c1

Técnicas Computacionais em Probabilidade e Estatística I

Aula X

Chang Chiann

MAE 5704- IME/USP

1º Sem/2008

Slide 1

c1 chang; 22/4/2008

Simulação Estática

<u>Objetivo</u>: Em análise estatística de dados, modelos estocásticos são adotados a dados reais e estudos de simulação são usados para análise do modelo adotado

(Por meio destas técnicas pode-se estudar as propriedades de um modelo estocástico)

$\Rightarrow$ Gerar números aleatórios (pseudo-aleatórios)

$\Rightarrow$ Gerar valores de variáveis aleatórias arbitrárias (discretas e contínuas)

$\Rightarrow$ Usar v.a. para gerar o comportamento de um modelo estocástico $\Rightarrow$ estudar o comportamento de estimadores

Motivação

A Natureza e Fenômenos Aletórios

$$\underset{\substack{\uparrow \\ \text{Fenômeno} \\ \text{Real}}}{Y} = \underbrace{f(X) + e}_{\substack{\text{Modelo estrutural estocástico adotado} \\ \text{para explicar a variação de Y em função} \\ \text{das variáveis X, além de um resíduo}}}$$

$\Rightarrow$ O Modelo estrutural e estocástico deve ser o mais próximo possível do fenômeno real, mas deve ser matematicamente tratável

$\Rightarrow$ Exemplo: tempo de vida, tempo entre chegadas, tempo de cura, tamanho da ferida, diâmetro da folha, comprimento do recém-nascido, número de acidentes, número de receituários emitidos, …

Modelo Estatístico

$$\underbrace{Y}_{\substack{\text{Var. aleatória} \\ \text{(dependente)}}} = \underbrace{f(X)}_{\substack{\text{Var. explicativa} \\ \text{(independente)}}} + \underbrace{e}_{\substack{\text{Var. aleatória} \\ \text{(resíduo)}}}$$

$$E(Y) = f(X) = \mu$$

$$V(Y) = V(e) = \sigma^2$$

Suposições comuns a conjuntos de dados:

- Amostra aleatória simples de tamanho n (“iid”)
- Distribuição de probabilidade “conhecida”
- Parâmetros de Locação e Escala comuns

$\Rightarrow$ Análise de diagnóstico (gráficos)

Motivação
ANOVA – Efeitos Fixos

Suposições: modelo estrutural e distribucional são adotados aos dados

$$y_{ij} = \boxed{\mu_j} + \varepsilon_{ij} \qquad \varepsilon_{ij} \overset{iid}{\sim} N(0;\sigma^2)$$

componente fixo

- ✓ *normalidade*
- ✓ *homocedasticidade*
- ✓ *independência*

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

efeito do tratamento

$$\begin{cases} \textbf{H}_0\textbf{:}\ \mu_1 = \mu_2 = \ldots = \mu_k = \mu \iff \tau_1 = \ldots = \tau_k = 0 \\ \textbf{H}_1\textbf{: existe pelo menos uma diferença} \end{cases}$$

$$F = \frac{SSM/(k-1)}{SSE/(n-k)} = \frac{QMM}{QME} \overset{\textit{sob } H_0}{\sim} F_{k-1,n-k}$$

Análise de Diagnóstico

Números Aleatórios

X : Variável aleatória

$X = x$: Valor da variável aleatória

Geradores de Números aleatórios: Tabelas, calculadoras, computadores

$\Rightarrow$ Computadores: geram # pseudo-aleatórios pois são obtidos deterministicamente, a partir de uma semente, e se "parecem" com U(0;1) independentes

$X \sim U(0\,;1)$ V.a. com distribuição Uniforme no intervalo (0,1)

$$X \sim U(a\,;b)$$

$$f_X(x) = \frac{1}{b-a}, \quad 0 < x < 1; \qquad E(X) = \int_a^b x\, f_X(x)\, dx = \frac{b+a}{2}$$

$$V(X) = E(X - E(X))^2 = \frac{1}{12}(b-a)^2$$

Geradores de Números Aleatórios

- Método Congruencial Multiplicativo

$$x_0: \text{ Semente (valor inicial)}$$

$$x_n = a\, x_{n-1} \quad \mathrm{mod}\, ulo \quad m; \quad n \geq 1$$

$a\, x_{n-1}$: é divisível por m $\quad x_n = R = 0,1,2,\ldots,m-1$: é o resto da divisão

$$\frac{x_n}{m} \sim U(0\,;1)$$

Ex: Computador 32 bits $\Rightarrow$ m=2^{31}-1 a=7^5=16.807

36 bits $\Rightarrow$ m=2^{35}-31 a=5^5

- Método Congruencial Misto (envolve termos multiplicativo e aditivo)

$$x_0; \quad x_n = a\, x_{n-1} + c \quad \mathrm{mod}\, ulo \quad m; \quad n \geq 1$$

Geradores de Números Aleatórios

- Maioria das linguagens já tem um gerador de # aleatórios implementado

No Aplicativo R:

```
> k<-10 # numero de amostras
> x<-runif(k)
> x
 [1] 0.97009484 0.65791503 0.68147775 0.24834995 0.20636895
0.97331990
 [7] 0.05213162 0.06984107 0.23460813 0.12680279
```

Aplicação

- Cálculo de Integrais $\theta = \int_0^1 g(x)\,dx$?

$$\Rightarrow U \sim U(0\,;1) \quad \Rightarrow \quad \theta = E(g(U))$$

Lei Forte dos Grandes Números:

$$\Rightarrow U_1, U_2, \ldots, U_k \overset{iid}{\sim} U(0\,;1)$$

$$\Rightarrow g(U_1), g(U_2), \ldots, g(U_k)$$ v.a. *iid* com valor esperado θ

Com probabilidade 1 tem-se:

$$\Rightarrow \sum_{j=1}^{k} \frac{g(U_j)}{k} \quad \overset{k \to \infty}{\longrightarrow} \quad E(g(U)) = \theta$$ Pesquisa de Monte Carlo!

Aplicação

- Cálculo de Integrais

Ex. 1: $$\theta = \int_0^1 \exp(e^x)\, dx$$

```
> k<-1000 # numero de amostras
> x<-runif(k)
> gx<-exp(exp(x))
> mean(gx)
[1] 6.205932
```

Ex. 1: $$\theta = \int_0^1 \ln(x+1)\, dx$$

```
> k<-1000 # numero de amostras
> x<-runif(k)
> gx<-log(x+1)
> mean(gx)
[1] 0.3935435
```

Aplicação

- Cálculo de Integrais

$$\theta = \int_a^b g(x)\,dx \qquad ?$$

$$\theta = \int_a^b g(x)\,dx = \boxed{\left\langle y = \frac{x-a}{b-a} \quad dy = \frac{1}{b-a}dx \right\rangle}$$

$$\theta = \int_0^1 g\left[y(b-a)+a\right](b-a)dy$$

$$= \int_0^1 h(y)dy; \qquad h(y) = (b-a)g\left(a+(b-a)y\right)$$

Aplicação

- Cálculo de Integrais

$$\theta = \int_0^{\infty} g(x)\,dx \qquad ?$$

$$\theta = \int_0^{\infty} g(x)\,dx = \boxed{\left\langle y = \frac{1}{x+1} \quad dy = \frac{-1}{(x+1)^2}dx = -y^2 dx \right\rangle}$$

$$\theta = \int_0^1 h(y)dy; \qquad h(y) = \frac{g\left(\frac{1}{y} - 1\right)}{y^2}$$

