Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1912/1914/1915/1916 SANNOLIKHETSTEORI OCH STATISTIK, FREDAG 21 OKTOBER 2022 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 2022. 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 datorlaborationen får 3 bonuspoäng på del II på ordinarie tentamenstillfället och vid omtentamen i december 2022.

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

Bestäm $E(X^2)$ för den stokastiska variabel X som har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0 & \text{om} \quad x < 0 \\ \frac{x(6-x)}{8} & \text{om} \quad 0 \le x < 2 \\ 1 & \text{om} \quad x \ge 2 \end{cases}$$

A: 0.833

B: 1.00

C: 1.50

D: 2.20

Uppgift 2

Den tvådimensionella stokastiska variabeln (X, Y) har den simultana täthetsfunktionen:

$$f_{X,Y}(x,y) = \begin{cases} xe^{-xy} & \text{för} \quad x \geq 0 \text{ och } y \geq 1 \\ 0 & \text{annars} \end{cases}$$

Beräkna $P(Y < 3)$

A: 1/9

B: 1/3

C: 4/9

D: 2/3

Uppgift 3

Låt X, Y och Z vara stokastiska variabler sådana att $D(X) = D(Y) = D(Z) = 2$ och att Z är oberoende av X och Y. Bestäm $C(X, Y)$ om $D(2X - Y + 3Z) = 8$

A: -4

B: -2

C: 4

D: 5

Uppgift 4

En påse frysta piroger innehåller piroger med tre smaker: räkor (5 st), svamp (7 st) och köttfärs (4 st), men alla piroger ser likadana ut. Emma tycker bäst om räkpiroger och minst om köttfärspiroger. Om hon tar tre piroger på måfå ur påsen, hur stor är sannolikheten att hon får minst två räkpiroger och ingen köttfärspirog.

A: 0.018

B: 0.125

C: 0.143

D: 0.200

Uppgift 5

En trädgårdsbutik säljer bl.a. hyacintlökar. Dessvärre har en del lökar hamnat fel, så i ett stort parti lökar som ska ge röda hyacinter kommer 15 % av lökarna istället att ge blå hyacinter. Hur stor är sannolikheten att en påse med 12 lökar innehåller minst 10 lökar som ger röda hyacinter? Antag att lökarna väljs slumpmässigt ur det stora partiet.

A: 0.264

B: 0.292

C: 0.557

D: 0.736

Uppgift 6

George jämför två aktier. Han räknar med att aktiernas värdeförändring i procent efter tre månader är normalfördelad, $\mathrm{N}(\mu_A, \sigma_A)$ med väntevärdet $\mu_A = 2.3$ och standardavvikelsen $\sigma_A = 1.5$ för aktie A och $\mathrm{N}(\mu_B, \sigma_B)$ med väntevärdet $\mu_B = 1.7$ och standardavvikelsen $\sigma_B = 2.1$ för aktie B. Hur stor är då sannolikheten att aktie B utvecklats mer än en procentenhet bättre än aktie A efter tre månader? (d.v.s. att B:s värdeförändring i procent minus A:s är minst en procentenhet).(Vi antar att A:s värdeförändring och B:s värdeförändring är oberoende.)

SVAR:..............

Uppgift 7

X_1 och X_2 är oberoende stokastiska variabler. $X_1 \in Bin(200, p)$, $X_2 \in Bin(400, p)$. Vi har fått utfallen $x_1 = 42$, $x_2 = 76$. Bestäm Minsta-kvadrat-skattningen av p.

A: 0.194

B: 0.197

C: 0.200

D: 0.203

Uppgift 8

Låt X och Y vara två oberoende s.v. sådana att $X \in U(0, 4\theta)$ och $Y \in U(0, 2\theta)$.

$$\theta^*_{obs} = \frac{x+y}{3}$$

är en skattning av θ.
Bestäm medelfelet för denna skattning av θ om vi fått utfallen $x = 7.3$ och $y = 3.2$.

A: 0.185

B: 0.648

C: 0.805

D: 1.506

Uppgift 9

Låt datamängden 2, 4, 11, 9, 4 vara given. Varje data anses vara ett utfall av en stokastisk variabel X_i, där X_i:na antas vara oberoende och Normalfördelade med väntevärde μ och med standardavvikelse σ. Ange det ensidigt uppåt begränsade konfidensintervallet för μ som har konfidensgraden 95 %.

