Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1920/SF1921 SANNOLIKHETSTEORI OCH STATISTIK, MÅNDAG 11 MARS 2019 KL 8.00–13.00.

Examinator: Björn-Olof Skytt, 08-790 86 49.

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik (utdelas vid tentamen), miniräknare.

Tentamen består av två delar, benämnda del I och del II. Del I består av uppgifterna 1-12. På denna del skall endast svar anges, antingen i form av ett numeriskt värde med tre värdesiffrors noggrannhet eller i form av val av ett av de möjliga svarsalternativen. Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1-3, utan får tillgodoräkna sig dessa tre uppgifter. Gränsen för godkänt är preliminärt 9 poäng. Möjlighet att komplettera ges för tentander med, preliminärt, 8 poäng. Tid och plats för komplettering kommer att anges på kursens hemsida. Del II består av uppgifterna 13-16 och varje korrekt lösning ger 10 poäng. Del II rättas bara för studenter som är godkända på del I och poäng på del II krävs för högre betyg än E. På denna del skall resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Införda beteckningar skall förklaras och definieras och numeriska svar skall anges med minst två värdesiffrors noggrannhet. Studenter som är godkända på datorlaborationen får 4 bonuspoäng på del II på ordinarie tentamenstillfället och det första omtentamenstillfället.

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

Del I

Uppgift 1

Låt X och Y vara stokastiska variabler med följande simultana sannolikhetsfunktion:

$$p_{X,Y}(-2,1) = \frac{1}{16}, \quad p_{X,Y}(-2,2) = \frac{4}{16}, \quad p_{X,Y}(2,1) = \frac{7}{16}, \quad p_{X,Y}(2,2) = \frac{4}{16}$$

Bestäm $C(X,Y)$ d.v.s. kovariansen mellan X och Y.

A: -0.405

B: -0.375

C: 0.375

D: 0.405

Uppgift 2

En stokastisk variabel X har täthetsfunktionen

$$f_X(x) = \begin{cases} 0, & x < 0 \\ x^3, & 0 \leq x \leq 1 \\ \frac{1}{4}, & 1 < x \leq 4 \\ 0, & x > 4 \end{cases}$$

Bestäm $P(\frac{1}{2} < X < 2)$.

A: 0.375

B: 0.484

C: 0.516

D: 0.625

Uppgift 3

Från en mellanstor flygplats opererar 3 flygbolag. Flygbolag A står för 60% av avgångarna, flygbolag B står för 30% av avgångarna, och flygbolag C står för 10% av avgångarna.
Det gäller att 8% av flygbolag A:s avgångar är försenade, 6% av flygbolag B:s avgångar är försenade, och 4.9% av flygbolag C:s avgångar är försenade.
Vad är sannolikheten att ett slumpmässigt plan är försenat?

Uppgift 4

En affär får in 14 datorer varav 5 redan har modem installerade enligt leverantören. Dock är det p.g.a. en miss i kommunikationen ej klarlagt vilka, och de 14 datorerna går ej att skilja på. Antag att vi plockar ut 4 av de 14 datorerna på måfå.
Bestäm sannolikheten att exakt 2 av de 4 utplockade datorerna har modem installerade.

A: 0.3012

B: 0.3288

C: 0.3596

D: 0.3704

Uppgift 5

Antalet spikar i en kartong kan antas vara normalfördelat med väntevärde 563.3 och standardavvikelse 33.2. Leverantören garanterar att antalet spikar överstiger ett visst antal n med sannolikheten 99%. Bestäm n.

