Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1920/SF1921 SANNOLIKHETSLÄRA OCH STATISTIK,
ONSDAG 8 JUNI 2022 KL 14.00–19.00

Examinator: Mykola Shykula, 08-790 6644.

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 ett av de möjliga svarsalternativen. Svaren anges på svarsblanketten.

Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1-3, utan får tillgodoräkna sig dessa tre uppgifter (i svarsblanketten anges ordet Bonus). Studenter som är godkända på den andra datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift (i svarsblanketten anges ordet Bonus). Detta gäller endast på det här tentamenstillfället. Gränsen för godkänt är preliminärt 9 poäng. Möjlighet att komplettera ges för tentander med 8 poäng.
Del II består av uppgifterna 13-16 och varje korrekt lösning ger 10 poäng. Del II rättas bara för studenter som är godkända på eller får komplettera del I och poäng på del II krävs för högre betyg än E. På denna del (dvs del II) 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 den andra datorlaborationen får 3 bonuspoäng på del II på det här tentamenstillfället.

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

Bestäm $P(A \cup B \mid A^* \cup B^*)$ om A och B är två händelser sådana att $P(A) = 0.5$, $P(B) = 0.3$ och $P(A \cap B) = 0.1$.

A: 7/9

B: 2/3

C: 1/3

D: 2/9

Uppgift 2

Den stokastiska variabeln X har täthetsfunktionen

$$f_X(x) = \begin{cases} 2\,xe^{1-x^2}, & \text{om} \quad x \geq c > 0, \\ 0, & \text{annars.} \end{cases}$$

Bestäm konstanten c.

A: $c = 0.5$

B: $c = 1$

C: $c = 1.5$

D: $c = 2$

Uppgift 3

De diskreta stokastiska variablerna X och Y har den simultana sannolikhetsfunktionen

$p_{X,Y}(j,k)$	0	1	2
0	$\frac{1}{12}$	$\frac{1}{3}$	0
1	0	$\frac{1}{2}$	$\frac{1}{12}$

Således antar X alltså värdena 0 eller 1, medan Y antar värdena 0, 1, eller 2.
Beräkna standardavvikelsen $D(XY)$.

Uppgift 4

Antalet samtal till ett företags callcenter under en tjugominuters period antas vara Poissonfördelad med väntevärdet 2.5. Vad är sannolikheten att det kommer minst 5 samtal under perioden $10.00-11.00$?

A: 0.24

B: 0.67

C: 0.76

D: 0.87

Uppgift 5

Vid en automatförpackning av kex placeras dessa intill varandra. Vidare placeras dessa kex mellan två stöd som ligger på avståndet 23 cm från varandra. Tjockleken hos kexen (i cm) antas vara normalfördelade med väntevärdet 2 och standardavvikelsen 0.3, dvs $N(2, 0.3)$, och är oberoende från varandra. Vad är sannolikheten att kexpaket kommer att innehålla åtminstone 12 kex ?

A: 0.001

B: 0.04

C: 0.17

D: 0.59

Uppgift 6

Låt A vara kvadraten $\{(x, y) \in \mathbb{R}^2 : -1 \leq x \leq 1, -1 \leq y \leq 1\}$. Låt den tvådimensionella s.v. (X, Y) vara likformigt fördelad på A. Beräkna sannolikheten $P(0 \leq Y - X \leq 1)$.

A: 0.125

B: 0.250

C: 0.375

D: 0.5

Uppgift 7

Man har två oberoende observationer $x_1 = 5$ och $x_2 = 12$ av en stokastisk variabel med täthetsfunktionen

$$f_X(x) = \frac{x}{a}e^{-(x/a)^2},\ x > 0,\ a > 0.$$

Bestäm Maximum-Likelihood skattningen av a.

A: 13

B: 18.38

C: 8.5

D: 4.12

Uppgift 8

Låt $x = 60$ vara ett utfall av en stokastisk variabel $X \in Bin(n, p)$, där $n = 600$ och p är okänd. Bestäm medelfelet för skattningen $p^* = X/n$.

