Estatística Computacional

Simulação de variáveis aleatórias - Parte 2

Prof. Paulo Cerqueira Jr
Faculdade de Estatística - FAEST
Programa de Pós-graduação em Matemática e Estatística - PPGME

https://github.com/paulocerqueirajr

Introdução

Introdução

- Vimos o método congruêncial para gerar valores de uma uniforme.
- Método da transformação inversa e da aceitação/rejeição.
- Vamos ver alguns métodos específicos.
- Até então, vimos somente caso contínuo como aplicação dos métodos.
- Dessa forma, iremos ver o caso discreto também.

Métodos específicos

Métodos específicos

1. Simulação e uma variável $X \sim U(a, b)$, onde $-\infty < a < b < \infty$, basta ver que

$$Y = (b - a)X + a \sim U(a, b), \text{ se } X \sim U(0, 1).$$

2. Para simular da distribuição de $Y \sim Exp(\lambda)$, onde $\lambda > 0$, basta ver que

$$Y = \frac{1}{\lambda} X \sim Exp(\lambda), \text{ se } X \sim Exp(1).$$

Métodos específicos

3. Para simular da distribuição de $Y \sim Gama(n, \lambda)$, onde $n \in \mathbb{N}$ e $\lambda > 0$, basta ver que

$$Y = \sum_{i=1}^{n} X_i \sim Gama(n, \lambda), \;\; \text{se} \;\; X \sim Exp(\lambda).$$

- Logo, segue que,

$$Y = \sum_{i=1}^{n} -\frac{1}{\lambda}\ln(U_i) = -\frac{1}{\lambda}\ln\left(\prod_{i=1}^{n} U_i\right), \;\; \text{se} \;\; U_i \sim U(0,1), \; i = 1, 2, \ldots, n.$$

Métodos específicos

4. Para simulação da distribuição $X \sim N(0,1)$, basta usar a transformação *Box-Muller*. Observe que $R^2 = -2\ln(U_1) \sim Exp(1/2)$ e $\Theta = 2\pi U_2$ se $U_1,\ U_2 \sim U(0,1)$. Logo

$$\begin{cases} X = R\cos\Theta = \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \sim N(0,1) \\ Y = R\sin\Theta = \sqrt{-2\ln(U_1)}\sin(2\pi U_2) \sim N(0,1) \end{cases}$$

5. Para simular da distribuição de $X \sim N(\mu, \sigma^2)$, onde $\mu \in \mathbb{R}$ e $\sigma^2 > 0$, basta ver que

$$Y = \sigma Z + \mu \sim N(\mu, \sigma^2), \quad \text{se} \quad Z \sim N(0,1).$$

Métodos específicos

- Para simular da distribuição $Y \sim \chi^2_{(n)}$, onde $n \in \mathbb{N}$, basta ver,

$$Y = \sum_{i=1}^{n} Z_i \sim \chi^2_{(n)}, \text{ se } Z_i \sim N(0,1), i = 1, 2, \ldots, n.$$

- Em particular, se n é par, ou seja, $n = 2k$ para $k = 1, 2 \ldots$, segue que $Y \sim \chi^2_{(2k)} = Gama(k, 1/2)$. Logo,

$$Y = -2 \ln\left(\prod_{i=1}^{n} U_i\right), \text{ se } U_i \sim U(0,1), i = 1, 2, \ldots, k.$$

- Se n é impar, ou seja, $n = 2k + 1$ para $k = 1, 2 \ldots$, segue que

$$Y = Z^2 + \chi^2_{(2k)} + \chi^2_{(2k+1)}, \text{ onde } Z \sim N(0,1).$$

Geração de VA - Caso discreto

Método da transformação inversa

- Suponha que se deseje simular valores da distribuição de uma variável aleatória discreta X com função massa de probabilidade dada por

$$p_j = f(x_j) = P(X = x_j),\ j = 1, 2, \ldots,$$

onde $\sum_j p_j = 1$. Para fazer isto, basta gerar uma $U \sim U(0,1)$ e definir

$$\begin{cases} x_0 & \text{se } U \leq p_0 \\ x_1 & \text{se } p_0 < U \leq p_0 + p_1 \\ x_2 & \text{se } p_0 + p_1 < U \leq p_0 + p_1 + p_2 \\ \vdots & \\ x_j & \text{se } \sum_{i=0}^{j-1} p_i < U \leq \sum_{i=0}^{j} p_i \\ \vdots & \end{cases}$$