A: 486

B: 503

C: 624

D: 641

Uppgift 6

En sur gubbe tar samma buss till jobbet varje morgon. Bussen går var tolfte minut, och antal minuter som sura gubben får vänta på bussen morgon nr i kan anses vara en stokastisk variabel X_i som är likformigt fördelad mellan 0 och 12. D.v.s. $X_i \in U[0, 12]$. Sura gubben hävdar att han under ett år tillbringar mer än ett dygn med att stå och vänta på bussen på morgonen. Låt oss anta att sura gubben tar bussen till jobbet på morgonen 225 gånger på ett år, och att X_i:na är oberoende.
Bestäm den approximativa sannolikheten att sura gubben under ett år tillbringar *mindre* än ett dygn(1440 minuter) med att vänta på bussen på morgonen.

Uppgift 7

En stokastisk variabel X har täthetsfunktionen

$$f_X(x) = \begin{cases} 0 & , \quad x < 0 \\ \dfrac{\theta}{(1+x)^{\theta+1}} & , \quad x > 0 \end{cases}$$

Vidare har vi två observationer av X, nämligen $x_1 = 0.3$ och $x_2 = 0.6$
Bestäm Maximum-Likelihood-skattningen av θ m.h.a. dessa två observationer, om vi vet att θ antingen är 2, 4, 6, eller 8.

Uppgift 8

Antag att $X_1, \ldots, X_n$ utgör ett stickprov på $N(\mu, \sigma)$, där σ är okänd. Ebba önskar testa nollhypotesen $H_0 : \mu = 2$ mot $H_1 : \mu > 2$ med hjälp av ett lämpligt konfidensintervall för μ.
Vilket av nedanstående konfidensintervall för μ skall väljas för att testets signaifikansnivå skall bli α?

A: $I_\mu = \left(-\infty,\ \overline{x} + \frac{s}{\sqrt{n}} \cdot t_\alpha(n-1)\right)$

B: $I_\mu = \left(-\infty,\ \overline{x} + \frac{s}{\sqrt{n}} \cdot t_{\alpha/2}(n-1)\right)$

C: $I_\mu = \left(\overline{x} - \frac{s}{\sqrt{n}} \cdot t_\alpha(n-1), \infty\right)$

D: $I_\mu = \left(\overline{x} - \frac{s}{\sqrt{n}} \cdot t_{\alpha/2}(n-1), \infty\right)$

Uppgift 9

Antag att $X \in Bin(n, p)$. Vi vill bilda ett approximativt konfidensintervall I_p för p.
Vi gör n= 500 försök och får $x = 62$.
Ange det tvåsidiga konfidensintervallets bredd då den approximativa konfidensgraden är 95%.

A: 0.0485

B: 0.0578

C: 0.0617

D: 1.38

Uppgift 10

Låt datamängden 1, 3, 8, 7, 5 vara given. Varje data anses vara ett utfall av en stokastisk variabel X_i, där X_i:na antas vara oberoende och Poissonfördelade med väntevärde μ. D.v.s. varje $X_i \in Po(\mu)$. Eftersom X_i:na är Poissonfördelade skattar vi $E(X_i) = \mu$ med $\bar{x}$ och $D(X_i) = \sqrt{\mu}$ med $\sqrt{\bar{x}}$.
Ange medelfelet för skattningen av μ.

A: 0.980

B: 1.10

C: 2.15

D: 2.19

Uppgift 11

Låt $\bar{x} = 137.0$, $\bar{y} = 208.0$, $\sigma_x^2 = 92.5$, $\sigma_y^2 = 163.0$, $n_x = 5$ och $n_y = 5$ vara givet och antag att $X_i \in N(\mu_x, \sigma_x)$ och $Y_i \in N(\mu_y, \sigma_y)$ samt att alla dessa stokastiska variabler är oberoende. Antag även att σ_x och σ_y är kända.
Ange undre gränsen för det 99%-iga tvåsidiga konfidensintervallet $I_{\mu_x - \mu_y}$.

A: -85.0

B: -89.4

C: -95.0

D: -98.8

Uppgift 12

Vid test av given fördelning har man följande nollhypotes H_0:

$$P(A_1) = 0.25, \quad P(A_2) = 0.50, \quad P(A_3) = 0.25.$$

I ett experiment gör man 140 försök och får då att antalet gånger x_1 som man får resultatet A_1 blir 25, antalet gånger x_2 som man får resultatet A_2 blir 85, och antalet gånger x_3 som man får resultatet A_3 blir 30. Vilken slutsats kan man dra från experimentet?