A: $(-\infty, 8.80)$

B: $(-\infty, 9.34)$

C: $(-\infty, 9.63)$

D: $(-\infty, 10.73)$

Uppgift 10

Inför valet publicerades det många opinionsundersökningar. Enligt en av dem hade det rödgröna blocket 49.2 % av sympatisörerna. Man angav också att felmarginalen(dvs. halva bredden av konfidensintervallet) var 2 procentenheter. Om vi antar att den stokastiska variabeln X är antalet som svarat att de sympatiserar med det rödgröna blocket, och att $X \in Bin(n, p)$, hur stort är då n ungefär? Ungefär hur många var alltså de som svarade i opinionsundersökningen? Vi antar att det är ett konfidensintervall med den approximativa konfidensgraden 95%.

A: ungefär 625

B: ungefär 1691

C: ungefär 2400

D: ungefär 4725

Uppgift 11

Anta att $X \in \text{ffg}(p)$ och låt H_0 vara att $p = 0.1$. Vi vill testa H_0 mot alternativet $H_1 : p = 0.25$ och förkastar H_0 till förmån för mothypotesen H_1 om vi får observationen $x \leq 3$. Bestäm testets styrka.

A: 0.27

B: 0.44

C: 0.58

D: 0.98

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 $\gamma^*_{obs} = -0.026$.
99%-iga konfidensintervall $I_\beta(0.99)$ och $I_\gamma(0.99)$ har beräknats, $I_\beta(0.99) = (-2.45, 9.30)$ samt $I_\gamma(0.99) = (-0.09, 0.04)$.
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 och därför är båda termerna (dvs både den linjära och den kvadratiska) i regressionsmodellen signifikanta på risknivån 1 %.

B: Båda P-värdena är större än 0.01 och därför är ingen av de två termerna (dvs varken den linjära eller den kvadratiska) i regressionsmodellen signifikanta på risknivån 1 %.

C: Båda P-värdena är större än 0.01 och därför är båda termerna (dvs både den linjära och den kvadratiska) i regressionsmodellen signifikanta på risknivån 1 %.

D: Båda P-värdena är mindre än 0.01 och därför är ingen av termerna (dvs varken den linjära eller den kvadratiska) i regressionsmodellen signifikanta på risknivån 1 %.

Del II

Uppgift 13

En optisk äppelsorteringsmaskin sorterar äpplen i två kategorier: felfria och defekta. Av inkommande äpplen betraktas 75 % som felfria och 25 % som defekta. Den betingade sannolikheten för att ett äpple sorteras som felfritt givet att det är felfritt är $1-p$ och den betingade sannolikheten för att ett äpple sorteras som defekt givet att det är defekt är $1-q$. Maskinen är justerbar och värdet på p kan ges ett godtyckligt värde mellan 0.01 och 0.1, men produkten $pq = 1.2 \cdot 10^{-3}$ är låst. Om p minskas, så ökar q och omvänt.

a) Beräkna den betingade sannolikheten för att ett äpple är felfritt givet att det sorterats som felfritt, om $p = 0.01$. (5 p)

b) Bestäm p så att sannolikheten att ett slumpvis valt äpple blir felsorterat är så liten som möjligt och ange denna sannolikhet. (5 p)

Uppgift 14

Ett elektroniskt instrument innehåller tio lika komponenter av något slag, vilkas livslängder $X_1, X_2, \ldots, X_{10}$ är oberoende och Weibullfördelade med parametrarna a och c, dvs. fördelningsfunktionen är

$$F_{X_i}(t) = \begin{cases} 1 - e^{-(t/a)^c} & \text{om} \quad t \geq 0, \\ 0 & \text{annars,} \end{cases}$$

där a och c är positiva konstanter, $i = 1, 2, \ldots, 10$. Om de tio komponenterna *seriekopplas*, visa då att hela systemets livslängd också blir Weibullfördelad. Ange fördelningens parametrar. (10 p)

