KTH Matematik

Avd. Matematisk statistik

TENTAMEN I SF1917/SF1919 SANNOLIKHETSTEORI OCH STATISTIK,
TISDAG 10 JANUARI 2023 KL 8.00–13.00.

Examinator: Mykola Shykula, 08-790 6644.

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 ordinarietentan i januari 2023 och vid omtentamen i april 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 datorlaborationen får 3 bonuspoäng på del II på ordinarie tentamenstillfället och vid omtentamen i april 2023.

Tentamen kommer att vara rättad inom tre arbetsveckor(15 arbetsdagar) från skrivningstillfället och kommer att finnas tillgänglig på studentexpeditionen minst sju veckor efter skrivningstillfället.

Del I

Uppgift 1

Antag att $A \cup B = \Omega$ samt $A \cap B = \emptyset$ medan $\mathrm{P}(A \cap C) = 0.2$, $\mathrm{P}(B \cap C^{\star}) = 0.4$ och $\mathrm{P}(A) = 0.3$. Beräkna $\mathrm{P}(C)$.

A: 0.6

B: 0.8

C: 0.2

D: 0.5

Uppgift 2

Den stokastiska variabeln Y har fördelningsfunktionen

$$F_Y(y) = \begin{cases} \frac{14}{5} y^{2/5}\left(\frac{1}{2} - \frac{y}{7}\right) & \text{om} \quad 0 < y \leq 1, \\ 0 & \text{om} \quad y \leq 0, \\ 1 & \text{om} \quad y > 1. \end{cases}$$

Bestäm väntevärdet $E(Y)$.

A: 2/5

B: 1/5

C: 1/6

D: 1/7

Uppgift 3

Den tvådimensionella stokastiska variabeln (X, Y) har sannolikhetsfunktionen:

$p_{X,Y}(x,y)$	$y = 0$	$y = 1$	$y = 2$
$x = 0$	0.1	0.2	0.3
$x = 1$	0.3	0.1	0

Bestäm korrelationskoefficienten $\rho(X, Y)$.

A: -0.64

B: -0.32

C: -0.01

D: 0.32

Uppgift 4

Antag att en påse lördagsgodis innehåller sju råttor, fyra geléhallon och fem centerpraliner. Om vi plockar ut 4 godisbitar på måfå, vad är sannolikheten att vi får två sega råttor, ett geléhallon och en centerpralin?

A: 0.179

B: 0.091

C: 0.019

D: 0.231

Uppgift 5

Antag att X och Y är oberoende stokastiska variabler med fördelning $\mathrm{Bin}(4, 0.3)$ respektive $\mathrm{Bin}(5, 0.3)$. Beräkna sannolikheten $\mathrm{P}(X + Y \geq 2)$.

A: 0.804

B: 0.896

C: 0.777

D: 0.537

Uppgift 6

När man analyserar löner (och inkomster i allmänhet) är en rimlig modell att dessa är lognormalfördelade, dvs. att deras logaritm är normalfördelat. I Sverige 2021 ger detta att om månadslönen ges av X har $\ln(X)$ fördelningen $\mathrm{N}(10.41, 0.3650)$. Bestäm, enligt denna modell, den minsta månadslönen bland de 1% bäst betalda arbetstagarna.

A: 102531 kr

B: 77583 kr

C: 84980 kr

D: 47810 kr

Uppgift 7

Antag att X och Y är oberoende och fördelade enligt $\mathrm{N}(\theta, 2)$ respektive $\mathrm{N}(2\theta, 1)$. Vi har då två punktskattingar

$$\theta^\star = \frac{X + 8Y}{17} \qquad \text{och} \qquad \widehat{\theta} = \frac{2X + Y}{4}$$

Vilket av följande påståenden stämmer?

A: $\theta^\star$ är väntevärdesriktig, men $\widehat{\theta}$ är inte väntevärdesriktig.

B: $\theta^\star$ är inte väntevärdesriktig, men $\widehat{\theta}$ är väntevärdesriktig.

C: $\theta^\star$ och $\widehat{\theta}$ är väntevärdesriktiga; $\theta^\star$ är effektivare än $\widehat{\theta}$.

D: $\theta^\star$ och $\widehat{\theta}$ är väntevärdesriktiga; $\widehat{\theta}$ är effektivare än $\theta^\star$.

