Tema 14

6. Fie densitatea de probabilitate

$$f(x) = \begin{cases} 2\theta x^{2\theta-1} & \text{daca } 0 \le x \le 1 \\ 0 & \text{altfel} \end{cases},$$

unde parametru $\theta > 0$ este necunoscut.
a). Sa se determine media teoretica si apoi estimatorul parametrului θ.
b). Sa se determine estimatorul verosimilitatii maxime al parametrului θ pe baza unui esantion oarecare de volum n.

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

$= \frac{2\theta}{2\theta+1}$

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

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

b) $L(\theta; x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} 2\theta \cdot x_i^{2\theta-1} = 2^n \theta^n (x_1 \dots x_n)^{2\theta-1}$

$\ell(\theta) = n\ln 2 + n\ln\theta + (2\theta-1)\ln(x_1 \cdot x_2 \cdot \dots \cdot x_n) =$

$= n\ln 2 + n\ln\theta + [(2\theta-1)\ln x_1 + \dots + (2\theta-1)\ln(x_n)]$

$\ell'(\theta) = \frac{n}{\theta} + 2(\ln x_1 + \dots + \ln x_n) = 0$

$\frac{n}{\theta} = -2(\ln x_1 + \dots + \ln x_n)$

$\theta = \frac{n}{-2\sum_{i=1}^{n} \ln x_i}$

$\ell''(\theta) = -\frac{n}{\theta^2} < 0$ ✓

7. Un simulator al distributiei geometrice de parametru $p \in (0,1)$ genereaza sirul de numere: 2, 3, 1, 14, 1, 3, 4, 1, 11, 10. $\to \bar{x} = \frac{50}{10} = 5$

Reamintim ca daca X este o variabila distribuita geometric, atunci $f(x) = P(X = x) = pq^{x-1}, x = 1, 2, \ldots$, unde p este probabilitatea succesului, iar $q = 1 - p$.

a). Sa se determine estimatorul verosimilitatii maxime al lui p pe baza unui esantion oarecare $x_1, \ldots, x_n$ de observatii ale variablei distribuite geometric X.

b). Sa se determine numeric un estimator nedeplasat al parametrul p pe baza esantionului din problema.

$$f(x) = P(X = x) = p \cdot q^{x-1} \quad , x > 0$$

a) $$L(p; x_1, x_2, \ldots, x_n) = f(p; x_1) \cdot f(p; x_2) \cdot \ldots \cdot f(p; x_n) =$$

$$= p \cdot q^{x_1 - 1} \cdot p \cdot q^{x_2 - 1} \cdot \ldots \cdot p \cdot q^{x_n - 1} = p^n \cdot q^{-n + \sum_{i=1}^{n} x_i} =$$

$$= p^n (1-p)^{-n + \sum_{i=1}^{n} x_i}$$

$$\ell(p) = n \ln p + \left(-n + \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

$$\ell'(p) = \frac{n}{p} - \frac{-n + \sum_{i=1}^{n} x_i}{1-p} = 0$$

$$\frac{n}{p} = \frac{-n + \sum x_i}{1-p}$$

$$p\left(-n + \sum x_i\right) = n - np$$

$$p\left(\cancel{-n} + \cancel{n} + \sum x_i\right) = n \Rightarrow p = \frac{n}{\sum_{i=1}^{n} x_i}$$

$$M(X) = \frac{1}{p} \qquad \frac{1}{p} = 5 \Rightarrow p = \frac{1}{5}$$

8. Se consideră o buclă `for`:

```
for(i=1;i<=n;i++) // n>30
   {
   executa blocul B;
   }
```

Ştiind că timpul de execuţie al blocului B este o variabilă aleatoare de distribuţie de probabilitate necunoscută, având media $m = 60ms$ şi abaterea standard de $\sigma = 8ms$, iar execuţiile succesive ale blocului sunt independente, să se determine distribuţia de probabilitate a timpului de execuţie a buclei `for`. Care este probabilitatea ca timpul de execuţie al buclei în cazul $n = 50$ să fie cuprins între $0.75s$ sec şi 1 sec?