Aplicação

- Cálculo de Integrais Multidimensionais

$$\theta = \int_0^1 \int_0^1 \dots \int_0^1 g(x_1, x_2, \dots, x_n)\, dx_1 dx_2, \dots, dx_n \qquad ?$$

$$\Rightarrow \theta = E\left(g(U_1, U_2, \dots, U_n)\right) \quad \Rightarrow \quad U_j \overset{iid}{\sim} U(0\,;1)$$

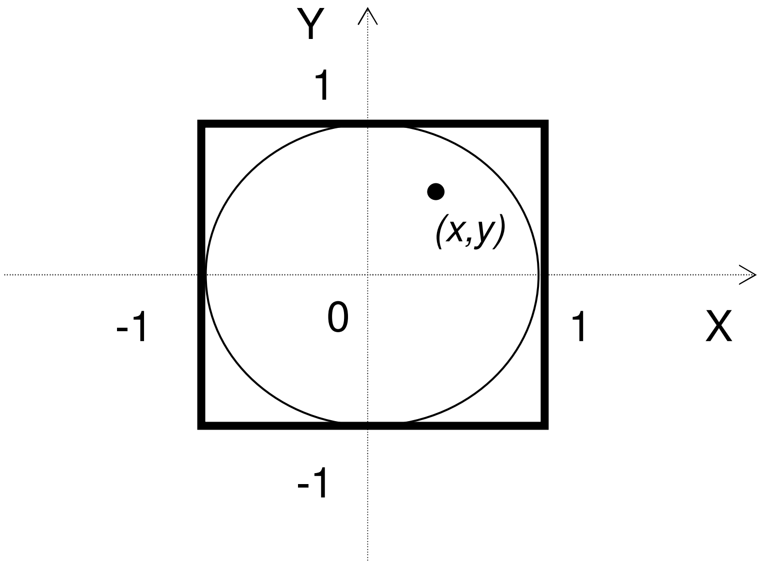
Cálculo aproximado: Gerar k conjuntos independentes de n U(0;1) independentes

$$\Rightarrow \begin{bmatrix} U_1^1 & U_2^1 & \dots & U_n^1 \\ U_1^2 & U_2^2 & \dots & U_n^2 \\ & & \dots & \\ U_1^k & U_2^k & \dots & U_n^k \end{bmatrix}; \quad \Rightarrow g\left(U_1^j, U_2^j, \dots, U_n^j\right)$$ são iid com média θ

$$\Rightarrow \sum_{j=1}^{k} \frac{g\left(U_1^j, U_2^j, \dots, U_n^j\right)}{k} \quad \overset{k \to \infty}{\longrightarrow} \quad \theta$$

Aplicação

- Estimação do Número π



$$\Rightarrow P((x,y) \in circulo) = P(x^2 + y^2 \leq 1) = \frac{\pi}{4}$$

Gerar pontos no quadrado (um grande número de vezes) e a proporção de pontos que caem dentro do círculo será aproximadamente $\pi/4$.

Aplicação

- Estimação do Número π

$$\Rightarrow X \sim U\left(-1;1\right) \quad Y \sim U\left(-1;1\right); \quad X \perp Y$$

$$\Rightarrow f_{XY}(x,y) = f_X(x) f_Y(y) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{1}; \quad -1 \le x, y \le 1$$

$$\Rightarrow U \sim U\left(0;1\right); \quad 2U \sim U\left(0;2\right); \quad (2U-1) \sim U\left(-1;1\right)$$

$$\Rightarrow X = 2U_1 - 1 \quad Y = 2U_2 - 1$$

$$\Rightarrow I = \begin{cases} 1 & se \quad x^2 + y^2 \le 1 \\ 0 & cc \end{cases} \qquad \Rightarrow E\left(I\right) = P\left(I=1\right) = P\left(X^2 + Y^2 \le 1\right) = \frac{\pi}{4}$$

Aplicação

- Estimação do Número π

```
> k<-1000 # numero de amostras
> n<-2  # tamanho da amostra
> z<-numeric(0)
> for(i in 1:k)
+ {
+   xy<-runif(n)          # amostra de tamanho n da U(0,1)
+   z[i]<-xy[1]^2+xy[2]^2 # z calculado para cada amostra
+ }
> vi<-ifelse(z<1,1,0) # varißvel indicadora
> sum(vi)/k
[1] 0.775
```

Conclusão:

Simulação Estocástica

<u>Gerando Variáveis Aleatórias Discretas</u>

$\Rightarrow$ Método da Transformada Inversa

$\Rightarrow$ Método da Rejeição (Aceitação/Rejeição)

$\Rightarrow$ Método da Composição

Método da Transformada Inversa

X: variável aleatória discreta

$$P(X = x_j) = p_j; \qquad j = 0,1,\ldots \quad \sum_j p_j = 1$$

$$\Rightarrow U \sim U(0;1) \qquad \Rightarrow X = \begin{cases} x_0 & se \quad U < p_0 \\ x_1 & se \quad p_0 \le U < p_0 + p_1 \\ \cdots & \\ x_j & se \quad \sum_{i=1}^{j-1} p_i \le U < \sum_{i=1}^{j} p_i \\ \cdots & \end{cases}$$

Desde que: $0 < a < b < 1 \quad P(a \le U < b) = b - a$

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

Método da Transformada Inversa

<u>ALGORITMO:</u>

$$P(X = x_j) = p_j; \qquad j = 0,1,2,\ldots$$

$$\Rightarrow Gerar \quad U \sim U(0;1)$$

$$\begin{cases} 1.\ \ Se \quad U < p_0 \Rightarrow X = x_0, \quad pare \\ 2.\ \ Se \quad U < p_0 + p_1 \Rightarrow X = x_1, \quad pare \\ 3.\ \ Se \quad U < p_0 + p_1 + p_2 \Rightarrow X = x_3, \quad pare \\ 4.\ \ \ldots \end{cases}$$

Método da Transformada Inversa

<u>Exemplo:</u> Simular X, tal que:

$$P(X = x_j) = p_j; \qquad j = 1,2,3,4 \qquad p_1 = 0,2 \quad p_2 = 0,15 \quad p_3 = 0,25 \quad p_4 = 0,4$$

$$\Rightarrow U \sim U(0;1) \quad \begin{cases} 1.\ \ Se \quad U < 0,20 \Rightarrow X = 1, \quad pare \\ 2.\ \ Se \quad U < 0,35 \Rightarrow X = 2, \quad pare \\ 3.\ \ Se \quad U < 0,60 \Rightarrow X = 3, \quad pare \\ 4.\ \ c.c. \quad X = 4 \end{cases}$$

Outra alternativa (mais eficiente):

$$\Rightarrow U \sim U(0;1) \quad \begin{cases} 1.\ \ Se \quad U < 0,4 \Rightarrow X = 4, \quad pare \\ 2.\ \ Se \quad U < 0,65 \Rightarrow X = 3, \quad pare \\ 3.\ \ Se \quad U < 0,85 \Rightarrow X = 1, \quad pare \\ 4.\ \ c.c. \quad X = 2 \end{cases}$$

