Labor 10 - 2017

I. *Deskriptive Statistik und Schätztheorie*

Bezeichnungen (Schätzfunktionen):

▶ $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ Stichprobenmittel; empirischer Mittelwert

▶ $\tilde{S}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2$ Stichprobenvarianz; empirische Varianz

▶ $\hat{F}_n : \mathbb{R} \to [0,1]$ ist die empirische Verteilungsfunktion, bezüglich der Stichprobenvariablen $X_1, \ldots, X_n$

$$\hat{F}_n(x) = \frac{\#\{i \in \{1, ..., n\} : X_i \le x\}}{n}, x \in \mathbb{R}.$$

Eigenschaften: Seien $X_1, \ldots, X_n$ unabhängige Zufallsgrößen, welche dieselbe Verteilung haben und welche den Erwartungswert $m = E(X_n)$ und die Varianz $\sigma^2 = Var(X_n)$ haben für $\forall\ n \ge 1$. Falls

▶ $X_1, \ldots, X_n \sim N(m, \sigma^2)$

oder

▶ $n > 30$ und $X_1, \ldots, X_n$ haben eine beliebige Verteilung,

$$\Longrightarrow \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \text{ und } \frac{\bar{X}_n - m}{\frac{\tilde{S}_n}{\sqrt{n}}} \sim Student(n-1).$$

Anwendung: Gegeben sind $m \in \mathbb{R}$ und $\sigma > 0$. Man generiere n (z.B. $n = 1000$) unabhängige Zufallswerte (welche den Stichprobenvariablen $X_1, ...X_n$ entsprechen)

a) für die $N(m, \sigma^2)$ Verteilung

b) eine andere Verteilung mit Erwartungswert gleich m und Varianz gleich σ^2.

1) Man zeichne das Histogramm der relativen Häufigkeiten für M Daten (z.B. $M = 5000$) $z_1, ..., z_M$ die der Zufallsgröße $\frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$ entsprechen. Auf demselben Bild stelle man die (theoretische) Dichtefunktion der $N(0,1)$-Verteilung dar (*normpdf*).

2) Man zeichne die empirische Verteilungsfunktion welche den M generierten Daten $z_1, ..., z_M$ entspricht, d.h.

$$\hat{F}_M(z) = \frac{\#\{i \in \{1, ..., M\} : z_i \le z\}}{M}, z \in \mathbb{R}.$$

Auf demselben Bild stelle man die (theoretische) Verteilungsfunktion der $N(0,1)$-Verteilung dar (*normcdf*).

3) Man zeichne das Histogramm der relativen Häufigkeiten für M Daten (z.B. $M = 5000$) $v_1, ..., v_M$ die der Zufallsgröße $\frac{\bar{X}_n - m}{\frac{\tilde{S}_n}{\sqrt{n}}}$ entsprechen. Auf demselben Bild stelle man die Dichtefunktion der $Student(n-1)$-Verteilung dar (*tpdf*).

4) Man zeichne die empirische Verteilungsfunktion welche den M generierten Daten $v_1, ..., v_M$ entspricht, d.h.

$$\hat{F}_M(v) = \frac{\#\{i \in \{1, ..., M\} : v_i \le v\}}{M}, v \in \mathbb{R}.$$

Auf demselben Bild stelle man die (theoretische) Verteilungsfunktion der $Student(n-1)$-Verteilung dar (*tcdf*).

5) Man wende die Kenngrößen der deskriptiven Statistik für die generierten Daten (aus 1) und 3)) an und man vergleiche diese mit den theoretischen Parametern der $N(0,1)$-Verteilung, bzw. $Student(n-1)$-Verteilung.

```
mean(x) %empirischer Mittelwert; Stichprobenmittel
var(x) % Stichprobenvarianz; empirische Varianz, normalisiert mit n-1
var(x,1) %Stichprobenvarianz; empirische Varianz, normalisiert mit n
std(x) % empirische Standardabweichung
median(x) %(Lageparameter) Median= teilt die Liste der Daten in 2 H\"alften
prctile(x,[25, 75])
%(Lageparameter) unteres Quartil = 25% der Daten liegen darunter
% oberes Quartil = 75% der Daten liegen darunter
```

