Universitatea din București
Facultatea de Matematică și Informatică

CURS nr. 2

METODE MODERNE DE CALCUL ȘI SIMULARE

1. Recapitulare: noțiuni de statistică

2. Generarea numerelor aleatoare

Lect. dr. Bianca Mogoș

Conținut

Partea I

- Scopul experimentului aleator. Statistică inferențială.
- Populație țintă. Eșantion
- Model probabilist. Selecție. Repartiția selecției
- Model statistic. Statistică
- Convergența in probabilitate
- Estimator. Estimație. Consistența unui estimator

Partea a II - a

- Numere aleatoare. Variabile uniforme
- Generarea numerelor aleatoare uniform distribuite

Scopul experimentului aleator

- *Experimentul aleator* se realizează pentru *colectarea de date* necesară pentru a obține informații privind un anumit fenomen de interes.
- Pe baza datelor se emit *concluzii* care, în general, ies din sfera experimentului particular.
- Cercetătorii *generalizează concluziile experimentului* pentru clasa tuturor experimentelor similare.
- Problema acestui demers este că *nu putem garanta corectitudinea concluziilor* obținute.
- Totuși, folosind tehnici statistice, putem măsura și administra *gradul de incertitudine* al rezultatelor.

Statistică inferențială

- *Statistica inferențială* este o colecție de *metode* care permit cercetătorilor să observe o submulțime a obiectelor de interes și folosind informația obținută pe baza acestor observații să facă afirmații sau inferențe privind întreaga populație.
- Câteva dintre aceste *metode* sunt:
  - estimarea parametrilor unei populații
  - verificarea ipotezelor statistice
  - estimarea densității de repartiție.

Populație țintă

- *Populația țintă* este definită ca fiind întreaga colecție de obiecte sau indivizi despre care vrem să obținem anumite informații.
- Populația țintă trebuie *bine definită* indicând
  - ce constituie *membrii* acesteia (de ex, populația unei zone geografice, o anumită firmă care construiește componente hardware, etc.)
  - *caracteristicile* populației (de ex, starea de sănătate, numărul de defecțiuni, etc.).
- In majoritatea cazurilor este imposibil sau nerealist să observăm întreaga populație; cercetătorii măsoară numai o parte a populației țintă, denumită *eșantion*.
- Pentru a face inferențe privind întreaga populație este important ca *mulțimea eșantion* să fie *reprezentativă* relativ la întreaga populație.

Model probabilist

- Fie X o v.a. cu densitatea $f(x,\theta), x \in \mathbb{R}, \theta \in \mathbb{R}$.

Definiție

Mulțimea densităților de repartiție $f(x,\theta), \theta \in \Theta \subset \mathbb{R}$, ce depind de parametrul θ se numește *model probabilist unidimensional*.

$$\{f(x,\theta) | x \in \mathbb{R}, \theta \in \Theta\}. \tag{1}$$

- Fie $X = (X_1, X_2, \ldots, X_n)$ un vector aleator cu densitatea de repartiție

$$f\left(x_1, x_2, \ldots, x_n; \theta\right), \left(x_1, x_2, \ldots, x_n\right) \in \mathbb{R}^n, \theta \in \mathbb{R}^k.$$

Definiție

Mulțimea densităților de repartiție $f\left(x_1, x_2, \ldots, x_n; \theta\right)$ cu parametrul $\theta \in \Theta \subset \mathbb{R}^k$ se numește *model probabilist multidimensional*.

$$\{f(x_1, x_2, \ldots, x_n; \theta) | \theta \in \Theta\}. \tag{2}$$

Selecție. Repartiția selecției

Definiție

O *selecție* este o mulțime de v.a. $X_1, X_2, \ldots, X_n$ având aceeași densitate de repartiție $f(x, \theta)$.

- Deoarece selecția este o mulțime de variabile aleatoare asociate unui model probabilist, selecția trebuie să aibă o repartiție, pe care o vom numi *repartiția selecției*.

Definiție

Repartiția selecției $X_1, X_2, \ldots, X_n$ este definită ca fiind repartiția vectorului $X = (X_1, X_2, \ldots, X_n)$, notată prin $f(x_1, x_2, \ldots, x_n; \theta)$.

- Cea mai folosită formă de selecție este selecția aleatoare și este bazată pe ideea experimentului aleator.

Selecție aleatoare

Definiție

Spunem că $X_1, X_2, \ldots, X_n$ este o *selecție aleatoare* asupra v.a. X care are densitatea de repartiție $f(x;\theta)$ dacă $X_1, X_2, \ldots, X_n$ sunt v.a. independente și identic repartizate ca X.

