Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1903 (f d 5B2501) SANNOLIKHETSLÄRA OCH STATISTIK FÖR 3-ÅRIG MEDIA LÖRDAGEN DEN 11 FEBRUARI 2012 KL 14.00–19.00.

Examinator: Gunnar Englund, tel. 073 3213745.

Tillåtna hjälpmedel: Läroboken, Formel- och tabellsamling i Matematisk statistik. Räknare.

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet. Tentamen består av 5 uppgifter. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 20 poäng. Möjlighet att komplettera ges för de tentander med 18–19 poäng. Tid och plats för komplettering kommer att anges på kursens hemsida. Det ankommer på dig själv att ta reda på om du har rätt att komplettera.

Tentamen kommer att finnas tillgänglig på elevexpeditionen sju veckor efter skrivningstillfället.

Uppgift 1

a) I ett kommunikationssystem sänder man ett meddelande om tre binära tecken (0 eller 1). Sannolikheten för felaktig överföring av ett tecken (0:a mottas som 1:a eller 1:a mottas som 0:a) antas vara 0.1. Händelserna att de olika tecknen överförs felaktigt antas vara oberoende av varandra.
Beräkna sannolikheten att meddelandet ej blir korrekt överfört. (5 p)

b) Petter ljuger med sannolikheten 0,8. Han kastar tärning och jag frågar honom om utfallet blev en sexa. Hur stor är sannolikheten att han svarar ja? (5 p)

Uppgift 2

Skattningarna θ^*_{obs} och $\widehat{\theta}_{\text{obs}}$ är väntevärdesriktiga med medelfelen 0.3 respektive 0.4. Man har föreslagit skattningarna

$0.3\theta^*_{\text{obs}} + 0.7\widehat{\theta}_{\text{obs}}$ samt
$0.4\theta^*_{\text{obs}} + 0.6\widehat{\theta}_{\text{obs}}$.

a) Vilken av dessa två skattningar är att föredra? (Motivering krävs naturligtvis). (5 p)

b) Vad är medelfelet för $\theta^*_{\text{obs}} - \widehat{\theta}_{\text{obs}}$? (5 p)

Uppgift 3

Vid en simtävling används samtidigt automatisk och manuell tidtagning för varje deltagare. Den automatiska betraktas som felfri, medan den manuella har en slumpmässig osäkerhet och ett systematiskt fel Δ. De manuella mätningarna uppfattas som utfall av oberoende normalfördelade stokastiska variabler med den sanna tiden plus det systematiska felet som väntevärde och med en okänd standardavvikelse σ.
Följande data erhölls (tidsenhet sek):

Deltagare:	1	2	3	4
Automatisk tid:	60.03	60.17	61.10	64.22
Manuell tid:	60.18	60.35	61.27	64.40

Beräkna ett 95 % konfidensintervall för det systematiska felet Δ. (10 p)

Uppgift 4

Till nästa års Vasalopp tänker man köpa 7100 liter blåbärssoppa. Man antar att olika åkare dricker blåbärssoppa oberoende av varandra och att den mängd blåbässoppa som en enskild åkare dricker kan beskrivas av en stokastisk variabel X med $E(X) = 0.7$ och $D(X) = 0.6$. Beräkna med lämplig och välmotiverad approximation sannolikheten för att blåbärsoppan räcker om 10000 åker Vasaloppet. (10 p)

Uppgift 5

Vid en undersökning av en viss cancersjukdom vill en forskare ta reda på om cancersjukdomen möjligen förorsakas av någon okänd miljöfaktor. Antag att risken för varje enskild person att få cancer av denna typ ett visst år är 0.0001. Forskaren delar in landet i 800 distrikt om 10000 invånare vardera och tar genom cancerregistret reda på hur många som fått cancerformen ett visst år i de 800 distrikten. I genomsnitt borde det bli ca 1 fall per distrikt, men i ett distrikt upptäcker han 6 fall. Frågan är om detta är oroande många fall.

a) Antag att man bara får reda på att det i ett visst givet distrikt om 10000 personer har uppträtt 6 fall i stället för det förväntade enda fallet. Undersök om detta är oroande många genom att beräkna sannolikheten för 6 eller fler fall i det givna distriktet. (5 p)

b) Om man också får reda på att forskaren valt ut det ”värsta” distriktet, är då också uppgiften om 6 fall oroande? För att bedöma detta skall Du beräkna sannolikheten för att det i åtminstone ett distrikt uppträder åtminstone 6 fall ett visst år. (5 p)

<u>OBS!</u> Lämpliga och väl motiverade approximationer är tillåtna i både a- och b-delen.

Den som ej klarat a-delen får i b-delen använda 10^{-3} som värde på den i a) sökta sannolikheten.

Avd. Matematisk statistik

KTH Matematik

LÖSNINGAR TILL
TENTAMEN I SF1903 (f d 5B2501) SANNOLIKHETSLÄRA OCH STATISTIK FÖR 3-ÅRIG MEDIA LÖRDAGEN DEN 11 FEBRUARI 2012 KL 14.00–19.00.

