Tentamentsskrivning i **Matematisk Statistik TMA321**

Tid: den 29 maj, 2017
Examinator och jour: Erik Broman, mob. 073 7320791,
Hjälpmedel: Typgodkänd miniräknare, egenhändigt skriven formelsamling om två A4 fram och bak (dvs 4 sidor), samt utdelade tabeller.

Tentamen består av 8 frågor om sammanlagt 50 poäng. Preliminära betygsgränser är satta till:
betyg "3": 20 till 29 poäng
betyg "4": 30 till 39 poäng
betyg "5": 40 eller fler poäng.

OBS! Alla lösningar skall vara väl redovisade och motiverade. Talen är ej ordnade efter svårighetsgrad.

1. Låt X vara en slumpvariabel med täthetsfunktion

$$f(x) = \begin{cases} c_1 x^2 & \text{om } 0 \le x < 1, \\ c_2\sqrt{x} & \text{om } 1 \le x \le 2. \end{cases}$$

(a) Bestäm konstanterna c_1, c_2 så att medianen för X blir 1. (2p)

(b) Beräkna väntevärdet av X. (2p)

(c) Hitta fördelningsfunktionen $F_X(x)$ för slumpvariabeln X. (3p)

Lösning:

(a) Vi har att

$$\frac{1}{2} = \int_0^1 c_1 x^2 dx = \frac{c_1}{3}$$

så att $c_1 = 3/2$ medans

$$\frac{1}{2} = \int_1^2 c_2\sqrt{x}dx = c_2\left[\frac{2x^{3/2}}{3}\right]_1^2 = \frac{2c_2}{3}(\sqrt{8}-1)$$

så att $c_2 = \frac{3}{4(\sqrt{8}-1)}$.

(b) Vi har att

$$\begin{aligned}\mathbb{E}[X] &= c_1\int_0^1 x^3 dx + c_2\int_1^2 x^{3/2}dx = \frac{c_1}{4} + c_2\left[\frac{2x^{5/2}}{5}\right]_1^2 \\ &= \frac{3}{8} + \frac{3}{4(\sqrt{8}-1)}\frac{2(\sqrt{32}-1)}{5} = \frac{3}{8} + \frac{3(\sqrt{32}-1)}{10(\sqrt{8}-1)}.\end{aligned}$$

(c) Vi har att för $x \in [0,1]$ gäller att

$$F(x) = \int_{-\infty}^{x} f(s)ds = \int_{0}^{x} c_1 s^2 ds = \frac{3}{2}\left[\frac{s^3}{3}\right]_0^x = \frac{x^3}{2}.$$

För $x \in [1,2]$ får vi då att

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(s)ds \\ &= \int_{0}^{1} f(s)ds + \int_{1}^{x} f(s)ds = \frac{1}{2} + c_2\left[\frac{2s^{3/2}}{3}\right]_1^x \\ &= \frac{1}{2} + \frac{3}{4(\sqrt{8}-1)}\frac{2x^{3/2}-2}{3} = \frac{1}{2} + \frac{x^{3/2}-1}{2(\sqrt{8}-1)} = \frac{x^{3/2}+\sqrt{8}-2}{2(\sqrt{8}-1)}. \end{aligned}$$

Dessutom gäller att $F(x) = 0$ om $x \leq 0$ och $F(x) = 1$ om $x \geq 2$.

2. Låt $X \sim U\{1,\dots,n\}$ och givet att $X = k$, låt $Y \sim$Poi(k).

(a) Hitta den gemensamma slf för X, Y och beräkna $\mathbb{E}[e^{X-Y}]$. (3p)

(b) Beräkna $\mathbb{E}[Y]$. (2p)

Lösning:

(a) Vi har att $\mathbb{P}(X = k) = \frac{1}{n}$ för $k = 1,\dots,n$ och att

$$\mathbb{P}(Y = l|X = k) = \frac{k^l}{l!}e^{-k} \text{ för } l = 0,1,2,\dots.$$

Vi får att

$$\mathbb{P}(X = k, Y = l) = \mathbb{P}(Y = l|X = k)\mathbb{P}(X = k) = \frac{k^l}{l!n}e^{-k}$$

för $l = 0,1,\dots$ och $k = 1,\dots,n$.
Därför blir

$$\begin{aligned} \mathbb{E}[e^{X-Y}] &= \sum_{k=1}^{n}\sum_{l=0}^{\infty} e^{k-l}\mathbb{P}(X = k, Y = l) \\ &= \sum_{k=1}^{n}\sum_{l=0}^{\infty} e^{k-l}\frac{k^l}{l!n}e^{-k} = \sum_{k=1}^{n}\frac{1}{n}\sum_{l=0}^{\infty}\frac{(k/e)^l}{l!} \\ &= \frac{1}{n}\sum_{k=1}^{n} e^{k/e} = \frac{e^{1/e}}{n}\sum_{k=0}^{n-1} e^{k/e} = \frac{e^{1/e}}{n}\frac{e^{n/e}-1}{e^{1/e}-1}. \end{aligned}$$

