

Matematisk statistik 2022-xx-yy – Lösningar

1. (a) Eftersom $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, följer det att

$$P(A) + P(B) = 0.6 + 0.3 = 0.9.$$

- (b) Om det är givet att $P(A) = 0.4$ söker vi $P(A|B^*)$. Eftersom $P(A) + P(B) = 0.9$ så måste $P(B) = 0.5$. Vi vet också enligt definitionen att

$$P(A|B^*) = \frac{P(A \cap B^*)}{P(B^*)} = \frac{P(A) - P(A \cap B)}{1 - P(B)} = \frac{0.4 - 0.3}{1 - 0.5} = 0.2.$$

- (c) Låt $P(A|B^*) = P(A)$. Eftersom $P(B^*) \neq 0$ gäller att

$$\begin{aligned} \frac{P(A \cap B^*)}{P(B^*)} = P(A) &\Leftrightarrow P(A \cap B^*) = P(B^*)P(A) \\ &\Leftrightarrow P(A) - P(A \cap B) = P(A)(1 - P(B)) \\ &\Leftrightarrow P(A \cap B) = P(A)P(B). \end{aligned}$$

Således är A och B oberoende.

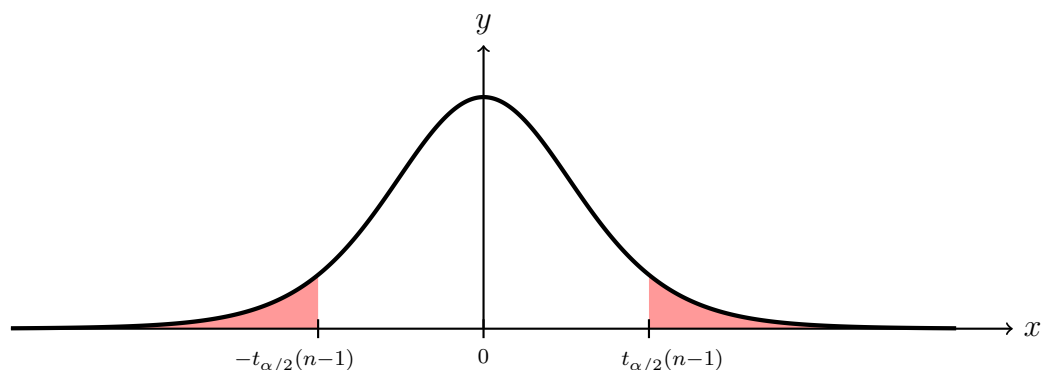
Svar: (a) 0.9. (b) 0.2. (c) Oberoende. Se ovan.

2. (a) Vi har $n = 16$ och bildar testvariabeln

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(15)$$

och ser att

$$P(-t_{\alpha/2}(15) \leq T \leq t_{\alpha/2}(15)) = 1 - \alpha.$$



Den markerade arean innehåller 99% av sannolikheten för $t(15)$ -fördelningen. Vi löser ut μ ur intervallet i sannolikhetsmåttet och erhåller

$$\bar{X} - \frac{S}{\sqrt{n}}t_{\alpha/2}(15) \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}}t_{\alpha/2}(15).$$

Vi ersätter $\bar{X}$ med den observerade punktskattningen $\bar{x} = 21$ och S med $s = \sqrt{7.2}$. Då erhåller vi ett konfidensintervall med konfidensgrad 99%:

$$I_\mu = [19.02, 22.98],$$

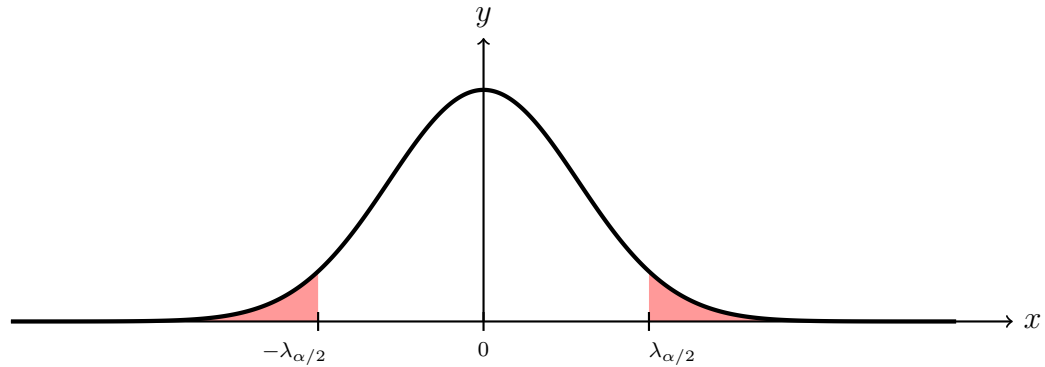
där vi använt $\alpha = 0.01$ och $t_{0.005}(15) = 2.947$.

