Matematisk statistik 2022-06-08 – Lösningar

1. (a) Eftersom $0.4 = P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$, så följer det att $P(A \cap B) = 0.12$.

   (b) Sannolikheten p att både A och B inträffar i ETT försök är $p = 0.12$ enligt föregående deluppgift. Vi upprepar 7 gånger och räknar antalet X gånger både A och B inträffar. Det följer (om vi antar oberoende) att $X \sim \text{Bin}(7, p = 0.12)$. Sålunda,

$$P(X \leq 3) = 0.88^7 + 7 \cdot 0.88^6 \cdot 0.12 + \binom{7}{2} 0.88^5 \cdot 0.12^2 + \binom{7}{3} 0.88^4 \cdot 0.12^3 = 0.9946$$

   (c) Med samma p som ovan undersöker vi nu variabeln Y som räknar antalet gånger vi upprepar försöket tills dess att både A och B inträffar för första gången. Alltså, $Y \sim \text{Ffg}(p = 0.12)$. Vi erhåller nu att

$$P(Y \leq 2) = 0.12 + 0.88 \cdot 0.12 = 0.2256$$

   (d) Eftersom $P(A \cap B) = 0.12$ så följer det att $P(B) = 0.4$ är nödvändigt för att

$$P(A \cap B) = P(A)P(B).$$

Svar: (a) 0.12. (b) 0.995. (c) 0.2256. (d) Ja, om $P(B) = 0.4$.

2. Modellen är stickprov i par. Vi betraktar skillnaden som ett enda stickprov:

$z_i = y_i - x_i$	-3.11	-2.31	-3.15	0.17	-0.65

och skattar väntevärdet med

$$\widehat{\delta} = \frac{1}{5} \sum_{i=1}^{5} (y_i - x_i) = -1.81$$

och standardavvikelsen med

$$s = \left(\frac{1}{5-1} \sum_{i=1}^{5} (y_i - x_i - \bar{\Delta})^2 \right)^{1/2} = 1.50.$$

   (a) En bra skattning av Δ är medelvärdet -1.81.

   (b) Låt

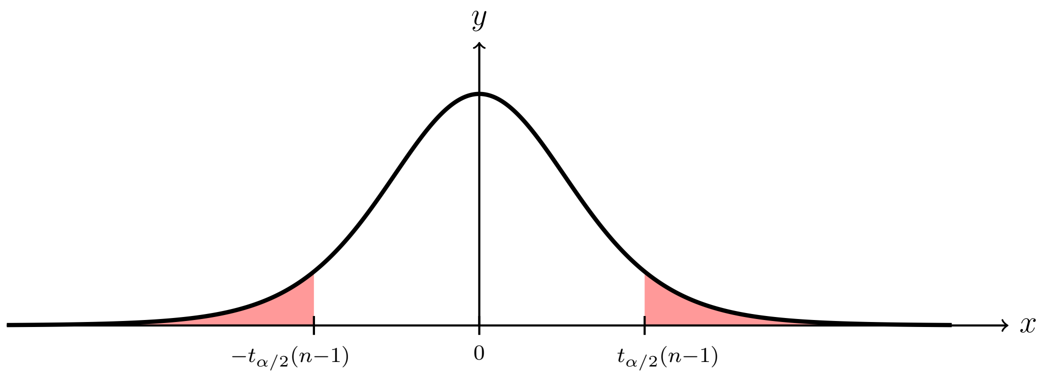
$$\widehat{\Delta} = \frac{1}{5} \sum_{i=1}^{5} (Y_i - X_i).$$

   Vi bildar testvariabeln

$$T = \frac{\widehat{\Delta} - \Delta}{S/\sqrt{n}} \sim t(4)$$

   där vi skattar S med s, så $s/\sqrt{n} = 0.6709$ och

$$P(-t_{\alpha/2}(4) < T < t_{\alpha/2}(4)) = 1 - \alpha.$$



Den markerade arean innehåller 90% av sannolikheten för $t(4)$-fördelningen. Vi löser ut Δ ur intervallet i sannolikhetsmåttet och får att

$$\widehat{\Delta} - 0.6709 \cdot t_{\alpha/2}(4) < \Delta < \widehat{\Delta} + 0.6709 \cdot t_{\alpha/2}(4).$$

Vi ersätter $\widehat{\Delta}$ med den observerade punktskattningen $\widehat{\delta} = 1.81$ och erhåller ett konfidensintervall av konfidensgrad 90%:

$$I_\Delta = [-3.24, -0.38],$$

där vi använt $\alpha = 0.10$ och $t_{0.05}(4) = 2.132$.