(b) Vi har att $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$ där vi vet att $\mathbb{E}[Y|X] = X$ ty $Y \sim$Poi(X). Därför blir $\mathbb{E}[Y] = \mathbb{E}[X] = \frac{n+1}{2}$.

3. Per och Lina har en kvadrat hemma hos sig. De vill skatta arean och mäter därför upp sidan på kvadraten två gånger. Dessa mätningar representeras av två oberoende slumpvariabler $S_1, S_2 \sim N(L, \sigma^2)$ där L är kvadratens sanna sidolängd.

   Per tycker att de skall skatta arean genom att ta medelvärdet av skattningarna för sidolängden och kvadrera, medans Lina tycker att det verkar bättre att helt sonika multiplicera de två skattade värdena.

   (a) Ange Pers respektive Linas skattningar av arean, och undersök väntevärdesriktigheten hos dessa. (3p)

   (b) För den eller de skattningar i uppgift (a) som är väntevärdesriktiga, beräkna MSE (i termer av σ^2 och L), dvs den förväntade kvadratiska avvikelsen från det sanna värdet på arean. (3p)

   Lösning:

   (a) Låt arean betecknas av A och låt

$$\hat{A}_P = \left(\frac{S_1 + S_2}{2}\right)^2 \text{ och } \hat{A}_L = S_1 S_2$$

   beteckna Pers respektive Linas skattningar av arean. Vi har då att

$$\mathbb{E}[\hat{A}_L] = \mathbb{E}[S_1]\mathbb{E}[S_2] = L^2 = A$$

   som därmed är VVR. Vidare gäller att $\mathbb{E}[S_1^2] = \mathrm{Var}(S_1) + \mathbb{E}[S_1]^2 = \sigma^2 + L^2$ så att

$$\begin{aligned}\mathbb{E}[\hat{A}_L] = \mathbb{E}\left[\frac{(S_1 + S_2)^2}{4}\right] &= \frac{\mathbb{E}[S_1^2 + S_2^2 + 2S_1S_2]}{4} \\ = \frac{\mathbb{E}[S_1^2] + L^2}{2} &= \frac{\sigma^2 + L^2 + L^2}{2} = L^2 + \sigma^2/2 \neq L^2.\end{aligned}$$

   Vi drar därför slutsatsen att Linas skattare är VVR medan Pers inte är det.

   (b) Vi skall beräkna $\mathbb{E}[(\hat{A}_L - L^2)^2]$. Vi får att

$$\begin{aligned}\mathbb{E}[(\hat{A}_L - L^2)^2] &= \mathbb{E}[(S_1S_2 - L^2)^2] \\ &= \mathbb{E}[S_1^2S_2^2 + L^4 - 2S_1S_2L^2] = \mathbb{E}[S_1^2]^2 + L^4 - 2L^4 \\ &= (\mathrm{Var}(S_1) + \mathbb{E}[S_1]^2)^2 - L^4 = (\sigma^2 + L^2)^2 - L^4 = \sigma^4 + 2\sigma^2L^2.\end{aligned}$$

4. Låt $X_1, \ldots, X_n$ vara i.i.d. med tfkn

$$f(x) = (\alpha/2 - 1)x^{-\alpha/2} \text{ för } x \geq 1, \text{ där } \alpha > 4.$$

   (a) Hitta MME (momentskattaren) för α. Vad händer om $\alpha \leq 2$? (3p)

   (b) Hitta MLE (maximum likelihood skattaren) för α. (3p)

Lösning:

(a) Vi har att

$$\mathbb{E}[X] = \int_1^\infty x(\alpha/2-1)x^{-\alpha/2}dx = (\alpha/2-1)\int_1^\infty x^{1-\alpha/2}dx$$
$$= (\alpha/2-1)\left[\frac{x^{2-\alpha/2}}{2-\alpha/2}\right]_1^\infty = \frac{\alpha/2-1}{\alpha/2-2} = \frac{\alpha-2}{\alpha-4}.$$

