Tentamentsskrivning i **Matematisk Statistik TMA321**

Tid: den 28 augusti, 2019
Hjälpmedel: Typgodkänd miniräknare, egenhändigt skriven formelsamling om två A4 fram och bak (dvs 4 sidor), samt utdelade tabeller.

Tentamen består av 8 frågor om sammanlagt 50 poäng. Preliminära betygsgränser är satta till:
betyg "3": 20 till 29 poäng
betyg "4": 30 till 39 poäng
betyg "5": 40 eller fler poäng.

OBS! Alla lösningar skall vara väl redovisade och motiverade. Talen är ej ordnade efter svårighetsgrad.

1. En bridgehand består av 13 kort dragna från en standardkortlek bestående av 52 kort. I systemet HCP (High Card Points) tilldelas ett äss 4 poäng, en kung 3 poäng, en dam 2 poäng och en knekt 1 poäng. Alla andra kort tilldelas noll poäng.

   (a) Hur många bridgehänder ger en totalpoäng som är exakt 3? (3p)

   (b) Hur många bridgehänder ger en totalpoäng som är större än eller lika med 3? (3p)

   Lösning:

   (a) En bridgehand kan ha 3 poäng på tre olika sätt. Antingen innehåller den bara en kung som det enda poängkortet, eller också en dam och en knekt, eller så innehåller den tre knektar.
   Antalet händer med en kung och inga andra poängkort är

   $$\binom{4}{1}\binom{36}{12},$$

   ty kungen kan väljas på fyra olika sätt, och de resterande 12 korten måste väljas bland de 52-16=36 poänglösa korten.
   Antalet händer med en dam och en knekt men inga andra poängkort är

   $$\binom{4}{1}\binom{4}{1}\binom{36}{11},$$

   ty damen och knekten kan väljas på fyra olika sätt, och de resterande 11 korten väljs som ovan.
   Till sist har vi fallet med tre knektar, och vi ser att antalet händer blir här

   $$\binom{4}{3}\binom{36}{10}$$

Tillsammans får vi då

$$\binom{4}{1}\binom{36}{12} + \binom{4}{1}\binom{4}{1}\binom{36}{11} + \binom{4}{3}\binom{36}{10}.$$

(b) Det är enklast att beräkna antalet händer med två, en eller noll poäng. Liknande som ovan ser vi att antalet händer som ger exakt två poäng är

$$\binom{4}{1}\binom{36}{12} + \binom{4}{2}\binom{36}{11},$$

att antalet händer som ger exakt en poäng är

$$\binom{4}{1}\binom{36}{12},$$

och att antalet händer som ger exakt noll poäng är

$$\binom{36}{13}.$$

Det sökta antalet händer blir därför

$$\binom{52}{13} - \binom{4}{1}\binom{36}{12} - \binom{4}{2}\binom{36}{11} - \binom{4}{1}\binom{36}{12} - \binom{36}{13}.$$

2. Låt $0 < \epsilon < 1$ och låt X, Y, B vara tre oberoende slumpvariabler med följande fördelningar: X är exponentialfördelad med parameter 1, Y är exponentialfördelad med parameter $1-\epsilon$ och B är Bernoulli-fördelad med parameter ϵ (dvs $\mathbb{P}(B=1)=\epsilon$ medans $\mathbb{P}(B=0)=1-\epsilon$).

   (a) Beräkna den momentgenererande funktionen för X (OBS: Tydlig redovisning krävs då man kan antas ha svaret på sin formelsamling). (2p)

   (b) Beräkna den momentgenererande funktionen för BY. (2p)

   (c) Vilken fördelning har slumpvariabeln Z om $Z = X + BY$? (2p)

Lösning:

(a) Vi har att

$$\begin{aligned} M_X(t) = \mathbb{E}\left[e^{Xt}\right] &= \int_0^\infty e^{xt}e^{-x}dx \\ &= \int_0^\infty e^{-(1-t)x}dx = \frac{1}{1-t}\left[-e^{-(1-t)x}\right]_0^\infty = \frac{1}{1-t}, \end{aligned}$$

för $t < 1$.

