Laborator 12 - solutii

Exercitiul 1

Un producator considera ca marfa livrata contine numai 5% produse cu defecte, dar un inspector suspecteaza ca aceasta proportie este de 10%. Trebuie sa decidem daca sa acceptam sau sa respingem marfa livrata depinzand de θ, procentul de produse cu defecte.

Inainte de a observa datele propriu-zise, sa presupunem ca ambele valori ale lui θ au probabilitate egala, $p(0.05) = p(0.10) = 0.5$.

O mostra aleatoare formata din 20 de produse are 3 produse cu defecte. Sa se calculeze distributia a posteriori a lui θ.

Solutie:

Daca am sti datele de dinainte (toate produsele), am constata urmatoarele:

- Faptul ca un produs este acceptat sau respins defineste o v.a. de tip Bernoulii(θ), unde θ ia cele 2 valori, 0.05 si 0.1 ce reprezinta opiniile producatorului, respectiv, inspectorului.
- Deci, verosimilitatile sunt distribuite astfel: $p(x = 1|\theta) = \theta$, mai exact $p(x = 1|0.05) = 0.05, p(x = 0|0.05) = 0.95, p(x = 1|0.1) = 0.1, p(x = 0|0.1) = 0.9$.

Insa nu stim in realitate setul de date. Nu stim toate produsele ca sa le putem atribui cate un numar 0 sau 1 in functie de daca are defect sau nu. Cunoastem doar cele 20 de valori din selectia respectiva. Deci, din setul de date putem cunoaste doar numarul de defecte dintr-un tuplu de 20 de produse (mai putem face astfel de selectii, dar acesta este singurul mod de a observa datele).

De asemenea, aici avem urmatoarea situatie: toate produsele sunt fie conform presupunerii producatorului, fie conform celei a inspectorului.

In momentul selectarii celor 20 de produse, ele fie respecta una din cele 2 presupuneri. Deci, numarul de produse defecte apartine fie repartitiei Binom$(20, 0.05)$, fie Binom$(20, 0, 1)$. Avem ca verosimititatile sunt definite astfel: $p(x = k|\theta) = \binom{20}{k} \cdot \theta^k \cdot (1 - \theta)^{20-k}$.

- $p(x = 3|\theta = 0.05) = \binom{20}{3} \cdot 0.05^3 \cdot (1 - 0.05)^{17} = 0.0596$.
- $p(x = 3|\theta = 0.1) = \binom{20}{3} \cdot 0.1^3 \cdot (1 - 0.1)^{17} = 0.1901$.

Calculam $p(\theta|x = 3) = \frac{p(x=3|\theta)\cdot p(\theta)}{p(x=3)}$. Distributia marginala a lui x este conform Legii Probabilitatii Totale, $p(x = 3) = p(x = 3|\theta = 0.05) \cdot p(\theta = 0.05) + p(x = 3|\theta = 0.1) \cdot p(\theta = 0.1) = 0.12485$.

Inlocuim in formula de mai sus si obtinem $p(\theta = 0.05|x = 3) = 0.2387, p(\theta = 0.05|x = 3) = 0.7613$.

Exercitiul 2

Dat fiind $Z \sim N(0,1)$, gasiti x_l, x_h asa incat $\mathbb{P}(x_l \le Z \le x_h) = 0.95$ (adica dati un interval de incredere de procent 95% pentru Z).

Solutie:

Aici, $1-\alpha = 0.95$ si notam cu ϕ cdf al lui Z.

O metoda este sa alegem x_l si x_h asa incat $\mathbb{P}(X \le x_l) = \alpha/2$ si $\mathbb{P}(X \ge x_h) = \alpha/2$. Echivalent, $x_l = \phi^{-1}(\alpha/2) = \phi^{-1}(0.025)$ si $x_2 = \phi^{-1}(1-\alpha/2) = \phi^{-1}(1-0.025)$.

Din tabelul cdf pentru $N(0,1)$, $x_l = -1.96, x_h = 1.96$.

Exercitiul 3

Fie $x_1, \dots x_n$ niste date dintr-o distributie cu varianta cunoscuta $\mathrm{Var}(X) = \sigma^2$ si medie necunoscuta $\mathbb{E}[X] = \theta$. Gasiti un interval de incredere de $1-\alpha$ pentru media θ. Presupunem ca n este suficient de mare.