Detta gäller dock bara om $\alpha > 4$, annars blir $\mathbb{E}[X] = \infty$ och momentskattare saknas. Momentmetoden ger att vi skall lösa ekvationen

$$\bar{X} = \frac{\hat{\alpha}-2}{\hat{\alpha}-4} \quad \Leftrightarrow \quad \hat{\alpha}(\bar{X}-1) = 4\bar{X}-2$$

vilket ger

$$\hat{\alpha} = \frac{4\bar{X}-2}{\bar{X}-1}.$$

(b) Likelihooden är

$$f(X_1,\dots,X_n|\alpha) = \prod_{k=1}^n f(X_k|\alpha) = (\alpha/2-1)^n \left(\prod_{k=1}^n X_k\right)^{-\alpha/2} I(X_1,\dots,X_n \geq 1),$$

så att

$$l(\alpha) = n\log(\alpha/2-1) - \frac{\alpha}{2}\sum_{k=1}^n \log(X_k) + \log I(X_1,\dots,X_n \geq 1).$$

Vi ser att

$$l'(\alpha) = \frac{n/2}{\alpha/2-1} - \frac{1}{2}\sum_{k=1}^n \log(X_k) = \frac{n}{\alpha-2} - \frac{1}{2}\sum_{k=1}^n \log(X_k),$$

och

$$l''(\alpha) = -\frac{n}{(\alpha-2)^2}.$$

Villkoret $l'(\alpha) = 0$ ger

$$\hat{\alpha} = 2 + \frac{2n}{\sum_{k=1}^n \log(X_k)},$$

vilket vi från uttrycket för $l''(\alpha)$ ser är ett maximum.

5. (a) Visa att den momentgenererande funktionen för $Z \sim N(0,1)$ är $e^{t^2/2}$. (2p)

(b) Använd (a) för att hitta den momentgenererande funktionen för $X \sim N(\mu,\sigma^2)$. (2p)

(c) Härled den momentgenererande funktionen för $Z_1 + \cdots + Z_n$ om $Z_1, \dots, Z_n$ är i.i.d. $N(0,1)$. (1p)

(d) Låt $N \sim$Poi(λ) för något λ, och låt sedan

$$Y = \sum_{k=1}^{N} Z_k,$$

där $Z_1, Z_2, \dots$ är i.i.d. $N(0,1)$. Bestäm den momengenereande funktionen för Y. Jämför ditt svar med det i (c) och diskutera kort. (3p)

Lösning:

(a) Vi har att

$$\begin{aligned} M_Z(t) = \mathbb{E}[e^{Zt}] &= \int_{-\infty}^{\infty} e^{st} \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-(s-t)^2+t^2}{2}} ds = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-(s-t)^2}{2}} ds \\ &= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-s^2}{2}} ds = e^{t^2/2}. \end{aligned}$$

(b) Vi kan skriva $X = \mu + \sigma Z$ så att

$$M_X(t) = \mathbb{E}[e^{(\mu+\sigma Z)t}] = e^{\mu t} M_Z(\sigma t) = e^{\mu t + \sigma^2 t^2/2}.$$

(c) Vi får att

$$M_{Z_1 + \cdots Z_n}(t) = \prod_{k=1}^{n} M_{Z_k}(t) = e^{nt^2/2}.$$

(d) Vi har att

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{Yt}] = \mathbb{E}[\mathbb{E}[e^{Yt}|N]] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[e^{Yt}|N=n]\mathbb{P}(N=n) = \sum_{n=0}^{\infty} \mathbb{E}[e^{(Z_1+\cdots Z_n)t}]\mathbb{P}(N=n) \\ &= \sum_{n=0}^{\infty} e^{nt^2/2} \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\left(\lambda e^{t^2/2}\right)^n}{n!} = e^{-\lambda} e^{\lambda e^{t^2/2}} = e^{\lambda(e^{t^2/2}-1)}. \end{aligned}$$

Detta är INTE en mgf för en normalfördelad slumpvariabel. En deterministisk summa av oberoende $N(0,1)$ är alltså normalfördelad, medans en slumpmässig summa ej behöver vara det!

6. Ämnet polonium är mycket giftigt men dödar relativt långsamt. FSB vill använda det för att ha ihjäl regimkritiker som bor utomlands. De är intresserade av att hitta lämplig dos för att deras agenter skall hinna åka

hem väl i tid innan offret avlider. FSB gjorde en förstudie på 9 råttor och fick följande resultat:

dos (μg):	10	11	12	13	14	15	16	17	18
överlevnadstid (h):	188	189	180	181	177	183	176	175	168

Data kunde sammanfattas med $S_{xx} = 60$, $S_{yy} = 348$ och $S_{xy} = -128$.

