Estadística: Problemes resolts

Mercè Claverol Aguas

Universitat Politècnica de Catalunya
Dept. Matemàtiques

Índex

1 Probabilitat

1. Els successos A i B que compleixen $P(A) = 33/39$, $P(B) = 32/39$ i $P(A \cap B) = 31/39$. Calculeu $P(A \cup B)$.

   Solució

   $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \dfrac{33 + 32 - 31}{39} = \dfrac{34}{39}$

2. Raoneu si és certa o falsa l'afirmació:

   Si $P(A) = 0.5$, $P(B) = 0.2$ i $P(A \cup B) = 0.45$, aleshores A i B són independents.

   Solució

   $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.2 - P(A \cap B) = 0.45 \implies P(A \cap B) = 0.5 + 0.2 - 0.45 = 0.25$

   $$P(A \cap B) = 0.25 \neq P(A)P(B) = (0.5)(0.2) = 0.1$$

   Per tant, A i B no són independents (l'afirmació era falsa).

3. Raoneu si és certa o falsa l'afirmació:

   Si $P(A) = 0.7$, $P(B) = 0.6$ i $P(\overline{A} \cup \overline{B}) = 0.58$, aleshores A i B són independents.

   Solució

   $P(\overline{A} \cup \overline{B}) = P(\overline{A \cap B}) = 1 - P(A \cap B) = 0.45 \implies P(A \cap B) = 1 - 0.58 = 0.42$

   $$P(A \cap B) = 0.52 \neq P(A)P(B) = (0.7)(0.6) = 0.42$$

   Per tant, A i B són independents (l'afirmació era certa).

4. Si A i B són independents, demostreu que A i $\overline{B}$ també ho són.

   Solució

   Sabem que A i B són independents i, per tant, $P(B/A) = P(B)$.

   $$P(A \cap \overline{B}) = P(A)\,P(\overline{B}/A) = P(A)(1 - P(B/A)) \overset{\downarrow}{=} P(A)(1 - P(B)) = P(A)P(\overline{B})$$

   Hem vist $P(A \cap \overline{B}) = P(A)P(\overline{B})$, és a dir, A i $\overline{B}$ són independents.

5. Sabent que A i B són dos successos independents tals que $P(A) = \dfrac{3}{11}$ i $P(B) = \dfrac{3}{10}$. Trobeu el valor de $P(A/B)$, $P(A \cap B)$ i $P(\overline{A}/B)$.

   Solució

   $P(A/B) = P(A) = \dfrac{3}{11}$

   $P(A \cap B) = P(A)P(B) = \dfrac{9}{110}$

   $P(\overline{A}/B) = 1 - P(A/B) = \dfrac{8}{11}$

6. Els successos A i B compleixen $P(B) = \dfrac{1}{13}$ i $P(A \cap B) = \dfrac{1}{22}$. Calculeu $P(A/B)$.

   Solució

   $P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{1/22}{1/13} = \dfrac{13}{22}$

7. Calculeu la probabilitat de C sabent que: $\Omega = A \cup \overline{A}$, $P(C \cap A) = 0.4$ i $P(C \cap \overline{A}) = 0.3$

Solució

Apliquem el teorema de probabilitats totals: $P(C) = P(C \cap A) + P(C \cap \overline{A}) = 0.4 + 0.3 = 0.7$

8. Calculeu la probabilitat de C sabent que:

   (a) $\Omega = A \cup \overline{A}$, $P(A) = 0.4$, $P(C/A) = 0.384$, $P(C/\overline{A}) = 0.243$

   (b) $\Omega = A_1 \cup A_2 \cup A_3$, els successos $A_i, i = 1, 2, 3$ són incompatibles dos a dos, $P(A_1) = 0.4$, $P(A_2) = 0.5$, i les probabilitats condicionades $P(C/A_1) = 0.3$, $P(C/A_2) = 0.2$, $P(C/A_3) = 0.7$.

Solució

(a) Apliquem el teorema de probabilitats totals:
$P(C) = P(C\cap A)+P(C\cap\overline{A}) = P(C/A)P(A)+P(C/\overline{A})P(\overline{A}) = (0.384)(0.4)+(0.243)(1-0.4) = 0.2994$

(b) Donat que $\Omega = A_1 \cup A_2 \cup A_3$ i els successos $A_i, i = 1, 2, 3$ són incompatibles dos a dos, aleshores $P(A_3) = 1 - (P(A_1) + P(A_2)) = 1 - (0.4 + 0.5) = 0.1$,
Apliquem ara el teorema de probabilitats totals:
$P(C) = P(C \cap A_1) + P(C \cap A_2) + P(C \cap A_3) = P(C/A_1)P(A_1) + P(C/A_2)P(A_2) + P(C/A_3)P(A_3) = (0.3)(0.4) + (0.2)(0.5) + (0.7)(0.1) = 0.29$

9. Sabem que $\Omega = A \cup B$, $P(A) = 0.4$, $P(B) = 0.6$, $P(C/A) = 0.384$ i $P(C/B) = 0.243$. Quant val la probabilitat $P(A/C)$?

Solució

Apliquem el teorema de Bayes,

$$P(A/C) = \frac{P(C \cap A)}{P(C)} = \frac{P(C/A)P(A)}{P(C/A)P(A) + P(C/B)P(B)} = \frac{(0.384)(0.4)}{(0.384)(0.4) + (0.243)(0.6)} = 0.513$$

10. Una empresa té dues màquines. La primera produeix el 40% de les peces i l'altra la resta. Sigui A l'esdeveniment "la peça es produïda per la primera màquina", i B "la peça es produïda per la segona màquina". Sigui D "la peça produïda és defectuosa". La probabilitat que una peça tingui defectes és de 0.2 si és de la primera màquina ($P(D/A) = 0.2$), i de 0.1 si és de la segona ($P(D/B) = 0.1$).

    (a) Quina és la probabilitat que una peça surti defectuosa?

    (b) Si una peça és defectuosa, quina és la probabilitat que hagi estat produïda per la primera màquina?

Solució

(a) La primera màquina produeix el 40% de les peces i la segona la resta, per tant, $P(A) = 0.4$ i $P(B) = 0.6$.
Aplicant probabilitats totals:
$P(D) = P(D \cap A) + P(D \cap B) = P(D/A)P(A) + P(D/B)P(B) = (0.2)(0.4) + (0.1)(0.6) = 0.14$

(b) Ens demanen $P(A/D)$. Apliquem el teorema de Bayes per calcular-ho:
$$P(A/D) = \frac{P(A \cap D)}{P(D)} = \frac{P(D/A)P(A)}{0.14} = \frac{(0.2)(0.4)}{0.14} = 0.5714$$

11. Només 1 de cada 1000 adults es veu afectat per una malaltia de la qual s'ha desenvolupat una prova diagnosi.
Si un individu té la malaltia, la prova dóna positiu el 99% de les vegades, i si no la té el 2%.
Apliquem la prova a un individu a l'atzar i dóna positiu, quina és la probabilitat que l'individu tingui la malaltia?

Solució

Apliquem el teorema de Bayes,

$$P(M/+) = \frac{P(+ \cap M)}{P(+)} = \frac{P(+/M)P(M)}{P(+/M)P(M) + P(+/\overline{M})P(\overline{M})} = \frac{(0.99)(0.001)}{(0.99)(0.001) + (0.02)(0.999)} = 0.0472$$

2 Combinatòria

1. Tenim n peces de les quals sabem que k són defectuoses.

   (a) De quantes maneres es poden escollir k peces de les n que hi ha? I si ho fem ordenadament?

   (b) Quina és la probabilitat que si agafem k **sense reemplaçament**, les k siguin defectuoses?

   Solució

   (a) Fer subconjunts de k del total de n peces és: $\binom{n}{k} = \dfrac{(n)!}{(n-k)!\,k!}$.

   Si les escollim ordenadament: $n(n-1)\cdots(n-k+1) = \dfrac{n!}{(n-k)!}$

   (b) Si no hi ha reemplaçament, no són successos independents.

$$P(d_1 \cap \cdots \cap d_k) = P(d_1)P(d_2 \,/\, d_1)P(d_3 \,/\, d_1 \cap d_2) \cdots P(d_k \,/\, d_1 \cap \cdots \cap d_{k-1}) =$$

$$= \frac{k}{n}\frac{k-1}{n-1} \cdots \frac{1}{n-k+1} = \frac{k!\,(n-k)!}{n!}$$

   També es pot pensar així: només un grup de k peces està format per les k defectuoses, per tant,

$$\frac{1}{\binom{n}{k}} = \frac{(n-k)!\,k!}{n!}$$

2. Volem formar paraules de longitud 10 en codi binari, és a dir amb zeros i uns.

   (a) Quantes es poden fer?

   (b) Quantes amb exactament 4 uns?
   (Indicació. Considereu el conjunt de 10 elements representant les posicions dels zeros i uns en la paraula i formeu subconjunts de 4 elements).

   (c) Quantes amb exactament 6 zeros?
   (Indicació. Observeu que es demana el mateix que l'apartat anterior però ara aplica permutacions amb repeticions fixades (4 uns i 6 zeros) per obtenir el mateix resultat.

   Solució

   (a) 2^{10}

   (b) $\binom{10}{4} = 210$

   (c) $\dfrac{10!}{6!\,4!} = 210$

3. Si llancem un dau 4 cops i anotem ordenadament la puntuació, quina és la probabilitat que:

   (a) El primer cop surti un sis i els altres diferent de sis.

   (b) Un cop surti un sis i els altres diferent de sis.

   (c) Surtin dos sisos i les altres diferent de sis.

   (d) Almenys una tirada surti diferent de 6.

   Solució

   (a) $\dfrac{1}{6}\left(\dfrac{5}{6}\right)^3$

(b) $\binom{4}{1}\frac{1}{6}\left(\frac{5}{6}\right)^3$ Observeu que $\binom{4}{1}$ compta de quantes maneres $\neq$ podem escollir en quina tirada sortirà el 6:

$$\frac{1}{6}\left(\frac{5}{6}\right)^3+\frac{5}{6}\frac{1}{6}\left(\frac{5}{6}\right)^2+\left(\frac{5}{6}\right)^2\frac{1}{6}\frac{5}{6}+\left(\frac{5}{6}\right)^3\frac{1}{6}$$

(c) $\binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2$ Observeu que $\binom{4}{2}$ compta de quantes maneres diferents podem escullir en quines dues tirades sortirà el 6.

(d) $1-\left(\frac{1}{6}\right)^4$ Hem calculat el complementari d'obtenir un 6 en totes les tirades ja que equival a obtenir almenys una diferent de sis: $P(\overline{S_1\cap S_2\cap\cdots\cap S_4})=P(\overline{S_1}\cup\overline{S_2}\cup\cdots\cup\overline{S_4})$.

4. Llancem una moneda n vegades i anotem, ordenadament, si surt cara o creu.

(a) Quin és l'espai mostral i el seu cardinal si $n=4$?

(b) Quina és la probabilitat que surtin k cares?

Solució

(a) $\Omega=\{cccc, cccx,\cdots,xxxx\}$, $|\Omega|=2^4$ (permutacions amb repetició).

(b) Aplicant la regla de Laplace, $\frac{\binom{n}{k}}{2^n}=\binom{n}{k}\left(\frac{1}{2}\right)^n$

Observeu que és el mateix que si ho comptem així: $\binom{n}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k}$

on $\binom{n}{k}$ compta de quantes maneres diferents podem escullir en quines k tirades sortirà cara, $\left(\frac{1}{2}\right)^k$ la probabilitat de sortir k cares i $\left(\frac{1}{2}\right)^{n-k}$ la de sortir $n-k$ creus.

5. Si la probabilitat de fer canasta és de 0.1 i els resultats en diferents llançaments són independents, calculeu la probabilitat d'encertar al quart intent.

Solució

$(1-0.1)^3(0.1)=0.0729$

6. En una urna amb 100=20+80 boles (20 negres i 80 blanques), s'extreu una mostra aleatòria de 4 boles **sense reemplaçament**.

(a) Quina és la probabilitat que cap d'elles sigui negra?

(b) Quina és la probabilitat que exactament una d'elles sigui negra?

Solució

(a) P(cap negra)$=\frac{\binom{20}{0}\binom{80}{4}}{\binom{100}{4}}=\frac{\binom{80}{4}}{\binom{100}{4}}=\frac{80!}{76!\,\cancel{4!}}\;\frac{96!\,\cancel{4!}}{100!}=\frac{80\cdots 77}{100\cdots 97}=0.4$ és el mateix que calcular:

$$P(B_1\cap B_2\cap B_3\cap B_4)=P(B_1)P(B_2\,/\,B_1)P(B_3\,/\,B_1\cap B_2)P(B_4\,/\,B_1\cap B_2\cap B_3)=\frac{80}{100}\cdots\frac{77}{97}=0.4$$

(b) P(una negra)$=\frac{\binom{20}{1}\binom{80}{3}}{\binom{100}{4}}=\frac{20\cdot 80!}{77!\,3!}\;\frac{96!\,4!}{100!}=\frac{20\cdot 80\cdots 78\cdot 4!}{3!\cdot 100\cdots 97}=\frac{20\cdot 80\cdots 78\cdot 4}{100\cdots 97}=0.419$

és el mateix que: $\binom{4}{1}P(N_1\cap B_2\cap B_3\cap B_4)=4\frac{20}{100}\frac{80}{99}\frac{79}{98}\frac{78}{97}=0.419$

3 Models de probabilitat discrets

1. Per a les distribucions de probabilitat següents, expliqueu els seus paràmetres, què descriuen les v.a. indicant el recorregut, la funció de probabilitat, i les corresponents esperances i variàncies.

   a) Bernouilli; b) Binomial; c) Binomial Negativa; d) Geomètrica; e) Hipergeomètrica; f) Poisson.

   Solució

   (a) **Bernoulli**, $X \hookrightarrow b(p)$ $\quad X(\Omega) = \{0, 1\}$
   $X =$ nombre d'èxits en fer l'experiència aleatòria un sol cop, és a dir, 0 o 1.
   Paràmetre: $p =$ probabilitat d'èxit.

   $$\boxed{P(X = k) = f(k) = p^k(1-p)^{1-k}}$$

   $E(X) = p, \quad VAR(X) = p(1-p)$

   (b) **Binomial**, $X \hookrightarrow B(n,p)$ $\quad (X = X_1 + \cdots + X_n$, on $X_i \hookrightarrow b(p))$, $\quad X(\Omega) = \{0, 1, \cdots, n\}$
   $X =$ nombre d'èxits en n experiències aleatòries independents.
   Paràmetres: $n =$ nombre de repeticions de l'experiència, i $p =$ probabilitat d'èxit.

   $$\boxed{P(X = k) = f(k) = \binom{n}{k}p^k(1-p)^{n-k}}$$

   $E(X) = np, \quad VAR(X) = np(1-p)$

   (c) **Binomial negativa**, $X \hookrightarrow BN(r,p)$, $\quad X(\Omega) = \{0, 1, \cdots\}$
   $X =$ nombre de fracassos *abans de* l'èxit r-èsim.
   Paràmetres: $r =$ nombre d'èxits, i $p =$ probabilitat d'èxit.

   $$\boxed{P(X = k) = f(k) = \binom{k+r-1}{r-1}p^r(1-p)^k}$$

   $E(X) = \frac{r(1-p)}{p}, \quad VAR(X) = \frac{r(1-p)}{p^2}$

   Binomial negativa modificada, $Y \hookrightarrow BN^*(r,p)$, $\quad Y(\Omega) = \{0, 1, \cdots\}$
   $Y = X + 1 =$ nombre de repeticions de l'experiència *fins a* l'èxit r-èsim.

   $$\boxed{P(Y = k) = f(k) = \binom{k-1}{r-1}p^r(1-p)^{k-r}}$$

   $E(Y) = \frac{r}{p}, \quad VAR(Y) = \frac{r(1-p)}{p^2}$

   (d) **Geomètrica**, $X \hookrightarrow G(p)$ $\quad (G(p) = BN(1,p))$, $\quad X(\Omega) = \{0, 1, \cdots\}$
   $X =$ nombre de fracassos *abans del* primer èxit.
   Paràmetre: $p =$ probabilitat d'èxit.

   $$\boxed{P(X = k) = f(k) = p(1-p)^k}$$

   $E(X) = \frac{1-p}{p}, \quad VAR(X) = \frac{1-p}{p^2}$

   Geomètrica modificada, $Y \hookrightarrow G^*(p)$, $\quad Y(\Omega) = \{0, 1, \cdots\}$
   $Y = X + 1 =$ nombre de repeticions de l'experiència *fins al* primer èxit.

   $$\boxed{P(X = k) = f(k) = p(1-p)^{k-1}}$$

   $E(Y) = E(X+1) = E(X) + 1 = \frac{1}{p}, \quad VAR(Y) = \frac{1-p}{p^2}$

(e) **Hipergeomètrica**, $X \hookrightarrow HG(N,n,p)$, $X(\Omega) = \{0,1,\cdots,D\}$, on D és el nombre d'individuos de la població del tipus "èxit" ($N-D$ el complementari).

