Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1917/SF1918/SF1919 SANNOLIKHETSTEORI OCH STATISTIK, ONSDAGEN DEN 17:E APRIL 2019 KL 8.00–13.00.

Examinator för SF1918: Camilla Landén, 08-790 61 97

Examinator för SF1917/SF1919: Jörgen Säve-Söderbergh, 08-790 65 85.

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 på uppgifterna i Del I (dvs uppgifterna 1-12) skall anges på den bifogade svarsblanketten!**
Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1-3, utan får tillgodoräkna sig dessa tre uppgifter. Gränsen för godkänt är preliminärt 9 poäng. Möjlighet att komplettera ges för tentander med, preliminärt, 8 poäng. Tid och plats för komplettering kommer att anges på kursens hemsida.
Del II består av uppgifterna 13-16 och varje korrekt lösning ger 10 poäng. Del II rättas bara för studenter som är godkända på 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 4 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

För händelserna A och B gäller att $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$ och $B \subset A$. Beräkna $P(B \cap A^*)$.

Uppgift 2

En stokastisk variabel X har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0, & x < -1 \\ \dfrac{x+1}{2}, & -1 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

Beräkna $D(X)$.

A: 0.208

B: 0.577

C: 0.333

D: 0.828

Uppgift 3

Ur en låda julpynt med tre guldfärgade och fyra röda julgranskulor dras två kulor utan återläggning. Bestäm sannolikheten att man får en kula av varje färg.

A: 0.286

B: 0.375

C: 0.444

D: 0.571

Uppgift 4

Låt X och Y vara två stokastiska variabler sådana att $X \in \mathrm{Bin}\left(10, \frac{1}{5}\right)$ och $Y \in \mathrm{Bin}\left(5, \frac{1}{5}\right)$. Dessutom är X och Y oberoende. Beräkna $P(X + Y = 3)$.

Uppgift 5

Tiden mellan två kunder är exponentialfördelad med väntevärde två minuter. Hur stor är sannolikheten att det dröjer mer än en minut mellan två kunder?

A: 0.632

B: 0.607

C: 0.393

D: 0.368

Uppgift 6

Låt X och Y vara två stokastiska variabler sådana att $V(X) = 9$, $V(Y) = 4$ och $\rho(X,Y) = \frac{1}{2}$. Beräkna $C(3X - Y, Y)$.

Uppgift 7

Antag att $X_1, X_2, \ldots X_5$ är oberoende stokastiska variabler sådana att $X_i \in N(\mu, \sigma)$. Två skattningar av μ har föreslagits; $\theta^*_{\text{obs}} = x_1$ och $\hat{\theta}_{\text{obs}} = \frac{1}{5}\sum_{i=1}^{5} x_i$. Vilket av nedanstående påståenden är sant?

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

B: $\hat{\theta}_{\text{obs}}$ är den effektivaste skattningen av μ.

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

Tidsomställningen mellan sommar och vintertid kommer att avskaffas år 2021. EU:s medlemsländer skall själva bestämma om man vill ha permanent sommartid eller vintertid. Antag att en preliminär undersökning av vad man tycker i Sverige har gjorts och att av $n = 1000$ tillfrågade svarade $x = 636$ att de vill ha permanent sommartid. Man skattar därför andelen i Sverige som vill ha sommartid, p, med $p^*_{obs} = 636/1000$. Vilken fördelning har stickprovsvaribeln p^*?

A: $\text{Bin}\,(1, p)$

B: $\text{Bin}\left(\frac{1}{n}, p\right)$

C: Approximativt $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

D: Approximativt $N\left(p, \sqrt{\frac{p}{n}}\right)$

Uppgift 9

För att jämföra tillförlitligheten hos avgasreningen på nya bilar av tre olika fabrikat tog en motortidning slumpmässigt ut 100 bilar av vardera sorten och utsatte dessa för ett grundligt test. Resultat:

Fabrikat	Antalet felfria	Antal med fel
Vozda	81	19
Maab	85	15
Salvo	74	26

För att testa nollhypotesen H_0 : Bilmärkena skiljer sig ej åt beträffande avgasreningen, beräknar man teststorheten Q och får $Q = 3.875$. Vilken slutsats kan man dra då man fått denna teststorhet?