Forskarna vid FSB ansatta en linjär regressionsmodell $y = \beta_0 + \beta_1 x$ där x är dosen och y överlevnadstiden i timmar. Hjälp dem med följande uppgifter:

(a) Skatta β_0, β_1. (2p)

(b) Skapa ett 99% konfidensintervall för β_1 och testa huruvida $\beta_1 = -2$ på 99%-nivån. (2p)

(c) Ange förklaringsgraden och beräkna residualerna. Verkar deras modell rimlig? (2p)

Lösning

(a) Vi har att

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \approx -2.133 \text{ och } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \approx 209.533$$

(b) Vi använder att

$$\frac{\hat{\beta}_1 - \beta_1}{s_r/\sqrt{S_{xx}}} \sim t(7)$$

och får med hjälp av tabell att $t_{0.005}(7) \approx 3.4995$ så att

$$\begin{aligned} 0.99 &= \mathbb{P}\left(-3.4995 \leq \frac{\hat{\beta}_1 - \beta_1}{s_r/\sqrt{S_{xx}}} \leq 3.4995\right) \\ &= \mathbb{P}\left(\hat{\beta}_1 - 3.4995\frac{s_r}{\sqrt{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + 3.4995\frac{s_r}{\sqrt{S_{xx}}}\right). \end{aligned}$$

Vi har att

$$s_r^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) \approx 10.7048$$

så att $s_r = 3.2718$. Ett 99% numeriskt K.I. för β_1 blir då

$$I_{\beta_1} = \hat{\beta}_1 \pm 3.4995\frac{s_r}{\sqrt{S_{xx}}} \approx [-3.6115, -0.6552].$$

Då $-2 \in I_{\beta_1}$ förkastar vi ej H_0 på 99% nivån.

(c) Förklaringsgraden blir

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \approx 0.7847$$

vilket kan betecknas som ok. Residualerna blir

e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9
-0.2000	2.933	-3.933	-0.800	-2.667	5.467	0.600	1.733	-3.133

vilka ser ok ut och inte tyder på något annat förhållande än linjärt. Större studier bör nog genomföras.

7. Låt $X_1, \ldots, X_n$ vara i.i.d. geometriskt fördelade slumpvariabler. Betrakta hypoteserna

$$H_0 : p = \frac{1}{2} \text{ och } H_1 : p = \frac{3}{4}.$$

(a) Sätt upp likelihood ration och härled ur denna enklaste möjliga test med tillhörande förkastningsområde. (3p)

(b) Om $n = 12$, bestäm förkastningsområde (RR) motsvarande signifikansnivå 1% för ditt härledda test. (3p)

(c) Vilken styrka får du för ditt test i (b)? (2p)

Lösning:

(a) Vi har att

$$\begin{aligned}\text{lik}(p) &= f(X_1, \ldots, X_n|p) \\ &= \prod_{k=1}^{n} f(X_k|p) = \prod_{k=1}^{n} (1-p)^{X_k - 1} p = \prod_{k=1}^{n} (1-p)^{X_k} \frac{p}{1-p},\end{aligned}$$

så den sökta LR kvoten blir

$$\begin{aligned}\frac{\text{lik}(1/2)}{\text{lik}(3/4)} &= \frac{\prod_{k=1}^{n}(1-1/2)^{X_k} \frac{1/2}{1-1/2}}{\prod_{k=1}^{n}(1-3/4)^{X_k} \frac{3/4}{1-3/4}} \\ &= \left(\frac{1}{2}\right)^{X_1+\cdots+X_n} 4^{X_1+\cdots+X_n} 3^{-n} = 2^{X_1+\cdots+X_n} 3^{-n}.\end{aligned}$$

Vi förkastar H_0 om denna kvot är liten, vilket är ekvivalent med att vi förkastar om $X_1 + \cdots + X_n$ är litet. Vårt RR blir nu $\{X_1 + \cdots X_n \le c_\alpha\}$ där konstanten c_α bestäms av signifikansnivån.