$X =$nombre d'èxits *de la mostra*.

Paràmetres: N nombre d'individus de la població, n nombre d'individus de la mostra, i p proporció d'èxits en la població ($p = D/N$).

$$\boxed{P(X=k) = f(k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} = \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}}$$

$$E(X) = np, \quad VAR(X) = np(1-p)\frac{N-n}{N-1}$$

(f) **Poisson**, $X \hookrightarrow P(\lambda)$, $X(\Omega) = \{0,1,\cdots\}$

$X =$nombre d'observacions d'un esdeveniment per unitat de magnitud emprada.

Paràmetre: $\lambda =$ nombre mitjà d'esdeveniments per unitat de magnitud emprada.

$$\boxed{P(X=k) = f(k) = \frac{e^{-\lambda}\lambda^k}{k!}}$$

$$E(X) = \lambda, \quad VAR(X) = \lambda$$

2. La probabilitat de rebre de manera errònia un bit enviat per un canal de transmissió digital és de 0.1. Indiqueu la v.a. i el model de distribució que feu servir a cada apartat.

(a) Si hem rebut 50 bits,
Quin és el nombre esperat de bits erronis?
Quina és la probabilitat de rebre: 1) exactament tres erronis, 2) com a molt tres erronis, 3) almenys tres erronis.

(b) Calculeu la probabilitat que hagi 15 bits correctament rebuts abans del primer erroni.
Quin és el nombre esperat de bits correctament rebuts abans del primer erroni?

(c) Quin és el nombre esperat de bits rebuts correctament abans del tercer erroni?
Quina és la probabilitat que arribin més de 5 bits correctes abans del tercer erroni?

Solució

(a) Sigui X la v.a. que compta el nombre de bits rebuts de forma errònia (observeu que considerem com a èxit rebre un bit erroni).

Aleshores, $X \hookrightarrow B(n,p)$ on $n = 50$ i $p = 0.1$.

$$E(X) = np = 50 \cdot 0.1 = 5$$

Ens demanen la probabilitat d'obtenir:

- Exactament 3 erronis: $P(X=3) = \binom{50}{3}(0.1)^3(0.9)^{50-3} = 0.1385651$

- Com a molt 3 erronis: $P(X \leq 3) = \sum_{i=0}^{3}\binom{50}{i}(0.1)^i(0.9)^{50-i} =$

$$= \left(\binom{50}{0}(0.1)^0(0.9)^{50} + \binom{50}{1}(0.1)(0.9)^{49} + \binom{50}{2}(0.1)^2(0.9)^{48} + \binom{50}{3}(0.1)^3(0.9)^{47}\right) = 0.2502939$$

- Almenys 3 erronis: $P(X \geq 3) = 1 - P(X \leq 2) = 0.8882712$

(b) Sigui Y la v.a. que compta el nombre de bits correctament rebuts abans del primer erroni. Aleshores, $Y \hookrightarrow G(0.1)$. La probabilitat de rebre 15 bits correctes abans del primer erroni:

$$P(Y = 15) = (0.9)^{15}(0.1) = 0.02058911$$

$$E(Y) = \frac{1-p}{p} = \frac{1-0.1}{0.1} = 9$$

(c) Sigui Z la v.a. que compta el nombre de bits correctament rebuts abans del tercer erroni. Aleshores, $Z \hookrightarrow BN(3, 0.1)$.

$$E(Z) = \frac{r(1-p)}{p} = \frac{3(1-0.1)}{0.1} = 27$$

- La probabilitat de rebre més de 5 bits correctes abans del 3 erroni:

$$P(Z > 5) = 1 - P(Z \leq 5) = 1 - \sum_{i=0}^{5} \binom{i+3-1}{3-1}(0.1)^3(0.9)^i = 0.9619082$$

3. Sabem que en un determinat col·lectiu de 100 persones 20 estan a favor d'una llei i la resta en contra. Quina és la probabilitat que a l'enquestar a 10 persones d'aquest col·lectiu, 2 d'elles estiguin a favor de la llei?

Solució

Sigui X la v.a. que compta el nombre de persones enquestades a favor de la llei.

Aleshores, $X \hookrightarrow HG(100, 10, 20/100)$.

$$P(X = 2) = \frac{\binom{20}{2}\binom{100-20}{10-2}}{\binom{100}{10}} = 0.3181706$$

Observació: Si no hi ha reemplaçament, la probabilitat de la intersecció es calculava amb probabilitats condicionades. Observeu que dóna el mateix: $\binom{10}{2}\frac{20}{100}\frac{19}{99}\frac{80}{98}\frac{79}{97}\frac{78}{96}\frac{77}{95}\frac{76}{94}\frac{75}{93}\frac{74}{92}\frac{73}{91} = 0.3181706$

Ara bé, si la població és molt gran (a vegades no es coneix el valor de N), aleshores es pot aproximar la hipergeomètrica per una binomial $B(n, p)$. Per exemple, suposeu que el col·lectiu és de 10000 persones de les quals un 20% (és a dir 2000) estan a favor de la llei. Compareu els resultats que en escollir 10 persones dues estiguin a favor en els dos models de probabilitat:

$$X \hookrightarrow HG(10000, 10, 20/100) \qquad P(X = 2) = \frac{\binom{2000}{2}\binom{10000-2000}{10-2}}{\binom{10000}{10}} = 0.302141$$

$$X \hookrightarrow B(10, 20/100) \qquad P(X = 2) = \binom{10}{2}\left(\frac{20}{100}\right)^2\left(\frac{80}{100}\right)^8 = 0.3019899$$

4. Sabem que el nombre d'errades per fulla segueix una distribució de Poisson. Si el nombre mitjà d'errades per fulla és de 0.2, quina és la probabilitat que una fulla no tingui errades? I en 10 fulles?

Sigui X la v.a. que compta el nombre d'errades per fulla. Aleshores, $X \hookrightarrow P(0.2)$.

$$P(X = 0) = \frac{(0.2)^0 e^{-0.2}}{0!} = e^{-0.2} = 0.8187$$

Sigui Y la v.a. que compta el nombre d'errades per 10 fulles.

Aleshores, $Y = X_1 + \cdots + X_{10} \hookrightarrow P(0.2 + \cdots + 0.2)) = P(2)$, on cada $X_i \hookrightarrow P(0.2)$.

$$P(Y = 0) = \frac{2^0 e^{-2}}{0!} = e^{-2} = 0.1353353$$

5. En una editorial la probabilitat que una pàgina tingui errors és $p = 0.005$ Tenint en compta que els errors en un llibre són independents d'una pàgina a una altra, i que el nombre de pàgines és $n = 400$, quina és la probabilitat que hi hagi almenys una fulla amb errors?

Solució

Sigui X la v.a. que compta el nombre de fulles amb errades. Aleshores, $X \hookrightarrow B(400, 0.005)$.

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = \binom{400}{0}(0.005)^0(1 - 0.005)^{400} = 0.865342$$

Observació: Quan n és gran i p petita (a la pràctica, $n \geq 50$ i $np < 5$), aleshores la binomial es pot aproximar per una Poisson amb $\lambda = np$.

Observeu que ens dóna en aquest cas aplicant la Poisson: $X \hookrightarrow P(400 * 0.005) = P(2)$

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \frac{2^0 e^{-2}}{0!} = 1 - e^{-2} = 0.8646647$$

6. Tornem al problema 2(b).

Havíem calculat la probabilitat de rebre 15 bits correctes (fracassos) abans del primer erroni (èxit), considerant $X \hookrightarrow G(0.1)$, on $p = 0.1$ és la probabilitat de rebre un bit erroni.

Però ara ho heu de tornar a calcular amb una informació addicional: se sap que s'han rebut més de 10 correctes (haureu d'usar probabilitats condicionades).

Indicació: $\displaystyle\sum_{i=0}^{k}(1-p)^i p = p\sum_{i=0}^{k}(1-p)^i = 1 - (1-p)^{k+1}$ (observeu que hem fet la suma d'una progressió geomètrica de raó 1-p). Això s'utilitza quan calculeu $P(X \leq k) = 1 - (1-p)^{k+1}$, tal i com ho podeu trobar indicat al formulari.

Solució

$$P(X \geq 15 / X > 10) = \frac{P(\,(X \geq 15) \cap (X > 10)\,)}{P(X > 10)} = \frac{P(X \geq 15)}{P(X > 10)}) = \frac{1 - P(X < 15)}{1 - P(X \leq 10)} =$$

$$= \frac{1 - P(X \leq 14)}{1 - P(X \leq 10)} = \frac{1 - \left(1 - (1-p)^{14+1}\right)}{1 - (1 - (1-p)^{10+1})} = \frac{(1-p)^{15}}{(1-p)^{11}} = (1-p)^4$$

El model geomètric compleix la propietat de *pèrdua de memòria* que diu:

$$P(X \geq s + t / X > t) = P(X \geq s)$$

Aplicant la propietat a l'exemple d'abans tenim:

$$P(X \geq 15 / X > 10) = P(X \geq 5) =$$

$$= P(X > 4) = 1 - P(X \leq 3) = (1-p)^4$$

4 Models de probabilitat continus

1. Per a les distribucions de probabilitat següents, digues quines són les funcions densitat de probabilitat i les corresponents esperances i variàncies.

   a) Uniforme contínua, b) exponencial, c) normal.

   Solució

   (a) **Uniforme contínua**, $X \hookrightarrow U(a,b)$

   $$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in (a,b) \\ 0 & \text{si } x \notin (a,b) \end{cases} \qquad E(X) = \frac{a+b}{2}, \quad VAR(X) = \frac{(b-a)^2}{12}$$

   (b) **Exponencial**, $X \hookrightarrow \epsilon(\lambda)$

   $$f(x) = \begin{cases} \lambda\, e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} \qquad E(X) = \frac{1}{\lambda}, \quad VAR(X) = \frac{1}{\lambda^2}$$

   (c) **Normal**, $X \hookrightarrow N(\mu, \sigma)$

   $$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty \qquad E(X) = \mu, \quad VAR(X) = \sigma^2$$

2. Per a les distribucions de probabilitat següents, digues quines són les funcions de distribució $F(X)$.

   a) Uniforme contínua, b) exponencial, c) normal.

   Solució

   (a) **Uniforme contínua**, $X \hookrightarrow U(a,b)$

   $$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \int_{a}^{x} \frac{1}{b-a}\,dt = \left[\frac{t}{b-a}\right]_a^x = \frac{x-a}{b-a}, \quad \text{per a } a < x < b.$$

   $$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ \dfrac{x-a}{b-a} & \text{si } x \in (a,b) \\ 1 & \text{si } x \geq b \end{cases}$$

   (b) **Exponencial**, $X \hookrightarrow \epsilon(\lambda)$

   $$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} \lambda\, e^{-\lambda t}\,dt = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}, \quad \text{per a } 0 < x < +\infty$$

   $$F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

   (c) **Normal**, $X \hookrightarrow N(\mu, \sigma)$

   $$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad dt$$

3. Donada la funció: $f(x) = \begin{cases} k\cos x & \text{si } x \in [0, \pi/2] \\ 0 & \text{altrament} \end{cases}$

   (a) Calculeu k perquè $f(x)$ sigui una funció de densitat.
   (b) Trobeu també la funció de distribució $F(x)$.
   (c) Si X és una variable amb aquesta densitat, calculeu $E(X)$.

   Solució

   (a) Perquè $f(x)$ sigui una funció de densitat, cal que $\int_{-\infty}^{+\infty} f(x)\,\mathrm{dx} = 1$.

$$1 = \int_{-\infty}^{+\infty} k\cos x\,\mathrm{dx} = \int_{0}^{\pi/2} k\cos x\,\mathrm{dx} = k\,[\sin x]_0^{\pi/2} = k(1-0) = k$$

   Per tant, $k = 1$.

$$F(x) = \int_{-\infty}^{x} \cos t\,\mathrm{dt} = \int_{0}^{x} \cos t\,\mathrm{dt} = [\sin t]_0^x = \sin x$$

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sin x & \text{si } x \in [0, \pi/2] \\ 1 & x > \pi/2 \end{cases}$$

$$E(X) = \int_{-\infty}^{+\infty} x\,\cos x\,\mathrm{dx} = \int_{0}^{\pi/2} x\cos x\,\mathrm{dx} = [x\sin x]_0^{\pi/2} - \int_{0}^{\pi/2} \sin x\,dx = \frac{\pi}{2} + [\cos x]_0^{\pi/2} = \frac{\pi}{2} - 1$$

$$\uparrow$$
$$u = x \Longrightarrow du = dx$$
$$dv = \cos x\,dx \Longrightarrow v = \sin x$$

4. El Tram passa de manera regular cada 8 minuts. Suposem que el temps d'espera en la parada és una v.a. que segueix una distribució uniforme entre 0 i 8 minuts. Quina és la probabilitat que arribant a la parada hagi d'esperar menys de 4 minuts? Si agafo el tram moltes vegades, quin és el temps mitjà d'espera?

   Solució Sigui X la v.a. que compta el temps d'espera en la parada del tram, $X \hookrightarrow U(0,8)$

$$P(X < 4) = \int_0^4 \frac{1}{8-0}\,dx = \left[\frac{x}{8}\right]_0^4 = \frac{4}{8} = \frac{1}{2}$$
$$E(X) = \frac{0+8}{2} = 4.$$

5. Suposem que en un servei tècnic el temps que transcorre entre l'arribada de dues persones es distribueix com una v.a. exponencial de mitjana 5 minuts.

   (a) Quin és el paràmetre de la v.a?
   (b) Quina és la probabilitat que transcorrin més de 6 minuts entre l'arribada de dues persones al servei tècnic?
   (c) Sabent que ja han transcorregut més de 2 minuts de l'arribada d'una persona, quina és la probabilitat que transcorrin 4 minuts més fins que arribi una altra?

   Solució

   (a) Sigui X la v.a. que compta el temps que transcorre entre l'arribada de dues persones, $X \hookrightarrow \epsilon(\lambda)$.
   Sabem que $E(X) = 5 \Longleftrightarrow \lambda = \dfrac{1}{5}$

(b) $P(X>6)=\int_6^{+\infty}\frac{1}{5}\,e^{-\frac{x}{5}}\,dx=\lim_{t\to+\infty}\left[-e^{-\frac{x}{5}}\right]_6^t=e^{-\frac{6}{5}}$

O bé, ho podríem calcular fent: $P(X>6)=1-P(X\leq 6)=1-F(6)=1-(1-e^{-\frac{6}{5}})=e^{-\frac{6}{5}}$.

(c) $P\left(X>4+2\,/\,X>2\right)=\dfrac{P\left((X>6)\cap(X>2)\right)}{P(X>2)}=\dfrac{P(X>6)}{P(X>2)}=\dfrac{1-P(X\leq 6)}{1-P(X\leq 2)}=\dfrac{e^{-\frac{6}{5}}}{e^{-\frac{2}{5}}}=e^{-\frac{4}{5}}$

Observació: Hem vist que es compleix: $P(X>4+2\,/\,X>2)=P(X>4)=e^{-\frac{4}{5}}$.

La v.a. exponencial té la propietat de la pèrdua de memòria: $P(X>s+t\,/\,X>t)=P(X>s)$.

6. Sigui $Z\hookrightarrow N(0,1)$. Calculeu les probabilitats següents.

Solució

(a) $P(Z<1.5)=\Phi(1.5)=0.9331928$

(b) $P(Z>1.5)=1-\Phi(1.5)=0.0668072$

(c) $P(Z<-1.5)=1-\Phi(|-1.5|)=1-\Phi(1.5)=0.0668072$

(d) $P(Z>-1.5)=\Phi(|-1.5|)=\Phi(1.5)=0.9331928$

(e) $P(-1.5<Z<1.5)=\Phi(1.5)-\Phi(-1.5)=\Phi(1.5)-(1-\Phi(1.5))=2\,\Phi(1.5)-1=0.8663856$

7. Sigui $Z\hookrightarrow N(0,1)$. Feu servir la taula de la normal per trobar el valor de z que verifiqui $P(Z>z)=0.95154$.

