Tentamensskrivning i Matematisk statistik F1
Onsdagen den 16/1, 2002.

Jour: Urban Hjorth, tel 772 5362, examinator Urban Hjorth.

Hjälpmedel: Tabeller som utdelats i kursen. Kalkylator med tomt textminne. Varje väl löst uppgift ger 3 poäng och betygskraven blir 11, 15, 19 för betygen 3, 4, 5. Dessutom krävs minst 3 poäng på vardera delen.

Del 1:

1. Definiera varians och härled sedan variansen för $aX - bY + c$ om X och Y är oberoende med $\mu_x, \mu_y, \sigma_x, \sigma_y$. Ange var oberoendet utnyttjas.

2. Den tvådimensionella variabeln (X, Y) har frekvensfunktionen

$$f(x,y) = \frac{1+x+y+cxy}{c+3}e^{-(x+y)}, x > 0, y > 0.$$

a) Beräkna den marginella frekvensfunktionen för X.

b) Bestäm väntevärdet för $X + Y$.

3. Ur en mängd med 7 blåa, 4 röda och 3 gröna föremål drar man slumpvis och utan återläggning 4 stycken. Bestäm sannolikheterna att få 1, 2 eller 3 färger.

4. Vid tillverkning av ett material skapas ett moln av fibrer i olika längder vilka blandas om helt slumpmässigt i en turbulent ström innan de sugs ner på en vira. För att studera fiberlängdsfördelningen indelas fibrerna i längdklasser. Studera ett litet delområde på viran. Antag att materialet från detta område skall detaljstuderas. Låt $X_i(t)$ beteckna antalet fibrer i längdklassen i som landat i delområdet fram till tiden t (fiberuppsamlingen startas vid tiden 0).

a) Föreslå en lämplig modell för $X_i(t)$. Motivera hur du resonerar, eventuella oberoendeförutsättningar etc.

b) Är det rimligt att anta att $X_i(t)$ är oberoende för olika i? Vilken fördelning får i så fall $\sum_{i=1}^{K} X_i(t)$ (det totala antalet fibrer i längdklasserna 1 till K i det detaljstuderade området).

Vänd!!

Del 2:

5. Variablerna X_i är oberoende $N(\mu_1, \sigma)$ och Y_i är oberoende $N(\mu_2, \sigma)$ (samma standardavvikelse). Alla variablerna är oberoende. Data:

x_i: 12.5 11.7 13.0 12.3 12.4

y_j: 10.8 12.2 9.9 10.5 11.3 11.3 10.6

Skatta parametrarna μ_1, μ_2, σ så bra som möjligt.

6. På 100 minuter registreras 22687 radioaktiva sönderfall med en apparatur som anses registrera samtliga sönderfall. Gör ett 99% konfidensintervall för sönderfallsintensiteten, d.v.s. det förväntade antalet sönderfall per minut.

7. För att visa att ett preparat har medicinsk verkan läter man 150 personer få preparatet och 150 andra får ett verkningslöst s.k. placebo med samma utseende och smak. Personerna utses slumpmässigt till de båda grupperna och patienterna och läkaren som bedömer utvecklingen är ovetande om vilken medicinering som getts så att inget av medlen gynnas. Data från försöket (antal patienter som blivit):

	Bättre	Oförändrat	Sämre
Preparat	94	33	23
Placebo	69	55	26

Gör ett hypotestest med signifikansnivån 0.01 (approximativt) och försök visa att medicinen är verksam. Observera att det bara är positiv effekt av medicinen som är intressant vilket man bör utnyttja.

8. I ett gammalt industriområde vill man studera halten av bly och blyföreningar och tar därför n jordprover, $x_1, \ldots, x_n$, i olika delar av området som sedan analyseras på ett lab. Som modell prövas frekvensfunktionen

$$f(x; \theta) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}, \; x > 0,$$

med en enda parameter θ och proverna bedöms som oberoende eftersom provtagningsplatserna slumpats ut. Finn maximum likelihoodskattningen av parametern och studera dess väntevärde.

Mat stat F1 16/1 2002. Lösningar.

1. Se kursmaterial

2. $f_X(x) = \int_0^\infty \frac{1+x+y+cxy}{c+3} e^{-(x+y)}\,dy$;

a) $\int_0^\infty e^{-y}dy = 1$, $\int_0^\infty y e^{-y} dy = 1$ gör $\frac{1+x+1+cx}{c+3} e^{-x} = f_X(x)$, $x>0$.

b) $E[X] = \int_0^\infty x f_X(x)dx = \int \frac{x+x^2+x+cx^2}{c+3} e^{-x}dx$

$\int_0^\infty x^2 e^{-x}dx = \left[-x^2e^{-x}\right]_0^\infty + \int_0^\infty 2xe^{-x}dx = 2 \Rightarrow E[X] = \frac{1+2+1+2c}{c+3} = \frac{4+2c}{c+3}$

$EY = EX$ (symmetri). $E[X+Y] = E[X] + E[Y] = \frac{8+4c}{c+3}$

---

3. X = antal färger

$P(X=1) = P(A) + P(B)$; $A = \{4 \text{ blåa}\}$, $B = \{4 \text{ röda}\}$ är disjunkta