• theoretische Parameter der $N(0,1)$-Verteilung:
Erwartungswert: 0 , Varianz: 1, Standardabweichng: 1, Median: 0, unteres Quartil: $norminv(0.25, 0, 1)$, oberes Quartil: $norminv(0.75, 0, 1)$
• theoretische Parameter der $Student(n)$-Verteilung ($n \geq 3$) :
Erwartungswert: 0 , Varianz: $\dfrac{n}{n-2}$, Standardabweichng: $\left(\dfrac{n}{n-2}\right)^{\frac{1}{2}}$, Median: 0, unteres Quartil: $tinv(0.25, n)$, oberes Quartil: $tinv(0.75, n)$

II. Monte-Carlo Methoden für die Approximation von Integralen

Seien folgende Funktionen:
a) $g : [0, 1] \to [0, \infty)$, $g(x) = x^3$, $x \in [0, 1]$.
b) $g : [2, 5] \to [0, \infty)$, $g(x) = \frac{1}{\sqrt{x-1}}$, $x \in [2, 5]$.
c) $g : [-1, 2] \to [0, \infty)$, $g(x) = \begin{cases} \frac{1}{1+x^2}, & x \in [-1, 0] \\ \frac{1}{x^2} \sin^2 \frac{x}{1-x}, & x \in (0, 1) \\ \sqrt{2x - x^2}, & x \in [1, 2]. \end{cases}$

1) Man schreibe eine Octave/Matlab Funktion, welche den Wert $g(x)$ berechnet für x aus dem Wertebereich der Funktion g.

2) Man implementiere in Octave/Matlab jede der folgenden Methoden für die gegebenen Funktionen.

Monte-Carlo Integration I

Sei $g : [a, b] \to [0, \infty)$ eine stetige Funktion und sei $M > 0$ so dass $g(x) \leq M$, $\forall\, x \in [a, b]$. Man approximiert den Wert des Integrals $\int_a^b g(x)\, dx$, wie folgt:
• $N \in \mathbb{N}$ ist gegeben ($N = 100, 1000, \ldots$)
• $x_1, x_2, \ldots, x_N$ sind gleichmäßig verteilt auf $[a, b]$ ($unifrnd$)
• $y_1, y_2, \ldots, y_N$ sind gleichmäßig verteilt auf $[0, M]$ ($unifrnd$)
• man berechnet die Anzahl P der Paare (x_i, y_i), welche die Ungleichung $y_i \leq g(x_i)$, $i = 1, 2, \ldots, N$, erfüllen
• der approximative Wert des Integrals ist $\displaystyle\int_a^b g(x)\, dx \approx M(b-a)\frac{P}{N}$.

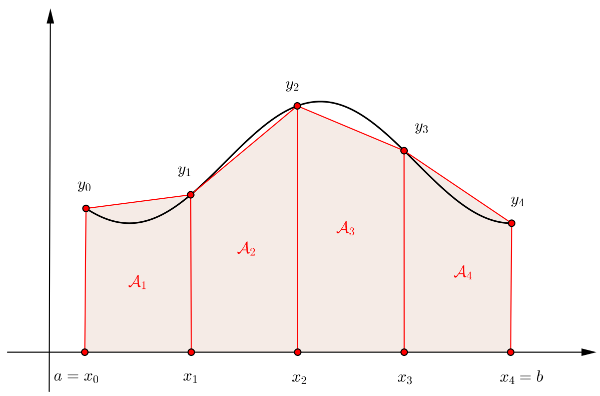
Monte-Carlo Integration II

Sei $g : [a, b] \to [0, \infty)$ eine stetige Funktion. Man approximiert den Wert des Integrals $\displaystyle\int_a^b g(x)\, dx$, wie folgt:
• $N \in \mathbb{N}$ ist gegeben ($N = 100, 1000, \ldots$)
• $x_1, x_2, \ldots, x_N$ sind gleichmäßig verteilt auf $[a, b]$ ($unifrnd$)
• man berechnet: $y_i = (b-a)g(x_i)$, $i = 1, 2, \ldots, N$
• der approximative Wert des Integrals ist $\displaystyle\int_a^b g(x)\, dx \approx \frac{1}{N}\big(y_1 + y_2 + \ldots + y_N\big)$.

