Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1912/1914/1915/1916 SANNOLIKHETSTEORI OCH STATISTIK, ONSDAG 20 DECEMBER 2023 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 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 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

Låt A, B och C vara tre händelser sådana att: A och C är oberoende, B och C är oberoende, A och B är disjunkta, $P(A \cup C) = \frac{2}{3}$, $P(B \cup C) = \frac{3}{4}$ och $P(A \cup B \cup C) = \frac{11}{12}$. Bestäm $P(C)$.

A: 0.50

B: 0.33

C: 0.25

D: 0.65

Uppgift 2

Låt X och Y vara två oberoende stokastiska variabler med $E(X) = E(Y) = 0$ och $V(X) = V(Y) = \sigma^2$. Beräkna korrelationskoefficienten $\rho(X + 2Y, 4X - 3Y)$.

A: 0.8660

B: -1.4142σ

C: $-2\sigma^2$

D: -0.1789

Uppgift 3

Låt X vara en kontinuerlig stokastisk variabel med täthetsfunktion

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

Bestäm variansen $V(X^2)$.

A: 0.5

B: 0.083

C: 0.055

D: 0.197

Uppgift 4

Låt X och Y vara kontinuerliga stokastiska variabler med den tvådimensionella täthetsfunktionen

$$f_{X,Y}(x, y) = \begin{cases} cx^2y(1 + y) \text{ om } 0 \leq x \leq 3, 0 \leq y \leq 3, \\ 0 \text{ annars} , \end{cases}$$

där c är en reell konstant. Bestäm $P(1 \leq X \leq 2, 0 \leq Y \leq 1)$.

A: 0.007

B: 0.016

C: 0.008

D: 0.019

Uppgift 5

I ett lotterispel är det vinst på var femtionde lott i genomsnitt. Om en person köper 10 lotter, vad är sannolikheten att denna person vinner?

A: 0.215

B: 0.200

C: 0.183

D: 0.020

Uppgift 6

Enligt ett företag som säljer blandat julgodis till storkunder ska 25% av skumtomtarna som levereras vara blåa. Ur en mycket stor leverans dras slumpmässigt 10000 skumtomtar. Av dessa 10000 skumtomtar visar sig 2550 vara blåa. Antag att den påstådda andelen blåa i populationen på 25% är korrekt. Vad är då den approximativa sannolikheten att ur ett slumpmässigt urval på 10000 skumtomtar få 2550 eller fler som är blåa?

A: 0.02

B: 0.13

C: 0.16

D: 0.25

Uppgift 7

Vi har en cirkel vars omkrets vi kallar för θ. Vi gör först en mätning av radien och får observationen $x_1 = 4.2$. Sedan gör vi en mätning av diametern och får observationen $x_2 = 8.8$.
Bestäm minsta-kvadrat-skattningen av θ utgående från dessa två observationer.

A: 27.0

B: 27.2

C: 27.4

D: 28.9

Uppgift 8

En karriärrådgivare är intresserad av att undersöka lönerna för de som varit ute i arbetslivet i ett år efter det att de tagit examen på en viss högskola. Speciellt är karriärrådgivaren intresserad av att undersöka om det är någon skillnad i lön mellan män och kvinnor. Från ett slumpmässigt urval av 20 män fås en månadslön på i genomsnitt 42780 kr med en skattad standardavvikelse på 5426 kr. Från ett slumpmässigt urval av 12 kvinnor fås en månadslön på i genomsnitt 40136 kr med en skattad standardavvikelse på 4383 kr. Antag att observationerna i respektive stickprov är utfall av oberoende stokastiska variabler X_i och Y_j där $X_i \in N(\mu_x, \sigma_x)$ och $Y_j \in N(\mu_y, \sigma_y)$ där $\sigma_x \neq \sigma_y$. Bilda ett tvåsidigt konfidensintervall med den approximativa konfidensgraden 95% för skillnaden i månadslön mellan män och kvinnor ett år efter examen från högskolan och ange konfidensintervallets övre gräns.

A: 5527 kr

B: 5790 kr

C: 6080 kr

D: 6420 kr

Uppgift 9

Antalet samtal som under en dag inkommer till en informationsavdelning antas vara Poissonfördelat med väntevärde μ. En samtalsräkning visade att under 25 dagar hade 900 samtal inkommit till informationsavdelningen. Antalet samtal som rings olika dagar antas vara oberoende. Bestäm ett tvåsidigt konfidensintervall för μ med den approximativa konfidensgraden 95%.

