Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1910 TILLÄMPAD STATISTIK,
TISDAG 8 JANUARI 2019 KL 8.00–13.00.

Examinator: Björn-Olof Skytt, 08 – 790 86 49.

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. 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 $P(A \cap B^*) = 0.2$, $P(A^* \cap B) = 0.4$ och $P(A \cup B) = 0.8$. Bestäm $P(A \mid B)$.

A: 0.25

B: 0.33

C: 0.50

D: 0.67

Uppgift 2

En stokastisk variabel X har fördelningsfunktionen

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

Bestäm $E\left(e^{-X}\right)$.

A: 0.432

B: 0.500

C: 0.568

D: 0.787

Uppgift 3

På julbordet ligger tre skivor kallrökt lax, fyra skivor gravad lax och fem skivor varmrökt lax. Lille Nisse tar två skivor helt på måfå. Vad är sannolikheten att Lille Nisse får två skivor gravad lax?

Uppgift 4

Låt X och Y vara två oberoende stokastiska variabler sådana att $X \in \mathrm{Po}(3)$ och $Y \in \mathrm{Po}(4)$. Beräkna $P(X + Y = 2)$.

Uppgift 5

Låt X och Y vara två oberoende stokastiska variabler där $X \in N(2, 3)$ och $Y \in N(4, 2)$. Låt $Z = 2Y - X$. Beräkna $P(Z > 4)$.

A: 0.345

B: 0.468

C: 0.532

D: 0.655

Uppgift 6

Låt $X \in \text{Exp}(5)$, d v s intensiteten är lika med fem. Bestäm $E(X^2)$

Uppgift 7

Låt X_1 och X_2 vara två oberoende likafördelade stokastiska variabler sådana att $X_i \in \text{Exp}(\lambda)$, d v s intensiteten är lika med λ. Beräkna maximum-likelihood-skattningen av λ då $x_1 = 4$ och $x_2 = 6$.

A: 0.082

B: 0.205

C: 0.200

D: 0.503

Uppgift 8

Antag att $X_1, \ldots, X_n$ utgör ett stickprov på $N(\mu, \sigma)$, där σ är känd. Tyko önskar testa nollhypotesen $H_0 : \mu = 2$ mot $H_1 : \mu < 2$ med hjälp av ett lämpligt konfidensintervall för μ. Vilket av nedanstående konfidensintervall för μ skall väljas för att testets signifikansnivå skall bli α?

A: $I_\mu = \left(-\infty,\ \overline{x} + \frac{\sigma}{\sqrt{n}} \cdot \lambda_\alpha\right)$

B: $I_\mu = \left(-\infty,\ \overline{x} + \frac{\sigma}{\sqrt{n}} \cdot \lambda_{\alpha/2}\right)$

C: $I_\mu = \left(\overline{x} + \frac{\sigma}{\sqrt{n}} \cdot \lambda_\alpha, \infty\right)$

D: $I_\mu = \left(\overline{x} + \frac{\sigma}{\sqrt{n}} \cdot \lambda_{\alpha/2}, \infty\right)$

Uppgift 9

Antag att $X_1, \ldots, X_n$ utgör ett stickprov på $N(\mu, \sigma)$. Från tjugo observationer erhölls följande värden $\overline{x} = 0.46$, samt $s = 0.43$. Ange nedre gränsen för det tvåsidiga konfidensintervallet för σ med konfidensgrad 99%.

A: 0.158

B: 0.302

C: 0.316

D: 0.000

Uppgift 10

Två stickprov från två populationer. Varje stickprov uppfattas som observationer på $N(\mu_i, \sigma_i)$, där vi antar att $\sigma_1 = \sigma_2$. Från de två stickproven beräknades följande sammanfattande mått:
från stickprov 1

$$n_1 = 4 \quad \overline{x}_1 = 1007.25 \quad s_1 = 143.66$$

från stickprov 2

$$n_2 = 4 \quad \overline{x}_2 = 817.75 \quad s_2 = 73.627$$

Beräkna den undre gränsen i ett 95%-igt tvåsidigt konfidensintervall för $\mu_1 - \mu_2$.

