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

Tid: Fredagen den 21 augusti, 2009 kl 14:30-18:30
Examinator och jour: Erik Broman, tel. 772-3514, mob. 073 7320791.
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 pdf (sannolikhetsfördelning):

$$f(x) = c(x-a)^3$$

för slumpvariabeln X, där $a > 0$.

a. Hitta c och bestäm gränser så att detta blir en pdf (sannolikhetsfördelning).

b. Beräkna $\mathbb{E}[X]$.

c. Beräkna $Var[X]$.

d. Bestäm ett a så att $\mathbb{P}(X > 1) = \frac{1}{3}$.

Lösning: Observera att det finns många möjligheter här. Följande svar är ett av dom.
a. Vi väljer gränserna $a \leq x \leq 2a$, då blir

$$\int_a^{2a} c(x-a)^3 dx = c\left[\frac{(x-a)^4}{4}\right]_a^{2a} = c\frac{a^4}{4}$$

så $c = \frac{4}{a^4}$.

b. Vi får att

$$\begin{aligned}
\mathbb{E}[X] &= \int_a^{2a} cx(x-a)^3 dx \\
&= \left[cx\frac{(x-a)^4}{4}\right]_a^{2a} - \int_a^{2a} c\frac{(x-a)^4}{4} dx \\
&= c2a\frac{a^4}{4} - c\left[\frac{(x-a)^5}{20}\right]_a^{2a} \\
&= 2a - \frac{4}{a^4}\frac{a^5}{20} = 2a - \frac{a}{5} = \frac{9a}{5}.
\end{aligned}$$

c. Vi får att

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_a^{2a} cx^2(x-a)^3 dx \\
&= \left[cx^2\frac{(x-a)^4}{4}\right]_a^{2a} - \int_a^{2a} c2x\frac{(x-a)^4}{4}dx \\
&= c4a^2\frac{a^4}{4} - c\left(\left[2x\frac{(x-a)^5}{20}\right]_a^{2a} - \int_a^{2a} 2\frac{(x-a)^5}{20}dx\right) \\
&= 4a^2 - c4a\frac{a^5}{20} + c\left[2\frac{(x-a)^6}{120}\right]_a^{2a} \\
&= 4a^2 - 4a\frac{a}{5} + \frac{4}{a^4}\frac{a^6}{60} \\
&= \frac{a^2(60-12+1)}{15} = \frac{49a^2}{15}.
\end{aligned}$$

Således får vi

$$Var[X] = \frac{49a^2}{15} - \frac{81a^2}{25} = \frac{a^2(49*5-81*3)}{75} = \frac{492a^2}{75}$$

d. Självklart måste a väljas så att $a < 1 < 2a$. Vi har:

$$\begin{aligned}
\mathbb{P}(X > 1) &= \int_1^{2a} c(x-a)^3 dx \\
&= \frac{4}{a^4}\left[\frac{(x-a)^4}{4}\right]_1^{2a} = 1 - \frac{4}{a^4}\frac{(1-a)^4}{4} = \frac{a^4-(1-a)^4}{a^4}.
\end{aligned}$$

Därför måste vi ha att

$$\frac{2}{3}a^4 = (1-a)^4,$$

vilket ger

$$a\left(\left(\frac{2}{3}\right)^{1/4} + 1\right) = 1,$$

så att

$$a = \frac{1}{\left(\frac{2}{3}\right)^{1/4} + 1}.$$

2. (4 poäng) Låt X_n vara en slumpvariabel med pmf

$$\mathbb{P}(X_n = \frac{k}{n}) = \frac{2k+1}{n^2} \quad 0 \le k \le n-1.$$

a. Hitta den momentgenererande funktionen för X_n. Tips:

$$\sum_{k=0}^{n-1} kx^k = \frac{nx^n}{x-1} + \frac{1-x^n}{(1-x)^2}.$$

b. Visa att $X_n \to X$ och bestäm pdf:en för X.