Uppgift 8

X_1, X_2, X_3 och X_4 är oberoende stokastiska variabler. $X_1, X_2, X_3, X_4 \in \text{Exp}(\lambda)$, dvs $E(X_i) = 1/\lambda$. Vi har fått utfallen $x_1 = 7.3$, $x_2 = 5.7$, $x_3 = 11.2$ och $x_4 = 3.8$. Bestäm MK-skattningen av λ.

A: 1/28

B: 1/4

C: 7

D: 1/7

Uppgift 9

Antalet restaurangbesökare hos en viss restaurang räknades under en vecka:

Dag	M	Ti	O	To	F
Besökare	53	32	36	36	38

Antag att antalet besökare per dag är poissonfördelat med parameter θ och att varje dag utgör en oberoende observation av denna stokastiska variabel. Bestäm den övre gränsen för ett tvåsidigt 95% approximativt konfidensintervall för θ.

A: 45.5

B: 44.5

C: 51.2

D: 39.0

Uppgift 10

Låt X vara fördelad enligt $\text{Exp}(\theta)$ och $H_0 : \theta = 1$ samt $H_1 : \theta = 5$. Vi har ett test som förkastar nollhypotesen om $X < 0.05$. Bestäm testets styrka.

A: 0.779

B: 0.221

C: 0.951

D: 0.049

Uppgift 11

Antag att två oberoende stickprov, $x_1, \ldots, x_{n_1}$ och $y_1, \ldots, y_{n_2}$ är givna och att man tänker använda följande formel för att beräkna ett konfidensintervall för skillnaden mellan väntevärdena:

$$I_{\mu_2-\mu_1} = \left(\bar{y} - \bar{x} \pm t_{\alpha/2}(n_1 + n_2 - 2)s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

där

$$s = \sqrt{\frac{\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{i=1}^{n_2}(y_i - \bar{y})^2}{n_1 + n_2 - 2}}$$

Antag att stickproven kommer från X_i, $i = 1, \ldots, n_1$, respektive Y_i, $i = 1, \ldots, n_2$. Vad skall gälla för dessa variabler för att konfidensgraden för det givna intervallet skall vara $1 - \alpha$?

A: $X_i \in \mathrm{N}(\mu_1, \sigma_1)$ och $Y_i \in \mathrm{N}(\mu_2, \sigma_2)$, där σ_1 och σ_2 är kända.

B: $X_i \in \mathrm{N}(\mu_1, \sigma_1)$ och $Y_i \in \mathrm{N}(\mu_2, \sigma_2)$, där $\sigma_1 = \sigma_2 = \sigma$ och σ är okänd.

C: $X_i \in \mathrm{N}(\mu_1, \sigma_1)$ och $Y_i \in \mathrm{N}(\mu_2, \sigma_2)$, där σ_1 och σ_2 båda är okända och olika.

D: X_i och Y_i kan ha godtycklig fördelning så länge $n_1 \leq 20$ och $n_2 \leq 20$.

Uppgift 12

Följande atmosfäriska koncentrationer av koldioxid uppmättes vid Mauna Loa-observatoriet i Hawaii, USA:

År	2017	2018	2019	2020	2021	2022
Koncentration (ppm)	406.0	407.8	410.7	413.2	415.1	417.8

Om vi låter x-axeln representera antalet år sedan 2000 och y-axeln representera koncentration kan datan sammanfattas enligt

$$\sum_{i=1}^{n} x_i = 117, \quad \sum_{i=1}^{n} x_i^2 = 2299, \quad \sum_{i=1}^{n} y_i = 2470.6, \quad \sum_{i=1}^{n} y_i^2 = 1017410.42$$

samt

$$\sum_{i=1}^{n} x_i y_i = 48218.4.$$

Ta fram en prognos för koncentrationen år 2030 med hjälp av en enkel linjär regressionsmodell.

