Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1920/SF1921 SANNOLIKHETSLÄRA OCH STATISTIK,
TISDAG 12 MARS 2024 KL 08.00–13.00

Examinator: Björn-Olof Skytt, tel 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). Detta gäller endast på den här tentan och vid omtentamen i juni 2023. 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 på det här ordinarie tentamenstillfället och omtentamenstillfället i juni 2023.

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

För händelserna A, B och C vet vi att $P(A) = 0.3$, $P(B) = 0.4$, $P(C) = 0.5$, $P(A|B) = 0.6$, och $P(C|A) = 0.2$. Vi vet också att händelserna B och C är disjunkta händelser.
Beräkna $P(A^* \cap B^* \cap C^*)$.

A: 0.1

B: 0.2

C: 0.3

D: 0.4

Uppgift 2

Låt X vara en stokastisk variabel med täthetsfunktionen $f_X(x)$. Låt $Y = \ln(2X)$. Bestäm täthetsfunktionen $f_Y(y)$ för Y.

A: $f_Y(y) = f_X(e^{y/2})$

B: $f_Y(y) = \frac{1}{2}e^{y/2} f_X(e^{y/2})$

C: $f_Y(y) = f_X(\frac{1}{2}e^y)$

D: $f_Y(y) = \frac{1}{2}e^y f_X(\frac{1}{2}e^y)$

Uppgift 3

En stokastisk variabel X har följande täthetsfunktion

$$f_X(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \le x \le 1 \\ \frac{2}{9}, & 1 < x \le 4 \\ 0, & x > 4. \end{cases}$$

Beräkna variansen $V(X)$.

A: 1.09

B: 1.19

C: 2.21

D: 4.87

Uppgift 4

I en bestickslåda med 18 bestick ligger knivar och gafflar huller om buller. Det finns lika många knivar som gafflar i lådan. En person sticker ned handen i lådan och tar upp 4 bestick på en gång. Vad är sannolikheten för att det är 2 knivar och 2 gafflar?

A: 0.296

B: 0.330

C: 0.375

D: 0.424

Uppgift 5

En basketspelare skjuter 10 straffkast. Antag att sannolikheten att spelaren sätter ett straffkast är 0.8 och att sannolikheten inte förändras för olika kast. Antag också att händelserna att sätta ett straffkast är oberoende av varandra. Bestäm sannolikheten att antalet satta straffkast överstiger 6 men ej 9.

A: 0.772

B: 0.860

C: 0.879

D: 0.967

Uppgift 6

$X \in N(\mu_x, \sigma_x)$ och $Y \in N(\mu_y, \sigma_y)$ där $\mu_x = 5$, $\sigma_x = 2$, $\mu_y = 8$, och $\sigma_y = 3$, och X och Y är oberoende stokastiska variabler.

Sök z så att $P(Y + 2X < z) = 0.10$

A: 9.78

B: 11.22

C: 11.59

D: 12.72

Uppgift 7

$X_1 \in Po(2\lambda)$ och $X_2 \in Po(3\lambda)$, där X_1 och X_2 är oberoende stokastiska variabler.

För att skatta λ kan man använda sig av

$\hat{\lambda}_{obs} = \frac{1}{2} \cdot (\frac{x_1}{2} + \frac{x_2}{3})$ och $\lambda^*_{obs} = \frac{x_1+x_2}{2+3}$

Vilket av nedanstående påståenden är sant?

A: λ^*_{obs} är den effektivaste skattningen av λ.

B: $\hat{\lambda}_{\mathrm{obs}}$ är den effektivaste skattningen av λ.

C: Bägge skattningarna är lika effektiva.

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

Uppgift 8

En stokastisk variabel X har följande fördelningsfunktion

$$F_X(x) = \begin{cases} x^{\theta+1}; & 0 \le x \le 1 \\ \\ 0; & annars \\ \end{cases}$$

Vi har gjort tre oberoende försök och fått utfallen $x_1 = 0.70$, $x_2 = 0.80$, och $x_3 = 0.50$. Bestäm Maximum-Likelihood-skattningen av θ.

