MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 24 oktober 2007 em M

Tillåtna hjälpmedel är räknedosa utan lagrad information om kursen, Beta, kursens formel- och tabellsamling.

Examinator är Tommy Norberg, ankn 3528 eller 0730 79 42 09.

Jour är Sofia Tapani, ankn 5336.

Maximalt antal tentamenspoäng är 30, av dessa krävs normalt 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar till tentamensproblemen går att ladda ner från kurshemsidan. Rättningsprotokoll anslås ej.

Svar och lösningar skall motiveras om ej annat sägs i uppgiften.

Uppgifter

1. Du drar två kort från en väl blandad kortlek.

   (a) Visa att sannolikheten att det andra kortet är ett ess är $1/13$ om du inte har sett det första.

   (b) Om du däremot har sett att det första kortet är ett ess, vad är då sannolikheten att det andra kortet är ett ess?

   (c) Hur stor är sannolikheten att exakt ett av korten är ett ess?

   (d) Givet att det andra kortet är ett ess, hur stor är sannolikheten att det första också är ett ess?

   Hjälp för den som behöver: En kortlek består av 52 kort. Av dessa är 4 ess. (3 p)

2. Antalet patogena (hälsofarliga) bakterier i Göta Älv antas vara Poissonfördelat med parameter $\lambda = 25$ st per m^3 vatten. I ditt mikroskop betraktar du en droppe innehållande 1 ml älvvatten. Beräkna sannolikheten att din droppe innehåller minst en patogen bakterie. Hjälp för den som behöver: 1 m^3 = 1000 l. (3 p)

3. Härled ett snyggt slutet analytiskt uttryck för mgf för Poi(λ). Visa m.h.a detta att $\mu = \sigma^2 = \lambda$. (4 p)

4. Anta att paret X, Y har en bivariat (tvådimensionell) fördelning med täthet

   $$f(x,y) = c \quad \text{för} \quad 0 < x \le y < 1$$

   där $c > 0$. Bestäm, för $0 < x < 1$, den betingade tätheten för Y givet $X = x$. (4 p)

5. Den s.k Paretofördelningen, Pareto(u, α), definieras av att fördelningsfunktionen är

   $$F(x) = 1 - \left(\frac{u}{x}\right)^{\alpha} \quad \text{för} \quad x > u$$

   där $u > 0$ är en känd nivå och $\alpha > 0$. Härled

   (a) momentskattningen (2 p)

   (b) trolighetsskattningen (2 p)

   av parametern α. Härvid förutsätts att du har givet data $x_1, \ldots, x_n$, som kan anses vara oberoende observationer av Pareto(u, α).

6. I en undersökning av partisympatier erhöll partierna A och B frekvenserna $f_A = 476$ resp $f_B = 119$. Totalt tillfrågades 1372 slumpmässigt utvalda väljare och av dessa valde $n = 1359$ att delta i undersökningen (d.v.s svara på frågan ”Vilket parti skulle du rösta på om det var val idag?”). Punkt- och intervallskatta de sanna proportionerna p_A resp p_B som skulle välja att rösta på A och B om det vore val idag (eller snarare den dag de tillfrågades). Varför är den statistiska felmarginalen mindre för parti B än för parti A? Skriv något klokt om bortfallet som bestod av 13 personer. (5 p)

7. I en mätning av tiden det tar för en robot att utföra ett visst produktionsmoment erhölls i $n = 25$ mätningar medelvärdet $\bar{x} = 47.3$ sekunder och variansen $s^2 = 22.84$. Det anses viktigt att variationen i utförandetiderna är liten. Beräkna därför ett 99% uppåt begränsat konfidensintervall för standardavvikelsen σ. Redovisa vilka antaganden du behöver göra för att kunna lösa uppgiften. (4 p)

8. Beräkna ett 95% konfidensintervall för robotens förväntade utförandetid μ. Redovisa vilka antaganden du behöver göra för att kunna lösa uppgiften. (3 p)

Lycka till!

Kortfattade svar eller lösningar till MVE090 den 24/10-07

1. (a) $\frac{4}{52}\frac{3}{51}+\frac{48}{52}\frac{4}{51}=\frac{1}{13}$ (lagen om total sannolikhet)

   (b) $\frac{3}{51}=\frac{1}{17}$

   (c) $\frac{4}{52}\frac{48}{51}+\frac{48}{52}\frac{4}{51}=\frac{32}{221}$ (lagen om total sannolikhet) eller $\frac{\binom{4}{1}\binom{48}{1}}{\binom{52}{2}}=\frac{32}{221}$

   (d) $\frac{\frac{4}{52}\frac{3}{51}}{\frac{4}{52}\frac{3}{51}+\frac{48}{52}\frac{4}{51}}=\frac{1}{17}$ (Bayes formel)

2. Antalet i en droppe är Poi(λ) med $\lambda=25\cdot10^{-6}$, så sannolikheten för minst en bakterie är $1-e^{-\lambda}\approx\lambda=0.000\,002\,5$ (då λ är mycket liten gäller ju med god noggranhet att $e^{-\lambda}=1-\lambda$)

3. M.g.f för Poisson är $m(t)=\sum_{k=0}^{\infty}e^{kt}e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!}=e^{-\lambda}e^{\lambda e^t}=e^{\lambda(e^t-1)}$
   Derivering: $m'(t)=\lambda e^t e^{\lambda(e^t-1)}=\lambda e^t m(t)$ och $m''(t)=\lambda e^t m(t)+\lambda e^t m'(t)$
   Så, $m'(0)=\lambda$ och $m''(0)=\lambda+\lambda^2$ och det följer att $\mu=\lambda$ samt $\sigma^2=\lambda+\lambda^2-\lambda^2=\lambda$

