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

Tid: Måndagen den 26 maj, 2008 kl 14:00-18:00
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 en slumpvariabel X med följande pmf (sannolikhetsfördelning):

$$p_k = \mathbb{P}(X = k) = c|k-1|^\alpha \quad \text{för } k = 0, 1, 2, 3.$$

a. Bestäm ett samband mellan c och α så att detta blir en pmf (sannolikhetsfördelning).

b. Antag att $\mathbb{E}[X] = 2$, bestäm exakta värden på c och α.

c. Beräkna $Var[X]$ med hjälp av dessa värden.

1. Lösning:
a.
Om pmf har vi

$$1 = p_0 + p_1 + p_2 + p_3 = c + c + c2^\alpha = c(2 + 2^\alpha) \quad \Rightarrow \quad c = \frac{1}{2 + 2^\alpha}.$$

b. Vi har

$$2 = \mathbb{E}[X] = \sum_{k=0}^{3} kc|k-1|^\alpha = 2c + 3c2^\alpha = \frac{2 + 32^\alpha}{2 + 2^\alpha} \quad \Rightarrow \quad 2 = 2^\alpha,$$

så att $\alpha = 1$ och $c = 1/4$.

c. Vi har

$$\mathbb{E}[X^2] = \sum_{k=0}^{3} k^2 \frac{|k-1|}{4} = 1 + \frac{9 * 2}{4} = \frac{11}{2}.$$

Därför blir

$$Var[X] = \frac{11}{2} - 2^2 = \frac{3}{2}$$

2. (5 poäng) År 1944 var det pga kriget ransonering av livsmedel i sverige. Bagaren Kålle i Göteborg hade direktiv att baka frallor med en genomsnittlig vikt av 100 gram per fralla. Emil gick varje dag i två veckors tid och köpte en fralla (per dag) hos Kålle. Emil fattade misstankar om att Kålle fuskade på så sätt att han sparade lite av mjölet som skulle vara i frallorna för att sälja på svarta marknaden. Låt X_i vara vikten av frallan som Emil köpte dag $i \in \{1, \ldots, 14\}$. Vi antar att vikten är normalfördelad. Då Emil vägde alla frallorna han köpte fick han följande resultat:

dag:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
vikt i gram:	101	89	96	105	91	93	84	106	79	109	86	83	98	104

a. Sätt upp ett hypotestest för att testa Emils misstankar.

b. Låt signifikansnivån $\alpha = 0.02$. Använd normalapproximation för att testa hypoteserna du satte upp i uppgift a. Skall du förkasta nollhypotesen på den givna signifikansnivån?

c. Låt signifikansnivån igen vara $\alpha = 0.02$. Använd den exakta nollfördelningen för att testa hypoteserna du satte upp i uppgift a. Skall du förkasta nollhypotesen på den givna signifikansnivån?

d. Är det någon skillnad mellan b. och c. Är det viktigt vilken metod som används? Vilket är p-värdet av testen i b. och c.? Motivera.

2. Lösning:

a. Använd $H_0 : \mu = 100$ och $H_1 : \mu < 100$.

b. Vi använder att under H_0 och med normalapproximation är

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \approx N(0, 1).$$

Vidare blir $\bar{X} \approx 94.57$ och $s \approx 9.61$. Därför blir

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \approx -2.11$$

Tabell ger

$$\mathbb{P}(\frac{\bar{X} - \mu}{s/\sqrt{n}} \leq -2.11) \approx 1 - 0.983 = 0.017.$$

Vi kan därför förkasta H_0 på signifikansnivån 0.02.

c. Den exakta nollfördelningen är:

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{13}.$$

Vi får ur tabell:

$$\mathbb{P}(\frac{\bar{X}-\mu}{s/\sqrt{n}} \leq -2.11) \approx 1 - 0.975 = 0.025,$$

vi kan alltså ej förkasta H_0 på signifikansnivån 0.02.

d. p-värdena är 0.017 och 0.025 (cirka). Vi ser att med vald signifikansnivå är det en stor skillnad (vi tar ju olika beslut). Det är viktigt vilken vi använder, konsekvenserna av att göra ett typ 1 fel kan vara förödande för Kålle.

3. (3 poäng) Låt X_n vara likformigt fördelad på mängden $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$.

a. Bestäm den momentgenererande funktionen för X_n och skriv den på enklaste form.

b. Använd resultatet i a. för att visa att $X_n \to X$, då $n \to \infty$ där X är likformigt fördelat på intervallet $[0,1]$.

c. Beräkna n:te momentet av X med hjälp av den momentgenererande funktionen för X. Finns det ett enklare sätt?

