Geração de números não uniformes

Variáveis discretas e contínuas

Prof. Walmes Zeviani
walmes@ufpr.br

Laboratório de Estatística e Geoinformação
Departamento de Estatística
Universidade Federal do Paraná

Atualizado em 2018-08-20

Justificativas

- A GNA Uniformes são o ponto de partida para GNA de outras distribuições.
- Do ponto de vista de simulação computacional, é importante gerar números de todas as distribuições de probabilidade.

Objetivos

- Mostrar a GNA de v.a. discretas.
- Apresentar o método da transformação integral de probabilidades.

A distribuição uniforme discreta

Se X tem distribuição uniforme discreta com k valores, o seu suporte é o conjunto $x \in \{1, 2, \ldots, k\}$, $k \geq 2$. A função de probabilidade é

$$p(x) = \frac{1}{k} \cdot I(x \in \{1, 2, \ldots, k\}). \tag{1}$$

A função de probabilidade acumulada é

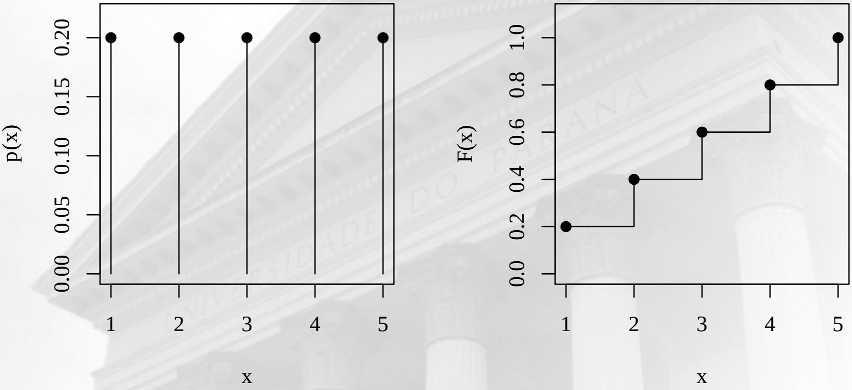
$$F(x) = \frac{x}{k}. \tag{2}$$

A média e variância são

$$\mathrm{E}(X) = \frac{k+1}{2} \quad \text{e} \quad \mathrm{V}(X) = \frac{k^2 - 1}{12}. \tag{3}$$

https://en.wikipedia.org/wiki/Discrete_uniform_distribution

Gráficos da uniforme discreta



Números da uniforme discreta

Como gerar números da uniforme discreta?

Por compartimentalização (*data binnig*).

Gera-se números da v.a. uniforme $U \sim \text{U}(0,1)$ e obtém-se números da v.a. X de distribuição uniforme discreta usando

$$X = \begin{cases} 1, & \text{se } U < 1/k \\ 2, & \text{se } 1/k \leq U < 2/k \\ \vdots & \vdots \\ k, & \text{se } (k-1)/k \leq U < 1. \end{cases} \tag{4}$$

★ Expresse a equação acima usando a função de distribuição $F(x)$.

A distribuição Bernoulli

Uma v.a. X tem distribuição Bernoulli, $X \sim \text{Ber}(p)$ se sua função de probabilidade é

$$p(x) = \begin{cases} p & \text{, se } x = 1 \\ 1 - p & \text{, se } x = 0, \end{cases} \tag{5}$$

em que $0 < p < 1$.

A média e variância são dados por

$$\text{E}(X) = p \quad \text{e} \quad \text{V}(X) = p(1 - p). \tag{6}$$

Números da Bernoulli
Como gerar números da Bernoulli?

Por dicotomização. Gere números $U \sim \mathrm{U}(0,1)$ e faça

$$X = \begin{cases} 1, & \text{se } U \leq p \\ 0, & \text{se } U > p. \end{cases} \qquad (7)$$

★ Expresse a equação acima usando a função de distribuição $F(x)$.