Solució

(a) Donat que la probabilitat que Z sigui més gran que un cert valor z és 0.95154 més gran que 0.5, deduïm que $z<0$. Aleshores, $P(Z>z)=P(Z\leq|z|)=0.95154$.

En la taula trobem que: $P(Z\leq 1.66)=0.95154$, per tant, el valor buscat és $z=-1.66$.

8. Sigui $X\hookrightarrow N(\mu,\sigma)$ amb $\mu=2$ i $\sigma=5$. Calculeu les probabilitats següents.

Solució

(a) $P(X<9.5)=P(\frac{X-2}{5}<\frac{9.5-2}{5})=P(Z<1.5)=\Phi(1.5)=0.9331928$

(b) $P(-1<X<9.5)=P(\frac{-1-2}{5}<\frac{X-2}{5}<\frac{9.5-2}{5})=P(-0.6<Z<1.5)=$

$=\Phi(1.5)-(1-\Phi(0.6))=\Phi(1.5)+\Phi(0.6)-1=0.6589397$

9. Interpolem quan:

- Volem calcular $\Phi(z)$ i coneixem z però té més de dues xifres decimals

$$\boxed{\Phi(z)=\Phi(a)+\frac{\Phi(b)-\Phi(a)}{b-a}(z-a)}$$

- Busquem z i coneixem $\Phi(z)$ però no surt a la taula de la normal estàndard.

$$\boxed{z=a+\frac{b-a}{\Phi(b)-\Phi(a)}(\Phi(z)-\Phi(a))}$$

(a) Aplica-ho per calcular $\Phi(1.332)$

(b) Aplica-ho per trobar z tal que $P(z) = 0.9$.

Solució

(a) $\Phi(1.332) = \Phi(1.33) + \dfrac{\Phi(1.34) - \Phi(1.33)}{1.34 - 1.33}(1.332 - 1.33) = 0.90854$

(b) $\Phi(1.28) < \Phi(z) = 0.9 < \Phi(1.29)$

$$z = 1.28 + \frac{1.29 - 1.28}{\Phi(1.29) - \Phi(1.28)}(0.9) - \Phi(1.28)) = 1.28 + \frac{1.29 - 1.28}{0.90147 - 0.89973}(0.9 - 0.90147) = 1.2884$$

10. L'organització Mensa requereix per poder ser membre tenir un qüocient intel·lectual igual o superior a 130.9. Indiqueu la v.a. i el model de distribució que feu servir a cada apartat.

    (a) Admetem que el qüocient intel·lectual es distribueix com una normal amb mitjana 100 i desviació tipus 15. Calculeu la probabilitat p que un individu, escollit a l'atzar, tingui un qüocient igual o superior a 130.9 (condició que es demana pertànyer a Mensa). Arrodoniu el resultat a dos decimals.

    (b) Tenim 50 individus triats aleatòriament.
    Quina és la probabilitat que com a mínim hi hagi dos d'ells que puguin pertànyer a Mensa?
    Quin és el nombre esperat de persones que podrien pertànyer a Mensa?

Solució

(a) $p = 1 - P(X < 130.9) = 1 - P(\dfrac{X - 100}{15} < \dfrac{130.9 - 100}{15}) = 1 - P(Z < 2.06) = 1 - 0.98 = 0.02$

(b) Considerem X la v.a. que compta el nombre de persones que poden pertànyer a Mensa.
$X \hookrightarrow B(50, p)$, amb $p = 0.02$.

$$P(X \geq 2) = 1 - (P(X = 0) + P(X = 1)) = 1 - \sum_{k=0}^{1} \binom{50}{k} p^k (1-p)^{50-k} =$$

$$1 - (\binom{50}{0} p^0 (1-p)^{50-0} + \binom{50}{1} p (1-p)^{50-1}) = 1 - ((1-p)^{50} + 50\, p(1-p)^{49}) =$$

$$1 - ((0.98)^{50} + 50\,(0.02)(0.98)^{49}) = 0.2642286$$

$$E(X) = np = 50(0.02) = 1$$

11. Sabem que el temps de funcionament d'un component electrònic segueix una distribució exponencial i és, en promig, de 10 dies.

    (a) Troba el valor del paràmetre de la v.a. que compta el temps de funcionament.

    (b) Quina és la probabilitat que el component electrònic funcioni més de 10 dies?

    (c) Quina és la probabilitat que com a màxim deixin de funcionar dos components en un mes.

    (d) Quin és el nombre esperat de components que deixaran de funcionar en un mes?

Solució

(a) $T \hookrightarrow \epsilon(\lambda)\,,\ E(T) = 10 = \dfrac{1}{\lambda} \Longrightarrow \lambda = \dfrac{1}{10} = 0.1.$

(b) $P(T > 10) = 1 - P(T \leq 10) = 1 - (1 - e^{-(0.1)10}) = e^{-1} = 0.368$

(c) Considerem $X \hookrightarrow P(30(0.1)) = P(3)$ la v.a. que compta el nombre de components electrònics que fallen al mes.

$P(X \le 2) = \sum_{i=0}^{2} e^{-3}\frac{3^i}{i!} = e^{-3}\left(\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right) = 0.42319$

(d) $E(X) = 3$

12. Una empresa constructora produeix rajoles de mig quilo que ven a pes. Se sap que el pes en grams de la rajola és una v.a. de distribució normal de mitja 498 i variància 16. El cost de la producció és de 1 euro cada 100 grams, mentre que es ven a 2 euros. Anomenen G la v.a. que compta els guanys.

(a) Calculeu la probabilitat que una rajola surti amb un pes inferior a 490 grams.

(b) Quina és la probabilitat que una rajola presenti un pes entre 480 i 490 grams?

(c) Expresseu la relació entre les variables G i X. Quin és el guany promig per rajola?

(d) Quin guany s'obté com a màxim en el 90% dels casos?

Solució

(a) $X \hookrightarrow N(498, \sqrt{16}) = N(498, 4)$.

$$P(X \le 490) = P(\frac{X-498}{4} < \frac{490-498}{4}) = P(Z \le -2) = \Phi(-2) = 1 - \Phi(2) = 1 - 0.97725 = 0.02275$$

(b)

$$P(480 \le X \le 490) = P\left(\frac{480-498}{4} \le \frac{X-498}{4} \le \frac{490-498}{4}\right) = P(-4.5 \le Z \le -2) =$$

$$= \Phi(-2) - \Phi(-4.5) = 1 - \Phi(2) - (1 - \Phi(4.5)) = 0.02275 - (1-1) = 0.02275$$

(c) S'inverteix 1 euro per a cada 100 grams, és a dir, 0.01 per gram.

$$G = (0.02 - 0.01)X = (0.01)X$$

El guany promig per rajola és: $E(G) = (0.01)E(X) = 4.98$

(d) Donat que $G = (0.01)X$, és combinació lineal d'X que és una v.a. normal, G també té una distribució normal. Ja coneixem la mitja ($E(G)$). Calculem la variància:

$$VAR(G) = (0.01)^2 VAR(X) = (0.0001)16 = 0.0016$$

Així doncs, $G \hookrightarrow N(4.98, \sqrt{0.0016}) = N(4.98, 0.04)$

Ara busquem el valor màxim a tal que: $P(G \le a) = 0.9$.

$$P(\frac{G-4.98}{0.04} \le \frac{a-4.98}{0.04}) = P(Z \le \frac{a-4.98}{0.04}) = 0.9$$

Mirant la taula de la normal standaritzada, $\Phi(1.285) = 0.9$

$$\frac{a-4.98}{0.04} = 1.285 \Longleftrightarrow a = 5.0314$$

5 Mostreig

Aproximant el model binomial pel normal

1. Es llença una moneda a l'aire 100 cops. Sigui X la variable aleatòria que compta el nombre de cares. Calculeu la probabilitat que en aquests 100 llençaments surtin 60 o més de 60 cares.

   Solució

   X segueix una distribució binomial de paràmetres $n = 100$ i $p = 0.5$, $X \hookrightarrow B(100, 0.5)$.

$$P(X \geq 60) = \sum_{k=60}^{100} \binom{100}{k} p^k (1-p)^{100-k}$$

   Calculant el complementari, els càlculs no se simplifiquen:

$$P(X \geq 60) = 1 - P(X < 60) = 1 - P(X \leq 59) = 1 - \sum_{k=0}^{\mathbf{59}} \binom{100}{k} p^k (1-p)^{100-k} =$$

$$1 - \left(\binom{100}{0} p^0 (1-p)^{100} + \cdots + \binom{100}{59} p^{59} (1-p)^{100-59} \right)$$

   Caldria fer els càlculs amb programari! Fent servir R:

```
n ← 100
p ← 0.5
pbinom(100, n, p)-pbinom(59, n, p)
0.02844397
1- pbinom( 59, n, p)
0.02844397
```

   És possible també, en aquest cas, obtenir el resultat aproximant la binomial per una normal?

   Observeu que: $n = 100$, $np = 100(0.5) = 50$ i $nq = n(1-p) = 50$.

   Aleshores, donat que $n > 30$, $np > 5$ i $nq > 5$, podem aproximar la binomial per una normal de paràmetres:

$$\mu = \mathrm{E}(X) = np = 50, \qquad \sigma = \sqrt{\mathrm{VAR(X)}} = \sqrt{npq} = 5$$

$$X \hookrightarrow B(100, 0.5) \sim \mathcal{N} \hookrightarrow N(50, 5)$$

   Estem aproximant una variable discreta per una contínua, per això cal fer l'anomenada **correcció per continuïtat**:

$$1 - P(0 \leq X \leq 59) \sim 1 - P(-\mathbf{0.5} \leq \mathcal{N} \leq 59 + \mathbf{0.5})$$

$$= 1 - P\left(\frac{-\mathbf{0.5} - 50}{5} \leq \frac{\mathcal{N} - 50}{5} \leq \frac{59 + \mathbf{0.5} - 50}{5} \right) =$$

$$= 1 - P(-10.1 \leq Z \leq 1.9) = 1 - (\Phi(1.9) - \Phi(-10.1)) =$$

$$= 1 - \Phi(1.9) + \Phi(-10.1) = 1 - \Phi(1.9) + 1 - \Phi(10.1) = 2 - 1.971283 = 0.02871656$$

   Amb R,

   1 − (pnorm(59+**0.5**, 50, 5) − pnorm(−**0.5**, 50, 5))

   0.02871656

2. Sigui X una v.a. amb $\mu = 25$ i $\sigma = 4$. Determineu l'esperança, la variància i la desviació típica de la mitjana aritmètica d'una mostra de mida 100, $\bar{X}_{100}$.

   Solució

   $\mu_{\bar{X}_{100}} = \mu_X = 25$

   $\sigma_{\bar{X}_{100}} = \frac{\sigma_X}{\sqrt{n}} = \frac{4}{\sqrt{100}} = 0.4$

3. Un tub està format per la unió de tres peces cilíndriques. La longitud de la primera peça es distribueix segons una normal $N(10,2)$ i les altres dues segons una normal $N(15,3)$. Suposant que les tres v.a. són independents, quina distribució tindrà la longitud del tub?

   Solució

   $X_1 \hookrightarrow N(10,2)$, $X_2 \hookrightarrow N(15,3)$ i $X_3 \hookrightarrow N(15,3)$ són v.a. independents.

   Sigui $Y = X_1 + X_2 + X_3$. La combinació lineal de v.a. normals també té una distribució normal.

   L'esperança de la suma és suma d'esperances, i la variància de la suma és suma de variàncies per ser independents.

$$Y \hookrightarrow N(40, \sqrt{22})$$

4. Estem construint una estructura formada per peces cilíndriques. La longitud de les peces cilíndriques segueix una mateixa distribució amb mitjana 10 cm i desviació típica 2 cm. Unint 31 peces, quina és la probabilitat que l'estructura resultant sigui superior a 301 cm?

   Solució

   Apliquem el teorema central del límit per aproximar la distribució de la longitud total de l'estructura:

$$L = L_1 + \cdots + L_n \sim N(n\mu, \sqrt{n}\sigma)$$

   Fent $n = 31$, $\mu = 10$, i $\sigma = 2$, $L \sim N(310\,,\, 2\sqrt{31})$

$$P(L > 301) = P\left(\frac{L-310}{2\sqrt{31}} > \frac{301-310}{2\sqrt{31}}\right) = P\left(Z > -0.81\right) = P\left(Z < 0.81\right) = 0.79103$$

5. Estem construint una estructura formada per peces cilíndriques. La longitud de les peces cilíndriques segueix una mateixa distribució amb mitjana 10 cm i desviació típica 2 cm. Determineu el nombre mínim de peces que és necessari unir perquè la longitud de l'estructura total sigui superior a 300 cm, amb una probabilitat superior a 95.154%.

   Solució

   Apliquem el teorema central del límit per aproximar la distribució de la longitud total de l'estructura:

$$L = L_1 + \cdots + L_n \sim N(n\mu, \sqrt{n}\sigma) = N(10\,n\,,\, 2\sqrt{n})$$

$$P(L \geq 300) > 0.95154 \Longleftrightarrow P(Z \geq \frac{300-10\,n}{2\sqrt{n}}) > 0.95154$$

   Recordeu el prob. 6 de les qüestions dels models continus per deduir que $\frac{300-10\,n}{2\sqrt{n}} < 0$ i, per tant,

$$P(Z \geq \frac{300-10\,n}{2\sqrt{n}}) = P(Z \leq |\frac{300-10\,n}{2\sqrt{n}}|) > 0.95154$$

   Donat que $P(Z \leq 1.66) = 0.95154$, imposem que $\frac{300-10\,n}{2\sqrt{n}} < -1.66$

Podem trobar el primer valor de n que verifica la desigualtat resolent l'equació de segon grau que s'obté substituint $\sqrt{n} = x$. Aleshores la solució positiva elevada al quadrat és: $(5.645741)^2 = 31.87439$.

Per tant, el nombre mínim de peces necessàries és $n = 32$.

Fent $n = 31$ obtenim $\dfrac{300 - 310}{2\sqrt{31}} = -0.8980265 \not< -1.66$.

Fent $n = 32$ obtenim $\dfrac{300 - 320}{2\sqrt{32}} = -1.767767 < -1.66$.

6. El temps de vida útil d'una màquina és una v.a. normal amb mitjana de 4 anys i desviació típica 1.5 anys.

   (a) Determineu la probabilitat que la vida mitjana d'una mostra aleatòria de 50 d'aquestes màquines estigui compresa entre 3.5 i 3.8 anys?

   (b) Determineu el valor θ pel qual el 10% de les mitjanes de mostres aleatòries de mida 9 siguin majors a θ (percentil 90).

Solució

(a) Tenim una mostra de tamany 50 d'una v.a. $X \hookrightarrow N(\mu, \sigma)$.
La distribució de probabilitat de la mitjana mostral corresponent a una m.a.s. d'una v.a. normal, també és normal (no importa el tamany de la mostra):

$$\bar{X}_{50} \hookrightarrow N(\mu_X, \frac{\sigma_X}{\sqrt{n}}) = N(4, \frac{1.5}{\sqrt{50}}) = N(4, 0.2121)$$

$$P(3.5 \leq \bar{X}_{50} \leq 3.8) = P(\frac{3.5-4}{0.2121} < Z < \frac{3.8-4}{0.2121}) = P(-2.35 < Z < -0.94) =$$

$$= \Phi(-0.94) - \Phi(-2.36) = 1 - \Phi(0.94) - (1 - \Phi(2.36)) = \Phi(2.36) - \Phi(0.94) = 0.99086 - 0.82639 = 0.16447$$

(b)

$$\bar{X}_9 \hookrightarrow N(\mu_X, \frac{\sigma_X}{\sqrt{n}}) = N(4, \frac{1.5}{\sqrt{9}}) = N(4, 0.5)$$

$$P(\bar{X}_9 > \theta) = P(\frac{\bar{X}_9 - 4}{0.5} > \frac{\theta - 4}{0.5}) = P(Z > 2(\theta - 4)) = 0.1$$

$$\Longleftrightarrow 1 - P(Z \leq 2(\theta - 4)) = 0.9$$

$\Phi(1.281) = 0.9$. Per tant, fem: $2(\theta - 4) = 1.281 \Longleftrightarrow \theta = 4.64$

6 Estimadors

1. (Ex.15-16-Q1) Sigui X una v.a. discreta que pren valors $0, 1, -1$ amb probabilitat:

$$P(X = 0) = 1/2,\ P(X = 1) = a,\ P(X = -1) = (1/2) - a, \text{ on } \quad 0 < a < 1/2 \quad \text{és conegut.}$$

Sigui $X_1, \cdots, X_n$ una m.a.s. d'X i sigui $\bar{X} = \dfrac{X_1 + \cdots + X_n}{n}$ la mitjana mostral. Definim l'estimador $T = \dfrac{\bar{X}}{2} + \dfrac{1}{4}$.