3. Lösning:
a. Vi har att:

$$M_{X_n}(t) = \mathbb{E}[e^{tX_n}] = \frac{1}{n+1}\sum_{k=0}^{n} e^{tk/n} = \frac{1}{n+1}\frac{1-e^{t\frac{n+1}{n}}}{1-e^{\frac{t}{n}}}.$$

b. Mgf för X är:

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_0^1 e^{tx}dx = \frac{e^t-1}{t}.$$

Vi får:

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

c. Vi har (Taylor-utveckling):

$$M_X(t) = \frac{\sum_{k=0}^{\infty}\frac{t^k}{k!}-1}{t} = \sum_{k=1}^{\infty}\frac{t^{k-1}}{k!} = \sum_{k=0}^{\infty}\frac{t^k}{(k+1)!}.$$

Om vi deriverar n gånger och sätter $t=0$ får vi:

$$M_X^{(n)}(0) = \frac{n!}{(n+1)!} = \frac{1}{n+1}.$$

4. (4 poäng) Den så kallade Pareto-fördelningen används ofta i ekonomi för att t.ex. beskriva fördelningen av pengar i en population. Den har pdf:

$$f(x \big| x_0, \theta) = \theta x_0^\theta x^{-\theta-1}, \quad x \geq x_0, \quad \theta > 1.$$

Antag att oberoende samplingar $X_1, \ldots, X_n$ tas från denna fördelning.

a. Bestäm MME (method of moments estimate) skattaren för θ.

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

c. Antag att $x_0 = 1$. Låt $Y = \sqrt{X}$, bestäm pdf:en (sannolikhetsfördelningen) för Y.

4. Lösning:
a. Vi har:

$$\mathbb{E}[X] = \int_{x_0}^{\infty} \theta x_0^\theta x x^{-\theta-1} dx = \theta x_0^\theta \left[\frac{x^{-\theta+1}}{1-\theta} \right]_{x_0}^{\infty}$$
$$= \theta x_0^\theta \left(0 - \frac{x_0^{-\theta+1}}{1-\theta} \right) = \frac{\theta x_0}{\theta - 1}.$$

MME ger:

$$\bar{X} = \hat{\theta} \frac{x_0}{\hat{\theta} - 1},$$

ur vilket följer att

$$\hat{\theta} = \frac{\bar{X}}{\bar{X} - x_0}.$$

b. Vi har att

$$f(X_1, \ldots, X_n \big| x_0, \theta) = \theta^n x_0^{n\theta} \prod_{i=1}^{n} X_i^{-\theta-1} 1_{\{X_i \geq x_0\}}$$
$$= 1_{\{\min(X_1, \ldots, X_n\}) \geq x_0} e^{n \log \theta} e^{n\theta \log x_0} e^{-(1+\theta) \sum_{i=1}^n \log X_i}.$$

Om vi deriverar och sätter likamed 0, får vi:

$$-n = \theta(n \log x_0 - \sum_{i=1}^{n} \log X_i),$$

ur vilket följer att vår MLE blir:

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n \log X_i - n \log x_0}.$$

c. Låt $F_X(x)$ vara cdf:en för X:

$$F_X(x) = \int_1^x \theta y^{-\theta-1} dy = [-y^{-\theta}]_1^x = -x^{-\theta} + 1, \quad x \geq 1.$$

Vi får:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq y^2) = 1 - y^{-2\theta},$$

så att

$$f_Y(y) = F_Y'(y) = 2\theta y^{-2\theta-1}.$$

5. (4 poäng) En undersökning genomfördes på chalmers för att uppskatta hur många skor en genomsnittlig teknolog ägde. 25 slumpvist utvalda teknologer ombeddes därför svara på frågan. "Hur många skor äger du?." Svaren blev som följer:

5	2	7	4	12	7	7	5	
6	5	3	4	9	11	3	7	
3	14	8	9	10	4	8	11	5

a. Låt μ vara medelantalet skor som en teknolog äger. Skatta μ.

b. Ge ett 95 % approximativt CI (konfidensintervall) för din skattning. Vilka antaganden använder du dig av?

c. Förklara i ord betydelsen av intervallet från uppgift b.

5. Lösning: a. Vi använder $\bar{X}$ (medelvärdet) som skattare och får

$$\bar{X} = 6.76$$

b. Vi antar $\bar{X}$ normalfördelat. Ett 95 % approximativt CI blir då

$$\bar{X} \pm z_{0.025} s_{\bar{X}},$$

där $z_{0.025} = 1.96$ och $s_{\bar{X}} = s/\sqrt{25}$ där

$$S^2 = \frac{1}{24} \sum_{i=1}^{25} (X_i - \bar{X})^2 \approx 10.03.$$

Vi får att

$$6.76 \pm 1.241$$

är ett 95 % approximativt CI.

c. Tolkningen är att med 95% sannolikhet kommer vårt CI att innehålla det verkliga värdet på μ.