Método da Transformada Inversa

X: variável aleatória discreta

$$P(X = x_j) = p_j; \qquad j = 0,1,\ldots \quad \sum_j p_j = 1$$

$$\Rightarrow U \sim U(0;1) \qquad \Rightarrow X = \begin{cases} x_0 & se \quad U < p_0 \\ x_1 & se \quad p_0 \le U < p_0 + p_1 \\ \cdots & \\ x_j & se \quad \sum_{i=1}^{j-1} p_i \le U < \sum_{i=1}^{j} p_i \\ \cdots & \end{cases}$$

Se: $x_0 < x_1 < x_2 < \ldots \quad \Rightarrow F(x_k) = P(X \le x_k) = \sum_{i=1}^{k} p_j$

$$\boxed{\Rightarrow X = x_j \quad se \quad F(x_{j-1}) \le U < F(x_j)}$$

Método da Transformada Inversa

<u>Exemplo:</u> Simular X: variável aleatória uniforme discreta

$$P(X = j) = \frac{1}{n}; \qquad j = 1,2,\ldots,n$$

$$\Rightarrow U \sim U(0;1)$$

$$\Rightarrow X = x_j \quad se \quad F(x_{j-1}) \leq U < F(x_j)$$

$$\Rightarrow X = j \quad se \quad \frac{j-1}{n} \leq U < \frac{j}{n} \quad \Rightarrow \quad j - 1 \leq nU < j$$

$$\Rightarrow X = [nU] + 1$$

Escreva o algoritmo!

Método da Transformada Inversa

<u>Exemplo:</u> Uso da v.a. uniforme discreta

Gerar uma permutação aleatória dos números 1, 2,…, n

$$\Rightarrow U \sim U(0;1) \qquad \Rightarrow I = [nU]+1 \quad \textit{ocupará a posição } n$$

$$\Rightarrow U \sim U(0;1) \qquad \Rightarrow I = [(n-1)U]+1 \quad \textit{ocupará a posição } n-1$$

$$\Rightarrow U \sim U(0;1) \qquad \Rightarrow I = [(n-2)U]+1 \quad \textit{ocupará a posição } n-2$$

…

Para um conjunto geral de n números seguir com a ordenação e atribuição da série 1,2,…,n $\Rightarrow$ aplicar o algoritmo acima.

Método da Transformada Inversa

<u>Aplicação:</u>

- Gerar a Geométrica: $X \sim Geom(p) \quad P(X = i) = pq^{i-1} \quad i \geq 1$

- Gerar a Poisson: $X \sim P(\lambda) \quad \Rightarrow \quad P(x = i) = \dfrac{e^{-\lambda}\lambda^{i}}{i\,!} \quad i \geq 0$

- Gerar a Binomial: $X \sim Bino(n; p) \quad \Rightarrow \quad P(x = i) = \dfrac{n!}{i!(n-i)!} p^{i} q^{n-i} \quad i = 0,1,\ldots,n$

Escreva os algoritmos!

Método da Transformada Inversa

Aplicação:

- Gerar a seguinte v.a. X tal que:

X=x	1	2	3	4	5	6	7	8	9	10
P(X=x)	0,11	0,12	0,09	0,08	0,12	0,1	0,09	0,09	0,1	0,1

$$\Rightarrow U \sim U(0;1) \quad \begin{cases} 1.\ \ Se \quad U < 0{,}08 \Rightarrow X = 4, \quad pare \\ 2.\ \ Se \quad U < 0{,}17 \Rightarrow X = 3, \quad pare \\ 3.\ \ Se \quad U < 0{,}26 \Rightarrow X = 7, \quad pare \\ 3.\ \ Se \quad U < 0{,}35 \Rightarrow X = 8, \quad pare \\ 3.\ \ Se \quad U < 0{,}45 \Rightarrow X = 6, \quad pare \\ 3.\ \ Se \quad U < 0{,}55 \Rightarrow X = 9, \quad pare \\ 3.\ \ Se \quad U < 0{,}65 \Rightarrow X = 10, \quad pare \\ 3.\ \ Se \quad U < 0{,}76 \Rightarrow X = 1, \quad pare \\ 3.\ \ Se \quad U < 0{,}88 \Rightarrow X = 2, \quad pare \\ 4.\ \ c.c. \quad X = 5 \end{cases}$$

Método da Rejeição

$\Rightarrow$ Alternativa mais eficiente:

- Gerar a seguinte v.a. X tal que:

X=x	1	2	3	4	5	6	7	8	9	10
P(X=x)	0,11	0,12	0,09	0,08	0,12	0,1	0,09	0,09	0,1	0,1

$$\Rightarrow U \sim U(0;1) \quad \Rightarrow \quad Y = [10 * U] + 1; \quad Y \sim U(1;10)$$

$$\Rightarrow X = y \quad se \quad \frac{P(X = y)}{P(Y = y)} = \frac{p_X(y)}{p_Y(y)} = \frac{p_X(y)}{1/10} \leq c; \quad c \geq 1$$

Método da Rejeição (Aceitação/Rejeição)

Suponha que a variável aleatória Y tem sido gerada eficientemente:

$$Y; \quad P(Y = y_j) = q_j \quad j \geq 0$$

Em geral Y é a uniforme discreta

Y é usada para gerar a variável aleatória X, tal que,

$$P(X = j) = p_j \quad j \geq 0$$

da seguinte forma:

$$\begin{cases} 1.\ \ Simule \quad Y; P(Y = j) = q_j \\ 2.\ \ Gere \quad U \sim U(0;1) \\ 3.\ \ Se \quad U < \dfrac{p_Y}{cq_Y} \Rightarrow X = Y, \quad pare \\ 4.\ \ c.c. \quad retorne \ \ ao \ \ passo \ \ 1 \end{cases}$$

Note que:

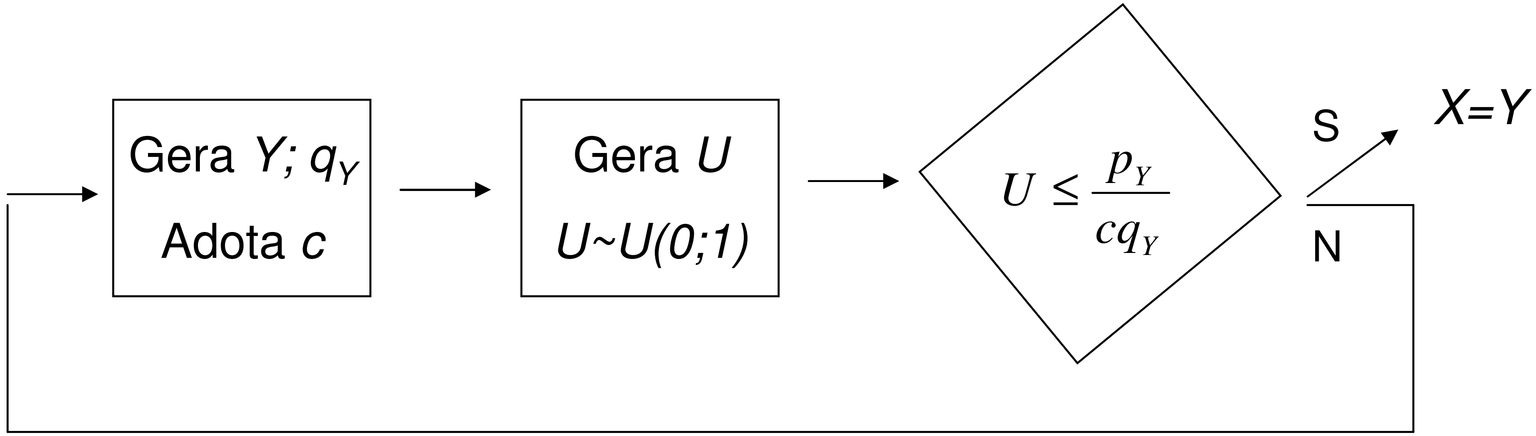
$$\frac{p_j}{q_j} \leq c$$

$$para \ \ todo \ \ j; p_j \neq 0; \ \ c \geq 1$$

Método da Rejeição

Algoritmo:

$$\begin{cases} 1.\ \ \textit{Simule}\quad Y;P(Y=j)=q_j \\ 2.\ \ \textit{Gere}\quad U \sim U(0;1) \\ 3.\ \ \textit{Se}\quad U < \dfrac{p_Y}{cq_Y} \Rightarrow X = Y,\quad \textit{pare} \\ 4.\ \ \textit{c.c.}\quad \textit{retorne}\ \ \textit{ao}\ \ \textit{passo}\ \ 1 \end{cases}$$



