(A-R)

## 1.3. Método da Aceitação-Rejeição (Newmann, 1951)

Uma Prova deste Algoritmo é Apresentada em Flury (1990) com base em 3 Lemas

Flury, Bernard (1990). Acceptance-Rejection Sampling made easy. SIAM REVIEW, 32, pg. 474-476

Vamos ver este algoritmo para simular valores de uma v.a. X contínua.

- X v.a. que se pretende simular, e que é "difícil", com f.d.p. $f(x)$
- Y outra v.a. fácil de simular com f.d.p. $g(x)$ (Normalmente usa-se a Uniforme ou Exponencial) v.a. positiva

Vamos Admitir que:

(1) $f_X(x) \leq g_Y(x) \cdot c$, $\forall x$ onde c é uma constante e $c \geq 1$

$$\Leftrightarrow 0 \leq \frac{f(x)}{c\, g(x)} \leq 1$$

**Algoritmo: (A-R)** Seja Y uma v.a. com f.d.p. $g(x)$ e U uma v.a. $U(0,1)$ com $U \perp\!\!\!\perp Y$. Assume-se que (1) é válido

**Passo 1:** obter um valor y da v.a. Y (candidato a ser Aceite/rejeitado para valor da v.a. X)

**Passo 2:** obter um valor $u$ da $U(0,1)$

**Passo 3:** Declarar y como um valor da v.a. X se

$$u \leq \frac{f(y)}{c\, g(y)} \quad (2) \quad \text{tomando } c = \max_x \frac{f(x)}{g(x)}$$

caso contrário voltar Passo 1.

Observações:

1. Na forma Aleatória temos $U \leq \frac{f(Y)}{g(Y)}$ e como $U \perp\!\!\!\perp Y$ então $U \perp\!\!\!\perp \frac{f(Y)}{c\,g(Y)}$ e $0 \leq \frac{f(Y)}{c\,g(Y)} \leq 1$

2. Dado $Y=y$, a prob. de o candidato $y$ a ser aceite para $x$ é:

$$P\left(U \leq \frac{f(Y)}{c\,g(Y)} \,\middle|\, Y=y\right) \underset{U \sim U(0,1)}{=} \frac{f(y)}{c\,g(y)} = P(\text{'Aceite'} \mid Y=y)$$

3. O nº de iterações dos passos 1 e 2 até obter um valor de $y$ Aceite para $x$ é v.a. Geométrica com parâmetro $p = P(\text{'Aceite'})$

Qual é $p = P(\overset{A}{\overline{\text{'Aceite'}}})$ ?

- Sabemos que $P(\text{'Aceite'} \mid Y=y) = \frac{f(y)}{c\,g(y)}$ e queremos $P(\text{'Aceite'})$

---

$$E[X] = E[E[X|Y]] = \begin{cases} \sum_y E[X|Y=y]\,P(Y=y) & Y \text{ v.a. discreta} \\ \int_{-\infty}^{+\infty} E[X|Y=y] \cdot f_Y(y)\,dy & Y \text{ v.a. contínua} \end{cases}$$

$\forall (X,Y)$

Considerar A um Acontecimento e X = 'v.a. Indicadora de realização de A'

$X \sim Ber(P(A))$

$E(X) = P(A)$ e $E[X|Y=y] = P(A|Y=y)$, $\forall$ v.a. $Y$

$$E(X) = P(A) = E[E[X|Y]] \underset{Y \text{ v.a. contínua}}{=} \int_{-\infty}^{+\infty} E[X|Y=y] \cdot f_Y(y)\,dy$$

$$= \int_{-\infty}^{+\infty} P(A|Y=y)\, f_Y(y) \cdot dy =$$

$= E[P(A|Y)]$

---

$$P(A) = P(\text{'Aceite'}) = E[P(\text{'Aceite'} \mid Y)] = \int_{-\infty}^{+\infty} \frac{f(y)}{c\,\cancel{g(y)}} \cdot \cancel{g(y)}\, dy$$

$$= \frac{1}{c} \underbrace{\int_{-\infty}^{+\infty} f(y)\, dy}_{1} = \frac{1}{c}, \quad c > 1$$

$W =$ 'v.a. nº de iterações até aceitar um valor y para x'

$W \sim Geo\left(\frac{1}{c}\right)$

- Este algoritmo funciona? Isto é, a dist. de $Y$ condicional a $U \leq \frac{f(Y)}{cg(Y)}$ é "igual" à dist. de $X$?

$$Y \mid \left(U \leq \frac{f(Y)}{cg(Y)}\right) \overset{?}{\underset{d}{=}} X$$

Vamos ver que sim!

temos que ver se $P(\underbrace{Y \leq y}_{A} \mid \overbrace{U \leq \frac{f(Y)}{cg(Y)}}^{B}) = F(y)$ onde