A: 13.14

B: -18450

C: 44.53

D: -8.25

Uppgift 11

Två undersökningar gjordes på två grupper av patienter som hade blivit vaccinerade respektive inte hade blivit vaccinerade. Femhundra vaccinerade undersöktes där fyrtionio hade blivit vinterkräksjuka. I den icke vaccinerade gruppen undersöktes sexhundra där femtioåtta hade fått vinterkräksjuka. Låt p_1 stå för andelen patienter som blivit vinterkräksjuka trots att de har vaccinerats och låt p_2 stå för andelen patienter som blivit vinterkräksjuka utan att ha vaccinerats. Vi är intresserade av parametern $p_1 - p_2$. Bestäm medelfelet för skattningen av $p_1 - p_2$.

A: 0.0180

B: 0.419

C: 0.297

D: 0.441

Uppgift 12

En forskare har gjort tio försök som anses vara oberoende av varandra där sannolikheten för lyckat försök är p. Låt X stå för antalet lyckade försök. Forskaren önskar pröva $H_0 : p = 1/2$ mot $H_1 : p > 1/2$. Resultatet av de tio försöken var åtta lyckade försök. Beräkna p-värdet!

A: 0.0107

B: 0.0440

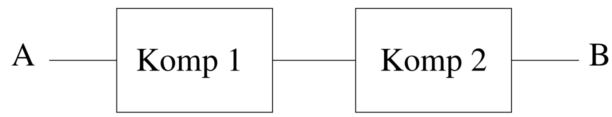
C: 0.100

D: 0.0547

Del II

Uppgift 13

a) I ett system är två komponenter kopplade enligt figuren. Systemet fungerar om både komponent 1 och 2 fungerar.



Antag att livslängderna T_1 och T_2 för komponent 1, respektive 2 är obeoroende stokastiska variabler med fördelningsfunktioner $F_1(x) = 1 - e^{-x/5}$ för $x \geq 0$, respektive $F_2(x) = 1 - e^{-x/8}$ för $x \geq 0$. Beräkna systemets förväntade livslängd. (4 p)

b) Man vill få reda på andelen p av personer i en stor population med en viss egenskap som ger svaret 'Ja' på en känslig fråga. (Ett par exempel: ”Har du under det senaste året använt narkotika?”eller ”Har du någon gång snattat?”).
För att få ökad personlig sekretess (och ett mer korrekt undersökningsresultat) lät man de tillfrågade först dra ett kort. Med sannolikheten 2/3 drar de ett kort av typ I som säger de skall svara ärligt Ja/Nej på den känsliga frågan och med sannolikheten 1/3 drar de ett kort av typ II som säger att de skall svara ärligt Ja/Nej på en irrelevant fråga, t ex “Är den sista siffran i ditt personnummer ett jämnt tal?”. Vilken typ av kort de får vet endast de tillfrågade själva. (De tillfrågade visar alltså inte kortet för någon annan och får själva dra ett kort på måfå).
Antag att man genomfört en undersökning enligt ovanstående princip och att man fick 40% Ja-svar. Vad är $p =$, dvs andelen individer som t ex under det senaste året använt narkotika? Du får anta att den irrelevanta frågan var “Är den sista siffran i ditt personnummer ett jämnt tal?” och att sannolikheten för ett jämnt tal som sista siffra i ett personnummer är lika stor som sannolikheten för ett udda tal som sista siffra. (6 p)

Ledning: Du får utgå från att totala sannolikheten för svaret Ja är/skattas som 0.4.