Método da transformação inversa

Observe que se $F^{-1} : (0, 1) \to \mathbb{R}$ é definida por

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}.$$

Então a equação, temos que

$$X = F^{-1}(U).$$

É simples mostrar que $P(X = x_j) = p_j$. De fato

$$P(X = x_j) = P\left(\sum_{i=1}^{j-1} p_i < U \leq \sum_{i=1}^{j} p_i\right) = \sum_{i=1}^{j} p_i - \sum_{i=1}^{j-1} p_i = p_j.$$

`O método acima representa a versão discreta do método da transformação inversa.`

Exemplo 1

> Seja X uma v.a. discreta assumindo valores em $\{1, 2, 3, 4\}$ com função massa de probabilidade dada por $p_1 = 0,20$, $p_2 = 0,15$, $p_3 = 0,25$ e $p_4 = 0,40$. Obtenha uma amostra de tamanho 20 da distribuição de X.

`Sol:` Usando a definição do método, temos

$$\begin{cases} 1 & \text{se } U \leq 0,2 \\ 2 & \text{se } 0,2 < U \leq 0,2 + 0,15 = 0,35 \\ 3 & \text{se } 0,35 < U \leq 0,35 + 0,25 = 0,6 \\ 4 & \text{se} U > 0,6 \end{cases}$$

Esses são os intervalos que iremos usar na implementação!

Exemplo 1

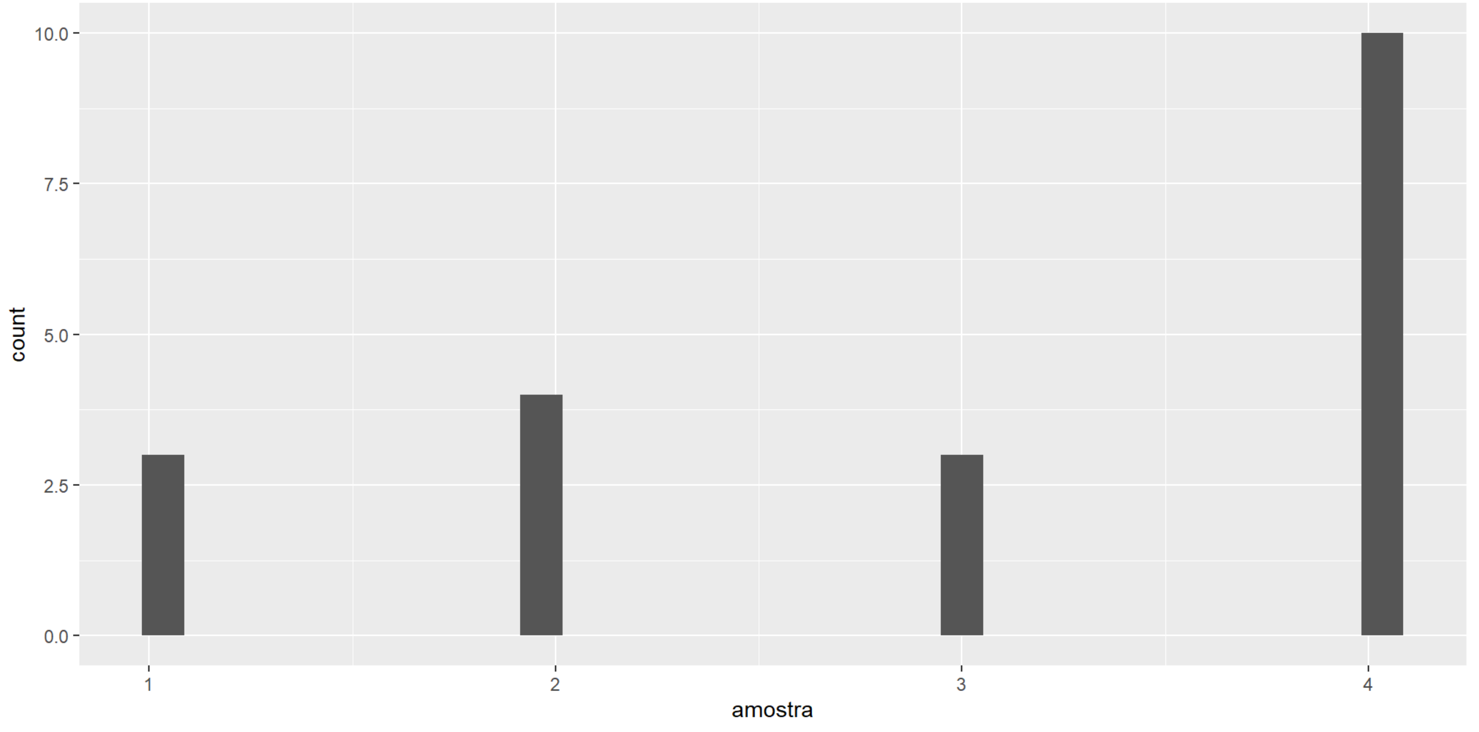
No R:

```
tam     <- 20
amostra <- NULL # Vetor de armazenamento
for(k in 1:tam){
    u <- runif(1,0,1)
    if(u <= 0.20) amostra[k] <- 1 # condição 1
    if((0.20 < u) & (u <= 0.35)) amostra[k] <- 2 # condição 2
    if((0.35 < u) & (u <= 0.60)) amostra[k] <- 3 # condição 3
    if(u > 0.60) amostra[k] <- 4 # condição 3
}
amostra
```