(b) Vi använder oss av testvariabeln

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Vi vet att $\sigma = \sqrt{7.0}$ och det följer att

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



På samma sätt som ovan löser vi ut μ ur intervallet i sannolikhetsmåttet och får att

$$\bar{X} - \lambda_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{X} + \lambda_{\alpha/2}\sigma/\sqrt{n}.$$

Vi ersätter $\bar{X}$ med den observerade punktskattningen $\bar{x} = 21$ och $\sigma = \sqrt{7.0}$. Då erhåller vi ett konfidensintervall av konfidensgrad 99%:

$$I_\mu = [19.30, 22.70],$$

där vi använt $\alpha = 0.01$ och $\lambda_{0.005} = 2.575$.

(c) Vi har

$$T = \frac{\bar{X} - \mu}{S/\sqrt{16}} \sim t(15).$$

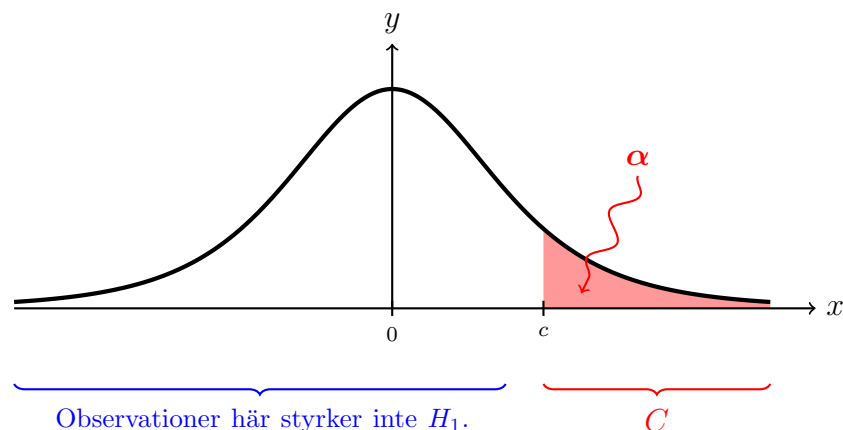
Då gäller att

$$P(T < c) = 0.99$$

om $c = 2.60$ (ur tabell). Vi låter $C = [2.60, \infty[$ vara det kritiska området. Med våra observationer finner vi att

$$t = \frac{21 - 20}{\sqrt{7.2}/\sqrt{16}} = 1.49 \notin C,$$

så vi kan inte förkasta H_0 . Det kan vara så att $\mu = 20$.



Svar: (a) $I_\mu = [19.02, 22.98]$. (b) $I_\mu = [19.30, 22.70]$. (c) Vi kan inte förkasta H_0 .

3. (a) Eftersom X och Y är oberoende så kommer sannolikhetsfunktionen att ges av produkten av de marginella sannolikhetsfunktionerna: $p_{X,Y}(j, k) = p_X(j)p_Y(k)$. Vi kan direkt läsa ut att

$$p_X(0) = \sum_{k=0}^2 p_{X,Y}(0, k) = 0.06 + 0.12 + 0.12 = 0.30.$$

Från detta följer att $p_{X,Y}(0, k) = p_X(0)p_Y(k) = 0.30 \cdot p_Y(k)$, och om vi utnyttjar informationen given i tabellen kan vi nu lösa ut att

$$\begin{aligned} p_Y(0) &= 0.06/0.30 = 0.2, \\ p_Y(1) &= p_Y(2) = 0.12/0.30 = 0.40. \end{aligned}$$

Vi kan även lösa ut att $p_X(1) = 0.08/0.40 = 0.20$, vilket vi sedan kan använda för att räkna ut att $p_{X,Y}(2, 0) = 0.30 \cdot 0.20 = 0.06$.