Lösning:

a. Vi har att

$$\begin{aligned} M_{X_n}(t) &= \mathbb{E}[e^{tX_n}] \\ &= \sum_{k=0}^{n-1} \frac{2k+1}{n^2} e^{tk/n} = \frac{1}{n^2}\left(2\frac{ne^t}{e^{t/n}-1} + 2\frac{1-e^t}{(1-e^{t/n})^2} + \frac{1-e^t}{1-e^{t/n}}\right). \end{aligned}$$

b. Vidare gäller att

$$\frac{1-e^t}{n^2(1-e^{t/n})} = \frac{1-e^t}{-nt+o(1)} \to 0,$$

$$2\frac{ne^t}{n^2(e^{t/n}-1)} = 2\frac{e^t}{t+O(1/n)} \to 2\frac{e^t}{t},$$

samt

$$2\frac{1-e^t}{n^2(1-e^{t/n})^2} = 2\frac{1-e^t}{t^2+O(1/n^2)} \to 2\frac{1-e^t}{t^2}.$$

Därför gäller att

$$M_{X_n}(t) \to \frac{2}{t^2}(te^t - e^t + 1).$$

Vad är dåX? Uppenbarligen lever den på intervallet $[0,1]$ och måste växa linjärt. Då finns bara ett alternativ nämligen pdf:en $2x$. Mycket riktigt har vi

$$\begin{aligned} &\int_0^1 e^{tx} 2x dx \\ &\quad = [\frac{2x}{t}e^{tx}]_0^1 - \frac{2}{t}\int_0^1 e^{tx}dx = \frac{2e^t}{t} - \frac{2}{t^2}[e^{tx}]_0^1 \\ &\quad = \frac{2e^t}{t} - \frac{2}{t^2}(e^t - 1) = \frac{2}{t^2}(te^t - e^t + 1). \end{aligned}$$

3. (4 poäng) Taro och Markus surfar tillsammans. Dom väntar på vågor som dom kan surfa samtidigt och tiden dom lyckas stå på en våg mäts av en person på stranden. Resultatet blev som följer:

våg nr:	1	2	3	4	5	6	7	8	9	10	11	12	13
tid Taro:	16.4	7.6	23.6	4.8	8.9	16.	19.2	34.2	24.0	20.3	23.1	18.9	36.4
tid Markus:	15.1	7.4	22.4	4.4	7.1	17.8	19.6	33.2	24.7	19.8	19.8	20.0	35.5

a. Beräkna samplingsmedlet och samplingsvariansen av skillnaderna.

b. Sätt upp ett lämpligt hypotestest för att mäta hurivida Taro och Markus är lika bra på att stå länge eller om det är någon skillnad. Antag att dom uppmätta skillnaderna är normalfördelade.

c. Kan du förkasta din nollhypotes på 5%-nivån?

d. Uppskatta p-värdet.

Lösning:

a. Skillnaderna blir:

$$-1.3 \quad -0.2 \quad -1.2 \quad -0.4 \quad -1.8 \quad 1.8 \quad 0.4 \quad -1.0 \quad 0.7 \quad -0.5 \quad -3.3 \quad 1.1 \quad -0.9$$

μ skattas med hjälp av $\bar{X} = -0.577$ och σ skattas med hjälp av s^2, där

$$s^2 = \frac{1}{12}\sum_{i=1}^{13}(X_i - \bar{X})^2 \approx 1.7724.$$

b. Låt H_0 vara $\mu = 0$ och H_1 vara $\mu \neq 0$.

c. Vi har att

$$T = \frac{\bar{X} - \mu}{s_{\bar{X}}} = \frac{\bar{X}}{s/\sqrt{13}} \sim t_{12},$$

under H_0. Vidare har vi att med siffrorna ovan

$$T = \frac{\bar{X}}{s/\sqrt{13}} \approx -1.5627.$$

Vi utläser ur tabell att

$$\mathbb{P}(-2.160 \leq T \leq 2.160) = 0.95,$$

vi kan därför inte förkasta H_0 på 5%-nivån.

d. Vi utläser ur tabell att

$$\mathbb{P}(-1.5627 \leq T \leq 1.5627) \approx 0.85,$$

så p-värdet blir därför ca 0.15.

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

$$f(x) = 2\alpha x e^{-\alpha x^2}, \quad 0 \leq x < \infty.$$

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

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

Lösning:

a. Uppenbarligen måste $\alpha > 0$. Vi har att

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty 2\alpha x^2 e^{-\alpha x^2} dx \\ &= [x(-e^{-\alpha x^2})]_0^\infty + \int_0^\infty e^{-\alpha x^2} dx = \{\sqrt{\alpha} x = y\} \\ &= \frac{1}{\sqrt{\alpha}} \int_0^\infty e^{-y^2} dy.\end{aligned}$$

MME ger då att $\bar{X} = \frac{1}{\sqrt{\hat{\alpha}}} \int_0^\infty e^{-y^2} dy$ d.v.s. $\hat{\alpha} = \left(\frac{\int_0^\infty e^{-y^2} dy}{\bar{X}}\right)^2$.

b. Vi har att

$$\begin{aligned}L(\alpha) &= f(X_1, \ldots, X_n \big| \alpha) \\ &= \prod_{i=1}^n f(X_i \big| \alpha) = \prod_{i=1}^n 2\alpha X_i e^{-\alpha X_i^2} \\ &= (2\alpha)^n \prod_{i=1}^n X_i e^{-\alpha X_i^2}.\end{aligned}$$

Om vi logaritmerar så får vi:

$$l(\alpha) = \log L(\alpha) = n \log(2\alpha) + \sum_{i=1}^n (\log(X_i) - \alpha X_i^2).$$

Vidare blir då

$$l'(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^n X_i^2,$$

så att

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n X_i^2}.$$

Dessutom är

$$l''(\alpha) = \frac{-n}{\alpha^2} < 0,$$

så vi har hittat ett maximum.

