

Labor 8 - 2017

Octave/Matlab Befehle (Beispiele):

```
>> A=[1 2 2 3 4 6]
>> B=[6 6 1 3 3]
>> ismember(A,B)
>> unique(A)
>> unique(B)
>> nchoosek(['a','b','c'],2)
>> nchoosek(5,2)
>> Y=[ 2 2 5 2 2 1 3 3 5 5 2]
>> unique(Y)
>> ny=hist(Y,unique(Y)) % absolute Haufigkeiten
```

Inversionsmethode für diskrete Verteilungen

Gegeben seien x_1, \dots, x_n (die Werte), p_1, \dots, p_n (ihre Wahrscheinlichkeiten). Man generiere Zufallszahlen, die folgende diskrete Verteilung haben

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix},$$

mit Hilfe auf $[0,1]$ gleichmäßig verteilten Zufallszahlen.

Sei $U \sim \mathcal{U}[0, 1]$; z.B. $U = rand$ oder $U = unifrnd(0, 1)$

Wenn $U \leq p_1 \Rightarrow Y = x_1$;

Wenn $p_1 < U \leq p_1 + p_2 \Rightarrow Y = x_2$;

Wenn $p_1 + p_2 < U \leq p_1 + p_2 + p_3 \Rightarrow Y = x_3$;

...

Wenn $p_1 + \dots + p_k < U \leq p_1 + \dots + p_{k+1} \Rightarrow Y = x_{k+1}$

...

display(Y)

P1. Laut medizinischen Studien haben 46% der Menschen die Blutgruppe **0**, 40% haben die Blutgruppe **A**, 10% haben die Blutgruppe **B** und 4% haben die Blutgruppe **AB**. Man simuliere $N(= 100, 1000)$ zufällige Werte mit der gegebenen Verteilung "Blutgruppe". Man gebe die relative Häufigkeit jeder Blutgruppe anhand der Simulationen an.

P2. Ein Kartenspiel (mit 52 Karten) enthält 4 Asse. Man entnimmt ohne Zurücklegen 4 Karten. Sei X die Zufallsgröße die Anzahl der gezogenen Asse angibt. Man simuliere mit

a) *randsample*

b) *hygernd*

c) der oben angegebenen Methode

zufällige Werte für X . Man vergleiche die erhaltenen Ergebnisse mit den theoretischen Werten (aus der Wahrscheinlichkeitsverteilung von X).

P3. Man weiss, dass 5% der hergestellten Produkte einer bestimmten Firma fehlerhaft sind. Die Produkte werden getestet bis man das erste fehlerhafte Produkt findet. Sei X die Anzahl der getesteten Produkte **bevor** man das erste fehlerhafte Produkt findet. Anhand der Inversionsmethode simuliere man zufällige

Werte für X . Man schätze die Wahrscheinlichkeit, dass man mehr als 5 Produkte testet, bevor man das erste fehlerhafte Produkt findet. Man vergleiche die erhaltenen Ergebnisse mit den theoretischen Werten.

Hinweis für P3: $X \sim Geo(p)$, $p = 0.05$, $P(X = k) = p(1 - p)^k$, $k \in \{0, 1, 2, \dots\}$, für $U \sim \mathcal{U}[0, 1]$

$$\sum_{j=0}^k P(X = j) = \sum_{j=0}^k p(1 - p)^j = 1 - (1 - p)^{k+1}$$

Mit Hilfe der Inversionsmethode:

$$X = 0 \Leftrightarrow U \leq P(X = 0)$$

$$X = k \ (k \geq 1) \Leftrightarrow P(X = 0) + \dots + P(X = k - 1) < U \leq P(X = 0) + \dots + P(X = k - 1) + P(X = k + 1)$$

$$X = k \ (k \geq 1) \Leftrightarrow 1 - (1 - p)^k < U \leq 1 - (1 - p)^{k+1}$$

$$X = k \ (k \geq 1) \Leftrightarrow k < \frac{\ln(1 - U)}{\ln(1 - p)} \leq k + 1$$

d.h. man berechnet $X = \left\lceil \frac{\ln(1 - U)}{\ln(1 - p)} \right\rceil - 1$, also

$V \sim \mathcal{U}[0, 1] \Rightarrow X = \text{ceil}(\ln(V)/\ln(1 - p)) - 1 \sim Geo(p)$ weil $V \sim \mathcal{U}[0, 1] \Leftrightarrow 1 - V \sim \mathcal{U}[0, 1]$.