Método da Rejeição

<u>Algoritmo:</u>

$$\begin{cases} 1.\ \ Simule \quad Y; P(Y=j)=q_j \\ 2.\ \ Gere \quad U \sim U(0;1) \\ 3.\ \ Se \quad U < \dfrac{p_Y}{cq_Y} \Rightarrow X = Y, \quad pare \\ 4.\ \ c.c. \quad retorne \ \ ao \ \ passo \ \ 1 \end{cases}$$

<u>Prova:</u>

$$P(X=i) = P(Y=i \mid Aceitação) = \frac{P(Y=i; Aceit)}{P(Aceit)}$$

$$= \frac{1}{k} P\left(Y = i; U \le \frac{p_i}{cq_i} \right) = \frac{1}{k} P(Y=i) P\left(U \le \frac{p_i}{cq_i} \right) = \frac{q_i}{k}\frac{p_i}{cq_i} = \frac{p_i}{kc} = p_i$$

$$\Rightarrow 1 = \sum_i P(X=i) = \sum_i \frac{p_i}{kc} = \frac{1}{kc}$$

Método da Rejeição

Simule a v.a. X tal que:

X=x	1	2	3	4	5	6	7	8	9	10
P(X=x)	0,11	0,12	0,09	0,08	0,12	0,1	0,09	0,09	0,1	0,1

Algoritmo:

$$c \equiv \max\left\{\frac{p_Y}{q_Y}\right\}$$

$$\begin{cases} 1.\ \ Gere U_1 \sim U(0;1) \Rightarrow Y = [10U_1]+1 \\ 2.\ \ Escolha \quad c = Max\dfrac{p_Y}{q_Y} = 1{,}2 \\ 3.\ \ Gere \quad U_2 \sim U(0;1) \\ 4.\ \ Se\ \ U_2 < \boxed{\dfrac{p_Y}{0{,}12}} \Rightarrow X = Y, \quad pare \\ 4.\ \ c.c. \quad retorne\ \ ao\ \ passo\ \ 1 \end{cases}$$

$$U < \frac{p_Y}{cq_Y}$$

Método da Rejeição

Simule a v.a. X tal que:

X=x	1	2	3	4	5	6	7	8	9	10
P(X=x)	0,11	0,12	0,09	0,08	0,12	0,1	0,09	0,09	0,1	0,1

Algoritmo:

$$\begin{cases} 1.\ \ Gere U_1 \sim U(0;1) \Rightarrow Y = [10U_1] + 1 \\ 2.\ \ Escolha \quad c = Max \dfrac{p_Y}{q_Y} = 1{,}2 \\ 3.\ \ Gere \quad U_2 \sim U(0;1) \\ 4.\ \ Se \ \ U_2 < \boxed{\dfrac{p_Y}{0{,}12}} \Rightarrow X = Y, \quad pare \\ 4.\ \ c.c. \quad retorne \ \ ao \ \ passo \ \ 1 \end{cases}$$

$$U < \frac{p_Y}{cq_Y}$$

Método da Composição (Mistura)

Suponha que podemos gerar eficientemente X_1 e X_2, tal que:

$$P(X_1 = j) = q_j \qquad P(X_2 = j) = p_j; \qquad j \geq 0$$

Então, para gerar X definida como:

$$P(X = j) = \alpha \; p_j + (1-\alpha)\, q_j; \qquad j \geq 0, \quad 0 < \alpha < 1$$

basta fazer:

$$X = \begin{cases} X_1 & com \quad probabilid\,ade \quad \alpha \\ X_2 & com \quad probabilid\,ade \quad (1-\alpha) \end{cases}$$

Método da Composição

Aplicação: Gerar X tal que,

$$P(X = j) = 0{,}5\ p_j + 0{,}5\ q_j; \qquad j = 1,2,\ldots,10$$

$$p_j = 0{,}10 \qquad j = 1,2,\ldots,10$$

$$q_j = \begin{cases} 0 & j = 1,2,3,4,5 \\ 0{,}2 & j = 6,7,8,9,10 \end{cases}$$

X=x	1	2	3	4	5	6	7	8	9	10
P(X=x)	0,05	0,05	0,05	0,05	0,05	0,15	0,15	0,15	0,15	0,15

Método da Composição

Aplicação: Gerar X tal que,

$$P(X = j) = 0{,}5\ p_j + 0{,}5\ q_j; \qquad j = 1,2,\ldots,10$$

$$p_j = 0{,}10 \qquad j = 1,2,\ldots,10 \qquad\qquad q_j = \begin{cases} 0 & j = 1,2,3,4,5 \\ 0{,}2 & j = 6,7,8,9,10 \end{cases}$$

Algoritmo:

$$\begin{cases} 1.\ \ Gere\ \ U_1 \sim U(0;1) \\ 2.\ \ Gere\ \ U_2 \sim U(0;1) \\ 3.\ \ Se\ \ U_1 < 0{,}5 \Rightarrow X = [10\,U_2] + 1, \quad pare \\ 4.\ \ c.c. \quad X = [5\,U_2] + 6 \end{cases}$$

Simulação Estocástica

<u>Gerando Variáveis Aleatórias Contínuas</u>

$\Rightarrow$ Método da Transformada Inversa

$\Rightarrow$ Método da Rejeição (Aceitação/Rejeição)

$\Rightarrow$ Método Polar (v.a. Normais)

Método da Transformada Inversa

Gerar X, variável aleatória contínua, tal que:

$$P(X \le x) = F_X(x) \qquad \Rightarrow \qquad x = F_X^{-1}$$

<u>ALGORITMO:</u>

$$\begin{cases} \Rightarrow Gerar \quad U \sim U(0;1) \\ \Rightarrow X = F^{-1}(U) \end{cases}$$

Prova:

$$\begin{aligned} F_X(x) &= P(X \le x) = P\left(F^{-1}(U) \le x\right) \\ &= P\left(F(F^{-1}(U)) \le F(x)\right) \\ &= P(U \le F(x)) = F(x) \end{aligned}$$

Método da Transformada Inversa

Gerar X, variável aleatória contínua, tal que:

$$P(X \le x) = F_X(x) = x^n; \qquad 0 < x < 1$$

$$U \sim U(0;1) \quad \Rightarrow \quad x = F^{-1}(u)$$

$$\Rightarrow \quad u = F(x) = x^n \quad \Leftrightarrow \quad x = u^{1/n}$$

Algoritmo:
$$\begin{cases} \Rightarrow Gerar \quad U \sim U(0;1) \\ \Rightarrow X = U^{1/n} \end{cases}$$