4. Givet $x\in(0,1)$ är inget y-värde mer troligt än något annat. Mao, $Y|x$ är likformigt fördelad, så vi måste ha att $f_{Y|x}(y)=\frac{1}{1-x}$ för $x\le y<1$
   Alternativ: $f_X(x)=\int_x^1 c\,dy=c(1-x)$ och $f_{Y|x}(y)=\frac{c}{c(1-x)}=\frac{1}{1-x}$ för $x\le y<1$

5. (a) Obs att $f(x)=F'(x)=\frac{\alpha}{u}\left(\frac{u}{x}\right)^{\alpha+1}$. I formelsamlingen ses att $\mu=\frac{\alpha}{\alpha-1}u$ så $\alpha=\frac{\mu}{\mu-u}$. Härur fås att momentskattningen är $\widehat{\alpha}=\frac{\bar{x}}{\bar{x}-u}$. En förutsättning för att detta ska fungera är att $\alpha>1$. Annars existerar ju inte väntevärdet.

   (b) Trolighetsfunktionen är $L(\alpha)=\prod_i\frac{\alpha}{u}\left(\frac{u}{x_i}\right)^{\alpha+1}=\left(\frac{\alpha}{u}\right)^n\left(\prod_i\frac{u}{x_i}\right)^{\alpha+1}$
   Logaritmera: $\mathcal{L}(\alpha)=n\ln\alpha-n\ln u+(\alpha+1)(n\ln u-\sum_i\ln x_i)$
   Derivera: $\mathcal{L}'(\alpha)=\frac{n}{\alpha}+n\ln u-\sum_i\ln x_i$
   Detta $=0$ precis då $\frac{n}{\alpha}=\sum_i\ln x_i-n\ln u$, d.v.s då $\alpha=\frac{n}{\sum_i\ln x_i-n\ln u}=\frac{n}{\sum_i\ln(x_i/u)}$
   Trolighetsskattningen är alltså $\widehat{\alpha}=\frac{n}{\sum_i\ln(x_i/u)}=\frac{1}{\frac{1}{n}\sum_i\ln(x_i/u)}$

   (Man ser att dessa skattningar ger olika resultat. Trolighetsskattningen rekommenderas. Den har bättre egenskaper.)

6. Bortfallet är litet. Om samtliga, t.ex, skulle svarat A, så skulle p_A skattats med $\frac{476+13}{1372}=0.356$ och p_B med $\frac{119-13}{1372}=0.0773$, istället för som nu $\hat{p}_A=\frac{476}{1359}=0.350$ resp $\hat{p}_B=\frac{119}{1359}=0.0876$, men att enbart parti A-sympatisörer avstått från att deltaga är knappast troligt. Antar man att bortfallet har ungefär samma sympatier som de svarande, så blir skattningarna ungefär detsamma som de skulle blivit om alla tillfrågade deltagit i undersökningen.
   Inget är sagt om konfidensgraden i uppgiften, så vi väljer att intervallskatta med säkerheten ca 95%. Då är $\alpha=0.05$ och $z_{\alpha/2}=z_{0.025}=1.96$ samt $p_A=0.350\pm1.96\sqrt{\frac{0.350\cdot0.650}{1359}}=0.350\pm0.025$ och $p_B=0.0876\pm1.96\sqrt{\frac{0.0876\cdot0.9124}{1359}}=0.0876\pm0.0150$.
   Att den statistiska felmarginalen är större för parti A än för parti B beror på att funktionen $p\curvearrowright p(1-p)$ växer från 0 till 0.25 då p växer från 0 till 0.5 och sedan avtar ned mot 0. M.a.o $p_B(1-p_B)<p_A(1-p_A)$. Vi vet ju att $p_B\approx0.09<0.35\approx p_A$.

7. Det är viktigt att data är approximativt normalfördelade. Centrala gränsvärdessatsen kan man inte luta sig så väldigt starkt emot när det är σ eller σ^2 man ska intervallskatta. (J.f.r läroboken, sida 281, nedersta stycket.) I formelsamlingen ser vi att $\frac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$, så sannolikheten att händelsen $\frac{(n-1)s^2}{\sigma^2}\ge\chi^2_{0.99}(24)=10.8564$ inträffat är 0.99. Vi vänder på olikheten och får $\sigma^2\le\frac{(n-1)s^2}{10.8564}=50.4919\Rightarrow\sigma\le7.11$. Notera att σ skattas av $s=\sqrt{22.84}=4.78$.

8. Här går det bättre att hänvisa till centrala gränsvärdessatsen och räkna som om data vore normalfördelade, trots att vi inte alls vet något om datas fördelning. Men obs att det finns besvärliga fördelningar som man behöver väldigt stora stickprov på för att

centrala gränsvärdessatsen ska fungera bra. Så vi kan inte vara bergsäkra på att $\mu = \bar{x} \pm t_{\alpha/2}(n-1)s/\sqrt{n} = 47.3 \pm 2.0639 \cdot \sqrt{22.84/25} = 47.3 \pm 1.97$ (eller $45.33 \leq \mu \leq 49.27$) har den önskade säkerheten ca 95%.