A: H_0 kan varken förkastas på risknivån 1% eller risknivån 5%

B: H_0 kan både förkastas på risknivån 1% och risknivån 5%

C: H_0 kan förkastas på risknivån 1% , men inte på risknivån 5%

D: H_0 kan förkastas på risknivån 5% , men inte på risknivån 1%

Del II

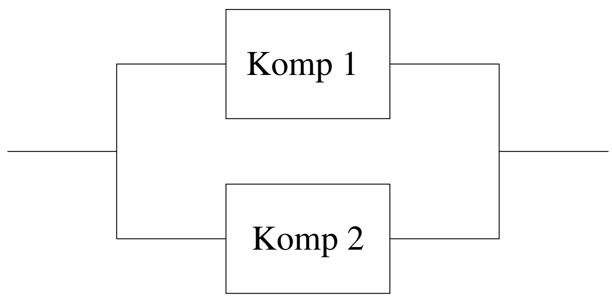
Uppgift 13

a) Låt A och B vara två händelser. Följande sannolikheter är kända

$$P(A \cup B) = 0.92, \quad P(A \cup B^*) = 0.88, \quad P(A^* \cup B) = 0.68.$$

Avgör om A och B är oberoende händelser. Noggrann motivering krävs. (5 p)

b) I ett system är två komponenter kopplade enligt figuren. Systemet fungerar om minst en av komponenterna 1 och 2 fungerar.



Antag att livslängderna T_1 och T_2 för komponent 1, respektive 2 är obeoroende stokastiska variabler med fördelningsfunktioner $F_1(x) = 1 - e^{-x/4}$ för $x \geq 0$, respektive $F_2(x) = 1 - e^{-x/7}$ för $x \geq 0$. Beräkna systemets förväntade livslängd. (5 p)

Uppgift 14

a) För att bestämma längden av en kvadrats sida θ har man gjort två mätningar av kvadratens sida och tre mätningar av dess diagonal. De fem mätningarna kan uppfattas som utfall av oberoende normalfördelade stokastiska variabler vilkas väntevärden överensstämmer med de sanna längderna och vilkas standardavvikelse är σ, där σ är ett känt tal. Detta svarar emot att mätutrustningen inte har något systematiskt fel och att den är beprövad så att man vet standardavvikelsen.

Skatta kvadratens sida θ med MK-metoden. (5 p)

Observera: För att lösningen ska ge poäng måste alla mätningarna utnyttjas.

b) Avgör om skattningen från a) är väntevärdesriktig. (2 p)

c) Antag att man istället mätt kvadratens sida fem gånger och fått observationer $y_1, \ldots, y_5$ och utgående från dessa gör den väntevärdesriktiga skattningen

$$\hat{\theta}_{obs} = \frac{1}{5} \sum_{i=1}^{5} y_i. \tag{1}$$

Avgör vilken av de två skattningarna, MK-skattningen som du tagit fram i uppgift a) eller den som givits i (1), som är effektivast. (3 p)

Uppgift 15

Vid en undersökning av böjhållfasthetens beroende av bränntemperaturen hos gult tegel erhölls följande observationsmaterial på 5 tegelbitar vid temperaturen 700^o och 5 bitar vid temperaturen 800^o.

