Tentamensskrivning i Matematisk statistik F1
Lördagen den 20/5, 2000.

Jour: Mikael Knutsson, tel 7725380, examinator Urban Hjorth.
Hjälpmedel: Tabeller och fördelningstablå som utdelas vid skrivningen. Kalkylator med tomt textminne.
Varje väl löst uppgift ger 3 poäng och betygskraven blir 11, 15, 20 (19 utan bonus) för betygen 3, 4, 5. Dessutom krävs minst 3 poäng på vardera delen. Bonus från kontrollskrivningen adderas till poängen på del 1.

Del 1:

1. Vid tiden noll frigörs en partikel i en punkt som är placerad på en talaxel enligt en normalfördelning med $\mu = 1$, $\sigma = 2$. Därifrån rör den sig slumpmässigt på talaxeln så den oberoende av startpunkt vid tiden ett har förflyttat sig X från utgångspunkten, där X är normalfördelad med väntev'ärde 2 och standardavvikelse 3. Finn sannolikheten att partikeln ligger på den negativa sidan vid tiden 1.

2. En detalj inhandlas från två olika leverantörer. Leverantör I står för 2/3 av enheterna och leverantör II står för 1/3 och enheterna blir slumpmässigt omblandade innan de säljs till kunder. En detalj från I har 90% chans att hålla i minst 5 år och en detalj från II har 60% sannolikhet att hålla så länge. Den enhet vi handlat går sönder före fem år. Hur stor är den betingade sannolikheten att den är från I.

3. Ange förutsättningarna för att en variabel skall bli binomialfördelad. Visa hur frekvensfunktionen följer och härled på enklast möjliga sätt väntevärde och varians för en sådan $Bi(n,p)$-fördelad variabel.

4. Ett radioaktivt preparat har halveringstiden 27 timmar och vid tiden 0 har det intensiteten 8000 sönderfall per sekund.
a) Ange frekvensfunktionen för en radioaktiv atoms livslängd.
b) Hur många radioaktiva atomer finns det i preparatet vid tiden 0?
c) Ange fördelningen för antalet radioaktiva atomer som ännu inte sönderfallit efter en månad (30 dagar). Finn speciellt ett numeriskt värde på sannolikheten att ingen atom är kvar.

Del 2:

5. En variabel X kan anta värdena 0, 1, 2, 3, 4, 5. Symmetri och variabelns konstruktion ger att dess frekvensfunktion uppfyller villkoren $f(0) = f(5) = p$, $f(1) = f(4) = p^2$, $f(2) = f(3) = \frac{1}{2}(1 - 2p - 2p^2)$. Från 200 oberoende observationer får man följande resultat:

x	antal observationer
0	53
1	24
2	17
3	26
4	19
5	61

a) Skatta parametern p med maximum likelihoodmetoden.

b) Beräkna $\mu = \mu(p)$ och motivera varför momentmetoden inte kan användas på sin ursprungliga form.

6. Två nya material har framställts och för att jämföra deras dragstyrka så fogar man ihop dem så att fogen inte brister och sedan drar man tills det ena materialet går sönder. Vid 30 sådana provningar vann material A över material B i 21 fall och förlorade i 9 fall. Räcker detta för att visa att material A är starkare? Gör en modell och analysera med ett lämpligt konfidensintervall med 95% konfidensgrad.

7. I ett utmattningsförsök för metallstavar studerar man effekten av att polera materialets yta. Tanken är att poleringen eventuellt minskar risken för brott genom att sprickbildning inte så lätt startar vid ytan. Antalet belastningscykler till materialbrott avläses och har en mycket skev sannolikhetsfördelning. Därför studerar man istället logaritmerna av antalet belastningscykler som har ett betydligt mer normalfördelningsliknande utseende. Data (log antal cykler):

Utan polering: 9.00, 6.16, 10.28, 10.66, 7.36, 12.73, 12.73, 9.91, 10.75, 10.40, 9.57, 11.66, 8.64, 15.02, 9.68, 10.26, 12.45, 10.13, 9.78, 8.08, 10.67, 6.92, 11.64, 13.73, 8.40. Medelv. 10.26, skattad standardavvikelse 2.10.

Med polering: 14.35, 15.30, 8.47, 8.84, 13.67, 11.34, 13.95, 14.25, 14.00, 15.39, 13.90, 15.15, 9.41, 12.25, 11.92, 8.45, 12.91, 9.76, 15.69, 10.36. Medelv. 12.47, skattad standardavvikelse 2.48.

a) Vi skall testa hypotesen att poleringen inte har någon effekt mot den ensidiga hypotesen att polering ökar livslängden. Visa teorikunskaper genom skriva upp modellen, formulera hypoteserna i modellens termer och skriva ut fördelningen för alla led i räkningarna och ange det kritiska området C. Lika varians kan förutsättas. b) Genomför testet på signifikansnivån $\alpha = 0.01$. (a och b kan göras tillsammans).