Uppgift 14

En stressad klassförälder hade lovat att se till att det fanns lussebullar till barnen i första klass som skulle gå luciatåg på luciadagen. De skulle gå ett luciatåg på morgonen och ett på eftermiddagen och fika efter båda tågen. Antalet barn i första klass på skolan är 55. På morgonen äter en elev i årskurs 1 ingen lussebulle med sannolikhet 0.1, en lussebulle med sannolikhet 0.7 och två lussebullar med sannolikhet 0.2. Fördelningen över antalet ätna lussebullar är densamma på eftermiddagen, dock finns det ett beroende mellan antal ätna lussebullar på morgonen och eftermiddagen som resulterar i en negativ korrelationskoefficient på $\rho = -7/29$. Det finns inget beroende mellan hur många bullar olika elever äter.
Beräkna approximativt antalet lussebullar som föräldern borde ha bakat för att alla barn med sannolikhet 95% skulle få så många bullar som de ville. Alla gjorda approximationer skall motiveras.

(10 p)

Uppgift 15

Man misstänkte att ett roulettebord på ett kasino var manipulerat och genomförde ett test med 8000 försök. Om rouletten är korrekt skall röd, svart och grön (nollan) komma upp i proportionerna 18:18:1. Testresultatet gav röd: 3751, svart: 4018, grön: 231. Avgör med felrisken 1% om rouletten är korrekt. Det måste klart framgå av svaret vad slutsatsen är. (10 p)

Uppgift 16

a) Täthetsfunktionen för $\chi^2(1)$-fördelningen ges av

$$f(t) = \frac{1}{\sqrt{2}\sqrt{\pi}} \cdot \frac{1}{\sqrt{t}} e^{-t/2} \qquad t \geq 0.$$

Visa att om $X \in N(0,1)$ så gäller att $Y = X^2 \in \chi^2(1)$. (3 p)

b) För att bestämma arean θ av en kvadrat gör man observationer $x_1, \ldots, x_n$ av stokastiska variabler X_i $i = 1, \ldots, n$, där $X_i \in N(\sqrt{\theta}, \sigma)$, σ är känd och X_i:na antas oberoende. MK-skattningen av θ baserat på observationerna $x_1, \ldots, x_n$ ges av

$$\theta^*_{obs} = \bar{x}^2 = \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^2.$$

Visa att om man tar fram ett dubbelsidigt konfidensintervall med konfidensgrad $1 - \alpha$ för $\sqrt{\theta}$ utgående från fördelningen för

$$\left(\frac{\bar{X} - \sqrt{\theta}}{\sigma/\sqrt{n}} \right)^2$$

så kommer man att få fram samma intervall som när man utgår från fördelningen för

$$\frac{\bar{X} - \sqrt{\theta}}{\sigma/\sqrt{n}}.$$

För att full poäng skall utdelas måste lösningen vara väl motiverad, speciellt måste sambandet mellan de två fördelningarnas kvantiler klart framgå. (7 p)

Ledning: Precis som vid χ^2-test måste man titta på variabeln och tänka efter vad som kan anses som bekymmersamma utfall.

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1910 TILLÄMPAD STATISTIK, TISDAG 8 JANUARI 2019 KL 8.00–13.00.

Del I

Uppgift 1

Vi ska beräkna $P(A \mid B) = P(A \cap B)/P(B)$.
Med hjälp av ett Venn-diagram är det lätt att se att

$$A \cup B = (A \cap B^*) \cup (A \cap B) \cup (A^* \cap B),$$

samt att de tre snitten är disjunkta. Enligt Kolmogorovs tredje axiom har vi därför

$$P(A \cup B) = P(A \cap B^*) + P(A \cap B) + P(A^* \cap B)$$

eller

$$P(A \cap B) = P(A \cup B) - P(A \cap B^*) - P(A^* \cap B) = 0.8 - 0.2 - 0.4 = 0.2$$

