Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1912/1914/1915/1916 SANNOLIKHETSTEORI OCH STATISTIK, ONSDAG 22 DECEMBER 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 den andra datorlaborationen behöver ej besvara uppgift 12, utan får tillgodoräkna sig denna uppgift (i svarsblanketten anges ordet Bonus). 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 den andra datorlaborationen får 3 bonuspoäng på del II.

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

Tre mynt har olika sannolikhet för att ge krona. Första myntet ger krona med sannolikhet 25%, andra myntet ger krona med sannolikhet 50% och tredje myntet ger krona med 80%. De tre mynten är oberoende av varandra. Två av mynten dras slumpmässigt utan återläggning. (dvs varje mynt har samma sannolikhet att väljas) och kastas. Vad är sannolikheten att de två mynten båda visar krona.

A: 29/40

B: 29/120

C: 1/10

D: 3/10

Uppgift 2

De kontinuerliga stokastiska variablerna X och Y har den simultana täthetsfunktionen

$$f_{X,Y}(x,y) = \begin{cases} \frac{3}{16}xy^2, & 0 \le x \le 2,\ 0 \le y \le 2, \\ 0, & \text{annars.} \end{cases}$$

Bestäm väntevärdet $E(X)$.

A: 4/3

B: 7/3

C: 1

D: 2

Uppgift 3

X, Y och Z är oberoende stokastiska variabler. X är exponentialfördelad med väntevärdet $E(X) = 0.5$, Y är normalfördelad med väntevärde $E(Y) = 2$ och standardavvikelse $D(Y) = 3$, och $Z \in Bin(10, 0.8)$. Bestäm variansen $V(2X + Y - 5Z)$.

Uppgift 4

För de två stokastiska variablerna X och Y gäller att

$$P(X = 0, Y = 0) = 0.10, \quad P(X = 1, Y = 0) = \alpha,$$
$$P(X = 0, Y = 1) = 0.20, \quad P(X = 1, Y = 1) = \beta,$$

där α och β är okända konstanter. Bestäm standardavvikelsen $D(X + 1)$. Svaret skall anges som ett numeriskt värde, dvs inte som en funktion av α och β.

Uppgift 5

Vid en verkstad utför man bilbesiktning av personbilar. Sannolikheten att en bil inte klarar besiktningen är 10%. I ett slumpmässigt urval av 5 bilar, vad är sannolikheten att minst 4 av dessa bilar klarar besiktningen?

A: 0.59

B: 0.66

C: 0.92

D: 0.99

Uppgift 6

För Bob är tiden mellan matchingar på en dejting-app exponentialfördelad med väntevärdet 4 timmar. Beräkna sannolikheten att Bob får vänta minst 8 timmar på nästa matchning.

A: $1.27 * 10^{-14}$

B: 0.082

C: 0.135

D: 0.223

Uppgift 7

En diskret stokastisk variabel X_i uppfyller

$$P(X_i = 0) = \frac{1}{4}, \quad P(X_i = 1) = \frac{3}{8} - \alpha, \quad P(X_i = 2) = \frac{3}{8} + \alpha,$$

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

A: $5/8$

B: $5/24$

C: $-5/24$

D: $-5/8$

Uppgift 8

I en undersökning samlar man in datapunkter som är från en $N(\mu, \sigma)$-fördelningen, där σ är okänd. Man fick de fem observationerna

$$4.12, \quad 3.87, \quad 3.76, \quad 4.05, \quad 4.18.$$

Från detta beräknas ett tvåsidigt konfidensintervall för μ, med konfidensgrad 80%. Det upptäcktes dock att det sista mätvärdet (dvs 4.18) var felaktigt antecknat och skulle vara 4.10. Hur påverkas konfidensintervallet om man ersätter det felaktiga mätvärdet 4.18 med det korrekta mätvärdet 4.10 ?

A: Stickprovets standardavvikelse s blir större och därmed intervallet blir längre.

B: Stickprovets standardavvikelse s blir större och därmed intervallet blir kortare.

