Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1912/1914/1915/1916 SANNOLIKHETSTEORI OCH STATISTIK, FREDAG 22 OKTOBER 2021 KL 8.00–13.00.

Examinator för SF1912: Mykola Shykula, 08-790 6644.
Examinator för SF1914/1915/1916: Björn-Olof Skytt, 08-790 8649.

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. 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 datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift (i svarsblanketten anges ordet Bonus). Detta gäller på den här tentan och vid omtentamen i december 2021. Gränsen för godkänt är preliminärt 9 poäng. Möjlighet att komplettera ges för tentander med, preliminärt, 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 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 3 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

Sex brickor, med talen $1, 1, 2, 2, 3, 3$ ligger i en påse. Två brickor dras slumpmässigt en i taget, utan återläggning. Vad är sannolikheten att talet på den första brickan som dras är större än på den andra brickan som dras?

A: 1/3

B: 4/15

C: 2/5

D: 1/2

Uppgift 2

En stokastisk variabel X har täthetsfunktionen

$$f_X(x) = \begin{cases} x^2/9, & 0 \leq x \leq 3, \\ 0, & \text{annars.} \end{cases}$$

Bestäm den betingade sannolikheten $P(X > 2 | X > 1)$.

A: 5/9

B: 19/27

C: 5/8

D: 19/26

Uppgift 3

Låt X, Y och Z vara oberoende stokastiska variabler sådana att $V(X) = 4$, $V(Y) = 1$, och $V(Z) = 2$. Bestäm standardavvikelsen $D(2X - 4Y + 3Z)$.

Uppgift 4

För de två stokastiska variablerna X och Y gäller att

$$P(X = 0, Y = 0) = 0.10, \quad P(X = 1, Y = 0) = 0.30,$$
$$P(X = 0, Y = 1) = 0.20, \quad P(X = 1, Y = 1) = 0.40.$$

Bestäm kovariansen $C(X, Y)$.

A: 0.42

B: 0.4

C: 0.0024

D: -0.02

Uppgift 5

En normalfördelad stokastisk variabel X med väntevarde 5, har egenskapen att $P(X < 0) = 0.05$. Bestäm standardavvikelsen $D(X - 5)$.

A: 3.04

B: 1.96

C: 1.6449

D: 5.0

Uppgift 6

Låt $X \in Po(2)$ och $Y \in Po(2)$ vara två oberoende stokastiska variabler. Låt $Z = \max(X, Y)$. Beräkna $P(Z \geq 2)$.

A: 0.323

B: 0.594

C: 0.835

D: 0.908

Uppgift 7

En stokastisk variabel X_i har fördelningsfunktionen

$$F_{X_i}(x) = \begin{cases} x^\alpha, & 0 \leq x \leq 1, \\ 0, & \text{annars,} \end{cases}$$

där $i = 1, 2, 3$. Låt $x_1 = 0.873$, $x_2 = 0.910$ och $x_3 = 0.820$ vara observationer på de oberoende stokastiska variablerna X_1, X_2, och X_3. Bestäm Maximum-likelihood-skattningen av α utgående från dessa tre observationer.

A: 4

B: 5

C: 6

D: 7

Uppgift 8

Låt x vara en observation på den stokastiska variabeln X. $X \in Po(\mu)$, där μ är en okänd parameter. Minsta-kvadrat-skattningen av μ ges då av

$$\mu^*_{obs} = x,$$

dvs $\mu^* = X$. Bestäm medelfelet för μ^*, dvs bestäm $d(\mu^*)$.

A: $d(\mu^*) = \mu$

B: $d(\mu^*) = \sqrt{\mu x}$

C: $d(\mu^*) = x$

D: $d(\mu^*) = \sqrt{x}$

E: $d(\mu^*) = \sqrt{\mu}$

Uppgift 9

Ett skevt mynt, där $P(\text{krona}) = p$, kastas 30 gånger. Krona inträffade 20 gånger. Bestäm ett tvåsidigt konfidensintervall för p med konfidensgrad 90%. Ange nedre gränsen.

A: 0.53

B: 0.56

C: 0.65

D: 0.67

Uppgift 10

Låt $X \in Po(\mu)$. Vi testar $H_0 : \mu = 3$. Vi förkastar H_0 om $x \geq 7$. Bestäm styrkan hos testet för alternativet $H_1 : \mu = 7$.