Återigen med samma metod ser vi att

$$B = (A \cap B) \cup (A^* \cap B),$$

och de är disjunkta. Så

$$P(B) = P(A \cap B) + P(A^* \cap B) = 0.2 + 0.4 = 0.6$$

Därmed blir

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.6} = \frac{1}{3} = 0.333$$

Uppgift 2

$$E\left(e^{-X}\right) = \int_0^2 e^{-x} \cdot \frac{1}{2}\, dx = \left[-\frac{1}{2}e^{-x}\right]_{x=0}^{x=2} = \frac{1}{2} - \frac{1}{2}e^{-2} = 0.432$$

Uppgift 3

$$P(\text{två gravade laxar}) = \frac{\binom{3}{0}\binom{4}{2}\binom{5}{0}}{\binom{12}{2}} = \frac{\binom{4}{2}}{\binom{12}{2}} = \frac{4 \cdot 3}{12 \cdot 11} = \frac{1}{11} = 0.0909$$

Uppgift 4

Då $X \in \mathrm{Po}(3)$ och oberoende av $Y \in \mathrm{Po}(4)$, uttalar additionsegenskapen att $X+Y \in \mathrm{Po}(7)$. Därmed blir

$$P(X+Y=2) = \frac{7^2}{2!}e^{-7} = 0.0223$$

Uppgift 5

Om vi använder sats 6.5 i Blom *et al* har vi att $2Y - X \in N(6,5)$.

$$\begin{aligned} P(Z>4) &= 1 - P(Z \le 4) \\ &= 1 - P\left(\frac{Z-6}{5} \le \frac{4-6}{5}\right) \\ &= 1 - \Phi(-0.40) \\ &= \Phi(0.40) = 0.655 \end{aligned}$$

Uppgift 6

Notationen $X \in \mathrm{Exp}(5)$ betyder att $E(X) = \frac{1}{5}$ och $\mathrm{Var}(X) = \frac{1}{5^2}$. Vi kan även skriva beräkningsformeln för variansen $\mathrm{Var}(X) = E(X^2) - (E(X))^2$ som

$$E\left(X^2\right) = \mathrm{Var}(X) + (E(X))^2 = \frac{1}{5^2} + \left(\frac{1}{5}\right)^2 = \frac{2}{25} = 0.08$$

Uppgift 7

Notationen $X_i \in \mathrm{Exp}(\lambda)$ betyder att tätheten för X_i är $f_{X_i}(x) = \lambda e^{-\lambda x_i}$. Därmed blir likelihood-funktionen

$$L(\lambda) = \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_1} = \lambda^2 e^{-\lambda(x_1+x_2)}$$

Då blir log-likelihoodfunktionen

$$\ln L(\lambda) = 2\ln\lambda - \lambda(x_1+x_2).$$

Om vi maximerar $\ln L(\lambda)$ m a p λ har vi

$$\frac{\ln L(\lambda)}{\lambda} = \frac{2}{\lambda} - (x_1+x_2) = 0 \Leftrightarrow \lambda = \frac{2}{x_1+x_2} = \frac{1}{\overline{x}}.$$

Då $\overline{x} = 5$ är ML-skattningen 1/5=0.2

Uppgift 8

Alternativ A är rätt. Se boken.

Uppgift 9

Den undre gränsen i ett konfidensintervall för σ ges av

$$k_1\, s = \sqrt{\frac{f}{\chi^2_{\alpha/2}(f)}}\, s \qquad f = n-1$$

Då $n = 20$ och $\alpha = 0.01$, blir $f = 19$ och $\chi^2_{0.005}(19) = 38.6$. Eftersom $s = 0.43$ blir den undre gränsen

$$k_1 \, s = \sqrt{\frac{f}{\chi^2_{\alpha/2}(f)}}\, s = \sqrt{\frac{19}{38.6}}\, 0.43 = 0.302$$

Uppgift 10

$$s = \sqrt{\frac{Q_1 + Q_2}{(n_1 - 1) + (n_2 - 1)}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}} = \sqrt{\frac{3 \cdot 143.66^2 + 3 \cdot 73.627^2}{3 + 3}} = 114.147$$