Med samma teknik som ovan kan vi räkna ut att $p_Y(1) = p_Y(2) = 0.40$, och därefter resterande fält i tabellen. Vi erhåller sålunda

j / k	$k = 0$	$k = 1$	$k = 2$	$p_X(j)$
$j = 0$	0.06	0.12	0.12	0.30
$j = 1$	0.08	0.16	0.16	0.40
$j = 2$	0.06	0.12	0.12	0.30
$p_Y(k)$	0.20	0.40	0.40	1.00

- (b) Vi söker sannolikheten att $X + Y < 2$. Om $X = j$ och $Y = k$ händer detta endast vid $(j, k) = (0, 0)$, $(0, 1)$ samt $(1, 0)$ med nollskild sannolikhet. Vi får alltså

$$P(X + Y < 2) = 0.06 + 0.12 + 0.08 = 0.26.$$

Svar: (a) Se tabell ovan. (b) 0.26.

4. (a) Det är klart att $f_X(x) \geq 0$ för alla x , och vi ser att

$$\int_{-\infty}^{\infty} f_X(x) dx = \frac{3}{4} \int_0^2 (2x - x^2) dx = \frac{3}{4} \left[x^2 - \frac{x^3}{3} \right]_0^2 = \frac{3}{4} \cdot \frac{4}{3} = 1,$$

så f_X är en täthetsfunktion. För att beräkna väntevärdet använder vi definitionen:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \frac{3}{4} \int_0^2 (2x^2 - x^3) dx = \frac{3}{4} \left[\frac{2x^3}{3} - \frac{x^4}{4} \right]_0^2 = 1.$$

Eftersom $V(X) = E(X^2) - E(X)^2$, och

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \frac{3}{4} \int_0^2 (2x^3 - x^4) dx = \frac{3}{4} \left[\frac{x^4}{2} - \frac{x^5}{5} \right]_0^2 = \frac{6}{5},$$

erhåller vi $V(X) = 6/5 - 1^2 = 1/5$.

- (b) Låt X_i , $i = 1, 2, \dots, 100$, vara oberoende s.v. med samma täthetsfunktion $f_X(x)$ som i uppgift (a). Vi definierar $Y = X_1 + X_2 + \dots + X_{100}$. Enligt CGS kommer Y att vara approximativt normalfördelad med väntevärde och varians enligt

$$\begin{aligned} E(Y) &= 100 \cdot E(X) = 100, \\ V(Y) &= 100 \cdot V(X) = 100/5. \end{aligned}$$

Vi beräknar approximativt

$$P(Y < 105) \approx \Phi\left(\frac{105 - 100}{\sqrt{100/5}}\right) = \Phi\left(\frac{5\sqrt{5}}{10}\right) = \Phi(\sqrt{5}/2) \approx \Phi(1.12) = 0.87.$$

Att det är ok att använda CGS följer av att vi summerar så många variabler, och att fördelningen för varje X_i redan är ganska symmetrisk (skissa $f_X(x)$).

Svar: (a) $E(X) = 1$ och $V(X) = 1/5$. (b) ≈ 0.87 .

5. (a) Låt X vara antalet blindgångare i Kalles test. Då blir X binomialfördelad enligt $X \sim \text{Bin}(n, p)$ med $n = 1000$ och $p = 0.01$ (vi antar att tillverkaren talat sanning). Vi antar även att varje skott Kalle tar är oberoende av övriga (kanske inte helt sant då det finns gemensamma faktorer så som Kalles vapen och Kalle själv). Enklast nu är att Poissonapproximera X : $X \stackrel{\text{appr.}}{\sim} \text{Po}(10)$, vilket är OK då $p \leq 0.1$ och $n \geq 10$. Vi finner sedan att

$$P(X = 10) = P(X \leq 10) - P(X \leq 9) \approx 0.5830 - 0.4579 = 0.1251$$

och att

$$P(X < 20) = P(X \leq 19) \approx 0.9965$$

ur tabell för $\text{Po}(10)$.