A: 0.99

B: 0.60

C: 0.55

D: 0.45

Uppgift 11

I ett laboratorium utförs dagligen mätningar för att bestämma halten μ av ett visst ämne i en råvara. Mätningarna kan betraktas som observationer från $N(\mu, \sigma)$-fördelningen, där σ är okänd. En dag fås följande sex mätvärden (enhet: %):

$$5.72 \quad 5.62 \quad 6.02 \quad 6.34 \quad 4.97 \quad 5.02$$

Bestäm ett tvåsidigt konfidensintervall för μ med konfidensgrad 99%. Ange intervallets övre gräns.

A: 6.05

B: 6.18

C: 6.51

D: 6.91

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)	105	108	107	100	89

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. Bestäm Minsta-kvadrat-skattningarna α^*_{obs} och β^*_{obs} av regressionskoefficienterna α resp β.

A: $\alpha^*_{obs} = 0.865, \ \ \beta^*_{obs} = 129.3$

B: $\alpha^*_{obs} = 129.3, \ \ \beta^*_{obs} = -0.865$

C: $\alpha^*_{obs} = 129.3, \ \ \beta^*_{obs} = 0.865$

D: $\alpha^*_{obs} = -0.865, \ \ \beta^*_{obs} = 129.3$

Del II

Uppgift 13

Alice är en rollspelsnörd och går alltid runt med en 6-sidig, en 12-sidig och en 20-sidig tärning i handväskan. Sidorna är numrerade $1, \ldots, 6$, $1, \ldots, 12$, samt $1, \ldots, 20$, respektive. Eftersom den 6-sidiga tärningen är lite större, väljer Alice denna med 50% sannolikhet från väskan, de övriga plockas med 25% sannolikhet var. Alice plockar en tärning på måfå och kastar. Resultatet blir en 5:a. Vad är sannolikheten att tärningen hon kastade är den 6-sidiga tärningen? (10 p)

Uppgift 14

Alice och Bob driver ett konsultbolag. Under en typisk vecka får de särskilda uppdrag med följande sannolikhet och utbetalning.

Sannolikhet	0.3	0.4	0.3
Alice	800	1000	1500
Bob	500	800	1200

Det vill säga, Bob får ett uppdrag som inbringar 800 kr, med sannolikhet 0.4 under en vecka. Hur stor är den approximativa sannolikheten att deras gemensamma inkomst för dessa uppdrag överstiger 80 000 kr under ett arbetsår om ett arbetsår är 40 veckor? Var tydlig med vilka antaganden som görs och vilka satser som används. (10 p)

Uppgift 15

Fem personer (A-E) är medlemmar på en nätdejtingsite. Under en månad använder de en neutral profilbild. Månaden senare byter de till att använda en profilbild där de håller i en hund. Antalet matchningar per månad och person presenteras nedan:

Person:	A	B	C	D	E
Utan hund:	89	112	97	75	102
Med hund:	102	125	107	88	109

Antag att båda series mätdata (dvs med och utan hund) är normalfördelade. Testa på signifikansnivån 5% om det finns en systematisk skillnad mellan antalet matchningar med hund och antalet matchningar utan hund. Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är. Slutsatsen och den tillämpade statistiska metoder bör motiveras. (10 p)

Uppgift 16

En 6-sidig tärning har sannolikheten θ_1 för att få 1, sannolikheten θ_2 för att få 2. Sannolikheten för att få 3 är lika med sannolikheten för att få 4 som är lika med sannolikheten för att få 5 som är lika med sannolikheten för att få 6. På 100 kast har vi fått följande utfall, för antalet ettor, tvåor, etc.

$$10, \ 20, \ 15, \ 25, \ 21, \ 9.$$

Finn MK-skattningen av θ_1 och θ_2, samt avgör på signifikansnivån 5% om tärningen har denna (MK-skattade) fördelning eller inte. Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är. (10 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1912/SF1914/SF1915/SF1916 SANNOLIKHETSTEORI OCH STATISTIK,
FREDAG 22 OKTOBER 2021 KL 8.00–13.00.

Del I

Svar:

1. C

2. D

3. $\sqrt{50} = 7.071$

4. D

5. A

6. C

7. D

8. D

9. A

10. C

11. C

12. B

Uppgift 1

Det finns tre sätt att få en större före en mindre; $(2, 1)$, $(3, 1)$ samt $(3, 2)$. På grund av symmetri, så är alla dessa lika sannolika, så det räcker att räkna ut sannolikheten för att få en 2a följt av en 1a. Att dra en tvåa sker med sannolikhet $2/6$, och en etta därefter sker med sannolikhet $2/5$. Så, $4/30$ för $(2, 1)$. Totala sannolikheten är då $3 \cdot 4/30 = 2/5 = 0.4$.