Uppgift 15

En elfirma har ansvaret att åtgärda elfelen i fyra områden:A, B ,C , och D. Vi antar att μ är väntevärdet för sammanlagda antalet elfel som inträffar i alla fyra områdena under en månad. Vi antar även att elfelen överallt uppträder oberoende av varandra och att antalet elfel i vart och ett av områdena kan anses vara Poissonfördelat.
Under en slumpvis vald månad inträffar 7 elfel i område A, 2 elfel i område B, 3 elfel i område C, och 9 elfel i område D.

a) Bestäm utifrån dessa data ett konfidensintervall för μ med en konfidensgrad på approximativt 95 %. Ange vilka statistiska modellantaganden du gör samt motivera varför du får göra detta approximativa konfidensintervall. (7 p)

b) Enligt en bemanningsplan som elfirman har planerar elfirman som om det vore i genomsnitt 14 elfel i månaden. Kan man på grundval av de data vi har förkasta nollhypotesen $H_0 : \mu = 14$ på den approximativa risknivån 5% om mothypotesen är $H_1 : \mu \neq 14$? Motivera din slutsats. (3 p)

Uppgift 16

För att skatta parametern b i den likformiga fördelningen $U(0, b)$ gör man observationerna $x_1, \ldots, x_n$ som är utfall av de stokastiska variablerna X_i, $i = 1, \ldots, n$, där varje $X_i \in U(0, b)$ och X_i:na antas oberoende.

a) Härled ML-skattningen b^*_{obs} av b. (4 p)

b) Låt

$$\hat{b}_{obs} = c(n) \cdot b^*_{obs}$$

vara den korrigerade ML-skattningen av b, där $c(n)$ är en funktion av n. Bestäm $c(n)$ sådan att $\hat{b}_{obs}$ blir en väntevärdesriktig skattning av b. (6 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1912/SF1914/SF1915/SF1916
SANNOLIKHETSTEORI OCH STATISTIK,
FREDAG 21 OKTOBER 2022 KL 8.00–13.00.

Del I, Svar

1. B

2. D

3. B

4. C

5. D

6. $1 - \Phi(0.620) \approx 26.8\%$

7. A

8. D

9. C

10. C

11. C

12. B

Del I, Lösningsförslag

Uppgift 1

Täthetsfunktionen fås genom att derivera fördelningsfunktionen, $f_X(x) = F'_X(x) = \frac{1}{4}(3-x)$ för $0 \le x \le 2$ och 0 annars. Väntevärdet av X^2 ges sedan av integralen $\int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^2 x^2 \cdot \frac{1}{4}(3-x)dx = \frac{1}{4}\left[x^3 - \frac{1}{4}x^4\right]_0^2 = 1$.

Uppgift 2

Den sökta sannolikheten fås genom att integrera täthetsfunktionen över området $x > 0, 1 < y < 3$.

$$P(Y<3) = \int_0^\infty dx \int_1^3 f_{X,Y}(x,y)dy = \int_0^\infty dx \int_1^3 xe^{-xy}dy =$$
$$\int_0^\infty [-e^{-xy}]_1^3 dx = \int_0^\infty e^{-x}dx - \int_0^\infty e^{-3x}dx = 1 - \frac{1}{3} = \frac{2}{3}$$

Uppgift 3

Om Z är oberoende av X och Y så är $C(X,Z) = C(Y,Z) = 0$. Alltså är $V(2X - Y + 3Z) = 4V(X) + V(Y) + 9V(Z) - 2 \cdot 2C(X,Y)$, så

$$C(X,Y) = \frac{1}{4}(4V(X) + V(Y) + 9V(Z) - V(2X - Y + 3Z)) =$$
$$\frac{1}{4}(4D(X)^2 + D(Y)^2 + 9D(Z)^2 - D(2X - Y + 3Z)^2) = -2$$

Uppgift 4

Om Emma ska få minst två räkpiroger, men ingen köttfärspirog, kan hon antingen få två räkpiroger och en svamppirog eller få tre räkpiroger. Antal gynnsamma utfall blir därmed

$$g = \binom{5}{2}\binom{7}{1} + \binom{5}{3} = 8\binom{5}{3}$$

Antal möjliga utfall är $m = \binom{16}{3}$ så sannolikheten att minst två räkpiroger, men ingen köttfärspirog blir

$$\frac{g}{m} = \frac{8\binom{5}{3}}{\binom{16}{3}} = \frac{1}{7} \approx 0.143$$

Uppgift 5

Att en påse innehåller minst tio "röda" lökar är detsamma som att den innehåller maximalt två "blå" lökar och antalet "blå" lökar i en påse har fördelningen Bin(12, 0.15). Den sökta sannolikheten är därför $F(2; 12, 0.15) \approx 0.736$ (hämtas från tabell 6).