A: H_0 kan varken förkastas på risknivån 1% eller risknivån 5%.

B: H_0 kan både förkastas på risknivån 1% och risknivån 5%.

C: H_0 kan förkastas på risknivån 1% , men inte på risknivån 5%.

D: H_0 kan förkastas på risknivån 5% , men inte på risknivån 1%.

Uppgift 10

Vi har sex oberoende observationer 64, 70, 78, 84, 100 och 102 från en $N(\mu, \sigma)$-fördelning där det är känt att $\sigma = 15$. Beräkna ett 95%-igt tvåsidigt konfidensintervall för μ.

A: $(67.3, 98.7)$

B: $(70.6, 95.4)$

C: $(71.0, 95.0)$

D: $(73.9, 92.1)$

Uppgift 11

Vi har två stickprov från två populationer. Varje stickprov uppfattas som observationer på $N(\mu_i, \sigma_i)$, $i = 1, 2$ där vi antar att $\sigma_1 = \sigma_2 = \sigma$, samt att de bägge stickproven är oberoende av varandra. För de två stickproven har följande sammanfattande mått beräknats:
från stickprov 1

$$n_1 = 5 \quad \overline{x}_1 = 23.4 \quad s_1^2 = 5.3$$

från stickprov 2

$$n_2 = 5 \quad \overline{x}_2 = 27.0 \quad s_2^2 = 8.5$$

Beräkna övre gränsen av ett 95%-igt tvåsidigt konfidensintervall för $\mu_2 - \mu_1$.

A: 9.67

B: 7.44

C: 6.69

D: 7.35

Uppgift 12

En psykolog hade gjort ett test på sexton personer. Resultaten av testet uppfattas som observationer x_i på $N(\mu, 10)$-fördelningen. Psykologen önskar testa nollhypotesen $H_0 : \mu = 110$ mot $H_1 : \mu > 110$. Stickprovsmedelvärdet beräknat på de sexton observationerna blev $\bar{x} = 113.5$. Hjälp forskaren med hypotesprövningen genom att beräkna testets P-värde.

A: 0.081

B: 0.363

C: 0.637

D: 0.919

Del II

Uppgift 13

a) För en stokastisk variabel X gäller att

$$P(X > x) = \begin{cases} 1 & \text{om } x < 1, \\ 1/x^4 & \text{om } x \geq 1. \end{cases}$$

Detta innebär att X är Paretofördelad, vilket är ett rimligt antagande i olika ekonomiska tillämpningar.
Beräkna $E(X)$. (3 p)

b) Nollor och ettor sänds i en brusig miljö. Sannolikheten att en nolla respektive etta sänds är 0.4 respektive 0.6. Den mottagna signalen kan uppfattas som en stokastisk variabel X som är $\mathrm{N}(0, 1)$ respektive $\mathrm{N}(1, 1)$ om noll respektive ett sänds. Mottagaren använder regeln: om $X < 0.20$ anses en nolla ha sänts, annars har en etta sänts.
Beräkna sannolikheten att en mottagen nolla har sänts som en nolla. (7 p)

Uppgift 14

För att undersöka om två olika metoder A och B att klorera avloppsvatten ger olika resultat gjorde man på följande sätt. Under 8 olika dagar tog man prov på avloppsvattnet. Varje dags prov delades upp på två behållare. Kloreringsmetod A användes sedan på en slumpvis vald behållare av de två. Metod B användes på den andra. Man mätte därefter logaritmen av densiteten per ml av de koliforma bakterierna och fick följande värden:

Dag	1	2	3	4	5	6	7	8
A	2.8	3.1	3.4	3.0	2.7	2.9	3.5	2.6
B	3.2	3.1	2.9	3.5	2.4	3.0	3.2	2.8

Ange en lämplig statistisk modell baserad på normalfördelningen som beskriver data och undersök med hjälp av denna om det finns någon systematisk skillnad mellan metoderna. Använd 5% signifikansnivå. Ange tydligt de uppställda hypoteserna och motivera tydligt vilken slutsats du drar. (10 p)

Uppgift 15

En viss händelse inträffar med intensitet λ (h^{-1}). För att skatta λ observerar man under tre olika tidsperioder antalet gånger händelsen inträffar.