Solutie:

Il estimam pe θ cu media de selectie a datelor, $\bar{X} = \frac{X_1 + \dots + X_n}{n}$. Pentru ca n este mare, aplicand Teorema Limita Centrala, v.a. $Z = \sqrt{n} \cdot \frac{\bar{X} - \theta}{\sigma}$ are distibutia $N(0,1)$. Pe noi ne intereseaza sa gasim un interval de incredere pentru Q, pe care il stim, $\mathbb{P}(-z_{\alpha/2} \le Q \le z_{\alpha/2}) = 1-\alpha$. Atunci, $\mathbb{P}\left(-z_{\alpha/2} \le \sqrt{n} \cdot \frac{\bar{X}-\theta}{\sigma} \le z_{\alpha/2}\right) = 1 - \alpha$ ceea ce este echivalent cu
$\mathbb{P}(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \theta \le \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$.

Exercitiul 4

Un inginer masoara o cantitate θ. Se presupune ca exista cate o eroare aleatoare in fiecare masuratoare deci inginerul va face n masuratori si va calcula media lor, θ. Stim ca n este suficient de mare. Notam cu x_i valoarea obtinuta la masuratoarea cu numarul i si presupunem ca $x_i = \theta + w_i$ unde w_i reprezinta eroarea. Presupunem ca w_i sunt independente, din aceeasi distributie, cu $\mathbb{E}[W] = 0, \mathrm{Var}[W] = 4$. Se calculeaza media de selectie a masuratorilor: $\bar{X} = \frac{X_1 + \dots + X_n}{n}$.

Cate masuratori trebuie sa se faca ca sa fie 90% sigur ca eroarea finala este mai mica de 0.25? Mai exact, cat ar trebui sa fie n asa incat $\mathbb{P}(\theta - 0.25 \le \bar{X} \le \theta + 0.25) \ge 0.9$?

Solutie:

Media v.a. X este $\mathbb{E}[X] = \theta + \mathbb{E}[W] = \theta$, si varianta este $\mathrm{Var}(X) = Var(W) = 4$. Vrem sa aflam n pentru care intervalul $[\bar{X} - 0.25, \bar{X} + 0.25]$ este un interval de incredere 90% pentru $\theta = \mathbb{E}[X]$.

Stim din Exercitiul anterior, ca un interval de incredere pentru $\theta = \mathbb{E}[X]$ este $[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}]$. Deci, avem $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 0.25$, unde $\sigma = 2, \alpha = 1 - 0.9 = 0.1$, $z_{\alpha/2} = z_{0.05} = \phi^{-1}(1 - 0.05) = 1.645$. Calculam si obtinem $n \geq 174$.

Exercitiul 5

Aratati ca daca $X \sim$ Poisson$(\alpha), Y \sim$ Poisson(β), X, Y independente, atunci $X + Y \sim$ Poisson$(\alpha + \beta)$;

Solutie:
$X + Y : \Omega \to \mathbb{N}$ este o variabila aleatoare discreta; Vrem sa aflam distributia acesteia deci calculam $\mathbb{P}(X + Y = k)$ pentru orice $k \in \mathbb{N}$.
$[X + Y = k] = \bigcup_{i \leq k}[X + Y = k] \cap [Y = i] = \bigcup_{i \leq k}[X = k - i] \cap [Y = i]$.
Pentru ca evenimentele $[X = k - i] \cap [Y = i], [X = k - j] \cap [Y = j]$ sunt disjuncte pentru i, j, avem ca $\mathbb{P}(X + Y = k) = \sum_{i=1}^{k} \mathbb{P}([X = k - i] \cap [Y = i])$.
Deoarece X si Y sunt independente, $\mathbb{P}([X = k - i] \cap [Y = i]) = \mathbb{P}(X = k - i) \cdot \mathbb{P}(Y = i) = e^{-\alpha} \cdot \frac{\alpha^{k-i}}{(k-i)!} \cdot e^{-\beta} \cdot \frac{\beta^i}{i!} = e^{-(\alpha-\beta)} \cdot \frac{\alpha^{k-i} . \beta^i}{(k-i)! \cdot i!} = \frac{e^{-(\alpha-\beta)}}{k!} \cdot \binom{k}{i} \cdot \alpha^{k-i} \cdot \beta^i$.
Deci, $\mathbb{P}(X + Y = k) = \frac{e^{-(\alpha-\beta)}}{k!} \cdot (\alpha + \beta)^k$, asa ca $X + Y \sim$ Poisson$(\alpha + \beta)$.

Exercitiul 6

Numarul de e-mailuri pe care il primesc intr-o zi de lucru este modelat de o v.a. Poisson de medie 0.2 e-mailuri pe minut.

1. Care este probabilitatea sa nu primesc niciun e-mail intr-o fereastra de timp de lungime de 5 minute?

2. Care este probabilitatea sa primesc mai mult de 3 e-mail uri intr-un interval de lungime 10 minute?

Solutie:
Numarul de e-mailuri primite intr-un minut este distribuit conform $f_\lambda(k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}$. Media unei astfel de v.a. este λ, deci $\lambda = 0.2$.