A: 36 ± 0.39

B: 36 ± 0.47

C: 36 ± 1.97

D: 36 ± 2.35

Uppgift 10

x_1 och x_2 är observationer av två oberoende stokastiska variabler $X_1 \in Bin(n_1, p)$ respektive $X_2 \in Bin(n_2, p)$. Vi sätter skattningen av p till

$$p^*_{obs} = \frac{x_1 + x_2}{n_1 + n_2}.$$

För $n_1 = 100$ och $n_2 = 300$ fick vi utfallen $x_1 = 12$ och $x_2 = 28$.
Bestäm medelfelet för skattningen av p. Dvs. bestäm $d(p^*)$.

A: 0.015

B: 0.024

C: 0.123

D: 0.154

Uppgift 11

Antag att $X \in Exp(\lambda)$ där λ är intensiteten, dvs $E(X) = 1/\lambda$. Låt H_0 vara att $\lambda = 0.2$. Vi vill testa H_0 mot alternativet $H_1 : \lambda = 0.5$. Tänk efter om det är för stora eller för små värden på x vi i detta fall förkastar H_0 till förmån för H_1. Vi får en observation: $x = 3.2$.
Bestäm testets P-värde.

A: 0.202

B: 0.473

C: 0.527

D: 0.798

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):

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

där y_i = försäljning (kkr) beror av x_i = reklamkostnad (kkr) och ε_i betecknar slumpmässiga fel. Minsta-kvadrat-skattningarna av regressionskoefficienterna α, β och γ blev $\alpha^*_{obs} = 239.3$, $\beta^*_{obs} = 3.42$ respektive $\gamma^*_{obs} = -0.026$.
95%-iga konfidensintervall $I_\beta(0.95)$ och $I_\gamma(0.95)$ har beräknats, $I_\beta(0.95) = (0.88, 5.97)$ samt $I_\gamma(0.95) = (-0.055, 0.0035)$.
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 0.05, och därför är det bara en av de två termerna (dvs den kvadratiska) i regressionsmodellen *signifikant* på risknivån 5 %.

B: Ett av de två P-värdena är större än 0.05, och därför är det bara en av de två termerna (dvs den kvadratiska) i regressionsmodellen *signifikant* på risknivån 5 %.

C: Båda P-värdena är mindre än 0.05, och därför är det bara en av de två termerna (dvs den linjära) i regressionsmodellen *signifikant* på risknivån 5 %.

D: Ett av de två P-värdena är större än 0.05, och därför är det bara en av de två termerna (dvs den linjära) i regressionsmodellen *signifikant* på risknivån 5 %.

Del II

Uppgift 13

I en låda ligger fem likadana batterier. Vi vet att två fungerar och att tre inte fungerar. Hur många batterier måste man i genomsnitt testa för att veta vilka två som fungerar? (10 p)

Uppgift 14

Antag att X är en diskret stokastisk variabel med följande sannolikhetsfunktion:

$$p_X(x) = \begin{cases} 2(1-\theta)/3, & \text{if } x = 0, \\ (1-\theta)/3, & \text{if } x = 1, \\ 2\theta/3, & \text{if } x = 2, \\ \theta/3, & \text{if } x = 3, \end{cases}$$

där $0 \leq \theta \leq 1$ är en parameter. Låt $(3, 0, 2, 0, 3, 1)$ vara 6 oberoende observationer av X. Utgående från dessa observationer, vad blir då maximum-likelihood-skattningen av θ? (10 p)

Uppgift 15

Någon hävdar att mäklarfirma A värderar en viss typ av lägenheter högre än mäklarfirma B gör. För att få en uppfattning om detta låter man de båda firmorna ovetandes om varandra värdera 5 olika lägenheter med följande resultat vad gäller priset i miljoner kr.