A: 0.0122

B: 0.09

C: 7.35

D: $1.5 \cdot 10^{-4}$

Uppgift 9

Från ett normalfördelat stickprov erhålls observationerna 4.82, 4.25, 3.96, 3.73, 3.32. Ange ett 99% nedåt begränsat konfidensintervall för väntevärdet.

A: $[3.57, \infty)$

B: $[3.07, \infty)$

C: $[2.82, \infty)$

D: $[3.48, \infty)$

Uppgift 10

Man är intresserad av den andel p av alla studenter på en högskola som dricker alkohol minst två gånger i veckan. För att testa

$$H_0 : p = 0.4 \text{ mot } H_1 : p > 0.4$$

tillfrågas 18 slumpmässigt utvalda studenter och H_0 förkastas om fler än 10 av de 18 dricker alkohol minst två gånger i veckan. Beräkna testets styrka då $p = 0.5$.

A: 0.94

B: 0.76

C: 0.24

D: 0.12

Uppgift 11

Man gör ett stickprov $x_1, x_2, \ldots, x_n$ där $X_i \in N(\mu, \sigma)$, där μ är okänd och $\sigma = 3$. Låt $\bar{x}$ vara stickprovsmedelvärdet. För att testa $H_0 : \mu = 10$ mot $H_1 : \mu > 10$ vill man använda beslutsregeln " förkasta H_0 om $\bar{x} > 10.5$". Hur stort måste n minst vara för att man ska kunna förkasta H_0 på signifikansnivån 5%, med hjälp av den ovan nämnda beslutsregeln?

A: 139

B: 98

C: 44

D: 10

Uppgift 12

Följande datamaterial beskriver hur försäljningen av en vara beror av hur mycket som spenderats på reklam.

Reklam (kkr)	22	29	31	35	42
Försäljning (kkr)	103	108	106	110	114

Det är rimligt att tro att det föreligger ett linjärt samband mellan variablerna. Utifrån datamaterialet ovan skattas en linjär regressionsmodell

$$y_i = \alpha + \beta x_i + \varepsilon_i, \; i = 1, \ldots, 5,$$

där y_i = försäljning (kkr) beror av x_i = reklamkostnad (kkr) och ε_i betecknar slumpmässiga fel. Minsta-kvadrat-skattningarna av regressionskoefficienterna α och β blev $\alpha^*_{obs} = 91.02$ respektive $\beta^*_{obs} = 0.54$. Man kontrollerar vidare om den effekt som reklamkostnaden har på försäljningen är signifikant, dvs man testar $H_0 : \beta = 0$ mot $H_1 : \beta \neq 0$. P-värdet för testet blev 0.008.
Vilken slutsats kan man dra då man fått det P-värdet?

A: Effekten är signifikant på både risknivån 1 % och risknivån 2.5 %.

B: Effekten är varken signifikant på risknivån 1 % eller risknivån 2.5 %.

C: Effekten är signifikant på risknivån 2.5 %, men inte på risknivån 1 %.

D: Effekten är signifikant på risknivån 1 %, men inte på risknivån 2.5 %.

Var god vänd!

Del II

Uppgift 13

Ett flygbolag vet av erfarenhet att 4% av resenärerna som köpt biljetter inte dyker upp. Bolaget vill undersöka möjligheten att sälja fler biljetter än vad som får plats i flygplanet. En Boeing777 har en kapacitet på 440 pasagerare. Antag att 440 st biljetter redan sålts. Antag vidare att pasagerare köper biljetter oberoende av varandra. Hur många extrabiljetter törs bolaget sälja om sannolikheten ska vara minst 99.9% att alla får plats? (10 p)

Ledning: lösningarna till ekvationen $ax^2 + bx + c = 0$, där a, b, c är konstanter, $a \neq 0$, ges av

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Uppgift 14

Slumptal genereras (oberoende av varandra) på en dator enligt en likformig fördelning på intervallet mellan 0 och 4.