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2}(f) s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 1007.25 - 817.75 - t_{0.025}(n_1 + n_2 - 2) \cdot 114.147 \cdot \sqrt{\tfrac{1}{4} + \tfrac{1}{4}} = 189.5 - 2.45 \cdot 114.147 \cdot \sqrt{\tfrac{1}{4} + \tfrac{1}{4}} = -8.25$$

Uppgift 11

$$\sqrt{\frac{\frac{x_1}{n_1}\left(1 - \frac{x_1}{n_1}\right)}{n_1} + \frac{\frac{x_2}{n_2}\left(1 - \frac{x_2}{n_2}\right)}{n_2}} = \sqrt{\frac{\frac{49}{500}\left(1 - \frac{49}{500}\right)}{500} + \frac{\frac{58}{600}\left(1 - \frac{58}{600}\right)}{600}} = 0.01795352 = 0.0180$$

Uppgift 12

$$P(X \geq 8) = 1 - P(X \leq 7) = \lceil \text{Tabellvärde där } p = \frac{1}{2} \rceil = 1 - 0.9453 = 0.0547$$

Del II

Uppgift 13

a) Låt T=systemets livslängd. Då gäller att $T = \min\{T_1, T_2\}$ och

$$\begin{aligned}
F_T(t) &= P(T \leq t) = P(\min\{T_1, T_2\} \leq t) = 1 - P(\min\{T_1, T_2\} > t) \\
&= 1 - P(T_1 > t, T_2 > t) = \{\text{oberoende}\} = 1 - P(T_1 > t)P(T_2 > t) \\
&= 1 - [1 - P(T_1 \leq t)]\,[1 - P(T_2 \leq t)] \\
&= 1 - \left[1 - (1 - e^{-t/5})\right]\left[1 - (1 - e^{-t/8})\right] \\
&= 1 - e^{-t(1/5+1/8)} = 1 - e^{-t\cdot 13/40}
\end{aligned}$$

Antingen känner man redan nu igen fördelningfunktionen för en exponentialfördelning med parameter $\lambda = 13/40$ eller också deriverar man för att få

$$f_T(t) = \frac{d}{dt}F_T(t) = \frac{d}{dt}(1 - e^{-t\cdot 13/40}) = \frac{13}{40}e^{-t\cdot 13/40}.$$

Känner man nu igen täthetsfunktionen för en exponentialfördelningen med parameter $\lambda = 13/40$ får man mha formelsamlingen att systemets förväntade livslängd är $\underline{E(T) = 40/13 \approx 3.0769}$ (annars kan den beräknas genom att man löser integralen

$$E(T) = \int_{-\infty}^{\infty} t f_T(t) dt).$$

b) Inför följande händelser

$$\begin{aligned}
JA &= \text{händelsen att man svarar ja på en fråga} \\
Nonsens &= \text{händelsen att man får svara på en nonsensfråga} \\
Kanslig &= \text{händelsen att man får svara på en känslig fråga}
\end{aligned}$$

Vi vet då att

$$P(JA) = 0.4, \quad P(Nonsens) = \frac{1}{3}, \quad P(Kanslig) = \frac{2}{3}$$

samt att $P(JA|Nonsens) = 1/2$. Vidare ger lagen om total sannolikhet att

$$P(JA) = P(JA|Nonsens)P(Nonsens) + P(JA|Kanslig)P(Kanslig)$$

Eftersökt är $p = P(JA|Kanslig)$ och med insatta värden fås

$$0.4 = \frac{1}{2}\cdot\frac{1}{3} + p\cdot\frac{2}{3}$$

varför $\underline{p=0.35}$.

