Estatística Computacional

Patrícia de Siqueira Ramos

PPGEAB
UNIFAL-MG

9 de Maio de 2018

Outras transformações

Há outras transformações, além da transformação inversa, que podem ser utilizadas para simular variáveis aleatórias:

1) Se $Z \sim N(0,1)$, então

$$V = Z^2 \sim \chi^2(1)$$

Outras transformações

Há outras transformações, além da transformação inversa, que podem ser utilizadas para simular variáveis aleatórias:

1) Se $Z \sim N(0,1)$, então

$$V = Z^2 \sim \chi^2(1)$$

2) Se $V \sim \chi^2(m)$ e $W \sim \chi^2(n)$ são independentes, então

$$F = \frac{V/m}{W/n} \sim F_{m,n}$$

Outras transformações

Há outras transformações, além da transformação inversa, que podem ser utilizadas para simular variáveis aleatórias:

1) Se $Z \sim N(0,1)$, então

$$V = Z^2 \sim \chi^2(1)$$

2) Se $V \sim \chi^2(m)$ e $W \sim \chi^2(n)$ são independentes, então

$$F = \frac{V/m}{W/n} \sim F_{m,n}$$

3) Se $Z \sim N(0,1)$ e $V \sim \chi^2(n)$ são independentes, então

$$T = \frac{Z}{\sqrt{V/n}} \sim t_n$$

Outras transformações

4) Se $U, V \sim U(0,1)$ são independentes, então

$$Z_1 = \sqrt{-2\log U}\cos(2\pi V)$$

$$Z_2 = \sqrt{-2\log V}\sin(2\pi U)$$

são variáveis normais padrão independentes (algoritmo Box-Müller)

Outras transformações

4) Se $U, V \sim U(0, 1)$ são independentes, então

$$Z_1 = \sqrt{-2 \log U} \cos(2\pi V)$$

$$Z_2 = \sqrt{-2 \log V} \sin(2\pi U)$$

são variáveis normais padrão independentes (algoritmo Box-Müller)

5) Se $U \sim gama(r, \lambda)$ e $V \sim gama(s, \lambda)$ são independentes, então

$$X = \frac{U}{U + V} \sim beta(r, s)$$

Outras transformações

4) Se $U, V \sim U(0,1)$ são independentes, então

$$Z_1 = \sqrt{-2\log U}\cos(2\pi V)$$

$$Z_2 = \sqrt{-2\log V}\sin(2\pi U)$$

são variáveis normais padrão independentes (algoritmo Box-Müller)

5) Se $U \sim gama(r, \lambda)$ e $V \sim gama(s, \lambda)$ são independentes, então

$$X = \frac{U}{U+V} \sim beta(r, s)$$

6) Se $U, V \sim U(0,1)$ são independentes, então

$$X = \left\lfloor 1 + \frac{\log V}{\log(1-(1-\theta)^U)} \right\rfloor \sim logaritmica(\theta),$$

em que $\lfloor x \rfloor$: parte inteira de x.

Outras transformações

Exemplo 6: Já vimos um gerador de realizações de v.a.s *beta* usando o método da aceitação/rejeição. A relação (5) entre as distribuições *gama* e *beta* provê outro gerador:

Se $U \sim gama(r, \lambda)$ e $V \sim gama(s, \lambda)$ são independentes, então

$$X = \frac{U}{U + V} \sim beta(r, s).$$

Outras transformações

Exemplo 6: Já vimos um gerador de realizações de v.a.s *beta* usando o método da aceitação/rejeição. A relação (5) entre as distribuições *gama* e *beta* provê outro gerador:

Se $U \sim gama(r, \lambda)$ e $V \sim gama(s, \lambda)$ são independentes, então

$$X = \frac{U}{U + V} \sim beta(r, s).$$

Essa transformação permite criar o algoritmo para gerar realizações $beta(a, b)$:

1. Gerar uma realização u de $gama(a, \lambda)$
2. Gerar uma realização v de $gama(b, \lambda)$
3. Fazer $x = \dfrac{u}{u + v}$

Outras transformações

Exemplo 6 no R: Gerar a.a. da *beta*$(3, 2)$

```
n = 1000
a = 3
b = 2
lambda = 1
u = rgamma(n, shape = a, rate = lambda)
v = rgamma(n, shape = b, rate = lambda)
x = u/(u + v)
x
hist(x, prob = T)
# densidade obtida e a amostra gerada pelo R - rbeta
lines(density(x), col = "blue")
lines(density(rbeta(n, a, b)), col = 'red')
# QQ plot da beta(3,2)
qqplot(qbeta(ppoints(n), a, b), x,
       cex = 0.25, main = 'QQ plot da beta(3,2)',
       xlab = 'quantis teóricos', ylab = 'quantis amostrais')
qqline(x, distribution = function(p) qbeta(p,a,b))

# ppoints(n): função do R que calcula o vetor de n posições
# p[i] = (i  - 3/8)/(n + 1/4)
```