$F(\cdot)$ é a função dist. de $X$ e $G(\cdot)$ a f. dist. de $Y$

$P(A|B) = ?$ é mais fácil de calcular $P(B/A)$

$$P(A|B) = \frac{P(B/A) \cdot P(A)}{P(B)}$$

$$P(B|A) = P\left(U \leq \frac{f(Y)}{cg(Y)} \,\middle|\, Y \leq y\right) = \frac{P\left(U \leq \frac{f(Y)}{cg(Y)}, Y \leq y\right)}{P(Y \leq y)} =$$

$$= \frac{P\left(U \leq \frac{f(Y)}{cg(Y)}, Y \leq y\right)}{G(y)} =$$

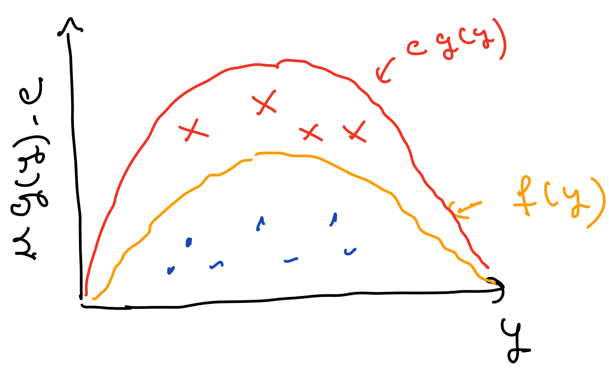
$$\underset{\text{condicionando em } Y}{\overset{\downarrow}{=}} \frac{1}{G(y)} \int_{-\infty}^{y} \underbrace{P\left(U \le \frac{f(y)}{c\,g(y)} \,\middle|\, Y=w\right)} g(w)\,dw =$$

$$= \frac{1}{G(y)} \underbrace{\int_{-\infty}^{y} \frac{f(w)}{c\,\cancel{g(w)}}\, \cancel{g(w)}\,dw} = \frac{F(y)}{c\,G(y)}$$

$$P(B/A) = \frac{F(y)}{c\,G(y)}$$

$$P(A|B) = \frac{P(B/A) \cdot P(A)}{P(B)} = \frac{F(y)}{c\,G(y)} \frac{G(y)}{\frac{1}{c}} = F(y)$$

$$\therefore\ Y \,\Big|\, \left(U \le \frac{f(y)}{c\,g(y)}\right) \overset{d}{=} X$$




• Aceitos: $u\,g(y).c \le f(y)$

× ñ Aceitos: $u\,g(y).c > f(y)$

pontos $(y,\ u\,g(y).c)$

Ex 1: Usar o Algoritmo A-R para gerar valores da dist. $N(\mu, \sigma^2)$. $X \sim N(\mu; \sigma^2)$

como $X = \sigma Z + \mu$ com $Z \sim N(0,1)$ é suficiente ter um Algoritmo para gerar valores de $Z$.

Como no método da A-R a v.a. Y Auxiliar (candidata) é uma v.a. positiva, vamos pensar em gerar $|Z|$ e depois por simetria obtem-se valores de $Z$ $(P(Z<0) = P(Z>0) = 1/2)$

$Z \sim N(0,1)$ ; $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, $z \in \mathbb{R}$

e $|Z| = X$ ; $f_X(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, $x \in \mathbb{R}^+$

Y v.a. auxiliar (positiva). Uma hipótese é considerar $Y: E(Y) = E(X)$ ou $\sigma_Y = \sigma_X$ ou uma v.a. cuja f.d.p. faça simplificar $\frac{f(x)}{g(x)}$

vamos escolher $Y \sim Exp(1)$ : $g(x) = e^{-x}$, $x \in \mathbb{R}^+$

$$h(x) = \frac{f(x)}{g(x)} = \frac{2}{\sqrt{2\pi}} e^{\underline{-\frac{x^2}{2} + x}}$$

x maximiza $h(x)$

$$\max_x h(x) = \max_x \left\{x - \frac{x^2}{2}\right\} \Rightarrow x = 1$$

$$c = h(x)\Big|_{x=1} = \frac{f(1)}{g(1)} = \frac{2}{\sqrt{2\pi}} e^{1-\frac{1}{2}} = \frac{2}{\sqrt{2\pi}} e^{1/2}$$

$$c = \frac{2}{\sqrt{2\pi}} e^{1/2}$$

Assim $\frac{f(x)}{c\,g(x)} = e^{x - \frac{x^2}{2} - 1/2} = e^{-\frac{(x-1)^2}{2}}$

Assim $|Z|$ é gerado com:

1. Gerar um $y$ da dist. Exp(1) i.e, $y = -\ln u_1$ com $u_1$ um valor da U(0,1)

2. Gerar um nº $u_2$ da U(0,1)

3. Se $u_2 \le \frac{f(y)}{c\,g(y)} = e^{-\frac{(y-1)}{2}} \Rightarrow x = y$, c.c. voltar passo 1

4. Gerar $u_3$ da U(0,1) se $u_3 \le 0.5 \Rightarrow z = x$ parar
c.c. $z = -x$ parar

Ex 2. X v.a. discreta com f. prob.