(a) Calculeu $E(T)$ i $VAR(T)$

(b) És T un estimador insesgat i consistent d'a? Justifica la resposta.

Solució

(a) $E(X) = \displaystyle\sum_{i=1}^{n} x_i P(X = x_i) = 0 \cdot \frac{1}{2} + 1 \cdot a + (-1)(\frac{1}{2} - a) = 2a - \frac{1}{2}$

$E(X^2) = \displaystyle\sum_{i=1}^{n} x_i^2 P(X = x_i) = 0^2 \cdot \frac{1}{2} + 1^2 \cdot a + (-1)^2(\frac{1}{2} - a) = \frac{1}{2}$

$VAR(X) = E(X^2) - (E(X))^2 = \dfrac{1}{2} - (2a - \dfrac{1}{2})^2 = \dfrac{1}{4} + 2a - 4a^2$

$$E(T) = \frac{1}{2}E(\bar{X}) + \frac{1}{4} = \frac{1}{2}E(X) + \frac{1}{4} = a, \qquad VAR(T) = \frac{1}{4}\frac{VAR(X)}{n} = \frac{1}{4n}\left(\frac{1}{4} + 2a - 4a^2\right)$$

(b) $E(T) = a \Longrightarrow T$ no té biaix.

$$\lim_{n\to\infty} VAR(T) = \lim_{n\to\infty} \frac{VAR(X)}{4n} = 0.$$

T és un estimador consistent donat que és asimptòticament sense biaix (de fet, no en té) i asimptòticament la variància és nul·la.

2. (Ex.15-16-Q2) Sigui $X_1, \cdots, X_n$ una m.a.s. d'X que té per funció densitat

$$f(x) = \begin{cases} (1+\theta)x^\theta & 0 < x < 1 \\ 0 & x \notin (0,1) \end{cases} \qquad \text{on} \quad \theta > -1.$$

Una mostra aleatòria produeix les dades:

$$x_1 = 0.92;\ x_2 = 0.79;\ x_3 = 0.90;\ x_4 = 0.65;\ x_5 = 0.86.$$

(a) Calculeu $E(X)$ en funció del paràmetre θ.

(b) Trobeu un estimador $\hat{\theta}$ del paràmetre θ pel mètode dels moments.

(c) Quina és l'estimació del paràmetre θ segons la mostra de cinc dades?

Solució

(a) $$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, \mathrm{dx} = \int_0^1 \mathrm{x\, f(x)\, dx} = \int_0^1 (1+\theta)\, \mathrm{x}^{\theta+1}\, \mathrm{dx} \overset{\theta > -1}{=} (1+\theta)\left[\frac{\mathrm{x}^{\theta+2}}{\theta+2}\right]_0^1 = \frac{1+\theta}{2+\theta}$$

(b) $E(X) = \bar{X} \Longleftrightarrow \dfrac{1+\theta}{2+\theta} = \bar{X} \Longleftrightarrow \theta = \dfrac{2\bar{X} - 1}{1 - \bar{X}}$. Per tant, $\hat{\theta} = \dfrac{2\bar{X} - 1}{1 - \bar{X}}$

(c) De la mostra obtenim $\bar{x} = 0.824$.

L'estimació del paràmetre és $\theta_{obs} = \dfrac{2\bar{x} - 1}{1 - \bar{x}} = 3.6818$

3. Apliqueu el mètode de la màxima versemblança per trobar l'estimador del paràmetre de cadascuna de les v.a. que teniu a continuació. Indiqueu la funció de màxima versemblança, l'estimador i el valor del mateix segons la mostra observada.

   (a) Bernouilli, $X \hookrightarrow b(p)$.

   $(X_1, \cdots, X_N)$ m.a.s. d'X. La mostra observada és $(X_1 = x_1, \cdots, X_{100} = x_{100})$, amb $\sum_{i=1}^{100} x_i = 20$, $(x_i \in \{0, 1\}$, segons indiqui un èxit o un fracàs).

   (b) Binomial $X \hookrightarrow B(n, p)$, coneixem $n = 100$ i es vol estimar p.

   $(X_1, \cdots, X_N)$ és una m.a.s. d'X i la mostra observada és $(X_1 = 20, X_2 = 22, X_3 = 18)$.

   (c) Poisson $X \hookrightarrow P(\lambda)$.

   $(X_1, \cdots, X_N)$ és una m.a.s. d'X i la mostra observada és $(X_1 = 5, X_2 = 4, X_3 = 6, X_4 = 6)$.

   (d) Exponencial $X \hookrightarrow \epsilon(\lambda)$.

   $(X_1, \cdots, X_N)$ és una m.a.s. d'X i la mostra observada és $(X_1 = 5, X_2 = 4, X_3 = 6, X_4 = 6)$.

Solució (feu els detalls!).

(a) Recordeu la funció de probabilitat de la v.a. Bernouilli: $f(x) = p^x(1-p)^{1-x}$, on $X(\Omega) = \{0, 1\}$.

$$v(p) = \prod_{i=1}^{N} P(X_i = x_i) = p^{\sum_i x_i}(1-p)^{N - \sum_i x_i}$$

$$\frac{d}{dp}(ln(v(p))) = 0 \Longrightarrow \hat{p} = \overline{X} \qquad \text{De la mostra obtenim: } p_{obs} = \bar{x} = \frac{20}{100} = 1/5$$

(b) Recordeu la funció de probabilitat de la v.a. Binomial: $f(x) = \binom{n}{x} p^x(1-p)^{1-x}$, on $X(\Omega) = \{0, 1, \cdots, n\}$.

$$v(p) = \prod_{i=1}^{N} P(X_i = x_i) = \prod_{i=1}^{N} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} = \prod_{i=1}^{N} \binom{n}{x_i} \left(p^{\sum_i x_i}(1-p)^{nN - \sum_i x_i}\right)$$

$$\frac{d}{dp}(ln(v(p))) = 0 \Longrightarrow \hat{p} = \frac{\overline{X}}{n} \qquad \text{De la mostra obtenim: } p_{obs} = \frac{\bar{x}}{n} = \frac{(20 + 22 + 18)/3}{100} = 1/5$$

(c) Recordeu la funció de probabilitat de la v.a. Poisson: $f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$, on $X(\Omega) = \{0, 1, \cdots, +\infty\}$.

$$v(\lambda) = \prod_{i=1}^{N} P(X_i = x_i) = \prod_{i=1}^{N} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-N\lambda} \frac{\lambda^{\sum_i x_i}}{x_1! \cdots x_N!}$$

$$\frac{d}{d\lambda}(ln(v(\lambda))) = 0 \Longrightarrow \hat{\lambda} = \overline{X} \qquad \text{De la mostra obtenim: } \lambda_{obs} = \bar{x} = \frac{5 + 4 + 6 + 6}{4} = 21/4$$

(d) Recordeu la funció de densitat de la v.a. exponencial: $f(x) = \lambda\, e^{-\lambda x}$, on $X(\Omega) = (0, +\infty)$.

$$v(\lambda) = \prod_{i=1}^{N} f(x_i) = \prod_{i=1}^{N} \lambda\, e^{-\lambda x_i} = \lambda^N e^{-\lambda \sum_i x_i}$$

$$\frac{d}{d\lambda}(ln(v(\lambda))) = 0 \Longrightarrow \hat{\lambda} = \frac{1}{\overline{X}} \qquad \text{De la mostra obtenim: } \lambda_{obs} = \frac{1}{\bar{x}} = \frac{4}{5 + 4 + 6 + 6} = 4/21$$

4. Sigui $X_1, \cdots, X_n$ una m.a.s. d'una variable exponencial de paràmetre $1/\beta$, $X \hookrightarrow \epsilon\ (1/\beta)$, que té per funció densitat $f(x) = \dfrac{1}{\beta} e^{-x/\beta}$, $x > 0$. Recordeu que l'esperança d'una v.a. exponencial és l'invers del seu paràmetre ($E(X) = \beta$), i la variància aquest valor al quadrat ($VAR(X) = \beta^2$).

Una mostra aleatòria produeix les dades:

$$x_1 = 0.81,\ x_2 = 0.68,\ x_3 = 0.89,\ x_4 = 0.54,\ x_5 = 0.75.$$

(a) Trobeu un estimador $\hat{\beta}$ del paràmetre β, pel mètode de la màxima versemblança.

(b) Quina és l'estimació del paràmetre β segons la mostra de cinc dades?

(c) Calculeu l'esperança i la variància de l'estimador trobat. Raoneu si és un estimador consistent.

Solució

(a) $v(\beta) = \displaystyle\prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\beta}\, e^{-x_i/\beta} = (1/\beta)^n e^{-(1/\beta)\sum_{i=1}^{n} x_i}$

$$ln(v(\beta) = n \ln(1/\beta) - (1/\beta) \sum_{i=1}^{n} x_i = -n \ln(\beta) - (1/\beta) \sum_{i=1}^{n} x_i$$

$$\frac{d}{d\beta}\left(ln(v(\beta))\right) = 0 \Longleftrightarrow \frac{-n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{n} x_i = 0 \Longleftrightarrow -\beta n + \sum_{i=1}^{n} x_i = 0$$

$\Longrightarrow \hat{\beta} = \bar{X}$

(b) De la mostra obtenim: $\beta_{obs} = \bar{x} = (0.81 + 0.68 + 0.89 + 0.54 + 0.75)/5 = 0.734$

(c) $E(\hat{\beta}) = E(\bar{X}) = \dfrac{1}{n} \displaystyle\sum_{=1}^{n} E(X_i) = E(X) = \beta$

$$VAR(\hat{\beta}) = VAR(\bar{X}) = \frac{1}{n^2} VAR(\sum_{i=1}^{n} X_i) \stackrel{*}{=} \frac{1}{n^2} \sum_{i=1}^{n} VAR(X_i) = \frac{1}{n} VAR(X) = \frac{1}{n}\beta^2$$

$*$ Aquesta igualtat és certa quan les v.a. són independents, i ho són per ser una m.a.s.

Així doncs, $\hat{\beta}$ és un estimador sense biaix ($E(\hat{\beta}) - \hat{\beta} = 0$). Per tant, també asimptòticament sense biaix.

D'altra banda, $\hat{\beta}$ és asimptòticament sense variància: $\displaystyle\lim_{n\to\infty} VAR(\hat{\beta}) = \frac{1}{n}\beta^2 = 0.$

Deduïm que $\hat{\beta}$ és un estimador consistent (asimptòticament sense biaix i sense variància).

5. Sigui X una variable contínua que té per funció densitat $f(x) = \dfrac{2(\theta - x)}{\theta^2}$, $0 \leq x \leq \theta$.

Sigui $X_1, \cdots, X_n$ una m.a.s de X.

(a) Trobeu un estimador $\hat{\theta}$ del paràmetre θ, pel mètode dels moments.

(b) Estudieu el biaix de l'estimador que heu trobat.

(c) Calculeu la variància de l'estimador trobat. Es tracta d'un estimador consistent?

Solució

(a) Per trobar l'estimador pel mètode dels moments igualem l'esperança a la mitjana mostral:

$$E(X) = \int_0^\theta x\, \frac{2(\theta - x)}{\theta^2} dx = \frac{2}{\theta^2} \left[\frac{\theta x^2}{2} - \frac{x^3}{3}\right]_0^\theta = \frac{\theta}{3}$$

$E(X) = \bar{X} \Longleftrightarrow \dfrac{\theta}{3} = \bar{X}$ Per tant, $\hat{\theta} = 3\bar{X}$

$E(\hat{\theta}) = E(3\bar{X}) = \dfrac{3}{n} E(\displaystyle\sum_{i=1}^{n} X_i) = 3E(X) = \theta \Longleftrightarrow E(\hat{\theta}) = \theta \Longleftrightarrow \hat{\theta}$ és un estimador sense biaix.

(b) $VAR(\hat{\theta}) = VAR(3\bar{X}) = \frac{9}{n^2} VAR(\sum_{i=1}^{n} X_i) \overset{*}{=} \frac{9}{n^2} \sum_{i=1}^{n} VAR(X_i) = \frac{9}{n} VAR(X)$

$*$ Aquesta igualtat és certa quan les v.a. són independents, com és en el cas d'una m.a.s.

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

$$VAR(X) = E(X^2) - (E(X))^2 = \frac{\theta^2}{6} - \left(\frac{\theta}{3} \right)^2 = \frac{\theta^2}{18}$$

$$VAR(\hat{\theta}) = \frac{9}{n} VAR(X) = \frac{9\theta^2}{18n} = \frac{\theta^2}{2n}, \qquad \lim_{n\to\infty} VAR(\hat{\theta}) = \frac{\theta^2}{2n} = 0 \Longrightarrow \hat{\theta} \text{ és asimptòticament sense variància.}$$

A més, és asimptòticament sense biaix (si no té biaix, també és asimptòticament sense biaix).
Donat que és asimptòticament sense variància i sense biaix deduïm que $\hat{\theta}$ és un estimador consistent.

7 Intervals de confiança

1. Descriviu els passos a seguir en l'estimació de paràmetres per intervals de confiança, indicant també les instruccions en R.

   Solució

   ### Estimació de la mitjana SUPOSANT NORMALITAT.

   No importa el tamany de la mostra.

   Suposem ara que SIGMA és CONEGUT.

   mean(x)+c(-1,1)* qnorm(1-alfa/2) *sigma/sqrt(n)

   Si el paràmetre SIGMA és DESCONEGUT,

   mean(x)+c(-1,1)* qt(1-alfa/2,df=n-1) *sd(x)/sqrt(n)

   t.test(x) Instrucció directa, on x=vector de la mostra.

   Per defecte, correct=TRUE, conf.level=0.95, alternative='two.sided' (en els contrastos d'hpòtesi).

   Per canviar-les cal especificar-ho.

   ### Estimació de la mitjana SENSE suposar NORMALITAT.

   Si el nombre de dades n és GRAN ($n > 30$), pel teorema central del límit es poden utilitzar les fórmules suposant normalitat.

   Si el nombre de dades n és PETIT ($n < 30$), no podem aplicar els mètodes fins ara utilitzats.

   ### Estimació de la variància de poblacions NORMALS.

   Donada la mostra x, de qualsevol mida,

   (n-1)*var(x)/qchisq(alfa/2,df=n-1)
   (n-1)*var(x)/qchisq(1-alfa/2,df=n-1)

   ### Estimació de la PROPORCIÓ.

   Necessitem una mostra de mida gran, $n > 30$, $n \cdot$pobs > 5, $n(1-$pobs$) > 5$ (les proporcions observades, pobs i $1-$ pobs, no molt petites).

   pobs**+c(-1,1)*qnorm(1-alfa/2)*sqrt((**pobs***(1-**pobs**))/n)**

   prop.test(x,n) Instrucció directa, on x=nombre d'èxits, n=tamany mostra.

   Per defecte, correct=TRUE, conf.level=0.95, alternative='two.sided' (en els contrastos d'hpòtesi).

   Si no es volen les opcions per defecte, cal especificar-les.

2. Suposem que el temps en hores dedicat pels estudiants d'una determinada assignatura a preparar l'examen final té una distribució normal amb desviació tipus sigma=4.2. Es pren una mostra aleatòria de 11 estudiants que obtenen els resultats següents: $12.2, 18.4, 23.1, 11.7, 8.2, 12.2, 18.4, 23.1, 11.7, 8.2, 24$. Calculeu un interval de confiança del 99% per a la mitjana poblacional.

Solució

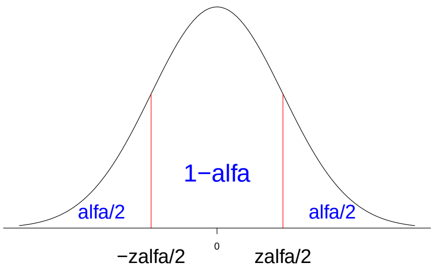
Suposem població amb distribució normal. Coneixem la variància. La mida de la mostra pot ser qualsevol, en aquest cas $n = 11$. Busquem interval de confiança per a la μ.