Uppgift 2

Vi har att $F_X(x) = \int x^2/9 \; dx = x^3/27$. Detta ger att $P(X > 2) = 1 - 8/27 = 19/27$, samt $P(X > 1) = 1 - 1/27 = 26/27$. Nu, $P(X > 2|X > 1) = P(X > 2)/P(X > 1)$, och den betingade sannolikheten blir $(19/27)/(26/27) = 19/26 = 0.73$.

Uppgift 3

Vi beräknar variansen, och utnyttjar att $V(aX + bY + cZ + d) = |a|^2V(X) + |b|^2V(Y) + |c|^2V(Z)$. $V(2X - 4Y + 3Z) = 4V(X) + 16V(Y) + 9V(Z) = 16 + 16 + 18 = 50$. Standardavvikelsen blir slutligen $\sqrt{50} \approx 7.07$.

Uppgift 4

Lagen om total sannolikhet ger att $P(X = 0) = 0.3$, $P(Y = 0) = 0.4$. Från detta kan vi sedan beräkna att $E[X] = 0.7$, $E[Y] = 0.6$. Formeln för kovarians ger nu att

$$Cov(X, Y) = E[(X - 0.7)(Y - 0.6)] = E[XY] - 0.6E[X] - 0.7E[Y] + 0.42.$$

Vi ser att $E[XY] = P(X = 1, Y = 1) = 0.4$, så

$$Cov(X, Y) = 0.4 - 0.6 \cdot 0.7 - 0.7 \cdot 0.6 + 0.42 = -0.02.$$

Uppgift 5

Låt $\sigma = D(X - 5)$. Vi har att $Y = (X - 5)/\sigma$ är $N(0, 1)$. Villkoret $P(X < 0) = 0.05$ ger att $P((X - 5)/\sigma < -5/\sigma) = 0.05$. Därför har vi att $P(Y < -5/\sigma) = 0.05$. För en $N(0, 1)$-normalfördelning gäller det att $P(Y < 1.6449) = 0.95$ (se tabellen), och därför $P(Y < -1.6449) = 0.05$. Allstå har vi att $1.6449 = 5/\sigma$ så $\sigma = 5/1.6449 = 3.04$.

Uppgift 6

Vi har: $P(Z \geq 2) = 1 - P(Z \leq 1) = 1 - P(\max(X, Y) \leq 1) = 1 - P(X \leq 1, Y \leq 1) =$ |X och Y är oberoende och likafördelade Po(2)| $= 1 - \big(P(X \leq 1)\big)^2 =$ |Tabell 5 för Po(2)| $= 1 - (0.40601)^2 = 0.835$

Uppgift 7

Vi deriverar Fördelningsfunktionen och får då täthetsfunktionen

$$f_{X_i}(x) = \begin{cases} ax^{a-1}, & 0 \leq x \leq 1, \\ 0, & \text{annars.} \end{cases}$$

Likelihoodfunktionen blir då $L(\alpha) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3) =$
$= \alpha x_1^{\alpha-1}\alpha x_2^{\alpha-1}\alpha x_3^{\alpha-1} = \alpha 0.873^{\alpha-1}\alpha 0.910^{\alpha-1}\alpha 0.820^{\alpha-1} =$
$= \alpha^3(0.873 \cdot 0.910 \cdot 0.820)^{\alpha-1}$

Logaritmering ger
$lnL(\alpha) = 3ln\alpha + (\alpha - 1)ln(0.873 \cdot 0.910 \cdot 0.820)$
Vi deriverar och sätter derivatan lika med 0.

$$\frac{dlnL(\alpha)}{d\alpha} = \frac{3}{\alpha} + ln(0.873 \cdot 0.910 \cdot 0.820) = 0$$

Vi får då

$$\alpha = \frac{-3}{ln(0.873 \cdot 0.910 \cdot 0.820)} \approx 7$$

Uppgift 8

Vi har: $\mu^*_{obs} = x$, dvs $\mu^* = X$, där $X \in Po(\mu)$.
$V(\mu^*) = V(X) = \mu$,
$D(\mu^*) = \sqrt{V(\mu^*)} = \sqrt{V(X)} = \sqrt{\mu}$.
Medelfelet $d(\mu^*)$ för μ^* är en skattning av $D(\mu^*) = \sqrt{\mu}$, dvs $d(\mu^*) = \sqrt{\mu^*_{obs}} = \sqrt{x}$.