5. (4 poäng) Den blåfenade tonfisken är utrotningshotad och därför mycket svårfångad. Under ett år genomfördes en undersökning där fiskare vid sammanlagt 971 tillfällen försökte fånga en blåfenad tonfisk. Låt X vara antalet tillfällen av dessa 971 där dom lyckades.

a. Vilken fördelning har X?

b. Approximera fördelning för X med en annan vanligt förekommande fördelning. Motivera ditt val med en lämplig beräkning!

c. Antag att $p = 0.01$ är sannolikheten att man lyckas fånga en tonfisk vid ett tillfälle. Vad är sannolikheten för att $X \leq 3$ vid användandet av dom två olika fördelningarna?

Lösning:

a. X är $\mathrm{Bin}(p, 971)$ fördelad.

b. X är Poisson(λ), där $\lambda = p971$. Detta kan motiveras av (t.ex.)

$$\begin{aligned}
&\mathbb{P}(X = k) \\
&= \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n!}{k!(n-k!)} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda},
\end{aligned}$$

för fixt λ. Högerledet är just sannolikheten att en Poisson(λ) antar värdet k.

c. Vi har att för $\mathrm{Bin}(0.01, 971)$ så blir

$$\begin{aligned}
&\mathbb{P}(X \leq 3) \\
&= \sum_{k=0}^{3} \binom{971}{k} 0.01^k (1 - 0.01)^{971-k} \approx 0.0125,
\end{aligned}$$

medan för Poisson(9.71)

$$\begin{aligned}
&\mathbb{P}(X \leq 3) \\
&= \sum_{k=0}^{3} \frac{9.71^k}{k!} e^{-9.71} \approx 0.128.
\end{aligned}$$

Alltså ganska så lika.

6. (4 poäng) Norge skall snart gå till val. En valundersökning genomförs för att se hur många som tänker rösta på dom olika partierna. Därför frågas 1000 personer om vilket parti dom tänker rösta för. I denna undersökning får bland andra Frp 28.1 % och V 4.9 %.

a. Ange 95-% CI (konfidensintervall) för proportionen som tänker rösta för Frp respektive V.

b. Antag att vi vill att vårt CI i båda fallen skall vara högst 0.5 % enheter brett. Ungefär hur många personer måste vi då minst fråga?

Lösning:

a. Vi kallar proportionerna p_f respektive p_v. Vi har då att $\hat{p_f} = 0.281$ och $\hat{p_v} = 0.049$. Dessutom blir da $s^2_{\hat{p_f}} = \hat{p_f}(1 - \hat{p_f})/n$ så att

$$s_{\hat{p_f}} \approx 0.0142, \quad s_{\hat{p_v}} \approx 0.0068.$$

Vi har att ett 95−% igt CI ges av

$$[\hat{p_f} - z_{0.975} s_{\hat{p_f}}, \hat{p_f} + z_{0.975} s_{\hat{p_f}}] \approx [0.281 - 1.96 * 0.0142, 0.281 + 1.96 * 0.0142] \approx [0.2531, 0.3089],$$

och på samma sätt blir

$$[\hat{p_v} - z_{0.975} s_{\hat{p_v}}, \hat{p_v} + z_{0.975} s_{\hat{p_v}}] \approx [0.0356, 0.0624].$$

b. Bredden på våra CI ges uppenbarligen av formeln

$$2 z_{0.975} s_{\hat{p_f}} = 2 * 1.96 * \sqrt{0.\hat{2}81(1 - 0.\hat{2}81)/n}.$$

Här använder vi oss av vår givna uppskattning av parametern p_f. Om detta skall vara högst 0.005 så måste vi välja

$$n_f \geq \frac{\hat{p_f}(1 - \hat{p_f})}{(0.005/(2 * 1.96))^2} = 124184.5,$$

dvs n måste minst vara 124185 st. På samma sätt fås att vi måste välja

$$n_v \geq \frac{\hat{p_v}(1 - \hat{p_v})}{(0.005/(2 * 1.96))^2} = 28642.3,$$

dvs 28643 st.

