MATEMATISKA VETENSKAPER
Chalmers Tekniska Högskola

Tentamen för MVE051/MSG810 – Matematisk statistik och diskret matematik

3/6 – 2023, 8.30–12.30

Hjälpmedel: Typgodkänd miniräknare, formelsamlingen BETA, ett A4-papper med valfritt handskrivet innehåll på båda sidor
Telefonvakt: Malin Palö Forsström, telefon 0721 63 32 79
Betygsgränser: Chalmers: 3: 12-17p, 4: 18-23p, 5: 24-30p, GU: G: 12-21, VG: 22-30.
Till alla uppgifter skall fullständiga lösningar inlämnas. Motivera och förklara så väl du kan. Poäng ges för lösningar som går att följa. Endast svar ger inga poäng.

---

1. Enligt SMHI var mängden regn som föll över Göteborg under dagarna februari 2023:

0.1	6.4	2.6	0.0	2.0	0.0	0.0	0.0	0.4	5.3	0.0	0.0	0.0	0.0
0.4	2.5	14.5	0.0	19.1	1.1	0.0	1.6	10.9	3.9	0.0	0.0	0.0	0.0

   Beräkna the dre kvantilerna och rita en boxplot som illustrerar hur mycket det regnade.

   (4p)

2. Tiden mellan två stjärnfall som kan ses med ett teleskop kan antas vara expenentialfördelat, och tiden mellan två sådana stjärnfall är i snitt 10 minuter.

   (a) Vad är sannolikheten att inget stjärnfall kan ses under en timma? (2p)

   (b) Hur många stjärnfall kan ses per timma i snitt? (2p)

   (c) Malin och Anna turas om att titta i teleskopet. Anna får börja, och de byter sedan plats varannan minut. Vad är sannolikheten att Anna ser ett stjärnfall före Malin? (4p)

3. Sommaren 2021 var medianpriset m_{2021} på piggelin i glasskioskerna i Göteborg 8.50 kr. En student tycker att det verkar som att glasspriserna är högre sommaren 2022 och kontrollerar därför priserna på Piggelin i 13 slumpvis valda glasskiosker en varm sommardag 2022.

   (a) Anta att medianpriset m_{2022} (för hela populationen) är 8.50 kr även 2022. Vad är i så fall sannolikheten att en slumpvis vald glass (sommaren 2022) har ett pris som är högre än medianen? (1p)

   (b) Låt X vara antalet glasskiosker i stickprovet som hade ett pris som var högre än 8.50 kr. Anta att medianpriset (för hela populationen) är 8.50 kr även 2022. Vad för fördelning har i så fall X? (1p)

(c) Studenten planerar att göra ett hypotestest för att testa $H_0\colon m_{2022} = 8.50$ mot $H_1\colon m_{2022} > 8.50$, och förkasta H_0 om Piggelinpriset är högre än 8.50 i minst 10 av glasskioskerna (det här kallas för ett teckentest). Hur stor är sannolikheten att studenten gör ett typ-I-fel om H_0 stämmer? (3p)

4. När man vill ha en slumpvariabel X som följer en viss fördelning är det vanligt att man först slumpar U från en likformig/uniform fördelning på intervallet $(0, 1)$ och sedan låter $X = f(U)$ för någon funktion f som ser till att X får rätt fördelning.

   Antag att U har en likformig/uniform fördelning på intervallet $(0, 1)$ och att $X = -\log(1 - U)$.

   (a) Är U och X oberoende? (1p)

   (b) Låt $t > 0$. Vad är $P(X < t)$? (2p)

   (c) Vad har X för fördelning? (1p)

5. Lös rekurrensekvationen $a_{j+2} = 2a_{j+1} + a_j$, $a_0 = 0$, $a_1 = 1$, dvs. ge en sluten formel för a_n. (Serien börjar med 0,1,2,5,12,29,...) (4p)

6. Malin har samlat in n positiva heltal genom en undersökning. För att testa hypotesen $H_0\colon \mu = 1$ mot $H_0\colon \mu \neq 1$ har hon tänkt göra ett t-test.

   (a) Vad behöver gälla för att Malin ska kunna lita på siffrorna hon får ut från testet, dvs. vilka antaganden krävs för att Malin ska kunna göra ett t-test? (3p)

   (b) Förklara till vad Malin använder centrala gränsvärdessatsen. (2p)

Lycka till!

Lösningar till

Tentamen för MVE051/MSG810 – Matematisk statistik och diskret matematik

3/4 – 2023, 14.00–18.00

1. Om vi sorterar datan får vi

0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.4	0.4	1.1	1.6	2.0	2.5	2.6	3.9	5.3	6.4	10.9	14.5	19.1

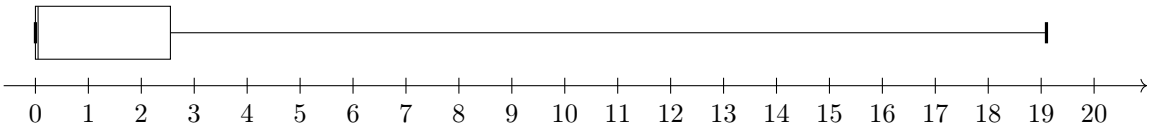
Om vi tar medelvärdet av de två talen i mitten får vi medianen $Q2 = (0.0+0.1)/2 = 0.05$.