Un estimador puntual de la mitjana és la mitjana mostral, que per a la mostra donada val: $\overline{x} = 15.56364$.

$$\overline{X} \hookrightarrow N(\mu, \frac{\sigma}{\sqrt{n}}) \Longleftrightarrow Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \hookrightarrow N(0,1)$$

El nivell de confiança és $nc = 0.99 = 1 - \alpha$, aleshores,

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 2\Phi(z_{\alpha/2}) - 1 = nc \Longleftrightarrow \frac{nc+1}{2} = 0.995 = \Phi(z_{\alpha/2})$$



Per tant, de la taula de la normal stàndard obtenim

$$z_{\alpha/2} = 2.58 \quad (\text{ambR, qnorm}(0.995) = 2.575829)$$

Observeu que el punt $-z_{\alpha/2}$ és el simètric de $z_{\alpha/2}$. I amb la notació del subíndex de $z_{\alpha/2}$ estem indicant la probabilitat que deixa a la dreta d'aquest punt la funció densitat, $P(Z > z_{\alpha/2}) = \alpha/2$.

L'interval de confiança per a μ, amb un nivell de confiança $nc = 0.99$ és:

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = [12.30174, 18.82553]$$

És a dir, hem obtingut que, al 99% de confiança, la mitjana poblacional es pot aproximar per $\overline{x}$, amb un marge d'error $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$:

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 15.56364 \pm 3.261895$$

Recordeu que el nivell de confiança vol dir que obtenim un interval que contindrà, en un $100(1-\alpha)\%$ dels casos, el valor del paràmetre desconegut.

3. Un fabricant de vehicles sap que el consum de gasolina dels seus vehicles es distribueix normalment. Se selecciona una mostra aleatòria simple de cotxes i s'observa el consum cada cent kilòmetres obtenint les següents observacions: (19.2, 19.4, 18.4, 18.6, 20.5, 20.8).

(a) Obteniu l'interval de confiança per al consum mitjà de gasolina de tots els vehicles d'aquest fabricant, al nivell de confiança del 99%.

(b) Obteniu l'interval de confiança per a la σ^2 de tots els vehicles d'aquest fabricant, al nivell de confiança del 95%.

Solució

(a) Suposem població amb distribució normal. No coneixem la variància (farem servir $\hat{S}$ i la t d'Student). La mida de la mostra pot ser qualsevol, en aquest cas, $n = 6$. Busquem interval de confiança per a la μ.

Per a la mostra donada,

la mitjana mostral val: $\overline{x} = 19.48333$,

i la variància mostral corregida val: $\hat{S}^2 = \frac{1}{5}\left(\sum_{i=1}^{6}(x_i - \overline{x})^2\right) = 0.9616667 \Longrightarrow \hat{S} = \sqrt{0.9616667} = 0.980646$

(O fent servir la fórmula de Steiner: $\hat{S}^2 = \frac{n}{n-1}(\overline{X^2} - \overline{X}^2) = \frac{6}{5}\left(\frac{1}{6}\sum_{i=1}^{6}(x_i)^2 - (19.48333)^2\right) = 0.9616667$)

Utilitzem l'estadístic: $T = \dfrac{\overline{X} - \mu}{\dfrac{\hat{S}}{\sqrt{n}}} \hookrightarrow t_{n-1}$

$$P(-t_{\alpha/2\,,n-1} < T < t_{\alpha/2\,,n-1}) = nc = 0.99 = 1 - \alpha \Longrightarrow \frac{\alpha}{2} = 0.005$$

Donat que $n = 6$, la t de Student té 5 graus de llibertat.

Recordeu que $t_{\alpha/2\,,n-1}$ és un valor que compleix $P(T > t_{\alpha/2\,,n-1}) = \frac{\alpha}{2}$.

De la taula de la t de Student obtenim:

$$t_{\alpha/2\,,n-1} = t_{0.005\,,5} = 4.0321 \quad (\text{ambR}, \text{qt}(1 - 0.005, \text{df} = 5) = 4.032143)$$

L'interval de confiança per a μ, amb un nivell de confiança $nc = 0.99$ és:

$$\left[\overline{x} - t_{\alpha/2\,,n-1}\frac{\hat{S}}{\sqrt{n}}\,,\, \overline{x} + t_{\alpha/2\,,n-1}\frac{\hat{S}}{\sqrt{n}}\right] = [17.86908, 21.09759]$$

(b) Suposem població amb distribució normal. La mida de la mostra pot ser qualsevol, en aquest cas, $n = 6$. Ara busquem un interval de confiança per a la σ^2.

A l'apartat anterior hem calculat la variància mostral corregida: $\hat{S}^2 = 0.9616667$

Utilitzem l'estadístic: $(n-1)\dfrac{\hat{S}^2}{\sigma^2} \hookrightarrow X^2_{n-1}$ distribució chi-quadrat amb $n-1$ graus de llibertat.

$$P\left(x^2_{1-\alpha/2\,,n-1} < X^2_{n-1} < x^2_{\alpha/2\,,n-1}\right) = P\left((n-1)\frac{\hat{S}^2}{x^2_{\alpha/2\,,n-1}} < \sigma^2 < (n-1)\frac{\hat{S}^2}{x^2_{1-\alpha/2\,,n-1}}\right) = nc = 1-\alpha$$

$$nc = 0.95 = 1 - \alpha \Longrightarrow \frac{\alpha}{2} = 0.025, \quad 1 - \frac{\alpha}{2} = 0.975$$

Recordeu que la distribució chi-quadrat no és simètrica. El supíndexs $1 - \alpha/2$ i $\alpha/2$ del punt x^2, indiquen respectivament l'àrea que deixen a la dreta en la gràfica de la funció densitat:

$$P(X^2_{n-1} > x^2_{1-\alpha/2\,,n-1}) = 1 - \alpha/2 \qquad P(X^2_{n-1} > x^2_{\alpha/2\,,n-1}) = \alpha/2$$

Amb les taules de la chi-quadrat obtenim:

$x^2_{1-\alpha/2\,,n-1} = x^2_{0.975\,,5} = 0.831$ (amb R, qchisq(0.025, df=5)=0.8312116).

$x^2_{\alpha/2\,,\,n-1} = x^2_{0.025\,,\,5} = 12.832$ (amb R, qchisq(0.975, df=5)=12.8325).

L'interval de confiança per a σ^2, amb un nivell de confiança $nc = 0.95$ és:

$$\left[(n-1)\frac{\hat{S}^2}{x^2_{\alpha/2\,,\,n-1}} \quad , \quad (n-1)\frac{\hat{S}^2}{x^2_{1-\alpha/2\,,\,n-1}}\right] = [0.3747143\,,\,5.786202]$$

4. A fí d'estudiar la reacció d'una mostra de 100 pacients a un nou fàrmac s'observa que 20 d'ells han presentat determinats efectes secundaris. Trobeu un interval de confiança per a la proporció amb un nivell de confiança del 95%.

Solució

Busquem interval de confiança per a la p, proporció de pacients amb efectes secundaris.

La proporció de pacients amb efectes secundaris que ens dóna la mostra és: $\hat{p} = \dfrac{20}{100} = 0.2$

Observem que: $n = 100 > 30$, $n\hat{p} = 20 > 5$ i $n(1-\hat{p}) = 80 > 5$. Aleshores, podem aproximar per la normal,

$$\hat{p} = \frac{X}{n} \sim N\left(p, \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

i emprear l'estadístic:

$$Z = \frac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

El nivell de confiança és $nc = 0.95 = 1-\alpha$, aleshores,

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 2\Phi(z_{\alpha/2}) - 1 = nc \Longleftrightarrow \frac{nc+1}{2} = 0.975 = \Phi(z_{\alpha/2})$$

Per tant, de la taula de la normal stàndard obtenim

$$z_{\alpha/2} = 1.96 \quad \text{(ambR, qnorm(0.975)} = 1.959964)$$

L'interval de confiança per a la proporció amb un nc=0.95 és:

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = [0.1216014, 0.2783986]$$

És a dir, amb un nivell de confiança del 95%, la proporció de pacients amb efectes secundaris es troba dins de l'interval $[0.1216014, 0.2783986]$.

5. Hem obtingut la següent mostra de valors: 348.3, 378.9, 329.6, 379.3, 348.8, 367.7, 358.4, 378.2, 377.9, 341.8,348.1, 378, 359.6, 379.2, 345.8, 377.7, 350.4, 378, 373.9, 345.8,365.3, 378, 339.6, 349.3, 368.1, 367.2, 358.1, 378.4, 366.9, 361.8,345.2, 308.9, 321.6, 374.3, 368.8, 363.7, 358.2, 378.2, 373.9, 355.8. Estimeu puntualment i mitjançant intervals de confiança, a nivells de confiança 0,95 i 0.99, la proporció dels valors per sobre de 362.

Solució

Busquem interval de confiança per a la p, proporció de valors superiors a 362.

Tenim una mostra de tamany $n = 40$, dels quals 21 són superiors a 362, aleshores la proporció de valors superiors a 362 donada per la mostra és: $\hat{p} = \dfrac{21}{40} = 0.525$.

Observem que: $n = 40 > 30$, $n\hat{p} = 21 > 5$ i $n(1-\hat{p}) = 19 > 5$. Aleshores, podem aproximar per la normal i emprear l'estadístic:

$$Z = \frac{\hat{p} - p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 2\Phi(z_{\alpha/2}) - 1 = nc$

Si $\quad nc = 0.95 = 1{-}\alpha$, aleshores : $\dfrac{nc+1}{2} = 0.975 = \Phi(z_{\alpha/2}) \quad \Longleftrightarrow z_{\alpha/2} = 1.96 \quad$ (ambR, qnorm(0.975) $= 1.959964$

L'interval de confiança per a la proporció amb un nc=0.95 és:

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = [0.370245, 0.679755]$$

Si $\quad nc = 0.99 = 1{-}\alpha$, aleshores : $\dfrac{nc+1}{2} = 0.995 = \Phi(z_{\alpha/2}) \quad \Longleftrightarrow z_{\alpha/2} = 2.58 \quad$ (ambR, qnorm(0.995) $= 2.575829$

L'interval de confiança per a la proporció amb un nc=0.99 és:

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = [0.3216175, 0.7283825]$$

6. Una empresa farmacèutica està produint comprimits i com a part del control de qualitat es desitja mesurar el diàmetre dels comprimits. Se sap que el diàmetre X segueix una distribució normal amb $\sigma = 1.3$ mm. L'empresa vol una mesura del diàmetre amb un error menor que 0.1 mm i un nivell de confiança del 99% Quin tamany de la mostra s'ha d'utilitzar per conseguir l'objectiu?

Solució

Suposem població amb distribució normal. Coneixem la variància. El nivell de confiança és $nc = 0.99$.

$\Phi(z_{\alpha/2}) = \dfrac{nc+1}{2} = 0.995 \Longrightarrow z_{\alpha/2} = 2.58$

L'error d'aproximar el diàmetre dels comprimits per la mitjana mostral és $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ (semi-amplada de l'interval de confiança).

Imposem que $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} < 0.1 \Longleftrightarrow z_{\alpha/2}\dfrac{\sigma}{0.1} < \sqrt{n} \Longleftrightarrow \left(2.58\,\dfrac{1.3}{0.1}\right)^2 < n \Longleftrightarrow 1121.3 < n$.

Per tant, agafem una mostra de 1122 comprimits.

7. (Ex.15-16-Q2) Una variable física X (amb les unitats corresponents) és aleatòria i sabem que segueix una distribució normal amb σ coneguda. Per estimar el valor esperat de X, amb un nivell de confiança del 95%, s'ha pres una m.a.s. de mida $n = 9$, i s'ha obtingut l'interval: $[118.25, 123.55]$.

   (a) Calculeu la mitjana artitmètica dels valors observats de la variable.

   (b) Calculeu la variància (poblacional) de la variable X.

   (c) Si es vols augmentar el nivell de confiança al 97%, determineu l'interval de confiança resultant fent servir la mateixa mostra dels apartats anteriors.

   (d) Es vol estudiar la influència que pot tenir la mida de la mostra. Determineu (de forma aproximada), quina hauria de ser la mida de la mostra perquè la longitud de l'interval de confiança del 97% sigui aproximadament igual a la meitat de la longitud de l'interval inicial.

Solució

(a) Sabem que l'interval de confiança, $[118.25, 123.55]$, està centrat en $\overline{x}$. Per tant,

$$\overline{x} = \frac{118.25 + 123.55}{2} = 120.9$$

(b)

$$1 - \alpha = \Phi(z_{\alpha/2}) = \frac{nc + 1}{2} = \frac{0.95 + 1}{2} = 0.975 \implies z_{\alpha/2} = 1.96$$

L'interval de confiança és:

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = [118.25, 123.55]$$

Observem que el radi val

$$z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} = \frac{123.55 - 118.25}{2} = 2.65$$

D'on obtenim la variància (poblacional) fent:

$$1.96\,\frac{\sigma}{\sqrt{9}} = 2.65 \iff \sigma = 4.056122 \iff \sigma^2 = 16.45213$$

(c) $1 - \alpha = \Phi(z_{\alpha/2}) = \dfrac{nc + 1}{2} = \dfrac{0.97 + 1}{2} = 0.985 \implies z_{\alpha/2} = 2.17$
L'interval de confiança és ara:

$$\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = [117.9661, 123.8339]$$

(d) Volem que el nou interval, amb un nivell de confiança del 97% ($\implies z_{\alpha/2} = 2.17$), sigui de longitud la meitat de la de l'interval inicial. Per tant, imposem la mateixa condició sobre els radis:

$$z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} = \frac{2.65}{2}$$

$$2.17\,\frac{4.056122}{\sqrt{n}} = \frac{2.65}{2} \implies n = 44.12754$$

Donat que $n \in \mathbb{N}$, aquest valor es pot aproximar raonablement per $n = 44$

8 Contrast d'Hipòtesi

1. Descriviu els passos a seguir per fer un **contrast per a la μ, en una població normal i σ coneguda**.

   Solució

   A partir de les dades d'una mostra, es plantegem dues hipòtesis, la hipòtesi nul·la H_0 i l'alternativa H_1, sobre el valor del paràmetre desconegut. Utilitzant l'estadístic apropiat per a la mostra decidirem si acceptem o rebutgem la hipòtesi nul·la. D'entrada suposarem que H_0 és certa.

   Suposem que volem fer un contrast d'hipòtesi pel paràmetre μ, la població és normal i σ coneguda.

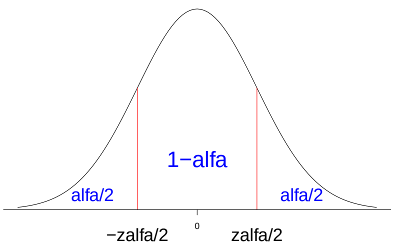
   Partim d'una mostra. Veiem quins passos cal fer:

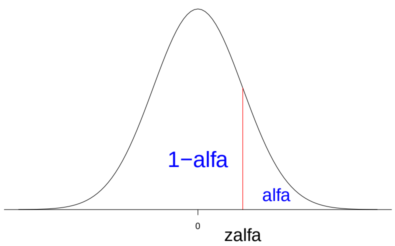
   (a) Fixem un valor del paràmetre, μ_0, i establim $\mathbf{H_0}$ i $\mathbf{H_1}$, observant el **tipus de contrast**:

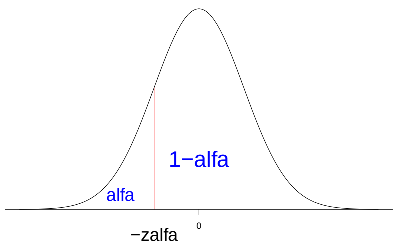
   Cas bilateral $H_0 : \mu = \mu_0$ $\quad H_1 : \mu \neq \mu_0$

   unilateral superior $H_0 : \mu = \mu_0$ $\quad H_1 : \mu > \mu_0$

   unilateral inferior $H_0 : \mu = \mu_0$ $\quad H_1 : \mu < \mu_0$







   (b) Triem l'**estadístic de contrast** adequat, amb les dades de la mostra.

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

   (c) Fixem el **nivell de significació**: α.