Observationstid t_i (h)	6	8	5
Antal händelser x_i	66	72	46

Antalet händelser x_i som inträffar i ett visst tidsintervall av längden t_i (enhet: timmar) kan anses vara $Po(\lambda t_i)$. Antalet händelser i skilda tidsintervall kan anses oberoende.

a) Bestäm Maximum-likelihood-skattningen av λ. (5 p)

b) Bestäm Minsta-kvadrat-skattningen av λ. (5 p)

Uppgift 16

En stokastisk variabel Y sägs vara *lognormalfördelad* om dess logaritm är normalfördelad, dvs den kan skrivas på formen $Y = e^X$, där $X \in N(\mu, \sigma)$. För att förenkla inför vi kodbeteckningen $Y \in Lognormal(\mu, \sigma)$.

a) Bestäm täthetsfunktionen för $Y \in Lognormal(\mu, \sigma)$. (3 p)

b) Beräkna väntevärdet för $Y \in Lognormal(\mu, \sigma)$. (4 p)

Ledning: För att beräkna en integral av typen

$$\int_{-\infty}^{\infty} e^x f_X(x)dx$$

där $X \in N(\mu, \sigma)$ kan man kvadratkomplettera exponenten.

c) Låt U och V vara två oberoende lognormalfördelade stokastiska variabler. Visa att $Z = U \cdot V$ också är lognormalfördelad och ange dess parametrar. (3 p)

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1917/SF1918/SF1919 SANNOLIKHETSTEORI OCH STATISTIK.
ONSDAGEN DEN 17:E APRIL 2019

Uppgift 1

$P(B \cap A^*) = P(B)\,P(A^* \mid B) = P(B)(1 - P(A \mid B)) = P(B)\left(1 - \frac{P(A\cap B)}{P(B)}\right) = P(B) - P(A \cap B)$, men då $B \subset A$ är $A \cap B = B$, så $P(B \cap A^*) = P(B) - P(B) = 0$.

Uppgift 2

Vi deriverar $F_X(x)$ för att få tätheten. $f_X(x) = F'_X(x) = \frac{1}{2}$, då $-1 \leq x \leq 1$.

$$E(X) = \int_{-1}^{1} x f_X(x)\,dx = \int_{-1}^{1} \frac{x}{2}\,dx = \left[\frac{x^2}{4}\right]_{-1}^{1} = \frac{1^2}{4} - \frac{(-1)^2}{4} = 0$$

$$E\left(X^2\right) = \int_{-1}^{1} x^2 f_X(x)\,dx = \int_{-1}^{1} \frac{x^2}{2}\,dx = \left[\frac{x^3}{6}\right]_{-1}^{1} = \frac{1^3}{6} - \frac{(-1)^3}{6} = \frac{1}{3}$$

Eftersom

$$\mathrm{Var}(X) = E\left(X^2\right) - (E(X))^2 = E\left(X^2\right) = \frac{1}{3}$$

Alltså blir $D(X) = \frac{1}{\sqrt{3}} = 0.577$

$$P\left(-\frac{1}{2} < X < \frac{1}{2}\right) = F_X\left(\frac{1}{2}\right) - F_X\left(-\frac{1}{2}\right) = \frac{\frac{1}{2}+1}{2} - \frac{-\frac{1}{2}+1}{2} = \frac{1}{2}$$

Uppgift 3

$$P(\text{en av varje färg}) = \frac{\binom{3}{1}\binom{4}{1}}{\binom{7}{2}} = \frac{3\cdot 4}{\frac{7\cdot 6}{2}} = \frac{4}{7} = 0.571$$