Temperatur	Böjhållfasthet				
700^o	147	140	121	138	139
800^o	193	227	201	212	207

Antag att slumpmässigheten i data kan beskrivas som normalfördelad med samma standardavvikelse vid båda temperaturerna och oberoende mellan samtliga 10 observationer.

a) Beräkna ett (exakt) 99% konfidensintervall för den systematiska skillnaden i böjhållfasthet för de två temperaturerna. (6 p)

b) Testa om en inverkan av temperaturen på böjhållfastheten kan påvisas. Välj själv en lämplig signifikansnivå. Ange den tydligt, likaså slutsatsen av testet. (4 p)

Uppgift 16

En person spelar ett spel på ett nätkasino. Kasinot hävdar att spelet ger vinst med sannolikhet $p = 0.3$ i varje spelomgång och att omgångarna är oberoende av varandra. Spelaren misstänker att vinstsannolikheten är lägre och nätkasinot har därför gått med på att göra ett hypotestest där nollhypotesten är H_0: $p = 0.3$, och mothypotesen är H_1: $p < 0.3$. Om H_0 förkastas till förmån för H_1 måste nätkasinot ändra sin marknadsföring. Spelaren får välja mellan två testmetoder:

1. Spela 18 omgångar. Om det blir vinst i högst 2 omgångar förkastas H_0 till förmån för H_1.
2. Spela tills det blir vinst i en omgång. Om vinsten inträffar i omgång 9 eller därefter förkastas H_0 till förmån för H_1.

a) Bestäm signifikansnivån (felrisken) för de två testen. (4 p)

b) Bestäm styrkan för de två testen då $p = 0.1$, dvs $h(0.1)$, och avgör utgående från dina resultat vilket test spelaren bör välja? Motivera din slutsats utförligt. (6 p)

Ledning: Man kan ha nytta av följande:

$$\sum_{i=0}^{n} aq^i = a\frac{q^{n+1} - 1}{q - 1}.$$

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1917/SF1918/SF1919 SANNOLIKHETSTEORI OCH STATISTIK,
TISDAG 8 JANUARI 2019 KL 8.00–13.00.

Del I

Uppgift 1

$$C(X,Y) = E(XY) - E(X)E(Y) = \sum_{x,y} x \cdot y \cdot p_{x,y}(x,y) - \sum_{x,y} x \cdot p_{x,y}(x,y) \cdot \sum_{x,y} y \cdot p_{x,y}(x,y) =$$

$$= (-2) \cdot 1 \cdot \frac{1}{16} + (-2) \cdot 2 \cdot \frac{4}{16} + 2 \cdot 1 \cdot \frac{7}{16} + 2 \cdot 2 \cdot \frac{4}{16} -$$

$$[(-2) \cdot \frac{1}{16} + (-2) \cdot \frac{4}{16} + 2 \cdot \frac{7}{16} + 2 \cdot \frac{4}{16}][1 \cdot \frac{1}{16} + 2 \cdot \frac{4}{16} + 1 \cdot \frac{7}{16} + 2 \cdot \frac{4}{16}] =$$

$$= \frac{3}{4} - \frac{3}{4} \cdot \frac{3}{2} = -\frac{3}{8} = -0.375$$

Uppgift 2

$$P(\frac{1}{2} \leq X \leq 2) = \int_{\frac{1}{2}}^{2} f_X(x)dx = \int_{\frac{1}{2}}^{1} x^3 dx + \int_{1}^{2} \frac{1}{4} dx =$$

$$= [\frac{x^4}{4}]_{\frac{1}{2}}^{1} + [\frac{x}{4}]_1^2 = \frac{1}{4} - \frac{1}{64} + \frac{2}{4} - \frac{1}{4} = \frac{1}{2} - \frac{1}{64} = \frac{31}{64} = 0.484$$

Uppgift 3

Händelsen A är att flygbolag A står för avgången, händelsen B är att flygbolag B står för avgången, och händelsen C är att flygbolag C står för avgången. Händelsen S är att avgången är försenad. Sökt:

$$P(S) = P(S|A)P(A) + P(S|B)P(B) + P(S|C)P(C) = 0.08 \cdot 0.6 + 0.06 \cdot 0.3 + 0.049 \cdot 0.1 = 0.0709$$