C: Stickprovets standardavvikelse s blir mindre och därmed intervallet blir längre.

D: Stickprovets standardavvikelse s blir mindre och därmed intervallet blir kortare.

Uppgift 9

I Sveriges befolkning anser man att en person gillar låten *Barbie Girl* med Aqua med sannolikhet p. I en undersökning bland n tillfrågade, skattade man $p^*_{obs} = x/n = 0.2$, där x är antalet som gillade Barbie Girl. Man beräknade även ett 95%-igt konfidensintervall för p till $I_p = (0.15, 0.25)$. Hur många personer n ingick i undersökningen?

A: $n = 7$

B: $n = 64$

C: $n = 176$

D: $n = 246$

Uppgift 10

Låt $X \in Po(\mu)$. Vi testar nollhypotesen $H_0 : \mu = 3$. Mothypotesen antas vara $H_1 : \mu > 3$. Vi gör ett försök och får observationen $x = 7$. Bestäm P-värdet för testet.

A: 0.988

B: 0.401

C: 0.034

D: 0.012

Uppgift 11

Låt X och Y vara två oberoende s.v. sådana att $X \in Po(2\mu)$ och $Y \in Po(\mu)$. Låt x och y vara observationer på de stokastiska variablerna X och Y, respektive.

$$\mu^*_{obs} = \frac{x + 2y}{4}$$

och

$$\hat{\mu}_{obs} = \frac{x + y}{3}$$

är två skattningar av μ. Vilket av de fyra följande alternativen stämmer?

A: Båda skattningarna är lika effektiva, dvs $V(\mu^*) = V(\hat{\mu})$.

B: Man kan inte avgöra vilken av skattningarna som är effektivast, eftersom minst en av skattningarna inte är väntevärdesriktig.

C: μ^*_{obs} är effektivare än $\hat{\mu}_{obs}$

D: $\hat{\mu}_{obs}$ är effektivare än μ^*_{obs}

Uppgift 12

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

Reklam (kkr)	28	39	45	53	59
Försäljning (kkr)	315	335	340	350	352

Utifrån datamaterialet ovan skattas följande regressionsmodell (dvs modell med både linjär och kvadratisk term):

$$y_i = \alpha + \beta x_i + \gamma x_i^2 + \varepsilon_i,\ i = 1, \dots, 5,$$

där y_i = försäljning (kkr) beror av x_i = reklamkostnad (kkr) och ε_i betecknar slumpmässiga fel. Minsta-kvadrat-skattningarna av regressionskoefficienterna α, β och γ blev $\alpha^*_{obs} = 239.3$, $\beta^*_{obs} = 3.42$ respektive $\gamma^*_{obs} = -0.026$. Vidare har 99%-iga konfidensintervall $I_\beta(0.99)$ och $I_\gamma(0.99)$ för β respektive γ beräknats: $I_\beta(0.99) = (-2.45, 9.30)$, $I_\gamma(0.99) = (-0.09, 0.04)$.

Man kontrollerar vidare om den effekt som reklamkostnaden har på försäljningen är signifikant, dvs man testar $H_{0,\beta} : \beta = 0$ mot $H_{1,\beta} : \beta \neq 0$ samt $H_{0,\gamma} : \gamma = 0$ mot $H_{1,\gamma} : \gamma \neq 0$. Vidare beräknas två P-värden för de två testen.

Vilken slutsats kan man dra?

A: Båda P-värdena är mindre än 0.01

B: Båda P-värdena är större än 0.01

C: De båda termerna (dvs. både den linjära och den kvadratiska) i regressionsmodellen är signifikanta på risknivån 1%.

D: Inget av alternativen A-C är korrekt.

Var god vänd!