- Numarul de e-mail-uri primite intr-un interval de 5 minute este $X_1 + \cdots + X_5$, unde X_i sunt v.a. Poisson de parametru λ, independente. Calculam $\mathbb{P}(X_1 + \cdots + X_5 = 0) = \mathbb{P}(X_1 = 0, \ldots X_5 = 0) = \mathbb{P}(X_1 = 0) \cdot \cdots \cdot \mathbb{P}(X_5 = 0) = \left(\frac{\lambda^0 \cdot e^{-\lambda}}{0!}\right)^5 = e^{-1}$.

- Numarul de e-mail-uri primite intr-un interval de 10 minute este $X_1 + \cdots + X_10$ unde X_i sunt v.a. Poisson de parametru λ, independente. Calculam $\mathbb{P}(X_1 + \cdots + X_{10} > 3) =$

$1 - \mathbb{P}(X_1 + \cdots + X_{10} \leq 3)$ (folosim tactica aceasta cand avem de-a face cu o v.a cu valori naturale, asa ca "$X \leq x$" poate fi calculat cu pmf, $\mathbb{P}(X \leq x) = \sum_{i=0}^{x} \mathbb{P}(X = i)$). Deci, $\mathbb{P}(X_1 + \cdots + X_{10} > 3) = 1 - (\mathbb{P}(X_1 + \cdots + X_{10} = 0) + \mathbb{P}(X_1 + \cdots + X_{10} = 1) + \mathbb{P}(X_1 + \cdots + X_{10} = 2) + \mathbb{P}(X_1 + \cdots + X_{10} = 3))$.

Observam ca fiecare dintre cele 3 valori poate fi calculata efectiv, insa calculul este unul ineficient. Putem folosi Exercitiul 5, pentru ca cele 10 v.a. sunt independente. Deci, $Y = X_1 + \cdots + X_{10} \sim \text{Poisson}(\lambda_Y = 10 \cdot 0.2 = 2)$. Astfel,
$\mathbb{P}(Y \leq 3) = \mathbb{P}(Y = 0) + \mathbb{P}(Y = 1) + \mathbb{P}(Y = 2) + \mathbb{P}(Y = 3) = \frac{\lambda_Y^0 \cdot e^{-\lambda_Y}}{0!} + \frac{\lambda_Y^1 \cdot e^{-\lambda_Y}}{1!} + \frac{\lambda_Y^2 \cdot e^{-\lambda_Y}}{2!} + \frac{\lambda_Y^k \cdot e^{-\lambda_Y}}{3!} = \frac{2^0 \cdot e^{-2}}{0!} + \frac{2^1 \cdot e^{-2}}{1!} + \frac{2^2 \cdot e^{-2}}{2!} + \frac{2^3 \cdot e^{-2}}{3!}$.

Exercitiul 7

Fie $X \sim N(-5, 4)$. Calculati:

- $\mathbb{P}(X < 0)$.
- $\mathbb{P}(-7 < X < -3)$.
- $\mathbb{P}(X > -3 | X > -5)$.

Solutie:

Ne intereseaza sa examinam cdf F_X a lui X in diverse puncte. Stim cdf a lui $Z \sim N(0,1)$ (tabel) asa ca il transformam pe X intr-o $N(0,1)$ folosind proprietatile repartitiilor normale. Daca $X \sim N(\mu, \sigma^2)$, atunci $aX + b \sim N(a\mu + b, a^2 \cdot \sigma^2)$. Calculam a, b asa incat $aX + b \sim N(0,1)$. $a\mu + b = 0, a^2 \cdot \sigma^2 = 1$, deci $a = 1/\sigma, b = -\mu/\sigma$.

$Z = 1/2 \cdot X + 5/2 \sim N(0,1)$, deci $X = 2Z - 5$.

- $\mathbb{P}(X < 0) = \mathbb{P}(2Z - 5 < 0) = \mathbb{P}(Z < 2.5) = \phi(2.5)$.
- $\mathbb{P}(-7 < X < -3) = 1 - \mathbb{P}(X < -7 \cup X > -3) = 1 - (\mathbb{P}(X < -7) + \mathbb{P}(X > -3)) = 1 - (\mathbb{P}(X < -7) + 1 - \mathbb{P}(X < -3)) = \mathbb{P}(X < -3) - \mathbb{P}(X < -7) = \phi(-3) - \phi(-7)$.
- $\mathbb{P}(X > -3 | X > -5) = \frac{\mathbb{P}(X > -3, X > -5)}{\mathbb{P}(X > -5)} = \frac{\mathbb{P}(X > -3)}{\mathbb{P}(X > -5)} = \frac{1 - \mathbb{P}(X < -3)}{1 - \mathbb{P}(X < -5)} = \frac{1 - \phi(-3)}{1 - \phi(-5)}$.