Método da Transformada Inversa

Gerar X, variável aleatória com distribuição exponencial de parâmetro λ=1

$$X \sim \exp(\lambda = 1); \qquad f_X(x) = \lambda e^{-\lambda x} = e^{-x}; \qquad x > 0$$

$$P(X \leq x) = F_X(x) = 1 - e^{-x}; \qquad x > 0$$

$$\Rightarrow \quad x = F^{-1}(u) \Rightarrow u = F(x) = 1 - e^{-x} \Rightarrow 1 - u = e^{-x} \quad \Rightarrow \quad x = -\ln(1-u)$$

Algoritmo:
$$\begin{cases} \Rightarrow Gerar \quad U \sim U(0;1) \\ \Rightarrow X = -\ln(1-U) \end{cases}$$

Método da Transformada Inversa

Gerar X, variável aleatória com distribuição exponencial de parâmetro λ=1

$$X \sim \exp(\lambda = 1); \qquad f_X(x) = \lambda e^{-\lambda x} = e^{-x}; \qquad x > 0$$

$$P(X \leq x) = F_X(x) = 1 - e^{-x}; \qquad x > 0$$

Algoritmo: $\begin{cases} \Rightarrow Gerar \quad U \sim U(0;1) \\ \Rightarrow X = -\ln(1-U) \end{cases}$

Algoritmo mais rápido: $\begin{cases} \Rightarrow Gerar \quad U \sim U(0;1) \\ \Rightarrow X = -\ln(U) \end{cases}$

$$Y \sim \exp(\lambda); \qquad f_Y(y) = \lambda e^{-\lambda y}; \qquad y > 0$$

$$P(Y \leq y) = F_Y(y) = 1 - e^{-\lambda y}; \qquad y > 0$$

$$Y = \lambda X \quad \Rightarrow \quad Y = -\frac{1}{\lambda} \ln U$$

Método da Transformada Inversa

Gerando o processo de Poisson a partir da soma de variáveis exponenciais:

$N \sim P(\lambda)$ $\quad$ Y_i : tempos entre chegadas sucessivas (i=1,…)

$\sum_{i=1}^{n} Y_i$: Tempo total até a n-ésima chegada

$$N(1) \sim Poisson(\lambda = 1) \quad \Rightarrow \quad N(1) = \max\left\{ n : \sum_{i=1}^{n} Y_i \leq 1 \right\}$$

Algoritmo:

$$\begin{cases} \Rightarrow Gerar \quad U_i \sim U(0;1), \quad i = 1,2,\ldots,n \\ N = \max\left\{ n; \sum_{i=1}^{n} -\frac{1}{\lambda} \ln U_i \leq 1 \right\} \\ \quad = \max\left\{ n; \sum_{i=1}^{n} \ln U_i \geq -\lambda \right\} = \max\{n; \ln(U_1 * U_2 * \ldots * U_n) \geq -\lambda\} \\ \quad = \max\{n; (U_1 * U_2 * \ldots * U_n) \geq e^{-\lambda}\} \end{cases}$$

Método da Transformada Inversa

Gerando o processo de Poisson a partir da soma de variáveis exponenciais:

Algoritmo mais eficiente:

$$\begin{cases} \Rightarrow Gerar \quad U_i \sim U(0;1), \quad i = 1,2,\ldots,n \\ N = \min\{n; (U_1 * U_2 * \ldots * U_n) < e^{-\lambda}\} - 1 \end{cases}$$

Método da Transformada Inversa

Gerando a distribuição Gama:

$$X \sim Gama(n,\lambda); \qquad f_X(x) = \lambda e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}; \qquad x > 0$$

$$P(X \le x) = F_X(x) = \int_0^x \lambda e^{-\lambda y}\frac{(\lambda y)^{n-1}}{(n-1)!}dy \qquad ?$$

$$X \sim Gama(n,\lambda) \quad \Rightarrow \quad X = \sum_{i=1}^{n} Y_i; \quad \Rightarrow \quad Y_i \sim \exp(\lambda)$$

Algoritmo:

$$\begin{cases} \Rightarrow Gerar \quad U_i \sim U(0;1), \quad i = 1,2,\dots,n \\ X = -\frac{1}{\lambda}\ln U_1 - \dots - \frac{1}{\lambda}\ln U_n = -\frac{1}{\lambda}\ln(U_1 \dots U_n) \end{cases}$$

Método da Rejeição

Suponha que a v.a. Y pode ser gerada eficientemente,

$\Rightarrow Y; \quad g_Y(y)$ Em geral, Y é Uniforme ou Exponencial

A variável X pode ser gerada a partir de Y, como:

$$\Rightarrow X = Y \quad se \quad \frac{f_X(y)}{g_Y(y)} \leq c \quad para \quad todo \quad y$$

Algoritmo:

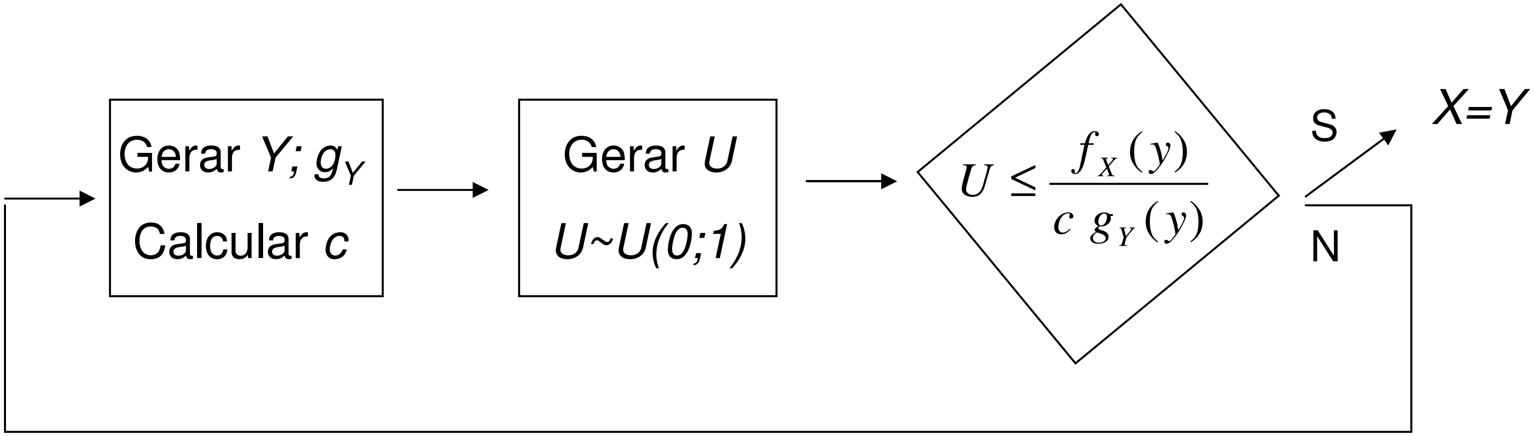
$$\begin{cases} 1.\, Gerar \quad Y; \quad g_Y(y) \\ 2.\, Gerar \quad U \sim U(0,1) \\ 3.\, Se \quad U \leq \dfrac{f_X(y)}{c\, g_Y(y)} \Rightarrow X = y \\ 4.\, c.c. \quad retornar\ ao\ passo\ 1 \end{cases}$$

