

APROXIMAÇÃO DE MONTE CARLO

Modelos Compartmentais em Epidemiologia e
Inferência Bayesiana

Gustavo Libotte e Regina Almeida

14 de janeiro de 2022

Aproximação de Monte Carlo

Visão geral



- ▶ Duas grandes classes de problemas numéricos que surgem na inferência estatística são: problemas de **otimização**, e problemas de **integração**.
- ▶ De fato, diversos exemplos mostram que nem sempre é possível **calcular analiticamente** os estimadores associados à um determinado problema.
- ▶ Experimentação de Monte Carlo significa o uso de valores aleatórios para a **estimação** de alguma função de uma distribuição de probabilidade.
- ▶ Um problema que não possui um componente estocástico pode ser colocado como um problema com um componente que pode ser identificado como a **esperança** de uma função de uma variável aleatória.
- ▶ Isso pode ser feito através da **decomposição** de uma função densidade de probabilidade.
- ▶ O problema é então resolvido pela **estimação do valor esperado** por meio do uso de uma amostra aleatória da distribuição da variável aleatória.

- ▶ Ao longo da década de 1990, a Pesquisa Social Geral coletou dados sobre o nível de escolaridade e o número de filhos de **155 mulheres** com 40 anos de idade na época de sua participação na pesquisa.
- ▶ Essas mulheres estavam na casa dos 20 anos durante a década de 1970, um período de taxas de fertilidade historicamente baixas nos Estados Unidos.
- ▶ Neste exemplo, compararemos as *mulheres com diploma universitário* àquelas *sem, em termos de número de filhos*.
- ▶ Denote como $Y_{1,1} \dots, Y_{n_1,1}$ o número de filhos para as n_1 mulheres **sem diploma universitário** e $Y_{1,2} \dots, Y_{n_2,2}$ os dados para **mulheres com diplomas**.
- ▶ Para este exemplo, usaremos os seguintes modelos amostrais:

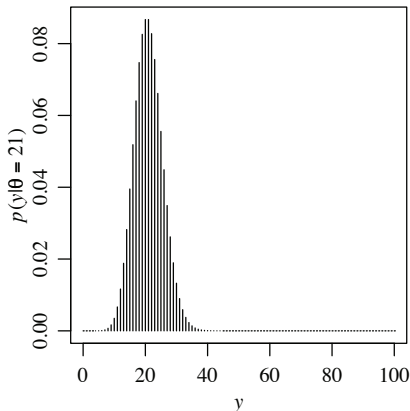
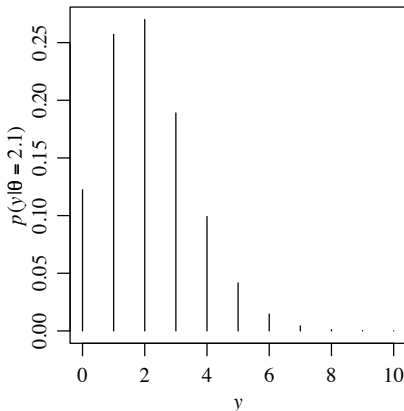
$$Y_{1,1} \dots, Y_{n_1,1} \mid \theta_1 \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\theta_1)$$

$$Y_{1,2} \dots, Y_{n_2,2} \mid \theta_2 \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\theta_2)$$

- ▶ Essas são as nossas **funções de verossimilhança**.