(b) Vi har att

$$\begin{aligned}
M_{BY}(t) &= \mathbb{E}\left[e^{BYt}\right] = \mathbb{E}\left[e^{BYt}|B=1\right]\mathbb{P}(B=1) + \mathbb{E}\left[e^{BYt}|B=0\right]\mathbb{P}(B=0) \\
&= \mathbb{E}\left[e^{Yt}\right]\epsilon + \mathbb{E}\left[e^{0t}\right](1-\epsilon) = \frac{1-\epsilon}{1-\epsilon-t}\epsilon + 1 - \epsilon = \frac{1-\epsilon}{1-\epsilon-t}(\epsilon + 1 - \epsilon - t) \\
&= \frac{(1-\epsilon)(1-t)}{1-\epsilon-t}
\end{aligned}$$

för $t < 1 - \epsilon$.

(c) Vi ser att

$$\begin{aligned}
M_{X+BY}(t) &= \mathbb{E}\left[e^{X+BYt}\right] = \mathbb{E}\left[e^{Xt}\right]\mathbb{E}\left[e^{BYt}\right] \\
&= \frac{1}{1-t}\frac{(1-\epsilon)(1-t)}{1-\epsilon-t} = \frac{1-\epsilon}{1-\epsilon-t}
\end{aligned}$$

och denna momentgenererande funktion känner vi igen och ser att $Z \sim$Exp$(1-\epsilon)$.

3. Låt $X_1, X_2, \ldots$ vara oberoende normalfördelade slumpvariabler där $X_n \sim$N$(n\mu, \sigma^2)$. Låt nu

$$Y = \frac{X_1 + X_2 + X_3}{3}, \quad Z = \frac{X_1 + X_2 + X_3}{6} \text{ och } W = \frac{X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3}{3}$$

(a) Vilka av Y, Z och W är väntevärdesriktiga skattare av μ? (2p)

(b) Vilken/vilka av de väntevärdesriktiga skattarna är effektivast? (2p)

(c) Låt

$$W_n = \frac{X_1 + \frac{1}{2}X_2 + \cdots + \frac{1}{n}X_n}{n}.$$

Är W_n en konsistent skattare? (2p)

Lösning:

(a) Vi har att

$$\mathbb{E}[Y] = \frac{\mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3]}{3} = \frac{\mu + 2\mu + 3\mu}{3} = 2\mu,$$

$$\mathbb{E}[Z] = \frac{\mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3]}{6} = \frac{\mu + 2\mu + 3\mu}{6} = \mu,$$

och

$$\mathbb{E}[W] = \frac{\mathbb{E}[X_1] + \frac{1}{2}\mathbb{E}[X_2] + \frac{1}{3}\mathbb{E}[X_3]}{3} = \frac{\mu + \mu + \mu}{3} = \mu,$$

så att Z, W är väntevärdesriktiga men inte Y.

(b) För att jämföra effektiviteten mellan väntevärdesriktiga skattare beräknar vi varianserna. Vi ser att

$$\mathrm{Var}(Z) = \frac{\mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \mathrm{Var}[X_3]}{6^2} = \frac{\sigma^2 + \sigma^2 + \sigma^2}{36} = \frac{\sigma^2}{12},$$

och

$$\mathrm{Var}(W) = \frac{\mathrm{Var}[X_1] + \frac{1}{4}\mathrm{Var}[X_2] + \frac{1}{9}\mathrm{Var}[X_3]}{9} = \frac{36\sigma^2 + 9\sigma^2 + 4\sigma^2}{36 \cdot 9} = \frac{\sigma^2}{12}\frac{49}{27},$$

så vi ser att Z är effektivast.