$$p_i = P(X=i) = \begin{cases} 0.11, & i=1 \\ 0.12, & i=2,5 \\ 0.09, & i=3,7,8 \\ 0.08, & i=4 \\ 0.1, & i=6,9,10 \\ 0, & \text{c.c.} \end{cases}$$

$R_X = \{1, 2, \dots, 10\}$

Escolher Y : $R_Y = R_X$

Escolher $Y \sim \text{Unif}\{1, 2, \dots, 10\}$

$Y : P(Y=j) = q_j = \frac{1}{10},\ j = 1, \dots, 10$

Encontrar $c$ $c = ?$ $\max \frac{p_j}{q_j} = \frac{0.12}{0.1} = 1.2$

$c = 1.2$

Algoritmo:

Passo 1: Gerar $y = [10 \times u_1] + 1$

Passo 2: Gerar $u_2$

Passo 3: se $u_2 \le \frac{P(X=y)}{1.2\, P(Y=y)} = \frac{P(X=y)}{1.2 \times 0.1}$ então $x = y$ e parar, c.c. voltar passo 1.

Ex 3. Gerar observações de $X \sim G(\overset{\alpha,\ \lambda}{3/2, 1})$

$E(X) = \frac{\alpha}{\lambda} = 3/2$ ; $f(x) = \frac{e^{-x} x^{1/2}}{\Gamma(3/2)} = \frac{e^{-x} x^{1/2}}{\frac{\sqrt{\pi}}{2}}, x > 0$

v.a. Y Auxiliar (candidata): $Y \sim \text{Exp}(2/3)$ ; $E(Y) = E(X)$

$g(x) = 2/3\, e^{-2/3 x}, x > 0$

$c = ?$ $\quad \max_x \frac{f(x)}{g(x)} = \frac{\frac{2\, e^{-x} x^{1/2}}{\sqrt{\pi}}}{2/3\, e^{-2/3 x}} = \max_x \frac{3\, x^{1/2} e^{-x/3}}{\sqrt{\pi}}$

o máximo é obtido em $x = 3/2$

Assim $c = \left.\frac{f(x)}{g(x)}\right|_{x=3/2} = \frac{3^{3/2}}{(2\pi e)^{1/2}}\, e$

$$\frac{f(x)}{c\, g(x)} = \left(\frac{2e}{3}\right)^{1/2} x^{1/2} e^{-x/3} = \left(\frac{2ex}{3}\right)^{1/2} e^{-x/3}$$

Algoritmo e

1. gerar $y = -3/2 \ln u_1$ (valor de $Y \sim Exp(2/3)$)
2. gerar $u_2$
3. Se $u_2 \leq \left(\frac{2ey}{3}\right)^{1/2} e^{-y/3}$ então $x = y$, c.c. voltar a 1

## 1.4. Uso de simulação em teste de hipóteses. Cálculo do valor-p por simulação.

como simular o valor-p num teste de hipóteses?

vamos supor de $X \sim F_X(x; \theta)$ conhecida a menos do parâmetro $\theta$.

<u>objectivo</u>: Inferência para o parâmetro $\theta$

teste de hipóteses:

$H_0: \theta \leq \theta_0$ vs $H_1: \theta > \theta_0$

considerar a est. de teste $T_0 = g(X_1, \ldots, X_n)$ onde $(X_1, \ldots, X_n)$ a.a. de $X$

supor que a Região de Rejeição é a cauda da direita da dist. de est. $T_0$, ie, rej $H_0$ se $T_0 \geq a$

valor-p = $P(T_0 \geq t_0)$ $\qquad T_0 \sim ?$

<u>Alternativa</u>: Simulação de Monte Carlo

$\underset{\sim}{x}^{(1)} = (x_1^{(1)}, \dots, x_n^{(1)}) \longrightarrow t_0^{(1)}$ $\qquad$ Amostra têm que ser independentes

$\vdots$

$\underset{\sim}{x}^{(r)} = (x_1^{(r)}, \dots, x_n^{(r)}) \longrightarrow t_0^{(r)}$

o valor-p do teste pode ser aproximado a:

$$\text{valor-p} \simeq \frac{n^{\underline{o}}\, j : T_0^{(j)} \geq a}{r} \qquad r \text{ grande}$$

$I\{T_0^{(i)} > a\}, \; i = 1, \dots, r$

## 1.5 - técnicas de redução de variância

<u>objectivo</u>: melhorar a eficiência de uma simulação

vamos supor que se pretende estimar $\theta = E[\underbrace{h(X)}_{Y}] = E(Y)$

onde $X$ é uma v.a.. Um Algoritmo usual e simples (simulação de Monte-Carlo usual) é:

- Gerar $x_1, \dots, x_n$
- Estimar $\theta$ usando o estimar $\hat{\theta} = \frac{\sum_{i=1}^{n} Y_i}{n}$ onde $Y_i = h(x_i)$
- Um I.C. para $\theta$, aproximado a $(1-\alpha) \times 100\%$, é:

$$\left( \hat{\theta} - \Phi^{-1}(1-\alpha/2) \frac{\hat{\sigma}}{\sqrt{n}}, \; \hat{\theta} + \Phi^{-1}(1-\alpha/2) \frac{\hat{\sigma}}{\sqrt{n}} \right) \text{ ou } (\text{t.e.e})$$