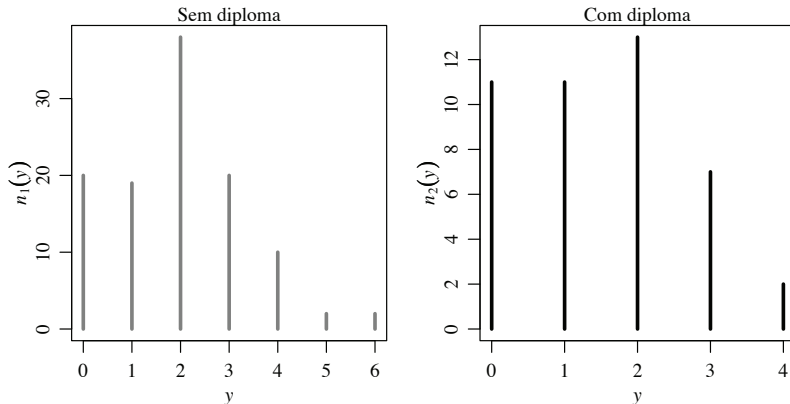
Motivação

- **Relembrando:** na distribuição de **Poisson**, $p(Y = y | \theta) = \frac{\theta^y e^{-\theta}}{y!}$ é a probabilidade de se obter y eventos, se o **número médio** esperado for θ . Veja alguns **exemplos** abaixo.



Motivação

- Distribuições empíricas para os **dados** são exibidas na figura abaixo.



- Alguns indicadores relacionados aos dados:

Sem diploma: $n_1 = 111$, $\sum_{i=1}^{n_1} Y_{i,1} = 217$, $\bar{Y}_1 = 1,95$

Com diploma: $n_2 = 44$, $\sum_{i=1}^{n_2} Y_{i,2} = 66$, $\bar{Y}_2 = 1,50$

Motivação



- ▶ No caso em que $\{\theta_1, \theta_2\} \stackrel{\text{i.i.d.}}{\sim} \text{Gama}(\alpha = 2, \beta = 1)$, temos as seguintes distribuições *a posteriori*:
 - ▶ $\theta_1 \mid \{n_1 = 111, \sum_{i=1}^{n_1} Y_{i,1} = 217\} \sim \text{Gama}(2 + 217, 1 + 111) = \text{Gama}(219, 112)$
 - ▶ $\theta_2 \mid \{n_2 = 44, \sum_{i=1}^{n_2} Y_{i,2} = 66\} \sim \text{Gama}(2 + 66, 1 + 44) = \text{Gama}(68, 45)$

O modelo conjugado Gama/Poisson

Se $Y_1, \dots, Y_n \mid \theta \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\theta)$ e o parâmetro $\theta \stackrel{\text{i.i.d.}}{\sim} \text{Gama}(\alpha, \beta)$, então temos

$$\text{Função de verossimilhança: } p(Y_1, \dots, Y_n \mid \theta) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{y_i}}{y_i!} = \frac{e^{-n\theta} \theta^{\sum y_i}}{\prod_{i=1}^n (y_i!)}$$

$$\text{Distribuição a priori: } p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \quad \theta > 0$$

$$\text{Distribuição a posteriori: } p(\theta \mid Y_1, \dots, Y_n) \propto \theta^{\sum y_i + \alpha - 1} e^{-(n + \beta)\theta}, \quad \theta > 0$$

Portanto $p(\theta \mid Y_1, \dots, Y_n) \sim \text{Gama}(\sum y_i + \alpha, n + \beta)$ é a distribuição *a posteriori* conjugada Gama/Poisson para este problema.

- Suponha que queiramos saber:

$$p \left(\theta_1 > \theta_2 \mid \sum_{i=1}^{n_1} Y_{i,1} = 217, \sum_{i=1}^{n_2} Y_{i,2} = 66 \right)$$

isto é, queremos saber a **probabilidade do valor esperado de filhos de uma mulher sem diploma (θ_1) ser maior do que o valor esperado de filhos de uma mulher com diploma (θ_2)**.

- Como isso pode ser calculado?

$$\begin{aligned} & P(\theta_1 > \theta_2 \mid y_{1,1}, \dots, y_{n_2,2}) \\ &= \int_0^\infty \int_0^{\theta_1} p(\theta_1, \theta_2 \mid y_{1,1}, \dots, y_{n_2,2}) d\theta_2 d\theta_1 \\ &= \int_0^\infty \int_0^{\theta_1} \text{Gama}(\theta_1 \mid 219, 112) \times \text{Gama}(\theta_2 \mid 68, 45) d\theta_2 d\theta_1 \\ &= \frac{112^{219} 45^{68}}{\Gamma(219)\Gamma(68)} \int_0^\infty \int_0^{\theta_1} \theta_1^{218} \theta_2^{67} e^{-112\theta_1 - 45\theta_2} d\theta_2 d\theta_1 \\ &= 0,97 \end{aligned}$$