A: 1.36

B: 2.14

C: 3.33

D: 4.83

Uppgift 9

Låt $\bar{x} = 65.0$, $\bar{y} = 96.0$, $s_x^2 = 56.5$, $s_y^2 = 82.1$, $n_x = 12$ och $n_y = 12$ vara givet och antag att $X_i \in N(\mu_x, \sigma)$ och $Y_i \in N(\mu_y, \sigma)$ samt att alla dessa stokastiska variabler är oberoende. Ange övre gränsen för det 99%-iga ensidigt uppåt begränsade konfidensintervallet $I_{\mu_y - \mu_x}$ för $\mu_y - \mu_x$.

A: 36.6

B: 38.9

C: 39.5

D: 44.4

Uppgift 10

Låt datamängden 6, 12, 12, 9, 7, 16, 10 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 nedre gränsen för det dubbelsidiga konfidensintervallet I_μ för μ som har den approximativa konfidensgraden 95%.

A: 2.67

B: 7.08

C: 7.72

D: 7.91

Uppgift 11

Antag att $X \in Exp(\lambda)$. Vi vill pröva nollhypotesen $H_0 : \lambda = 0.2$. Vi förkastar H_0 till förmån för mothypotesen $H_1 : \lambda = 0.6$ om vi får en observation:$x < 3$. Vi får en observation: $x = 1.2$. Bestäm testets styrka då $\lambda = 0.6$.

A: 0.22

B: 0.45

C: 0.51

D: 0.83

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 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 på 5% signifikansnivå?

A: P-värdet för testet $H_{0,\beta} : \beta = 0$ är mindre än 0.05, och P-värdet för testet $H_{0,\gamma} : \gamma = 0$ är större än 0.05 och därmed är den linjära termen signifikant medan den kvadratiska termen inte är signifikant .

B: P-värdet för testet $H_{0,\beta} : \beta = 0$ är större än 0.05, och P-värdet för testet $H_{0,\gamma} : \gamma = 0$ är mindre än 0.05 och därmed är den linjära termen signifikant medan den kvadratiska termen inte är signifikant .

C: P-värdet för testet $H_{0,\beta} : \beta = 0$ är större än 0.05, och P-värdet för testet $H_{0,\gamma} : \gamma = 0$ är mindre än 0.05 och därmed är den linjära termen inte signifikant medan den kvadratiska termen är signifikant .

D: P-värdet för testet $H_{0,\beta} : \beta = 0$ är mindre än 0.05, och P-värdet för testet $H_{0,\gamma} : \gamma = 0$ är större än 0.05 och därmed är den linjära termen inte signifikant medan den kvadratiska termen är signifikant .

Del II

Uppgift 13

En digitalt kommunikationssystem fungerar så att en sändare skickar en spänning som är antingen 0 volt (för en digital 0:a) eller 1.8 volt (för en digital 1:a). På vägen till mottagaren störs signalen av additivt normalfördelat brus med väntevärde 0 volt och standardavvikelse 0.45 volt. Den *mottagna* signalen X är alltså sådan att betingat att en 0:a har sänts är X normalfördelad med väntevärde 0 volt, och betingat att en 1:a har sänts är X normalfördelad med väntevärde 1.8 volt. Standardavvikelsen för inspänningen är 0.45 volt i båda fallen.

