Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1901, SF1905 SANNOLIKHETSTEORI OCH STATISTIK, MÅNDAGEN DEN 12:E JANUARI 2015 KL 14.00–19.00.

Kursledare: Gunnar Englund, 073 321 37 45

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik, Mathematics Handbook (Beta), Hjälpreda för miniräknare, 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 6 uppgifter. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 24 poäng. Möjlighet att komplettera ges för tentander med, preliminärt, 22–23 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.

Poäng från kontrollskrivning och laborationer under kursomgång period 2 HT 2014 tillgodoräknas.

Tentamen kommer att vara rättad inom tre arbetsveckor från skrivningstillfället och kommer att finnas tillgänglig på studentexpeditionen minst sju veckor efter skrivningstillfället.

Uppgift 1

a) En stokastisk variabel X har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0 & \text{om } x < 0, \\ x^3 & \text{om } 0 \leq x \leq 1, \\ 1 & \text{om } x > 1. \end{cases}$$

Beräkna $E(X)$ och $D(X)$. (5 p)

b) A och B är två oberoende händelser med sannolikheterna 0.5 resp. 0.4. Beräkna den betingade sannolikheten för att både A och B inträffar, då man vet att minst en av händelserna A och B har inträffat. (5 p)

Uppgift 2

Ett flygplan av viss typ tar 100 passagerare. Passagerarnas vikter kan ses som oberoende stokastiska variabler med väntevärde 75 (kg) och standardavvikelse 10. Vikterna på passagerarnas bagage är också oberoende stokastiska variabler men med väntevärde 15 och standardavvikelse 5. Korrelationskoefficienten mellan en passagerares vikt och vikten av hans bagage är 0.1. Det finns alltså ett visst beroende mellan en passagerares vikt och vikten av hans/hennes bagage.

a) Beräkna väntevärde och varians för den sammanlagda vikten av en passagerare och hans/hennes bagage. (5 p)

b) Beräkna sannolikheten att den sammanlagda vikten för de 100 passagerarna och bagaget överstiger 9300 kg. Rimliga och väl motiverade approximationer får göras. (5 p)

Uppgift 3

De stokastiska variablerna X och Y är oberoende och Poissonfördelade; X har väntevärdet 1 och Y väntevärdet 1/2. Bilda $Z = X + 2Y$.

a) Beräkna väntevärde och varians för Z. (5 p)

b) Beräkna sannolikheten att Z antar värdet 2. (5 p)

Uppgift 4

Man har observerat följande mätvärden

$$x_1 = 15, x_2 = 12, x_3 = 15, x_4 = 14, x_5 = 19.$$

De är oberoende observationer från den diskreta fördelningen med sannolikhetsfunktionen $p_X(k) = \theta(1-\theta)^{k-1},\ k = 1, 2, , 3\ldots$ där $0 < \theta < 1$

a) Bestäm Maximum Likelihood-skattningen (ML-skattningen) av θ. (5 p)

b) Bestäm Minsta Kvadrat-skattningen (MK-skattningen) av θ. (5 p)

Uppgift 5

Följande data visar antalet genomsnittliga förluster i arbetstimmar per vecka på grund av olyckor från tio olika industrianläggningar före respektive efter att ett säkerhetsprogram implementerats.

före :	45	73	46	74	33	57	83	34	26	17
efter:	36	60	44	69	35	51	77	29	24	11

Man kan anta att antalet förlorade arbetstimmar i en given industrianläggning är (approximativt) normalfördelad (både med och utan säkerhetsprogram).

a) Bestäm ett relevant konfidensintervall med konfidensgraden 99% som beskriver effekten av säkerhetsprogrammet. (7 p)

b) Testa på nivån 1% om säkerhetsprogrammet har någon effekt. (3 p)

Uppgift 6

Vid ett tillfälle önskade man undersöka hur småföretagare av olika typ förhöll sig till vissa EU-frågor. Ett slumpmässigt urval om 40 företagare togs ur en viss stor population och av dessa hade 32 åsikten A och 8 åsikten B. Ur en annan stor population togs på motsvarande sätt 39 ut varav 23 tyckte A och 16 B. Ger detta belägg för att de båda typerna av småföretagare är olika i sin benägenhet att stödja A eller B eller kan olikheten förklaras som en slumpens skörd?