A: 435.2 ppm

B: 436.8 ppm

C: 434.8 ppm

D: 436.5 ppm

Del II

Uppgift 13

Sannolikheten är 0.9984 att träffa minst en gång när man skjuter fyra gånger. Antag att varje skott är oberoende av de andra. Beräkna sannolikheten att man träffar vid ett enskilt skott. (10 p)

Uppgift 14

För vissa tillämpningar (såsom komplicerade aposteriorifördelningar inom bayesiansk inferens) erhålls täthetsfunktioner vars väntevärden kan vara svåra att beräkna för hand då integralerna inte har någon analytisk lösning. Ett alternativ är då att simulera oberoende utfall $X_1, X_2, \ldots, X_n$ från fördelningen och bilda det aritmetiska medelvärdet

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Enligt stora talens lag så approximerar denna det sökta väntevärdet μ när n går mot oändligheten, dvs. att för alla $\epsilon > 0$ har vi

$$\lim_{n\to\infty} \mathrm{P}\left(|\overline{X} - \mu| < \epsilon\right) = 1.$$

Anta att vi börjar med att simulera 1000 utfall och får

$$\overline{x} = 10.144 \quad \text{och} \quad s^2 = 103.067.$$

Hur många utfall måste vi simulera för att det aritmetiska medelvärdet $\overline{X}$ ska ligga högst 0.1 från väntevärdet med sannolikhet 0.95? Motivera eventuella antaganden och approximationer utförligt. (10 p)

Uppgift 15

Antag att vi har en stokastisk variabel X med täthetsfunktion

$$f(x) = \begin{cases} \alpha x^{\alpha-1}, & 0 < x < 1, \\ 0, & \text{annars,} \end{cases}$$

för något $\alpha > 0$.

(a) Visa att $-\ln(X)$ är exponentialfördelad och bestäm dess parameter. (4 p)

(b) Beräkna ML-skattningen för α givet ett stickprov $X_1, X_2, \ldots, X_n$ från denna fördelning. (6 p)

Uppgift 16

För en viss mätapparatur misstänker man att mätfelen inte enbart är slumpmässiga utan även systematiska. För att undersöka detta mäter man 9 gånger en storhet vars värde man känner exakt. Låt $x_1, x_2, \ldots, x_9$ beteckna mätfelen vid dessa 9 mätningar. Erfarenhetsmässigt kan man anta att $x_1, x_2, \ldots, x_9$ är observationer från en normalfördelning $N(\Delta, \sigma)$, $\sigma = 0.4$. Nu misstänker man att $\Delta > 0$ och inte $\Delta = 0$. Därför skall man testa $H_0 : \Delta = 0$ mot $H_1 : \Delta > 0$ och väljer mellan två test. Det ena testet är det vanliga baserat på medelvärdet $\bar{x}$, och beslutsregeln blir: förkasta H_0 om $\bar{x} \geq 0.179$. Det andra testet är baserat på d, där d är antalet positiva x_i-värden, och beslutsregeln blir: förkasta H_0 om $d \geq 7$. Signifikansnivån är approximativt 0.09 i båda.

(a) Beräkna styrkan i $\Delta = 0.34$ för de båda testen. (8 p)

(b) Vilket av de två testen bör man välja? Motivera svaret. (2 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1917/SF1919 SANNOLIKHETSTEORI OCH STATISTIK, TISDAG 10 JANUARI 2023 KL 8.00–13.00.

Del I, Svar

1. D

2. C

3. A

4. D

5. A

6. B

7. C

8. D

9. B

10. B

11. B

12. B

Del I, Lösningsförslag:

Uppgift 1

Eftersom $A \cup B = \Omega$ och $A \cap B = \emptyset$ har vi

$$\mathrm{P}(C) = \mathrm{P}(A \cap C) + \mathrm{P}(B \cap C)$$

samt

$$\mathrm{P}(B) = \mathrm{P}(B \cap C) + \mathrm{P}(B \cap C^{\star}),$$

vilket ger

$$\mathrm{P}(B \cap C) = \mathrm{P}(B) - \mathrm{P}(B \cap C^{\star}).$$

Komplementsatsen ger i sin tur $\mathrm{P}(B) = 1 - \mathrm{P}(B^\star) = 1 - \mathrm{P}(A)$, så om vi sätter ihop allting får vi

$$\begin{aligned}\mathrm{P}(C) &= \mathrm{P}(A \cap C) + \mathrm{P}(B) - \mathrm{P}(B \cap C^\star) \\ &= \mathrm{P}(A \cap C) + 1 - \mathrm{P}(A) - \mathrm{P}(B \cap C^\star) \\ &= 0.2 + 1 - 0.3 - 0.4 = 0.5.\end{aligned}$$

Svaret är alltså D.