(c) Vi har att för varje $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}(|W_n - \mu| > \epsilon) &\leq \frac{\mathrm{Var}(W_n)}{\epsilon^2} \\ &= \frac{1}{\epsilon^2}\frac{\mathrm{Var}(X_1) + \frac{1}{4}\mathrm{Var}(X_2) + \cdots + \frac{1}{n^2}\mathrm{Var}(X_n)}{n^2} \\ &= \frac{1}{\epsilon^2}\frac{\sigma^2}{n^2}\sum_{k=1}^{n}\frac{1}{k^2} \leq \frac{1}{\epsilon^2}\frac{\sigma^2}{n^2}\sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{C}{\epsilon^2}\frac{\sigma^2}{n^2}\end{aligned}$$

där vi använt Chebyshevs olikhet och att $\sum_{k=1}^{\infty}\frac{1}{k^2}$ är ändlig. Vi ser att $\lim_{n\to\infty}\mathbb{P}(|W_n - \mu| > \epsilon) = 0$ och därför är W_n en konsistent skattare av μ.

4. Företaget Georgios solceller tillverkar solceller. Georgios vill testa hur stor effekt hans solceller producerar som funktion av ljusstyrkan. Han samlar därför ihop data och sammanställer i följande tabell.

Ljusstyrka (Lumen)	100	150	200	250	300	350	400	450
Effekt (W)	0.9	2.3	4.3	5.6	5.7	7.8	8.9	10.9

Georgios ansätter en linjär regressionsmodell $y = \beta_0 + \beta_1 x$ där y är effekten och x är ljusstyrkan. Data sammanfattas med att $S_{xx} = 105000$, $S_{yy} \approx 78.18$ och $S_{xy} = 2840$.

Hjälp Georgios med att lösa följande uppgifter:

(a) Skatta β_0 och β_1. (2p)

(b) Skapa ett 95% konfidensintervall för β_1 och testa huruvida $\beta_1 = 0$ på 95%-nivån. (2p)

(c) Ange förklaringsgraden och beräkna residualerna. Kommentera ditt resultat. (2p)

(d) Vilken effekt bör man förvänta sig om lusstyrkan uppgår till 500 lumen? Vad händer vid 20 lumen? Kommentera dina resultat. (1p)

Lösning:

(a) Vi har att

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \approx 0.027 \text{ och } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \approx -1.64.$$

(b) Vi betraktar hypteserna $H_0 : \beta_1 = 0$ och $H_1 : \beta_1 \neq 0$. Vi använder att

$$\frac{\hat{\beta}_1 - \beta_1}{s_r/\sqrt{S_{xx}}} \sim t(6)$$

och får med hjälp av tabell att $t_{0.025}(6) \approx 2.447$ så att

$$\begin{aligned} 0.95 &= \mathbb{P}\left(-2.447 \leq \frac{\hat{\beta}_1 - \beta_1}{s_r/\sqrt{S_{xx}}} \leq 2.447\right) \\ &= \mathbb{P}\left(\hat{\beta}_1 - 2.447\frac{s_r}{\sqrt{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + 2.447\frac{s_r}{\sqrt{S_{xx}}}\right). \end{aligned}$$

Vi har att

$$s_r^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) \approx 0.2275$$

så att $s_r = 0.477$. Ett 95% numeriskt K.I. för β_1 blir då

$$I_{\beta_1} = \hat{\beta}_1 \pm 2.447\frac{s_r}{\sqrt{S_{xx}}} \approx [0.0234, 0.0306].$$

Då $0 \notin I_{\beta_1}$ förkastar vi H_0 på 95% nivån.

(c) Förklaringsgraden blir

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \approx 0.9825$$

vilket är mycket bra. Residualerna är $e_k = y_k - (\hat{\beta}_0 + \hat{\beta}_1 x_k)$ och vi får att

e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
-0.17	-0.12	0.53	0.48	-0.78	-0.03	-0.28	0.37

Det är något anmärkningsvärt att de stora residualerna alla hamnar i mitten, men inga tydliga mönster kan observeras så vi kan inte dra slutsatsen att ett specifikt annat samband skulle passa bättre. Man bör dock vara något försiktig då antalet mätpunkter är så lågt.

(d) Enligt vår modell får vi att med $x = 500$ så bör effekten bli ca 11.89 W medans för 20 lumen så blir den ca -1.10 W. Den senare siffran är anmärkningsvärd då den är fullständigt orimlig. Man skall alltid vara försiktig med extrapolera utanför sitt testområde och här är det rimligt att tro att solcellerna kanske behöver en minsta ljusstyrka för att överhuvudtaget fungera. Om man då går under den så får man bara nonsens.