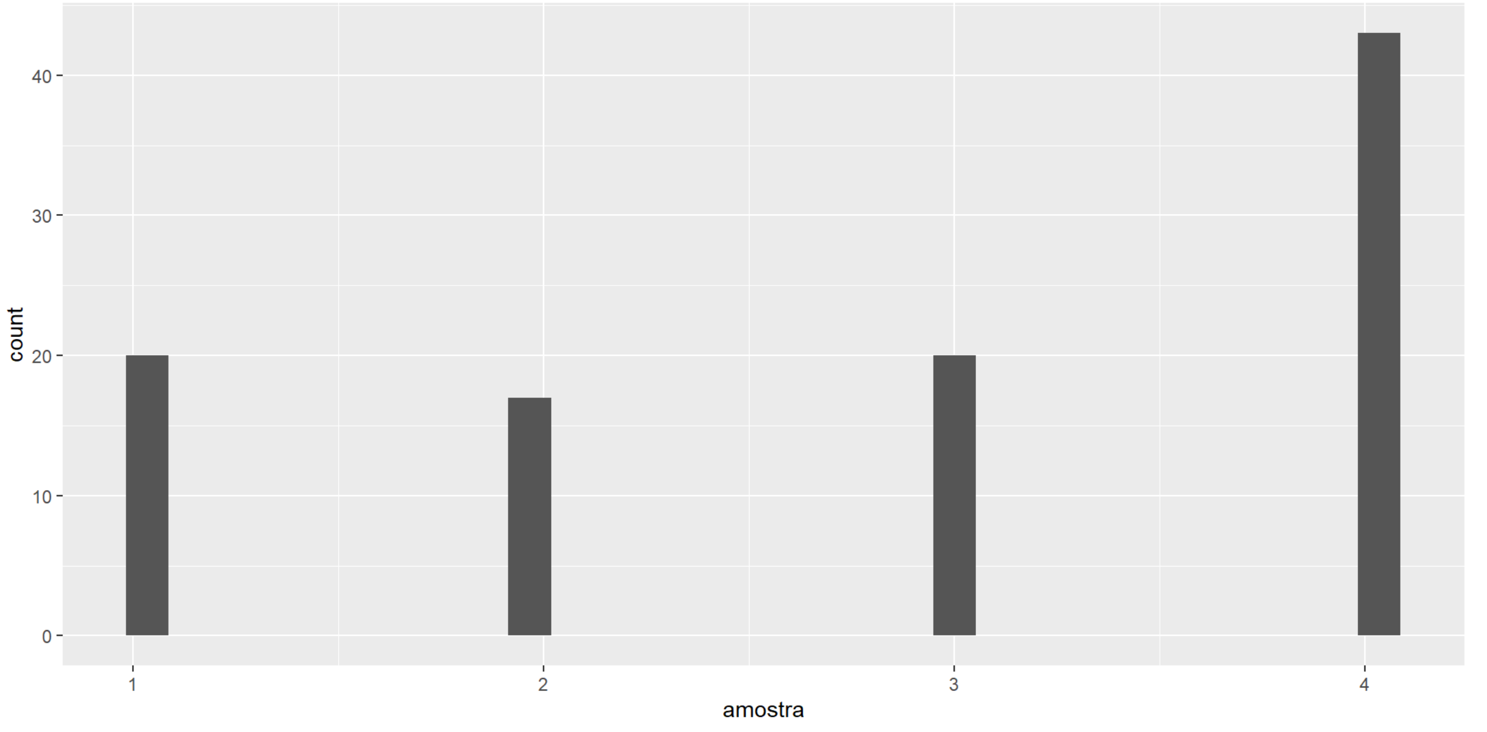
```
[1] 4 3 4 2 4 1 3 4 4 1 2 4 4 2 4 4 1 3 4 2
```

Exemplo 1

Tamanho 20.