Uppgift 2

Täthetsfunktionen för Y är:

$$f_Y(y) = F_Y'(y) = \begin{cases} \frac{14}{25}\left(y^{-3/5} - y^{2/5}\right) & \text{om} \quad 0 < y \leq 1, \\ 0 & \text{om} \quad y \leq 0 \text{ eller } y > 1. \end{cases}$$

Vi har:

$$E(Y) = \frac{14}{25}\int_0^1 y\left(y^{-3/5} - y^{2/5}\right)dy = \frac{14}{25}\int_0^1 \left(y^{2/5} - y^{7/5}\right)dy = \frac{14}{25}\left(\frac{y^{7/5}}{7/5} - \frac{y^{12/5}}{12/5}\right)\Bigg|_0^1 =$$

$$= \frac{14}{5}\left(\frac{1}{7} - \frac{1}{12}\right) = \frac{14}{5} \cdot \frac{12-7}{7 \cdot 12} = \frac{14}{7 \cdot 12} = \frac{1}{6}.$$

Svaret är alltså C.

Uppgift 3

Korrelationskoefficienten

$$\rho(X,Y) = \frac{C(X,Y)}{D(X) \cdot D(Y)} = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X) \cdot V(Y)}}.$$

Vi har:

$$E(XY) = 1 \cdot 1 \cdot (0.1) = 0.1,$$

$$E(X) = 1 \cdot (0.3 + 0.1) = 0.4,$$

$$E(Y) = 1 \cdot (0.3) + 2 \cdot (0.3) = 0.9,$$

$$V(X) = E\left(X^2\right) - \left(E(X)\right)^2 = 0.4 - (0.4)^2 = (0.4)(0.6) = 0.24,$$

$$V(Y) = E\left(Y^2\right) - \left(E(Y)\right)^2 = (0.3 + 4 \cdot (0.3)) - (0.9)^2 = 1.5 - 0.81 = 0.69.$$

Nu får vi:

$$\rho(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{0.1 - (0.4)(0.9)}{\sqrt{(0.24)(0.69)}} = -0.6389.$$

Svaret är alltså A.

Uppgift 4

Vi har dragning utan återläggning ur en urna eftersom vi behåller de godisbitar som vi plockar ut. Antal möjliga utfall är $m = \binom{16}{4} = 16 \cdot 15 \cdot 14 \cdot 13 / \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 1820$ då vi har 16 olika godisbitar att välja mellan och vill plocka ut 4 stycken.

Eftersom det finns 7 sega råttor och vi vill plocka ut 2 stycken finns det $\binom{7}{2} = 7 \cdot 6/2 = 21$ sätt att göra detta på. På samma sätt finns det $\binom{4}{1} = 4$ sätt att välja ett geléhallon och $\binom{5}{1} = 5$ sätt att välja en pralin. Enligt multiplikationsprincipen är totala antalet gynnsamma utfall

$$g = 21 \cdot 4 \cdot 5 = 420,$$

vilket ger sannolikheten

$$\frac{g}{m} = \frac{420}{1820} \approx 0.231.$$

Svaret är alltså D.

Uppgift 5

Om $Z = X + Y$ har vi att Z är Bin$(9, 0.3)$-fördelad. Vidare har vi att

$$\begin{aligned} \mathrm{P}\left(X + Y \geq 2\right) &= \mathrm{P}\left(Z \geq 2\right) = 1 - \mathrm{P}\left(Z < 2\right) = 1 - \mathrm{P}\left(Z = 0\right) - \mathrm{P}\left(Z = 1\right) \\ &= 1 - \binom{9}{0} \cdot 0.3^0 \cdot 0.7^9 - \binom{9}{1} \cdot 0.3^1 \cdot 0.7^8 \\ &\approx 0.804. \end{aligned}$$

Svaret är alltså A.

Uppgift 6

Vi har alltså, eftersom $\ln(x)$ är strikt växande,

$$\begin{aligned} \mathrm{P}\left(X > a\right) &= 0.01 \\ \mathrm{P}\left(\ln(X) > \ln(a)\right) &= 0.01. \end{aligned}$$

Eftersom $\ln(X)$ är N$(10.41, 0.3650)$-fördelad uppfylls denna ekvation av $\ln(a) = 10.41 + 0.3650\lambda_{0.01}$, vilket ger

$$a = \exp(10.41 + 0.3650\lambda_{0.01}) = \exp(10.41 + 0.3650 \cdot 2.3263) \approx 77583.$$

Svaret är alltså B.