(c) Vi har

$$T = \frac{\overline{\Delta} - \delta}{S/\sqrt{5}} \sim t(4).$$

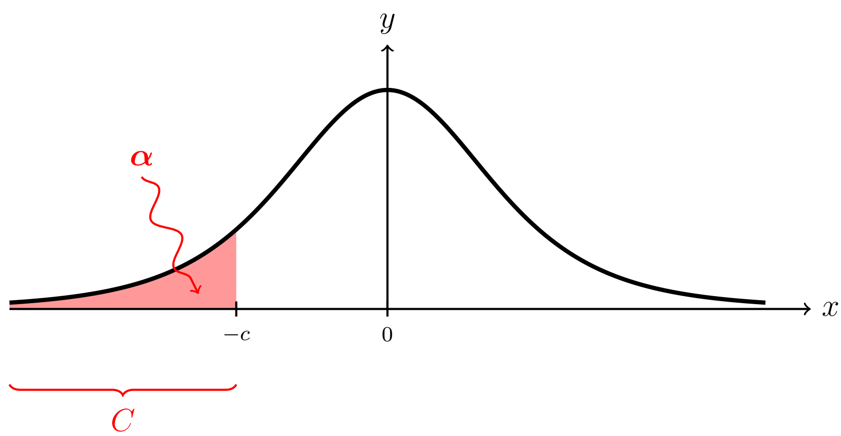
Då gäller att

$$P(T > -c) = 0.90$$

om $c = 1.53$ (ur tabell). Vi låter $C =]-\infty, -1.53[$ vara det kritiska området. Med våra observationer finner vi att

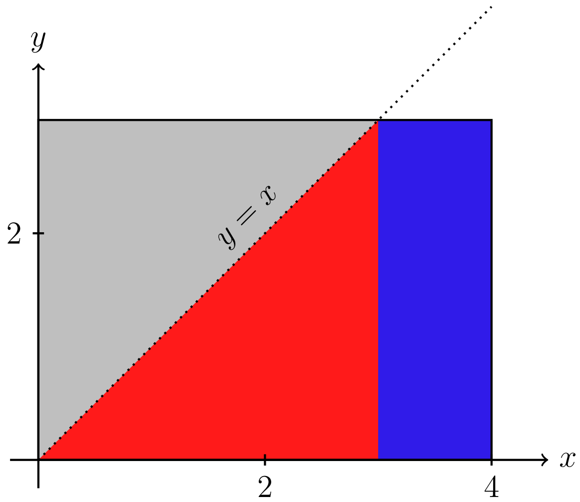
$$t = \frac{-1.81 - 0}{0.6709} = -2.70 \in C,$$

så vi kan förkasta H_0 och hävda att H_1 är mer trolig.



Svar: (a) T.ex. -1.81 (eller $e^{\sqrt{\pi}}$?). (b) $[-3.24, -0.38]$; (c) Hypotesen kan förkastas på 10%-nivån.

3. Vi börjar med att skissa området.



(a) Eftersom arean av rektangeln är 12 måste $c = 1/12$ för att dubbelintegralen av $f_{X,Y}$ över området skall bli ett.

(b) Sannolikheterna kan enkelt beräknas med kvoten av arean av det sökta området och rektangeln. Området då $X > 3$ ges av den blå rektangeln med arean $1 \cdot 3 = 3$, så

$$P(X > 3) = \frac{3}{12} = \frac{1}{4}.$$

Det finns inga punkter i rektangeln där $Y > 3$, så

$$P(Y > 3) = 0.$$

Området då $X > Y$ ges av den blå rektangeln tillsammans med den röda triangeln, så

$$P(X > Y) = \frac{1 \cdot 3 + 3 \cdot 3/2}{12} = \frac{15}{24} = \frac{5}{8}.$$

Alternativt kan komplementet betraktas och vi tar 1 minus kvoten som ligger i den grå triangeln:

$$P(X > Y) = 1 - \frac{3 \cdot 3/2}{12} = \frac{5}{8}.$$

Svar: (a) $c = 12$ (b) $P(X > 3) = 1/4$, $P(Y > 3) = 0$ och $P(X > Y) = 5/8$.

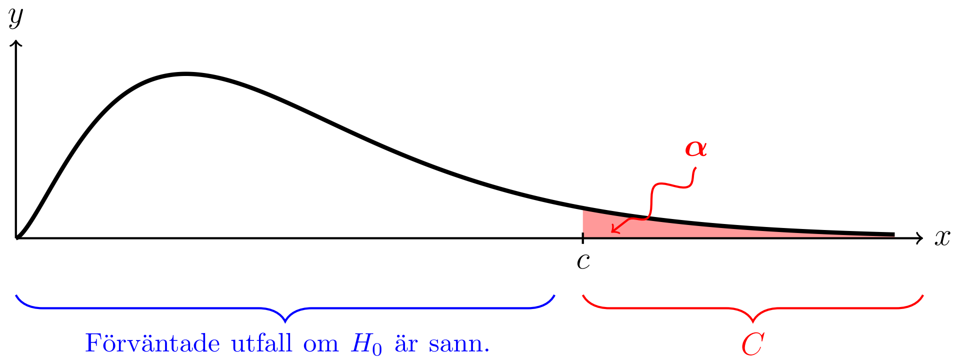
4. Låt H_0 vara hypotesen att datan är homogen mellan de två sidorna (dvs att de tycker likadant) och H_1 att detta inte är sant. Totalt sett har vi $n = 289$ observationer. Vi ser direkt att de sista fyra skivorna har för liten $n_i\widehat{p_j}$ (betydligt mindre än 5), så vi kombinerar dessa för att få ett användbart test. Detta ändrar det faktiska hypotestestet men det kan inte hjälpas. Med datan given kan vi utföra följande kalkyler.