Uppgift 6

En linjärkombination av oberoende normalfördelade stokastiska variabler är normalfördelad. Om X_A och X_B är värdeförändringen hos aktie A respektive aktie B, så gäller att $X_B - X_A \in \mathrm{N}(\mu_B - \mu_A, \sqrt{\sigma_B^2 + \sigma_A^2})$. Därmed fås $P(X_B - X_A > 1) = 1 - P(X_B - X_A \le 1) = 1 - \Phi(\frac{1-(\mu_B-\mu_A)}{\sqrt{\sigma_B^2+\sigma_A^2}}) = 1 - \Phi(0.62) \approx 26.76\%$.

Uppgift 7

$X_1 \in Bin(200, p)$ och $X_2 \in Bin(400, p)$ och X_1 och X_2 är oberoende.
Vi vill skatta p med Minsta-kvadratmetoden och använder §9.2 i formelsamlingen.

$$Q = \sum_{alla\ x_i} (x_i - \mu_i(p))^2 = \sum_{alla\ x_i} (x_i - E(X_i))^2 = (x_1 - n_1 p)^2 + (x_2 - n_2 p)^2 = (42 - 200p)^2 + (76 - 400p)^2$$

Nu deriverar vi Q m.a.p p för att hitta det p som minimerar Q.

$$\frac{dQ}{dp} = -2(42 - 200p)200 + -2(76 - 400p)400 = 0$$

$$42 \cdot 2 - 400p + 76 \cdot 4 - 1600p = 0$$

$$\Rightarrow p^*_{obs_{MK}} = \frac{42 \cdot 2 + 76 \cdot 4}{400 + 1600} = 0.194$$

Uppgift 8

X och Y är två oberoende s.v. sådana att $X \in U(0, 4\theta)$ och $Y \in U(0, 2\theta)$.

$$\theta^*_{obs} = \frac{x + y}{3}$$

är vår skattning av θ.
Se F.S. §9.3. Medelfelet är den skattade standardavvikelsen av skattningen.

$$V[\theta^*] = V[\frac{X+Y}{3}] = \frac{V[X+Y]}{9} = \textbf{ober} = \frac{V[X] + V[Y]}{9} = \frac{\frac{(4\theta-0)^2}{12} + \frac{(2\theta-0)^2}{12}}{9} = \frac{5\theta^2}{27}$$

$$D[\theta^*] = \theta\sqrt{\frac{5}{27}}$$

Medelfelet för denna skattning är då

$$D^*_{obs}[\theta^*] = \theta^*_{obs}\sqrt{\frac{5}{27}} = \frac{x+y}{3}\sqrt{\frac{5}{27}} = \frac{7.3 + 3.2}{3}\sqrt{\frac{5}{27}} = 1.506$$

om vi fått utfallen $x = 7.3$ och $y = 3.2$.

Uppgift 9

Ett tvåsidigt konfidensintervall för μ då ett stickprov betraktas som observationer på $N(\mu.\sigma)$ och σ är okänd tas fram m.h.a §12.2 och §11.1d i formelsamlingen.
Då blir det tvåsidiga 95%-iga konfidensintervallet $I_\mu = \bar{x} \pm t_{0.025}(n-1)\frac{s}{\sqrt{n}}$.
Då blir det ensidigt uppåt begränsade konfidensintervallet $I_\mu = (-\infty, \bar{x} + t_{0.05}(n-1)\frac{s}{\sqrt{n}}\ )$
Vi har datamängden 2, 4, 11, 9, 4.