Uppgift 7

Väntevärdena fås enligt

$$E\left(\theta^\star\right) = \frac{1}{17}E\left(X\right) + \frac{8}{17}E\left(Y\right) = \frac{1}{17}\theta + \frac{8}{17} \cdot 2\theta = \theta$$

och

$$E\left(\widehat{\theta}\right) = \frac{2}{4}E\left(X\right) + \frac{1}{4}E\left(Y\right) = \frac{2}{4}\theta + \frac{1}{4} \cdot 2\theta = \theta,$$

så båda skattningarna är väntevärdesriktiga.
För varianserna får vi

$$V\left(\theta^\star\right) = \frac{1}{17^2}V\left(X\right) + \frac{8^2}{17^2}V\left(Y\right) = \frac{1}{289} \cdot 2 + \frac{64}{289} \cdot 1 = \frac{2 + 64}{289} = \frac{66}{289} \approx 0.228$$

och

$$V\left(\widehat{\theta}\right) = \frac{2^2}{4^2}V\left(X\right) + \frac{1}{4^2}V\left(Y\right) = \frac{4}{16} \cdot 2 + \frac{1}{16} \cdot 1 = \frac{8 + 1}{16} = \frac{9}{16} = 0.5625.$$

Alltså har $\theta^\star$ lägre varians jämfört med $\widehat{\theta}$ och är därför effektivare. Svaret är C.

Uppgift 8

Vi sätter upp funktionen $Q = Q(\lambda)$:

$$Q = Q(\lambda) = \sum_{i=1}^{4} \Big(x_i - E(X_i)\Big)^2 = \sum_{i=1}^{4} \Big(x_i - \frac{1}{\lambda}\Big)^2$$

Vi deriverar Q map λ, sätter sedan till 0 och löser ekvationen för λ. Vi har:

$$\frac{dQ}{d\lambda} = \sum_{i=1}^{4} 2\Big(x_i - \frac{1}{\lambda}\Big)\frac{1}{\lambda^2} = 0.$$

Detta ger $\sum_{i=1}^{4}\Big(x_i - \frac{1}{\lambda}\Big) = 0$, vilket blir

$$\sum_{i=1}^{4} x_i - \frac{4}{\lambda} = 0,$$

. där lösningen är

$$\lambda^*_{MK} = \frac{4}{\sum_{i=1}^{4} x_i} = \frac{4}{7.3 + 5.7 + 11.2 + 3.8} = 0.142857 = \frac{1}{7}.$$

Svaret är alltså D.

Uppgift 9

ML-skattningen $\theta^\star$ för θ fås av det aritmetiska medelvärdet

$$\theta^\star = \frac{1}{5}\sum_{i=1}^{5} X_i$$

där X_i är antalet besökare under dag i. Med de givna värdena får vi skattningen $\theta^\star_{\text{obs}} = 39$. Då detta värde är tillräckligt högt kan vi tillämpa normalapproximation för poissonfördelningen, vilket ger att $\theta^\star$ har den approximativa fördelningen $\mathrm{N}(\theta, \sqrt{\theta/5})$ (standardavvikelsen för X_i är $\sqrt{\theta}$ vilket ger standardavvikelsen $\sqrt{\theta/5}$ för $\overline{X}$). Stoppar vi in den observerade punktskattningen $\theta^\star_{\text{obs}} = 39$ får vi medelfelet $d(\theta^\star) = \sqrt{\theta^\star_{\text{obs}}/5} \approx 2.793$. Det approximativa konfidensintervallet blir då

$$39 \pm \lambda_{0.025} 2.793 \approx [33.5, 44.5].$$

Svaret är alltså B.

Uppgift 10

Styrkan är

$$h(5) = \mathrm{P}\left(X < 0.05; \theta = 5\right) = \int_0^0 .055e^{-5x}dx = \left[-e^{-5x}\right]_{x=0}^{0.05} = 1 - e^{-5\cdot 0.05} \approx 0.221.$$

Svaret är alltså B.

Uppgift 11

Eftersom det är t-baserad k.i., samt att σ_1 och σ_2 är okända och skattas med samma *pooled* stickprovs standardavvikelse s, dvs $\sigma_1^* = \sigma_2^* = \sigma^* = s$; så måste vi ha normalfördelningsantagande i botten, där σ_1 och σ_2 är både okända dock kan antas vara lika.
Svaret är alltså B.