Beslutskretsen i mottagaren fungerar enligt följande. Om inspänningen är större än 0.9 volt fattas beslutet ”1:a sänd”, och om den är mindre fattas beslutet att ”0:a sänd”.

a) Bestäm den betingade felsannolikheten för att beslutskretsen tar beslutet ”0:a sänd” givet att en 1:a i själva verket sändes. (3 p)

b) Låt $p_0 = 0.4$ vara sannolikheten att en nolla är sänd och $p_1 = 0.6$ vara sannolikheten att en etta är sänd. Bestäm sannolikheten för att en etta faktiskt sändes givet att beslutskretsen tar beslutet ”1:a sänd”. (7 p)

Uppgift 14

Nedanstående visar avkastningarna, i procent, för aktiefonderna *Handelsbanken Sverigefond* och *SEB Sverigefond* för åren 2003–2010.
De numeriska värdena (i %, från fondsidorna är enligt följande tabell:

år	2003	2004	2005	2006	2007	2008	2009	2010
Handelsbanken	25.4	10.2	32.2	22.9	–3.8	–41.0	40.7	23.9
SEB	29.5	14.5	27.3	22.8	–6.2	–41.8	44.0	21.9

Det kan vara intressant att undersöka, speciellt eftersom båda fonderna har samma inriktning (sverigefonder), om det finns någon systematisk skillnad mellan deras förväntade avkastningar. Gör detta, utgående från de givna data, med ett test på signifikansnivån 0.05 i en modell baserad på normalfördelningsantaganden. Var noga med att ange vilken modell du arbetar med och vilka antaganden du gör. (10 p)

Uppgift 15

Till en servicestation anländer anrop enligt en poissonprocess med intensitet 2 anrop per minut. Dvs. tiderna mellan anrop är exponentialfördelade med intensitet $\lambda = 2$ per minut. Låt T vara den längsta tiden mellan två på varandra följande anrop fram till anrop nummer 401. Beräkna sannolikheten att T är större än 3 minuter. (10 p)

Uppgift 16

a) Beräkna Minsta-Kvadrat skattningen θ^*_{obs} av θ då man erhållit observationerna $x_1, x_2, \ldots, x_n$ på en stokastisk variabel med täthetsfunktion

$$f_X(x) = \begin{cases} \theta x^{\theta-1} & \text{om } 0 \leq x \leq 1 \\ 0 & \text{annars} \end{cases}$$

Beräkna även det numeriska värdet på θ^*_{obs} då man erhållit observationerna 0.15, 0.40, 0.28 och 0.70. (5 p)

b) Man erhöll $\theta^*_{obs} = 0.58$ baserat på $n = 100$ observationer. Ge ett approximativt 95 % konfidensintervall för väntevärdet på fördelningen i a) och beräkna från detta ett approximativt 95 % konfidensintervall för θ. (5 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG
TENTAMEN I SF1920/SF1921 SANNOLIKHETSLÄRA OCH STATISTIK,
TISDAG 12 MARS 2024 KL 8.00–13.00.

Del I, Svar

1. A
2. D
3. B
4. D
5. A
6. C
7. A
8. A
9. C
10. D
11. D
12. A

Del I, Lösningsförslag

Uppgift 1

För händelserna A, B och C vet vi att $P(A) = 0.3$, $P(B) = 0.4$, $P(C) = 0.5$, $P(A|B) = 0.6$, och $P(C|A) = 0.2$. Vi vet också att händelserna B och C är disjunkta händelser.

$P(A^* \cap B^* \cap C^*) = 1 - P(A \cup B \cup C)$.

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) =$
$=$ händelserna B och C är disjunkta händelser $= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) =$
$= P(A) + P(B) + P(C) - P(A|B)P(B) - P(C|A)P(A) =$
$= 0.3 + 0.4 + 0.5 - 0.6 \cdot 0.4 - 0.2 \cdot 0.3 = 0.9$

$\Rightarrow P(A^* \cap B^* \cap C^*) = 0.1$

Uppgift 2

$Y = \ln(2X)$
$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{dx}{dy}\frac{d}{dx} F_Y(y)$

$y = ln(2x) \Rightarrow e^y = 2x \Rightarrow x = \frac{1}{2}e^y \Rightarrow \frac{dx}{dy} = \frac{1}{2}e^y$

$F_Y(y) = P(Y \leq y) = P(\ln(2X) \leq y) = P(X \leq \frac{1}{2}e^y) = F_X(\frac{1}{2}e^y)$