Lägenhet nr	1	2	3	4	5
Mäklarfirma A	1.19	1.15	1.32	1.34	1.20
Mäklarfirma B	1.06	0.99	1.06	1.06	1.07

Resultat k, $k = 1, 2, 3, 4, 5$, från mäklarfirma A betraktas som ett utfall av $N(\mu_k + \Delta, \sigma_1)$ och resultat k från mäklarfirma B betraktas som ett utfall av $N(\mu_k, \sigma_2)$. Alla stokastiska variabler antas oberoende. Avgör på risknivån 5% om vi kan dra slutsatsen att mäklarfirma A systematiskt värderar lägenheterna högre än vad mäklarfirma B gör. Dina hypoteser och slutsatser skall vara tydligt angivna och motiverade. (10 p)

Uppgift 16

Sambandet mellan normalfördelningen och χ^2-fördelningen framgår av följande: om $X_1, X_2, \ldots, X_n$ är oberoende $N(0, 1)$, så gäller det att $\sum_{i=1}^{n} X_i^2 \in \chi^2(n)$, dvs χ^2-fördelad med n frihetsgrader. Härled täthetsfunktionen för $\chi^2(1)$-fördelningen. (10 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1912/SF1914/SF1915/SF1916
SANNOLIKHETSTEORI OCH STATISTIK,
ONSDAG 20 DECEMBER 2023 KL 8.00–13.00.

Del I, Svar

1. A

2. D

3. B

4. B

5. C

6. B

7. C

8. C

9. D

10. A

11. B

12. D

Del I, Lösningsförslag

Uppgift 1

Eftersom A and C är oberoende och B och C är oberoende, så har vi:

$$\frac{2}{3} = P(A \cup C) = P(A) + P(C) - P(A)P(C) \tag{1}$$

and

$$\frac{3}{4} = P(B \cup C) = P(B) + P(C) - P(B)P(C). \tag{2}$$

Eftersom A och B är disjunkta, har vi också

$$\begin{aligned} \frac{11}{12} &= P(A \cup B \cup C) = P(A \cup B) + P(C) - P((A \cup B) \cap C) \qquad (3) \\ &= P(A) + P(B) + P(C) - P((A \cap C) \cup (B \cap C)) \\ &= P(A) + P(B) + P(C) - (P(A \cap C) + P(B \cap C)) = \\ &= P(A) + P(B) + P(C) - P(A)P(C) - P(B)P(C). \end{aligned}$$

Om vi summerar(1) och (2) så får vi:

$$\frac{2}{3} + \frac{3}{4} = P(A) + P(B) + 2P(C) - P(A)P(C) - P(B)P(C)$$

Nu subtraherar vi ekvation (3) från (1)+(2) och får

$$\frac{17}{12} - \frac{11}{12} = P(C) = \frac{6}{12} = 0.50.$$

Uppgift 2

Vi har

$$\rho(X + 2Y, 4X - 3Y) = \frac{C(X + 2Y, 4X - 3Y)}{D(X + 2Y)D(4X - 3Y)}.$$

Först beräknar vi

$$C(X + 2Y, 4X - 3Y) = 4C(X, X) - 3C(X, Y) + 2 \cdot 4C(Y, X) - 2 \cdot 3C(Y, Y).$$

Eftersom$X$ och Y är oberoende så gäller att $C(X, Y) = C(Y, X) = 0$. Därmed fås

$$C(X + 2Y, 4X - 3Y) = 4C(X, X) - 2 \cdot 3C(Y, Y) = 4V(X) - 6V(Y) = -2\sigma^2.$$

Återigen, eftersom X and Y är oberoende , så har vi:

$$V(X + 2Y) = V(X) + 2^2V(Y) = 5\sigma^2, \text{ and } V(4X - 3Y) = 4^2V(X) + (-3)^2V(Y) = 25\sigma^2,$$

Alltså gäller att

$$\rho(X + 2Y, 4X - 3Y) = \frac{-2\sigma^2}{\sqrt{5\sigma^2}\sqrt{25\sigma^2}} = -\frac{2}{5\sqrt{5}} = -0.1789.$$

Uppgift 3

Vi vet att

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

$$V(X^2) = E((X^2)^2) - E(X^2)^2$$

där

$$E(X^4) = \int_0^1 x^4 \cdot 2x dx = 2\frac{x^6}{6}\Big|_0^1 = \frac{1}{3}$$

och

$$E(X^2) = \int_0^1 x^2 \cdot 2x dx = 2\frac{x^4}{4}\Big|_0^1 = \frac{1}{2}.$$

Därmed fås, $V(X^2) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12} \approx 0.083$

Uppgift 4

Vi vet att

$$\begin{aligned} 1 &= \int_0^3 \int_0^3 f_{X,Y}(x,y)dydx = \int_0^3 \int_0^3 cx^2y(1+y)dydx = \int_0^3 \int_0^3 c(x^2y + x^2y^2)dydx \\ &= \int_0^3 c\left(x^2\frac{y^2}{2} + x^2\frac{y^3}{3}\right)\Big|_{y=0}^{y=3} dx = c\left(\frac{9}{2} + 9\right)\int_0^3 x^2 dx = c\left(\frac{9}{2} + 9\right)\frac{x^3}{3}\Big|_{x=0}^{x=3} \\ &= c\left(\frac{9}{2} + 9\right)9 = c\left(\frac{27}{2}\right)9 = c\frac{243}{2}. \end{aligned}$$