Besvara frågan genom att utföra ett lämpligt signifikanstest på (approximativ) nivå 5%. (10 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGAR TILL
TENTAMEN I SF1901, SF1905 SANNOLIKHETSTEORI OCH STATISTIK I
MÅNDAGEN DEN 12 JANUARI 2015 KL 14.00–19.00.

Uppgift 1

a) Vi har

$$f_X(x) = \begin{cases} 3x^2 & \text{om } 0 \leq x \leq 1, \\ 0 & \text{annars.} \end{cases}$$

Detta ger

$$E(X) = \int_0^1 x \cdot 3x^2 \, dx = \left[\frac{3}{4} x^4 \right]_0^1 = \underline{\frac{3}{4}} \quad \text{och} \quad E(X^2) = \int_0^1 x^2 \cdot 3x^2 \, dx = \left[\frac{3}{5} x^5 \right]_0^1 = \frac{3}{5},$$

vilket ger

$$V(X) = E(X^2) - (E(X))^2 = \frac{3}{5} - \frac{9}{16} = \frac{3}{80} \text{ och således } D(X) = \sqrt{\frac{3}{80}} = \underline{0.194}.$$

b) Vi har

$$P(A \cap B \mid A \cup B) = \frac{P((A \cap B) \cap (A \cup B))}{P(A \cup B)} = \left(\begin{matrix} \text{eftersom} \\ (A \cap B) \subset (A \cup B) \end{matrix} \right) = \frac{P(A \cap B)}{P(A \cup B)}$$

$$= \frac{P(A \cap B)}{P(A) + P(B) - P(A \cap B)} = \left(\begin{matrix} \text{eftersom } A \text{ och } B \\ \text{är oberoende} \end{matrix} \right) = \frac{P(A) \cdot P(B)}{P(A) + P(B) - P(A) \cdot P(B)}$$

$$= \frac{0.5 \cdot 0.4}{0.5 + 0.4 - 0.5 \cdot 0.4} = \underline{\frac{2}{7}}.$$

Uppgift 2

a)
Låt X_j=person j:s vikt och Y_j=vikten av person j:s bagage. Låt vidare $Z_j = X_j + Y_j$. Vi får $E(Z_j) = E(X_j) + E(Y_j) = 75 + 15 = \underline{90}$ för $j = 1, 2, \cdots, 100$. Vi får $V(Z_j) = V(X_j + Y_j) = V(X_j) + V(Y_j) + 2C(X_j, Y_j) = 10^2 + 5^2 + 2\varrho(X_j, Y_j) D(X_j) D(Y_j) = 100 + 25 + 2 \cdot 0.1 \cdot 10 \cdot 5 = \underline{135}$

Totala vikten blir $S = Z_1 + Z_2 + \cdots + Z_{100}$ som är approximativt normalfördelad enligt Centrala gränsvärdessatsen ty $Z_1, \cdots, Z_{100}$ är oberoende och likafördelade.
S är approximativt $N(100 \cdot 90, \sqrt{100 \cdot 135})$ och vi får

$$P(S > 9300) = P\left(\frac{S - 9000}{\sqrt{13500}} > \frac{9300 - 9000}{\sqrt{13500}} \right) \approx$$

$$1 - \Phi(300/\sqrt{13500}) \approx 1 - \Phi(2.58) \approx 1 - 0.99506 \approx \underline{4.94 \cdot 10^{-3}}.$$