5. Låt $X_1, \dots, X_n$ vara i.i.d. (dvs oberoende och likafördelade) med fördelning
$$f(x) = \frac{2m^2}{x^3} \text{ för } x \geq m,$$
där $m > 0$ är en parameter vars värde är okänt. Betrakta hypoteserna
$$H_0 : m = 2 \text{ och } H_1 : m \neq 2.$$
   (a) Bestäm den generaliserade likelihood-ration och motsvarande förkastningsregion. (3p)
   (b) Bestäm enklast möjliga test utifrån ditt svar i (a). Bestäm dessutom en lämplig förkastningsregion för ditt test om signifikansnivån skall vara 1% och $n = 10$. (3p)

Lösning:

(a) Den generaliserade likelihood ration ges av
$$\Lambda(m) = \frac{\text{lik}(2)}{\max_{m>0} \text{lik}(m)}.$$
Vi kan skriva täthetsfunktionen som
$$\frac{2m^2}{x^3} I(x \geq m),$$
så att likelihooden blir
$$\text{lik}(m) = \prod_{k=1}^{n} \frac{2m^2}{X_k^3} I(X_k \geq m) = 2^n m^{2n} I(\min(X_1, \dots, X_n) \geq m) \left(\prod_{k=1}^{n} X_k \right)^{-3}.$$
Denna maximeras då $\hat{m} = \min(X_1, \dots, X_n)$ (vilket för övrigt är en ML-skattare av m).
Den generaliserade likelihood ration blir då
$$\begin{aligned}
\Lambda(m) &= \frac{\text{lik}(2)}{\text{lik}(\hat{m})} \\
&= \frac{2^n 2^{2n} I(\min(X_1, \dots, X_n) \geq 2) \left(\prod_{k=1}^{n} X_k\right)^{-3}}{2^n \hat{m}^{2n} I(\min(X_1, \dots, X_n) \geq \hat{m}) \left(\prod_{k=1}^{n} X_k\right)^{-3}} \\
&= \left(\frac{2}{\hat{m}} \right)^{2n} I(\min(X_1, \dots, X_n) \geq 2).
\end{aligned}$$
Observera att $I(\min(X_1, \dots, X_n) \geq \hat{m}) = I(\min(X_1, \dots, X_n) \geq \min(X_1, \dots, X_n)) = 1$ pga definitionen av $\hat{m}$. Paradigmen säger att vi skall förkasta H_0 om Λ är liten, så vår förkastningsregion kan skrivas på formen
$$\left(\frac{2}{\hat{m}} \right)^{2n} I(\min(X_1, \dots, X_n) \geq 2) = \left(\frac{2}{\hat{m}} \right)^{2n} I(\hat{m} \geq 2) \leq c, \tag{1}$$
där c väljs utefter önskad signifikansnivå.

(b) Det blir enklast att formulera testet i termer av $\hat{m}$. Vi ser från (**??**) att vi förkastar H_0 om antingen $\hat{m} < 2$ eller om $\hat{m}$ "blir för stor", dvs om $\hat{m} \geq \gamma$ där γ skall väljas utifrån föreskriven signifikansnivå.

För att beräkna γ börjar vi med att observera att under $H_0 : m = 2$ så är villkoret $\min(X_1, \ldots, X_n) \geq 2$ automatiskt uppfyllt. Sannolikheten att göra ett Typ I fel ges därför av

$$\alpha = \mathbb{P}(\hat{m} \geq \gamma | m = 2)$$

(Observera att $\mathbb{P}(\hat{m} < 2 | m = 2) = 0$). Eftersom signifikansnivån är 1% måste då γ väljas så att

$$\begin{aligned} 0.01 &= \mathbb{P}(\hat{m} \geq \gamma | m = 2) = \mathbb{P}(\min(X_1, \ldots, X_n) \geq \gamma | m = 2) \\ &= \mathbb{P}(X_1 \geq \gamma, \ldots, X_n \geq \gamma | m = 2) = \mathbb{P}(X_1 \geq \gamma | m = 2)^n \\ &= \left(\int_\gamma^\infty \frac{2 \cdot 2^2}{x^3} dx \right)^n = \left(8 \left[\frac{x^{-2}}{-2} \right]_\gamma^\infty \right)^n = \left(\frac{2}{\gamma} \right)^{2n}. \end{aligned}$$