Då blir

$$n = 5, \bar{x} = 6, s = \sqrt{\frac{1}{n-1}\sum_{alla\ x_i}(x_i - \bar{x})^2} = \sqrt{11.6}$$

Då $t_{0.05}(4) = 2.13$ blir den övre gränsen för μ

$$6 + \frac{\sqrt{11.6}}{\sqrt{5}}2.13 = 9.63$$

och då fås konfidensintervallet $I_\mu = (-\infty,\ 9.63\ )$

Uppgift 10

Låt X vara antalet som svarade att de sympatiserade med det rödgröna blocket. $X \in Bin(n,p)$.
Vi får då det tvåsidiga konfidensintervallet till

$$I_p = p^*_{obs} \pm \sqrt{\frac{p^*_{obs}\cdot(1-p^*_{obs})}{n}} \cdot \lambda_{\frac{\alpha}{2}}$$

och felmarginalen är då lika med

$$\sqrt{\frac{p^*_{obs}\cdot(1-p^*_{obs})}{n}} \cdot \lambda_{\frac{\alpha}{2}} = 0.02$$

Detta ger att

$$n = (\frac{\lambda_{\frac{\alpha}{2}}}{0.02})^2 \cdot p^*_{obs}\cdot(1-p^*_{obs}) = (\frac{1.96}{0.02})^2 \cdot 0.492 \cdot (1-0.492) \approx 2400$$

Uppgift 11

Styrkan hos testet är $P(\text{förkasta}H_0)$ om H_1 är sann.
Dvs. vi ska ta fram $P(X \le 3)$ om $X \in ffg(0.25)$
$P(X \le 3) = P(X=1) + P(X=2) + P(X=3) =$ [se F.S § 3 : ffg-fördeln] $=$
$= 0.25 + 0.75 \cdot 0.25 + 0.75^2 \cdot 0.25 = \frac{37}{64} = 0.58.$

Uppgift 12

$\beta = 0 \in I_\beta(0.99)$, så vi kan inte förkasta att $\beta = 0$ på 1%-nivån. Dämed antar vi att den linjära termen inte är signifikant på 1%-nivån. När vi inte kan förkasta H_0:$\beta = 0$ så vet vi också att P-värdet> 0.01

$\gamma = 0 \in I_\beta(0.99)$, så vi kan inte heller förkasta att $\gamma = 0$ på 1%-nivån. Dämed antar vi att den kvadratiska termen inte är signifikant på 1%-nivån. När vi inte kan förkasta H_0:$\gamma = 0$ så vet vi också även här att P-värdet> 0.01

Alltså är alternativ B det rätta alternativet.

Del II, Lösningsförslag

Uppgift 13

Låt A vara händelsen att ett äpple är felfritt och B vara händelsen att ett äpple sorteras som felfritt. Det är givet att $P(B|A) = 1 - p$ och att $P(B^*|A^*) = 1 - \frac{1.2 \cdot 10^{-3}}{p}$.

a) Om $p = 0.01$ så blir $q = 0.12$. Använd nu Bayes sats:
$P(A|B) = P(B|A)\frac{P(A)}{P(B)} = P(B|A)\frac{P(A)}{P(B|A)P(A)+P(B|A^*)P(A^*)} \approx 0.961$.

b) Sannolikheten att ett äpple blir felsorterat är $P(B \cap A^* \cup B^* \cap A) = P(A^*)P(B|A^*) + P(A)P(B^*|A) = 0.25\frac{1.2 \cdot 10^{-3}}{p} + 0.75p$. Minimum inträffar då $0.25\frac{1.2 \cdot 10^{-3}}{p} = 0.75p$ (Sätt t.ex. derivatan till 0), vilket ger $p = 0.02$ och felsannolikheten blir 0.03.

Uppgift 14

Lösningen: $Y = min(X_1, ..., X_{10})$ blir hela systemets livslängd. Därmed blir fördelningsfunktionen för Y: $F_Y(x) = 1 - (1 - F_{X_1}(x)) * ...(1 - F_{X_{10}}(x)) = ... = 1 - e^{-10(x/a)^c}$, vilket också är en Weibullfördelning med parametrarna $a_Y = a/(10^{1/c})$ och $c_Y = c$.