Numerische Integration - Trapezregel

Sei $g : [a,b] \to [0,\infty)$ eine stetige Funktion. Man approximiert den Wert des Integrals $\int_a^b g(x)\,dx$, wie folgt:

- $N \in \mathbb{N}$ ist gegeben ($N = 100, 1000, \ldots$)
- man betrachtet eine äquidistante Unterteilung des Intervalls $[a,b]$: $a = x_1 < x_2 < \cdots < x_N < x_{N+1} = b$
- man berechnet $y_i = g(x_i)$, $i = 1, 2, \ldots, N+1$
- man berechnet die Fläche $\mathcal{A}_i$ des Trapezes gebildet mit den Punkten, welche die Koordinaten $(x_i, 0)$, $(x_{i+1}, 0)$, (x_{i+1}, y_{i+1}), (x_i, y_i) haben, $i = 1, 2, \ldots, N$.
- der approximative Wert des Integrals ist $\int_a^b g(x)\,dx \approx \mathcal{A}_1 + \mathcal{A}_2 + \ldots + \mathcal{A}_N$.




```
%%%%%%%%% Funktion fur die drei Beispiele  Skript ->  g.m
function [rez]=g(x,tipex)
switch tipex
    case 'a'
     rez=x.^3;
    case 'b'
     rez=1./sqrt(x-1);
    case 'c'
%      if x<=0
%        rez=1./(1+x.^2);
%      elseif x>=1
%        rez=sqrt(2*x-x.^2);
%      else
%         rez=(sin(x./(1-x))).^2./x.^2;
%      end

     for i=1:length(x)
         if (x(i)<=0) %&(x(i)>=-1)
            rez(i)=1/(1+x(i)^2);
         elseif (x(i)>=1)%&(x(i<=2))
            rez(i)=sqrt(2*x(i)-x(i)^2);
         else
```

```
            rez(i)=(1/x(i)^2)*(sin(x(i)/(1-x(i))))^2;
         end
     end
end
end
%%%%%%%%%%%%%%%%%
```

```
%Trapezregel
function [rez]=TrapezMatlab(N,a,b,tipex)
h=(b-a)/N;
x=a:h:b;
val=g(x,tipex);
rez=trapz(x,val); %Octave/Matlab Befehl fur die Trapezregel
```

Laboratorul 10 - 2017

I. *Elemente de statistică descriptivă şi teoria estimaţiei*

Notaţii: ▶ $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ media de selecţie

▶ $\tilde{S}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2$ varianţa de selecţie

▶ $\hat{F}_n : \mathbb{R} \to [0,1]$ este funcţia de repartiţie empirică, corespunzătoare variabilelor de selecţie $X_1, \ldots, X_n$

$$\hat{F}_n(x) = \frac{\#\{i \in \{1, ..., n\} : X_i \le x\}}{n}, x \in \mathbb{R}.$$

Proprietăţi: Fie $X_1, \ldots, X_n$ variabile aleatoare independente, care au aceeaşi distribuţie, şi au media $m = E(X_n)$ şi varianţa $\sigma^2 = Var(X_n)$ pentru orice $n \ge 1$. Dacă

▶ $X_1, \ldots, X_n \sim N(m, \sigma^2)$

sau

▶ $n > 30$ şi $X_1, \ldots, X_n$ urmează o altă distribuţie decât cea normală,

$$\Longrightarrow \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \text{ şi } \frac{\bar{X}_n - m}{\frac{\tilde{S}_n}{\sqrt{n}}} \sim Student(n-1).$$

Aplicaţie: Se dau $m \in \mathbb{R}$ şi $\sigma > 0$. Să se genereze n (=1000,...) valori aleatoare (corespunzătoare variabilelor de selecţie $X_1, ...X_n$) care urmează

a) legea $N(m, \sigma^2)$

b) o altă lege de distribuţie cu media m şi varianţa σ^2.

1) Să se deseneze histograma frecvenţelor relative pentru M date (de ex. $M = 5000$) $z_1, ..., z_M$ corespunzătoare variabilei aleatoare $\frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$. Pe aceeaşi figură să se reprezinte grafic funcţia de densitate a distribuţiei normale standard (*normpdf*).