Vi ser att

$$\frac{2}{\gamma} = 0.01^{1/(2n)} \text{ så att } \gamma = \frac{2}{0.01^{1/(2n)}}.$$

Med $n = 10$ ser vi då att

$$\gamma = \frac{2}{0.01^{1/20}} \approx 2.52.$$

Vi förkastar därför H_0 om

$$\hat{m} \in (-\infty, 2) \cup [2.52, \infty).$$

6. Låt (X, Y) vara två slumpvariabler där X är Poissonfördelad med parameter λ där $0 < \lambda < 1$, och givet att $X = k$ så är Y Binomialfördelad med parametrar k och λ.

(a) Bestäm den gemensamma sannolikhetsfunktionen för (X, Y) och marginalsannolikhetsfunktionen för Y. (3p)

(b) Beräkna $\mathbb{E}[Y]$. (2p)

(c) Beräkna $\mathbb{E}[Y^2]$ (2p)

Lösning:

(a) Vi har att

$$\mathbb{P}(X = k, Y = l) = \mathbb{P}(Y = l | X = k)\mathbb{P}(X = k) = \binom{k}{l} \lambda^l (1-\lambda)^{k-l} \frac{\lambda^k}{k!} e^{-\lambda}$$

där det krävs att $k \in \{0, 1, \ldots\}$ och att $l \in \{0, \ldots, k\}$.

Vidare har vi att

$$\begin{aligned}\mathbb{P}(Y=l) &= \sum_{k=l}^{\infty}\mathbb{P}(X=k,Y=l) = \sum_{k=l}^{\infty}\binom{k}{l}\lambda^l(1-\lambda)^{k-l}\frac{\lambda^k}{k!}e^{-\lambda} \\ &= \sum_{k=l}^{\infty}\frac{k!}{l!(k-l)!}\lambda^l(1-\lambda)^{k-l}\frac{\lambda^k}{k!}e^{-\lambda} = \frac{\lambda^{2l}e^{-\lambda}}{l!}\sum_{k=l}^{\infty}\frac{(\lambda(1-\lambda))^{k-l}}{(k-l)!} \\ &= \frac{\lambda^{2l}e^{-\lambda}}{l!}\sum_{k=0}^{\infty}\frac{(\lambda(1-\lambda))^{k}}{k!} = \frac{\lambda^{2l}e^{-\lambda}}{l!}e^{\lambda(1-\lambda)} = \frac{\lambda^{2l}e^{-\lambda^2}}{l!}\end{aligned}$$

där $l \in \{0,1,\ldots\}$. Vi ser därför att $Y \sim$Poi(λ^2).

(b) Om man lyckats fullt ut med uppgift (a) räcker det att använda att man kan fördelningen för Y för att se att $\mathbb{E}[Y]=\lambda^2$.

Alternativt använder vi oss av att $\mathbb{E}[Y]=\mathbb{E}\left[\mathbb{E}[Y|X]\right]$ och eftersom $\mathbb{E}[Y|X=k]=\lambda k$ så ser vi att $\mathbb{E}[Y|X]=\lambda X$. Därför så blir

$$\mathbb{E}[Y]=\mathbb{E}\left[\mathbb{E}[Y|X]\right]=\mathbb{E}\left[\lambda X\right]=\lambda\mathbb{E}\left[X\right]=\lambda^2$$

där vi använder oss av att X är Poissonfördelad med parameter λ så att $\mathbb{E}\left[X\right]=\lambda$.

(c) Återigen kan vi använda (a) för att se att

$$\mathbb{E}[Y^2]=\mathrm{Var}(Y)+\mathbb{E}[Y]^2=\lambda^2+\lambda^4,$$

där vi använder att en Poisson-fördelad slumpvariabel med parameter γ har varians γ.