Därmed gäller att $c = \frac{2}{243} = 0.008$. Alltså får vi

$$\begin{aligned} P(1 < X < 2, 0 < Y < 1) &= \int_1^2 \int_0^1 f_{X,Y}(x,y)dydx = \int_1^2 \frac{2}{243}\left(x^2\frac{y^2}{2} + x^2\frac{y^3}{3}\right)\Big|_{y=0}^{y=1} dx = \\ &= \frac{2}{243}\left(\frac{1}{2} + \frac{1}{3}\right)\frac{x^3}{3}\Big|_{x=1}^{x=2} = \frac{2}{243}\left(\frac{1}{2} + \frac{1}{3}\right)\left(\frac{2^3}{3} - \frac{1^3}{3}\right) = \\ &= \frac{2}{243} \cdot \frac{5}{6} \cdot \frac{7}{3} \approx 0.016. \end{aligned}$$

Uppgift 5

Låt A vara händelsen att en viss lott ger vinst. Då är

$$P(A^*) = 1 - \frac{1}{50} = \frac{49}{50}$$

sannolikheten att denna lott inte ger vinst.

Då blir sannolikheten att inte få vinst på någon av de tio lotterna

$$(P(A^*))^{10} = \left(\frac{49}{50}\right)^{10} = 0.817.$$

Alltså blir sannolikheten att åtminstone en av de tio lotterna ger vinst

$$1 - (P(A^*))^{10} = 1 - 0.817 = 0.183.$$

Uppgift 6

Låt X vara antalet blåa skumtomtar vi drar. Då gäller att $X \sim Bin(n, p)$ med $n = 10000$ och $p = 0.25$.

Vi vet att $Bin(n, p) \approx N(np, np(1-p))$ (notera att $np(1-p) = 10000 \cdot 0.25 \cdot 0.75 = 1875 > 10$ och $np = 2500$). Därmed gäller att $X \approx N(2500, \sqrt{1875})$. Därför, har vi nu:

$$P(X \geq 2550) = 1 - P(X \leq 2549) \approx 1 - P\left(Z \leq \frac{2549 - 2500}{\sqrt{1875}}\right) = 1 - P(Z \leq 1.13) = 1 - 0.87 = 0.13$$

Uppgift 7

MK-skattning: Se §9.2 i F.S.

$$Q = \sum_{i=1}^{2}(x_i - \mu_i(\theta))^2 = (4.2 - \frac{\theta}{2\pi})^2 + (8.8 - \frac{\theta}{\pi})^2$$

$$\frac{dQ}{d\theta} = 2(4.2 - \frac{\theta}{2\pi})(\frac{-1}{2\pi}) + 2(8.8 - \frac{\theta}{\pi})(\frac{-1}{\pi})$$

$$\frac{dQ}{d\theta} = 0 \Rightarrow 4.2 + 2 \cdot 8.8 = \frac{\theta}{2\pi} + \frac{2\theta}{\pi} = \frac{5\theta}{2\pi}$$

$$\Rightarrow \theta^*_{obs_{MK}} \approx 27.39 \approx 27.4$$

Uppgift 8

Eftersom standardavvikelserna är okända och olika fås enligt §12.3 följande approximativa konfidensintervall:

$$I_{\mu_x-\mu_y} = \bar{x} - \bar{y} \pm \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} \cdot \lambda_{\frac{\alpha}{2}} = 42780 - 40136 \pm \sqrt{\frac{5426^2}{20} + \frac{4383^2}{12}} \cdot \lambda_{0.025} =$$

$$= 42780 - 40136 \pm \sqrt{\frac{5426^2}{20} + \frac{4383^2}{12}} \cdot 1.96$$

Övre gränsen blir alltså 6080 kr.