Uppgift 4

$$P\,(\text{två installerade modem}) = \frac{\binom{5}{2}\binom{9}{2}}{\binom{14}{4}} = 0.3596$$

Uppgift 5

$X \in N(563.3, 33.2)$ och vi söker n så att $P(X > n) = 0.99$ Gör om till $N(0,1)$

$$P\left(\frac{X - 563.3}{33.2} \geq \frac{n - 566.3}{33.2}\right) = 0.99 = P\left(Y \geq \frac{n - 563.3}{33.2}\right) = 0.99$$

där $Y \in N(0,1)$

$$P\left(Y \geq \frac{n - 563.3}{33.2}\right) = 0.99 \Leftrightarrow P\left(Y \geq \frac{563.3 - n}{33.2}\right) = 0.01$$

D.v.s.

$$\frac{566.3 - n}{33.2} = \lambda_{0.01} = 2.3263$$

$\Rightarrow n = 563.3 - 33.2 \cdot 2.3263 = 486$

Uppgift 6

X_i:na är oberoende, likafördelade och många. C.G.S. ger då att

$$Y = \sum_{i=1}^{225} X_i \in N(n \cdot E(X_i), D(X_i) \cdot \sqrt{n})$$

Vi har att X_i:na $\in U(0,12)$ F.S.§4 $\Rightarrow E(X_i) = 6$ och$V(X_i) = \frac{12^2}{12} \Rightarrow D(X_i) = \sqrt{12}$

$$\Rightarrow Y \in N(225 \cdot 6, \sqrt{12} \cdot \sqrt{225}) = N(1350, 51.96)$$

Vi söker nu $P(Y < 1440)$ =[Gör om till N(0,1)]=

$$= P(\frac{Y - 1350}{51.96} < \frac{1440 - 1350}{51.96}) = \Phi(1.73) = 0.9582$$

Uppgift 7

Eftersom vi har två observationer kommer Likelihoodfunktionen att se ut på följande sätt.

$$L(\theta) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) = \frac{\theta}{(1+x_1)^{\theta+1}} \cdot \frac{\theta}{(1+x_2)^{\theta+1}} =$$

$$= \frac{\theta}{(1+0.3)^{\theta+1}} \cdot \frac{\theta}{(1+0.6)^{\theta+1}} = \frac{\theta^2}{[(1+0.3)(1+0.6)]^{\theta+1}} = \frac{\theta^2}{2.08^{\theta+1}}$$

Eftersom vi vet att θ bara kan anta värdena 2, 4, 6, eller 8, och vi ska se för vilket θ som $L(\theta)$ är störst så sätter vi in dessa värden och ser vad $L(\theta)$ blir för respektive värde på θ och får följande.

$$L(2) = 0.444, L(4) = 0.411, L(6) = 0.214, L(8) = 0.087$$

Alltså är ML-skattningen av θ lika med 2.

Uppgift 8

Alternativ C är rätt. Se boken.

Uppgift 9

$X \in Bin(n,p) = Bin(500,p)$ Vi får då konfidensintervallet

$$I_p = p^*_{obs} \pm \sqrt{\frac{p^*_{obs} \cdot (1-p^*_{obs})}{n}} \cdot \lambda_{\frac{\alpha}{2}} = \frac{62}{500} \pm \sqrt{\left(\frac{\frac{62}{500} \cdot (1-\frac{62}{500})}{500}\right)} \cdot \lambda_{0.025}$$

Konfidensintervallets bredd blir alltså

$$2 \cdot \sqrt{\left(\frac{\frac{62}{500} \cdot (1-\frac{62}{500})}{500}\right)} \cdot 1.96 = 0.0578$$

Uppgift 10

Medelfelet är den skattade standardavvikelsen för skattningen μ. D.v.s. $D^*_{obs}(\bar{X})$.