Alternativt börjar vi med att dra oss till minne att om $Z \sim$Bin(n,p) så får vi att $\mathrm{Var}(Z)=np(1-p)$. Dessutom gäller att $\mathrm{Var}(Z)=\mathbb{E}[Z^2]-\mathbb{E}[Z]^2$ så att (liknande ovan)

$$\mathbb{E}[Z^2]=\mathrm{Var}(Z)+\mathbb{E}[Z]^2=np(1-p)+(np)^2.$$

Detta innebär att $\mathbb{E}[Y^2|X=k]=k\lambda(1-\lambda)+(k\lambda)^2$ så att

$$\mathbb{E}[Y^2|X]=X\lambda(1-\lambda)+(X\lambda)^2.$$

Vi ser därför att

$$\mathbb{E}[Y^2]=\mathbb{E}\left[\mathbb{E}[Y^2|X]\right]=\mathbb{E}\left[X\lambda(1-\lambda)+(X\lambda)^2\right]=\lambda(1-\lambda)\mathbb{E}[X]+\lambda^2\mathbb{E}[X^2].$$

Vi vet att $\mathbb{E}[X]=\lambda$ och som ovan ser vi att

$$\mathbb{E}[X^2]=\mathrm{Var}(X)+\mathbb{E}[X]^2=\lambda+\lambda^2.$$

Slutligen får vi då att

$$\mathbb{E}[Y^2]=\mathbb{E}\left[\mathbb{E}[Y^2|X]\right]=\lambda(1-\lambda)\mathbb{E}[X]+\lambda^2\mathbb{E}[X^2]=\lambda^2(1-\lambda)+\lambda^2(\lambda+\lambda^2)=\lambda^2+\lambda^4.$$

7. Företaget Rörtillverkarna tillverkar rör. För att rören inte skall läcka är det viktigt att måtten är exakta. I tillverkningsprocessen finns det dock små osäkerheter vilket gör att rörens dimensioner inte blir precis så som man vill ha dem.

   Rörtillverkarna skall testa sin nya rörtillverkningsmaskin och mäter därför skillnaden mellan den faktiska diametern på rören och den föreskrivna diametern. De gör det på 10 rör och får följande tabell över avvikelserna (i μm).

rör nr:	1	2	3	4	5	6	7	8	9	10
avvikelse:	96	-19	-8	18	16	47	41	55	-21	74

   Vi kan anta att avvikelserna är oberoende och normalfördelade med parametrar μ, σ^2.

   (a) Hitta ett 95% K.I. för μ om σ^2 är okänd. (2p)

   (b) Använd ditt svar i (a) för att testa huruvida det kan finnas några systematiska tillverkningsfel av rörens diameter. (2p)

   (c) Vilket p-värde har ditt test i (b)? (2p)

Lösning:

(a) Då σ^2 är okänd får vi skatta σ^2 med

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2 \approx 1568$$

så att $s \approx 39.6$. Vi har att

$$R = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(9)$$

och vi ser från tabell att $t_{0.025}(9) \approx 2.262$. Vi får därför att

$$0.95 = \mathbb{P}(-2.262 \leq R \leq 2.262) = \mathbb{P}\left(\bar{X} - 2.262\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + 2.262\frac{s}{\sqrt{n}}\right),$$

så att ett 95% numeriskt K.I. ges av

$$I_\mu = \bar{x} \pm 2.262\frac{s}{\sqrt{n}} \approx 29.9 \pm 2.262\frac{39.6}{\sqrt{10}} \approx [1.57, 58.2].$$

(b) Då $\mu = 0$ motsvarar att de inte har några systematiska tillverkningsfel så får vi testa $H_0 : \mu = 0$ mot $H_1 : \mu \neq 0$. Genom att använda oss av att $\mu = 0$ inte tillhör det 95% konfidensintervallet ovan så förkastar vi H_0 på 95% nivån och det verkar som om rörtillverkarna faktiskt har ett systematiskt fel i sin tillverkningsprocess.

