

Генерисање случајних величина (1. део)

2.час

8. март 2016.

Моделовање дискретних случајних променљивих са коначно много вредности

Нека је X дискретна случајна величина са коначно много вредности и нека је познат закон расподеле:

$$P\{X = x_k\} = p_k, \quad k = 1, 2, \dots, n.$$

Нека је γ један (псеудо)случајни број.

Ако је:

- $\gamma \leq p_1$, сматра се да се реализовала вредност x_1 ,
- $p_1 < \gamma \leq p_1 + p_2$, сматра се да се реализовала вредност x_2 ,
- .
- .
- .
- $p_1 + p_2 + \dots + p_{n-1} < \gamma$, реализовала се вредност x_n .

Значи да се за добијање једне реализације случајне величине користи један (псеудо)случајан број.

Моделовање случајних догађаја

На основу поступка за моделовање дискретних случајних величина могу се моделирати и реализације случајних догађаја. Ако је вероватноћа догађаја једнака p , тада је индикатор догађаја случајна величина:

$$I_A : \begin{pmatrix} 1 & 0 \\ p & 1 - p \end{pmatrix}$$

Нека је γ један (псеудо)случајни број.

Ако је:

- $\gamma \leq p$, реализовала се вредност 1 индикатора I_A
- $p < \gamma$, реализовала се вредност 0 индикатора I_A .

Моделовањем те случајне величине I_A добија се да се догађај реализовао ако је добијена моделирана вредност 1, а да се догађај није реализовао ако је добијена моделована вредност 0.

Моделовање случајних догађаја

```
indikator <- function(p) # p je verovatnoca realizacije  
dogadjaja A  
{  
  gama<-runif(1)  
  if(gama > p)  
    x<-0  
  if(gama <= p)  
    x<-1  
  return (x)  
}
```

```
> indikator(0.1)  
[1] 0  
> indikator(0.6)  
[1] 1  
> indikator(0.2)  
[1] 0
```

Моделовање биномне расподеле

Нека је вероватноћа реализације неког догађаја у сваком експерименту једнака p . Случајна величина S_n која је једнака броју реализација догађаја у n независних експеримената, има биномну расподелу $B(n, p)$. Њен закон расподеле је

$$P\{S_n = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

На основу дефиниције, S_n се може представити као збир независних случајних величина I_k које све имају исту расподелу и представљају индикатор догађаја у k -том експерименту, односно

$$S_n = \sum_{k=0}^n I_k.$$

Моделирају се вредности за I_k , $k = 1, 2, \dots, n$ и саберу се. Добијени број је реализована вредност случајне величине S_n .

Моделовање биномне расподеле

```
binomna <- function(n,p)
{
  x <- vector()
  for(i in 1:n) {
    x[i] <- indikator(p)
  }
  Sn <- sum(x)
  return (Sn)
}
```

```
> binomna(20,0.3)
```

```
[1] 7
```

```
> binomna(130,0.4)
```

```
[1] 50
```

```
> binomna(15,0.7)
```

```
[1] 9
```

```
> binomna(55,0.1)
```

```
[1] 5
```

Предност овог поступка је што се не морају рачунати вероватноће p_k из закона расподеле за S_n , а недостатак је што је потребно n (псеудо) случајних бројева за добијање једне реализације случајне величине.

Моделовање дискретних случајних величина са пребројиво много вредности

Нека је X дискретна случајна величина са пребројиво много вредности и нека је познат закон расподеле:

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

Нека је n природни број такав да је $p_{n_0+1} + p_{n_0+2} + \dots < \delta$, где је δ произвољно мали (унапред изабрани) позитиван реални број. Уместо случајне величине X посматра се засечена случајна величина:

$$X^* = \begin{pmatrix} x_1 & x_2 & \dots & x_{n_0} \\ p_1 & p_2 & \dots & p_{n_0}^* \end{pmatrix}$$

где је $p_{n_0}^* = 1 - p_1 - \dots - p_{n_0-1}$.