   (d) Busquem el **valor crític**: límit de la **regió de rebuig de** H_0. Ho fem segons el tipus de contrast:

   - cas bilateral:
     La regió de rebuig de H_0 és: $\quad (-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, +\infty) \Longrightarrow$ valor crític: $z_{\alpha/2}$

     L'obtenim de: $\Phi(z_{\alpha/2}) = 1 - \alpha/2 \quad$ (en R, $z_{\alpha/2} =$qnorm$(1 - \alpha/2)$)

   - cas unilateral superior:
     La regió de rebuig de H_0 és: $\quad (z_\alpha, +\infty) \Longrightarrow$ valor crític: z_α

     L'obtenim de: $\Phi(z_\alpha) = 1 - \alpha \quad$ (en R, $z_\alpha =$qnorm$(1 - \alpha)$)

   - cas unilateral inferior:
     La regió de rebuig de H_0 és: $\quad (-\infty, -z_\alpha) \Longrightarrow$ valor crític: $-z_\alpha$

     L'obtenim de: $\Phi(-z_\alpha) = \alpha \Longleftrightarrow \Phi(z_\alpha) = 1 - \alpha \quad$ (en R, $-z_\alpha =$-qnorm$(1 - \alpha)$)

(e) **Criteri de decisió basat en el valor crític**:
Calculem l'estadístic observat amb les dades de la mostra i suposant H_0 certa.

$$z_{obs} = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Si cau en la regió de rebuig de H_0, acceptem H_1. Altrament, mantindrem H_0!

- Cas bilateral ($H_1 : \mu \neq \mu_0$) $\quad |z_{obs}| > z_{\alpha/2} \Longrightarrow$ acceptem H_1
- Cas unilateral superior ($H_1 : \mu > \mu_0$) $\quad z_{obs} > z_\alpha \Longrightarrow$ acceptem H_1
- Cas unilateral inferior ($H_1 : \mu < \mu_0$) $\quad z_{obs} < -z_\alpha \Longrightarrow$ acceptem H_1

(f) **Criteri de decisió basat en el p-valor**:

- p-valor $< \alpha \Longrightarrow$ acceptem H_1
- p-valor $\geq \alpha \Longrightarrow$ mantenim H_0

Càlcul del p-valor (probabilitat d'obtenir els resultats de la mostra o d'altres més favorables a H_1, suposant la hipòtesi nul·la certa):

- Cas bilateral ($H_1 : \mu \neq \mu_0$): $\quad p$-valor$= P(|Z| > |z_{obs}|) = 2(1 - \Phi(|z_{obs}|))$
  (en R, p-valor$=2 * (1-\text{pnorm}(abs(z_{obs}))))$

- Unilateral superior ($H_1 : \mu > \mu_0$): $\quad p$-valor$= P(Z > z_{obs}) = 1 - \Phi(z_{obs})$
  (en R, p-valor$=1-\text{pnorm}(z_{obs}))$

- Unilateral inferior ($H_1 : \mu < \mu_0$): $\quad p$-valor$= P(Z < z_{obs}) = \Phi(z_{obs})$
  (en R, p-valor$=\text{pnorm}(z_{obs}))$

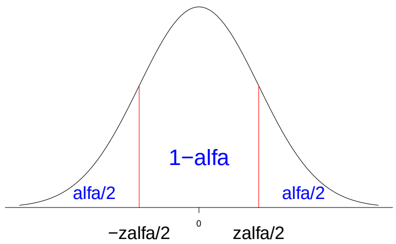
2. Un proveïdor ens proporciona un lot de peces assegurant-nos que la proporció de defectuoses és del 3%. Tenim una mostra de 1000 peces i 20 d'elles han resultat defectuoses. Realitzeu un contrast d'hipòtesi, amb nivell de significació 0.05, per decidir si es té la proporció (contrast bilateral).

Solució

$H_0 : p = 0.03$

$H_1 : p \neq 0.03$

Contrast bilateral.
Nivell de significació: $\alpha = 0.05 \Longrightarrow \alpha/2 = 0.025$
(prop.test(x=20,n=1000,p=0.03,alt="two.sided")



Observem que: $n = 1000 > 30$, $n\hat{p} = 20 > 5$ i $n(1-\hat{p}) = 980 > 5$. Per tant, podem aplicar una aproximació normal.

L'estadístic de contrast és

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

El contrast sempre es basa en suposar que la hipòtesi nul·la és certa, per tant, utilitzem p_0 i $1 - p_0$, en lloc de $\hat{p}$ i $1 - \hat{p}$. Suposant doncs que H_0 és cert, l'estadístic Z té una distribució aproximada $N(0, 1)$.

Criteri de decisió basat en el valor crític (límit de la regió de rebuig de H_0):

Nivell de significació: $\alpha = 0.05$ ($\Longrightarrow \alpha/2 = 0.025$)

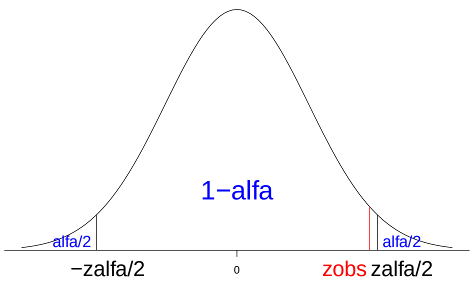
$\Phi(z_{\alpha/2}) = 1 - \alpha/2$

$\Phi(z_{0.025}) = 1 - 0.025 \Longrightarrow z_{0.025} = 1.959964$

Regió de rebuig de H_0:

$(-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, +\infty) =$

$(-\infty, -1.959964) \cup (1.959964, +\infty)$



Si $|z_{obs}| \leq z_{\alpha/2}$, mantenim H_0
ja que queda fora de la regió de rebuig.
Si $|z_{obs}| > z_{\alpha/2}$, rebutgem H_0.

$$z_{obs} = \frac{\hat{p}_{obs} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

$p_0 = 0.03 \quad \hat{p}_{obs} = 20/1000 \quad n = 1000$

$z_{obs} = -1.85376$ queda fora de la regió de rebuig

($|z_{obs}| = 1.85376 \leq z_{\alpha/2} = 1.959964$).

Per tant, no hi ha prou evidències per rebutjar H_0. Mantenim H_0 amb un nivell de significació del 5%.

Criteri de decisió basat en el p-valor:

Busquem el p-valor per determinar el grau d'encaix entre les dades i la H_0 (el p-valor és la probabilitat d'obtenir els resultats de la mostra o d'altres més favorables a H_1, suposant la hipòtesi nul·la certa):

$$P(|Z| > |z_{obs}|) = 2(1 - \Phi(|z_{obs}|)) = 2(1 - \Phi(1.85376)) = 0.0637735$$

- Si p-valor $< \alpha$ aleshores acceptem H_1
- Si p-valor $\geq \alpha$ aleshores mantenim H_0

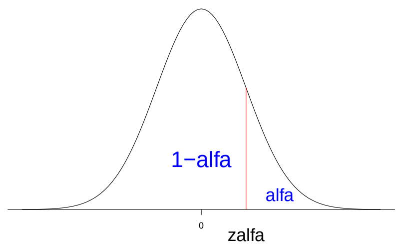
p-valor$= 0.0637735 > 0.05 = \alpha$, per tant, mantenim H_0.

3. Tenim un candidat electoral i l'objectiu és que el coneguin més del 25% dels electors. Es fa una enquesta amb 2000 persones i el nombre de persones que el coneixen és 450. Realitzeu un contrast d'hipòtesi per decidir si amb nivell de significació del 3% s'ha assolit l'objectiu.

Solució

$H_0 : p = 0.25$

$H_1 : p > 0.25$



Contrast unilateral superior.
Nivell de significació: $\alpha = 0.03$
(prop.test(x=450,n=2000,p=0.25,alt="greater",conf.level=0.97)

Observem que: $n = 2000 > 30, n\hat{p} = 450 > 5$ i $n(1-\hat{p}) = 1550 > 5$, per tant, podrem aproximar per la normal.

L'estadístic de contrast és

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

El contrast sempre es basa en suposar que la hipòtesi nul·la és certa, per tant, utilitzem p_0 i $1-p_0$, en lloc de $\hat{p}$ i $1-\hat{p}$. Suposant doncs que H_0 és cert, l'estadístic Z té una distribució aproximada $N(0,1)$.

Criteri de decisió basat en el valor crític (límit de la regió de rebuig de H_0):

Nivell de significació: $\alpha = 0.03$

$\Phi(z_\alpha) = 1 - \alpha$

$\Phi(z_{0.03}) = 1 - 0.03 \Longrightarrow z_{0.03} = 1.880794$

Regió de rebuig de H_0: $\quad (z_\alpha, +\infty) = (1.880794, +\infty)$

$$z_{obs} = \frac{\hat{p}_{obs} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

$p_0 = 0.25 \quad \hat{p}_{obs} = 450/2000 \quad n = 2000$

$z_{obs} = -2.581989$ queda fora de la regió de rebuig.

Per tant, no hi ha prou evidències per garantir que s'ha assolit l'objectiu. Mantenim H_0 amb un nivell de significació del 3%.

Criteri de decisió basat en el p-valor:

Busquem el p-valor (probabilitat d'obtenir els resultats de la mostra o d'altres més favorables a H_1, suposant la hipòtesi nul·la certa):

$$P(Z > z_{obs}) = 1 - \Phi(z_{obs}) = 1 - \Phi(-2.581989) = \Phi(2.581989) = 0.9950884$$

- Si p-valor $< \alpha$ aleshores acceptem H_1
- Si p-valor $\geq \alpha$ aleshores mantenim H_0

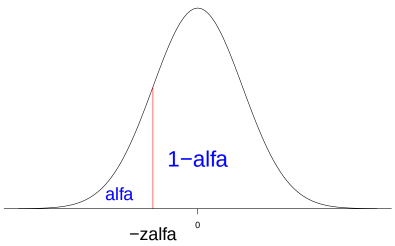
p-valor$= 0.9950884 > 0.03 = \alpha$, per tant, es manté H_0.

4. En una mostra de 2452 persones 568 d'elles van declarar que no coneixien un determinat fàrmac. Usa aquestes dades per contrastar, al 5% de nivell de significació, la hipòtesi que la proporció dels que no coneixen el fàrmac és superior o igual al 25% (hipòtesi nul·la).

Solució

$H_0 : p = 0.25$

$H_1 : p < 0.25$



Contrast unilateral inferior.
Nivell de significació: $\alpha = 0.05$
(prop.test(x=568,n=2452,p=0.25,alt="less")

Observem que: $n = 2452 > 30, n\hat{p} = 568 > 5$ i $n(1-\hat{p}) = 1884 > 5$, per tant, podrem aproximar per la normal.

L'estadístic de contrast és

$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

Recordeu que el contrast sempre es basa en suposar que la hipòtesi nul·la és certa, per tant, utilitzem p_0 i $1 - p_0$, en lloc de $\hat{p}$ i $1 - \hat{p}$. Suposant doncs que H_0 és cert, l'estadístic Z té una distribució aproximada $N(0, 1)$.

Criteri de decisió basat en el valor crític (límit de la regió de rebuig de H_0):

$\alpha = 0.05$ nivell de significació

$\Phi(-z_\alpha) = \alpha \iff \Phi(z_\alpha) = 1 - \alpha$

$$\Phi(z_{0.05}) = 0.95 \iff z_{0.05} = 1.644854$$

Regió de rebuig: $(-\infty, -z_\alpha) = (-\infty, -1.644854)$

$$z_{obs} = \frac{\hat{p}_{obs} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

$p_0 = 0.25 \quad \hat{p}_{obs} = 568/2452 \quad n = 2452$

$z_{obs} = -2.098706$ queda dins de la regió de rebuig de H_0.

La mostra ens dóna prou evidències per acceptar la hipòtesi alternativa que la proporció de persones que no coneixen un determinat fàrmac està per sota del 25%.

Per tant, acceptem H_1 amb un nivell de significació del 5%.

Criteri de decisió basat en el p-valor:

Busquem el p-valor (probabilitat d'obtenir els resultats de la mostra o d'altres més favorables a H_1, suposant la hipòtesi nul·la certa):

$$P(Z < z_{obs}) = \Phi(z_{obs}) = \Phi(-2.098706) = 1 - \Phi(2.098706) = 0.01792141$$

- Si p-valor $< \alpha$ aleshores acceptem H_1
- Si p-valor $\geq \alpha$ aleshores mantenim H_0

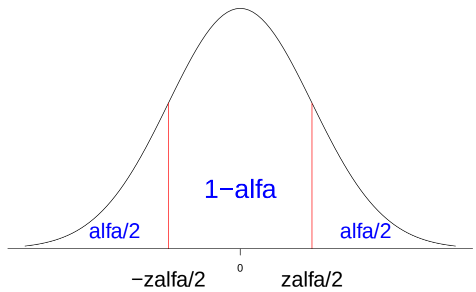
p-valor$= 0.0179214 < 0.05 = \alpha$, per tant, acceptem H_1.

5. Una màquina està ajustada de manera que la quantitat de líquid que expulsa és distribueix aproximadament segons una llei normal amb desviació típica igual a 0.15 decilitres. Una mostra de 36 expulsions ha donat un promig de 2.25 decilitres. Contrasteu amb $\alpha = 0.05$ les hipòtesis:

$H_0 : \mu_0 = 2.5$

$H_1 : \mu \neq 2.5$

Solució



Contrast bilateral.
Nivell de significació: $\alpha = 0.05 \Longrightarrow \alpha/2 = 0.025$

L'estadístic de contrast és: $Z = \dfrac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

Criteri de decisió basat en el valor crític (límit de la regió de rebuig H_0):

$\Phi(z_{\alpha/2}) = 1 - \alpha/2$

$\Phi(z_{0.025}) = 0.975 \Longrightarrow z_{\alpha/2} = 1.959964$

Regió de rebuig de H_0 (o crítica): $(-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, +\infty) = (-\infty, -1.959964) \cup (1.959964, +\infty)$

$$z_{obs} = \frac{\overline{x}_{obs} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$\mu_0 = 2.5 \quad \sigma = 0.15 \quad \overline{x}_{obs} = 2.25 \quad n = 36$

$z_{obs} = -10$ pertany a la regió de rebuig de H_0 ja que: $|z_{obs}| = |-10| = 10 > z_{\alpha/2} = 1.959964$

La mostra ens dóna prou evidències per rebutjar la hipòtesi nul·la que el promig de les expulsions és de 2.25 decilitres, amb un nivell de significació del 5%.

Per tant, acceptem H_1 amb un nivell de significació del 5%.

Criteri de decisió basat en el p-valor:

Busquem el p-valor per determinar el grau d'encaix entre les dades i la H_0 (el p-valor és la probabilitat d'obtenir els resultats de la mostra o d'altres més favorables a H_1, suposant la hipòtesi nul·la certa):

$$P(|Z| > |z_{obs}|) = 2(1 - \Phi(|z_{obs}|)) = 2(1 - 1) = 0$$

- Si p-valor $< \alpha$ aleshores acceptem H_1
- Si p-valor $\geq \alpha$ aleshores mantenim H_0

p-valor$= 0 < 0.05 = \alpha$, per tant, acceptem H_1.

6. S'està desenvolupant una màquina per tallar automàticament barres d'acer. La longitud de les barres és aproximadament normal. Es volen contrastar les hipòtesis $H_0 : \mu = 175$ contra $H_1 : \mu > 175$ amb una mostra de mida $n = 10$.

   (a) Digueu quin és l'estadístic de prova i quina la regió crítica per a $\alpha = 0.05$.

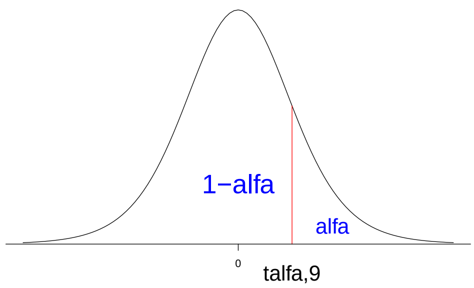
   (b) A quina conclusió es pot arribar si la mitjana mostral $\overline{x}_{obs} = 190$ mm i la desviació típica corregida $\hat{S} = 8.1$ mm?

Solució

(a)

$H_0 : \mu = 175$

$H_1 : \mu > 175$



Contrast unilateral superior.
Nivell de significació: $\alpha = 0.05$

L'estadístic de contrast és: $\quad T = \dfrac{\overline{X} - \mu_0}{\dfrac{\hat{S}}{\sqrt{n}}}$

Mida de la mostra: $n = 10 \Longrightarrow 9$ graus de llibertat.