A distribuição geométrica

Uma v.a. X tem distribuição Geométrica se é o número tentativas até que o primeiro sucesso seja obtido em uma série de provas independentes de Bernoulli (p constante). Assim

$$p(x) = (1-p)^{x-1} \cdot p, \quad x \in 1, 2, \ldots ; 0 < p < 1. \tag{8}$$

Representamos por $X \sim \text{Geo}(p)$ onde

$$\text{E}(X) = 1/p \quad \text{e} \quad \text{V}(X) = (1-p)/p^2. \tag{9}$$

Números da geométrica

A função de distribuição da geométrica é

$$F(x) = 1 - (1-p)^x \tag{10}$$

x	$p(x)$	$\sum_x p(x)$	$F(x) = 1 - (1-p)^2$
1	p	p	p
2	pq	$p + pq$	$1 - q^2$
3	pq^2	$p + pq + pq^2$	$1 - q^3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

★ Reconhece algum padrão algébrico?

Gerar números da geométrica

```r
random_geo <- function(p) {
    q <- 1 - p
    pq <- p * q
    x <- 1
    Fx <- p
    u <- runif(1)
    while (u > Fx) {
        x <- x + 1    # Incrementa x.
        Fx <- Fx + pq # Calcula F(x).
        pq <- pq * q  # Atualiza.
    }
    return(x)
}

x <- replicate(10000, random_geo(p = 0.5))
plot(ecdf(x))
curve(pgeom(x - 1, p = 0.5), add = TRUE, type = "s", col = 2)
```

Uma implementação que visa eficiência

```r
random_geo <- function(n, p) {
    q <- 1 - p
    pq <- p * q
    x <- 1
    Fx <- p
    u_vec <- sort(runif(n))
    x_vec <- integer(n)
    for (i in 1:n) {
        while (u_vec[i] > Fx) {
            x <- x + 1
            Fx <- Fx + pq
            pq <- pq * q
        }
        x_vec[i] <- x
    }
    x_vec <- sample(x_vec)
    return(x_vec)
}
```

GNA de v.a. contínuas

- A GNA de v.a. discretas é feito por busca direta no intervalo ao qual o valor uniforme pertence na distribuição acumulada.
- Em v.a. contínuas o mesmo raciocínio aplica-se.
- Se a função de distribuição $F(x)$ tiver inversa $F^{-1}(u)$, então a geração de números da v.a. X com distribuição $F(x)$ é muito simples.
- Essa abordagem é conhecida como **método da transformação integral da probabilidade** (MTIP).

O método da transformação integral da probabilidade

- Considere que X é uma v.a. contínua com função de distribuição $F(x) = \Pr(X \leq x)$ para todo x.
- Gera-se $U \sim \text{U}(0, 1)$.
- O método da transformação integral da probabilidade estabelece que a variável aleatória $W = F^{-1}(U)$ terá distribuição $F(x)$. Ou seja, W é na realidade X.
- Para verificar isso, note que

$$\Pr(X \leq x) = \Pr(F^{-1}(U) \leq x) = \Pr(U \leq F(x)) = F(x) \tag{11}$$

considerando que $F^{-1}(u) \leq x$ e $u \leq F(x)$ coincidem para todo u e x.

http://www.portalaction.com.br/simulacao-monte-carlo/metodo-da-transforma-inversa
https://en.wikipedia.org/wiki/Probability_integral_transform

Distribuição exponencial

Uma v.a. X segue o modelo exponencial com parâmetro $\lambda > 0$ se sua densidade é dada por

$$f(x) = \lambda e^{-\lambda x} \cdot I(x > 0), \tag{12}$$

sendo $\lambda > 0$. Denota-se por $X \sim \text{Exp}(\lambda)$.