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 $1, \dots, 6$, $1, \dots, 12$, samt $1, \dots, 20$, 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. Hon kastar den 4 gånger. Vad är sannolikheten att minst två av kasten visar ettor? (10 p)

Uppgift 14

De oberoende stokastiska variablerna $X_1, \dots, X_5$ har täthetsfunktionen

$$f_X(x) = \begin{cases} \frac{1+2t}{2}(x^2)^t & \text{då } 0 \leq x \leq e^{\frac{ln2}{2t+1}} \\ 0 & \text{annars,} \end{cases}$$

där $t > -\frac{1}{2}$. Vi har observationerna $x_1 = 0.12$, $x_2 = 0.21$, $x_3 = 0.76$, $x_4 = 0.52$ samt $x_5 = 0.84$, från respektive $X_1, \dots, X_5$. Använd maximum-likelihood metoden för att skatta t från dessa data. (10 p)

Uppgift 15

För 20 år sedan, gjorde man en undersökning där man bad de tillfrågade att ange vilken chokladpralin i Aladdinasken som var godast. Resultaten blev att 20% svarade trillingnöt, 30% svarade nöttryffel, 10% svarade romrussin samt att 40% svarade höstnougat.
Vid årets undersökning, så tillfrågades 100 slumpmässigt utvalda personer om deras preferenser. Där svarade 25 personer att trillingnöt var godast, 37 personer att nöttryffel var godast, 7 personer att romrussin var godast samt 31 personer att höstnougat var godast.
Avgör på 5% signifikansnivå, om populationens preferenser har ändrats eller inte. Ange tydligt vilka hypoteser som antas samt slutsats. (10 p)

Uppgift 16

Alice slumpar ett tal $X_1 \in U(0,1)$. Hon fortsätter därefter att slumpmässigt plocka $X_2, X_3, \dots \in U(0,1)$ (oberoende), så länge det nya observerade talet x_i är större än (eller lika med) det föregående, x_{i-1}. Till sist har hon en lista av n oberoende observationer

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_{n-1} > x_n,$$

där n är utfallet av den diskreta stokastiska variabeln N. Beräkna väntevärdet $E(N)$, det vill säga det förväntade antalet tal som Alice slumpar. (10 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

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

Del I

Svar:

1. B

2. A

3. 50

4. 0.458

5. C

6. C

7. B

8. D

9. D

10. C

11. D

12. B

Uppgift 1

Det finns tre oberoende fall, var och en förekommer med sannolikhet $\frac{1}{3}$. Om vi plockar 25% och 50%-mynten, får vi krona på båda med sannolikhet $(1/4)(1/2) = 1/8$. På samma sätt, 25% och 80%-mynten ger $(1/4)(4/5) = 1/5$, och 50% och 80%-mynten ger $(1/2)(4/5) = 2/5$. Lagen om total sannolikhet ger att den sammanlagda sannolikheten för att båda slumpmässigt valda mynten visar krona är $(1/3)(1/8 + 1/5 + 2/5) = 29/120$.

Uppgift 2

$f_X(x) = \int_0^2 f_{X,Y}(x,y)dy = \int_0^2 \frac{3}{16}xy^2dy = (x/2)$. $E(X) = \int_0^2 x \cdot (x/2)dx = [x^3/6]_0^2 = 8/6 = 4/3$. Väntevärdet är alltså 4/3.

Uppgift 3

Vi vet från formelbladet att $V(X) = 0.5^2 = 0.25$, $V(Y) = 3^2$ och $V(Z) = 10 \cdot 0.8(1 - 0.8) = 1.6$. Detta ger att $V(2X + Y - 5Z) = 2^2V(X) + V(Y) + 5^2V(Z) = 4 \cdot 0.25 + 9 + 25 \cdot 1.6 = 50$.

Uppgift 4

Först, notera att $D(X + 1) = D(X) = \sqrt{V(X)}$. Nu, första kolonnen ger att $P(X = 0) = 0.3$, så detta medför att $P(X = 1) = 0.7$. Nu kan vi lätt beräkna variansen för X:

$$V(X) = E(X^2) - E(X)^2 = 0.7 - 0.7^2 = 0.21.$$

Alltså är $D(X) = \sqrt{0.21} = 0.458$.