Método da Rejeição

Algoritmo:

$$\begin{cases} 1.\ Gerar \quad Y; \quad g_Y(y) \\ 2.\ Gerar \quad U \sim U(0,1) \\ 3.\ Se \quad U \le \dfrac{f_X(y)}{c\ g_Y(y)} \Rightarrow X = y \\ 4.\ c.c. \quad retornar\ ao\ passo\ 1 \end{cases}$$

$$c \equiv \max\left\{\frac{f_X}{g_Y}\right\}$$



$\Rightarrow$ O número de iterações necessárias é uma variável exponencial com média c

Método da Rejeição

Aplicação: Gerar X, tal que, $f_X(x) = 20\,x(1-x)^3 \qquad 0 < x < 1$

$$\Rightarrow Seja \quad Y \sim U(0;1) \Rightarrow g_Y(y) = 1 \quad 0 < y < 1$$

$$\Rightarrow c? \qquad \max \frac{f_X(y)}{g_Y(y)} = \max\left\{20x(1-x)^3\right\} \qquad \begin{cases} 20(1-x)^3 - 20x\left[3(1-x)^2\right] = 0 \\ \Rightarrow x = 1/4 \end{cases}$$

Diferenciar e igualar a zero

$$\Rightarrow c \equiv \frac{f_X(1/4)}{g_Y(1/4)} = 20\frac{1}{4}\left(\frac{3}{4}\right) = \frac{135}{64}$$

$$\Rightarrow \frac{f_X(y)}{c\; g_Y(y)} = \frac{135}{64}\left\{20x(1-x)^3\right\} = \frac{256}{27}x(1-x)^3$$

Método da Rejeição

Aplicação: Gerar X, tal que, $f_X(x) = 20\,x\,(1-x)^3 \qquad 0 < x < 1$

$$\Rightarrow Y \sim U(0;1) \Rightarrow \frac{f_X(y)}{c\ g_Y(y)} = \frac{135}{64}\left\{20x(1-x)^3\right\} = \frac{256}{27}x(1-x)^3$$

Algoritmo:

$$\begin{cases} 1.\ Gerar \quad U_1 \sim U(0;1) \\ 2.\ Gerar \quad U_2 \sim U(0;1) \\ 3.\ Se \quad U_2 \leq \dfrac{256}{27} U_1 (1-U_1)^3 \Rightarrow X = U_1 \\ 4.\ c.c. \quad retornar\ ao\ passo\ 1 \end{cases}$$

Método da Rejeição

<u>Aplicação</u>: $Gerar \quad X \sim Gama\left(\frac{3}{2};1\right) \Rightarrow f_X(x) = e^{-x}\frac{(x)^{\frac{3}{2}-1}}{\Gamma\left(\frac{3}{2}\right)} \qquad x > 0$

$$\Rightarrow Seja \quad Y \sim \exp\left(\lambda = \tfrac{2}{3}\right) \Rightarrow g_Y(y) = \frac{2}{3}e^{-\frac{2}{3}y} \qquad y > 0$$

$$\Rightarrow c\,? \quad \max\frac{f_X(x)}{g_Y(x)} = \max\left\{\frac{1/\Gamma(3/2)\;x^{1/2}e^{-x}}{3/2\;e^{-(2/3)x}}\right\} = \max\left\{\frac{3}{2\Gamma(3/2)}x^{1/2}e^{-(1/3)x}\right\}$$

$$\Rightarrow x = 3/2 \qquad \Rightarrow c \equiv \frac{3}{2\Gamma(3/2)}\left(\frac{3}{2}\right)^{1/2} e^{-(1/3)(3/2)}$$

$$\Rightarrow \frac{f_X(x)}{c\;g_Y(x)} = \left(\frac{2e}{3}\right)^{1/2} x^{1/2}e^{-x/3}$$

Método da Rejeição

Aplicação: $Gerar \quad X \sim Gama\left(\frac{3}{2};1\right) \Rightarrow f_X(x) = e^{-x}\frac{(x)^{\frac{3}{2}-1}}{\Gamma\left(\frac{3}{2}\right)} \qquad x > 0$

$$\Rightarrow Y \sim \exp\left(\lambda = \tfrac{2}{3}\right) \Rightarrow \frac{f_X(x)}{c\ g_Y(x)} = \left(\frac{2e}{3}\right)^{1/2} x^{1/2} e^{-x/3}$$

Algoritmo:

$$\begin{cases} 1.\ Gerar \quad U_1 \sim U(0;1) \Rightarrow Y = -\tfrac{3}{2}\ln U_1 \\ 2.\ Gerar \quad U_2 \sim U(0;1) \\ 3.\ Se \quad U_2 \leq \left(\dfrac{2e}{3}\right)^{1/2} Y^{1/2} e^{-Y/3} \Rightarrow X = Y, \quad pare \\ 4.\ c.c. \quad retornar\ ao\ passo\ 1 \end{cases}$$

$\Rightarrow$ Neste caso, o número de iterações necessárias é c=1,257

Método da Rejeição

Aplicação: $Gerar \quad Z \sim N(0;1) \Rightarrow f_Z(z) = \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty$

$Gerar \quad |Z|; \quad f_{|Z|}(x) = \frac{2}{(2\pi)^{1/2}} e^{-\frac{1}{2}x^2}; \quad x > 0$

$\Rightarrow$ Obter Z, tal que,
P(Z=x)=P(Z=-x)=0,5

$$\Rightarrow Y \sim \exp(\lambda = 1); \quad g_Y(y) = e^{-y} \quad y > 0$$

$$\Rightarrow c\,? \quad \max \frac{f_{|Z|}(x)}{g_Y(x)} = \max\left\{(2/\pi)^{1/2} e^{x - x^2/2}\right\} \Rightarrow x = 1 \quad \Rightarrow c \equiv (2/\pi)^{1/2} e^{1-1/2}$$

$$\Rightarrow \frac{f_X(x)}{c\, g_Y(x)} = \exp\left\{x - \frac{x^2}{2} - \frac{1}{2}\right\} = \exp\left\{-\frac{(x-1)^2}{2}\right\}$$

#médio de iterações

Método da Rejeição

<u>Aplicação</u>: $Gerar \quad Z \sim N(0;1) \Rightarrow f_Z(z) = \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty$

$Gerar \quad X = |Z|; \quad f_X(x) = \frac{2}{(2\pi)^{1/2}} e^{-\frac{1}{2}x^2}; \quad x > 0$ $\Rightarrow$ Obter Z, tal que, P(Z=x)=P(Z=-x)=0,5

Algoritmo:

$$\begin{cases} 1.\ Gerar \quad U_1 \sim U(0;1) \Rightarrow Y = -\ln U_1 \\ 2.\ Gerar \quad U_2 \sim U(0;1) \\ 3.\ Se \quad U_2 \leq \exp\left\{-\frac{(Y-1)^2}{2}\right\} \Rightarrow X = Y,\ vá\ para\ passo\ 5 \\ 4.\ c.c. \quad retornar\ ao\ passo\ 1 \\ 5.\ Gerar \quad U_3 \sim U(0;1) \Rightarrow Se\ \ U_3 \leq 0{,}5 \Rightarrow Z = X,\ pare \\ 6.\ c.c. \quad Z = -X,\ pare \end{cases}$$

