TMS136: Dataanalys och statistik – Tentamen 2011-08-20

Examinator och jour: Erik Jakobsson, tel. 031-772 53 79
Hjälpmedel: Chalmersgodkänd räknare och formelsamling (formelsamling delas ut med tentan).
Betygsgränser: För betyg **3** krävs 14 poäng, för **4** krävs 20 poäng och för betyg **5** krävs 27 poäng.
Fullständiga och välmotiverade lösningar skall ges till varje uppgift.

---

1. Definiera följande begrepp inom punktskattning.
   (a) Stickprov. (1)
   (b) Skattare. (1)
   (c) En skattares bias. (1)
   (d) Förklara hur en icke väntevärdesriktig skattare kan vara att föredra framför en väntevärdesriktig. (1)

2. Du spelar YATZY med en kompis och det enda du har kvar att kasta är liten och stor stege. (Spelet YATZY spelas med fem vanliga tärningar och det gäller att få olika kombinationer som ger olika mycket poäng. Man har tre kast på sig och man får välja att spara tärningar mellan kasten och bara kasta om de man vill. En liten stege är när tärningarna visar 1, 2, 3, 4 ,5 och den är värd 15 poäng. En stor stege är när de visar 2, 3, 4, 5, 6 och är värd 20 poäng.)
   (a) Antag att du kastat för andra gången och ser 1, 2, 2, 2, 2. Om du vill ha en liten stege sparar du alltså 1an och en av 2orna, och kastar om de andra 3. Om du vill ha en stor kastar du om 1an och tre av 2orna. Du är ju ute efter att få så mycket poäng som möjligt, vilket alternativ bör du välja för att maximera din förväntade vinst? Och varför? (2)
   (b) Om du efter andra kastet ser 1, 1, 1, 1, 1. Vad bör du välja då? (1)

3. Antag att du utför ett experiment där en händelse A kan antingen inträffa, eller inte inträffa.
   (a) Visa att om du gör n oberoende upprepningar av experimentet och låter K vara antalet gånger A inträffade, så har K massfunktionen

$$f(k) = \binom{n}{k} P(A)^k P(A^c)^{n-k}. \ (2)$$

   (b) Förklara varför man (om n är stort nog och $P(A)$ inte är för nära 0 eller 1) kan approximera fördelningen för K med normalfördelningen. Och förresten, varför får inte $P(A)$ vara nära 0 eller 1? (2)

4. Visa att exponentialfördelningen har minneslöshetsegenskapen. (3)

5. Låt X vara en kontinuerlig stokastisk variabel med täthet $f(x) = cx^{\alpha}$, för $0 \leq x \leq 1$ och $\alpha \neq -1$. Låt även 0.15, 0.07, 0.22, 0.24, 0.34, 0.43 vara ett stickprov från X.
   (a) Bestäm konstanten c. (1)
   (b) Hitta Maximum likelihood skattaren för α och skatta med stickprovet. (2)
   (c) Hitta Momentskattaren för α och skatta med den. (2)
   (d) Vilka fördelar/nackdelar finns, i allmänhet, med dessa två olika metoder att hitta skattare för en parameter. (1)

6. En (symmetrisk) slumpvandring på heltalen kan beskrivas som att du med början i origo tar ett steg åt höger (till +1), med sannolikhet 1/2, eller åt vänster (till -1), också med sannolikhet 1/2, och att du sedan förtsätter därifrån du hamnade på precis samma sätt, oberoende av tidigare steg. (Detta kallas ibland "The drunkards walk" efter hur du liksom raglar fram och tillbaka helt oordnat över tal-linjen.) Låt

$$X_i = \begin{cases} 1, & \text{om vandringen tar det i:te steget till höger} \\ -1, & \text{om vandringen tar det i:te steget till vänster} \end{cases} \qquad \text{för } i = 1, 2, \ldots$$

$$S_n = \sum_{1}^{n} X_i = \text{ vandringens position efter n steg}$$

   (a) Beräkna väntevärde och varians för S_n. (1)
   (b) Uppskatta sannolikheten att efter 1 miljon steg är vandringen tillbaka (eller fortfarande kvar) i intervallet $[-1000, 1000]$ (3)
   (c) Låt $n > 0$ vara ett jämnt tal. Vad är sannolikheten att $S_n = 0$? (Tips! För att vandringen skall vara tillbaka på 0 måste antalet hopp åt höger vara lika med antalet hopp åt vänster... och hoppen är oberoende...) (3)

7. För att ta reda på hur mycket jag vägde inför beach 2011, mätte jag min vikt på morgonen varje dag under en vecka i början på maj. Vågen visade (i kg såklart) 91.5, 90.0, 90.1, 90.7, 90.8, 89.6 och 90.5. Gör ett konfidensintervall för min vikt. Glöm inte att tydligt nämna eventuella antaganden du gör. (3)

8. Under ett antal år på nittonhundratalet och början av tjugohundratalet inträffade 26 stormar i sverige som fällde mer än 1 miljon m^3 skog. Man antar att sådana här stormar inträffar enligt en Poissonfördelning med okänd intensitet. Tiderna mellan stormarna (i dagar) har varit

158	47	4033	13	979	2976	4378	655	705	7
33	730	3383	66	224	770	661	1132	1140	365
1035	1476	712	75	1075					

   Testa om antagandet om Poissonfördelningen verkar rimligt. (5p)

Lycka till!