$\Rightarrow f_Y(y) = \frac{dx}{dy}\frac{d}{dx} F_Y(y) = \frac{1}{2}e^y f_X(\frac{1}{2}e^y)$

Uppgift 3

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x^3 dx + \int_1^4 \frac{2}{9} x dx = \left[\frac{x^4}{4}\right]_0^1 + \frac{2}{9}\left[\frac{x^2}{2}\right]_1^4 = \frac{1}{4} + \frac{16}{9} - \frac{1}{9} = \frac{69}{36}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^4 dx + \int_1^4 \frac{2}{9} x^2 dx = \left[\frac{x^5}{5}\right]_0^1 + \frac{2}{9}\left[\frac{x^3}{3}\right]_1^4 = \frac{1}{5} + \frac{128}{27} - \frac{2}{27} = \frac{2}{10} + \frac{126}{27} =$$

$$= \frac{54 + 1260}{270} = \frac{1314}{270}$$

$$V(X) = E(X^2) - E^2(X) = \frac{1314}{270} - \left(\frac{69}{36}\right)^2 = 1.19$$

.

Uppgift 4

Låt X vara antalet knivar vi drar. Då tillhör X den Hypergeometriska fördelningen. Sannolikheten för två knivar (och därmed två gafflar) är därmed

$$P(X=2) = \frac{\binom{9}{2}\binom{9}{2}}{\binom{18}{4}} = \frac{\frac{9\cdot 8}{2\cdot 1}\cdot\frac{9\cdot 8}{2\cdot 1}}{\frac{18\cdot 17\cdot 16\cdot 15}{4\cdot 3\cdot 2\cdot 1}} = 0.42$$

Uppgift 5

Låt X vara Antalet satta straffkast. Vi söker $P(6 < X \leq 9) = p_X(7) + p_X(8) + p_X(9)$ där $X \in Bin(10, 0.8)$

För att kunna använda tabell 6 i F.S. vänder vi på det hela och inför Y som är antalet missade straffkast där $Y \in Bin(10, 0.2)$.

Nu söker vi $p_Y(1) + p_Y(2) + p_Y(3) = P(Y \leq 3) - P(Y \leq 0) =$ [se tabell 6] $= 0.87913 - 0.10737 = 0.772$

Uppgift 6

$X \in N(\mu_x, \sigma_x)$ och $Y \in N(\mu_y, \sigma_y)$ där $\mu_x = 5$, $\sigma_x = 2$, $\mu_y = 8$, och $\sigma_y = 3$, och X och Y är oberoende stokastiska variabler.

Sök z så att $P(Y + 2X < z) = 0.10$

Låt $Z = Y + 2X$. Då är Z en linjärkombination av två oberoende Normalfördelade stokastiska variabler och är då således också Normalfördelad.
$E(Z) = E(Y) + 2E(X) = 8 + 2\cdot 5 = 18$. $V(Z) = V(Y) + 4V(X) = 9 + 4\cdot 4 = 25$. Således gäller att $Z \in N(18, 5)$

$P(Z \leq z) = 0.10$.

$$\text{Vi gör om till } N(0,1) \Rightarrow P\left(\frac{Z - E(Z)}{D(Z)} \leq \frac{z - E(Z)}{D(Z)}\right) = 0.10$$

$$\Rightarrow P\left(\frac{Z - 18}{5} = W \leq \frac{z - 18}{5}\right) = 0.10$$

Symmetri ger då att

$$P\left(W \geq \frac{18 - z}{5}\right) = 0.10$$

Dvs.

$$\frac{18 - z}{5} = \lambda_{0.10} = 1.2816$$

$$\Rightarrow z = 18 - 5\cdot 1.2816 = 11.59$$

Uppgift 7

$X_1 \in Po(2\lambda)$ och $X_2 \in Po(3\lambda)$, där X_1 och X_2 är oberoende stokastiska variabler.