7. (3 poäng) Låt X vara Geom(p)-fördelad och låt Y vara summan av X oberoende normalfördelningar med parametrar μ, σ.

a. Beräkna $\mathbb{E}[Y]$

b. Beräkna $\mathbb{P}(Y \leq 1)$ (till en bra approximation) om $p = 0.2$, $\mu = 0$ och $\sigma = 1$.

Lösning:
a. Vi har att

$$\mathbb{E}[Y] = \sum_{k=1}^{\infty} \mathbb{E}[Y \Big| X = k]\mathbb{P}(X = k) = \sum_{k=1}^{\infty} k\mu\mathbb{P}(X = k) = \frac{\mu}{p}.$$

b. Vi har att

$$\mathbb{P}[Y \leq 1] = \sum_{k=0}^{\infty} \mathbb{P}[Y \leq 1 \Big| X = k]\mathbb{P}(X = k).$$

För att kunna lösa detta konstaterar vi att summan av k stycken oberoende $N(0,1)$ är $N(0,k)$ fördelad. Därför blir $\mathbb{P}[Y \leq 1 \Big| X = k] = \mathbb{P}[Y/\sqrt{k} \leq 1/\sqrt{k} \Big| X = k] = \mathbb{P}[Z \leq 1/\sqrt{k}]$, där $Z \sim N(0,1)$.

Dessutom konstaterar vi att

$$\sum_{k=0}^{3} 0.2^k * 0.8 \approx 0.9984,$$

så vi approximerar $\mathbb{P}[Y \leq 1]$ med

$$\begin{aligned}
&\sum_{k=0}^{2} \mathbb{P}[Y \leq 1 \Big| X = k]\mathbb{P}(X = k) \\
&\quad = \mathbb{P}(X = 0) + \sum_{k=1}^{3} \mathbb{P}[Z \leq 1/\sqrt{k}]\mathbb{P}(X = k) \\
&\quad = 0.8 + \mathbb{P}[Z \leq 1]0.2 * 0.8 + \mathbb{P}[Z \leq 1/\sqrt{2}]0.2^2 * 0.8 + \mathbb{P}[Z \leq 1/\sqrt{3}]0.2^3 * 0.8 \\
&\quad \approx 0.8 * (1 + 0.8413 * 0.2 + 0.7794 * 0.2^2 + 0.718 * 0.2^3) \approx 0.9641.
\end{aligned}$$

För att veta hur noggrant svaret är kan vi konstatera att

$$0.0008 = 1/2*(1-0.9984) \leq \sum_{k=4}^{\infty} \mathbb{P}[Y \leq 1 \Big| X = k]\mathbb{P}(X = k) \leq (1-0.9984) = 0.0016,$$

eftersom $1/2 \leq \mathbb{P}[Y \leq 1 \Big| X = k] \leq 1$.

8. (3 poäng) Låt (X, Y, Z) vara likformigt fördelade på området $\Omega = \{x^2 + y^2 + z^2 \leq 1, z \geq 0\}$.

a. Bestäm pdf:en $f(x, y, z)$ för (X, Y, Z).

b. Bestäm den betingade pdf:en $f(x \big| y, z)$

c. Bestäm $\mathbb{E}[X \big| Y = y, Z = z]$

Lösning:

a. Vi har att

$$f(x, y, z) = \frac{1}{V} = \frac{6}{4\pi},$$

b. Vi beräknar först marginalfördelningen,

$$f(y, z) = \int_{-\sqrt{1-y^2-z^2}}^{\sqrt{1-y^2-z^2}} \frac{1}{V} dx = \frac{6}{4\pi} 2\sqrt{1 - y^2 - z^2} = \frac{3\sqrt{1 - y^2 - z^2}}{\pi}.$$

Därför blir

$$f(x \big| y, z) = \frac{f(x, y, z)}{f(y, z)} = \frac{6}{4\pi} \frac{\pi}{3\sqrt{1 - y^2 - z^2}} = \frac{1}{2\sqrt{1 - y^2 - z^2}}.$$

Med gränserna $-\sqrt{1 - y^2 - z^2} \leq x \leq \sqrt{1 - y^2 - z^2}$ för $y^2 + z^2 \leq 1$ och $z \geq 0$.

c. Vi har

$$\mathbb{E}[X \big| Y = y, Z = z] = \int_{-\sqrt{1-y^2-z^2}}^{\sqrt{1-y^2-z^2}} x f(x \big| y, z) dx = 0,$$

pga symmetrier.