Tenim normalitat amb σ desconeguda, i suposem la hipòtesi nul·la, H_0, certa. Per tant, $T \hookrightarrow t_{n-1}$.

$\alpha = P(T > t_{\alpha\,,n-1}) = P(T > t_{0.05\,,9})$ (en R, $t_{\alpha\,,n-1} = qt(1 - \alpha\,,\, df = n - 1)$)

Regió crítica o de rebuig de H_0: $\quad (t_{0.05\,,9}, +\infty) = (1.833113, +\infty)$

(b) $\mu_0 = 175 \quad \hat{S} = 8.1 \quad \overline{x}_{obs} = 190 \quad n = 10$

$$t_{obs} = \frac{\overline{x}_{obs} - \mu_0}{\dfrac{\hat{S}}{\sqrt{n}}}$$

$t_{obs} = 5.85607$ pertany a la regió de rebuig de H_0.

La mostra ens dóna prou evidències per acceptar, amb un nivell de significació del 5%, que la mitjana de les longituds de les barres d'acer està per sobre de 175.

Per tant, acceptem H_1 amb un nivell de significació del 5%.

Criteri de decisió basat en el p-valor:
Busquem el p-valor (probabilitat d'obtenir els resultats de la mostra o d'altres més favorables a H_1, suposant la hipòtesi nul·la certa):

$$P(T > t_{obs}) = P(T > 5.85607) = 0.0001209612$$

(en la taula de la t de Student per 9 graus de llibertat, el valor més gran que apareix és 3.2498 per al qual l'àrea que queda a la dreta en la funció densitat és de 0.005, per tant, podem dir que $P(T > 5.85607) < 0.005$. Amb R, $1 - pt(t_{obs}, df = 9) = 0.0001209612$).

- Si p-valor $< \alpha$ aleshores acceptem H_1
- Si p-valor $\geq \alpha$ aleshores mantenim H_0

p-valor$= 0.0001209612 < 0.05 = \alpha$, per tant, s'accepta H_1.

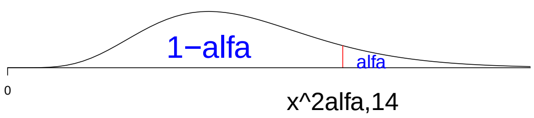
7. Un lot de cargols és acceptable, si la desviació típica de les seves longituds no supera els 0.2 mm. Per examinar un lot de 500 cargols, hem pres una mostra aleatòria de 15 cargols i hem obtingut una desviació típica corregida $\hat{S} = 0.24$ mm. Estem justificats a rebutjar la hipòtesi nul·la que la variància sigui menor o igual a 0.04, amb un nivell de significació del 0.05?

Solució

$H_0 : \sigma^2 = 0.04$

$H_1 : \sigma^2 > 0.04$

Hem de fer un contrast unilateral superior.
Nivell de significació: $\alpha = 0.05$



L'estadístic de contrast és: $X^2 = (n-1)\,\dfrac{\hat{S}^2}{\sigma_0^2}$

Mida de la mostra: $n = 15 \Longrightarrow 14$ graus de llibertat.

Criteri de decisió basat en el valor crític (límit de la regió de rebuig de H_0):

$\alpha = P(X^2 > x^2_{\alpha\,,\,n-1}) = P(X^2 > x^2_{0.05\,,\,14})$ (en R, $x^2_{\alpha\,,\,n-1} =$ qchisq$(1-\alpha, df = n-1)$).

Regió crítica o de rebuig de H_0: $(x^2_{\alpha\,,\,n-1}, +\infty) = (23.68479, +\infty)$

$$x^2_{obs} = (n-1)\,\frac{\hat{S}^2}{\sigma_0^2}$$

$\sigma_0^2 = 0.04 \quad n = 15 \quad \hat{S}^2 = (0.24)^2$
$x^2_{obs} = 20.16$ queda fora de la regió de rebuig.
Per tant, no hi ha prou evidències per rebutjar H_0. Mantenim H_0 amb un nivell de significació del 5%.

Criteri de decisió basat en el p-valor:

Busquem el p-valor (probabilitat d'obtenir els resultats de la mostra o d'altres més favorables a H_1, suposant la hipòtesi nul·la certa):

$$P(X^2 > x^2_{obs}) = P(X^2 > 20.16) = 0.1251772 \quad \text{(en R, } 1-\text{pchisq}(x^2_{obs}, df = 14)\text{)}$$

- Si p-valor $< \alpha$ aleshores acceptem H_1
- Si p-valor $\geq \alpha$ aleshores mantenim H_0

p-valor$= 0.1251772 > 0.05 = \alpha$, per tant, mantenim H_0.

9 Exàmens

Primer Parcial d'Estadística, 14/4/2016

1. **3p**

   (a) Raoneu si és certa o falsa l'afirmació:
   Si $P(A) = 0.5$, $P(B) = 0.2$ i $P(A \cup B) = 0.45$, aleshores A i B són independents.

   (b) Si A i B són independents, demostreu que A i $\overline{B}$ també ho són.

   (c) Si $f(x) = ke^{-x}$ per a $x \geq 11$, calculeu k perquè $f(x)$ sigui una funció de densitat.

   (d) Considereu la v.a. $Y = X_1 - 2X_2$, on $X_1 \hookrightarrow N(1,5)$ i $X_2 \hookrightarrow N(1,2)$ són dues v.a. independents. Trobeu $E(Y)$ i $VAR(Y)$ i digueu quin és el model de distribució de Y.

   Solució

   (a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.2 - P(A \cap B) = 0.45 \implies P(A \cap B) = 0.5 + 0.2 - 0.45 = 0.25$

   $$P(A \cap B) = 0.25 \neq P(A)P(B) = (0.5)(0.2) = 0.1$$

   Per tant, A i B no són independents.

   (b) Sabem que A i B són independents i, per tant, $P(B/A) = P(B)$.

   $$\downarrow$$

   $$P(A \cap \overline{B}) = P(A)\, P(\overline{B}/A) = P(A)(1 - P(B/A)) = P(A)(1 - P(B)) = P(A)P(\overline{B})$$

   Hem vist $P(A \cap \overline{B}) = P(A)P(\overline{B})$, és a dir, A i $\overline{B}$ són independents.

   (c) Perquè $f(x)$ sigui una funció de densitat, cal que $\displaystyle\int_{-\infty}^{+\infty} f(x)\,\mathrm{dx} = 1$.

   $$\int_{-\infty}^{+\infty} k\,e^{-x}\,\mathrm{dx} = \int_{11}^{+\infty} k\,e^{-x}\,\mathrm{dx} \lim_{b\to+\infty} \int_{11}^{b} k\,e^{-x}\,\mathrm{dx} = \lim_{b\to+\infty} -k\,[e^{-x}]_{11}^{b} = -k(\lim_{b\to+\infty}[e^{-b} - e^{-11}]) = k\,e^{-11}$$

   $$1 = \int_{11}^{+\infty} k\,e^{-x}\,\mathrm{dx} = k\,e^{-11} \implies k = 1/(e^{-11}) = e^{11}$$

   (d) $Y = X_1 - 2X_2$, on $X_1 \hookrightarrow N(1,5)$ i $X_2 \hookrightarrow N(1,2)$ són dues v.a. independents.
   Aplicant les propietats de l'esperança,

   $$E(Y) = E(X_1 - 2X_2) = E(X_1) - 2E(X_2) = 1 - 2 = -1$$

   Per al càlcul de la variància tindrem en compte que:

   - $VAR(-2\,X_2) = (-2)^2\,VAR(X_2)$,
   - si les v.a. són independents, la variància de la suma és suma de variàncies.

   $$VAR(Y) = VAR(X_1 - 2X_2) = VAR(X_1) + VAR(-2X_2) = VAR(X_1) + (-2)^2\,VAR(X_2) = 5^2 + 4(2)^2 = 25+16$$

   Finalment, la combinació lineal de v.a. normals independents també és normal, $Y \hookrightarrow N(-1, 41)$.

2. **2p** Una empresa té dues màquines. La primera produeix el 40% de les peces i l'altra la resta. Sigui A l'esdeveniment "la peça es produïda per la primera màquina", i B "la peça es produïda per la segona màquina". La probabilitat que una peça tingui defectes és de 0.2 si és de la primera màquina, i de 0.1 si és de la segona.

(a) Quina és la probabilitat que una peça surti defectuosa?

(b) Sigui C l'esdeveniment "una mostra de 3 peces de la mateixa màquina té exactament una defectuosa". Sabem que $P(\overline{C}/A) = 0.616$ i $P(C/B) = 0.243$. Quina és la probabilitat que la mostra hagi estat produïda per la primera màquina, és a dir $P(A/C)$?

Solució

(a) Sigui D el conjunt de peces defectuoses. Aleshores,

$$P(D) = P(D \cap A) + P(D \cap B) = P(D/A)P(A) + P(D/B)P(B) = (0.2)(0.4) + (0.1)(0.6) = 0.14$$

(b) Sigui C l'esdeveniment "una mostra de 3 peces de la mateixa màquina té exactament una defectuosa". Sabem que $P(\overline{C}/A) = 0.616 \Longrightarrow P(C/A) = 1 - 0.616 = 0.384$, i $P(C/B) = 0.243$.

$$P(A/C) = \frac{P(A \cap C)}{P(C)} = \frac{P(C/A)P(A)}{P(C/A)P(A) + P(C/B)P(B)} = \frac{(0.384)(0.4)}{(0.384)(0.4) + (0.243)(0.6)} = 0.5130261$$

3. **3p** L'organització Mensa requereix per poder ser membre tenir un qüocient intel·lectual igual o superior a 130.9.

Indiqueu la v.a. i el model de distribució que feu servir a cada apartat.

(a) Admetem que el qüocient intel·lectual es distribueix com una normal amb mitjana 100 i desviació tipus 15. Calculeu la probabilitat p que un individu, escollit a l'atzar, tingui un qüocient igual o superior a 130.9 (condició que es demana pertànyer a Mensa). Arrodoniu el resultat a dos decimals.

(b) Tenim 50 individus triats aleatòriament.
Quina és la probabilitat que com a mínim hi hagi dos d'ells que puguin pertànyer a Mensa?
Quin és el nombre esperat de persones que podrien pertànyer a Mensa?

(c) Sabem que en una reunió de 100 persones de la Smart-City 25 poden pertànyer a Mensa i 75 no. Si d'aquests 100 en seleccionem 2 a l'atzar, quina és la probabilitat que cap pugui pertànyer a Mensa?

Solució

(a) $p = 1 - P(X < 130.9) = 1 - P(\frac{X - 100}{15} < \frac{130.9 - 100}{15}) = 1 - P(Z < 2.06) = 1 - 0.98 = 0.02$

(b) Considerem X la v.a. que compta el nombre de persones que poden pertànyer a Mensa.
$X \hookrightarrow B(50, p)$, amb $p = 0.02$.

$$P(X \geq 2) = 1 - (P(X = 0) + P(X = 1)) = 1 - \sum_{k=0}^{1} \binom{50}{k} p^k (1-p)^{50-k} =$$

$$1 - (\binom{50}{0} p^0 (1-p)^{50-0} + \binom{50}{1} p (1-p)^{50-1}) = 1 - ((1-p)^{50} + 50\, p(1-p)^{49}) =$$

$$1 - ((0.88)^{50} + 50\,(0.02)(0.88)^{49}) = 0.2646026$$

$$E(X) = np = 50(0.02) = 1$$

(c) Ara tenim un total de 100 persones de dos tipus: els 25 que poden pertànyer a Mensa, i la resta.
Considerem X la v.a. que compta el nombre de les que poden pertànyer a Mensa en una mostra de 2 persones.
$X \hookrightarrow HG(N, D, n)$, amb $N = 100, D = 25$ i $n = 2$.

$$P(X = 0) = \frac{\binom{25}{0}\binom{75}{2}}{\binom{100}{2}} = \frac{75\,74\,\cancel{73!}}{\cancel{73!}\,2!} \, \frac{\cancel{98!}\,2!}{100\,99\,\cancel{98!}} = 0.5606061$$

4. **2p** Es té coneixement que el nombre de persones que arriben a un servei mèdic d'urgències per dia s'ajusta a un model de Poisson.

(a) Sabem que el percentatge de dies en què arriben tres persones és la meitat del percentatge de dies en què arriben dos. Trobeu el valor del paràmetre.

(b) Quin és el temps promig en dies que transcorre entre l'arribada de dues persones al servei mèdic?

(c) Quina és la probabilitat que com a màxim arribin dos persones en 2 dies?

Solució

(a) $P(X=3) = \frac{1}{2}P(X=2) \iff e^{-\lambda}\frac{\lambda^3}{3!} = \frac{1}{2}e^{-\lambda}\frac{\lambda^2}{2!} \iff \lambda = \frac{3}{2}$

(b) Hem vist que $X \hookrightarrow P(3/2)$ és la v.a. que compta les arribades per dia.
Considerem ara T la v.a. que compta el temps transcorregut entre l'arribada d'una persona i la següent.
Aleshores T segueix un model de distribució exponencial $T \hookrightarrow \epsilon(3/2)$.
$E(T) = 2/3$ és el temps promig en dies.

(c) Recordem que $X \hookrightarrow P(\lambda)$, on la unitat de temps és un dia.
Considerem $Y \hookrightarrow P(2\lambda)$, és a dir $Y \hookrightarrow P(3)$, on ara la unitat de temps és de 2 dies.

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = e^{-3}\sum_{k=0}^{2}\frac{3^k}{k!} = e^{-3}(1+3+\frac{9}{2}) = 0.42319$$

Primer Parcial d'Estadística, 10/11/2016

1. **2p**

   (a) Si $P(A) \neq 0$, $P(B) \neq 0$ i $A \subset B$, aleshores A i B poden ser independents? Justifica la resposta.

   (b) Sigui X una v.a. de funció densitat $f(x) = \begin{cases} \dfrac{x^3}{4} & \text{si } x \in [0,2] \\ 0 & \text{si } x \notin [0,2] \end{cases}$ Calculeu $E(X)$.

   Solució

   (a) Si $A \subset B \Longrightarrow A \cap B = A$, aleshores $P(A \cap B) = P(A)$.
   D'altra banda, A i B són independents si $P(A \cap B) = P(A)P(B)$.

   Per tant, tenim que $P(A) = P(A \cap B) = P(A)P(B) \underset{\substack{\uparrow \\ P(A)\neq 0,\ P(B)\neq 0,}}{\Longrightarrow} P(B) = 1 \Longrightarrow B = \Omega$.

   Hem obtingut que, en les condicions de l'enunciat, A i B són independents si $B = \Omega$.

   (b)
   $$E(X) = \int_{-\infty}^{+\infty} x\,\frac{x^3}{4}\,\mathrm{dx} = \frac{1}{4}\int_0^2 x^4\,\mathrm{dx} = \frac{1}{4}\left[\frac{x^5}{5}\right]_0^2 = \frac{1}{20}\left(2^5 - 0^5\right) = \frac{8}{5} = 1.6$$

2. **3p** Una fàbrica té dos operaris, A i B, inspeccionant la qualitat dels productes de manera que cada producte de la fàbrica és acceptat o rebutjat. La probabilitat que un producte sigui inspeccionat és de 0.32 pel primer operari i de 0.68 pel segon. El primer operari accepta el 81% dels productes que inspecciona i el segon un 90%.

   (a) Quina és la probabilitat que una peça sigui rebutjada?

   (b) Si un producte és rebutjat, quina és la probabilitat que hagi estat inspeccionat pel primer operari?

   (c) Si en un lot de 15 productes dels quals 5 són dels rebutjats. Quina és la probabilitat que escollint 3 sense reemplaçament tots tres siguin dels rebutjats?

