Tentamentsskrivning i **Matematisk statistik TMA321, 4.5 hp.**

Tid: Måndagen den 25 maj, 2009 kl 08:30-12:30
Examinator och jour: Erik Broman, tel. 772-3514, mob. 073 7320791, MV-huset rum L3080.
Hjälpmedel: valfri räknare, egen formelsamling (4 sidor på 2 blad A4) samt utdelade tabeller.

Tentamen består av 8 frågor om sammanlagt 30 poäng. Preliminära betygsgränser är satta till (inklusive bonuspoäng):
betyg "3": 12 till 17 poäng
betyg "4": 18 till 23 poäng
betyg "5": 24 eller fler poäng.

---

1. (4 poäng) Betrakta följande gemensamma pdf (sannolikhetsfördelning):

$$f(x,y) = k(x-y) \quad 0 \leq y \leq x \leq 1$$

för slumpvariablerna (X, Y).

a. Hitta k så att detta blir en gemensam pdf (sannolikhetsfördelning).

b. Bestäm den betingade fördelningen $f_{Y|X}(y|x)$.

c. Använd svaret i uppgift b. för att beräkna

$$\mathbb{E}[Y \Big| X = x].$$

d. Bestäm $\mathbb{P}(X > Y^2)$.

Lösning:
a. Vi har att

$$\int_0^1 \int_0^x k(x-y)dydx$$
$$= \int_0^1 k\left[\frac{(x-y)^2}{-2}\right]_0^x dx = \int_0^1 k\left[\frac{x^2}{2}\right]_0^x dx = \frac{k}{6},$$

så $k = 6$.

b. Vi kan utläsa ur räkningarna ovan att

$$f_X(x) = \frac{kx^2}{2} = 3x^2.$$

Därför blir

$$f_{Y|X}(y|x) = \frac{2(x-y)}{x^2}.$$

c. Vi får att

$$\mathbb{E}[Y \,\big|\, X = x] = \int_0^x y f_{Y|X}(y|x)dy = \frac{2}{x^2}\int_0^x y(x-y)dy = \frac{2}{x^2}\left[\frac{xy^2}{2} - \frac{y^3}{3}\right]_o^x = \frac{2}{x^2}\frac{x^3}{6} = \frac{x}{3}.$$

d. Vi har:

$$\begin{aligned}\mathbb{P}(X > Y^2) &= \int_0^1\int_0^x 1_{\{x>y^2\}}6(x-y)dydx \\ &= \int_0^1\int_0^{\sqrt{x}} 6(x-y)dydx = \int_0^1 6\left[\frac{(x-y)^2}{-2}\right]_0^{\sqrt{x}} dx \\ &= 3\int_0^1 x^2 - (x-\sqrt{x})^2 dx = 3\int_0^1 2x^{3/2} - xdx = 3\left[\frac{2x^{5/2}}{5/2} - \frac{x^2}{2}\right]_o^1 = \frac{9}{10}.\end{aligned}$$

2. (4 poäng) Låt $X_1, \ldots, X_n$ vara oberoende och likafördelade slumpvariabler med pmf

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}.$$

a. Hitta den momentgenererande funktionen för X_i.

b. Låt $S_n = (X_1 + \cdots + X_n)/\sqrt{n}$. Bestäm den momentgenererande funktionen för S_n.

c. Visa att $S_n \to Z$ där $Z \sim N(0, 1)$.

Lösning:

a. Vi har att

$$M_X(t) = \mathbb{E}[e^{tX}] = \frac{1}{2}(e^t + e^{-t}).$$

b. Vi har att

$$M_{S_n}(t) = \mathbb{E}[e^{tS_n}] = \prod_{i=1}^n \mathbb{E}[e^{tX_i/\sqrt{n}}] = \left(\frac{1}{2}(e^{t/\sqrt{n}} + e^{-t/\sqrt{n}})\right)^n.$$

c. Vi har att

$$\frac{1}{2}(e^{t/\sqrt{n}} + e^{-t/\sqrt{n}}) = \frac{1}{2}\left(\sum_{k=0}^{\infty}\frac{t^k}{k!n^{k/2}} - \sum_{k=0}^{\infty}\frac{(-t)^k}{k!n^{k/2}}\right) = 1 + \frac{t^2}{2n} + O\left(\frac{t^4}{n^2}\right).$$