Uppgift 4

Då X och Y är oberoende och $X \in \mathrm{bin}\left(10, \frac{1}{5}\right)$ samt $Y \in \mathrm{bin}\left(5, \frac{1}{5}\right)$, uttalar additionsegenskapen att $X + Y \in \mathrm{bin}\left(15, \frac{1}{5}\right)$. Därmed blir

$$P(X+Y=3) = \binom{15}{3}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^{12} = 455\cdot\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^{12} = 0.250$$

Uppgift 5

Sätt $X =$ «tiden mellan två kunder» med täthetsfunktionen $f_X(x) = \frac{1}{2}e^{-\frac{x}{2}}$ för $x \geq 0$ så att

$$P(X>1)=\int_1^\infty f_X(x)dx=\int_1^\infty \frac{1}{2}e^{-\frac{x}{2}}dx=\left[-e^{-\frac{x}{2}}\right]_1^\infty=e^{-\frac{1}{2}}=0.607$$

Uppgift 6

Korrelationen definieras som

$$\rho(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Om vi löser ut $\mathrm{Cov}(X,Y)$ har vi

$$\mathrm{Cov}(X,Y)=\rho(X,Y)\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}=\frac{1}{2}\cdot 3\cdot 2=3$$

$$\mathrm{Cov}(3X-Y,Y)=3\,\mathrm{Cov}(X,Y)-\mathrm{Cov}(Y,Y)=3\,\mathrm{Cov}(X,Y)-\mathrm{Var}(Y)=3\times 3-4=5$$

Uppgift 7

Det gäller att

$$\begin{aligned}
E\left[\theta^*_{\mathrm{obs}}\right] &= E[X_1]=\mu\\
E\left[\hat{\theta}_{\mathrm{obs}}\right] &= E\left[\frac{1}{5}\sum_{i=1}^{5}X_i\right]=\frac{1}{5}\sum_{i=1}^{5}E[X_i]=\frac{1}{5}5\mu=\mu\\
V\left(\theta^*_{\mathrm{obs}}\right) &= V(X_1)=\sigma^2\\
V\left(\hat{\theta}_{\mathrm{obs}}\right) &= V\left(\frac{1}{5}\sum_{i=1}^{5}X_i\right)=\{\text{oberoende}\}=\frac{1}{5^2}\sum_{i=1}^{5}V(X_i)=\frac{1}{5^2}5\sigma^2=\frac{\sigma^2}{5}
\end{aligned}$$

Således är bägge skattningarna väntevärdesriktiga och eftersom $V(\hat{\theta}_{\mathrm{obs}})<V(\theta^*_{\mathrm{obs}})$ är $\hat{\theta}_{\mathrm{obs}}$ är effektivast.

Uppgift 8

Eftersom Sveriges befolkning är stor i förhållande till stickprovsstorleken n=1000 kan x=636 ses som en observation av $X\in Bin(n,p)$. Då $np^*_{obs}(1-p^*_{obs})\approx 232>>10$ kan normalapproximation av X göras och därmed blir även $p^*=X/n$ approximativt normalfördelad med parametrar

$$\begin{aligned}
E\left[\frac{X}{n}\right] &= \frac{1}{n}E[X]=\frac{1}{n}np=p\\
V\left(\frac{X}{n}\right) &= \frac{1}{n^2}V(X)=\frac{1}{n^2}np(1-p)=\frac{p(1-p)}{n}\\
D\left(\frac{X}{n}\right) &= \sqrt{V\left(\frac{X}{n}\right)}=\sqrt{\frac{p(1-p)}{n}}
\end{aligned}$$

Uppgift 9

Om H_0 är sann så är 3.875 ett utfall av en approximativt $\chi^2((3-1)(2-1))$-fördelad stokastisk variabel. Eftersom $\chi^2_{0.05}(2)=5.99>3.875$ så kan H_0 inte förkastas på nivån 5% och därmed inte heller på nivån 1%. Data ger alltså inte belägg för att bilmärkena skiljer sig åt i det undersökta avseendet. Detta innebär inte att man visat att bilmärkena är likvärdiga.