8. Man har gjort regressionsanalys med modellen $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

a) Ange modellens förutsättningar om ε_i. Ge sedan modellen på vektorform och härled kovariansmatrisen för $\hat{\beta}$.

b) Gör ett konfidensintervall för lutningen β_1. Lämplig statistisk felrisk 0.01. Från en datorkörning får vi bl.a. följande resultat:

Datamaterial: $n = 20$ data. $\hat{\beta}_0 = 4.6761$, $\hat{\beta}_1 = 0.7249$, $s = 11.3926$, $X'X = \begin{pmatrix} 20 & 540 \\ 540 & 17538 \end{pmatrix}$, $(X'X)^{-1} = \begin{pmatrix} 0.2965 & -0.0091 \\ -0.0091 & 0.00034 \end{pmatrix}$.

Lösningar

Tenta i matematisk statistik för F1, 2000-05-20

UPPG 1: Y = startpunkt $\sim N(\mu_Y = 1, \sigma_Y^2 = 4)$, $X \sim N(\mu_X = 2, \sigma_X^2 = 9) \Rightarrow X + Y \sim N(\mu = 3, \sigma^2 = 13)$

$$P(\text{negativa sidan}) = P(X + Y < 0) = P\left(\frac{X+Y-3}{\sqrt{13}} < \frac{-3}{\sqrt{13}}\right) = \Phi(-0.83) = 0.2033$$

UPPG 2:
$$P(I \mid \text{sönder}) = \frac{P(I \cap \text{sönder})}{P(\text{sönder})} = \frac{P(\text{sönder} \mid I)P(I)}{P(\text{sönder} \mid I)P(I) + P(\text{sönder} \mid II)P(II)} =$$
$$= \frac{0.1 \cdot 2/3}{0.1 \cdot 2/3 + 0.4 \cdot 1/3} = \frac{1}{3}$$

UPPG 3: Om n oberoende försök utförs och var och ett lyckas med sannolikhet p, så är antalet lyckade $\sim Bin(n, p)$.

$$X \sim Bin(n,p) \Rightarrow f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \; x = 0, 1, \ldots, n,$$

ty det finns $\binom{n}{x}$ sätt att välja x lyckade ur n, var och ett av dessa x stycken lyckas med slh p och var och ett av övriga $n-x$ misslyckas med slh $(1-p)$.

Låt $Y_i = \begin{cases} 1, & \text{om försök } i \text{ lyckas} \\ 0, & \text{annars} \end{cases} \Rightarrow \begin{cases} E[Y_i] = 0 \cdot (1-p) + 1 \cdot p = p \\ E[Y_i^2] = 0^2 \cdot (1-p) + 1^2 \cdot p = p \end{cases}$

$$\Rightarrow Var(Y_i) = E[Y_i^2] - (E[Y_i])^2 = p - p^2 = p(1-p)$$

$$X = \sum_{i=1}^n Y_i \Rightarrow E[X] = E\left[\sum_{i=1}^n Y_i\right] = \sum_{i=1}^n E[Y_i] = \sum_{i=1}^n p = np$$

$$Var(X) = Var\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n Var(Y_i) = \sum_{i=1}^n p(1-p) = np(1-p)$$

UPPG 4: (a) Låt X = atomlivslängd (i sekunder)

$$f(x) = \lambda e^{-\lambda x}, \; x > 0 \Rightarrow F(x) = 1 - e^{-\lambda x}, \; P(X > x) = e^{-\lambda x}$$

Vet att $P(X > 27 \cdot 3600) = 0.5 \Leftrightarrow e^{-\lambda \cdot 27 \cdot 3600} = 0.5$

$$\Leftrightarrow -\lambda \cdot 27 \cdot 3600 = \ln 0.5 \Rightarrow \lambda = -\frac{\ln 0.5}{27 \cdot 3600} \approx 7.13 \cdot 10^{-6}$$

(b) Låt n = # atomer vid $t = 0$, Y = # sönderfall under första sekun[den]

$$\Rightarrow Y \sim Bin(n, p), \text{ där } p = P(X \le 1) = 1 - e^{-\lambda}$$

$$E[Y] = np = 8000 \Rightarrow n = \frac{8000}{p} = \frac{8000}{1 - e^{-\lambda}} = 1121844986$$