Uppgift 14

Låt X_j=antal lussebullar elev j äter på morgonen och Y_j=antal lussebullar elev j äter på eftermiddagen. Vi får

$$\begin{aligned}
E(X_j) &= 0\cdot 0.1 + 1\cdot 0.7 + 2\cdot 0.2 = 1.1 \\
E(X_j^2) &= 0^2\cdot 0.1 + 1^2\cdot 0.7 + 2^2\cdot 0.2 = 1.5 \\
V(X_j) &= E(X_j^2) - [E(X_j)]^2 = 1.5 - 1.1^2 = 0.29 \\
D(X_j) &= \sqrt{0.29}
\end{aligned}$$

och då X_j och Y_j är likafördelade har de samma väntevärde och varians.
Låt vidare $Z_j = X_j + Y_j$. Då gäller att $E(Z_j) = E(X_j) + E(Y_j) = 1.1 + 1.1 = \underline{2.2}$ för $j = 1, 2, \cdots, 55$.
Vi får $V(Z_j) = V(X_j + Y_j) = V(X_j) + V(Y_j) + 2C(X_j, Y_j) = 0.29 + 0.29 + 2\rho(X_j, Y_j)D(X_j)D(Y_j) = 0.29 + 0.29 + 2 \cdot (-7/29) \cdot \sqrt{0.29} \cdot \sqrt{0.29} = \underline{0.44}$

Totala antalet lussebullar som äts blir $S = Z_1 + Z_2 + \cdots + Z_{55}$ som är approximativt normalfördelad enligt Centrala gränsvärdessatsen ty $Z_1, \cdots, Z_{55}$ är oberoende och likafördelade och ganska många. S är approximativt $N(55 \cdot 2.2, \sqrt{55 \cdot 0.44})$ och för att lussebullarna skall räcka med sannolikhet 95% måste vi ha, med $x=$ antal lussebullar som bakas, att

$$P(S \leq x) = 0.95$$

eller

$$P\left(\frac{S - 121}{\sqrt{24.2}} \leq \frac{x - 121}{\sqrt{24.2}}\right) = 0.95$$

Vi ser nu att

$$\frac{x - 121}{\sqrt{24.2}} = \lambda_{0.05} = 1.6449$$

dvs man bör baka

$$x = 121 + 1.6449\sqrt{24.2} \approx \underline{130 \text{ bullar.}}$$

Här har vi avrundat uppåt till hela bullar för att vara säkra på att det skall räcka till.

Uppgift 15

Vi gör ett χ^2-test av H_0 : ”Rouletten är korrekt”. Vi har
$8000 \cdot P(\text{röd}) = 8000 \cdot 18/37 \approx 3891.89$, $8000 \cdot P(\text{svart}) \approx 3891.89$ och $8000 \cdot P(\text{grön}) = 8000/37 \approx 216.22$. Villkoret $np_i > 5$ är alltså uppfyllt med råge. Testvariabeln är alltså

$$Q = \frac{(3751 - 3891.89)^2}{3891.89} + \frac{(4018 - 3891.89)^2}{3891.89} + \frac{(231 - 216.22)^2}{216.22} \approx 10.20$$

Vi testar en given fördelning, så detta är en observation av en χ^2-fördelad variabel med 3-1=2 frihetsgrader. Eftersom $Q > \chi^2_\alpha(2) = 9.21$ förkastar vi H_0 med felrisken 1%, dvs. vi drar slutsatsen att rouletten är felaktig.

Uppgift 16

a) Vi har att fördelningsfunktionen för Y ges av

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \{\text{kontinuerlig fördelning}\} = P(-\sqrt{y} < X \leq \sqrt{y}) \\ &= P(X \leq \sqrt{y}) - P(X \leq -\sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

Här krävs uppenbarligen att $y \geq 0$.
Genom att derivara får vi täthetsfunktionen för Y som

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left[F_X(\sqrt{y}) - F_X(-\sqrt{y})\right] \\ &= f_X(\sqrt{y}) \cdot \frac{1}{2}\frac{1}{\sqrt{y}} - f_X(-\sqrt{y}) \cdot \left(-\frac{1}{2}\frac{1}{\sqrt{y}}\right) \\ &= \frac{1}{2}\frac{1}{\sqrt{y}}\varphi(\sqrt{y}) + \frac{1}{2}\frac{1}{\sqrt{y}}\varphi(-\sqrt{y}) \end{aligned}$$

för $y \geq 0$. Ur formelsamlingen hämtar man att täthetsfunktionen för en standard normalfördelning ges av