Uppgift 12

Från datat har vi först de aritmetiska medelvärdena

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{117}{6} = 19.5 \quad \text{och} \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{2470.6}{6} \approx 411.7667.$$

Med dessa ger formelsamlingen

$$s_{xy} = \sum_{i=1}^{n} x_i y_i - n\overline{xy} \approx 48218.4 - 6 \cdot 19.5 \cdot 411.7667 \approx 41.6961$$

och

$$s_{xx} = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2 = 2299 - 6 \cdot 19.5^2 = 17.5.$$

Detta ger då punktskattningarna

$$\beta^\star_{\text{obs}} = \frac{s_{xy}}{s_{xx}} \approx \frac{41.6961}{17.5} \approx 2.38263$$

och

$$\alpha^\star_{\text{obs}} = \overline{y} - \beta^\star_{\text{obs}}\overline{x} \approx 411.7667 - 2.38263 \cdot 19.5 \approx 365.3054.$$

Om vi nu stoppar in $x = 30$ (motsvarande år 2030) får vi

$$\alpha^\star_{\text{obs}} + \beta^\star_{\text{obs}} \cdot 30 \approx 365.3054 + 2.38263 \cdot 30 \approx 436.8.$$

Svaret är alltså B.

Del II, Lösningsförslag:

Uppgift 13

Låt p vara sannolikheten att man träffar vid ett enskilt skott.
Sannolikheten att inte träffa vid ett enskilt skott är därför lika med $(1-p)$. Och eftersom varje skott är oberoende av de andrea, så är $(1-p)^4$ sannolikhet att inte träffa alls när man skjuter fyra gånger. Sannolikheten alltså att träffa minst en gång när man skjuter fyra gånger blir

$$1-\left(1-p\right)^4,$$

vilket enligt lydelsen är 0.9984.
Dvs vi har:

$$1-\left(1-p\right)^4=0.9984.$$

Eller, när vi löser denna ekvation för p, vi får:

$$p=1-(1-0.9984)^{1/4}=0.8$$

Att man träffar vid ett enskilt skott är sannolikheten alltså lika med 0.8.

Uppgift 14

Vi söker alltså n så att

$$\mathrm{P}\left(|\overline{X}-\mu|<0.1\right)=0.95.$$

Eftersom n går mot oändligheten kan vi anta att n är tillräckligt stort för att centrala gränsvärdensatsen ska hålla. Vi har ju också att $X_1, X_2, \ldots, X_n$ är oberoende och likafördelade, vilket ger att

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

är approximativt normalfördelad med väntevärde $E\left(\overline{X}\right)=E\left(X_1\right)=\mu$ samt $D\left(\overline{X}\right)=D\left(X_1\right)/\sqrt{n}$.
Med andra ord har vi att

$$\mathrm{P}\left(|\overline{X}-\mu|<0.1\right)=\mathrm{P}\left(-\frac{0.1\sqrt{n}}{D\left(X_1\right)}<\frac{|\overline{X}-\mu|}{D\left(X_1\right)/\sqrt{n}}<\frac{0.1\sqrt{n}}{D\left(X_1\right)}\right)\approx\mathrm{P}\left(-\frac{0.1\sqrt{n}}{D\left(X_1\right)}<Z<\frac{0.1\sqrt{n}}{D\left(X_1\right)}\right),$$

där Z är en $\mathrm{N}(0,1)$-fördelad stokastisk variabel då $|\overline{X}-\mu|/(D\left(X_1\right)/\sqrt{n})$ är approximativt normalfördelad med väntevärde noll och standardavvikelse ett. Vi vet att sannolikheten ovan är lika med 0.95 då

$$\frac{0.1\sqrt{n}}{D\left(X_1\right)}=\lambda_{0.025},$$

så vi löser ut n och får

$$n=\frac{V\left(X_1\right)\lambda_{0.025}^2}{0.1^2}.$$

Eftersom $V\left(X_1\right)$ inte är känt måste vi skatta denna. Från datan har vi $s^2\approx 103.067$ så vi använder denna approximation då 1000 anses vara tillräckligt stort (vi ser t.ex. att den skattade standardavvikelsen $s/\sqrt{n}\approx 0.321$ för $\overline{X}$ är liten jämfört med det observerade värdet $\overline{x}\approx 10.144$). Vi får alltså

$$n\approx\frac{103.067\cdot 1.96^2}{0.1^2}\approx 39594.$$

Vi avrundar uppåt (för att sannolikheten ska överstiga 0.95) och får då 39595 som svar.