Indicaţie Notând cu T_i timpul celei de-a i-a execuţii a blocului B, timpul total de execuţie

$$T = T_1 + T_2 + \cdots + T_{50}$$

este aproximativ normal distribuit.

m, σ

$[0,75, 1]$

$n = 50$

$$T \sim ApN\left(m \cdot n,\ \delta^2 = n\sigma^2\right)$$

$$\left(60 \cdot 50 \cdot 10^{-3},\ \delta^2 = 50 \cdot 8 \cdot 10^{-3}\right)$$

$$\left(3,\ \delta^2 = 0{,}4\right)$$

$$P(T > 0{,}75,\ T < 1) = P(T > 0{,}75 \cap T < 1) = P(0{,}75 < T < 1)$$

$$= P\left(\frac{0{,}75-3}{0{,}4} < \frac{T-3}{0{,}4} < \frac{1-3}{0{,}4}\right) = P\left(-5{,}625 < \frac{T-3}{0{,}4} < -5\right) =$$

$$= \Phi(-5) - \Phi(-5{,}625) = \ldots$$

9. In primul an de operare Dropbox Romania va accepta un milion de clienti. Se estimeaza ca cererea de memorie de stocare X_i, de catre un user, i, $i = 1 \ldots 10^6$, are media $m = 1.5Gb$ si abaterea standard de $\sigma = 0.5Gb$.
Ce volum, x, de Gb trebuie asigurat, daca cu o probabilitate de $p = 0.9$, cererea totala, C, va fi de cel putin x Gb? Se va folosi $z_{0.1} = -1.28$.

$$\sim ApN\left(10^6 \cdot 1{,}5\ Gb,\ \sigma^2 = 10^6 \cdot 0{,}5^2\right)$$

$$P(C \le x) = 0{,}9$$

$$P\left(\frac{C - 10^6 \cdot 1{,}5}{10^3 \cdot 0{,}5} \le \frac{x - 10^6 \cdot 1{,}5}{500}\right) =$$

Avem probabilitatea $P(C \le x) = 0.9$. Pentru a folosi valoarea critică $z_{0.1} = -1.28$, folosim formula de standardizare:

$$P\left(\frac{C - 1.5 \cdot 10^6}{500} \le -1.28\right) = 0.9$$

Rezolvăm pentru x:

$$\frac{x - 1.5 \cdot 10^6}{500} = -1.28$$

$$x - 1.5 \cdot 10^6 = -1.28 \cdot 500$$

$$x - 1.5 \cdot 10^6 = -640$$

$$x = 1.5 \cdot 10^6 - 640$$

$$x \approx 1499360 \text{ GB}$$

10. Fie $X_1, X_2, \ldots, X_{50}$ variabile aleatoare discrete i.i.d. având distributia de probabilitate Poisson de parametrul $\lambda = 2.5$. Care este distributia de probabilitate a variabilei aleatoare medie aritmetica $\overline{X}_{50} = (X_1 + X_2 + \cdots + X_{50})/50$?

$$S \sim Pois(50 \cdot 2{,}5)$$

parametru $\lambda = 2.5$. Dorim să găsim distribuția variabilei aleatoare media aritmetică $\overline{X}_{50} = \frac{X_1+X_2+\cdots+X_{50}}{50}$.

Suma variabilelor Poisson independente are tot o distribuție Poisson, cu parametru egal cu suma parametrilor fiecărei variabile:

$$\sum_{i=1}^{50} X_i \sim Poisson(50 \cdot 2.5) = Poisson(125)$$

Media aritmetică a acestor variabile va avea o distribuție normală conform teoremei centrale a limitelor (pentru un n mare):

$$\overline{X}_{50} \approx N\left(\lambda, \frac{\lambda}{n}\right)$$

unde:

$$\mu_{\overline{X}_{50}} = \lambda = 2.5$$

$$\sigma^2_{\overline{X}_{50}} = \frac{\lambda}{n} = \frac{2.5}{50} = 0.05$$

Deci, distribuția lui $\overline{X}_{50}$ este aproximativ normală:

$$\overline{X}_{50} \sim N\left(2.5, \sqrt{0.05}\right)$$