$\hat{\sigma}$ estimador $\sqrt{\text{var}(Y)} = \sigma_Y$

A qualidade da estimação intervalar pode ser medida usando a metade do comprimento do I.C.. Para $\alpha$ fixo, tem-se

$$\varepsilon = \Phi^{-1}(1-\alpha/2)\sqrt{\frac{var(Y)}{n}} \quad , \ \varepsilon = \text{'precisão'}$$

O que se pode fazer para melhorar a eficiência de uma simulação?

- Desenvolver um bom Algoritmo de simulação minimizando o tempo de execução e os valores de Armazenamento;
- Diminuir a variabilidade dos resultados de simulação utilizados na estimação de $\theta$. Para isso deve-se utilizar métodos de redução de variância

Porquê reduzir a variância dos valores simulados?

vamos supor $\exists$ v.a's $W$ e $Z$ : $E(W) = E(Z) = \theta$
(estimadores centrados de $\theta$)

e o nosso objectivo é estimar $\theta$.

O que devemos fazer para estimar $\theta$; Simular valores de $W$ ou de $Z$?

vamos denotar por: $\begin{cases} M_W - \text{Método de Simulação de } \theta \text{ usando } W_i \\ M_Z - \quad " \quad\quad " \quad " \quad " \quad Z_i \end{cases}$

Qual é mais eficiente / preciso em termos da medida $\varepsilon$?

Seja $n_W$ e $n_Z$ as dimensões das Amostras de $W$ e de $Z$ para conseguir a precisão $\varepsilon$. temos que

$$n_W = \left(\frac{\Phi^{-1}(1-\alpha/2)}{\varepsilon}\right)^2 var(W)$$

$$n_Z = \left(\frac{\Phi^{-1}(1-\alpha/2)}{\varepsilon}\right)^2 var(Z)$$

$E_W$ e $E_Z$ o "esforço" computacional requerido para produzir as Amostras de dimensão $n_W$ e $n_Z$ das v.a's $W$ e $Z$

usando $M_W$ : $n_W \times E_W = \left( \frac{\Phi^{-1}(1-\alpha/2)}{\varepsilon} \right)^2 . var(W) . E_W$

" $M_Z$ : $n_Z \times E_Z = \left( \frac{\Phi^{-1}(1-\alpha/2)}{\varepsilon} \right)^2 . var(Z) . E_Z$

o Método $M_W$ é mais eficiente que $M_Z$ se

$$n_W \times E_W < n_Z E_Z$$

Admitindo q: $E_W \simeq E_Z$ queremos que $var(W) < var(Z)$ !!

Queremos reduzir a variância do estimador!

vamos ver 4 técnicas de redução de variância

(1) usando variáveis Antitéticas

(2) usando uma variável de controle

(3) redução de variância por condicionamento

(4) Amostragem por Importância

## (1) Variáveis Antitéticas

Propõe a redução de variância através da introdução de uma correlação negativa entre as estimativas.

Supor que se pretende estimar $\theta = E[h(X)] = E(Y)$ e se forem geradas 2 Amostras identicamente distribuídas mas não Independentes, i.e., $\{Y_{1i}\}_{i=1,\dots,n}$ e $\{Y_{2i}\}_{i=1,\dots,n}$ que conduzirão aos estimadores $\overline{Y}_1$ e $\overline{Y}_2$ de $\theta$

Um possível estimador centrado de $\theta$ é:

$$\hat{\theta} = \frac{\bar{Y}_1 + \bar{Y}_2}{2}$$

$$var(\hat{\theta}) = var\left(\frac{\bar{Y}_1+\bar{Y}_2}{2}\right) = \frac{1}{4}\left[var(\bar{Y}_1) + var(\bar{Y}_2) + 2\underbrace{cov(\bar{Y}_1,\bar{Y}_2)}\right]$$

$$cov(\bar{Y}_1,\bar{Y}_2) = \rho\sqrt{var(\bar{Y}_1)var(\bar{Y}_2)}$$

$\bar{Y}_1 \not\perp\!\!\!\perp \bar{Y}_2$

$$\underset{i.d.}{Y_1 e Y_2} = \rho\, var(\bar{Y}_1)$$

$$var(a_1X_1 + a_2X_2) =$$
$$= a_1^2 var(X_1) + a_2^2 var(X_2) + 2a_1a_2 cov(X_1,X_2)$$
$\forall a_1, a_2 \in \mathbb{R}$

$$var(\bar{Y}_1) \underset{i.d.}{=} var(\bar{Y}_2)$$

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X\sigma_Y} =$$
$$= cov\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right)$$
$$cov(X-a) = cov(X)$$

$$var(\hat{\theta}) = \frac{2\,var(\bar{Y}_1) + 2\rho\, var(\bar{Y}_1)}{4}$$
$$= \frac{1}{2} var(\bar{Y}_1)[1+\rho] =$$
$$var(\bar{Y}) = \frac{var(Y)}{n}$$
$$= \frac{1}{2n} var(Y_1)(1+\rho)$$

Assim, considerando $\bar{Y}_1$ e $\bar{Y}_2$ correlacionados negativamente consegue-se reduzir a variância do estimador.