För att skatta λ kan man använda sig av

$\hat{\lambda}_{obs} = \frac{1}{2} \cdot (\frac{x_1}{2} + \frac{x_2}{3})$ och $\lambda^*_{obs} = \frac{x_1+x_2}{2+3}$

$E(\hat{\lambda}) = E(\frac{1}{2} \cdot (\frac{X_1}{2} + \frac{X_2}{3})) = \frac{1}{2} \cdot (\frac{2\lambda}{2} + \frac{3\lambda}{3}) = \lambda$

$E(\lambda^*) = E(\frac{X_1+X_2}{2+3}) = \frac{2\lambda+3\lambda}{2+3} = \lambda$

Således är båda skattningarna väntevärdesriktiga.

$V(\hat{\lambda}) = V(\frac{1}{2} \cdot (\frac{X_1}{2} + \frac{X_2}{3})) = [\text{oberoende}] = \frac{1}{4} \cdot (\frac{2\lambda}{4} + \frac{3\lambda}{9}) = \frac{5}{24} \cdot \lambda$

$V(\lambda^*) = V(\frac{X_1+X_2}{2+3}) = [\text{oberoende}] = \frac{2\lambda+3\lambda}{25} = \frac{5}{25} \cdot \lambda$

$V(\lambda^*) < V(\hat{\lambda})$. Alltså gäller att λ^*_{obs} är den effektivaste skattningen av λ.

Uppgift 8

En stokastisk variabel X har följande fördelningsfunktion

$$F_X(x) = \begin{cases} x^{\theta+1}; & 0 \le x \le 1 \\ \\ 0; & annars \\ \end{cases}$$

$$\Rightarrow f_X(x) = \begin{cases} (\theta+1) \cdot x^{\theta}; & 0 \le x \le 1 \\ \\ 0; & annars \\ \end{cases}$$

Vi har gjort tre oberoende försök och fått utfallen $x_1 = 0.70$, $x_2 = 0.80$, och $x_3 = 0.50$. Detta leder till att Likelihood-funktionen $L(\theta)$ blir

$$(\theta+1) \cdot 0.7^{\theta} \cdot (\theta+1) \cdot 0.8^{\theta} \cdot (\theta+1) \cdot 0.5^{\theta} = (\theta+1)^3(0.7 \cdot 0.8 \cdot 0.5)^{\theta} = (\theta+1)^3 \cdot 0.28^{\theta}$$

Vi ska maximera $L(\theta)$ m.a.p. θ, men det θ som maximerar $lnL(\theta)$ maximerar även $L(\theta)$.

$$lnL(\theta) = 3ln(\theta+1) + \theta ln0.28$$

$$\frac{d}{d\theta} lnL(\theta) = \frac{3}{\theta+1} + ln0.28 = 0 \Rightarrow \theta^*_{obs_{ML}} = -\frac{3}{ln0.28} - 1 = 1.36$$

(Andra-derivatan är uppenbart negativ. Alltså har vi maximum.)

Uppgift 9

Låt $\bar{x} = 65.0$, $\bar{y} = 96.0$, $s_x^2 = 56.5$, $s_y^2 = 82.1$, $n_x = 12$ och $n_y = 12$ vara givet och antag att $X_i \in N(\mu_x, \sigma)$ och $Y_i \in N(\mu_y, \sigma)$ samt att alla dessa stokastiska variabler är oberoende. Ange övre gränsen för det 99%-iga ensidigt uppåt begränsade konfidensintervallet $I_{\mu_y - \mu_x}$ för $\mu_y - \mu_x$. Standardavvikelserna antas vara okända och lika.

$$s = [\text{Se § 11.2b}] = \sqrt{\frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{(n_x-1)+(n_y-1)}} = \sqrt{\frac{11 \cdot 56.5^2 + 11 \cdot 82.1^2}{11+11}} = \sqrt{69.3}$$

$$I_{\mu_y-\mu_x} = \left(-\infty, \overline{y} - \overline{x} + t_\alpha(f)\, s\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}\right) = \left(-\infty, 96.0 - 65.0 + t_{0.01}(11+11-2)\, s\sqrt{\frac{1}{12} + \frac{1}{12}}\right)$$

$$I_{\mu_y-\mu_x} = \left(-\infty, 31.0 + 2.51\sqrt{\frac{69.3}{6}}\right) = (-\infty, 39.5)$$