In Octave/Matlab $\log(x)$ entspricht $\ln(x)$.

P4. Mit Hilfe der Inversionsmethode schreibe man ein Programm, welches N zufällige Werte für die *Poisson Verteilung* mit Parameter $\lambda > 0$ generiert. Man schätze anhand der Simulationen den Erwartungswert einer Poisson verteilten Zufallsgröße mit Parameter λ .

$$X \sim \left(\frac{\lambda^j}{j!} e^{-\lambda} \right)_{j=0,1,2,\dots} \Rightarrow E(X) = \sum_{j=0}^{\infty} j P(X = j) = \lambda$$

I. Inversionsmethode: Sei $U \sim \mathcal{U}[0, 1]$, $P_0 = e^{-\lambda}$, $P_{j+1} = P_j \cdot \frac{\lambda}{j+1}$

$$X = k \ (k = 0, 1, 2, \dots) \Leftrightarrow k \text{ ist der kleinste Index, so dass } U \leq \sum_{j=0}^k P_j$$

II. Methode der Poissonprozesse: Seien $U_1, U_2, \dots \sim \mathcal{U}[0, 1]$

$$X = k - 1 \ (k = 1, 2, \dots) \Leftrightarrow k \text{ ist der kleinste Index, so dass } U_1 \cdot U_2 \cdot \dots \cdot U_k < e^{-\lambda}.$$

Anwendung: Die Anzahl der Kunden, welche eine Telefonzentrale innerhalb einer Stunde anrufen, ist Poisson verteilt mit Erwartungswert 15 ($\Rightarrow \lambda = 15$). Man simuliere die Anzahl der Anrufe in der Telefonzentrale und man gebe an wie oft genau k Kunden (z.B. $k = 5; 10$), in einer Stunde angerufen haben. Man vergleiche die erhaltenen Ergebnisse mit den theoretischen Werten.

P5. Mit Hilfe der Inversionsmethode schreibe man ein Programm, welches N zufällige Werte für die *gleichmäßige Verteilung* $Unif(n)$ ($n \in \mathbb{N}^*$ gegeben) generiert.

$$X \sim Unif(n) \Leftrightarrow X \sim \left(\begin{array}{cccc} 1 & 2 & \dots & n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{array} \right)$$

Sei $U \sim \mathcal{U}[0, 1]$

$$X = k \ (k = 1, 2, \dots, n) \Leftrightarrow k \text{ ist der kleinste Index, so dass } U \leq \sum_{j=0}^k \frac{1}{n} = \frac{k}{n}$$

$$X = k \ (k = 1, 2, \dots, n) \Leftrightarrow k - 1 < nU \leq k$$

Inversionsmethode $\Rightarrow X = \lceil nU \rceil$,

d.h. $U \sim \mathcal{U}[0, 1] \Rightarrow X = \text{ceil}(nU) \sim \text{Unif}(n)$.

Anwendung: Sei die Zufallsgröße $Y \sim \mathcal{U}[0, 1]$. Welche Verteilung hat $\lfloor 10 \cdot Y \rfloor$ ($\lfloor x \rfloor = \text{floor}(x)$ ganzer Teil der positiven Zahl x)?

P6. Mit Hilfe der Inversionsmethode schreibe man ein Programm, welches N zufällige Werte für die *binomiale Verteilung* $\text{Bino}(n, p)$.

$$X \sim \text{Bino}(n, p) \Leftrightarrow X \sim \begin{pmatrix} 0 & 1 & \dots & k & \dots & n \\ (1-p)^n & np(1-p)^{n-1} & \dots & C_n^k p^k (1-p)^{n-k} & \dots & p^n \end{pmatrix}.$$

Anwendung: In einem Multiple-Choice-Test gibt es 10 Fragen, bei denen man aus drei möglichen Antworten die richtige ankreuzen muss. Ein Schüler hat sich nicht auf den Test vorbereitet und hat sich entschieden die Fragen zufällig zu beantworten. Mit welcher Wahrscheinlichkeit wird er trotzdem

- a) genau 6 Fragen richtig beantworten?
- b) mindestens 6 Fragen richtig beantworten?