Uma possível abordagem para obter estimadores negativamente correlacionados é o seguinte:

vamos supor que $\theta = E[h(U)]$ onde $U \sim U(0,1)$ e que gerando 2 amostras i.d. com $\rho < 0$ consegue o nosso objectivo. Basta considerar que

se $U \sim U(0,1)$ e $(1-U) \sim U(0,1)$

$$cov(U, 1-U) = -\frac{1}{12} = -var(U)$$

e definindo $Y_{1i} = h(U_i)$ e $Y_{2i} = h(1-U_i)$;

as v.a's $U_i$ e $(1-U_i)$ (negativa) correlacionadas

chama-se variáveis Antitéticas

**Algoritmo - usando variáveis Antitéticas:**

Para $i = 1, \ldots, n$

gerar $u_i$

$y_{1i} = h(u_i)$ e $y_{2i} = h(1-u_i)$ (2n gerações)

$$z_i = \frac{y_{1i} + y_{2i}}{2}$$

Para

$$\hat{\theta} = \bar{Z} = \frac{\sum_{i=1}^{n} z_i}{n} = \frac{\sum_{i=1}^{n} (y_{1i} + y_{2i})}{2n}$$

**Observações:**

1) Pode mostrar-se que $Y_{1i}$ e $Y_{2i}$ estão correlacionados negativamente se $h$ for uma função monótona;

2) Esta técnica das variáveis antitéticas pode ser usada para v.a's não uniformes que são simuladas pelo método de transformação Inversa.

**Exemplo:** $W \sim G(r, \lambda)$. Criar um Algoritmo com N gerações de W para estimar $P(W > 8)$ reduzindo a variância da estimação com variáveis Antitética.

Para $i = 1, \ldots, N/2$

Gerar $u_i$

$$W_{1i} = \sum_{j=1}^{r} -\frac{\ln u_j}{\lambda} \quad \text{e} \quad W_{2i} = \sum_{j=1}^{r} -\frac{\ln(1-u_j)}{\lambda}$$

$$I_{1i} = I\{W_{1i} > 8\} \quad e \quad I_{2i} = I\{W_{2i} > 8\}$$

$$\hat{P}(W_{>8}) = \hat{p} = \sum_{i=1}^{N/2} \frac{(I_{1i} + I_{2i})}{N}$$

Reducção de variância

② variável de controle

Supor novamente que queremos estimar $\theta = E(h(x)) = E(Y)$ onde $Y = h(x)$ é o resultado da simulação Monte-Carlo.

Considerar outra v.a. $Z$ (variável de controle) fácil de simular com $E(Z)$ conhecido. Então:

$T_c = Y + c(Z - E(Z))$ é também um estimador centrado de $\theta$.

objectivo : Selecionar $c$ : $var(T_c) < var(Y)$

Qual é o $c$?

$$var(T_c) = var(Y + c(Z - E(Z))) = var(Y + cZ)$$
$$= var(Y) + c^2 var(Z) + 2c\,cov(Y, Z) \quad (1)$$

$c$ : $var(T_c)$ seja mínima

$$\min_{c=c^*} \left( var(Y) + c^2 var(Z) + 2c\,cov(Y,Z) \right)$$

$$\left.\frac{d\,var(T_c)}{dc}\right|_{c=c^*} = 0 \iff 2c^* var(Z) + 2\,cov(Y,Z) = 0 \iff$$

$$c^* = -\frac{cov(Y,Z)}{var(Z)}$$

$$\left.\frac{d^2 \operatorname{var}(T_c)}{dc^2}\right|_{c=c^*} = 2\operatorname{var}(Z) > 0 \quad \therefore \quad c^* = -\frac{\operatorname{cov}(Y,Z)}{\operatorname{var}(Z)} \text{ é}$$

mínimo!

$$\operatorname{var}(T_{c^*}) = \operatorname{var}(Y) + \frac{\operatorname{cov}(Y,Z)^2}{\operatorname{var}(Z)} - 2\,\frac{\operatorname{cov}(Y,Z)^2}{\operatorname{var}(Z)} =$$

coloca $c^*$ em (1)

$$= \operatorname{var}(Y) - \frac{\operatorname{cov}(Y,Z)^2}{\operatorname{var}(Z)}$$

$$\frac{\operatorname{var}(T_{c^*})}{\operatorname{var}(Y)} = 1 - \underbrace{\frac{\operatorname{cov}^2(Y,Z)}{\operatorname{var}(Y)\operatorname{var}(Z)}}_{\rho^2} = 1 - \rho^2,$$

Ie, A percentagem da redução de variância obtida usando $Z$ como v. de controle é $100\rho^2\,\%$

Algumas destas quantidades devem ser desconhecidas!