Uppgift 9

Låt X vara antal kronor på n myntkast, $X \in Bin(n, p)$, och vi observerar $x = 20$. Konfindensintervallet för p blir: $I_p(0.9) = \left(p^*_{obs} \pm \lambda_{0.05}\sqrt{p^*_{obs}(1 - p^*_{obs})/n)}\right)$, där $n = 30$, $p^*_{obs} = x/n = 20/30$, $\lambda_{0.05} = 1.6449$ från Tab 2. Nedre gränsen fås då till 0.525.

Uppgift 10

Vi har: $h(7) = P(\text{förskasta nollhypoptes}|H_1) = P(X \geq 7|X \in Po(7))$. Från Tabell 5 för $\mu = 7$, den sökta styrkan blir:

$$h(7) = P(X \geq 7|X \in Po(7)) = 1 - P(X \leq 6) = 1 - 0.44971 \approx 0.55.$$

Uppgift 11

Enligt t-metoden (eftersom σ är okänd): $I_\mu(0.99) = \left(\bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right)$, där $n = 6$, medelvärdet $\bar{x}$ beräknas till 5.615, stickprovs standardavvikelse $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{6}(x_i - \bar{x})^2}$ beräknas till ca 0.542, och $t_{\alpha/2}(n-1) = t_{0.005}(5) = 4.03$ från Tabell 3.
Intervallets övre gräns blir då: $5.615 + 4.03 * 0.542/\sqrt{6} \approx 6.5067$

Uppgift 12

Om man gör en skiss på datamaterialet så ser man att den skattade lutningen (dvs β^*_{obs}) ska vara negativ i den linjära regressions modellen. Det finns endast ett svarsalternativ med en negativ lutning.

Del II

Uppgift 13

a) Låt A vara händelsen *tärningen visar 5*, och låt D_6, D_{12} och D_{20} vara händelserna att Alice plockade respektive tärning. Vi har

$$P(A|D_6) = 1/6, \quad P(A|D_{12}) = 1/12, \quad P(A|D_{20}) = 1/20.$$

Vidare, $P(D_6) = 1/2$, $P(D_{12}) = P(D_{20}) = 1/4$. Bayes sats ger då att

$$\begin{aligned} P(D_6|A) &= \frac{P(A|D_6)P(D_6)}{P(A|D_6)P(D_6) + P(A|D_{12})P(D_{12}) + P(A|D_{20})P(D_{20})} \\ &= \frac{(1/6)(1/2)}{(1/6)(1/2) + (1/12)(1/4) + (1/20)(1/4)} = \frac{5}{7} = 0.71. \end{aligned}$$

Uppgift 14

Inkomsten för Alice under vecka i, X_i, har egenskaperna

$$\begin{aligned} E(X_i) &= 0.3 \cdot 800 + 0.4 \cdot 1000 + 0.3 \cdot 1500 = 1090, \\ E(X_i^2) &= 0.3 \cdot 800^2 + 0.4 \cdot 1000^2 + 0.3 \cdot 1500^2 = 1267000, \\ D(X_i) &= \sqrt{E(X^2) - E(X)^2} = 280.9. \end{aligned}$$

Liknande för Bob, Y_i, ger $E(Y_i) = 830$, $D(Y_i) = 272.2$. Deras totala inkomst under 40 veckor, $\sum_{i=1}^{40}(X_i + Y_i)$ kan (med centrala gränsvärdessatsen) då approximeras med en normalfördelning,

$$Z = N(40 \cdot 1090 + 40 \cdot 830, \sqrt{40 \cdot 280.9^2 + 40 \cdot 272.2^2}) = N(76800, 2473.86).$$

Sannolikheten $P(Z > 80000) = 1 - P(Z < 80000)$. Nu,

$$P(Z < 80000) = P\left(\frac{Z - 76800}{2473.86} < \frac{80000 - 76800}{2473.86}\right) = P\left(\frac{Z - 76800}{2473.86} < 1.29\right) = \Phi(1.29) = 0.90.$$

Alltså är $P(Z > 80000) = 0.1$, så med 10% sannolikhet överstiger inkomsten 80000.