Laboratorul 8 - 2017

Comenzi Octave/Matlab (exemple):

```
>> A=[1 2 2 3 4 6]
>> B=[6 6 1 3 3]
>> ismember(A,B)
>> unique(A)
>> unique(B)
>> nchoosek(['a','b','c'],2)
>> nchoosek(5,2)
>> Y=[ 2 2 5 2 2 1 3 3 5 5 2]
>> unique(Y)
>> ny=hist(Y,unique(Y)) % frecvente absolute
```

Generarea de numere pseudo-aleatoare ce urmează o distribuție discretă dată (metoda inversei)

Se dau (x_1, \dots, x_n) (valorile) și (p_1, \dots, p_n) (probabilitățile lor). Realizați un program care generează N numere pseudo-aleatoare, care urmează *distribuția discretă*

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix},$$

folosind numere aleatoare uniform distribuite pe $[0,1]$.

Procedul de generare al numerelor aleatoare $Y(i)$, $i = \overline{1, N}$, este:

- Se citesc valorile x_1, x_2, \dots, x_n și probabilitățile corespunzătoare p_1, p_2, \dots, p_n , precum și numărul N . Fie $p_0 = 0$.
- Se generează N numere aleatoare uniform distribuite pe $[0, 1]$: $U(i)$, $i = \overline{1, N}$.
- Pentru fiecare $i = \overline{1, N}$: $Y(i) = x_k$ dacă și numai dacă

$$p_0 + p_1 + \dots + p_{k-1} < U(i) \leq p_0 + p_1 + \dots + p_k, \quad k \in \{1, \dots, n\}.$$

- Se returnează: $Y(i)$, $i = \overline{1, N}$.

Verificarea procedurii: deoarece U urmează legea uniformă, avem pe baza procedurii de mai sus: $P(\text{"se generează } x_k") = P(p_0 + p_1 + \dots + p_{k-1} < U \leq p_0 + p_1 + \dots + p_k) = p_k$, $k = 1, \dots, n$, deci numerele generate urmează legea de distribuție discretă dată.

P1. Conform statisticilor medicale 46% din oameni au grupa sanguină **0**, 40% au grupa sanguină **A**, 10% au grupa sanguină **B** și 4% au grupa sanguină **AB**. Simulați de $N (= 100, 1000)$ ori stabilirea grupei sanguine a unei persoane alese aleator și afișați frecvența de apariție a fiecărei grupe sanguine. Comparați rezultatele obținute cu cele teoretice.

P2. Un pachet de cărți de joc (52 de cărți) conține 4 ași. Se extrag fără reținere 4 cărți de joc. Fie X variabila aleatoare, care indică numărul de ași extrași. Folosind

a) *randsample*

b) *hygernd*

c) metoda prezentată mai sus

să se simuleze valori aleatoare ale lui X .

Comparați rezultatele obținute cu cele teoretice.

P3. Se știe că 5% din produsele unei anumite companii sunt defecte. Produsele sunt testate pe rând până la detectarea primului produs defect. Fie X variabila aleatoare, care indică numărul de produse testate până la detectarea primului produs defect. Folosind metoda prezentată mai sus, să se simuleze valori aleatoare pentru X . Care este probabilitatea de a se verifica mai mult de 5 produse, până la detectarea primului produs defect? ($P(X > 5)$)

Comparați rezultatele obținute prin simulare cu cele teoretice.

Indicație pentru P3: $X \sim Geo(p)$, $p = 0.05$, $P(X = k) = p(1 - p)^k$, $k \in \{0, 1, 2, \dots\}$, fie $U \sim \mathcal{U}[0, 1]$

$$\sum_{j=0}^k P(X = j) = \sum_{j=0}^k p(1 - p)^j = 1 - (1 - p)^{k+1}$$

Calcule pentru metoda inversei:

$$X = 0 \Leftrightarrow U \leq P(X = 0)$$

$$X = k \ (k \geq 1) \Leftrightarrow P(X = 0) + \dots + P(X = k - 1) < U \leq P(X = 0) + \dots + P(X = k - 1) + P(X = k + 1)$$

$$X = k \ (k \geq 1) \Leftrightarrow 1 - (1 - p)^k < U \leq 1 - (1 - p)^{k+1}$$

$$X = k \ (k \geq 1) \Leftrightarrow k < \frac{\ln(1 - U)}{\ln(1 - p)} \leq k + 1$$

Simularea de la metoda inversei revine la calculul valorilor $X = \left\lceil \frac{\ln(1 - U)}{\ln(1 - p)} \right\rceil - 1$, adică

$$V \sim \mathcal{U}[0, 1] \Rightarrow X = \text{ceil}(\ln(V)/\ln(1 - p)) - 1 \sim Geo(p) \text{ pentru că } V \sim \mathcal{U}[0, 1] \Leftrightarrow 1 - V \sim \mathcal{U}[0, 1].$$

În Octave/Matlab *log* corespunde logaritmului natural.