Anmärkning: Redan här ser vi att $p = 0.01$ inte kommer att förefalla rimligt med resultatet av Kalles försök i deluppgift (b). Om p verkligen har det värdet är resultatet $X = 22$ antagligen orimligt.

- (b) Vi söker konfidensintervall för felrisken vid Kalles test. Kalle har testat $n = 1000$ stycken kulor, och fått $X = 22$ stycken blindgångare. En naturlig skattning på den verkliga andelen ges av

$$\hat{P} = \frac{X}{n},$$

och vårt observerade värde är $\hat{p} = 22/1000 = 0.022$. Vi vet att variabeln X är binomialfördelad: $X \sim \text{Bin}(n, p)$. Vidare ser vi att

$$np(1 - p) \approx 1000 \cdot 0.022 \cdot (1 - 0.022) = 21.516,$$

så vi borde kunna göra en normalapproximation: $X \stackrel{\text{appr.}}{\sim} N(np, D)$ där $D = \sqrt{np(1 - p)}$. Alltså blir $\hat{P} \stackrel{\text{appr.}}{\sim} N(p, d)$, där

$$d = \sqrt{\hat{p}(1 - \hat{p})/n} = \sqrt{21.516/1000^2} = 0.00464.$$

Vår testvariabel blir alltså

$$Z = \frac{\hat{P} - p}{d} \stackrel{\text{appr.}}{\sim} N(0, 1).$$

Vi stänger in Z :

$$P(-\lambda_{\alpha/2} \leq Z \leq \lambda_{\alpha/2}) \approx 1 - \alpha,$$

där λ_α är normalfördelningens α -kvantil. Vi har $\alpha = 0.05$ och $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$. Vi löser ut p ur olikheten inuti sannolikhetsmättet, och erhåller då

$$\hat{P} - \lambda_{0.025}d \leq p \leq \hat{P} + \lambda_{0.025}d.$$

Detta ger oss konfidensintervallet $I_p = [0.017, 0.027]$ om vi ersätter $\hat{P}$ med det observerade värdet $\hat{p} = 0.022$. Vi ser att $p = 0.01 \notin I_p$, så tillverkarens påstående verkar inte stämma.

Svar: (a) $P(X = 10) \approx 0.13$ och $P(X < 20) \approx 0.997$.

(b) $I_p = [0.017, 0.027]$. Det verkar inte som företaget varit helt ärliga.

6. Låt $x_1, x_2, \dots, x_n$ vara ett stickprov av n variabler med gammafördelningen given i uppgiften.

(a) Låt oss bilda likelihoodfunktionen $L(\theta)$:

$$L(\theta) = \prod_{i=1}^n \frac{\theta(\theta x_i)^{v-1} e^{-\theta x_i}}{\Gamma(v)}.$$

Vi logaritmerar och deriverar sedan:

$$\begin{aligned} \ln L(\theta) &= \sum_{i=1}^n (v \ln \theta + (v-1) \ln x_i - \theta x_i - \ln \Gamma(v)), \\ \frac{d \ln L(\theta)}{d\theta} &= \frac{n}{\theta} + \frac{n(v-1)}{\theta} - \sum_{i=1}^n x_i. \end{aligned}$$

Om vi sätter derivatan till noll och löser ut θ erhåller vi

$$\frac{n + n(v-1)}{\theta} - n\bar{x} = 0 \quad \Leftrightarrow \quad \theta = \frac{v}{\bar{x}},$$

där $\bar{x}$ är medelvärdet av våra x_j . Teckenväxlingen för derivatan vid punkten $\theta^* = v/\bar{x}$ är $+0-$, så det rör sig om en maxpunkt. Vi har alltså funnit ML-skattningen.

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

$$E\left(\frac{1}{\hat{\theta}}\right) = E\left(\frac{\bar{X}}{v}\right) = \frac{1}{nv} \sum_{i=1}^n E(X_i) = \frac{1}{nv} \frac{nv}{\theta} = \frac{1}{\theta},$$

där vi utnyttjat tipset som påstod att väntevärdet av en gammafördelad variabel är v/θ (försök gärna visa detta).

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