Uppgift 5

Vi har att $p = 0.9$ är sannolikheten att en bil klarar besiktningen. Sannolikheten att minst fyra klarar sig är då

$$\binom{5}{4}p^4(1-p) + \binom{5}{5}p^5 = 5 \cdot 0.9^4 \cdot 0.1 + 0.9^5 \approx 0.92.$$

Uppgift 6

Exponentialfördelningen är minneslös, så frågan är samma som att beräkna sannolikheten att Bob får vänta i 8 timmar utan matchning, eller i 8 timmar efter den senaste matchingen. Vi vill alltså bestämma $P(X \geq 8)$ där X har fördelningsfunktionen $F(x) = 1 - e^{-x/4}$. Detta är samma som $1 - F(8) = 1 - (1 - e^{-8/4}) = e^{-2} \approx 0.135$.

Uppgift 7

Likafördelningen för X_1, X_2 och X_3 ger att väntevärdet $\mu = E(X_1) = E(X_2) = E(X_3)$:

$$\mu = 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) + 2 \cdot P(X_i = 2) = (\frac{3}{8} - \alpha) + 2 \cdot (\frac{3}{8} + \alpha) = \frac{9}{8} + \alpha,$$

där $i = 1, 2, 3$. Vi vill nu minimera summan $Q = Q(\alpha)$ av kvadraterna av skillnaderna av de faktiskt observerade datapunkterna:

$$Q(\alpha) = (x_1 - \mu_0(\alpha))^2 + (x_2 - \mu_1(\alpha))^2 + (x_3 - \mu_2(\alpha))^2 = (x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 = \sum_{k=1}^{3} (x_k - \frac{9}{8} - \alpha)^2.$$

För att minimera $Q(\alpha)$, deriverar vi $Q(\alpha)$ map α, sätter derivatan till 0, och sedan löser akvationen map α för att få den sökta MK-sakttningen α^*_{MK} av α. Vi har:

$$\frac{dQ(\alpha)}{d\alpha} = (-2) \sum_{k=1}^{3} (x_k - \frac{9}{8} - \alpha) = (-2) \cdot (\sum_{k=1}^{3} x_k - \frac{27}{8} - 3\alpha) = 0.$$

Vi löser denna ekvation och vi får:

$$\alpha^*_{MK} = \frac{x_1 + x_2 + x_3}{3} - \frac{9}{8} = (1 + 1 + 2)/3 - 9/8 = 4/3 - 9/8 = (32 - 27)/24 = 5/24.$$

Uppgift 8

Eftersom datapunkterna kommer närmare varandra, så minskar standardavvikelsen. Konfidensintervallet (vars längd är proportioneligt med standardavvikelsen) blir kortare.

Uppgift 9

Vi har ett tvåsidigt konfidensintervall, med konfidensgraden är 95%. Vidare, konfidensintervallets totala bredd är $0.25 - 0.15 = 0.1$. Vi måste lösa ut n från $0.1 = 2 \cdot \lambda_{0.025} \cdot \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{n}}$, där $p^*_{obs} = 0.2$. Detta ger $n = 246$.

Uppgift 10

P-värdet ges av $P(X \geq 7)$, där $X \in Po(3)$. Detta är samma som att beräkna $1 - P(X \leq 6)$. Tabell 5 ger att detta är $1 - 0.96649 \approx 0.034$.

Uppgift 11

Vi har:

$$V(\hat{\mu}) = \frac{1}{9}(V(X) + V(Y)) = \frac{2\mu + \mu}{9} = \frac{\mu}{3},$$
$$V(\mu^*) = \frac{1}{16}(V(X) + 4V(Y)) = \frac{2\mu + 4\mu}{16} = \frac{6\mu}{16} = \frac{3\mu}{8}.$$

Skattningen $\hat{\mu}_{obs}$ är effektivare än μ^*_{obs}, för att $V(\hat{\mu}) = \frac{\mu}{3} < \frac{3\mu}{8} = V(\mu^*)$.