a) Vad är sannolikheten att av 10 genererade mätvärden inget överstiger 3. (3 p)

b) Vad är sannolikheten att 4 genererade slumptal hamnar i var sitt heltalsintervall, dvs att man får exakt ett tal i vart och ett av intervallen $(0, 1)$, $(1, 2)$, $(2, 3)$, $(3, 4)$. (4 p)

c) Antag att man genererar slumpmässigt nu ett antal sådana tal, dvs likformigt fördelade mellan 0 och 4. Antag att man stryker alla tal som överstiger 3. Hur stor andel av återstående tal ligger mellan 1 och 2? (3 p)

Uppgift 15

Livslängd (i mil) för ett visst bildäck kan betraktas som en normalfördelad stokastisk variabel med väntevärde μ och standardavvikelse $\sigma = 1000$. Tillverkaren hävdar att livslängden har ökat genom en ny tillverkningsprocess. Tidigare var $\mu = 6000$ medan man nu har $\mu > 6000$ och det är mycket troligt att $\mu = 7000$. För att bevisa påståendet tänker man mäta livslängden hos n däck (dvs man ska ta ett stickprov av storlek n) och sedan testa $H_0 : \mu = 6000$ mot $H_1 : \mu > 6000$. Man vill använda beslutsregeln "förkasta H_0 om $\bar{x} > k$", där $\bar{x}$ är stickprovsmedelvärdet. Bestäm n och k så att signifikansnivå blir 5% och testets styrka i $\mu = 7000$ blir nära 90%. (10 p)

Uppgift 16

De två stokastiska variablerna X och Y är oberoende och normalfördelade med väntevärdet 0 och standardavvikelsen 1, dvs $N(0, 1)$-fördelade. Beräkna sannolikheten $P(0 \leq Y \leq 4X)$. (10 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG
TENTAMEN I SF1920/SF1921 SANNOLIKHETSLÄRA OCH STATISTIK,
ONSDAG 8 JUNI 2022 KL 14.00–19.00.

Del I, Svar

1. B

2. B

3. 0.624 (eller $\frac{\sqrt{14}}{6}$)

4. D

5. C

6. C

7. A

8. A

9. B

10. C

11. B

12. A

Del II, Svar

13. 5 st, man kan tex approximera med CGS

14a. $(3/4)^{10}$; 14b. $24/(4^4)$; 14c. $1/3$

15. $n = 9$ och testet blir förkasta nollhypotes om stickprovetsmedelvärde är större än $k = 6562$

16. 0.211 $(= tan^{-1}(4)/(2\pi) = arctan(4)/(2\pi) = (75.96)/360))$

Del I, Lösningsförslag

Uppgift 1

Venndiagram metoden tillsammans med masstolkningen ger:

$$P(A \cup B \mid A^* \cup B^*) = |Def| = \frac{P((A \cup B) \cap (A^* \cup B^*))}{P(A^* \cup B^*)} =$$

$$= \frac{P(A \cap B^*) + P(B \cap A^*)}{1 - P(A \cap B)} = \frac{P(A) - P(A \cap B) + P(B) - P(A \cap B)}{1 - P(A \cap B)} =$$

$$= \frac{0.5 + 0.3 - 2 \cdot (0.1)}{1 - 0.1} = \frac{0.6}{0.9} = \frac{2}{3}$$

Uppgift 2

Följande gäller: $\int_c^\infty f_X(x)dx = 1$. Vi utvecklar det uttrycket, integrerar och sedan löser ut ekvationen för $c, c > 0$:

$$\int_c^\infty 2xe^{1-x^2} dx = e \int_c^\infty 2xe^{-x^2} dx = 1,$$

$$1 = e \int_c^\infty 2xe^{-x^2} = (-e) \int_c^\infty e^{-x^2} d(-x^2) = (-e) \Big[e^{-x^2} \Big]_c^\infty = e \cdot e^{-c^2},$$

och eftersom $c > 0$, får vi att $c = 1$.