Uppgift 3

a) Ur antagandet om Poissonfördelning

$$E(Z) = E(X) + 2E(Y) = 1 + 2 \cdot 1/2 = 2,$$

$$V(Z) = V(X) + 4V(Y) = 1 + 4 \cdot 1/2 = 3.$$

b) Ur oberoendet och Poissonfördelningen

$$P(Z = 2) = P(X + 2Y = 2) = P(X = 0, Y = 1) + P(X = 2, Y = 0)$$

$$= P(X = 0)P(Y = 1) + P(X = 2)P(Y = 0)$$

$$= \frac{1^0 e^{-1}}{0!} \frac{(1/2)^1 e^{-1/2}}{1!} + \frac{1^2 e^{-1}}{2!} \frac{(1/2)^0 e^{-1/2}}{0!} = e^{-3/2} = 0.22.$$

Uppgift 4

a) Låt $x_1, \ldots, x_5$ vara observationerna. Log-likelihood-funktionen är

$$\mathrm{LnL}(\theta) = \sum_{i=1}^{5} \left[\ln(\theta) + (x_i - 1) \ln(1 - \theta) \right]$$

Vi söker maximum och sätter derivatan map. $\theta = 0$:

$$0 = \sum_{1=1}^{5} \left(\tfrac{1}{\theta} - (x_i - 1) \tfrac{1}{1-\theta} \right) = \tfrac{5}{\theta} - \left((\sum_1^5 x_i) - 5 \right) \tfrac{1}{1-\theta} = \frac{5}{\theta} + \frac{5}{1-\theta} - \frac{1}{1-\theta} \sum_1^5 x_i.$$

Detta ger $\theta^*_{\text{obs}} = \dfrac{5}{\Sigma x_i} = \dfrac{1}{\bar{x}} = \dfrac{1}{15}$.

b) Vi känner igen att X är ffg-fördelad (se formelsamlingen), och följaktligen $E(X) = \dfrac{1}{\theta}$. Vi skall minimera

$$Q(\theta) = \sum_1^5 (x_i - \frac{1}{\theta})^2$$

och får

$$\frac{dQ}{d\theta} = \frac{2}{\theta^2} \sum_1^5 (x_i - \frac{1}{\theta}) = 0$$

som ger

$$\theta^*_{\text{obs}} = 1/\bar{x} = 1/15.$$

Uppgift 5

a) Vi har observationer i par, och antar att differensen Z före $-$ efter är $N(\mu, \sigma)$. Observationerna av Z är $9, 13, 2, 5, -2, 6, 6, 5, 2, 6$. Vi får $\bar{z} = 5.2$ och $s_z = 4.077$.

Nu är alltså $\dfrac{\bar{z}-\mu}{s_z/\sqrt{10}} = \dfrac{5.2-\mu}{4.077/\sqrt{10}}$ en observation ur en $t(9)$-fördelning, dvs

$$\frac{5.2-\mu}{4.077/\sqrt{10}} = \pm 3.2498$$

som ger

$$\mu = 5.2 \pm 4.19$$

b) Eftersom intervallet för μ bara innehåller positiva tal drar vi slutsatsen att säkerhetsprogrammet minskar risken för olyckor.

Uppgift 6

Använd ett χ^2–homogenitetstest. Bilda

$$Q = \frac{(32-55\cdot 40/79)^2}{55\cdot 40/79} + \frac{(23-55\cdot 39/79)^2}{55\cdot 39/79} + \frac{(8-24\cdot 40/79)^2}{24\cdot 40/79} + \frac{(16-24\cdot 39/79)^2}{24\cdot 39/79}$$

$$= 4.13 > \chi^2_{0.05}((2-1)(2-1)) = \chi^2_{0.05}(1) = 3.84.$$

Hypotesen att benägenheten är densamma kan därmed förkastas på nivån 5%.

Uppgiften skulle kunna lösas som en jämförelse mellan två binomial-parametrar, men villkoret för att approximera med normalfördelning ($np(1-p) \geq 10$) är inte uppfyllt.