(c) p-värdet ges av den minsta signifikansnivån för vilket vi förkastar nollhypotesen givet våra data. Om vi låter $T \sim t(9)$ så ser vi att

$$\text{p-värde} = \mathbb{P}\left(|T| \geq \frac{\bar{x}}{s/\sqrt{10}}\right) \approx 2\mathbb{P}(T \geq 2.39).$$

Då 2.39 inte finns i tabellen så måste vi interpolera mellan de närliggande värdena. Vi får då att

$$\mathbb{P}(T \leq 2.39) \approx 0.975 + \frac{2.39 - 2.26}{2.82 - 2.26}(0.99 - 0.975) \approx 0.9785$$

så att vårt p-värde blir ca $2 * (1 - 0.9785) = 0.043$.

8. Jöns-Harald har ställts inför några olika experimentella situationer som lämpligen modelleras som slumpmässiga skeenden. Jöns-Harald behöver hjälp med att identifiera lämplig slumpmässig modell. I varje deluppgift nedan skall ni ange en relevant slumpvariabel med specificerad fördelning. Ni skall även motivera kort men kärnfullt ert val av slumpvariabel/fördelning. Vilka antaganden gör ni? (OBS: Ni behöver inte härleda sannolikhetsfunktionerna/täthetsfunktionerna, bara motivera era val. Uppgiften går ut på att känna igen situationer.)

(a) Jöns-Harald kastar pil på en tavla 100 gånger. Tavlan har fält numrerade mellan 1 och 20 och Jöns-Harald låter X_1 vara det slumpmässiga antalet gånger som han träffar området markerat 18. Vilken fördelning har X_1? (1p)

(b) Återigen kastar Jöns-Harald pil, men nu låter ha X_2 vara det totala antalet kast som behövs till och med att han träffar fältet markerat 18 för första gången. Vilken fördelning har X_2? (1p)

(c) Denna gång räknar Jöns-Harald det totala antalet kast som behövs till och med att han träffat fältet markerat 18 sammanlagt 5 gånger. Om X_3 är detta antalet, vilken fördelning har då X_3? (1p)

(d) Jöns-Harald räknar antalet sönderfall N från ett radioaktivt prov som sker under en timme. Ange en lämplig fördelning för N. (1p)

(e) Jöns-Harald väntar på att en radioaktiv atomkärna skall sönderfalla. Hur skall han modellera väntetiden T? (1p)

(f) Jöns-Harald försöker mäta reaktionstiden hos ett elektriskt relä. Mätningen av reaktionstiden är behäftat med mätfel. Hur skall Jöns-Harald modellera detta mätfel M_{fel}? (1p)

Lösning:

(a) Om kasten antas vara oberoende och att varje kast träffar nr 18 med sannolikhet p så är det totala antalet träffar X_1 binomialfördelat med parametrar 100 och p. Om man dessutom antar att Jöns-Harald träffar alla fält med samma sannolikhet och aldrig missar tavlan helt så är $p = 1/20$.

(b) Med samma antaganden som i (a) så är X_2 =antalet försök totalt geometriskt fördelat med parameter p. Detta då vi räknar ”till första träffen”.

(c) Detta är en generalisering av situationen i (b) och med samma antaganden som ovan är X_3 =antalet försök totalt negativt binomialfördelad med parametrar $r = 5$ och p.

(d) Situationer som dessa modelleras med hjälp av en Poisson-fördelning. Man kan se det som en approximation till en binomialfördelning med stort n och litet p, men binomialfördelningen är opraktisk här.

(e) Sönderfallstider är exponentialfördelade. Det är svårt att motivera detta på annat sätt än att man empiriskt vet att det stämmer bra. Alternativt kan man säga att om den ej söderfallit innan tiden t så påverkar inte detta sannolikheten att den sönderfaller inom säg en sekund från t, och detta leder till att det måste vara en exponentialfördelning.

(f) Lämpligen använder han normalfördelning här. Anledninge till att detta ofta är det bästa valet är att felen orsakas av en mängd olika källor som var för sig ger oberoende, små bidrag till det totala felet. En sådan summa kan enligt Centrala Gränsvärdessatsen approximeras med en normalfördelning. Om instrumentet är rätt kalibrerat bör då $M_{fel} \sim N(0, \sigma^2)$.

Andra fördelningar kan dock också komma ifråga (situationsberoende).