Uppgift 3

Det gäller att $D(XY) = \sqrt{V(XY)}$. Vi har enl sannolikhetsfunktionen:
$V(XY) = E(X^2Y^2) - (E(XY))^2 = 1^2 \cdot 1^2 \cdot \frac{1}{2} + 1^2 \cdot 2^2 \cdot \frac{1}{12} - \Big(1 \cdot 1 \cdot \frac{1}{2} + 1 \cdot 2 \cdot \frac{1}{12}\Big)^2 = \frac{1}{2} + \frac{1}{3} - \Big(\frac{1}{2} + \frac{1}{6}\Big)^2 = \frac{5}{6} - \Big(\frac{2}{3}\Big)^2 = \frac{5}{6} - \frac{4}{9} = \frac{15-8}{18} = \frac{7}{18} = \frac{14}{36}$.
Alltså, $D(XY) = \sqrt{V(XY)} = \sqrt{\frac{14}{36}} = \frac{\sqrt{14}}{6} = 0.624$

Uppgift 4

Vi har $X \in Po(2.5)$, där s.v. X anger antal samtal till ett företags callcenter under 20-min period. Därför har vi att $Y \in Po(7.5)$, där s.v. Y i sin tur anger antal samtal till företagets callcenter under en timmes period. Slutligen,

$$P(Y \geq 5) = 1 - P(Y \leq 4) = |\text{Tabell 5}| = 1 - 0.13206 \approx 0.87$$

Uppgift 5

Låt $X_i \in N(2, 0.3)$ ange tjockleken hos de 12 olika kexen, $i = 1, 2, \ldots, 12$.
Vi har $Y = \sum_{i=1}^{12} X_i \in N(24, 0.3 \cdot \sqrt{12})$, och den sökta sannolikheten blir:

$$P\Big(\sum_{i=1}^{12} X_i \leq 23\Big) = \Phi\Big(\frac{23 - 24}{0.3 \cdot \sqrt{12}}\Big) = \Phi(-0.96) = 1 - \Phi(0.96) = |\text{Tabell 1}| \approx 0.17$$

Uppgift 6

Enl den geometriska sannolikheten:

$$P(0 \le Y - X \le 1) = \frac{\text{arean av S}}{\text{arean av A}},$$

där A är kvadraten, $A = \{(x, y) \in \mathbb{R}^2 : -1 \le x \le 1, -1 \le y \le 1\}$, och S är den delen av A som ligger mellan linjerna $y = x$ och $y = x + 1$. Därför,

$$P(0 \le Y - X \le 1) = \frac{\text{arean av S}}{\text{arean av A}} = \frac{\frac{2\cdot 2}{2} - \frac{1\cdot 1}{2}}{2^2} = \frac{2 - \frac{1}{2}}{4} = \frac{3}{8} = 0.375$$

Uppgift 7

Observera först att $a > 0$. Vi logarithmerar och deriverar the likelihoodfunktionen

$$L(a) = f_X(x_1; a) \cdot f_X(x_2; a) = \frac{x_1 x_2}{a^2} e^{-(x_1^2 + x_2^2)/a^2}.$$

Vi har

$$\ln L(a) = \ln(x_1 x_2) - 2 \ln a - \frac{x_1^2 + x_2^2}{a^2}.$$

Sedan deriverar man $\ln L(a)$ mha på a:

$$\frac{d \ln L(a)}{da} = -\frac{2}{a} + \frac{2(x_1^2 + x_2^2)}{a^3}.$$

Vi löser ekvation $\frac{d \ln L(a)}{da} = 0$ och får $a^2 = x_1^2 + x_2^2$, som leder till $a = \sqrt{x_1^2 + x_2^2} = \sqrt{5^2 + 12^2} = 13$.