$= \frac{\binom{7}{4}}{\binom{14}{4}} + \frac{\binom{4}{4}}{\binom{14}{4}} = \frac{7\cdot6\cdot5\cdot4 + 4321}{14\cdot13\cdot12\cdot11} = \underline{\underline{0{,}036}}$

$P(X=3) = P(2 \text{ blå}, 1 \text{ röd}, 1 \text{ grön}) + P(\text{en blå}, 2 \text{ röda}, \text{en grön})$

$+ P(1 \text{ blå}, 1 \text{ röd}, 2 \text{ gröna}) = \frac{\binom{7}{2}\cdot4\cdot3 + 7\binom{4}{2}3 + 7\cdot4\binom{3}{2}}{\binom{14}{4}} =$

$= \frac{21\cdot4\cdot3 + 7\cdot6\cdot3 + 7\cdot4\cdot3}{1001} = \underline{\underline{0{,}4615}}$

$P(X=2) = 1 - 0{,}036 - 0{,}4615 = \underline{\underline{0{,}5025}}$

---

4. a) $X_i(t)$ bör vara en Poissonprocess med intensitet c_i

Det är ständigt samma intensitet att en fiber av längdklass i träffar området.

b) Oberoendet verkar rimligt eftersom fibrer av annan längd kan landa där oberoende av $X_i(t)$.

$\sum_1^k X_i(t)$ blir en summa av ober. Poissonvariabler och alltså Poissonfördelad $Po(\sum_1^k c_i t)$.

5. $\hat{\mu}_1 = \bar{x} = 12.38 \quad \hat{\mu}_2 = \bar{y} = 10.94\,; \quad s_x^2 = \frac{1}{n_x - 1}\sum(x_i - \bar{x})^2$

$$s^2 = \frac{4s_x^2 + 6s_y^2}{4+6} = 0.412\,; \quad \hat{\sigma} = s = \sqrt{0.412} = 0.642$$

---

6. X = antalet sönderfall på 100 minuter.

$X \in Po(100c) \approx N(100c, \sigma = \sqrt{100c})$

$\frac{X - 100c}{\sqrt{100c}}$ är $N(0,1)$; $\quad P\left(-2.58 < \frac{X - 100c}{\sqrt{100c}} < 2.58\right) = 0.99$

Om detta gäller för vårt observerade x så är

$$-2.58 < \frac{22687 - 100c}{\sqrt{100c}} < 2.58$$

Exakt lösn: $22687 - 385 < 100c < 22687 + 392$

$$223.0 < c < 230.8$$

Enkel approx $100c = 22687 \pm \underbrace{2.58\sqrt{22687}}_{388.6}\,; \quad 223.0 < c < 230.8$

---

7. Jämför antalet som blivit bättre.

X = antal med prep. som är bättre; $\quad X \in Bi(150, p_1)$

Y = placebo $\quad Y \in Bi(150, p_2)$

$H_0: p_1 = p_2\,; \quad H_1: p_1 > p_2.$

$X - Y$ är approx $N\left(150p_1 - 150p_2, \sqrt{150p_1(1-p_1) + 150p_2(1-p_2)}\right)$

$\frac{X - Y - 150(p_1 - p_2)}{\sqrt{\ldots}}$ är $N(0,1)$

H_0 ger att $T = \frac{X - Y - 0}{\sqrt{150\hat{p}_1(1-\hat{p}_1) + 150\hat{p}_2(1-\hat{p}_2)}}$ approx $N(0,1)$

$\hat{p}_i$ kan antingen skattas individuellt $\frac{94}{150}$ resp $\frac{69}{150}$ eller lika $\frac{163}{300}$

Individuella $\hat{p}_i$ ger $\quad T = \frac{94 - 69}{\sqrt{72.35}} = 2.94$

H_1: ensidig $\Rightarrow$ 1% risk i högra kanten $\Rightarrow$ Kritiskt område $\underline{\underline{T > 2.33}}$

$\underline{\underline{H_0 \text{ förkastas}}}$ på signifikansnivån 1%.

---

8. $\ln L = \sum \ln x_i - n\ln\theta - \frac{\sum x_i^2}{2\theta}\,;$

$\frac{\partial \ln L}{\partial \theta} = \frac{-n}{\theta} + \frac{\sum x_i^2}{2\theta^2} = 0$ ger $\hat{\theta} = \frac{\sum x_i^2}{2n}$ (maximum).

$E[\theta^*] = E\left[\frac{1}{2n}\sum X_i^2\right] = \frac{1}{2}E[X^2] = \frac{1}{2}\int_0^\infty x^2 \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}\,dx = \left| \begin{array}{l} x = y\sqrt{\theta} \\ dx = dy\cdot\sqrt{\theta} \end{array} \right|$

$= \frac{\theta}{2}\int_0^\infty y^3 e^{-y^2/2}\,dy = \frac{\theta}{2}\left\{\cancel{\left[-y^2 e^{-y^2/2}\right]_0^\infty} + \int_0^\infty 2y e^{-y^2/2}\,dy\right\} = \theta\left[-e^{-y^2/2}\right]_0^\infty = \underline{\underline{\theta}}$