Uppgift 12

De båda P-värden är större än 0.01. Detta eftersom både k.i. $I_\beta(0.99)$ och $I_\gamma(0.99)$ överlappar 0, vilket innebär att man **inte** kan förkasta $H_{0,\beta} : \beta = 0$, och **inte** heller $H_{0,\gamma} : \gamma = 0$, på signifikansnivån 1%. OBS! Nollhypotesen kan förkastas om och endast om P-värdet är **mindre** än signifikansnivån.

Del II

Uppgift 13

Låt B vara händelsen att Alice producerar minst två ettor. Vi har att $P(B) = P(B|D_6)P(D_6) + P(B|D_{12})P(D_{12}) + P(B|D_{20})P(D_{20})$. Nu, $P(B|D_6)$ beräknas genom en binomialfördelning, $P(B|D_6) = P(X \geq 2)$ där $X \in Bin(\frac{1}{6}, 4)$. Detta ger (med komplementhändelserna 0 eller 1 etta)

$$P(B|D_6) = 1 - \binom{4}{0}\left(\frac{5}{6}\right)^4 - \binom{4}{1}\left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right) = 0.132.$$

På samma sätt för de andra tärningarna,

$$P(B|D_{12}) = 1 - \binom{4}{0}\left(\frac{11}{12}\right)^4 - \binom{4}{1}\left(\frac{11}{12}\right)^3\left(\frac{1}{12}\right) = 0.037,$$

och

$$P(B|D_{20}) = 1 - \binom{4}{0}\left(\frac{19}{20}\right)^4 - \binom{4}{1}\left(\frac{19}{20}\right)^3\left(\frac{1}{20}\right) = 0.014.$$

Totalt har vi nu $P(B) = 0.5 \cdot 0.132 + 0.037 \cdot 0.25 + 0.014 \cdot 0.25 = 0.079$.

Uppgift 14

Vi har att

$$L(t) = f_X(x_1) \cdots f_X(x_5),$$

så $L(t) = \frac{(1+2t)^5}{2^5} \cdot (x_1^2 x_2^2 \cdots x_5^2)^t$. Logaritmering ger uttrycket

$$\ln(L(t)) = 5(\ln(1+2t) - \ln(2)) + t\ln(x_1^2 x_2^2 \cdots x_5^2).$$

Löser vi $d \ln(L(t))/dt = 0$ får vi ekvationen

$$2\frac{5}{1+2t} + \ln(x_1^2 x_2^2 \cdots x_5^2) = 0.$$

Insättning av de värden vi fick ger

$$2\frac{5}{1+2t} - 9.56726 = 0,$$

som har lösningen $t = 0.0226$. Genom att undersöka andraderivatan, kan man enkelt verifiera att detta är ett maximum. Dvs. vi skattar parametern som $t^* = 0.0226$.

Uppgift 15

Vi utför ett χ^2-test. Vi har att

$$Q = (25-20)^2/20 + (37-30)^2/30 + (7-10)^2/10 + (31-40)^2/40 = 5.8$$

och $\chi^2_{0.05}(3) = 7.81$, så noll-hypotesen, att ingen förandring skett, kan ej förkastas.

Uppgift 16

Sannolikheten att processen tar minst $n+1$ steg är $1/n!$, eftersom det finns $n!$ sätt att ordna talen $x_1, \ldots, x_n$, men bara ett av dessa sätt är växande (och alla sätt är lika sannolika).
Därmed gäller att

$$P(N = n+1) = \frac{1}{n!} - \frac{1}{(n+1)!} = \frac{(n+1)-1}{(n+1)!}$$

Det förväntade antalet steg är därmed

$$E(N) = \sum_{n=2}^{\infty} n \cdot p_N(n) = \sum_{n=1}^{\infty} (n+1) \cdot p_N(n+1) = \sum_{n=1}^{\infty} (n+1) \cdot \frac{n}{(n+1)!} = \sum_{n=1}^{\infty} \frac{1}{(n-1)!} = e \approx 2.72$$