Uppgift 8

$p^*_{obs} = \frac{x}{n} = \frac{60}{600} = 0.1$. Eftersom X är binomialfördelad så är $V(X) = np(1-p)$ (se §3 i FS), och därför har vi:

$$V(p^*) = V\Big(\frac{X}{n}\Big) = \frac{V(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Vilket i sin tur leder till att medelfelet blir

$$d(p^*) = \big(D(p^*)\big)^*_{obs} = \left(\sqrt{V(p^*)}\right)^*_{obs} = \sqrt{\frac{p^*_{obs}(1 - p^*_{obs})}{n}} = \sqrt{\frac{0.1 \cdot 0.9}{600}} \approx 0.0122$$

Uppgift 9

Vi har $\bar{x} = 4.016$ och $s = 0.5636$. Eftersom σ är okänd, får vi den nedersta gränsen till

$$\bar{x} - t_{0.01}(4) * s/\sqrt{5} = 4.016 - 3.75 * 0.5636/\sqrt{5} = 3.07,$$

där $t_{0.01}(4) = 3.75$ från Tabell 3.

Uppgift 10

Låt X vara en s.v. som anger antal studenter av 18 som dricker alkohol minst två gånger i veckan. Vi har $X \in Bin(18, p)$. Vi testar $H_0 : p = 0.4$ mot $H_1 : p > 0.4$. Testets styrka i punkt $p = 0.5$ är

$$h(0.5) = P(\text{förkasta } H_0 | H_1 \text{ är sann med } p = 0.5) = P(X \geq 11 | X \in Bin(18, 0.5)) =$$

$$= 1 - P(X \leq 10) = |\text{Tabell 6 för Bin}(18, 0.5)| = 1 - 0.75966 \approx 0.24$$

Uppgift 11

Vi har

$$P(\bar{X} > 10.5 | H_0 \text{ är sann}) \leq 0.05.$$

Vidare, om H_0 är sann, $\bar{X} \in N(10, 3/\sqrt{n})$. Därför,

$$P(\bar{X} > 10.5 | H_0 \text{ är sann}) = P(\bar{X} > 10.5 | \bar{X} \in N(10, 3/\sqrt{n})) \leq 0.05.$$

From Tabell 2, har vi nu

$$\frac{10.5 - 10}{3/\sqrt{n}} \geq \lambda_{0.05} = 1.6449,$$

vilket leder i sin tur till $\sqrt{n} \geq (3 \cdot 1.6449)/0.5 = 9.8694$, eller $n \geq (9.8694)^2 = 97.405$. Alltså, n måste vara minst 98.

Uppgift 12

Effekten är signifikant på både risknivån 1% och risknivån 2.5%, eftersom P-vardet$= 0.008$ är mindre än både 0.01 och 0.025.

Del II, Lösningsförslag

Uppgift 13

Låt $X_i \in Bin(1, 0.96)$ vara en s.v. som anger om resenär i dyker upp eller ej dyker upp, $i = 1, 2, \ldots, n$. Dvs om resenär i dyker upp så har vi $X_i = 1$, annars, $X_i = 0$, och $P(X_i = 1) = 0.96$ enl lydelsen. Vi söker n (eller, det är $n - 440$ som söks egentligen, att vara mer precis) så att

$$P(X_1 + X_2 + \ldots + X_n \leq 440) \geq 0.999$$

Vi kan använda oss av CGS (centrala gränsvärdesatsen), eftersom n är stor (minst 440).
Vi har, $\mu = E(X_i) = 0.96$ och $\sigma = \sqrt{V(X_i)} = \sqrt{p(1-p)} = \sqrt{0.96(0.04)}$, $i = 1, \ldots, n$.
Enligt CGS, $X_1 + X_2 + \ldots + X_n \in ApproxN(n\mu, \sigma\sqrt{n})$. Detta tillsammans med det ovan nämnda $P(X_1 + X_2 + \ldots + X_n \leq 440) \geq 0.999$, ger följande:

$$\frac{440 - n\mu}{\sigma\sqrt{n}} \geq 3.0902,$$

där $\mu = 0.96$ och $\sigma = \sqrt{0.96(0.04)} \approx 0.196$.
Efter det förenklas och omflyttas, blir det

$$440 - 0.96\big(\sqrt{n}\big)^2 - 3.0902 \cdot (0.196)\sqrt{n} \geq 0,$$

eller

$$0.96\big(\sqrt{n}\big)^2 + 0.60568\sqrt{n} - 440 \leq 0.$$

Lösningen till denne olikhet blir $0 \leq \sqrt{n} \leq 21.0956$, vilket innebär, $0 \leq n \leq 445.024$.
Dvs att flygbolaget törs sälja bara 5 st extrabiljetter.

