MS

(seminar 14 - S14)

1. Cererea de memorie pentru o aplicaţie, ca proporţie din memoria ce poate fi alocată de un utilizator, este o variabilă aleatoare X ce are densitatea de probabilitate

$$f(x)=\begin{cases}(\theta+1)x^{\theta}, & 0<x<1,\\ 0, & \text{în rest.}\end{cases}$$

a) Să se determine media teoretică $M(X)$ a variabilei aleatoare X şi apoi să se estimeze θ în funcţie de media de selecţie $\overline{x}$ a unei selecţii aleatoare de volum n.
b) Să se determine un estimator al parametrului θ din selecţia următoare:

$$0.2, 0.4, 0.5, 0.7, 0.8, 0.9, 0.9, 0.6, 0.6, 0.4, \longrightarrow \overline{x}=\frac{0{,}2+0{,}4+0{,}5+\dots+0{,}4}{10}=0{,}6$$

rezultată în urma rulării aplicaţiei cu diferite date de intrare.

a) $$M(X)=\int_{-\infty}^{\infty}x f(x)\,dx=\int_0^1(\theta+1)x^{\theta+1}\,dx=(\theta+1)\left.\frac{x^{\theta+2}}{\theta+2}\right|_0^1=\frac{\theta+1}{\theta+2}$$

$$\frac{\hat\theta+1}{\hat\theta+2}=\overline{x}\Rightarrow\hat\theta=\frac{2\overline{x}-1}{1-\overline{x}}$$

b) $$\overline{x}=0{,}6\Rightarrow\hat\theta=\frac{2\cdot 0{,}6-1}{1-0{,}6}=0{,}5$$

2. Pentru a estima rata sosirii λ a cererilor de acces la o bază de date s-au monitorizat intervalele de timp dintre 10 cereri consecutive şi s-au înregistrat valorile:

0,3 0,4 0,45 0,5 0,7 1,5 2 2,2 3

0.2, 0.1, 0.1, 0.05, 0.05, 0.2, 0.8, 0.5, 0.2, 0.8. $\rightarrow \overline{x}=\frac{3}{10}=0{,}3$

1 2 3 4 5 6 7 8 9 10

Care este estimatorul ratei sosirilor, $\hat\lambda$?

$$\theta=\frac{1}{\lambda}\Rightarrow\lambda=\frac{1}{\theta}=\frac{1}{0{,}3}=\frac{10}{3}=3{,}33$$

3. Fie populaţia $\mathcal{P}$ formată dintr-un tip de circuite. Caracteristica ce dorim să o investigăm prin sondaj statistic este durata de viaţă a acestor circuite, ştiind că aceasta este exponenţial distribuită, cu parametrul θ necunoscut. Măsurând timpul de viaţă (în ani) a 10 circuite, se obţin valorile:

0.8830, 1.96511, 1.9189, 4.8448, 0.9208 3.4377, 1.7162, 4.2327, 5.9435, 8.3128.

Să determinăm estimatorul de verosimilitate maximă pentru θ (adică pentru media duratei de viaţă a acestui tip de circuite).

$$f_\theta(x)=\begin{cases}0, & x<0\\ \frac{1}{\theta}\cdot e^{-\frac{x}{\theta}}, & x\geq 0\end{cases}$$

Astfel, funcţia de verosimilitate este

$$L(\theta;x_1,x_2,\dots,x_n)=\prod_{i=1}^{n}\frac{1}{\theta}e^{-x_i/\theta}=\frac{1}{\theta^n}e^{-\sum_{i=1}^{n}x_i/\theta}. \tag{14.2}$$

Funcţia logaritmică cu bază mai mare ca 1 are derivata întâi pozitivă. Notând cu $h=\ln$ şi cu $\ell(\theta)=h(L(\theta))$, avem că funcţia $\ell'(\theta)=h(L(\theta))L'(\theta)$ are acelaşi semn ca derivata lui L, deci ℓ şi L au aceleaşi puncte de extrem şi de aceeaşi natură. Pentru simplitatea calculelor, determinăm punctul de maxim absolut (dacă acesta există) pentru ℓ şi acesta va fi punct de maxim absolut şi pentru L:

$$\ell(\theta)=\ln L(\theta;x_1,x_2,\dots,x_n)=-n\ln\theta-\frac{\sum_{i=1}^{n}x_i}{\theta}. \tag{14.3}$$

Avem

$$\ell'(\theta)=-\frac{n}{\theta}+\frac{\sum_{i=1}^{n}x_i}{\theta^2}. \tag{14.4}$$

Rezolvând ecuaţia $\ell'(\theta) = 0$ în raport cu θ, obţinem punctul $\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$, care este maxim absolut pentru ℓ, deci şi pentru L. Reamintim condiţia suficientă de maxim: $l^{(n)}(x_0) < 0, n-$par, unde x_0 este punct critic, i.e. $l'(x_0) = 0$. Aici, $l''(\overline{x}) = -\frac{n}{\overline{x}^2} < 0$.