Uppgift 15

Detta är stickprov i par. Vi bildar $Z_i = Y_i - X_i$, och har att $\bar{z} = 11.2$ samt stickprovsvarians $s = 2.68$, $d = 2.68/\sqrt{5} = 1.1985$. Ett tvåsidigt konfidensintervall ges då av

$$(\bar{z} \pm d \cdot t_{0.05/2}(4)) = (11.2 \pm 1.1985 \cdot 2.78) = (7.87, 14.58).$$

Eftersom intervallet inte innehåller 0, så föreligger en skillnad på 95% konfidensgrad.

Uppgift 16

Enligt §9.2 i formelsamlingen fås Minstakvadrat-skattningarna av θ_1 och θ_2 genom att minimera kvadratsumman $Q = Q(\theta_1, \theta_2) = \sum_{i=1}^{n}(x_i - \mu_i(\theta_1, \theta_2))^2 =$

$$(10 - 100 \cdot \theta_1)^2 + (20 - 100 \cdot \theta_2)^2 + (15 - 100 \cdot \frac{(1 - \theta_1 - \theta_2)}{4})^2 +$$

$$+(25 - 100 \cdot \frac{(1-\theta_1-\theta_2)}{4})^2 + (21 - 100 \cdot \frac{(1-\theta_1-\theta_2)}{4})^2 + (9 - 100 \cdot \frac{(1-\theta_1-\theta_2)}{4})^2 =$$

$$= (10 - 100 \cdot \theta_1)^2 + (20 - 100 \cdot \theta_2)^2 + (-10 + 25\theta_1 + 25\theta_2)^2 +$$
$$+(25\theta_1 + 25\theta_2)^2 + (-4 + 25\theta_1 + 25\theta_2)^2 + (-16 + 25\theta_1 + 25\theta_2)^2$$

Vi ska nu minimera Q med avseende på θ_1 och θ_2.
Minimum fås där $gradQ = (\frac{\partial Q}{\partial \theta_1}, \frac{\partial Q}{\partial \theta_2}) = (0, 0)$
Vi får alltså ett ekvationssystem med två ekvationer och två obekanta.

$$\frac{\partial Q}{\partial \theta_1} = 2(10 - 100 \cdot \theta_1)(-100) + 2(-10 + 25\theta_1 + 25\theta_2)25 + 2(25\theta_1 + 25\theta_2)25 +$$

$$+2(-4 + 25\theta_1 + 25\theta_2)25 + 2(-16 + 25\theta_1 + 25\theta_2)25 = 0$$

$$\frac{\partial Q}{\partial \theta_2} = 2(20 - 100 \cdot \theta_2)(-100) + 2(-10 + 25\theta_1 + 25\theta_2)25 + 2(25\theta_1 + 25\theta_2)25 +$$

$$+2(-4 + 25\theta_1 + 25\theta_2)25 + 2(-16 + 25\theta_1 + 25\theta_2)25 = 0$$

Lite förenklingar ger

$$-70 + 500\theta_1 + 100\theta_2 = 0$$

$$-110 + 100\theta_1 + 500\theta_2 = 0$$

Löser man ut θ_1 och θ_2 ur detta fås Minsta-kvadratskattningarna till

$$\theta^*_{1_{obs}} = 0.1$$

och

$$\theta^*_{2_{obs}} = 0.2$$

Detta medför att den givna fördelningen för tärningens resultat enligt MK-skattningen blir

$$(p_1, p_2, p_3, p_4, p_5, p_6) = (0.1, 0.2, \frac{0.7}{4}, \frac{0.7}{4}, \frac{0.7}{4}, \frac{0.7}{4})$$

Vi genomför nu test av given fördelning m.h.a §14.3 där nollhypotesen H_0 är att tärningens resultat följer den fördelning vi fått genom att ta fram θ_1 och θ_2 m.h.a. minsta-kvadratskattning.

$$Q = \sum_j^r \frac{(x_j - np_j)^2}{np_j} = \frac{(10-10)^2}{10} + \frac{(20-20)^2}{20} + \frac{(15 - \frac{70}{4})^2}{\frac{70}{4}} +$$

$$+\frac{(25 - \frac{70}{4})^2}{\frac{70}{4}} + \frac{(21 - \frac{70}{4})^2}{\frac{70}{4}} + \frac{(9 - \frac{70}{4})^2}{\frac{70}{4}} = 8.4$$

Som vi ser är alla $np_j \geq 5$, så χ^2-test är O.K.

$$\chi^2_\alpha(r - k - 1) = \chi^2_{0.05}(6 - 2 - 1) = 7.81$$

Eftersom $Q > \chi^2_{0.05}(6 - 2 - 1)$ kan vi förkasta H_0 på risknivån 5%.