$X_1, X_2, \ldots, X_n$ se numesc *variabile de selecție*.

- În cazul selecției aleatoare, densitatea de repartiție comună a variabilelor de selecție este

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta)$$

- O selecție aleatoare poate fi construită prin repetarea unui experiment aleator de n ori.
- Un *rezultat al selecției aleatoare* se notează prin $(x_1, x_2, \ldots, x_n)$ și mulțimea tuturor rezultatelor definesc *spațiul observațiilor* $S \equiv \mathbb{R}^n$.

Model statistic. Statistică

Definiție

Modelul probabilist

$$\{f(x;\theta), \theta \in \Theta)\}$$

împreună cu selecția $X = (X_1, X_2, \ldots, X_n)$ definesc *modelul statistic*.

Definiție

Statistica este o funcție $t_n : S \to \Theta \subset \mathbb{R}^k$ care nu conține niciun parametru necunoscut.

- Cele mai utilizate statistici sunt momentele de selecție.

Momente de selecție

- *Momentul de selecție de ordinul r*

$$m'_r(x_1, x_2, \ldots, x_n) = \frac{1}{n} \sum_{j=1}^{n} x_j^r \tag{3}$$

- *Momentul de selecție de ordin 1* – *media de selecție*

$$m'_1(x_1, x_2, \ldots, x_n) = \overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} x_j \tag{4}$$

- *Momentul centrat de selecție de ordin r*

$$m_r(x_1, x_2, \ldots, x_n) = \frac{1}{n} \sum_{j=1}^{n} (x_j - \overline{X}_n)^r. \tag{5}$$

- *Momentul centrat de selecție de ordin 2* – *dispersia de selecție*

$$m_2(x_1, x_2, \ldots, x_n) = \frac{1}{n} \sum_{j=1}^{n} (x_j - \overline{X}_n)^2. \tag{6}$$

Convergență în probabilitate

Definiție

Șirul de v.a. $(X_n)_n$ *converge în probabilitate* la v.a. X dacă

$$\lim_{n\to\infty} P\left(\{\omega \in \Omega, |X_n(\omega) - X(\omega)| < \epsilon\}\right) = 1. \tag{7}$$

Propoziție

Avem relația

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow[n\to\infty]{P} E[X] = \mu$$

Estimator. Estimație

Definiție

Se numește *estimator*, variabila aleatoare

$$t_n(X) : \Omega \to \Theta, \tag{8}$$

unde $X = (X_1, X_2, \ldots, X_n)$, Ω este spațiul de selecție; $t_n(x), x \in S$, S spațiul observațiilor, se numește *estimație*.

Definiție

Un estimator $t_n = t_n(X)$ se numește *consistent* pentru θ dacă

$$\lim_{n\to\infty} P(|t_n - \theta| < \epsilon) = 1 \tag{9}$$

și notăm $t_n \xrightarrow{P} \theta$.

Consistența unui estimator

- Consistența unui estimator reprezintă o proprietate asimptotică a estimatorului.
- Un estimator bun pentru parametrul θ trebuie să aibă o repartiție cu o valoare centrală în vecinătatea lui θ.
- Definiția următoare cere ca estimatorul să aibă o valoare centrală în vecinătatea lui θ nu numai pentru valori mari ale lui n, ci pentru orice n.

Definiție

Estimatorul t_n se numește *nedeplasat* pentru θ dacă

$$E[t_n] = \theta. \tag{10}$$

Numere aleatoare

- Majoritatea metodelor de generare a v.a. se bazează pe generarea unor numere aleatoare uniform distribuite pe intervalul (0,1).
- Datorită calculatorului avem posibilitatea de a genera foarte ușor numere aleatoare uniforme.
- Totuși, trebuie cunoscut faptul că numerele generate de calculator sunt pseudo-aleatoare, deoarece acestea sunt generate cu un algoritm determinist.

Numere aleatoare uniform distribuite

Fie $(\Omega, \mathcal{B}, P)$ un câmp de probabilitate.