Uppgift 1

a) Låt A_i = tecken i är felaktigt överfört", $i = 1, 2, 3$. Vi söker

$$P(A_1 \cup A_2 \cup A_3) = 1 - P(A_1^* \cap A_2^* \cap A_3^*) = \text{(oberoendet!)} = 1 - P(A_1^*)P(A_2^*)P(A_3^*) =$$

$$= 1 - (1 - 0.1)^3 = 1 - 0.9^3 = \underline{0.271}$$

b)
Låt A=Petter svarar ja, och B händelsen 6:a upp. Enligt satsen om total sannolikhet är $P(A) = P(A|B)P(B) + P(A|B^*)P(B^*) = 0{,}2 \cdot 1/6 + 0{,}8 \cdot 5/6 = 4{,}2/6 = \underline{0{,}7}$.

Uppgift 2

a) Båda är väntevärdesriktiga eftersom koefficienterna summerar sig till 1 för båda två. Vi jämför alltså varianserna och erhåller

$$V(0.3\theta^* + 0.7\widehat{\theta}) = 0.3^2 V(\theta^*) + 0.7^2 V(\widehat{\theta}).$$

Vi skattar varianserna med medelfelen^2 och erhåller $0.3^2 0.3^2 + 0.7^2 0.4^2 = 0.0865$ respektive $0.4^2 0.3^2 + 0.6^2 0.4^2 = 0.072$ dvs man föredrar $0.4\theta^*_{\text{obs}} + 0.6\widehat{\theta}_{\text{obs}}$ eftersom den har mindre varians.

b) Vi får $V(\theta^* - \widehat{\theta}) = V(\theta^*) + V(\widehat{\theta})$ som ger $0.3^2 + 0.4^2 = 0.25$ som ger medelfelet $\sqrt{0.25} = 0.5$.

Uppgift 3

Låt m_k vara den sanna tiden enligt den automatiska mätningen för simmare nr k och låt X_k vara den stokastiska variabel som beskriver den manuella mätningen. Sätt $Z_k = X_k - m_k$. Då gäller att Z_k är $N(\Delta, \sigma)$-fördelad. Vi har 4 observationer z_1, z_2, z_3, z_4 där $z_1 = 60.18 - 60.03 = 0.15$
$z_2 = 60.35 - 60.17 = 0.18$
$z_3 = 61.27 - 61.10 = 0.17$
$z_4 = 64.40 - 64.22 = 0.18$. Detta ger

$$\bar{z} = \frac{1}{4}\sum_{i=1}^{4} z_i = 0.17, \quad s^2 = \frac{1}{3}\left(\sum_{i=1}^{4} z_i^2 - 4\bar{z}^2\right) = \frac{1}{3}(0.1162 - 4 \cdot 0.17^2) = 0.0002 = 0.014^2.$$

Vi får

$$I_\Delta = \bar{z} \pm t_{0.025}(3)\frac{s}{\sqrt{4}} = 0.17 \pm 3.18 \cdot \frac{0.014}{2} = \underline{0.17 \pm 0.02}$$

Anmärkning: Det går även bra att räkna som i ”parvisa observationer”.

Uppgift 4

Låt X_i vara mängden blåbärssoppa som åkare nr i dricker. Då är $Y = \sum_{i=1}^{10000} X_i$ den mängd soppa som behövs. Vi har

$$E(Y) = \sum_{i=1}^{10000} E(X_i) = 10000 \cdot 0.7 = 7000$$

och

$$V(Y) = \sum_{i=1}^{10000} V(X_i) = 10000 \cdot 0.6^2 = 60^2 \text{ vilket ger } D(Y) = 60.$$

Då 10000 är ett mycket stort antal så följer av CGS att Y är approximativt $N(7000, 60)$-fördelad. Detta ger

$$P(\text{soppan räcker}) = P(Y \leq 7100) = P\left(\frac{Y-7000}{60} \leq \frac{7100-7000}{60}\right)$$

$$= P\left(\frac{Y-7000}{60} \leq 1.667\right) \approx \Phi(1.667) = \underline{0.952}.$$

Uppgift 5

Låt X = antalet fall i det givna distriktet. X är $\text{Bin}(10000, 0.0001) \approx \text{Po}(10000 \cdot 0.0001) = \text{Po}(1)$ eftersom $p = 0.0001 \leq 0.1$ så Poisson-approximation är tillåten.

a)

$$P(X \geq 6) = 1 - P(X \leq 5) \approx 1 - e^{-1}\left(\frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!} + \frac{1^3}{3!} + \frac{1^4}{4!} + \frac{1^5}{5!}\right) \approx \underline{5.9 \cdot 10^{-4}}.$$

Detta kan också erhållas ur Tabell 7 i Formelsamlingen.

b) Låt N = antalet distrikt med 6 eller fler fall. Vi ser att N är

$$\text{Bin}\left(800, P(X \geq 6)\right) \approx \text{Bin}(800, 5.9 \cdot 10^{-4}) \approx \text{Po}(800 \cdot 5.9 \cdot 10^{-4}) = \text{Po}(0.472)$$

där Poisson-approximationen är tillåten ty $5.9 \cdot 10^{-4} \leq 0.1$. Alltså får vi

$$P(N \geq 1) = 1 - P(N = 0) \approx 1 - \frac{0.472^0}{0!}e^{-0.472} \approx \underline{0.376}$$

Slutsatsen är alltså att om forskaren valt ut distriktet på förhand (kanske därför att han misstänker miljöpåverkan just där) så är de 6 fallen oroande många. Om han i stället valt ut distriktet just för att det var speciellt många fall där, så är det inte alls uppseendeväckande många fall just i det distriktet.