Uppgift 10

Eftersom $\overline{x} = 83.0$ och $\sigma = 15$, samt $n = 6$ blir intervallet

$$I_\mu = \left(\overline{x} \pm \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(83.0 \pm \underbrace{\lambda_{0.025}}_{1.96} \cdot \frac{15}{\sqrt{6}}\right) = (71.0, 95.0)$$

Uppgift 11

Eftersom $n_1 = 5$ och $s_1^2 = 5.3$, samt $n_2 = 5$ och $s_2^2 = 8.5$, så har vi

$$s^2 = \frac{Q_1 + Q_2}{(n_1 - 1) + (n_2 - 1)} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{4 \cdot 5.3 + 4 \cdot 8.5}{4 + 4} = 6.9$$

M.h.a §12.2 och §11.2d i Formelsamlingen fås sedan övre gränsen till

$$I_{\mu_2 - \mu_1} = \overline{x}_2 - \overline{x}_1 + s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) = 3.6 + \sqrt{6.9 \cdot 0.4} \cdot 2.31 = 7.44$$

Uppgift 12

$$\begin{aligned}
P\left(\overline{X} > 113.5\right) &= 1 - P\left(\overline{X} \leq 113.5\right) \\
&= 1 - P\left(\frac{\overline{X} - 100}{\frac{10}{\sqrt{16}}} \leq \frac{113.5 - 100}{\frac{10}{\sqrt{16}}}\right) \\
&= 1 - \Phi(1.40) \\
&= 1 - 0.9192 \\
&= 0.0808
\end{aligned}$$

Uppgift 13

a) Vi har

$$F_X(x) = \begin{cases} 0 & \text{om } x < 1, \\ 1 - x^{-4} & \text{om } x \geq 1. \end{cases}$$

vilket innebär att

$$f_X(x) = F_X'(x) = \begin{cases} 0 & \text{om } x < 1, \\ 4x^{-5} & \text{om } x \geq 1. \end{cases}$$

Detta ger

$$E(X) = \int_1^\infty x \cdot \frac{4}{x^5}\, dx = \int_1^\infty \frac{4}{x^4}\, dx = 4\left[\frac{1}{3x^3}\right]_1^\infty = \frac{4}{3}$$

b) Inför $S_0 =$ "nolla sänds", $S_1 =$ "etta sänds", $M_0 =$ "nolla mottages". Med Bayes formel erhålls den sökta sannolikheten:

$$P(S_0|M_0) = \frac{P(S_0)P(M_0|S_0)}{P(S_0)P(M_0|S_0) + P(S_1)P(M_0|S_1)}$$

$$= \frac{0.4P(X < 0.2 \text{ givet att } X \in N(0,1))}{0.4P(X < 0.2 \text{ givet att } X \in N(0,1)) + 0.6P(X < 0.2 \text{ givet att } X \in N(1,1))}$$

$$= \frac{1}{1 + 1.5\Phi(-0.8)/\Phi(0.2)} = 0.65.$$

Uppgift 14

Uppgiften handlar om jämförelse av väntevärden med *stickprov i par*.
Hypoteserna bör formuleras som $H_0 : \Delta = 0$ mot $H_1 : \Delta \neq 0$. Om det existerar en systematisk skillnad som vi fångar med parametern Δ, så kan Δ i medeltal vara positiv eller negativ.
Vi bildar differenser $z_i = y_i - x_i$. Vi betraktar z_i, $i = 1, \ldots, 12$, som utfall av oberoende $\mathrm{N}(\Delta, \sigma)$-fördelade stokastiska variabler. Δ skattas med $\overline{z} = 0.0125$ (se nedan) som är ett utfall av en $\mathrm{N}(\Delta, \sigma/\sqrt{n})$-fördelad stokastisk variabel $\overline{Z}$.
Testvariabel är

$$u = \frac{\overline{Z} - \Delta}{S/\sqrt{n}}$$

som är t-fördelad med $(n-1)$ frihetsgrader, om H_0 är sann.
Antalet frihetsgrader är $f = n - 1 = 7$. Testet är tvåsidigt. Då signifikansnivån är $\alpha = 0.05$, blir gränsen för det kritiska området $t_{\alpha/2}(n-1) = t_{0.025}(7) = 2.36$. Alltså, förkastar vi H_0 om $|u| > 2.36$.
Vi beräknar $\overline{z}$ och s från data:

Dag	1	2	3	4	5	6	7	8
A (x_i)	2.8	3.1	3.4	3.0	2.7	2.9	3.5	2.6
B (y_i)	3.2	3.1	2.9	3.5	2.4	3.0	3.2	2.8
$z_i = y_i - x_i$	0.4	0.0	-0.5	0.5	-0.3	0.1	-0.3	0.2
z_i^2	0.16	0.0	0.25	0.25	0.09	0.01	0.09	0.04

Summan av differenserna blir $\sum_{i=1}^n z_i = 0.1$. Då blir $\overline{z} = \frac{1}{n}\sum_{i=1}^n z_i = \frac{1}{8}\, 0.1 = 0.0125$. Vidare blir $\sum_{i=1}^n z_i^2 = 0.89$ och därmed har vi att

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^n z_i^2 - \frac{1}{n}\left(\sum_{i=1}^n z_i\right)^2\right] = \frac{1}{7}\left[0.89 - \frac{1}{8}\, 0.1^2\right] = 0.126964$$

Alltså blir $s = 0.3563$.
Därmed blir testvariabeln

$$|u| = \left|\frac{\overline{z}}{s/\sqrt{n}}\right| = \left|\frac{0.0125}{0.3563/\sqrt{8}}\right| = 0.10 < 2.36$$

Slutsatsen är att H_0 inte förkastas på 5% signifikansnivå.

Uppgift 15

a) Det gäller att x_1, x_2, x_3 är observationer av oberoende Po-fördelade stokastiska variabler X_1, X_2, X_3 med sannolikhetsfunktion

$$p_{X_i}(x_i;\lambda) = \frac{(\lambda t_i)^{x_i} e^{-\lambda t_i}}{x_i!}, \quad i = 1, 2, 3.$$

Likelihoodfunktionen ges av

$$L(\lambda) = p_{X_1}(x_1;\lambda)p_{X_2}(x_2;\lambda)p_{X_3}(x_3;\lambda) = \frac{t_1^{x_1} t_2^{x_2} t_3^{x_3} \lambda^{\sum_{i=1}^3 x_i} e^{-\lambda \sum_{i=1}^3 t_i}}{x_1! x_2! x_3!}.$$

ML-skattningen ges av det värde λ^*_{ML} som maximerar $L(\lambda)$. Logaritmering ger

$$\ln L(\lambda) = \ln(\lambda) \sum_{i=1}^3 x_i - \lambda \sum_{i=1}^3 t_i + \ln\left(\frac{t_1^{x_1} t_2^{x_2} t_3^{x_3}}{x_1! x_2! x_3!}\right).$$

Derivering m.a.p. λ ger

$$\frac{d \ln L(\lambda)}{d\lambda} = \frac{\sum_{i=1}^3 x_i}{\lambda} - \sum_{i=1}^3 t_i.$$

Denna derivata satt $= 0$ ger ML-skattningen $\underline{\lambda^*_{ML} = \frac{\sum_{i=1}^3 x_i}{\sum_{i=1}^3 t_i}}$. Med de erhållna värdena insatta så fås $\lambda^*_{ML} = \frac{66+72+46}{6+8+5} = \frac{184}{19} = \underline{9.684}$

b) Det gäller att x_1, x_2, x_3 är observationer av oberoende Po-fördelade stokastiska variabler X_1, X_2, X_3 med väntevärden

$$E[X_i] = \lambda t_i, \quad 1 = 1, 2, 3.$$

Vi får därför att

$$Q(\lambda) = \sum_{i=1}^3 (x_i - \lambda t_i)^2$$

MK-skattningen ges av det värde λ^*_{MK} som minimerar $Q(\lambda)$.
Derivering m.a.p. λ ger