Uppgift 9

Låt X vara antalet samtal som inkommer under 25 dagar. $X \in Po(25\mu)$. Observationen $x = 900$. X är appr. $N(25\mu, \sqrt{25\mu})$ om $25\mu > 15$. Då 25μ lämpligen skattas med 900 så torde villkoret vara uppfyllt med god marginal.
§12.3 ger då

$$\mu^*_{obs} \pm \lambda_{0.025} D^*(\mu^*)_{obs}$$

Detta ger konfidensintervallet

$$25\mu^*_{obs} \pm \lambda_{0.025}\sqrt{25\mu^*_{obs}} = 900 \pm 1.96 \cdot 30 = 900 \pm 58.8$$

för parametern 25μ och således fås konfidensintervallet

$$\mu^*_{obs} \pm \lambda_{0.025}\frac{\sqrt{25\mu^*_{obs}}}{25} = 36 \pm 2.35 = \underline{(33.6,\ 38.4)}$$

för den efterfrågade parametern μ.

Uppgift 10

$$V(p^*) = V\left(\frac{X_1 + X_2}{n_1 + n_2}\right) = \frac{1}{(n_1+n_2)^2}[V(X_1) + V(X_2)] = \frac{1}{(n_1+n_2)^2}[n_1p(1-p) + n_2p(1-p)] =$$

$$= \frac{1}{(n_1+n_2)^2}(n_1+n_2)p(1-p) = \frac{p(1-p)}{(n_1+n_2)} \Rightarrow D(p^*) = \sqrt{\frac{p(1-p)}{(n_1+n_2)}}$$

Alltså blir medelfelet

$$d(p*) = D^*(p^*_{obs}) = \sqrt{\frac{p^*_{obs}(1-p^*_{obs})}{(n_1+n_2)}} = \left[p^*_{obs} = \frac{x_1+x_2}{n_1+n_2} = \frac{12+28}{100+300} = \frac{1}{10}\right] =$$

$$= \sqrt{\frac{1}{10} \cdot \frac{9}{10} \cdot \frac{1}{400}} = \frac{3}{200} = 0.015$$

Uppgift 11

Definitionen av P-värde är P(förkasta H_0) om H_0 sann och om vi har fått ett visst värde på testvariabeln vilket är P(få det observerade värdet eller något extremare värde $|H_0$ sann).
Vi förkastar alltså H_0 till förmån för H_1 för avvikande låga värden på utfallet av X i vårt fall.

$$\Rightarrow P(X < 3.2) \text{ om } X \in Exp(0.2) = \int_0^{3.2} 0.2e^{-0.2x}dx = 1 - e^{-0.2 \cdot 3.2} = 0.473$$

Uppgift 12

$\beta = 0 \notin I_\beta(0.95)$, så vi kan förkasta att $\beta = 0$ på 5%-nivån. När vi kan förkasta H_0:$\beta = 0$ på 5%-nivån så vet vi att motsvarande P-värdet< 0.05. Därmed kan vi påstå att den *linjära termen är signifikant* på 5%-nivån.

$\gamma = 0 \in I_\gamma(0.95)$, dvs kan vi inte förkasta att $\gamma = 0$ på 5%-nivån. När vi inte kan förkasta H_0:$\gamma = 0$ på 5%-nivån så vet vi att motsvarande P-värdet> 0.05. Därmed kan vi påstå att den *kvadratiska termen INTE är signifikant* på 5%-nivån.

Alltså, D är det rätta alternativet.

Del II, Lösningsförslag

Uppgift 13

Låt händelsen F_i vara händelsen att batteri nr i fungerar och låt den stokastiska variabeln X vara antalet batterier vi måste testa.

$$P(X=2) = P(F_1 \cap F_2) = P(F_1)P(F_2|F_1) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}$$

$$\begin{aligned} P(X=3) &= P(F_1^* \cap F_2^* \cap F_3^*) + P(F_1^* \cap F_2 \cap F_3) + P(F_1 \cap F_2^* \cap F_3) = \\ &= \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} + \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} + \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = \frac{3}{10} \end{aligned}$$

Eftersom vi inte behöver testa alla fem batterierna måste då gälla att

$$P(X=4) = 1 - \frac{1}{10} - \frac{3}{10} = \frac{6}{10}$$

$$\Rightarrow E(X) = 2P(X=2) + 3P(X=3) + 4P(X=4) = \frac{2}{10} + \frac{9}{10} + \frac{24}{10} = 3.5$$

<u>Vi måste i genomsnitt testa 3.5 batterier för att veta vilka två som fungerar</u>