Tamanho 100



Exemplo 2

> Seja X uma v.a. aleatória discreta uniformemente distribuída em $\{1, \ldots, n\}$. Obtenha uma amostra de tamanho 100 da distribuição de X.

`Sol:` Note que,

$$X = j \quad \Leftrightarrow \frac{j-1}{n} \leq U < \frac{j}{n}$$

Portanto $X = [nU] + 1$, onde $[x]$ representa o maior inteiro menor que ou igual a x. No R a operação $[x]$ é realizada com a função `floor`:

```r
floor(4.9)
```

```
[1] 4
```

```r
floor(6.2)
```

```
[1] 6
```

Exemplo 2

Segue um código em R para simular de uma uniforme discreta em $\{1, \ldots, 10\}$:

```r
n <- 10
u <- runif(100,0,1)
amostra <- floor(n*u) + 1
table(amostra)
```

```
amostra
 1  2  3  4  5  6  7  8  9 10
13 13 11 11 10  7  8  9  9  9
```

Método de aceitação/rejeição

Método de aceitação/rejeição

- Seja X uma v.a. discreta com função massa de probabilidade dada por

$$q_j = g(x_j) = P(X = x_j), \quad j = 0, 1, 2, \dots,$$

e assuma que existe um método para simular da distribuição de X.

- É possível usar este método como base para simular de uma outra v.a. discreta Y com função massa de probabilidade dada por

$$p_j = f(x_j) = P(Y = x_j), \quad j = 0, 1, 2, \dots$$

Método de aceitação/rejeição

- A idéia fundamental é inicialmente gerar da distribuição de X e então aceitar o valor gerado, com probabilidade proporcional a p_j/q_j , como sendo um valor gerado da distribuição de Y.
- Este método representa a versão discreta do método da aceitação-rejeição (A-R) e pode ser descrito nos seguintes passos:

  1. Simule $X \sim q_j = P(X = x_j)$ e $U \sim U(0, 1)$;
  2. Aceite $Y = X$ se $U \leq p_j/cq_j$. Caso contrário, volte ao passo anterior.

- A constante c é tal que $p_j \leq cq_j$ para todo j no suporte comum da distribuição de X e Y.
- A função cq_j é chamada de envelope.

Exemplo

> Seja Y uma v.a. assumindo valores em $\{1, 2, \ldots, 10\}$ com as seguintes probabilidades
>
> $$p = (0.11, 0.12, 0.09, 0.08, 0.12, 0.10, 0.09, 0.09, 0.10, 0.10),$$
>
> respectivamente. Utilize o método da aceitação e rejeição para obter uma amostra de tamanho 5000 da distribuição de Y.

Exemplo

- Vamos usar como base a uniforme discreta em $\{1, 2, \dots, 10\}$, ou seja,

$$q = (1/10, 1/10, \dots, 1/10)$$

- Precisamos encontrar o valor de c tal que $p_j \leq cq_j$ para todo $j = 1, \dots, 10$. Basta fazer

$$c = \max\left(p_1/q_1, p_2/q_2, \dots, p_{10}/q_{10}\right) = 1.2.$$

- Agora, siga os seguintes passos:

1. Simule $U_1 \sim U(0,1)$ e faça $X = [10U_1] + 1$ e simule $U_2 \sim U(0,1)$;
2. Aceite $Y = X$ se $U_2 \leq (10p_X/1.2)$. Caso contrário, volte ao passo anterior.

Exemplo

Segue agora um código em R para simular os 5000 valores da distribuição de Y.

```
n <- 5000
p <- c(0.11,0.12,0.09,0.08,0.12,
       0.10,0.09,0.09,0.10,0.10)
amostra<- NULL
k <- 0
while(k <= n-1){
u.1 <- runif(1,0,1)
x <- floor(10*u.1) + 1
u.2 <- runif(1,0,1)
pass1 <- (10*p[x])/1.2
if (u.2 <= pass1){
  ((k <- k + 1) & (amostra[k] <- x))}
}
length(amostra)
```

```
[1] 5000
```

```
table(amostra)
```

```
amostra
  1   2   3   4   5   6   7   8   9  10 
541 571 426 404 637 515 450 440 531 485 
```

```
#amostra
freq.rel <- table(amostra)/n
#freq.rel
```

Exemplo

```r
# Plota o gráfico de barras da amostra e a distribuição de Y
sup <- 1:10
barplot(freq.rel,xlab="j",ylab="pj.hat")
plot(sup,p,type="h",xlab="j",ylab="pj",ylim=c(0,0.15))
```