În concluzie,

$$\text{argmax} L(\theta; x_1, x_2, \ldots, x_n) = \overline{x},$$

estimatorul de verosimilitate maximă a parametrului θ a distribuţiei exponenţiale este media de selecţie. În cazul exemplului dat, estimatorul verosimilităţii maxime a mediei de viaţă a circuitelor este media selecţiei:

$$\hat{\theta} = \frac{x_1 + x_2 + \ldots + x_{10}}{10} = 3.525.$$

Observaţie: Valorile de selecţie au fost generate simulând o variabilă $X \sim Exp(\theta = 3.5)$, deci estimatorul verosimilitaţii maxime $\hat{\theta} = 3.525$ este "destul de bun".

4. Un simulator al distribuţiei Bernoulli

$$X = \begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix},$$

de parametru p necunoscut, generează stringul de biţi:

1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0.

Să se determine estimatorul verosimilitătii maxime al parametrului p pe baza eşantionului de biţi.

$$p_X(\theta, b) = P(X = b) \quad b \in \{0, 1\}$$

$$p_X(\theta, b) = \begin{cases} \theta, & b = 1 \\ 1-\theta, & b = 0 \end{cases}$$

$$L(\theta; b_1, b_2, \ldots, b_{26}) = p_X(\theta; b_1) \cdot p_X(\theta; b_2) \cdot \ldots \cdot p_X(\theta; b_{26}) =$$
$$= \theta^{12} (1-\theta)^{14}$$

$$\ell(\theta) = \ln(L(\theta)) = 12 \ln \theta + 14 \ln(1-\theta)$$

$$\ell'(\theta) = \frac{12}{\theta} + \frac{14}{1-\theta} = \frac{12(1-\theta) + 14\theta}{\text{produs}} = 0 \Rightarrow \hat{\theta} = \frac{12}{26}$$

$$\ell''(\theta) = 12\left(-\frac{1}{\theta^2}\right) + 14 \frac{1}{(1-\theta)^2}(-1) < 0 \text{ pt } \theta = \frac{12}{26}$$

$$\Rightarrow \hat{\theta} = \frac{12}{26} \text{ - estimator de verosimilitate maximă}$$

într-adevăr, $l''(\frac{12}{26}) < 0$. Se observă că $\hat{\theta}$ este egal cu numărul biţilor 1 din string supra numărul total de biţi. Acest estimator al lui p este de fapt probabilitatea intuitivă de a obţine bitul 1: numărul cazurilor favorabile din string supra numărul cazurilor posibile.

5. Variabila aleatoare X, care dă numărul de zone defecte ale unui CD de un anumit tip, are următoarea distribuţie de probabilitate: $p(x) = P(X = x)$, unde

x	$p(x)$
0	0.75
1	0.15
2	0.10

x^2	
0	0,75
1	0,15
4	0,1

$M(x^2) = 0,15 + 0,4 = 0,55$

a) Să se calculeze media şi abaterea standard a numărului de zone defecte ale CD-ului.
b) Ce distribuţie de probabilitate are media de selecţie a unui eşantion de 400 de CD-uri din tipul investigat? Care este media şi dispersia acestei distribuţii?
c) Care este probabilitatea ca media numărului de zone defecte/CD într-un lot de 400 de CD-uri să fie mai mică decât 0.3?

a) $M(x) = 0 \cdot 0,75 + 1 \cdot 0,15 + 2 \cdot 0,1 = 0,35$

$\sigma^2(x) = M(x^2) - M^2(x) = 0,55 - (0,35)^2 = 0,4275$

$\sigma(x) = \sqrt{0,4275} = 0,6538$

b) $\overline{X_{400}} \sim ApN(m, \sigma^2/400)$

$m = 0,35$

$b^2 = \frac{\sigma^2}{400} = 0,00106$

$b = 0,032$

c) $P(\overline{X_{400}} < 0,3) = F_{\overline{X}}(0,3) = \Phi\left(\frac{0,3 - m}{b}\right) = \Phi(-1,5625) = 1 - \Phi(1,5625) =$

$= 1 - 0,94 = 0,06$