(b) Vi får räkna lite. Vi kan observera att summan av n oberoende geomteriskt fördelade slumpvariabler är negativt binomialfördelade. Låt $X = X_1 + \cdots + X_{12}$ så att

$$\mathbb{P}(X = k) = \binom{k-1}{12-1} \left(\frac{1}{2}\right)^1 2 \left(1 - \frac{1}{2}\right)^{k-12} = \binom{k-1}{11} 2^{-k}.$$

Vi har att

$$\mathbb{P}(X \leq 14) = \binom{11}{11} * 2^{-12} + \binom{12}{11} 2^{-13} + \binom{13}{11} 2^{-14} \approx 0.0065,$$

medans

$$\mathbb{P}(X \leq 15) = \mathbb{P}(X \leq 14) + \binom{14}{11} 2^{-15} \approx 0.0176.$$

Vårt RR blir därför $\{X \leq 14\}$.

(c) Styrkan ges av

$$\begin{aligned} 1 - \beta &= \mathbb{P}(\text{ förkasta } H_0 | H_1 \text{ sann }) = \mathbb{P}(X \leq 14 | X \sim \text{NegBin}(14, 3/4)) \\ &= \left(\frac{3}{4}\right)^{12} \left(1 - \frac{3}{4}\right)^0 + \binom{12}{11} \left(\frac{3}{4}\right)^{12} \left(1 - \frac{3}{4}\right)^1 + \binom{13}{11} \left(\frac{3}{4}\right)^{12} \left(1 - \frac{3}{4}\right)^2 \\ &\approx 0.2811. \end{aligned}$$

8. Trettio elever skall delas in i tre klasser, klass A,B och C, med tio elever i varje. Urvalet till klasserna är slumpmässigt.

(a) De tre vännerna Jenny, Penny och Lenny vill gärna hamna i samma klass. Vad är sannolikheten att exakt två av kompisarna hamnar i samma klass? (2p)

(b) Ingen av de tre vännerna gilla Kenny. På hur många sätt kan klasserna delas upp så att ingen av de tre hamnar med Kenny? (2p)

Lösning:

(a) Sannolikheten att Jenny och Penny hamnar i klass A medans Lenny inte hamnar där är

$$\frac{\binom{27}{8}}{\binom{30}{10}} \approx 0.0739.$$

Sannolikheten att Jenny och Penny hamnar i samma klass är därför

$$3\frac{\binom{27}{8}}{\binom{30}{10}} \approx 0.2217.$$

Det finns sedan 3 olika val av duo som skall väljas till samma klass så den sökta sannolikheten blir

$$9\frac{\binom{27}{8}}{\binom{30}{10}} \approx 0.665.$$

Alternativa och ekvivalenta uttryck kan t.ex. vara

$$\frac{6\binom{3}{2}\binom{27}{8}\binom{19}{9}}{\binom{30}{10,10,10}} = \frac{18\binom{27}{8}\frac{19!}{10!9!}}{\binom{30}{10}\frac{20!}{10!10!}} = 9\frac{\binom{27}{8}}{\binom{30}{10}}$$

som man kan komma fram till med ett liknande resonemang ($\binom{27}{8}$ för två i klass A, $\binom{19}{9}$ för den tredje i klass B, $\binom{3}{2}$ för att välja två av de tre, och 6 för antalet permutaioner av de tre klasserna).

(b) Om Kenny hamnar i klass A finns det

$$\binom{26}{9}\binom{20}{10}$$

sätt som klasserna kan fyllas upp på och som inte har J,P eller L i klass A. Detta gäller för alla de tre klasserna så vi får till slut

$$3\binom{26}{9}\binom{20}{10}\left(\quad = 3\frac{26!20!}{9!17!10!10!} = 3\frac{26!}{9!10!10!}20 * 19 * 18\right).$$

Ett något längre resonemang kan ge uttrycket

$$\begin{aligned}
6\left(\binom{26}{7}\binom{19}{9} + 3\binom{26}{8}\binom{18}{9}\right) &= 6\left(\frac{26!19!}{7!19!9!10!} + 2\frac{26!18!}{8!18!9!9!}\right) \\
&= 6\frac{26!}{9!10!10!}(8 * 9 * 10 + 3 * 9 * 10 * 10) = 3\frac{26!}{9!10!10!}(2 * 9 * 10 * 38) \\
&= \frac{26!}{9!10!10!}20 * 19 * 18 (\approx 1.731838079400000e + 12)
\end{aligned}$$

Detta resonemang bygger på att vi delar upp i fallen att alla JPL alla tre är i samma klass och att två av dem är det. Den "extra" koefficienten 3 i den andra termen kommer från det att de två kan väljas på $\binom{3}{2}$ olika sätt.