Därför gäller att

$$M_{S_n}(t) = \left(1 + \frac{t^2}{2n} + O\left(\frac{t^4}{n^2}\right)\right)^n \to e^{t^2/2}.$$

3. (4 poäng) I ett försök att hitta guld togs elva stycken bergsprover från närliggande områden och halten guld uppmättes i dessa. Halterna antogs vara normalfördelade med okänt μ och σ. Resultatet blev som följer:

test nr:	1	2	3	4	5	6	7	8	9	10	11
resultat (ppm):	236	171	309	193	206	253	203	195	158	215	133

a. Skatta μ och σ.

b. Ge ett 99 % CI (konfidensintervall) för din skattning av μ.

Lösning:

a. μ skattas med hjälp av $\bar{X} = 206.5$ och σ skattas med hjälp av s^2, där

$$s^2 = \frac{1}{10}\sum_{i=1}^{11}(X_i - \bar{X})^2 \approx 2295.$$

b. Vi har att

$$T = \frac{\bar{X} - \mu}{s_{\bar{X}}} = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{10}.$$

11 prover är för få för att kunna använda normalapproximation. Därför blir vårt 99 % CI för μ:

$$[\bar{X} - \frac{s}{\sqrt{11}}t_{\alpha/2,10}, \bar{X} + \frac{s}{\sqrt{11}}t_{\alpha/2,10}]$$
$$\approx [206.5 - 14.45 * 3.169, 206.5 + 14.45 * 3.169] \approx [160.8, 252.3].$$

4. (4 poäng) Antag att $X_1, \ldots, X_n$ är oberoende mätdata från en fördelning med pdf

$$f(x) = \frac{\theta}{(1+x)^{\theta+1}}, \quad 0 < \theta < \infty, \quad 0 \le x < \infty.$$

a. Bestäm MME (method of moments estimate) skattaren för θ. För vilka θ är detta möjligt?

b. Bestäm MLE (maximum likelihood estimate) skattaren för θ.

Lösning:
a. Vi har att

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty \frac{\theta x}{(1+x)^{\theta+1}}dx \\ &= \left[-x(1+x)^{-\theta}\right]_0^\infty - \int_0^\infty -(1+x)^{-\theta}dx \\ &= 0 + \left[\frac{1}{1-\theta}(1+x)^{1-\theta}\right]_0^\infty = 0 - \frac{1}{1-\theta} = \frac{1}{\theta-1},\end{aligned}$$

om $\theta > 1$. MME ger då att $\bar{X} = 1/(\hat{\theta}-1)$ d.v.s. $\hat{\theta} = 1 + \frac{1}{\bar{X}}$.

b. Vi har att

$$\begin{aligned}L(\theta) &= f(X_1,\ldots,X_n\big|\theta) \\ &= \prod_{i=1}^n f(X_i\big|\theta) = \prod_{i=1}^n \frac{\theta}{(1+X_i)^{\theta+1}} = \frac{\theta^n}{\prod_{i=1}^n (1+X_i)^{\theta+1}} \\ &= \frac{\theta^n}{\prod_{i=1}^n e^{(\theta+1)\log(1+X_i)}} = \frac{\theta^n}{e^{(\theta+1)\sum_{i=1}^n \log(1+X_i)}} = \left(\frac{\theta}{e^{\frac{(\theta+1)}{n}\sum_{i=1}^n \log(1+X_i)}}\right)^n.\end{aligned}$$