Uppgift 10

Eftersom väntevärdet $E(X) = \mu$ så skattar vi μ med $\bar{x} = \frac{6+12+1+9+7+16+10}{7} = \frac{72}{7}$

Eftersom summan av oberoende Poissonfördelade variabler
är Poissonfördelad gäller i vårt fall att $\sum X_i \in Po(7\mu)$.

Eftersom $7\mu^*_{obs} = 72 \geq 15$ så är $Y = \sum X_i$ approximativt Normalfördelad och då är även $\bar{X}$ det eftersom $\bar{X}$ då är en linjärkombination av en Normalfördelad stokastisk variabel. Alltså kan vi bilda ett konfidensintervall med approximativ konfidensgrad enligt § 12.3.

Då kan vi approximera $Po(\mu)$ till $N(\mu, \sqrt{\mu})$ enl §6 i Formelsamlingen och därmed bilda ett konfidensintervall med approximativ konfidensgrad enl. §12.3.

Enligt §12.3 blir då konfidensintervallet $I_\mu = \mu^*_{obs} \pm D^*_{obs}(\mu^*) \cdot \lambda_{\frac{\alpha}{2}}$.

$$\mu^*_{obs} = \bar{x} \Rightarrow D(\mu^*) = D(\bar{X}) = \frac{D(X)}{\sqrt{7}} = \frac{\sqrt{\mu}}{\sqrt{7}} \Rightarrow D^*_{obs}(\mu^*) = \frac{\sqrt{\bar{x}}}{\sqrt{7}}$$

Dvs. det dubbelsidiga konfidensintervallet blir

$$I_\mu = \bar{x} \pm \frac{\sqrt{\bar{x}}}{\sqrt{7}} \cdot \lambda_{0.025} = \frac{72}{7} \pm \frac{\sqrt{72}}{\sqrt{7 \cdot 7}} \cdot 1.96 \Rightarrow \text{nedre gränsen blir } 7.91$$

Uppgift 11

Styrkan hos testet är $P(\text{förkasta } H_0)$ om H_1 är sann.

Dvs. Vi ska räkna ut $P(X < 3)$ då $\lambda = 0.6$

$P(X < 3) = \int_{-\infty}^{3} f_X(x)dx = \int_0^3 0.6e^{-0.6x}dx = [-e^{-0.6x}]_0^3 = 1 - e^{-0.6\cdot 3} = 0.83$

Uppgift 12

Nollhypoteserna $H_{\beta,0} = 0$ och $H_{\gamma,0} = 0$ är ju att $\beta = 0$ respektive $\gamma = 0$.

$0 \notin I_\beta$ så vi förkastar att $\beta = 0$ på risknivån 5% och då antas att den linjära termen är signifikant. Eftersom vi kan förkasta nollhypotesen $\beta = 0$ på risknivån 5%, så vet vi också att P-värdet för testet $H_{\beta,0} = 0$ då är mindre än 0.05.

$0 \in I_\gamma$ så vi förkastar inte att $\gamma = 0$ på risknivån 5% och då antas att den kvadratiska termen inte är signifikant. Eftersom vi inte kan förkasta nollhypotesen $\gamma = 0$ på risknivån 5%, så vet vi också att P-värdet för testet $H_{\beta,0} = 0$ då är större än 0.05.

Alltså är A rätt svarsalternativ.