   Solució

   (a) Sigui R el conjunt de peces rebutjades. Donat que el primer operari accepta el 81% de les peces inspeccionades, $P(R/A) = 1 - (0.81) = 0.19$, i el segon accepta el 90%, és a dir, $P(R/B) = 1 - (0.9) = 0.1$. Aleshores,

   $$P(R) = P(R/A)P(A) + P(R/B)P(B) = (0.19)(0.32) + (0.1)(0.68) = 0.1288$$

   (b)
   $$P(A/R) = \frac{P(A \cap R)}{P(R)} = \frac{P(R/A)P(A)}{P(R/A)P(A) + P(R/B)P(B)} = \frac{(0.19)(0.32)}{0.1288} = 0.4720497$$

   (c)
   $$\frac{\binom{5}{3}}{\binom{15}{3}} = \frac{5 \cdot 4 \cdot \cancel{3!}}{2!\,\cancel{3!}}\,\frac{\cancel{12!}\,3!}{15 \cdot 14 \cdot 13 \cdot \cancel{12!}} = \frac{5}{15}\,\frac{4}{14}\,\frac{3}{13} = 0.02197802$$

3. **3p** Els comprimits d'un analgèsic han de tenir per normativa entre 599 i 601 mg. de principi actiu. Acceptem que la quantitat de principi actiu es distribueix normalment amb mitjana 600 mg. i desviació típica 0.5 mg.

   Indiqueu la v.a. i el model de distribució que feu servir a cada apartat. Arrodoniu els resultats a dos decimals.

   (a) Trobeu la probabilitat que un comprimit sigui conforme, és a dir, compleixi la normativa.

   (b) Si examinem 10 comprimits, quina és la probabilitat que almenys 9 siguin conformes?

Solució

(a) Considerem X la v.a. que indica la quantitat de principi actiu en un comprimit, $X \hookrightarrow N(600, 0.5)$.

$$p = P(599 < X < 601) = P(\frac{599-600}{0.5} < \frac{X-600}{0.5} < \frac{601-600}{0.5}) = P(-2 < Z < 2) =$$

$$\Phi(2) - \Phi(-2) = \Phi(2) - (1 - \Phi(|-2|) = 2\Phi(2) - 1 = 0.9544997$$

(b) Considerem X la v.a. que compta el nombre de comprimits conformes.
$X \hookrightarrow B(10, p)$, amb $p = 0.95$ (calculada a l'apartat anterior).

$$P(X \geq 9) = (P(X = 9) + P(X = 10)) = (\binom{10}{9} p^9 (1-p) + \binom{10}{10} p^{10} (1-p)^0 =$$

$$= 10(0.95)^9 (0.05) + (0.95)^{10} = 0.9138616$$

4. **2p** Se sap que el nombre de vegades que es penja un determinat software segueix un model de distribució de Poisson de mitja 2 cops per setmana.

(a) Calculeu la probabilitat que com a mínim es pengi 1 cop en una setmana.

(b) Quina és la probabilitat que passin més de 5 setmanes sense penjar-se si ja en porta més de 2 des de l'últim cop que es va penjar?

Solució

(a) $X \hookrightarrow P(2)$, on la unitat de temps és una setmana.
$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - e^{-\lambda} \frac{\lambda^0}{0!} = 1 - e^{-2} = 0.8646647$

(b) Considerem ara T la v.a. que compta el temps transcorregut entre dues penjades del software, $T \hookrightarrow \epsilon(2)$.
Apliquem la propietat de la pèrdua de memòria del model exponencial.

$$P(T > 5 \,/\, T > 2) = P(T > 3) = 1 - P(T \leq 3) = 1 - (1 - e^{-\lambda 3}) = e^{-6} = 0.002478752$$

Primer Parcial d'Estadística, 9/11/2016

1. **2.5p**

   (a) Siguin A i B dos sucessos amb $P(A) > 0$ i $P(B) > 0$ i $A \cap B = \emptyset$. Són independents? Raona la resposta.

   (b) Calculeu k perquè $f(x)$ sigui una funció de densitat. Trobeu també la funció de distribució $F(x)$.

   $$f(x) = \begin{cases} k \cos x & \text{si } x \in [0, \pi/2] \\ 0 & \text{altrament} \end{cases}$$

   Solució

   (a) Sabem que $P(A \cap B) = P(\emptyset) = 0$.
   D'altra banda, $P(A)P(B) > 0$, ja que $P(A) > 0$ i $P(B) > 0$.

   Per tant, $P(A \cap B) \neq P(A)P(B)$, és a dir, A i B no poden ser independents.

   (b) Perquè $f(x)$ sigui una funció de densitat, cal que $\int_{-\infty}^{+\infty} f(x)\,\mathrm{dx} = 1$.

   $$1 = \int_{-\infty}^{+\infty} k \cos x\,\mathrm{dx} = \int_0^{\pi/2} k \cos x\,\mathrm{dx} = k\ [\sin x]_0^{\pi/2} = k(1-0) = k$$

   Per tant, $k = 1$.

   $$F(x) = \int_{-\infty}^{x} \cos t\,\mathrm{dt} = \int_0^x \cos t\,\mathrm{dt} = [\sin t]_0^x = \sin x$$

   $$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sin x & \text{si } x \in [0, \pi/2] \\ 1 & x > \pi/2 \end{cases}$$

2. **3p** Una planta industrial produeix un component als torns de matí i de tarda. El 50% de les unitats es produeixen al torn de matí i les altres al torn de tarda.

   La probabilitat que un component inspeccionat sigui rebutjat és del 3% al torn de matí i del 4% al de tarda.

   (a) Quina és la probabilitat que un component inspeccionat a la fàbrica, sigui rebutjat?

   (b) Suposem que un component ha estat rebutjat, quina és la probabilitat que hagi estat produït al torn de tarda?

   (c) Tenim un lot de 25 components dels quals 5 són dels rebutjats. Quina és la probabilitat que escollint 3 sense reemplaçament cap d'ells sigui dels rebutjats?

   Solució

   (a) Sigui R el conjunt de peces inspeccionades rebutjades. Aleshores,

   $$P(R) = P(R \cap M) + P(R \cap T) = P(R/M)P(M) + P(R/T)P(T) = (0.03)(0.5) + (0.04)(0.5) = 0.035$$

   (b)

   $$P(T/R) = \frac{P(T \cap R)}{P(R)} = \frac{P(R/T)P(T)}{P(R/T)P(T) + P(R/M)P(M)} = \frac{(0.04)(0.5)}{0.035} = 0.5714286$$

   (c)

   $$\frac{\binom{20}{3}}{\binom{25}{3}} = \frac{20 \cdot 19 \cdot 18 \cdot \cancel{17!}}{3!\,\cancel{17!}} \frac{\cancel{22!}\,3!}{25 \cdot 24 \cdot 23 \cdot \cancel{22!}} = \frac{20}{25}\frac{19}{24}\frac{18}{23} = 0.4956522$$

3. **3p** Les especificacions de qualitat d'un circuit elèctric indiquen que ha de tenir una resistència de entre 290 i 310 Ohms. Suposem que la resistència d'aquests circuits es distribueix normalment amb una mitjana de 300 Ohms i una desviació típica de 8 Ohms.

   Indiqueu la v.a. i el model de distribució que feu servir a cada apartat. Arrodoniu el resultat a dos decimals.

   (a) Calculeu la probabilitat que un d'aquests circuits compleixi les especificacions de qualitat.
   (b) Si examinem 5 circuits, quina és la probabilitat que almenys 4 compleixin les especificacions de qualitat?

   Solució

   (a) Sigui X la v.a. que dóna el valor de la resistència dels circuits, $X \hookrightarrow N(300, 8)$,

$$p = P(290 < X < 310) = P(\frac{290-300}{8} < \frac{X-300}{8} < \frac{310-300}{8}) = P(-1.25 < Z < 1.25) =$$

$$= \Phi(1.25) - \Phi(-1.25) = \Phi(1.25) - (1 - \Phi(|-1.25|) = 2\Phi(1.25) - 1 = 0.7887005$$

   (b) Considerem X la v.a. que compta el nombre de circuits complint les especificacions de qualitat. $X \hookrightarrow B(5, p)$, amb $p = 0.79$ (trobada a l'apartat anterior).

$$P(X \geq 4) = P(X = 4) + P(X = 5) = \binom{5}{4} p^4 (1-p)^{5-4} + \binom{5}{5} p^5 (1-p)^{5-5} = 5p^4(1-p) + p^5 = 0.7166815$$

4. **1.5p** Sigui X una v.a. amb distribució exponencial de paràmetre 2.

   Calculeu la probabilitat: $P(X > 3 \,|\, X > 2)$

   Solució

   Considerem $X \hookrightarrow \epsilon(2)$. Aplicant la propietat de la pèrdua de memòria,
   $P(X > 3 \,|\, X > 2) = P(X > 1) = 1 - P(X \leq 1) = 1 - (1 - e^{-2}) = e^{-2} = 0.1353353$

Primer Parcial 19/04/2017

1. **2.5p**

   (a) Calculeu $P(A \cup B)$ sabent que A i B són independents, $P(\bar{A}/B) = 0.4$ i $P(A \cap B) = 0.3$.

   (b) Una pregunta tipus test té n respostes possibles, de les quals exactament una és correcta. Sigui X la v.a. que indica la nota obtinguda en una pregunta al contestar a l'atzar.
   Per una resposta correcta obtenim x punts, i per una resposta errònia $-y$ punts.
   Descriviu el recorregut de la v.a. X i la funció de probabilitat.
   Calculeu el valor de y (en funció de x) per tal que la nota esperada al contestar a l'atzar sigui un zero.

   (c) Sigui X una v.a. contínua amb funció de distribució $F(x) = \begin{cases} 0 & \text{si } x < 0 \\ a + x^2 - 3 & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$
   Trobeu el valor de $a \in \mathbb{R}$ perquè $F(x)$ pugui ser una funció de distribució i calculeu la funció densitat $f(x)$.

   Solució

   (a) $P(\bar{A}/B) = 0.4 \Longrightarrow P(A/B) = 0.6$
   $0.6 = P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.3}{P(B)} \Longrightarrow P(B) = \dfrac{0.3}{0.6} = \dfrac{1}{2}$
   A més, pel fet de ser A i B independents, sabem que $P(A \cap B) = P(A)P(B)$. Per tant, $0.3 = (1/2)P(A) \Longrightarrow P(A) = 0.6$.
   Finalment $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.6 - 0.3 = 0.8$

   (b) El recorregut de X és: $X(\Omega) = \{x, -y\}$, la variable només pren dos valors $x_1 = x, x_2 = -y$.
   Calculem la funció de probabilitat $f(x)$:
   $f(x) = P(X = x) = \dfrac{1}{n}$, donat que de les n respostes exactament una és correcta,
   $f(-y) = P(X = -y) = \dfrac{n-1}{n}$.
   Calculem l'esperança i imposant que sigui zero obtindrem el valor de y:
   $E(X) = \sum_i x_i f(x_i) = x\dfrac{1}{n} + (-y)\dfrac{n-1}{n} = 0 \Longrightarrow y = \dfrac{x}{n-1}$.

   (c) Cal que $F(x)$ sigui contínua perquè sigui una funció de distribució: $0 = F(0) = a - 3 \Longrightarrow a = 3$.
   La funció de densitat l'obtenim derivant: $F'(x) = 2x$. Per tant, $f(x) = 2x$ a l'interval $[0, 1]$.

2. **2.5p** Sobre un determinat conjunt de pacients s'està experimentant amb tres tractaments diferents. Cada pacient rep un sol tractament: un 40% rep el tractament A, un 50% el B i la resta rep el C. La probabilitat de millorar és de 0.8 pel tractament A, 0.6 pel B i 0.7 pel C.

   (a) Quina és la probabilitat que un d'aquests pacients millori?

   (b) Si un pacient no millora, quina és la probabilitat que se li hagi aplicat el tractament C?

   (c) Sabem que 15 pacients han rebut el tractament C. Quina és la probabilitat que millorin tots o tots menys un d'aquests pacients?

   Solució

   (a) $P(C) = 1 - (0.4 + 0.5) = 0.1$, aleshores,

   $$P(M) = P(M/A)P(A) + P(M/B)P(B) + P(M/C)P(C) = (0.8)(0.4) + (0.6)(0.5) + (0.7)(0.1) = 0.69$$

   (b)
   $$P(C/\bar{M}) = \frac{P(C \cap \bar{M})}{P(\bar{M})} = \frac{P(\bar{M}/C)P(C)}{1 - 0.69} = \frac{(1-0.7)(0.1)}{0.31} = 0.097$$

(c) Sigui $X \hookrightarrow B(15, 0.7)$ la v.a. que compta el nombre de pacients que milloren.
Ens demanen calcular: $P(X = 15) + P(X = 14) =$

$$= \binom{15}{15}(0.7)^{15}(0.3)^0 + \binom{15}{14}(0.7)^{14}(0.3) = (0.7)^{15} + 15(0.7)^{14}(0.3) = 0.0352$$

3. **2.5p** Els comprimits d'un medicament per millorar el rendiment han de tenir per normativa entre 599 i 601 mg. de principi actiu. Acceptem que la quantitat de principi actiu es distribueix normalment amb mitjana 600 mg. i desviació típica 0.5 mg.

   Indiqueu la v.a. i el model de distribució que feu servir a cada apartat. Arrodoniu els resultats a dos decimals.

   (a) Trobeu la probabilitat que un comprimit sigui conforme, és a dir, compleixi la normativa.
   (b) Quina és la probabilitat d'haver d'examinar 10 comprimits abans de trobar un no conforme?

   Solució

   (a) Considerem X la v.a. que indica la quantitat de principi actiu en un comprimit, $X \hookrightarrow N(600, 0.5)$.

$$p = P(599 < X < 601) = P(\frac{599-600}{0.5} < \frac{X-600}{0.5} < \frac{601-600}{0.5}) = P(-2 < Z < 2) =$$

$$\Phi(2) - \Phi(-2) = \Phi(2) - (1 - \Phi(|-2|) = 2\Phi(2) - 1 = 0.9544997 \sim 0.95$$

   (b) Considerem X la v.a. que compta el nombre de comprimits conformes abans del primer no conforme. $X \hookrightarrow G(1-p)$, amb p la probabilitat de conforme calculada en l'apartat (a).

$$P(X = 10) = (0.95)^{10}(0.05) = 0.02993685 \sim 0.03$$

4. **2.5p** Se sap que el nombre de vegades que es produeix un tall d'electricitat en una població rural segueix un model de distribució de Poisson de mitja 2 cops per setmana.

   (a) Calculeu la probabilitat que es talli almenys un cop en tres setmanes.
   (b) Quina és la probabilitat que passin més de 5 setmanes sense produir-se cap tall si ja en porta més de 2 des de l'últim cop que hi va haver un?

   Solució

   (a) Sigui X la v.a. discreta que compta el nombre de talls per setmana, $X \hookrightarrow P(2)$.
   Sigui $Y = X_1 + X_2 + X_3$, la v.a. que compta el nombre de talls cada tres setmanes, $Y \hookrightarrow P(6)$.
   $P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0) = 1 - e^{-\lambda}\frac{\lambda^0}{0!} = 1 - e^{-6} = 0.9975212$

   (b) Considerem ara T la v.a. contínua que compta el temps (en setmanes) transcorregut entre dos talls d'electricitat (successos consecutius del model de Poisson de paràmetre $\lambda = 2$). Aleshores T segueix un model exponencial amb el mateix paràmetre, $T \hookrightarrow \epsilon(2)$.
   Apliquem la propietat de la pèrdua de memòria del model exponencial.

$$P(T > 5 \,/\, T > 2) = P(T > 3) = 1 - P(T \leq 3) = 1 - (1 - e^{-\lambda 3}) \overset{\lambda=2}{=} e^{-6} = 0.002478752$$

   Observació: sense aplicar la propietat de la pèrdua de memòria calcularíem la probabilitat condicionada obtenint el mateix resultat:

$$P(T > 5 \,/\, T > 2) = \frac{P((T > 5) \cap (T > 2))}{T > 2} = \frac{P(T > 5)}{P(T > 2)} = \frac{1 - (1 - e^{-\lambda 5})}{1 - (1 - e^{-\lambda 2})} = \frac{e^{-5\lambda}}{e^{-2\lambda}} = e^{-3\lambda} \overset{\lambda=2}{=} e^{-6}$$

10 Bibliografia

Pozo Montero, F. [et al.]. Probabilitat i estadística matemàtica : teoria i problemes resolts [en línea]. Barcelona: Iniciativa digital politècnica, 2010. Disponible a: <http://hdl.handle.net/2099.3/36649>. ISBN 9788476535295.

Pujol Vázquez, G.; Gibergans Bàguena, J.; García Ciaurri, F. Problemes d'estadística amb aplicació a l'enginyeria. Barcelona: UOC, 2009. ISBN 9788497887748

San Segundo, F., Marvá, M. Postdata Un curso de introducción a la Estadística, pensado para principiantes. http://www.postdata-statistics.com/IntroEstadistica/Curso/000-CursoEstadistica-color.pdf