2) Să se deseneze funcţia de repartiţie empirică corespunzătoare celor M date generate $z_1, ..., z_M$, adică funcţia

$$\hat{F}_M(z) = \frac{\#\{i \in \{1, ..., M\} : z_i \le z\}}{M}, z \in \mathbb{R}.$$

Pe aceeaşi figură să se reprezinte grafic funcţia de repartiţie a distribuţiei normale standard (*normcdf*).
3) Să se deseneze histograma frecvenţelor relative pentru M date (de ex. $M = 5000$) $v_1, ..., v_M$ corespunzătoare variabilei aleatoare $\dfrac{\bar{X}_n - m}{\frac{\tilde{S}_n}{\sqrt{n}}}$. Pe aceeaşi figură să se reprezinte grafic funcţia de densitate a distribuţiei $Student(n-1)$ (*tpdf*).
4) Să se deseneze funcţia de repartiţie empirică corespunzătoare celor M date generate $v_1, ..., v_M$, adică funcţia

$$\hat{F}_M(v) = \frac{\#\{i \in \{1, ..., M\} : v_i \leq v\}}{M}, v \in \mathbb{R}.$$

Pe aceeaşi figură să se reprezinte grafic funcţia de repartiţie a distribuţiei $Student(n-1)$ (*tcdf*).
5) Să se aplice elementele de statistică descriptivă datelor generate şi să se compare cu parametrii teoretici ai distribuţiilor considerate ($N(0,1)$, respectiv $Student(n-1)$).

```
mean(x) %media
var(x) % dispersia normalizata cu n-1
var(x,1) %dispersia normalizata cu n
std(x) %abaterea standard
median(x) %mediana
prctile(x,[25, 75]) %cuartila inferioara, cuartila superioara
```

• parameterii teoretici pentru distribuţia $N(0,1)$:
valoarea medie: 0 , varianţa: 1, abaterea standard: 1, mediana: 0, cuartila inferioară: $norminv(0.25, 0, 1)$, cuartila superioară: $norminv(0.75, 0, 1)$
• parameterii teoretici pentru distribuţia $Student(n)$ $(n \geq 3)$:
valoarea medie: 0 , varianţa: $\dfrac{n}{n-2}$, abaterea standard: $\left(\dfrac{n}{n-2}\right)^{\frac{1}{2}}$, mediana: 0, cuartila inferioară: $tinv(0.25, n)$, cuartila superioară: $tinv(0.75, n)$

II. Metode Monte-Carlo pentru estimarea unor integrale

Se consideră următoarele funcţii:
a) $g : [0,1] \to [0,\infty)$, $g(x) = x^3$, $x \in [0,1]$.
b) $g : [2,5] \to [0,\infty)$, $g(x) = \frac{1}{\sqrt{x-1}}$, $x \in [2,5]$.
c) $g : [-1,2] \to [0,\infty)$, $g(x) = \begin{cases} \frac{1}{1+x^2}, & x \in [-1,0] \\ \frac{1}{x^2}\sin^2\frac{x}{1-x}, & x \in (0,1) \\ \sqrt{2x - x^2}, & x \in [1,2]. \end{cases}$

1) Realizaţi pentru fiecare funcţie de mai sus un program în Octave/Matlab care returnează valoarea $g(x)$ pentru x dat din domeniul de definiţie al funcţiei g.

2) Implementaţi în Octave/Matlab fiecare din metodele descrise mai jos pentru funcţiile date. Comparaţi rezultatele obţinute cu metode diferite pentru aceeaşi funcţie.

Integrare Monte-Carlo - versiunea I

Fie $g : [a,b] \to [0,\infty)$ o funcţie integrabilă dată şi $M > 0$ astfel încât $g(x) \leq M$, oricare ar fi $x \in [a,b]$. Considerăm următorii paşi pentru aproximarea integralei $\displaystyle\int_a^b g(x)\,dx$:

• se generează N numere aleatoare uniform distribuite pe intervalul $[a,b]$: $x_1, x_2, \ldots, x_N \in [a,b]$, unde $N \in \mathbb{N}$ este dat ($N = 100, 1000, \ldots$) ($unifrnd$)

• se generează N numere aleatoare uniform distribuite pe intervalul $[0, M]$: $y_1, y_2, \ldots, y_N \in [0, M]$ ($unifrnd$)

• se calculează numărul P de perechi (x_i, y_i) care verifică inegalitatea: $y_i \le g(x_i)$, pentru $i = 1, 2, \ldots, N$.

• se calculează valoarea aproximativă a integralei $\displaystyle\int_a^b g(x)\,dx \approx M(b-a)\frac{P}{N}$.