Моделовање дискретних случајних величина са пребројиво много вредности

Вредност ове случајне величине X^* се моделира на начин који је већ описан за моделовање вредности дискретне случајне величине са коначно много вредности.

На тај начин се неће добити ни једна од вредности случајне величине X која је већа од x_n .

Међутим, вероватноћа добијања било које вредности веће од x_n је мања од δ , а како је δ унапред изабрани врло мали позитиван број, то су и вероватноће добијања вредности случајне величине X које су веће од x_n занемарљиво мале.

У зависности од задатка који се решава и броја вредности које је потребно моделирати бира се δ .

Моделовање геометријске расподеле

Нека је вероватноћа реализације неког догађаја у сваком експерименту једнака p и нека се експерименти понављају (при истим условима) док се први пут не оствари догађај. Случајна величина X која представља број изведених експеримената, има геометријску расподелу. Њен закон расподеле је:

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p & (1-p)p & \dots & (1-p)^{n-1}p & \dots \end{pmatrix}$$

Ова случајна величина се моделира тако што се, за изабрани позитивни број δ и познату вероватноћу p , одреди најмањи природни број који задовољава неједнакост:

$$n_0 > \frac{\ln \delta}{\ln(1-p)}.$$

Формира се засечена случајна величина:

$$X^* = \begin{pmatrix} x_1 & x_2 & \dots & x_{n_0} \\ p_1 & p_2 & \dots & p_{n_0}^* \end{pmatrix}$$

Моделовање геометријске расподеле

```
geometrijska <- function(p, delta){  
  n0 <- trunc(log(delta)/log(1-p))+1  
  X <- c(1:n0)  
  P <- c()  
  for(i in 1:n0)  
    P[i]<-((1-p)^(i-1))*p  
  gama<-runif(1)  
  if(gama<=P[1])  
    x<-X[1]  
  pom<-P[1]  
  for(i in 2:n0){  
    if(gama>pom && gama<=pom+P[i])  
      x<-X[i]  
    pom<-pom+P[i]  
  }  
  return (x)  
}
```

```
> geometrijska(0.3,0.0001)  
[1] 10  
> geometrijska(0.6,0.0001)  
[1] 3
```

Моделовање Пуасонове расподеле

Случајна величина X има Пуасонову расподелу са параметром λ , $\lambda > 0$ ако је њен закон расподеле облика

$$p_k = P\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Ова случајна величина се моделира тако што се, за изабрани позитивни број n_0 и познати параметар λ , одреди најмањи природни број који задовољава неједнакост:

$$\exp^{-\lambda} \frac{\lambda^{n_0+1}}{(n_0+1)!} \frac{n_0+2}{n_0+2-\lambda} < \delta.$$

Формира се засечена случајна величина:

$$X^* = \begin{pmatrix} x_1 & x_2 & \dots & x_{n_0} \\ p_1 & p_2 & \dots & p_{n_0}^* \end{pmatrix} \quad (1)$$

где је $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, \dots, n_0 - 1$, $p_{n_0}^* = 1 - \sum_{k=0}^{n_0-1} p_k$, а затим се моделира коришћењем поступка за моделовање дискретних случајних величина са коначно много вредности.

Моделовање непрекидних случајних величина

Метода инверзне функције

Теорема: Нека је дата случајна величина X чија је функција расподеле F строго монотона и непрекидна и нека је F^{-1} њена инверзна функција. Нека је Y случајна величина са униформном расподелом $U(0, 1)$, тада случајна величина $F^{-1}(Y)$ има функцију расподеле F .

Дакле, реализована вредност x случајне величине X се добија помоћу једног (псеудо)случајног броја λ , по формули $x = F^{-1}(\lambda)$.

Метода инверзне функције може се применити на мали број густина расподеле, јер су у већини случајева густине расподела сложене функције, па није могуће аналитички добити инверзну функцију. Зато у пракси доминира примена нумеричких метода за интеграцију и инверзију. Ова метода се углавном примењује за генерисање случајних променљивих са експоненцијалном расподелом.

