Universitatea din Bucureşti
Facultatea de Matematică şi Informatică

Probabilităţi şi statistică
2024–2025

LABORATOR #11

EX#1 Fie seturile de date ataşate

(i) `sample_Bernoulli_1.npy`;
(ii) `sample_Bernoulli_2.npy`;
(iii) `sample_Poisson.npy`;
(iv) `sample_Exp.npy`;

Creaţi un fişier în Python® prin care, pentru fiecare set de date $(x_1, x_2, \ldots, x_n)$ de la (i)-(iv):

(a) să se afişeze într-o figură histograma datelor;
(b) folosind (a), să se sugereze o estimare pentru parametrul distribuţiei.

EX#2 Fie seturile de date ataşate

(i) `sample_Bernoulli.npy`;
(ii) `sample_Poisson.npy`;
(iii) `sample_Geom.npy`;
(iv) `sample_Exp.npy`;

Creaţi un fişier în Python® prin care, pentru fiecare set de date $(x_1, x_2, \ldots, x_n)$ de la (i)-(iv):

(a) să se afişeze într-o figură histograma datelor;
(b) să se afişeze într-o figură graficul funcţiei log-verosimilitate $\log L(x_1, x_2, \ldots x_n, \theta)$ corespunzătoare setului de date $(x_1, x_2, \ldots, x_n)$:
  (i) $\log L(x_1, x_2, \ldots x_n, \theta) = \log\theta \sum_{i=1}^n x_i + \log(1-\theta)(n - \sum_{i=1}^n x_i)$, $\theta \in (0,1)$;
  (ii) $\log L(x_1, x_2, \ldots x_n, \theta) = -n\,\theta + \log\theta \sum_{i=1}^n x_i - \sum_{i=1}^n \log(x_i!)$, $\theta \in (0,50)$;
  (iii) $\log L(x_1, x_2, \ldots x_n, \theta) = n\log\theta + \log(1-\theta)(\sum_{i=1}^n x_i - n)$, $\theta \in (0,1)$;
  (iv) $\log L(x_1, x_2, \ldots x_n, \theta) = -n\log(1/\theta) - \theta \sum_{i=1}^n x_i$, $\theta \in (0,50)$;
(c) să se determine estimarea $\hat{\theta}$ a parametrului distribuţiei care maximizează funcţia de log-verosimilate corespunzătoare asociată setului de date:
  (i) $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n x_i$;
  (ii) $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n x_i$;
  (iii) $\hat{\theta} = \frac{n}{\sum_{i=1}^n x_i}$;
  (iv) $\hat{\theta} = \frac{n}{\sum_{i=1}^n x_i}$;
(d) să se afişeze în figura de la (b) graficul punctului $(\hat{\theta}, \log L(x_1, x_2, \ldots x_n, \hat{\theta}))$.

EX#3 Fie setul de date $(x_1, x_2, \ldots, x_n)$ din fişierul ataşat `sample_Normal.npy`. Creaţi un fişier în Python® prin care

(a) să se afişeze într-o figură histograma datelor;

(b) ştiind că primul parametru al distribuţiei este $\mu = 0$, să se afişeze într-o figură graficul funcţiei log-verosimilitate

$$\log L_2(x_1, x_2, \dots x_n, \theta_2) = -n \log(\sqrt{2\pi\theta_2}) - \frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \mu)^2$$

corespunzătoare setului de date $(x_1, x_2, \dots, x_n)$, pentru $\theta_2 \in (0, 0.1)$;

(c) să se determine estimarea $\hat{\theta_2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$ pentru al doilea parametru al distribuţiei care maximizează funcţia de log-verosimilate corespunzătoare asociată setului de date;

(d) să se afişeze în figura de la (b) graficul punctului $(\hat{\theta_2}, \log L_2(x_1, x_2, \dots x_n, \hat{\theta_2}))$;

(e) ştiind că al doilea parametru al distribuţiei este $\sigma^2 = 0.01$, să se afişeze într-o figură graficul funcţiei log-verosimilitate

$$\log L_1(x_1, x_2, \dots x_n, \theta_1) = -n \log(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta_1)^2$$

corespunzătoare setului de date $(x_1, x_2, \dots, x_n)$, pentru $\theta_1 \in (-1, 1)$;

(f) să se determine estimarea $\hat{\theta_1} = \frac{1}{n} \sum_{i=1}^{n} x_i$ pentru primul parametru al distribuţiei care maximizează funcţia de log-verosimilate corespunzătoare asociată setului de date;

(g) să se afişeze în figura de la (b) graficul punctului $(\hat{\theta_1}, \log L_1(x_1, x_2, \dots x_n, \hat{\theta_1}))$.

Indicaţii Python®: `numpy`, `scipy.stats`, `matplotlib.pyplot`, `matplotlib.pyplot.hist`