As quantidades $\operatorname{cov}(Y,Z)$ e $\operatorname{var}(Z)$ são desconhecidas

Para resolver isto é feito um estudo de simulação prévio (simulação piloto) com $\underline{\underline{K}}$ simulações e:

$$\widehat{\operatorname{cov}}(Y,Z) = \frac{\sum_{j=1}^{K} (Y_j - \bar{Y})(Z_j - E(Z))}{K-1}$$

$$\widehat{\operatorname{var}}(Z) = \frac{\sum_{j=1}^{K} (Z_j - E(Z))^2}{K-1} \quad \text{e estimar}$$

$$\hat{c}^* = -\frac{\widehat{\operatorname{cov}}(Y,Z)}{\widehat{\operatorname{var}}(Z)}$$

# Algoritmo com variável de controle

## 1. Estudo Piloto

Para $i = 1$ até $k < n$

gera $(y_i, z_i)$ para obter $\hat{c}^* = -\dfrac{\widehat{cov}(Y,Z)}{\widehat{var}(Y)}$

para

## 2. Simulação Principal

Para $i = 1$ até $n$

gera $(y_i, z_i)$

$t_i = y_i + \hat{c}^* (z_i - E(Z))$

Parar

$$\hat{\theta} = \bar{T} = \frac{\sum_{i=1}^{n} t_i}{n}$$

menor variância que $\hat{\theta}^* = \bar{Y}$

Exemplo: $X \sim U(0,1)$ e seja $Y = I(X \leq a)$ onde $I$ é a função indicadora da realização de $X \leq a$, $0 < a < 1$. Supõe que se pretende estimar, usando a simulação de Monte-Carlo, $p = P(X \leq a)$ e vamos considerar 2 estimadores: $Y$ e $T = Y + c(X - E(X))$

a) $T$ é um estimador de que tipo? As v.a's $X$ e $Y$ são correlacionadas?

$T = Y + c(X - \underbrace{E(X)}_{\frac{1}{2}})$ é um estimador que utiliza

X como variável de controle.

$X \sim U(0,1)$ e $Y = I(X \leq a) = \begin{cases} 1, & \text{se } X \leq a \\ 0, & \text{c.c.} \end{cases}$

$Y \sim Ben(p = \underbrace{P(X \leq a)}_{a})$ $\qquad P(X \leq a) = a$

• $cov(X,Y) = ?$

$cov(X,Y) = E(X.Y) - E(X)E(Y)$

$E(X.Y) = E[\underbrace{E[X.Y|Y]}]$

$\boxed{\begin{array}{l} E(X) = E[E[X|Y]] \\ Y = 0 \text{ se } X > a \\ Y = 1 \text{ se } X \leq a \end{array}}$

$= \sum_{y=0}^{1} E[X.Y|Y=y].P(Y=y)$

$= E[X.\underset{0}{Y}|Y=0).P(Y=0) +$

$+ E[X.Y|Y=1).P(Y=1)$

$= E(X.\underset{1}{Y}|Y=1).P(Y=1) = E[X.\underset{1}{Y}|X \leq a].P(X \leq a)$

$= E[X|X \leq a].\underbrace{P(X \leq a)}_{a}$

$E[X|X \leq a] = ?$

$(X|X \leq a) \sim ?$ $\qquad P(X \leq \underset{x < a}{x}|X \leq a) = \frac{P(X \leq x)}{P(X \leq a)} = \frac{x}{a}$

$F_{X|X \leq a}(x) = \frac{x}{a}$ $\qquad \therefore (X|X \leq a) \sim U(0,a)$

$E(X|X \leq a) = \frac{a}{2}$

$cov(X,Y) = E(X.Y) - E(X)E(Y) = \underbrace{\frac{a}{2} \times a}_{E(X.Y)} - \frac{1}{2} \times a =$

$= \frac{a}{2}(a-1) < 0$ porque $a < 1$, ie

X e Y estão negativamente correlacionados

b) Determinar c: var(T) seja mínima

$$T = Y + c(X - \tfrac{1}{2})$$

$$var(T) = var(Y + c(X - \tfrac{1}{2})) = var(Y + cX) =$$

$$= var(Y) + c^2 var(X) + 2c\,cov(Y,X)$$

$Y \sim Ber(a)$

$var(Y) = a(1-a)$

$X \sim U(0,1)$

$var(Y) = \frac{1}{12}$

$\frac{(b-a)^2}{12}$

$$= a(1-a) + \frac{c^2}{12} + \cancel{2}c\,\frac{a}{\cancel{2}}(a-1)$$

$$= a(1-a) + \frac{c^2}{12} + ca^2 - ca$$

$$\frac{d}{dc} var(T) = 0 \Leftrightarrow \frac{2c}{12} + a^2 - a = 0 \Leftrightarrow$$