Del II, Lösningsförslag

Uppgift 13

Om en 1:a sändes är X N(1.8, 0.45) och

$$\begin{aligned} p_{10} &= P\left(X < 0.9\right) = P\left(\frac{X-1.8}{0.45} < \frac{0.9-1.8}{0.45}\right) = \Phi\left(\frac{0.9-1.8}{0.45}\right) = \Phi(-2) \\ &= 1-\Phi(2) = 0.0228. \end{aligned}$$

Detta är den sökta sannolikheten i a).
Vidare, $p_{11} = 1 - p_{10} = 0.9772$. Givet att en 0:a sändes är X N(0, 0.45) och

$$p_{00} = P\left(X < 0.9\right) = P\left(\frac{X-0}{0.45} < \frac{0.9-0}{0.45}\right) = \Phi\left(\frac{0.9-1.8}{0.45}\right) = \Phi(2) = 0.9772$$

och $p_{01} = 1 - p_{00} = 0.0228$. (Vi har samma felsannolikhet oavsett utsänt bit.) Givet är att

$$p_0 = P\left(0\text{:a sänd}\right) = 0.4 \qquad p_1 = P\left(1\text{:a sänd}\right) = 0.6.$$

Alltså,

$$\begin{aligned} P\left(1\text{:a sänd}|\text{beslut: "1:a sänd"}\right) &= \frac{P\left(1\text{:a sänd och beslut: "1:a"}\right)}{P\left(\text{beslut: "1:a"}\right)} \\ &= \frac{p_1 \cdot p_{11}}{p_0 \cdot p_{01} + p_1 \cdot p_{11}} \\ &= \frac{0.60 \cdot 0.9772}{0.40 \cdot 0.0228 + 0.60 \cdot 0.9772} = 0.9847. \end{aligned}$$

Uppgift 14

Låt $x_1, \ldots, x_8$ och $y_1, \ldots, y_8$ beteckna avkastningarna för Handelsbanken respektive SEB under de åtta åren. Eftersom det finns en kraftig samvaration (som är börsens generella utveckling under de olika åren) så är det lämpligt att arbeta med modellen *stickprov i par*. Med normalfördelningsantaganden blir alltså modellen att med $z_i = y_i - x_i$ är $z_1, \ldots, z_8$ oberoende observationer från $N(\mu, \sigma)$. Uppgiften är att undersöka om en systematisk skillnad finns, dvs om $\mu \neq 0$.
De $n = 8$ z-värdena är 4.1, 4.3, –4.9, –0.1, –2.4, –0.8, 3.3, –2.0, vilket ger $\sum z_i = 1.5$, $\sum z_i^2 = 80.61$, $\overline{z} = 0.1875$, $Q_z = \sum z_i^2 - n^{-1}(\sum z_i)^2 = 80.3288$, $s^2 = Q_z/(n-1) = 11.4755$, $s = 3.3876$.

Vi vill nu testa $H_0 : \mu = 0$ mot $H_1 : \mu \neq 0$. Under våra antaganden är $(\overline{z} - \mu)/(s/\sqrt{n})$ en observation från en t-fördelning med $n - 1 = 7$ frihetsgrader. Sätter vi här $\mu = 0$ får vi den observerade teststorheten 0.1566. Vi skall förkasta H_0 om detta tal är långt ute i svansarna på t-fördelningen, närmare bestämt om $|0.1566| > t_{0.025}(7) = 2.36$. Detta är uppenbarligen inte fallet, så H_0 kan inte förkastas på nivån 5%; vi kan inte hitta någon systematisk skillnad i förväntad avkastning mellan fonderna.

Alternativt så bildar vi ett tvåsidigt konfidensintervall för μ
och får enl. §12.2 $I_\mu = \bar{z} \pm \frac{s}{\sqrt{n}} t_{0.025}(7) = 0.1875 \pm \frac{3.3876}{\sqrt{8}} \cdot 2.36 = 0.1875 \pm 2.8266$

Vi konstaterar att $0 \in I_\mu$, så H_0 kan inte förkastas på nivån 5%; vi kan inte hitta någon systematisk skillnad i förväntad avkastning mellan fonderna.