Outras transformações

Exemplo 7: A relação (6) entre as distribuições uniforme e logarítmica é:

Se $U, V \sim U(0,1)$ *são independentes, então*

$$X = \left\lfloor 1 + \frac{\log V}{\log(1-(1-\theta)^U)} \right\rfloor \sim logaritmica(\theta).$$

Algoritmo para gerar realizações da distribuição logarítmica(θ):

1. Gerar uma realização u de $U(0,1)$
2. Gerar uma realização v de $U(0,1)$
3. Fazer $x = \lfloor (1 + \log v)/\log(1-(1-\theta)^u) \rfloor$

Outras transformações

Exemplo 7 no R:

```
n = 1000
theta = 0.5
u = runif(n)
v = runif(n)
x = floor(1 + log(v)/log(1 - (1 - theta) ** u))
hist(x, prob = T, breaks = seq(0.5, max(x) + 0.5))

# comparar com resultado teórico
k = 1:max(x)
p = -1 / log(1 - theta) * theta ** k / k  # f(x) teórica
pch = tabulate(x)/n                        # observada
round(rbind(pch, p), 3)
```

Outras transformações

Exemplo 7 no R por meio de uma função que retorna a amostra da distribuição logarítmica:

```
rlogaritmica = function(n, theta) {
  stopifnot(theta > 0 & theta < 1)
  u = runif(n)
  v = runif(n)
  x = floor(1 + log(v)/log(1 - (1 - theta) ** u))
  return(x)
}
x = rlogaritmica(1000, theta)
hist(x, prob = T, breaks = seq(0.5, max(x) + 0.5))
```

Método da convolução (soma) e misturas

- Somas e misturas de variáveis aleatórias são tipos especiais de transformações
- Primeiro trataremos da soma de v.a.s independentes (convoluções) e alguns exemplos

Sejam $X_1, \ldots, X_n$ v.a.s independentes e identicamente distribuídas e seja $S = X_1 + \cdots + X_n$. A função distribuição da soma S é chamada convolução de grau n de X e denota-se $F_n^{*(n)}$. Assim, para simular uma convolução, basta gerar $X_1, \ldots, X_n$ e obter a soma.

Convolução

Muitas distribuições são obtidas por convolução. Alguns exemplos incluem:

- A convolução de n v.a.s $Bernoulli(p)$ tem distribuição $bin(n, p)$
- Se $\nu > 0$ é um inteiro, a distribuição χ^2 com ν g.l. é a convolução de ν variáveis normais padrão i.i.d. ao quadrado
- A distribuição binomial negativa $binN(r, p)$ é a convolução de r variáveis aleatórias geométricas, $geom(p)$, i.i.d.
- A convolução de r variáveis aleatórias i.i.d. $exp(\lambda)$ tem distribuição $gama(r, \lambda)$

Obviamente, é mais fácil usar as funções `rchisq`, `rgeom`, `rnbinom` etc. para gerar a.a. dessas distribuições (χ^2, *binN*, *geom* etc.), mas vamos ilustrar sua obtenção pelo método da convolução.

Convolução

Exemplo 8: Este exemplo gera uma amostra da $\chi^2(\nu)$ a partir do quadrado de ν normais. Se $Z_1, \ldots, Z_\nu$ são v.a.s $N(0,1)$ i.i.d., então $V = Z_1^2 + \cdots + Z_\nu^2$ tem distribuição $\chi^2(\nu)$.
Passos para gerar uma amostra de tamanho n da $\chi^2(\nu)$:

1. Preencha uma matriz $n \times \nu$ com $n\nu$ variáveis $N(0,1)$
2. Eleve ao quadrado cada valor da matriz
3. Obtenha as somas das linhas das normais ao quadrado. Cada soma de linha será uma observação aleatória da $\chi^2(\nu)$
4. Retorne o vetor das somas das linhas

Convolução

Exemplo 8 no R:

```
n = 1000
nu = 2
Z = matrix(rnorm(n*nu), n, nu) ** 2 # matriz de N(0,1)
v = rowSums(Z)  # soma das linhas - método 1
v = apply(Z, 1, sum)    # soma das linhas - método 2

# conferindo
mean(v)    # E(V) = nu
mean(v ** 2)  # E(V**2) = 2nu + nu**2
var(v)     # V(V) = 2nu
```

Misturas

Uma v.a. X é uma mistura discreta de modelos se

$$F_X(x) = \sum_{i=1}^{k} \theta_i F_i(x), \ \ com \ \theta_i > 0 \ \ e \ \sum_{i=1}^{k} \theta_i = 1,$$

em que as constantes θ_i são chamadas pesos ou probabilidades da mistura (notação parecida com convolução, mas as distribuições são diferentes).