P4. Realizați un program care generează N valori aleatoare care urmează *legea Poisson* cu parametrul $\lambda > 0$, folosind metoda inversei. Estimați pe baza simulărilor valoarea medie pentru această variabilă aleatoare.

$$X \sim \left(\frac{\lambda^j}{j!} e^{-\lambda} \right)_{j=0,1,2,\dots} \Rightarrow E(X) = \sum_{j=0}^{\infty} j P(X=j) = \lambda$$

I. Metoda inversei: Fie $U \sim \mathcal{U}[0, 1]$, $P_0 = e^{-\lambda}$, $P_{j+1} = P_j \cdot \frac{\lambda}{j+1}$

$$X = k \ (k = 0, 1, 2, \dots) \Leftrightarrow k \text{ este cel mai mic indice a.î. } U \leq \sum_{j=0}^k P_j$$

II. Metoda proceselor Poisson: Fie $U_1, U_2, \dots \sim \mathcal{U}[0, 1]$

$$X = k - 1 \ (k = 1, 2, \dots) \Leftrightarrow k \text{ este cel mai mic indice a.î. } U_1 \cdot U_2 \cdot \dots \cdot U_k < e^{-\lambda}.$$

Aplicație: Numărul de clienți care sună la o centrală telefonică în decurs de o oră urmează distribuția Poisson cu media de 15 clienți pe oră ($\Rightarrow \lambda = 15$). Simulați numărul de apeluri în centrală și afișați frecvența de apariție a exact k clienți, unde $k \in \{5, 10\}$. Comparați rezultatele obținute prin simulare cu cele teoretice.

P5. Realizați un program care generează N valori aleatoare care urmează *legea uniformă discretă* $Unif(n)$ ($n \in \mathbb{N}^*$ parametru dat), folosind metoda inversei.

$$X \sim Unif(n) \Leftrightarrow X \sim \begin{pmatrix} 1 & 2 & \dots & n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

Fie $U \sim \mathcal{U}[0, 1]$

$$X = k \ (k = 1, 2, \dots, n) \Leftrightarrow k \text{ este cel mai mic indice a.î. } U \leq \sum_{j=0}^k \frac{1}{n} = \frac{k}{n}$$

$$X = k \ (k = 1, 2, \dots, n) \Leftrightarrow k - 1 < nU \leq k$$

Simularea de la metoda inversei revine la calculul valorilor $X = \lceil nU \rceil$,

adică $U \sim \mathcal{U}[0, 1] \Rightarrow X = \text{ceil}(nU) \sim Unif(n)$.

Aplicație: Se ia o variabilă aleatoare $Y \sim \mathcal{U}[0, 1]$. Ce distribuție are variabila aleatoare $\lfloor 10 \cdot Y \rfloor$ ($\lfloor x \rfloor = \text{floor}(x)$ partea întreagă a numărului pozitiv x)?

P6. Realizați un program care generează N valori aleatoare care urmează *legea binomială* $Binomial(n, p)$, folosind metoda inversei.

$$X \sim Binomial(n, p) \Leftrightarrow X \sim \begin{pmatrix} 0 & 1 & \dots & k & \dots & n \\ (1-p)^n & np(1-p)^{n-1} & \dots & C_n^k p^k (1-p)^{n-k} & \dots & p^n \end{pmatrix}.$$

Aplicație: Într-un test de tip grilă (multiple-choice) sunt formulate 10 întrebări; la fiecare întrebare sunt date trei răspunsuri posibile, dintre care doar unul este corect. Un elev nu s-a pregătit pentru test și s-a decis că va răspunde aleator la întrebări. Cu ce probabilitate va răspunde corect la

- exact 6 întrebări?
- cel puțin 6 întrebări?