Uppgift 14

a) Sannolikheten att ett slumpmässigt genererat tal, som är $U(0, 4)$, inte överstiger 3 är $3/4$. Multiplikationsprincipen och oberoendet ger att på 10 tal en sådan sannolikhet blir $(3/4)^{10}$.

b) Att 4 genererade slumptal på $(0, 4)$ hamnar i var sitt av de angivna fyra interval är ju samma sak som att välja slumpmässigt något/några av de angivna intervallen. Den klassiska definitionen av sannolikheten används här. Det blir $4 \cdot 3 \cdot 2 \cdot 1 = 24$ antal fall som gynnar att man får exakt ett tal i vart och ett av intervallen $(0, 1)$, $(1, 2)$, $(2, 3)$, $(3, 4)$ (räknar med ordningen). Och det blir $4 \cdot 4 \cdot 4 \cdot 4 = 4^4$ totalt antal fall på hur de fyra slumpmässigt genererade tal hamnar i de fyra angivna intervallen (räknar med ordningen också). Dvs den sökta sannolikheten blir $24/4^4$.

c) De tallen som blir kvar är då likformig fördelade på $(0, 3)$. Vilket leder i sin tur till att mellan 1 och 2 bör ligga $1/3$ del av talen som är kvar, efter att alla som överstiger 3 blev strukna.

Uppgift 15

Vi utnyttjar att signifikansnivån ska bli 5% och att testets styrka i $\mu = 7000$ ska bli nära 90%. Det första resp det andra ger två ekvationer med två okända (nämligen n och k är okända):

$$\begin{cases} P(\bar{X} > k|\mu = 6000) = 0.05, \\ P(\bar{X} > k|\mu = 7000) \geq 0.90. \end{cases}$$

Eftersom $\bar{X} \in N(\mu, \sigma/\sqrt{n})$, $\sigma = 1000$, (detta pga normalfördelnings antagande från lydelsen), vi har:

$$\begin{cases} \frac{k-6000}{1000/\sqrt{n}} = 1.6449, \\ \frac{k-7000}{1000/\sqrt{n}} = -1.2816, \end{cases}$$

vilket leder till ekvation

$$\frac{k - 6000}{1.6449} = \frac{k - 7000}{-1.2816}$$

med den approx lösningen $k = 6562$.
Vidare, från den första ekvationen, har vi

$$\sqrt{n} = \frac{(1.6449)(1000)}{(6562 - 6000)} \approx 2.93,$$

och $n = 2.93^2 = 8.585$. Antalet observationer n är ju heltal och vi väljer den närmaste uppåt, dvs, $n = 9$ för att inte minska styrkan.

Uppgift 16

Integrera den simultana täthetsfunktionen över ett lämpligt område, och sedan polära koordinater ska användas. Vi har $f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$.
Den sökta sannolikheten är

$$P(0 \leq Y \leq 4X) = \frac{1}{2\pi} \int_0^{\infty} dx \int_0^{4x} e^{-\frac{x^2+y^2}{2}} dy = \Big| x = r\cos(\theta), y = r\sin(\theta), |J| = r \Big| =$$

$$= \frac{1}{2\pi} \int_0^{\infty} dr \int_0^{\arctan(4)} re^{-\frac{r^2}{2}} d\theta = \frac{1}{2\pi} \left(\int_0^{\arctan(4)} d\theta \right) \int_0^{\infty} re^{-\frac{r^2}{2}} dr = \frac{\arctan(4)}{2\pi} =$$

$$= |radianer| = \frac{75.964}{360} = 0.211$$