Моделовање експоненцијалне расподеле

Случајна величина X има експоненцијалну расподелу са параметром λ , $\lambda > 0$ ако је њена функција расподеле облика:

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0.$$

Ако је γ (псеудо)случајни број, тада је моделирана вредност случајне величине X једнака $x = \frac{-1}{\lambda} \ln(1 - \gamma)$.

Овај израз се може поједноставити коришћењем особине униформне расподеле: Ако случајна величина Y има униформну расподелу $U(0, 1)$, тада и случајна величина $1 - Y$ има униформну расподелу $U(0, 1)$.

На тај начин се добија једноставна формула за моделовање вредности експоненцијалне расподеле

$$x = \frac{-1}{\lambda} \ln \gamma,$$

где је γ (псеудо)случајни број.

Моделовање експоненцијалне расподеле

```
eksponencijalna <- function(lambda)
{
  gama<-runif(1)
  x<-(-1/lambda)*log(1-gama)
  return(x)
}
```

```
> eksponencijalna(4)
[1] 0.1073578
> eksponencijalna(2)
[1] 0.1871785
> eksponencijalna(9)
[1] 0.1641344
> eksponencijalna(12)
[1] 0.001741015
```

Моделовање Ерлангове (гама) расподеле

Случајна величина X има гама расподелу са параметром λ , $\lambda > 0$ ако је њена густина расподеле облика:

$$f(x) = \frac{x^{a-1}e^{-x}}{\Gamma(a)}, \quad x > 0.$$

Ова расподела се може представити као збир независних једнако расподељених случајних величина са експоненцијалном $\epsilon(\lambda)$ расподелом.

Ако имамо n (псеудо)случајних бројева $\gamma_1, \gamma_2, \dots, \gamma_n$, тада се моделирана вредност случајне променљиве X добија по формули:

$$x = -\ln(\gamma_1 \cdot \dots \cdot \gamma_n).$$

Моделовање Ерлангове (гама) расподеле

```
gama_1 <- function(a){  
  y<-runif(20)  
  x <- -log(prod(y))  
  return(x)  
}
```

```
> gama_1(3)  
[1] 22.02735  
> gama_1(15)  
[1] 23.74357  
> gama_1(30)  
[1] 18.8909
```

Моделовање Гумбелове расподеле

Случајна променљива X има Гумбелову (двоструку експоненцијалну) расподелу са параметрима $u, u > 0$ и $a, a > 0$ ако је њена функција расподеле

$$F(x) = e^{-e^{-\frac{x-u}{a}}}, \quad x \in \mathbb{R}.$$

Може се применити метода инверзне функције.

Ако је γ (псеудо)случајни број, тада је моделована вредност случајне променљиве X једнака $x = u - a \ln(-\ln \gamma)$.

Моделовање Гумбелове расподеле

```
gumbel <- function(a,b){  
  u <- runif(1)  
  x <- a-b*log(-log(u))  
  return(x)  
}
```

```
> gumbel(2,6)  
[1] 8.829906  
> gumbel(12,15)  
[1] 3.397028  
> gumbel(8,22)  
[1] 2.981671
```

Моделовање Фрешеове расподеле

Фрешеова расподела са параметрима $a \in \mathbb{R}$, $b > 0$ и $k \geq 0$ одређена је функцијом расподеле

$$F(x) = e^{-\left(\frac{x-a}{b}\right)^k}, \quad x > a.$$

Примена методе инверзне функције даје $x = a + b(-\ln \gamma)^{-\frac{1}{k}}$, где је γ (псеудо)случајни број.

Моделовање Фрешеове расподеле

```
freseova <- function(a,b,k){  
  gama <- runif(1)  
  x <- a+b*(-log(gama))^((-1)/k)  
  return(x)  
}
```

```
> freseova(2,5,4)  
[1] 6.046877  
> freseova(31,45,7)  
[1] 72.74598  
> freseova(143,453,10)  
[1] 718.0458
```