För att maximera $L(\theta)$ räcker det att maximera

$$\frac{\theta}{e^{\frac{(\theta+1)}{n}\sum_{i=1}^n \log(1+X_i)}}.$$

Låt därför

$$g(\theta) := \theta e^{-(\theta+1)\alpha},$$

där $\alpha = \frac{1}{n}\sum_{i=1}^n \log(1+X_i)$. Vi har att

$$g'(\theta) = e^{-(\theta+1)\alpha} - \alpha\theta e^{-(\theta+1)\alpha} = 0,$$

ger villkoret $1 = \alpha\theta$. Vi ser att g bara har en extrempunkt och då g i denna antar ett positivt värde så måste detta vara ett maximum då

$$\lim_{\theta\to 0} \theta e^{-(\theta+1)\alpha} = \lim_{\theta\to\infty} \theta e^{-(\theta+1)\alpha} = 0,$$

då $\alpha > 0$. Vi får att MLE'n för θ blir

$$\hat{\theta} = \frac{1}{\alpha} = \frac{1}{\frac{1}{n}\sum_{i=1}^n \log(1+X_i)}.$$

5. (4 poäng) Liza som är en fårfarmare i Wales har N får som är svarta eller vita. Till den årliga "fårätarfestivalen" väljer hon ut n stycken får, alla med samma sannolikhet. Proportionen svarta får är p. Låt X vara antalet svarta får av dom n som hon väljer ut.

a. Vilken fördelning har X?

b. Ange sannolikheten för att $\mathbb{P}(X = k)$ och förklara varför den blir sådan.

Fåret Jöns-Harald ser vad som är på gång att hända och bestämmer sig för att rymma. Jöns-Harald försöker hoppa över ett stängsel, något som lyckas med sannolikhet $p > 0$. Dessutom måste Jöns-Harald hoppa över r stycken stängsel på sin väg till frihet.

c. Antag att Jöns-Harald fortsätter hoppa tills han lyckas rymma och antag att varje hopp är oberoende av dom föregående. Låt Y vara antalet hopp totalt innan han har rymt, vad är fördelningen för Y?

d. Ange sannolikheten för att $\mathbb{P}(Y = k)$ och förklara varför den blir sådan.

Lösning:

a. X är hypergeometriskt fördelad.

b.
$$\mathbb{P}(X = k) = \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}.$$
Här är $\binom{Np}{k}$ antal sätt att välja ut k svarta av totalt Np svarta, $\binom{N(1-p)}{n-k}$ antal sätt att välja ut $n - k$ vita av totalt $N(1 - p)$ vita och $\binom{N}{n}$ antal sätt att välja ut n stycken får av totalt N stycken sammanlagt. Vi har att
$$\mathbb{P}(X = k) = \frac{\#\text{ sätt att välja ut k svarta och n-k vita av N}}{\#\text{ sätt att välja ut n får av N}}.$$

c. Y är negativt binomialfördelad med parametrar r, p.

d. Varje sekvens av k försök varav r är lyckade (och den siste är en av dom) har sannolikhet $p^r(1-p)^{k-r}$. Anatalet sådan sekvenser är $\binom{k-1}{r-1}$, därför blir
$$\mathbb{P}(Y = k) = \binom{k-1}{r-1}p^r(1-p)^{k-r}.$$

6. (4 poäng) Tony och Tom singlar slant 50 gånger. Tony vet om att Tom har en fuskslant som gör att Tom vinner med sannolikhet 0.7 och misstänker att Tom använder sig av den.

a. Välj en lämplig nollhypotes och alternativ hypotes. Bilda likelihood ratio LR.

b. Antag att $RR = \{LR \leq c\}$, hur skall vi välja c så att vi förkastar vår nollhypotes på signifikansnivån $\alpha = 0.05$?

c. Antag att Tony bara tror att slanten är manipulerad men inte har någon aning om hur. Välj en lämplig nollhypotes och alternativ hypotes och bilda en genraliserad likelihood ratio.