Definiție

(Variabilă aleatoare uniformă)
Spunem că *variabila aleatoare continuă* $U : \Omega \to \mathbb{R}$ este *repartizată uniform* pe intervalul $[a, b]$ și scriem $U \sim \mathcal{U}(a, b)$ dacă *densitatea sa de repartiție* este de forma:

$$f(u) = \begin{cases} \dfrac{1}{b-a}, & \text{dacă } u \in [a, b] \\ 0, & \text{altfel} \end{cases} \tag{11}$$

Funcția de repartiție $F(x)$ a v.a. $U \sim \mathcal{U}(a, b)$ este definită prin

$$F(x) = \begin{cases} 0, & \text{dacă } x \leq a \\ \dfrac{x-a}{b-a}, & \text{dacă } x \in (a, b) \\ 1, & \text{dacă } x \geq b \end{cases} \tag{12}$$

Observații referitoare la repartiția uniformă și v.a. uniforme

- Probabilitatea ca v.a. uniformă $U \sim \mathcal{U}(a, b)$ să cadă într-un interval $[u_1, u_2]$, $a \leq u_1 < u_2 \leq b$ este proporțională cu lungimea intervalului:

$$P(u_1 < U \leq u_2) = \frac{u_2 - u_1}{b - a}. \tag{13}$$

- Pentru a genera numere aleatoare uniform distribuite pe un interval [a,b], și scriem $X \sim \mathcal{U}(a, b)$, pornind de la numere generate uniform pe intervalul [0,1] se poate folosi transformarea

$$X = (b - a) \cdot U + a, \tag{14}$$

unde $U \sim \mathcal{U}(0, 1)$.

Generarea numerelor uniforme în Matlab (1)

- În Matlab există funcția **rand** pentru generarea variabilelor aleatoare uniforme.

- **rand**(n), unde n este un număr natural, returnează o matrice de dimensiune $n \times n$ având ca elemente numere aleatoare uniform distribuite între 0 și 1.

- **rand**(m, n), unde m, n sunt numere naturale, returnează o matrice de dimensiune $m \times n$ având ca elemente numere aleatoare uniform distribuite între 0 și 1.

Generarea numerelor uniforme în Matlab (2)

- O secvență de numere aleatoare generată în Matlab depinde de "sămânța" sau starea generatorului. Starea este resetată la valoarea implicită în momentul pornirii Matlab-ului, astfel aceeași secvență de variabile aleatoare este generată la o nouă pornire a Matlab-ului. Acesta poate fi un avantaj în situațiile în care analistul are nevoie să reproducă rezultatele unei simulări pentru a verifica anumite concluzii.
- **rng(SD)** setează sămânța generatorului pe baza întregului nenegativ *SD*
- **rng('shuffle')** setează sămânța generatorului pe baza momentului curent de timp
- **rng('default')** resetează setările generatorului la valorile inițiale, și anume, se folosește generatorul Mersenne Twister cu sămânța 0.

Generarea numerelor uniforme în Matlab (4)

Exemplu care ilustrează utilizarea funcției **rand**.

% Generam un vector de numere aleatoare pe intervalul (0,1).
x = rand(1,1000);
% Histograma eșantionului generat in x.
[N, C] = hist(x,15);
% x: mulțimea eșantion
% 15 reprezintă numărul de dreptunghiuri ale histogramei
% N: vector conținând numărul de elemente din fiecare dintre dreptunghiuri.
% C: vector conținând centrele dreptunghiurilor
% Folosirea funcției bar pentru reprezentarea grafică a histogramei.
bar(C,N,1,'w')
title('Histograma asociata unei variabile aleatoare uniforme')
xlabel('X')
ylabel('Frecventa absoluta')

Histograma rezultată este prezentată în Figura 1.

Generarea numerelor uniforme în Matlab (5)

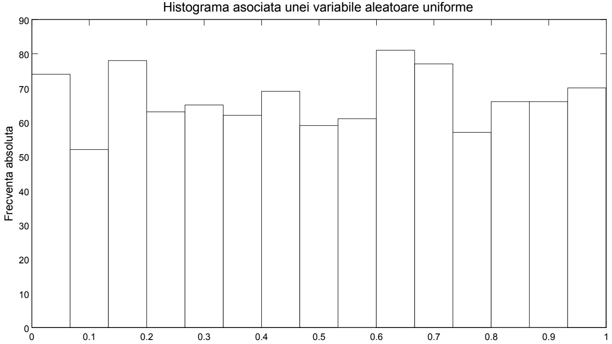

Figura : 1 – Histograma asociată unui eșantion de numere aleatoare uniform distribuite

Bibliografie I

- M. Craiu (1998), *Statistică matematică: teorie și probleme*, Editura Matrix Rom, București
- W. L. Martinez, A. R. Martinez (2002), *Computational Statistics Handbook with MATLAB*, Chapman & Hall/CRC, Boca Raton London New York Washington, D.C.
- I. Văduva (2004), *Modele de simulare: note de curs*, Editura Universității din București, București