Album Title	Web page		Sum	$\widehat{p_j}$
	Nuclear War Now!	Metalstorm.net		
Altars of Madness	67	82	149	0.516
Blessed Are The Sick	18	34	52	0.180
Covenant	11	32	43	0.149
D–H	8	37	45	0.156
n_i	104	185	289	

Den vanliga teststorheten finner vi nu enligt

$$\begin{aligned} q = \sum_{i=0}^{1}\sum_{j=0}^{3}\frac{(N_{ij}-n_i\widehat{p}_j)^2}{n_i\widehat{p}_j} &= \frac{(67-53.62)^2}{53.62}+\frac{(82-95.38)^2}{95.38}+\frac{(18-18.72)^2}{18.72} \\ &+\frac{(34-33.29)^2}{33.29}+\frac{(11-15.47)^2}{15.47}+\frac{(32-27.53)^2}{27.53} \\ &+\frac{(8-16.19)^2}{16.19}+\frac{(37-28.81)^2}{28.81} \\ &= 13.76. \end{aligned}$$

Om H_0 är sann så är q en observation av $Q \overset{\text{appr.}}{\sim} \chi^2((2-1)(4-1)) = \chi^2(3)$. Vi förkastar H_0 om q är stor så vi väljer det kritiska området på formen $C = [c, \infty)$. Från tabell finner vi att $c = \chi^2_{0.01}(3) = 11.34$. Om $q \geq c$ så förkastar vi H_0 (och anser H_1 styrkt).



Eftersom $q \in C$ så förkastar vi H_0. Det är antagligen en skillnad i hur åsikterna ser ut på de olika sidorna (vi förkastar åtminstone hypotesen att de tycker likadant).

Answer: Förkasta H_0.

5. (a) Vi behöver sannolikheterna att hamna i de olika fallen:

$$\begin{aligned} P(X=-1) &= P(Y<0) = \Phi((0-4)/2) = 1-\Phi(2) = 0.0228 \\ P(X=0) &= P(0 \leq Y < 3) = \Phi((3-4)/2) - \Phi((0-4)/2) = 0.2858 \\ P(X=1) &= P(3 \leq Y < 5) = \Phi((5-4)/2) - \Phi((3-4)/2) = 0.3829 \\ P(X=2) &= P(5 \leq Y < 6) = \Phi((6-4)/2) - \Phi((5-4)/2) = 0.1499 \\ P(X=3) &= P(Y \geq 6) = 1-\Phi((6-4)/2) = 0.1586 \end{aligned}$$

Kontrollera att sannolikheterna summerar till ett! Eftersom $p_X(k) = P(X=k)$ är vi klara med denna deluppgift.

(b) Vi använder definitionen och finner att

$$E(X) = \sum_{k=-1}^{3} k p_X(k) = 1.1357.$$

Vidare kan vi räkna ut att

$$E(X^2) = \sum_{k=-1}^{3} k^2 p_X(k) = 2.4327.$$

Steiners sats implicerar nu att

$$V(X) = E(X^2) - E(X)^2 = 2.4327 - 1.2898 = 1.1429$$

Svar: (a) Se ovan. (b) $E(X) = 1.1357$ och $V(X) = 1.1429$.

6. (a) Täthetsfunktionen ges av $f(x) = \theta e^{-\theta x}$, $x \geq 0$, så

$$L(\theta) = \prod_{k=1}^{n} \theta e^{-\theta x_k} = \theta^n \exp\left(-\theta \sum_{k=1}^{n} x_k\right) \quad \Rightarrow \quad l(\theta) = \ln L(\theta) = n \ln \theta - \theta \sum_{k=1}^{n} x_k.$$

Vi undersöker var det finns extrempunkter och finner att

$$0 = l'(\theta) = \frac{n}{\theta} - \sum_{k=1}^{n} x_k \quad \Leftrightarrow \quad \theta = \frac{n}{\sum_{k=1}^{n} x_k} = \frac{1}{\overline{x}},$$

under förutsättning att $\overline{x} \neq 0$. Är detta ett maximum? Använd det ni lärt er i envariabelanalysen! Till exempel ser vi att

$$l''(\theta) = -\frac{n}{\theta^2},$$

så $l''(\theta) < 0$ för alla $\theta > 0$. Således är det ett maximum vi funnit.

(b) Vi ersätter $\bar{x}$ med den stokastiska variabeln $\bar{X}$, där $\{x_j\}$ är ett stickprov av variabler X_j som är fördelade enligt uppgiften. Vi får

$$E\left(\frac{1}{\widehat{\Theta}}\right) = E\left(\bar{X}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \int_0^\infty x\theta\, e^{-\theta x}\, dx = \ldots = \frac{1}{\theta},$$

vilket visar att vi har en väntevärdesriktig skattning av $1/\theta$.

Svar: a) ML-skattningen ges av $\widehat{\theta} = 1/\bar{x}$. b) Se ovan.