Vi har totalt 28 tal, och $28/4 = 7$. Både det 7:e och det 8:onde talet är lika med 0.0, så vi får $Q1 = 0.0$.

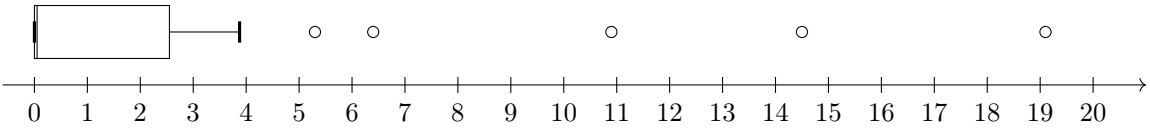
Det 21:e och 22:e talen är 2.5 och 2.6, så därför är $Q3 = (2.5 + 2.6)/2 = 2.55$.

Vi har $Q_2 + (Q_3 - Q_1) \cdot 1.5 = 0.05 + (2.55 - 0) \cdot 1.5 = 3.875$ och $Q_3 + (Q_3 - Q_1) \cdot 1.5 = 2.55 + (2.55 - 0) \cdot 1.5 = 6.375$

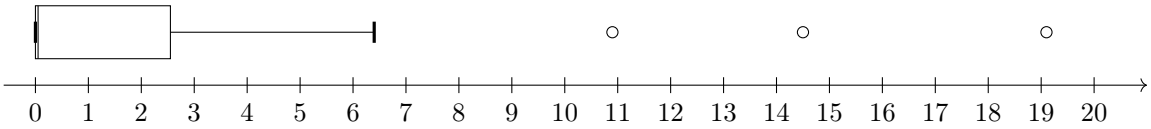
En boxplot ser typiskt ut så här (horisontell för att spara plats):



eller så här:



eller så här:



2. (a) Tiden T mellan två stjärnfall är exponentialfördelad med parameter $\lambda = 1/10$. Sannolikheten att inga stjärnfall kan ses under en timma är därför

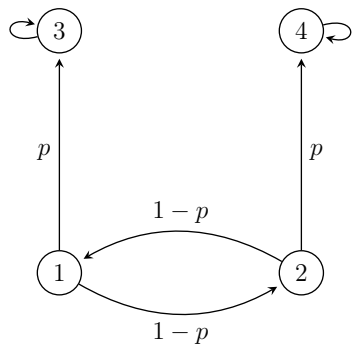
$$P(T > 60) = e^{-\lambda 60} = e^{-6}.$$

(b) Eftersom att $T \sim \exp(\lambda)$ är antalet stjärnfall under en timma Poissonfördelat med parameter $\lambda' = 60/10$. Antalet stjärnfall i snitt per timma är därför 60/10.

(c) Den här uppgiften går att lösa på flera olika sätt:

Sannolikheten för att man kan se ett stjärnfall inom de närmsta två minuterna är $p = 1 - e^{-2\lambda} = 1 - e^{-2/10}$.

Alternativ 1: Problemet kan beskrivas som en Markovkedja med fyra tillstånd: (1) Anna tittar, (2) Malin tittar, (3) Anna ser stjärnfall, (4) Malin ser stjärnfall. Övergångsmatrisen blir då



Övergångsmatrisen blir

$$P = \begin{pmatrix} 0 & 1-p & p & 0 \\ 1-p & 0 & 0 & p \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Så vi får

$$Q = \begin{pmatrix} 0 & 1-p \\ 1-p & 0 \end{pmatrix} \quad N = (I-Q)^{-1} = \frac{1}{1-(1-p)^2}\begin{pmatrix} 1 & 1-p \\ 1-p & 1 \end{pmatrix}$$

och

$$NR = \frac{1}{1-(1-p)^2}\begin{pmatrix} 1 & 1-p \\ 1-p & 1 \end{pmatrix}\begin{pmatrix} p & 0 \\ 0 & p \end{pmatrix} = \frac{1}{1-(1-p)^2}\begin{pmatrix} p & p(1-p) \\ p(1-p) & p \end{pmatrix}$$

så svaret blir $\frac{p}{1+(1-p)^2}$

Alternativ 2: Låt p_A vara sannolikheten att Anna ser ett stjärnfall först. Då är

$$p_A = p + (1-p)^2 p_A \Leftrightarrow p_A = \frac{p}{1-(1-p)^2}.$$

Alternativ 3: "Numret" X på det första två-minuters-intervall när man kan se ett stjärnfall följer en geometrisk fördelning med parameter p. Anna ser det första stjärnfallet om X är udda. Sannolikheten blir därför

$$\sum_{j=1}^{\infty}(1-p)^{j-1}p \cdot \mathbb{1}(j \text{ udda}) = p\sum_{i=0}^{\infty}(1-p)^{2i} = p\sum_{i=0}^{\infty}\big((1-p)^2\big)^i = \frac{p}{1-(1-p)^2}.$$