$$D(\bar{X}) = \frac{D(X_i)}{\sqrt{n}} = \frac{\sqrt{\mu}}{\sqrt{n}} \Rightarrow D^*_{obs}(\bar{X}) = \frac{\sqrt{\bar{x}}}{\sqrt{n}} = \frac{\sqrt{\frac{1+3+7+8+5}{5}}}{\sqrt{5}} = 0.980$$

Uppgift 11

$$I_{\mu_x-\mu_y} = \bar{x} - \bar{y} \pm \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \cdot \lambda_{\frac{\alpha}{2}} = 137 - 208 \pm \sqrt{\frac{92.5}{5} + \frac{163.0}{5}} \cdot \lambda_{0.005} =$$

$$= -71 \pm \sqrt{\frac{92.5}{5} + \frac{163.0}{5}} \cdot 2.5758 = -71 \pm 18.5$$

Nedre gränsen är alltså -89.5.

Uppgift 12

$$Q = \sum_{i=1}^{3} \frac{(x_i - n \cdot p_i)^2}{n \cdot p_i} = 6.79$$

$$\chi^2_{0.05}(2) = 5.99 < 6.79 < 9.21 = \chi^2_{0.01}(2)$$

Alltså är D rätt svar.

Del II

Uppgift 13

a) A och B oberoende $\Leftrightarrow P(A \cap B) = P(A)P(B)$. Givet är

$$P(A \cup B) = 0.92, \quad P(A \cup B^*) = 0.88, \quad P(A^* \cup B) = 0.68.$$

Detta ger (rita figur!)

$$\begin{aligned}
P(A) &= P(A \cup B) - P(A^* \cap B) = P(A \cup B) - [1 - P(A \cup B^*)] = 0.92 - 1 + 0.88 = 0.8 \\
P(B) &= P(B \cup A) - P(B^* \cap A) = P(B \cup A) - [1 - P(B \cup A^*)] = 0.92 - 1 + 0.68 = 0.6 \\
P(A \cap B) &= P(A \cup B) - P(A^* \cap B) - P(B^* \cap A) = 0.92 - [1 - P(A \cup B^*)] - [1 - P(B \cup A^*)] \\
&= 0.92 - [1 - 0.88] - [1 - 0.68] = 0.48.
\end{aligned}$$

Nu gäller att

$$P(A)P(B) = 0.8 \cdot 0.6 = 0.48 = P(A \cap B)$$

dvs A och B oberoende.

b) Låt T=systemets livslängd. Då gäller att $T = \max\{T_1, T_2\}$ och

$$\begin{aligned}
F_T(t) &= P(T \leq t) = P(\max\{T_1, T_2\} \leq t) = P(T_1 \leq t, T_2 \leq t) \\
&= \{\text{oberoende}\} = P(T_1 \leq t)P(T_2 \leq t) \\
&= \left[(1 - e^{-t/4})\right]\left[(1 - e^{-t/7})\right] \\
&= 1 + e^{-t(1/4+1/7)} - e^{-t(1/4)} - e^{-t(1/7)}
\end{aligned}$$

Nu deriverar man för att få

$$\begin{aligned}
f_T(t) &= \frac{d}{dt} F_T(t) \\
&= -(1/4 + 1/7) \cdot e^{-t(1/4+1/7)} + 1/4 \cdot e^{-t(1/4)} + 1/7 \cdot e^{-t(1/7)} \\
&= -(11/28) \cdot e^{-t(11/28)} + 1/4 \cdot e^{-t(1/4)} + 1/7 \cdot e^{-t(1/7)}
\end{aligned}$$

Känner man nu igen täthetsfunktionen $f_T(t) = \lambda \cdot e^{-\lambda \cdot t}$ för exponentialfördelningen får man mha formelsamlingen att systemets förväntade livslängd är $\underline{E(T) = -28/11 + 4 + 7 \approx 8.45}$ (annars kan den beräknas genom att man löser integralen

$$E(T) = \int_{-\infty}^{\infty} t f_T(t) dt).$$