Uppgift 15

Tiderna mellan anrop är exponentialfördelade med intensitet $\lambda = 2$ per minut. Om X är exponentialfördelad med intensitet λ så är

$$P(X > t) = \int_t^\infty \lambda \mathrm{e}^{-\lambda x}\, dx = \left[-\mathrm{e}^{-\lambda x}\right]_t^\infty = \mathrm{e}^{-\lambda t}$$

vilket med $\lambda = 2$ och $t = 3$ ger $p = P(X > 3) = \mathrm{e}^{-3\cdot 2} = \mathrm{e}^{-6} \approx 0.00248$. Vi söker

$$\begin{aligned} P(\max(X_1, \ldots, X_n) > t) &= 1 - P(\max(X_1, \ldots, X_n) \leq t) = 1 - P(X_1 \leq t \cap \ldots \cap X_n \leq t) \\ &= 1 - P(X_1 \leq t) \cdots P(X_n \leq t) = 1 - (1-p)\cdots(1-p) \\ &= 1 - (1-p)^n. \end{aligned}$$

Med $n = 400$ blir sannolikheten $1 - (1 - \mathrm{e}^{-6})^{400} \approx \underline{0.63}$.

Uppgift 16

Sätt fördelningens väntevärde till μ. Kvadratsumman $Q = \sum_{i=1}^n (x_i - \mu)^2$ skall minimeras. Derivatan $Q' = -2\sum_{i=1}^n (x_i - \mu)$ är 0 för $\mu = \mu^*_{obs} = \bar{x}$, vilket ger minimum. Eftersom $\mu = \int_0^1 x \cdot \theta x^{\theta-1} dx = \int_0^1 \theta x^\theta dx = \theta[\frac{x^{\theta+1}}{\theta+1}]_0^1 = \frac{\theta}{\theta+1}$ erhåller vi att MK-skattningen av θ uppfyller $\frac{\theta^*_{obs}}{\theta^*_{obs}+1} = \bar{x}$, vilket innebär att $\theta^*_{obs} = \frac{\bar{x}}{1-\bar{x}}$. Med siffror insatta erhålls $\bar{x} = 0.3825$ och alltså $\theta^*_{obs} = 0.6194$.

b) Om X har fördelningen i a) är $E(X^2) = \int_0^1 x^2 \theta x^{\theta-1} dx = \frac{\theta}{\theta+2}$ och variansen är $\sigma^2 = E(X^2) - \mu^2 = (\frac{\theta}{\theta+2}) - (\frac{\theta}{\theta+1})^2 = \frac{\theta}{(\theta+1)^2(\theta+2)}$. Enligt centrala gränsvärdessatsen är $\bar{X}$ approximativt $N(\mu, \sigma/\sqrt{n})$ och ett konfidensintervall för μ ges av $\bar{x} \pm \lambda_{\alpha/2} d$ där d är en skattning av standardavvikelsen $\sigma/\sqrt{n}$. Men skattningen av σ ges av $\sigma^* = \sqrt{\frac{\theta^*}{(\theta^*+1)^2(\theta^*+2)}}$. Eftersom $\bar{x} = \frac{\theta^*}{\theta^*+1}$ erhålls numeriskt att $\bar{x} = 0.3671$ och $\sigma^* = 0.3001$. Antalet observationer $n = 100$ och $\lambda_{0.025} = 1.96$. Detta ger att intervallet $0.3671 \pm 1.96 \cdot 0.3001/10 = 0.3671 \pm 0.0588 = (0.3083,\ 0.4259)$ är ett approximativt 95 % konfidensintervall för μ. Men att $0.3083 \leq \mu \leq 0.4259$ är ekvivalent med att $0.3083 \leq \frac{\theta}{\theta+1} \leq 0.4259$, d.v.s. $0.3083(1+\theta) \leq \theta \leq 0.4259(1+\theta)$ vilket i sin tur är ekvivalent med att $0.4457 \leq \theta \leq 0.7419$ varför detta sista intervall är ett approximativt 95 % konfidensintervall för θ.