(c) Låt W = # atomer som ej sönderfallit efter en månad (30 da[gar])

$$\Rightarrow W \sim Bin(n = 1121844986, p),$$

$$\text{där } p = P(X > 30 \cdot 24 \cdot 3600) = e^{-2592000\lambda}$$

$$P(W = 0) = (1-p)^n \approx 2.67 \cdot 10^{-5}$$

UPPG5: (a) $L(p) = p^{53+61}(p^2)^{24+19}\left(\frac{1}{2}(1-2p-2p^2)\right)^{17+26} = p^{200}\cdot\left(\frac{1}{2}\right)^{43}(1-2p-2p^2)$

$\ln L(p) = 200\ln p + 43\ln 0.5 + 43\ln(1-2p-2p^2)$

$\frac{\partial \ln L(p)}{\partial p} = 0 \Leftrightarrow \frac{200}{p} - \frac{86(1+2p)}{1-2p-2p^2} = 0 \Leftrightarrow 200(1-2p-2p^2) - 86p(1+2p) = 0$

$\Leftrightarrow 592p^2 + 486p - 200 = 0 \Rightarrow \hat{p} = -\frac{243}{592} + \sqrt{\left(\frac{243}{592}\right)^2 + \frac{200}{592}} \approx 0.30$

(b) $\mu = \mu(p) = \sum_{x=0}^{5} x f(x) = 0\cdot p + 1\cdot p^2 + 2\cdot\frac{1}{2}(1-2p-2p^2) + 3\cdot\frac{1}{2}(1-2p-2p^2)$
$+ 4\cdot p^2 + 5\cdot p = p^2 + 1 - 2p - 2p^2 + 1.5 - 3p - 3p^2 + 4p^2 + 5p = 2.5$

Momentmetoden skattar $E[X]$ med $\bar{X}$, dvs 2.5 med 2.585 ?

UPPG6: Låt X = # "A-vinster" $\Rightarrow X \sim Bin(n=30, p = P(\text{"A-vinst"}))$

Om A och B lika starka är $p = 0.5$, Om A starkare är $p > 0.5$.

$X \overset{appr.}{\sim} N(\mu = np, \sigma^2 = np(1-p)) \Rightarrow \hat{p} = \frac{X}{n} \sim N\left(\mu = p, \sigma^2 = \frac{p(1-p)}{n}\right)$

$\Rightarrow P\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq 1.645\right) = 0.95 \Leftrightarrow P\left(p \geq \hat{p} - 1.645\sqrt{\frac{p(1-p)}{n}}\right) = 0.95$

$\Rightarrow \left(\hat{p} - 1.645\sqrt{\frac{p(1-p)}{n}}, 1\right) \approx \left(\hat{p} - 1.645\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, 1\right)$ är ett nedåt begränsat 95% k.i. för p. Vi får intervallet

$\left(\frac{21}{30} - 1.645\sqrt{\frac{21/30\cdot 9/30}{30}}, 1\right) = (0.56, 1)$, vilket tyder på att A är starkare.

UPPG7: $X_1 \sim N(\mu_1, \sigma^2)$, $X_2 \sim N(\mu_2, \sigma^2) \Rightarrow \bar{X}_1 \sim N\left(\mu_1, \frac{\sigma^2}{25}\right)$, $\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma^2}{20}\right)$

$\Rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{25} + \frac{1}{20}\right)\right) \Rightarrow \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{1/25 + 1/20}} \sim N(0,1) \Rightarrow \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{25}+\frac{1}{20}}} \sim t_{43}$

Testa $H_0: \mu_1 = \mu_2$ mot $H_1: \mu_1 < \mu_2$. Om H_0 sann är $T = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{1/25 + 1/20}} \sim t_{23}$. Förkasta på α-nivå om $T < -t_{\alpha, 23}$, dvs på 0.01-nivå om $T < -2.416$. Vi får $S_p^2 = \frac{24\cdot 2.1^2 + 19\cdot 2.48^2}{43}$ och

$t = -3.24 \Rightarrow$ Förk. H_0 på 0.01-nivå.

UPPG8: (a) ε_i oberoende $\sim N(0, \sigma^2)$. $Y = X\beta + \varepsilon$, där

$Y = [Y_1, \ldots, Y_n]^T$, $\beta = [\beta_0, \beta_1]^T$, $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_n]^T$, $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$. $\hat{\beta} = (X^TX)^{-1}X^TY$

$Var(\hat{\beta}) = (X^TX)^{-1}X^T Var(Y)((X^TX)^{-1}X^T)^T = (X^TX)^{-1}X^TX(X^TX)^{-1}\sigma^2 = (X^TX)^{-1}\sigma^2$

(b) 99% k.i. för β_1:

$\hat{\beta}_1 \pm t_{0.005,18}\cdot S\sqrt{0.00034} = 0.7249 \pm 2.878\cdot 11.3926\sqrt{0.00034} =$

$= 0.7249 \pm 0.6046 \quad (0.1203, 1.3295)$