6. (4 poäng) Aktievärdet för ett litet läkemedelsföretag jämfördes mot branchindex (ett index baserat på värdet av all läkemedelsföretag på hela börsen) i en undersökning. Syftet var att se hur mycket av aktievärdet som berodde på hur branchen som helhet gick och hur mycket som berodde på faktorer som bara påverkade det enskilda företaget. Följande data registrerades vid slutet av varje månad (12 månader):

index (x)	1	1.03	1.09	1.02	0.95	0.92
aktievärdet (y)	78.10	81.85	84.35	78.21	78.70	74.68
index (x)	1.05	1.17	1.21	1.12	1.16	1.06
aktievärdet (y)	85.64	86.76	91.28	86.49	86.06	78.94

a. Plotta data i en figur.

b. Använd minsta kvadratmetoden för att bestämma regressionslinjen $y = b_0 + b_1 x$. Rita även in denna i din figur.

c. Skulle du säga att ett linjärt samband är rimligt? Förklara utifrån resultaten i a och b.

6. Lösning: a. Plotten ses i separat figur.

b. Vi använder oss av:

$$b_1 = \frac{12\sum_{i=1}^{12} x_i y_i - \sum_{i=1}^{12} x_i \sum_{i=1}^{12} y_i}{12\sum_{i=1}^{12} x_i^2 - (\sum_{i=1}^{12} x_i)^2}$$

och

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

Vi får att $b_1 \approx 50.4$ och $b_0 \approx 28.9$, för plotten se separat figur.

c. Från figuren ses att ett linjärt samband inte verkar vara helt orimligt. (Däremot kan man ifrågasatta modellens riktighet i just det här sammanhanget.)

7. (3 poäng) Antag att (X, Y) är likformigt fördelad på området A som ges av

$$A = [0, 2] \times [0, 1] \cup [1, 2] \times [1, 2].$$

a. Ange den gemensamma pdf:en (sannolikhetsfördelningen) av (X, Y).

b. Ange marginalfördelningarna for X och Y.

c. Ange dom betingade fördelningarna for X givet Y och vice versa.

7. Lösning:

a. Den gemensamma pdf:en ges av

$$f_{X,Y}(x,y) = \frac{1}{3} 1_{(x,y) \in A},$$

eller:

$$f_{X,Y}(x,y) = \begin{cases} 1/3 & \text{om } (x,y) \in A \\ 0 & \text{annars} \end{cases}$$

b. Vi har:

$$f_X(x) = \int_0^2 f_{X,Y}(x,y)dy = \begin{cases} 1/3 & \text{om } 0 \leq x \leq 1 \\ 2/3 & \text{om } 1 \leq x \leq 2. \end{cases}$$

På samma sätt:

$$f_Y(y) = \int_0^2 f_{X,Y}(x,y)dx = \begin{cases} 2/3 & \text{om } 0 \leq y \leq 1 \\ 1/3 & \text{om } 1 \leq y \leq 2. \end{cases}$$

c. Vi får här:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} 1/2 & \text{om } 0 \leq y \leq 1 & \text{för } 0 \leq x \leq 2 \\ 1 & \text{om } 1 \leq y \leq 2 & \text{för } 1 \leq x \leq 2. \end{cases}$$

På samma sätt:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} 1 & \text{om } 0 \leq x \leq 1 & \text{för } 0 \leq y \leq 1 \\ 1/2 & \text{om } 1 \leq x \leq 2 & \text{för } 0 \leq y \leq 2. \end{cases}$$

8. (3 poäng) I ett experiment får en fondförvaltare gissa vilken av tio aktier som kommer att gå bäst under en veckas tid. Detta upprepas 50 gånger under ett år.

a. Vi antar att antalet gånger som förvaltaren gissar rätt är binomialfördelat med parametrar $n = 50$ och något p. Vad måste vi anta om sambandet mellan förvaltarens gissningar för att detta antagande skall vara giltigt?

b. I slutet av året visade det sig att förvaltaren hade fått rätt 15 gånger. Ange skattningen $\hat{p}$ för p.

c. Ange skattningen för standardfelet av $\hat{p}$. Rättfärdiga formeln som du använder dig av.

8. Lösning:

a. Vi måste anta att gissningarna är oberoende och att alla har samma sannolikhet p att lyckas.

b. Vi har att

$$\hat{p} = 15/50.$$

c. Vi använder

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 0.0648.$$

Anledninger är att om $X \sim Bin(n,p)$ och X_i är i.i.d. $Bern(p)$

$$\sigma^2_{\hat{p}} = Var(\hat{p}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{p(1-p)}{n}.$$

Sedan ersätter vi p med det skattade värdet $\hat{p}$.