Integrare Monte-Carlo - versiunea II

Fie $g : [a, b] \to [0, \infty)$ o funcţie integrabilă dată. Considerăm următorii paşi pentru aproximarea integralei $\int_a^b g(x)\,dx$:

• se generează N numere aleatoare uniform distribuite pe intervalul $[a, b]$: $x_1, x_2, \ldots, x_N \in [a, b]$, unde $N \in \mathbb{N}$ este dat $(N = 100, 1000, \ldots)$

• se notează: $y_i = (b - a)g(x_i)$, pentru $i = 1, 2, \ldots, N$.

• se calculează valoarea aproximativă a integralei $\displaystyle\int_a^b g(x)\,dx \approx= \frac{1}{N}\big(y_1 + y_2 + \ldots + y_N\big)$.

Integrare numerică - regula trapezului

Fie $g : [a, b] \to [0, \infty)$ o funcţie integrabilă dată. Considerăm următorii paşi pentru aproximarea integralei $\int_a^b g(x)\,dx$:

• se consideră o diviziune echidistantă a intervalului $[a, b]$:

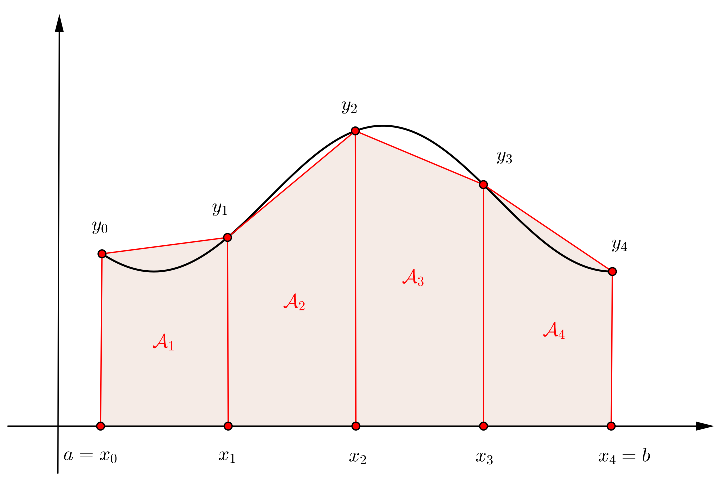
$$a = x_1 < x_2 < \cdots < x_N < x_{N+1} = b,$$

unde $N \in \mathbb{N}$ este dat $(N = 100, 1000, \ldots)$.

• se notează: $y_i = g(x_i)$, pentru $i = 1, 2, \ldots, N + 1$.

• se calculează aria $\mathcal{A}_i$ a trapezului cu vârfurile în punctele de coordonate $(x_i, 0)$, $(x_{i+1}, 0)$, (x_{i+1}, y_{i+1}) şi (x_i, y_i), pentru $i = 1, 2, \ldots, N$.

• se calculează valoarea aproximativă a integralei $\displaystyle\int_a^b g(x)\,dx \approx \mathcal{A}_1 + \mathcal{A}_2 + \ldots + \mathcal{A}_N$.




```
%%%%%%%%% Functie pt. cele trei exemple  ->  g.m
function [rez]=g(x,tipex)
```

```
switch tipex
    case 'a'
     rez=x.^3;
    case 'b'
     rez=1./sqrt(x-1);
    case 'c'
%      if x<=0
%        rez=1./(1+x.^2);
%      elseif x>=1
%        rez=sqrt(2*x-x.^2);
%      else
%         rez=(sin(x./(1-x))).^2./x.^2;
%      end

     for i=1:length(x)
         if (x(i)<=0) %&(x(i)>=-1)
            rez(i)=1/(1+x(i)^2);
         elseif (x(i)>=1)%&(x(i<=2))
            rez(i)=sqrt(2*x(i)-x(i)^2);
         else
            rez(i)=(1/x(i)^2)*(sin(x(i)/(1-x(i))))^2;
         end
     end
end
end
%%%%%%%%%%%%%%%%%


%Regula trapezului
function [rez]=TrapezMatlab(N,a,b,tipex)
h=(b-a)/N;
x=a:h:b;
val=g(x,tipex);
rez=trapz(x,val); % functie predefinita in Octave/Matlab
```