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

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

$$f_Y(y) = \frac{1}{2}\frac{1}{\sqrt{y}}\frac{1}{\sqrt{2\pi}} e^{-\sqrt{y}^2/2} + \frac{1}{2}\frac{1}{\sqrt{y}}\frac{1}{\sqrt{2\pi}} e^{-(-\sqrt{y})^2/2} = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}} e^{-y/2}$$

för $y \geq 0$, vilket överensstämmer med det givna uttrycket för täthetsfunktionen för en $\chi^2(1)$-fördelning.

b) Det gäller att

$$\frac{\bar{X} - \sqrt{\theta}}{\sigma/\sqrt{n}} \in N(0,1)$$

Det dubbelsidiga konfidensintervallet för $\sqrt{\theta}$ med konfidensgrad $1-\alpha$ man får baserat på fördelningen för ovanstående variabel är det vanliga

$$I_{\sqrt{\theta}} = \left(\bar{x} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \tag{1}$$

(för en härledning se läroboken avsnitt 12.3 a)).
Det gäller (se del a) av denna uppgift) att

$$\left(\frac{\bar{X} - \sqrt{\theta}}{\sigma/\sqrt{n}}\right)^2 \in \chi^2(1)$$

Vi får därför att

$$P\left(\left(\frac{\bar{X} - \sqrt{\theta}}{\sigma/\sqrt{n}}\right)^2 \leq \chi^2_\alpha(1)\right) = 1 - \alpha,$$

eller

$$P\left(\left|\frac{\bar{X} - \sqrt{\theta}}{\sigma/\sqrt{n}}\right| \leq \sqrt{\chi^2_\alpha(1)}\right) = 1 - \alpha.$$

Vi kan ta bort beloppstecknet och får då

$$P\left(-\sqrt{\chi^2_\alpha(1)} \leq \frac{\bar{X} - \sqrt{\theta}}{\sigma/\sqrt{n}} \leq \sqrt{\chi^2_\alpha(1)}\right) = 1 - \alpha.$$

Detta kan omformas till (se till att få $\sqrt{\theta}$ ensamt i mitten)

$$P\left(\bar{X} - \sqrt{\chi^2_\alpha(1)}\frac{\sigma}{\sqrt{n}} \leq \sqrt{\theta} \leq \bar{X} + \sqrt{\chi^2_\alpha(1)}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Ett dubbelsidigt konfidensintervall för $\sqrt{\theta}$ med konfidensgrad $1 - \alpha$ ges alltså av

$$I_{\sqrt{\theta}} = \left(\bar{x} - \sqrt{\chi^2_\alpha(1)}\frac{\sigma}{\sqrt{n}}, \bar{x} + \sqrt{\chi^2_\alpha(1)}\frac{\sigma}{\sqrt{n}}\right)$$

Genom att jämföra med intervallet i (1) ser vi att $\sqrt{\chi^2_\alpha(1)} = \lambda_{\alpha/2}$, vilket man också inser eftersom det, med beteckningar från deluppgift a), gäller att

$$P(Y \leq \chi^2_\alpha(1)) = 1 - \alpha,$$

vilket är detsamma som

$$P(|X| \leq \sqrt{\chi^2_\alpha(1)}) = 1 - \alpha$$

eller

$$P(-\sqrt{\chi^2_\alpha(1)} \leq X \leq \sqrt{\chi^2_\alpha(1)}) = 1 - \alpha.$$

Då $X \in N(0,1)$ gäller också att

$$P(-\lambda_{\alpha/2} \leq X \leq \lambda_{\alpha/2}) = 1 - \alpha$$

och vi måste ha att $\sqrt{\chi^2_\alpha(1)} = \lambda_{\alpha/2}$.