A média e variância são

$$\text{E}(X) = \frac{1}{\lambda} \quad \text{e} \quad \text{V}(X) = \frac{1}{\lambda^2}. \tag{13}$$

Geração de números da Exponencial

A função de distribuição da Exponencial é

$$F(x) = \int_0^x f(v)\,\mathrm{d}v = 1 - \exp\{-\lambda x\}. \tag{14}$$

Dessa forma, a inversa de $F(x)$ é

$$\begin{aligned} u &= 1 - \exp\{-\lambda x\} \\ 1 - u &= \exp\{-\lambda x\} \\ \log(1-u) &= -\lambda x \\ -\frac{\log(1-u)}{\lambda} &= x \\ \therefore \quad x &= F^{-1}(u) = -\frac{\log(1-u)}{\lambda}. \end{aligned}$$

Implementação em R

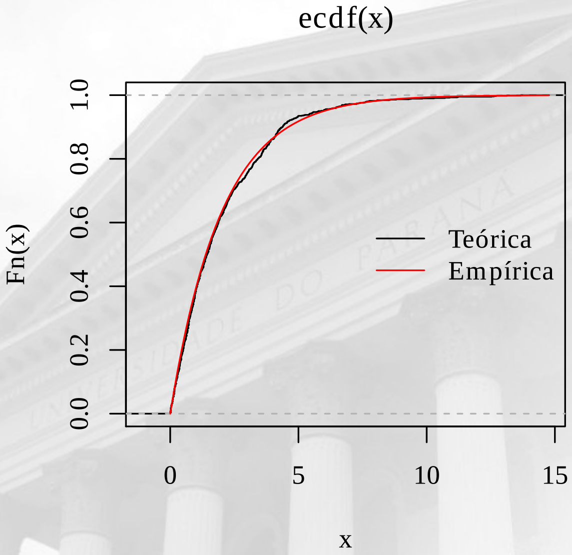
```r
random_exp <- function(n, lambda) {
    u <- runif(n)
    x <- -log(1 - u)/lambda
    return(x)
}
```

```r
x <- random_exp(n = 1000, lambda = 0.5)

Fx <- function(x, lambda) {
    1 - exp(-lambda * x)
}

plot(ecdf(x))
curve(Fx(x, lambda = 0.5), add = TRUE, col = 2, from = 0)
legend("right", legend = c("Teórica", "Empírica"),
       lty = 1, col = 1:2, bty = "n")
```

Implementação em R



Exercícios

Obter o GNA das distribuições especificadas abaixo usando o MTIP.

- $F(x) = 1 - (x-1)^2, \quad 0 < x < 1.$
- $F(x) = x^\theta, \quad 0 < x < 1, \theta > 1.$
- $F(x) = 1 - \exp\{-x^2/2\tau^2\}, \quad x > 0, \tau > 0.$
- $f(x) = \frac{1}{2}\sin(x), \quad 0 < x < \pi.$
- $f(x) = \frac{\sec^2(x)}{\sqrt{3}}, \quad 0 < x < \pi/3.$

Dica: use programas de matemática simbólica para verificar os resultados.

Wolfram Alpha: http://www.wolframalpha.com/.

Solução

- $F^{-1}(u) = 1 - \sqrt{1-u}$ ou apenas $F^{-1}(u) = 1 - \sqrt{u}$ por simetria.
- $F^{-1}(u) = u^{1/\theta}$
- $F^{-1}(u) = \sqrt{-2\tau \cdot \log(1-u)}$.
- $F(x) = \frac{1-\cos(x)}{2}$ e $F^{-1}(u) = \arccos(1-2u)$.
- $F(x) = \frac{\tan(x)}{\sqrt{3}}$ e $F^{-1}(u) = \arctan(u\sqrt{3})$.

Quando usar o MTIP para GNA?

Sempre que possível.

- O método funciona quando $F^{-1}(u)$ existe.
- Existem distribuições onde $F(x)$ é não inversão analítica.
- Existem distribuições onde $F(x)$ não existe (apenas $f(x)$).

Alternativas

- Pode-se aproximar numericamente $F(x)$ e/ou $F^{-1}(u)$ para GNA.
- Pode-se considerar métodos não baseados em $F^{-1}(u)$.

Próxima aula

- O método da aceitação-rejeição.

Avisos

- Sabatina 03 já está no Moodle!