Uppgift 15

(a) **Lösning:Alt 1:** Vi har alltså $y = g(x)$ där $g(x) = -\ln(x)$. Inversen ges av $x = g^{-1}(y) = e^{-y}$ och vi kan då använda formeln för $Y = g(X)$ som ger täthetsfunktionen

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}}{dy}(y) \right|.$$

Absolutvärdet av derivatan ges av

$$\left| \frac{dg^{-1}}{dy}(y) \right| = \left| -e^{-y} \right| = e^{-y}.$$

Eftersom $f_X(x)$ endast är nollskild då $x \in (0,1)$ så är $f_X(g^{-1}(y)) = f_X(e^{-y})$ nollskild då $e^{-y} \in (0,1)$, dvs. då $y \in (0, +\infty)$. Vi får då alltså

$$f_Y(y) = \begin{cases} \alpha \left(e^{-y}\right)^{\alpha-1} \cdot e^{-y}, & y > 0, \\ 0, & \text{annars} \end{cases} = \begin{cases} \alpha e^{-\alpha y}, & y > 0, \\ 0, & \text{annars.} \end{cases}$$

Detta ger att Y är exponentialfördelad med parameter α.

Lösning:Alt 2: $Y = -lnX$.

$$F_Y(y) = P(Y \leq y) = P(-lnX \leq y) = P(lnX \geq -y) =$$

$$= P(X \geq e^{-y}) = [0 < x < 1] = \int_{e^{-y}}^{1} \alpha x^{\alpha-1} dx = [x^\alpha]_{e^{-y}}^{1} = 1 - e^{-\alpha y}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = [0 < x < 1] = \alpha e^{-\alpha y}$$

Vi får då alltså

$$f_Y(y) = \begin{cases} \alpha e^{-\alpha y}, & y > 0, \\ 0, & \text{annars.} \end{cases}$$

Detta ger att Y är exponentialfördelad med parameter α.

(b) Först bildar vi log-likelihoodfunktionen

$$\ln L(\alpha) = \ln \left(\alpha^n \prod_{i=1}^{n} x_i^{\alpha-1} \right) = n \ln \alpha + (\alpha - 1) \sum_{i=1}^{n} \ln x_i.$$

Vi deriverar sedan och sätter till noll, vilket ger

$$\frac{d}{d\alpha} \ln L(\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n} \ln x_i = 0 \Rightarrow \alpha = -\frac{n}{\sum_{i=1}^{n} \log x_i}.$$

Då andraderivatan ges av

$$\frac{d^2}{d\alpha^2} \ln L(\alpha) = -\frac{n}{\alpha^2}$$

och således är negativ för alla $\alpha > 0$ har vi att vår kritiska punkt är det enda lokala maximat och därför också det globala maximat. ML-skattningen ges då av

$$\alpha^\star_{\text{obs}} = -\frac{n}{\sum_{i=1}^{n} \ln x_i}.$$

Uppgift 16

a) Styrkan i $\Delta = 0.34$ för både testen. Vi testar alltså $H_0 : \Delta = 0$ versus $H_1 : \Delta = 0.34$.
Det ena testets styrka:

$$h_1(0.34) = P(\bar{X} \geq 0.179|\Delta = 0.34) = 1 - \Phi\Big(\frac{0.179 - 0.34}{0.4/\sqrt{9}}\Big) = 0.8869 \approx 0.89$$

Och det andra testets styrka:

$$h_2(0.34) = P(D \geq 7|\Delta = 0.34) = P(D \geq 7|D \in Bin(9, 0.8023)) = 0.7382 \approx 0.74,$$

där, enl Tabell 1, $0.8023 = P(X_i \geq 0) = |X_i \in N(0.34, 0.4)| = 1 - \Phi\big(\frac{0-0.34}{0.4}\big) = \Phi(0.85)$, och d är ett observerat värde på s.v. D.

b) Test 1, dvd testet baserat på medelvärdet $\bar{x}$, ty den har högre styrkan (ca 89%).