Uppgift 15

a) Vi antar att antalet elfel som inträffar i område A,B,C respektive D under månaden är X_A, X_B, X_C
respektive X_D, och att $X_A \in Po(\mu_A)$,$X_B \in Po(\mu_B)$,$X_C \in Po(\mu_C)$ och $X_D \in Po(\mu_D)$. Dessutom antar vi att X_A, X_B, X_C och X_D är oberoende.
Då gäller att totala antalet elfel som vi kallar $Y = X_A + X_B + X_C + X_D \in Po(\mu_A + \mu_B + \mu_C + \mu_D) = Po(\mu)$. Villkoret för att få göra ett konfidensintervall för μ är att Poissonfördelningen kan approximeras till en Normalfördelning.Villkoret för detta är att $\mu \geq 15$ enligt §6 i Formelsamlingen. Vi har inte μ.
Men eftersom $E(Y) = \mu$ så skattar vi μ med $\mu^*_{obs} = y = x_A + x_B + x_C + x_D = 3+2+7+9 = 21 > 15$, vilket gör att vilkoret är uppfyllt. Sedan bildar vi konfidensintervallet enligt §12.3 i formelsamlingen:

$I_\mu = \mu^*_{obs} \pm D^*_{obs} \cdot \lambda_{\alpha/2}$

Eftersom $E(Y) = \mu$ så skattar vi μ med $\mu^*_{obs} = y$.
Eftersom $D(Y) = \sqrt{\mu}$ så skattar vi $D(Y)$ med $D^*_{obs} = \sqrt{\mu^*_{obs}} = \sqrt{y}$.

$$\underline{I_\mu = \left(y \pm \lambda_{\alpha/2}\sqrt{y}\right) = \left(21 \pm 1.96\sqrt{21}\right) = (12.02, 29.98)}$$

b) Eftersom konfidensintervallet i a-uppgiften innehåller 14, så kan vi inte förkasta $H_0 : \mu = 14$ på den approximativa risknivån 5%.

Uppgift 16

a) Eftersom X_i:na antas vara oberoende, blir likelihood-funktionen

$$L(b) = L(x_1, x_2, \ldots, x_n; b) = \frac{1}{b^n} \prod_{i=1}^{n} \mathbf{1}_{(0,b)}(x_i),$$

där

$$\mathbf{1}_{(0,b)}(x_i) = \begin{cases} 1 & \text{för} \quad 0 < x_i \leq b, \\ 0 & \text{annars.} \end{cases}$$

Likelihood-funktionen $L(b)$ maximeras i det minsta värdet på b som samtidigt är sådant att $\mathbf{1}_{(0,b)}(x_i) \neq 0$ för alla x_i, $i = 1, 2, \ldots, n$. Således, ML-skattningen $b^*_{obs} = b^*_{ML,obs} = \max_i \{x_i\}$ av b.

b) ML-skattningen av b är alltså: $b^* = \max_i \{X_i\}$. Vi studerar den korrigerade skattningen $\hat{b} = c(n) \cdot b^*$. Och $\hat{b}_{obs}$ blir väntevärdesriktig skatting av b om $E(\hat{b}) = b$. Vi har:

$$E(\hat{b}) = c(n)E(b^*).$$

Vi studerar $E(b^*)$. Först, härleder vi fördelningsfunktionen för b^*.

$$F_{b^*}(y) = P(b^* \leq y) = P\Big(\max_i \{X_i\} \leq y\Big) = P\Big(\cap_{i=1}^{n} \{X_i \leq y\}\Big) = |\text{oberoende}| =$$

$$\prod_{i=1}^{n} P(X_i \leq y) = |X_i \in U(0,b)| = \prod_{i=1}^{n} \left(\frac{y}{b}\right) = \frac{y^n}{b^n}.$$

Om vi deriverar fördelningsfunktionen $F_{b^*}(y)$, har vi följande täthetsfunktion

$$f_{b^*}(y) = F'_{b^*}(y) = \begin{cases} \frac{ny^{n-1}}{b^n} & \text{för} \quad 0 < y \leq b, \\ 0 & \text{annars.} \end{cases}$$

Och således har vi för $E(b^*)$:

$$E(b^*) = \int_0^b y f_{b^*}(y) dy = \int_0^b y \frac{ny^{n-1}}{b^n} dy = \frac{n}{b^n} \int_0^b y^n dy = \frac{nb^{n+1}}{(n+1)b^n} = \frac{n}{n+1} b.$$

Vilket innebär att $c(n)$ måste vara lika med $\frac{n+1}{n}$, för att $\hat{b}_{obs}$ ska bli en väntevärdesriktig skattning av b. Alltså fås att $c(n) = \frac{n+1}{n}$.