Uma v.a. X é uma mistura contínua de modelos se

$$F_X(x) = \int_{-\infty}^{\infty} F_{X|Y=y}(x) f_Y(y) dy, \ \ com \ f_Y(y) > 0 \ \ e \ \int_{-\infty}^{\infty} f_Y(y) dy = 1,$$

em que $f_Y(y)$ é a função peso.

Misturas

Exemplo 9: Compare os métodos de simulação de uma convolução e de uma mistura de variáveis normais. Suponha $X_1 \sim N(0,1)$ e $X_2 \sim N(3,1)$ independentes.

$S = X_1 + X_2$ denota a convolução. A distribuição de S é normal com média $\mu_1 + \mu_2 = 3$ e variância $\sigma_1^2 + \sigma_2^2 = 2$.

Passos para simular a convolução:

1. Gerar $x_1 \sim N(0,1)$
2. Gerar $x_2 \sim N(3,1)$
3. Retornar $s = x_1 + x_2$

Misturas

Exemplo 9 (cont.): Podemos definir uma mistura normal 50%, denotada por $F_X(x) = 0,5F_{X_1}(x) + 0,5F_{X_2}(x)$. Diferente da convolução anterior, a distribuição da mistura é não normal, é bimodal.

Passos para simular a mistura:

1. Gerar um inteiro $k \in \{1, 2\}$, em que $P(1) = P(2) = 0,50$
2. Se $k = 1$, gerar $x_1 \sim N(0, 1)$
3. Se $k = 2$, gerar $x_2 \sim N(3, 1)$

Misturas

Exemplo 10: Sejam $X_1 \sim gama(2, 2)$ e $X_2 \sim gama(4, 2)$ independentes. Compare os histogramas das amostras geradas pela convolução

$$S = X_1 + X_2$$

Misturas

Exemplo 10: Sejam $X_1 \sim gama(2,2)$ e $X_2 \sim gama(4,2)$ independentes. Compare os histogramas das amostras geradas pela convolução

$$S = X_1 + X_2$$

Se $X_1 \sim gama(r,\lambda)$ e $X_2 \sim gama(s,\lambda)$, então

$$S = X_1 + X_2 \sim gama(r+s,\lambda).$$

e pela mistura

$$F_X(x) = 0,5F_{X_1}(x) + 0,5F_{X_2}(x).$$

Misturas

Exemplo 10 no R:

```
n = 1000
r = 2; s = 4; lambda = 2
x1 = rgamma(n, r, lambda)  # gama(2,2)
x2 = rgamma(n, s, lambda)  # gama(4,2)
s = x1 + x2                # a convolução
u = runif(n)
k = sample(0:1, n, replace = T)
x = k * x1 + (1 - k) * x2  # a mistura

par(mfrow=c(1,2))     # dois gráficos em uma visualização
hist(s, prob=TRUE)
hist(x, prob=TRUE)

# gama(r,lambda) + gama(s,lambda) = gama(r+s,lambda)
# densidade obtida de s e a amostra gerada pelo R
par(mfrow=c(1,1))     # voltar a ser 1 gráfico
plot(density(s), col = "blue")
lines(density(rgamma(n, r+s, lambda)), col='red')
```

Exercício

Implementar no `R` o ex. 9 (convolução e mistura de duas normais).

a) Compare as duas distribuições obtidas (de S e de X) por meio dos histogramas dispostos lado a lado.

b) Obtenha os QQplots para as variáveis S e X. Elas parecem seguir a distribuição normal? Os comandos

```
qqnorm(x);qqline(x)
```

constroem o QQplot da amostra x.

c) Obtenha a média e a variância da variável S obtida e compare seus valores com o que seria esperado.

d) Compare a densidade de S com a obtida gerando-se valores da normal com os parâmetros μ e σ adequados (atenção, pois `rnorm` possui como parâmetros a média e o desvio padrão).