Lösning:

a. Vi låter $H_0 : p = 1/2$ och $H_1 : p = 0.7$. Vi får att

$$LR = \frac{f(X|H_o)}{f(X|H_1)} = \frac{(1-0.5)^{50-X}(0.5)^X}{(1-0.7)^{50-X}(0.7)^X}$$

b.Villkoret $RR = \{LR \leq c\}$ ger

$$\frac{0.5^{50}}{(0.3)^{50-X}(0.7)^X} \leq c \quad \Leftrightarrow \quad \left(\frac{0.3}{0.7}\right)^X \leq c\left(\frac{0.3}{0.5}\right)^{50},$$

som i sin tur ger villkoret

$$X \geq \frac{\log c + \log\left(\frac{0.3}{0.5}\right)^{50}}{\log\left(\frac{0.3}{0.7}\right)}.$$

Vi använder oss av normalapproximation så att vi antar att $X \sim N(25, 25/2)$.

$$0.05 \approx \mathbb{P}\left(\frac{X-25}{\sqrt{25/2}} \geq 1.645\right),$$

vilket ger $X \geq 25 + 1.645 * \sqrt{25/2} \approx 30.8$, men då X är ett heltal får vi villkoret $X \geq 31$. För att få fram c måste vi då lösa

$$31 = \frac{\log c + \log\left(\frac{0.3}{0.5}\right)^{50}}{\log\left(\frac{0.3}{0.7}\right)},$$

som ger

$$\log c = 31 * \log\left(\frac{0.3}{0.7}\right) - \log\left(\frac{0.3}{0.5}\right)^{50},$$

så att

$$c = \left(\frac{0.3}{0.7}\right)^{31}\left(\frac{0.3}{0.5}\right)^{-50} \approx 0.484$$

c. Vi får

$$GLR = \frac{0.5^{50}}{(1 - X/n)^{n-X}(X/n)^X},$$

därför att $\hat{p} = X/n$ MLE'n för p.

7. (3 poäng) Några forskare mäter medeltemperaturen under en vintervecka vid nordpolen. Detta upprepas under 14 års tid och följande data mättes upp:

år:	77	78	79	80	81	82	83
medeltemperatur:	−30.06	−30.37	−29.33	−29.71	−29.39	−29.79	−29.66
år:	84	85	86	87	88	89	90
medeltemperatur:	−29.17	−29.16	−28.36	−29.11	−28.01	−28.20	−29.48

a. Plotta data i figur.

b. Bestäm regressionslinjen $y = b_0 + b_1 x$ och rita in den i samma figur.

c. Verkar sambandet linjärt? Är detta rimligt?
Lösning:

a.

b. Vi använder oss av:

$$b_1 = \frac{14\sum_{i=1}^{14} x_i y_i - \sum_{i=1}^{14} x_i \sum_{i=1}^{14} y_i}{\sum_{i=1}^{14} x_i^2 - \left(\sum_{i=1}^{14} x_i y_i\right)^2},$$

och

$$b_0 = \bar{Y} - b_1 \bar{X}.$$

Vi får resultatet $b_0 \approx -39.37$ och $b_1 \approx 0.121$.

c. Från figuren ser ett linjärt samband rimligt ut. Dock kan den skarpa trenden eventuellt ifrågasättas.

8. (3 poäng) Låt X vara likformigt fördelad på mängden $\{-3, -2, -1, 0, 1, 2, 3, \}$ och Y vara likformigt fördelad på $\{-2, 2\}$. Antag att X, Y är oberoende och sätt

$$U = X + Y$$

och

$$V = X - Y.$$

a. Beräkna covariansen mellan U, V.

b. Är U, V oberoende?

c. Antag att $U = a + \sum_{i=1}^{n} b_i X_i$ och att $V = c + \sum_{i=1}^{m} d_i Y_i$, där $X_1, \ldots, X_n, Y_1, \ldots Y_m$ alla är oberoende. Visa att

$$Cov(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i d_j Cov(X_i, Y_j)$$

Lösning:

a. Vi har att

$Cov(X + Y, X - Y) = Var(X) - Var(Y) = (3^2 + 2^2 + 1) * 2/7 - 4 = 0.$

b. Nej, t.ex. gäller att

$\mathbb{P}(U = 5, V = 1) = \mathbb{P}(X = 3, Y = 2) = \mathbb{P}(U = 5) \neq \mathbb{P}(U = 5)\mathbb{P}(V = 1).$

c. Se bok sid 139-140.