$$\frac{dQ(\lambda)}{d\lambda} = -2 \sum_{i=1}^3 t_i (x_i - \lambda t_i)$$

Denna derivata satt $= 0$ ger MK-skattningen $\underline{\lambda^*_{MK} = \frac{\sum_{i=1}^3 t_i x_i}{\sum_{i=1}^3 t_i^2}}$. Med de erhållna värdena insatta så fås $\lambda^*_{MK} = \frac{6 \cdot 66 + 8 \cdot 72 + 5 \cdot 46}{6^2 + 8^2 + 5^2} = \frac{1202}{125} = \underline{9.616}$

Uppgift 16

a) Vi har att fördelningsfunktionen för Y ges av

$$F_Y(y) = P(Y \le y) = P(e^X \le y) = P(X \le \ln y) = F_X(\ln y).$$

Här krävs uppenbarligen att $y > 0$.
Genom att derivera får vi täthetsfunktionen för Y som

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\ln y) = f_X(\ln y) \cdot \frac{1}{y}$$

för $y > 0$. Ur formelsamlingen hämtar man att täthetsfunktionen för en normalfördelning ges av

$$f_X(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Insatt i uttrycket för $f_Y(y)$ ger detta att

$$f_Y(y) = \frac{1}{y} \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}$$

för $y > 0$, vilket är täthetsfunktionen $Y \in Lognormal(\mu, \sigma)$.

b) Låt $Y \in Lognormal(\mu, \sigma)$, dvs $Y = e^X$ där $X \in N(\mu, \sigma)$. Då gäller att

$$E[Y] = E[e^X] = \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Kvadratkomplettering av exponenten ger nu

$$\begin{aligned} & x - \frac{x^2 - 2\mu x + \mu^2}{2\sigma^2} = \frac{2\sigma^2 x - x^2 + 2\mu x - \mu^2}{2\sigma^2} = \\ & = \frac{-x^2 + 2(\mu + \sigma^2)x - \mu^2}{2\sigma^2} = \frac{-x^2 + 2(\mu + \sigma^2)x - (\mu + \sigma^2)^2 + 2\mu\sigma^2 + \sigma^4}{2\sigma^2} = \\ & = -\frac{[x - (\mu + \sigma^2)]^2}{2\sigma^2} + \mu + \frac{\sigma^2}{2} \end{aligned}$$

Således gäller att

$$E[Y] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{\mu+\sigma^2/2} e^{-\frac{[x-(\mu+\sigma^2)]^2}{2\sigma^2}} dx = e^{\mu+\sigma^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{[x-(\mu+\sigma^2)]^2}{2\sigma^2}} dx = e^{\mu+\sigma^2/2}$$

eftersom integralen är över täthetsfunktionen för $N(\mu+\sigma^2, \sigma^2)$. Väntevärdet för $Y \in Lognormal(\mu, \sigma)$ är alltså $e^{\mu+\sigma^2/2}$.

c) Vi vet att vi kan skriva U och V som

$$U = e^X, \qquad \text{och} \qquad V = e^Y,$$

där $X \in N(\mu_1, \sigma_1)$ och $Y \in N(\mu_2, \sigma_2)$ och X och Y är oberoende. Således har vi att

$$Z = U \cdot V = e^X \cdot e^Y = e^{X+Y}$$

och då linjärkombinationer av oberoende normalfördelade stokastiska variabler är normalfördelade är normalfördelade gäller att $X + Y$ är normalfördelad med paramtrar

$$\begin{aligned} E[X+Y] &= E[X] + E[Y] = \mu_1 + \mu_2 \\ V(X+Y) &= \{\text{oberoende}\} = V(X) + V(Y) = \sigma_1^2 + \sigma_2^2 \\ D(X+Y) &= \sqrt{V(X+Y)} = \sqrt{\sigma_1^2 + \sigma_2^2} \end{aligned}$$

dvs $Z \in Lognormal(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.