Uppgift 14

Med givna observationer $(3, 0, 2, 0, 3, 1)$ kommer Likelihoodfunktionen att bli:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{6} p_X(x_i;\theta) = P(X=3)P(X=0)P(X=2)P(X=0)P(X=3)P(X=1) \\ &= \left(\frac{2(1-\theta)}{3}\right)^2 \left(\frac{1-\theta}{3}\right) \left(\frac{2\theta}{3}\right) \left(\frac{\theta}{3}\right)^2. \end{aligned}$$

För att maximera $L(\theta)$, kan vi logaritmera:

$$\begin{aligned} \ln L(\theta) &= \sum_{i=1}^{6} \ln p_X(x_i;\theta) \\ &= 2\left(\ln\frac{2}{3} + \ln(1-\theta)\right) + \left(\ln\frac{1}{3} + \ln(1-\theta)\right) \\ &+ \left(\ln\frac{2}{3} + \ln(\theta)\right) + 2\left(\ln\frac{1}{3} + \ln(\theta)\right) \\ &= C + 3\ln(1-\theta) + 3\ln(\theta), \end{aligned}$$

där C är en konstant som inte beror av θ.

Genom att maximera $\ln L(\theta)$ får vi:

$$\frac{d\ \ln L(\theta)}{d\theta} = -\frac{3}{1-\theta} + \frac{3}{\theta} = 0,$$

vilket ger

$$\theta^*_{obs_{ML}} = 0.5.$$

Alternativ lösning: $L(\theta) = c \cdot \theta^3 \cdot (1-\theta)^3$
$\Rightarrow \frac{d}{d\theta}L(\theta) = c(3\theta^2 \cdot (1-\theta)^3 - 3(1-\theta)^2 \cdot \theta^3) = 0$
$\Rightarrow 1 - \theta - \theta = 0 \Rightarrow \theta^*_{obs_{ML}} = 0.5.$

Uppgift 15

Stickprov i par. De parvisa skillnaderna

Lägenhet nr	1	2	3	4	5
y_i : (Mäklarfirma A – Mäklarfirma B)	0.13	0.16	0.26	0.28	0.13

är utfall av $N(\Delta, \sigma)$-fördelade stokastiska variabler, där Δ skattas med $\overline{y} = 0.192$ och σ med $s = 0.072595$.
Hypoteserna blir
$H_0 : \Delta = 0$
$H_1 : \Delta > 0$

Vi gör då p.g.a.$H_1 : \Delta > 0$ ett ensidigt nedåt begränsat konfidensintervall för Δ, med risknivån $\alpha = 0.05$, då standardavvikelsen är okänd och får att

$$\begin{aligned} I_\Delta &= (\overline{y} - t_{0.05}(4) \cdot \frac{s}{\sqrt{5}}, \infty) \\ &= (0.192 - 2.13 \cdot \frac{0.072595}{\sqrt{5}}, \infty) \\ &= (0.123, \infty) \end{aligned}$$

Uppgift 16

Vi har $n = 1$. Definiera en ny s.v. $Y = \sum_{i=1}^n X_i^2 = X_1^2$, där $X_1 \in N(0,1)$.
Fördelningsfunktionen $F_Y(y)$, $y \geq 0$, blir:

$$F_Y(y) = P(Y \leq y) = |Y = X_1^2| = P(X_1^2 \leq y) = P(-\sqrt{y} \leq X_1 \leq \sqrt{y}) = F_{X_1}(\sqrt{y}) - F_{X_1}(-\sqrt{y}).$$

För $y < 0$, har vi $F_Y(y) = P(X_1^2 \leq y) = 0$.
Vidare, genom att derivera $F_Y(y)$ får vi täthetsfunktionen för $Y = X_1^2 \in \chi^2(1)$:

$$f_Y(y) = F'_Y(y) = \left(F_{X_1}(\sqrt{y}) - F_{X_1}(-\sqrt{y})\right)' = f_{X_1}(\sqrt{y})\frac{1}{2\sqrt{y}} + f_{X_1}(-\sqrt{y})\frac{1}{2\sqrt{y}} =$$

$$= \ |\text{ eftersom } X_1 \in N(0,1)| \ =$$

$$= \frac{1}{\sqrt{2\pi}}e^{-(\sqrt{y})^2/2}\frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}}e^{-(-\sqrt{y})^2/2}\frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}\,e^{-y/2},\ y \geq 0.$$

Alltså, täthetsfunktionen för $\chi^2(1)$-fördelningen är:

$$f(y) = f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}\,e^{-y/2}, & \text{om } \ y \geq 0, \\ 0, & \text{om } \ y < 0. \end{cases}$$