.025}s_{\hat{\mu}}, \hat{\mu} + z_{0.025}s_{\hat{\mu}}]$ der

$$\begin{aligned} \sigma_{\hat{\mu}} =& \sqrt{k}b \\ s_{\hat{\mu}} =& \sqrt{k}\hat{b} = 4.32/6 = 0.72 \end{aligned}$$

Dermed blir 95% konfidensintervallet

$$[5.23 - 1.96 * 0.72/\sqrt{30}, 5.23 + 1.96 * 0.72/\sqrt{30}] = [4.97, 5.49]$$

Tilsvarende blir et 99% konfidensintervallet

$$[5.23 - 2.58 * 0.72/\sqrt{30}, 5.23 + 2.58 * 0.72/\sqrt{30}] = [4.89, 5.57]$$

Fortolkning: Hvis vi bruker slike intervaller gjentatte ganger, vil vi i tilnærmet 95% (99%) av tilfellene "treffe", dvs μ ligger innenfor intervallet.

Merk at det *ikke* er riktig at sannsynligheten for at μ ligger innenfor intervallet $[4.97, 5.49]$ er 95%.