$$\Leftrightarrow c = 6a - 6a^2$$

$$\frac{d^2 var(T)}{dc^2} > 0 \Leftrightarrow \frac{1}{6} > 0 \therefore c = 6a - 6a^2$$

$$= 6a(1-a)$$

c) Estudar a redução de variância na estimação de $p$ consequência da utilização de T em vez de Y

$$\frac{var(T)}{var(Y)} = 1 - \frac{cov^2(X,Y)}{var(X)\,var(Y)} = 1 - \rho^2$$

$$\rho^2 = \frac{cov^2(X,Y)}{var(X)\,var(Y)} = \frac{\left(\frac{a^2}{2} - \frac{a}{2}\right)^2}{\frac{a(1-a)}{12}} = \cdots$$

$$= \frac{3(a^3 - 2a^2 + a)}{(1-a)}$$

Concretizar com $a = 0.9$

$$\rho^2 = \frac{3(0.9^3 - 2 \times 0.9 + 0.9)}{0.1} = 0.27$$

% redução de variância é de 27%

## 3) Redução de variância por condicionamento (Simulação de Monte-Carlo condicional)

Recordar:

- valor esperado condicional

$$E[X \mid Y = y] = \begin{cases} \sum_x x\, P(X = x \mid Y = y) & \text{Discreto} \\ \int_{-\infty}^{+\infty} x\, f_{X|Y=y}(x)\, dx & \text{contínuo} \end{cases}$$

$g(Y) = E[X|Y]$ é uma v.a. que pode tomar diferentes valores consoante o valor fixado para a v.a $Y$

$$E[E[X|Y]] \underset{\forall (X,Y)}{\overset{?}{=}} E(X) = \begin{cases} \sum_y \sum_x x\, P(X=x|Y=y) . P(Y=y) = \\ \qquad = \sum_y E[X|Y=y] . P(Y=y) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x\, f_{X|Y=y}(x)\, f_Y(y)\, dx\, dy = \\ \qquad \underset{(2)}{=} \int_{-\infty}^{+\infty} E[X|Y=y] . f_Y(y)\, dy \end{cases}$$

Dem (Caso contínuo (2))

$$E[g(Y)] = E[\underbrace{E[X|Y]}_{g(Y)}] = \int_{-\infty}^{+\infty} g(y)\, f_Y(y)\, dy =$$

$$= \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{+\infty} x\, f_{X|Y=y}(x)\, dx \right] f_Y(y)\, dy =$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x\, \frac{f(x,y)}{\cancel{f_Y(y)}}\, \cancel{f_Y}(y)\, dx\, dy =$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x\, f_{X,Y}(x,y)\, dx\, dy = E(X)$$

$E[E[X|Y]] = E(X)$ ✓

- variância condicional = $var[E[X|Y]] = ?$

Será que $var(E[X|Y])$ é menor que $var(X)$ ??

$$E(X) = E[E[X|Y]]$$

① $\underline{var(X|Y)}$

(*) $var(X|Y) = E[X^2|Y] - E^2[X|Y]$

$$E[\underbrace{var(X|Y)}_{(*)}] = E[E[X^2|Y]] - E[E^2[X|Y]]$$

$$= E(X^2) - E[E^2(X|Y)] =$$

* Aparecer var(X)

$$= \underbrace{E(X^2) - E^2(X)} + E^2(X) - E[E^2[X|Y]]$$

$$= var(X) - [E[E^2[X|Y]] - E^2[E[X|Y]]]$$

$$= var(X) - var(E[X|Y])$$

$$\therefore \quad var(E[X|Y]) = var(X) - E[\underbrace{var(X|Y)}_{>0}]$$

$$Var(X) = var(E[X|Y]) + E[var(X|Y)]$$

$$var(X) > var(E[X|Y])$$

Como usar estes resultado na simulação?

Objectivo: Efectuar uma simulação para estimar $\theta = E(X)$. Sem simulação condicional, gera-se valores da v.a. X e efetua-se a média

Usando esta técnica com simulação condicional:

- Admitir que $\exists$ outra v.a. $Y$, tal que $E[X|Y]$ é conhecida e fácil de simular. Como $E[E[X|Y]] = E(X) = \theta$, $E[X|Y]$ é estimador centrado de $\theta$, e com tem variância inferior à $var(X)$, temos vantagem em utilizar a v.a. $E[X|Y]$

Algoritmo (Monte-Carlo condicional)

Para $i = 1$ até $n$

gerar $y_i$

calcular $v_i = g(y_i) = E[X|y_i]$

Parar

$$\hat{\theta} = \frac{\sum_{i=1}^{n} v_i}{n}$$

Exemplo: vamos supor se pretende estimar

$\theta = P(U+Z>4)$ onde $U \sim Exp(1)$ e $Z \sim Exp(1/2)$

Considerar $Y = I_{(U+Z>4)} = \begin{cases} 1 & \text{se } U+Z>4 \\ 0 & \text{e.c.} \end{cases}$

$Y \sim Ber(\theta = P(U+Z>4))$

Portanto $\theta = E(Y)$

1) Método de simulação de Monte-Carlo usual:

1. Gerar $u_1, \dots, u_n, z_1, \dots, z_n$ (método da transformação inversa indep.)

2. $y_i = I\{u_i + z_i > 4\}$, $i = 1, \dots, n$

Para

$$\hat{\theta} = \frac{\sum_{i=1}^{n} y_i}{n}$$

2) Método de simulação de Monte-Carlo condicional

Seja $V = E[Y|Z]$

$$E[Y|Z=z] = P(U+Z>4 \mid Z=z) = P(U>4-z) =$$

$$= 1 - F_U(\underbrace{4-z}_{\geq 0})$$

$$= \begin{cases} e^{-(4-z)} & 0 \leq z < 4 \\ 1 & z \geq 4 \end{cases}$$

Assim $V = E[Y|Z] = \begin{cases} e^{-(4-Z)} & 0 \leq Z < 4 \\ 1 & Z \geq 4 \end{cases}$

Algoritmo para estimar $\theta = E(V) = E[E[Y|Z]]$

1. Gerar $z_1, \dots, z_n$ indep

2. $v_i = E[Y|z_i] = e^{-(4-z_i)}$

Para

$$\hat{\theta} = \frac{\sum_{i=1}^{n} v_i}{n}$$

## 4 Amostragem por Importância

Existem situações que envolvem a estimação de Acontecimentos raros

Ex 1. $X \sim N(0,1)$ estimar $\theta = P(X > 20) = E(I\{X \geq 20\})$

Monte-Carlo usual:

1. Gerar $x_1, \dots, x_n$ i.i.d. $N(0,1)$
2. $I_j = I\{x_i \geq 20\}$ $j = 1, \dots, n$

Para
$$\hat{\theta} = \sum_{j=1}^{n} \frac{I_j}{n}$$

Só vai funcionar $n$ muito grande, caso contrário $\hat{\theta} \cong 0$

Podemos usar a Amostragem por Importância

Vamos supor que se $\theta = E_f(h(X))$ onde $X$ é uma v.a. com f.d.p. $f(\cdot)$. Seja $g(\cdot)$ outra f.d.p. onde $g(x) \neq 0$ sempre que $f(x) \neq 0$ (com o mesmo suporte que $f(x)$)

$$\theta = E_f[h(x)] = \int h(x) f(x) dx = \int \overbrace{h(x) \frac{f(x)}{g(x)}}^{h^*(x)} g(x) dx$$

$$= E_g\left[\frac{h(x) f(x)}{g(x)}\right] = E_g[h^*(x)]$$

onde $E_g[\cdot]$ representa o valor esperado calculado usando a f.d.p. $g(\cdot)$

Assim, para estimar $\theta$ pode-se simular $x_i$ a partir da densidade $g(\cdot)$ e fazer $\hat{\theta} = \sum_{j=1}^{n} \frac{h(x_j) f(x_j)}{n\, g(x_j)}$

onde $\hat{\theta}$ é designado por estimador de Amostragem por Importância de $\theta$.

Como se pode tentar reduzir a variância?

Se possível selecionar uma densidade $g(\cdot)$ tal que

$$h^*(x) = \frac{h(x) f(x)}{g(x)}$$

tenha menor variância que a do estimador original.

Um possível critério é selecionar $g(\cdot)$ tal que o quociente $\frac{f(x)}{g(x)}$ seja grande para valores pequenos de $h(x)$ e vice-versa, de forma a tentar compensar a variabilidade de $h(x)$

Ex. 1 (novamente)

Estimar $\theta = P(X \geq 20)$ onde $X \sim N(0,1)$

$$\theta = E\left[I_{\{X \geq 20\}}\right] \qquad I_{\{X \geq 20\}} \sim Ber(\theta)$$

$$\theta = E\left[I_{\{X \geq 20\}}\right] = \int_{-\infty}^{+\infty} I_{\{x \geq 20\}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx =$$

considerar $g(\cdot)$ f.d.p. de outra Normal com valor $\mu$ ($\mu > 20$)

$$\underset{g(\cdot) \equiv N(\mu,1)}{=} \int_{-\infty}^{+\infty} I_{\{x \geq 20\}} \frac{\cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{x^2}{2}}}{\underbrace{\cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{(x-\mu)^2}{2}}}_{g(\cdot)}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} dx$$

$$= \int_{-\infty}^{+\infty} \underbrace{I\{x \geq 20\}\, e^{-\mu x + \frac{\mu^2}{2}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} \, dx$$

$$= E_g\left[ I\{X \geq 20\}\, e^{-\mu X + \frac{\mu^2}{2}} \right]$$

Algoritmo:

1. Gerar $x_1, \dots, x_n$ i.i.d. de $N(\mu, 1)$
2. $I_j = I\{x_j \geq 20\}\, e^{\mu x_j + \frac{\mu}{2}}$ $\quad j = 1, \dots, n$

Para

$$\hat{\Theta} = \frac{\sum_{j=1}^{n} I_j}{n}$$

FIM capítulo Simulação

//