Método da Rejeição

Um algoritmo mais eficiente pode ser obtido, por notar que:

$$
\begin{cases}
1.\ Gerar \quad U_1 \sim U(0;1) \Rightarrow Y = -\ln U_1 \\
2.\ Gerar \quad U_2 \sim U(0;1) \\
\boxed{3.\ Se \quad U_2 \leq \exp\left\{-\dfrac{(Y-1)^2}{2}\right\} \Rightarrow X = Y,\ vá\ para\ passo\ 5} \\
4.\ c.c. \quad retornar\ ao\ passo\ 1 \\
5.\ Gerar \quad U_3 \sim U(0;1) \Rightarrow Se\ \ U_3 \leq 0{,}5 \Rightarrow Z = X,\ pare \\
6.\ c.c. \quad Z = -X,\ pare
\end{cases}
$$

$$
-\ln U_2 \geq \frac{(Y-1)^2}{2} \quad ; \quad -\ln U_2 \sim \exp(\lambda = 1)
$$

Método da Rejeição

<u>Aplicação</u>: $Gerar \quad Z \sim N(0;1) \Rightarrow f_Z(z) = \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty$

Algoritmo mais eficiente:

$$\begin{cases} 1.\ Gerar \quad U_1 \sim U(0;1) \Rightarrow Y_1 = -\ln U_1 \\ 2.\ Gerar \quad U_2 \sim U(0;1) \Rightarrow Y_2 = -\ln U_2 \\ 3.\ Se \quad Y_2 \geq \frac{(Y_1 - 1)^2}{2} \Rightarrow X = Y_1, vá\ para\ passo 5 \\ 4.\ c.c \quad retornar ao\ passo 1 \\ 5.\ Gerar \quad U_3 \sim U(0;1) \Rightarrow Se U_3 \leq 0{,}5 \Rightarrow Z = X, pare. \\ 6.\ c.c. Z = -X, pare. \end{cases}$$

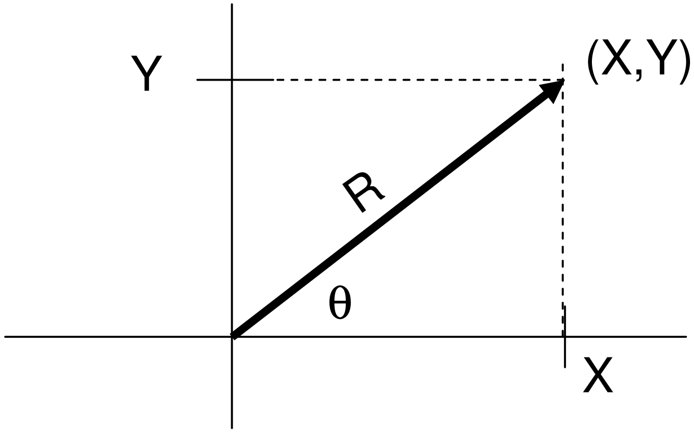
Simulação de Variáveis Normais

$$Gerar \quad X \sim N\left(\mu;\sigma^2\right) \Rightarrow f_X(x) = \frac{1}{(2\pi\sigma)^{1/2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

$$\begin{cases} Gerar \quad Z \sim N(0;1) \Rightarrow f_Z(z) = \dfrac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty \\ \\ Obter \quad X; \qquad X = \sigma\, Z + \mu \end{cases}$$

Método Polar para Gerar Normais

Coordenadas cartesianas e coordenadas polares:



$$\Rightarrow \begin{cases} R^2 = X^2 + Y^2 \\ \tan\theta = \dfrac{Y}{X} \qquad \theta = \tan^{-1}\left(\dfrac{Y}{X}\right) \end{cases}$$

$$\Rightarrow \begin{cases} X = R\cos\theta \\ Y = R\ sen\theta \end{cases}$$

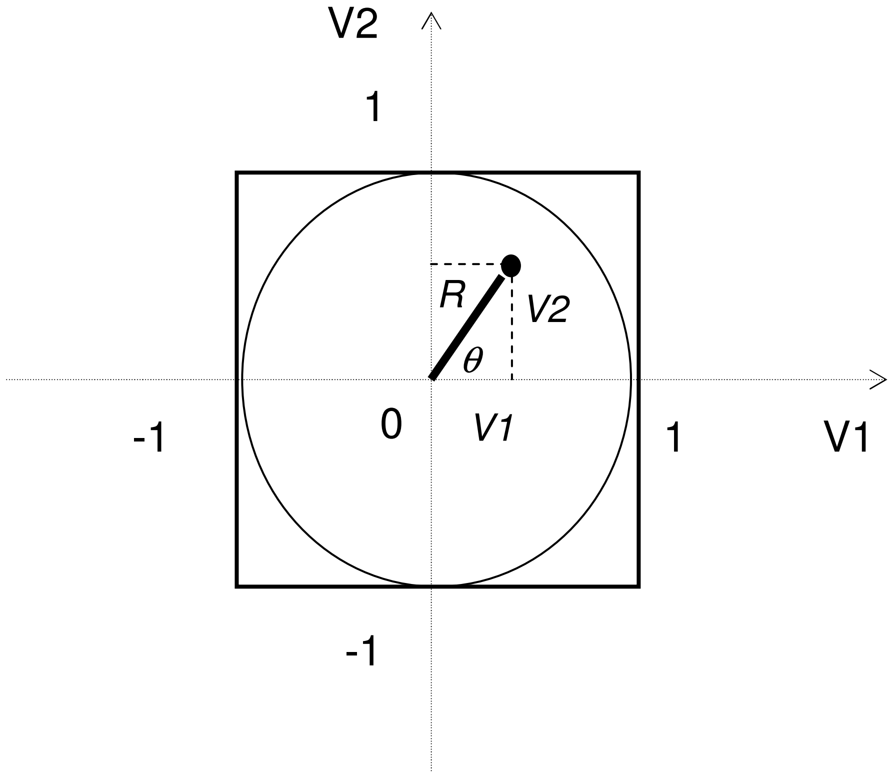
Transformação de Box-Miller: Obter variáveis Normais padrão independentes (X e Y) a partir de variáveis Uniformes

$$\Rightarrow \quad X = R\cos\theta = \sqrt{-2\ln U_1}\cos(2\pi U_2)$$

$$Y = R\ sen\theta = \sqrt{-2\ln U_1}\, sen(2\pi U_2)$$

Simulação: limitações com o cálculo de senos e cossenos

Método Polar para Gerar Normais



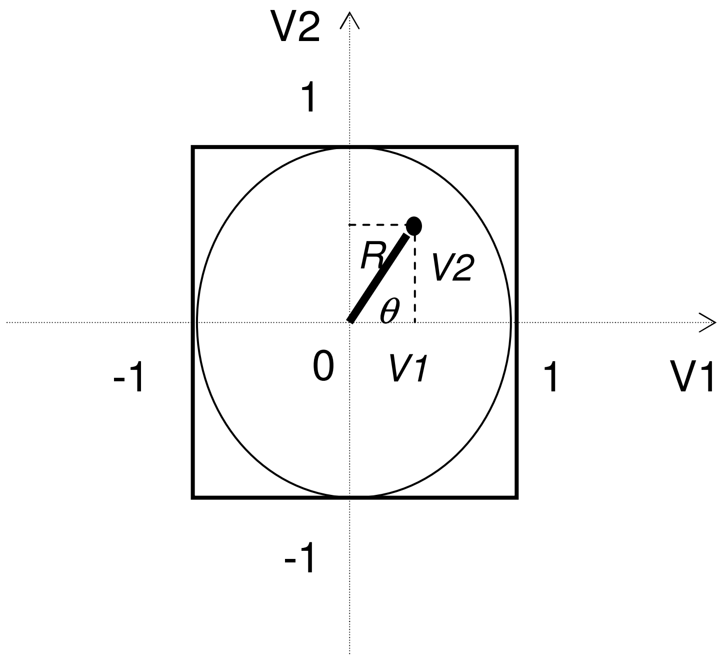
$$\Rightarrow \quad X = R\cos\theta$$

$$Y = R\, sen\theta$$

$$\begin{cases} R = \left(V_1^{\,2} + V_2^{\,2}\right)^{1/2} & S = R^2 \\ sen\theta = \dfrac{V_2}{R} & \cos\theta = \dfrac{V_1}{R} \end{cases}$$