3. (a) 1/2 (enligt definitionen av medianen)
   (b) Binomialfördelad med $n = 13$ och $p = 1/2$.
   (c)

$$
\begin{aligned}
P(X \geqslant 10) &= P(X=10) + P(X=11) + P(X=12) + P(X=13) \\
&= \binom{13}{10}2^{-13} + \binom{13}{11}2^{-13} + \binom{13}{12}2^{-13} + \binom{13}{13}2^{-13} \\
&= \left(\binom{13}{10} + \binom{13}{11} + \binom{13}{12} + \binom{13}{13}\right)2^{-13} \\
&= \left(\binom{13}{3} + \binom{13}{2} + \binom{13}{1} + \binom{13}{0}\right)2^{-13} = \frac{\frac{13\cdot 12\cdot 11}{3\cdot 2\cdot 1} + \frac{13\cdot 12}{2\cdot 1} + \frac{13}{1} + 1}{2^{13}} \\
&= 189/4096 = 0.0461426.
\end{aligned}
$$

Svar: Ungefär 4.6%.

4. Låt $A(x) = \sum_{j=0}^{\infty} a_j x^j$. Då är

$$
\begin{aligned}
a_{j+2} = 2a_{j+1} + a_j &\Leftrightarrow a_{j+2}x^{j+2} = 2a_{j+1}x^{j+2} + a_j x^{j+2} \\
&\Leftrightarrow \sum_{j=0}^{\infty} a_{j+2}x^{j+2} = 2\sum_{j=0}^{\infty} a_{j+1}x^{j+2} + \sum_{j=0}^{\infty} a_j x^{j+2} \Leftrightarrow \sum_{j=2}^{\infty} a_j x^j = 2x\sum_{j=1}^{\infty} a_j x^j + x^2\sum_{j=0}^{\infty} a_j x^j \\
&\Leftrightarrow A(x) - x = 2xA(x) + x^2A(x) \Leftrightarrow -x = A(x)(x^2+2x-1) \\
&\Leftrightarrow A(x) = \frac{-x}{x^2+2x-1} = \frac{-x}{(x+1)^2-2} = \frac{-x}{(x+1+\sqrt{2})(x+1-\sqrt{2})} \\
&= -x\left(\frac{1}{x+1+\sqrt{2}} + \frac{1}{x+1-\sqrt{2}}\right) = \frac{-x}{2\sqrt{2}}\left(\frac{1}{x+1-\sqrt{2}} - \frac{1}{x+1+\sqrt{2}}\right) \\
&= \frac{-x}{2\sqrt{2}}\left(\frac{1}{1-\sqrt{2}}\frac{1}{x/(1-\sqrt{2})+1} - \frac{1}{1+\sqrt{2}}\frac{1}{x/(1+\sqrt{2})+1}\right) \\
&= \frac{-x}{2\sqrt{2}}\left(\frac{1}{1-\sqrt{2}}\sum_{j=0}^{\infty}\left(-x/(1-\sqrt{2})\right)^j - \frac{1}{1+\sqrt{2}}\sum_{j=0}^{\infty}\left(-x/(1+\sqrt{2})\right)^j\right) \\
&= \sum_{j=0}^{\infty} x^{j+1}\cdot\frac{1}{2\sqrt{2}}\left(\frac{1}{(\sqrt{2}-1)^{j+1}} - \frac{(-1)^{j+1}}{(\sqrt{2}+1)^{j+1}}\right) \\
&= \sum_{j=1}^{\infty} x^j \cdot \underbrace{\frac{1}{2\sqrt{2}}\left(\frac{1}{(\sqrt{2}-1)^j} - \frac{(-1)^j}{(\sqrt{2}+1)^j}\right)}_{a_j}.
\end{aligned}
$$

5. (a) Nej, för om vi vet vad U är så vet vi exakt vad X är.

   (b)
$$
\begin{aligned}
P(X<t) &= P(-\log(1-U) < t) = P(\log(1-U) > -t) = P(1-U > e^{-t}) \\
&= P(U < 1-e^{-t}) = 1-e^{-t}
\end{aligned}
$$

   (c) X är exponentialfördelad med parameter 1.

6. (a) De n talen i stickprovet behöver vara oberoende och ha samma fördelning, och ha ändligt medelvärde μ och ändlig varians σ^2. För att approximationen med

CGS ska vara ok behöver $\mathrm{Var}(X_1 + \cdots + X_n) = n^2\sigma^2$ vara någorlunda stort (dvs. om σ är mindre måste n vara större).

(b) Malin använder centrala gränsvärdessatsen för att dra slutsatsen att $\bar{X}$ är approximativt normalfördelad, vilket är precis vad som behövs för att teststatistiskan approximativt ska följa en t-fördelning.