$$\Rightarrow \quad X = V_1\left(\frac{-2\ln S}{S}\right)^{1/2} \qquad \Rightarrow \quad Y = V_2\left(\frac{-2\ln S}{S}\right)^{1/2}$$

Método Polar



Pontos (v1,v2) estão distribuídos uniformemente no círculo de raio unitário. Logo:

$$\Rightarrow \quad U_1 \sim U(0;1)$$
$$V_1 = 2U_1 - 1; \quad V_1 \sim U(-1;1)$$

$$\Rightarrow \quad U_2 \sim U(0;1)$$
$$V_2 = 2U_2 - 1; \quad V_2 \sim U(-1;1)$$

$$\Rightarrow \quad X = V_1\left(\frac{-2\ln S}{S}\right)^{1/2} \qquad Y = V_2\left(\frac{-2\ln S}{S}\right)^{1/2} \qquad S = V_1^2 + V_2^2$$

X e Y são variáveis aleatórias independentes com distribuição Normal Padrão:

$$X \sim N(0;1) \qquad Y \sim N(0;1)$$

Validação Estatística de Simulação Estocástica

Geração de Variáveis Aleatórias Discretas:

$\Rightarrow$ Gráficos Q-Q

$\Rightarrow$ Testes de Bondade de Ajuste (Aderência): testes Qui-Quadrado

Geração de Variáveis Aleatórias Contínuas:

$\Rightarrow$ Gráficos Q-Q

$\Rightarrow$ Testes de Bondade de Ajuste (Aderência): categorizando a variável em pequenos intervalos

$\Rightarrow$ Testes de Kolmogorov-Smirnov

Validação Estatística de Simulação Estocástica

$\Rightarrow$ Testes de Bondade de Ajuste (Aderência):

$$Q = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{K} \frac{(N_i - np_i)^2}{np_i} \quad \sim \quad \chi^2_{K-1}$$

O_i: frequência observada (N_i) na “casela” *I*

E_i: frequência esperada (np_i) de observações na “casela” *I*

n observações foram geradas

K valores foram observados da variável nas n simulações

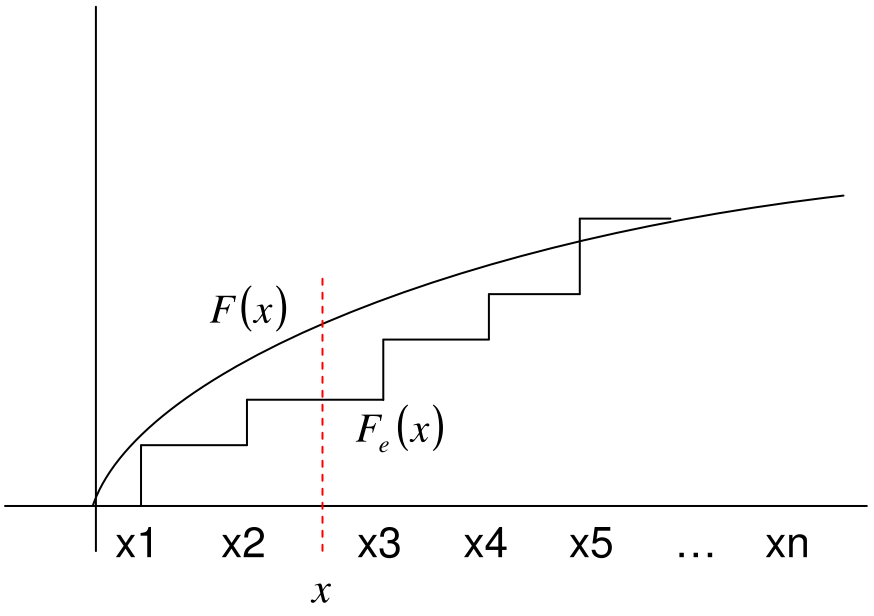
Validação Estatística de Simulação Estocástica

⇒ Testes Kolmogorov-Smirnov:

$$H_0 : F_e(x) = F(x)$$

$$D = \max_x \left| F_e(x) - F(x) \right|$$

$$F_e(x) = \begin{cases} 0 \quad se \quad x < x_1 \\ \dfrac{1}{n} \quad se \quad x_1 \leq x < x_2 \\ \dots \\ \dfrac{j}{n} \quad se \quad x_j \leq x < x_{j+1} \\ \dots \\ 1 \quad se \quad x_n \leq x \end{cases}$$



Valores ordenados das simulações

Validação Estatística de Simulação Estocástica

$\Rightarrow$ Testes Kolmogorov-Smirnov:

$$D = \max_x \left| F_e(x) - F(x) \right|$$

$$D = \max\left\{ \frac{j}{n} - F(x_j), F(x_j) - \frac{j-1}{n}; j = 1,\ldots,n \right\}; \quad x_1 < x_2 < \ldots < x_n$$

Suponha que para o conjunto de n valores simulados $\Rightarrow$ *D=d*

Cálculo do nível descritivo do teste:

$$p = P(D \geq d)$$

Pode ser mostrado que a distribuição da estatística D "não depende de F"

$\Rightarrow$ logo o valor p é calculado via simulação a partir da distribuição Uniforme como veremos a seguir.

Validação Estatística de Simulações

<u>Aplicação</u>: Verifique se os dados a seguir podem ser assumidos como tendo uma distribuição exponencial com média 100:

66 72 94 112 116 124 140 145 155

$$F(x) = 1 - e^{-x/100} \qquad \Rightarrow \text{d=0,4831}$$

Cálculo do nível descritivo do teste de Kolmogorov-Smirnov:

Algoritmo:

$$\begin{cases} 1.\ n = 9 \\ 2.\ Gerar: \quad U_1, \ldots, U_n \\ 3.\ Ordenar: \quad U_{(1)}, \ldots, U_{(n)} \\ 3.\ D = \max\left\{\frac{1}{n} - U_{(1)}, \frac{2}{n} - U_{(2)}, \ldots, \frac{j}{n} - U_{(j)}, \ldots, 1 - U_{(1)}\right\} \\ 4.\ retorne\ ao\ passo\ 1\ K\ vezes \end{cases}$$

$p = P(D \geq d)$: proporção dos K valores que são maiores ou iguais a d

Bibliografia

- Ross, S.M. (1997). *Simulation*. Academic Press.