Motivação



- ▶ Existem **várias maneiras** de calcular essa integral.
- ▶ Isso pode ser feito com lápis e papel usando resultados de cálculo e também pode ser calculado numericamente em muitos pacotes de *software* matemático.
- ▶ No entanto, a viabilidade desses métodos de integração **depende muito** dos detalhes particulares desse modelo, distribuição *a priori* e declaração de probabilidade que estamos tentando calcular.
- ▶ Como alternativa, usaremos um método de integração para o qual os princípios e procedimentos gerais permanecem relativamente constantes em uma ampla classe de problemas.
- ▶ O método, conhecido como **aproximação de Monte Carlo**, é baseado em amostragem aleatória e sua implementação não requer um conhecimento profundo de cálculo ou análise numérica.

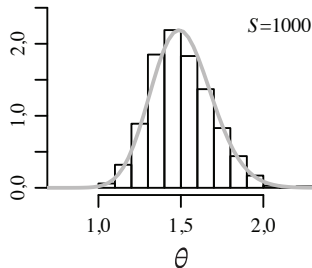
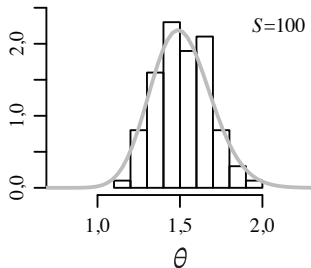
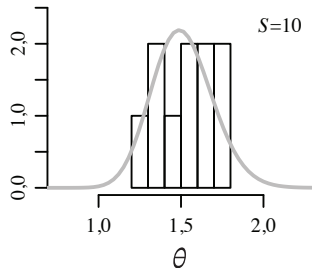
- ▶ Seja θ o parâmetro de interesse e seja y_1, \dots, y_n ser os valores numéricos de uma amostra da distribuição $p(y_1, \dots, y_n \mid \theta)$.
- ▶ Suponha que pudéssemos amostrar algum número S independente e aleatório de valores θ da distribuição *a posteriori* $p(\theta \mid y_1, \dots, y_n)$:

$$\theta^{(1)}, \dots, \theta^{(S)} \stackrel{\text{i.i.d}}{\sim} p(\theta \mid y_1, \dots, y_n)$$

- ▶ Então a distribuição empírica das amostras $\{\theta^{(1)}, \dots, \theta^{(S)}\}$ aproximaria $p(\theta \mid y_1, \dots, y_n)$, com a aproximação **melhorando** com o aumento S .
- ▶ A distribuição empírica de $\{\theta^{(1)}, \dots, \theta^{(S)}\}$ é conhecida como uma **APROXIMAÇÃO DE MONTE CARLO** para $p(\theta \mid y_1, \dots, y_n)$.

Motivação

- A figura a seguir mostra sucessivas aproximações de Monte Carlo para a densidade da distribuição Gama (68, 45), juntamente com a função de densidade verdadeira para comparação.



- Como vemos, a distribuição empírica das amostras de Monte Carlo fornece uma aproximação **cada vez mais próxima** da densidade real à medida que S aumenta.

Aproximação de Monte Carlo



- ▶ Seja $g(\theta)$ uma função arbitrária. Se $\theta^{(1)}, \dots, \theta^{(S)}$ são amostras i.i.d. de $p(\theta | y_1, \dots, y_n)$, então

$$\frac{1}{S} \sum_{s=1}^S g(\theta^{(s)}) \rightarrow \mathbb{E}[g(\theta) | y_1, \dots, y_n] = \int g(\theta) p(\theta | y_1, \dots, y_n) d\theta \quad \text{quando } S \rightarrow \infty.$$

- ▶ Isso implica que como $S \rightarrow \infty$:

- ▶ $\bar{\theta} = \sum_{s=1}^S \theta^{(s)} / S \rightarrow \mathbb{E}[\theta | y_1, \dots, y_n];$
 - ▶ $\sum_{s=1}^S (\theta^{(s)} - \bar{\theta})^2 / (S - 1) \rightarrow \text{Var}[\theta | y_1, \dots, y_n];$
 - ▶ $\#(\theta^{(s)} \leq c) / S \rightarrow P(\theta \leq c | y_1, \dots, y_n)$
 - ▶ a distribuição empírica de $\{\theta^{(1)}, \dots, \theta^{(S)}\} \rightarrow p(\theta | y_1, \dots, y_n);$
 - ▶ a mediana de $\{\theta^{(1)}, \dots, \theta^{(S)}\} \rightarrow \theta_{1/2}$
- ▶ Praticamente qualquer aspecto de uma distribuição *a posteriori* em que possamos estar interessados pode ser aproximado arbitrariamente exatamente com uma amostra de Monte Carlo **grande o suficiente**.

Aproximação de Monte Carlo

Exemplo ilustrativo



- ▶ Um exemplo clássico de usar um método de Monte Carlo para calcular uma quantidade de interesse é a **estimativa numérica do número π** .
- ▶ Podemos estimar o valor de π com o seguinte procedimento:
 1. Amostre N pontos aleatoriamente em um quadrado de lado igual a $2r$.
 2. Desenhe um círculo de raio r inscrito no quadrado e conte o número de pontos que estão dentro desse círculo.
 3. Estime $\hat{\pi}$ como a proporção $4d/N$, onde d é o número de pontos dentro do círculo.
- ▶ Entendendo melhor:
 - ▶ A área do círculo é πr^2 , onde r é o raio do círculo.
 - ▶ A área do quadrado é $\ell^2 = 4r^2$, onde ℓ é o comprimento de um dos lados do quadrado.
 - ▶ Se dividirmos a área do círculo pela área do quadrado, obtemos $\pi/4$.
 - ▶ A área do círculo e o quadrado são proporcionais ao número de pontos dentro do círculo e ao total de N pontos, respectivamente.
 - ▶ Sabemos que um ponto está dentro de um círculo se a relação $\sqrt{(x^2 + y^2)} \leq R$ for válida.

Aproximação de Monte Carlo

Exemplo ilustrativo



```
N = 1000

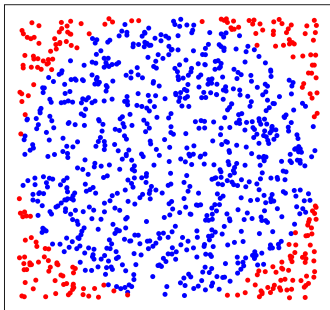
x, y = np.random.uniform(-1, 1, size=(2, N))
inside = (x**2 + y**2) <= 1
pi = inside.sum()*4/N
error = abs((pi - np.pi) / pi) * 100
outside = np.invert(inside)

plt.figure(figsize=(4, 4))
plt.plot(x[inside], y[inside], 'b.')
plt.plot(x[outside], y[outside], 'r.')
plt.plot(0, 0, markersize=1, alpha=0)
plt.xticks([])
plt.yticks([])
plt.title(f'N = {N}, n* = {pi:4.3f}, erro = {error:4.3f}', fontsize=12);
```

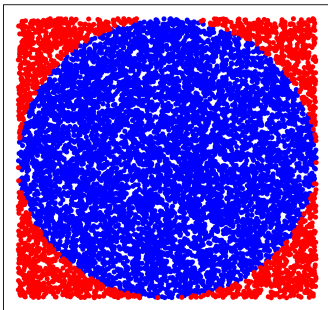
Aproximação de Monte Carlo

Exemplo ilustrativo

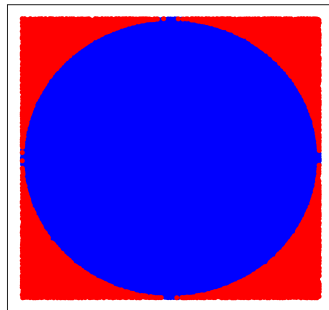
$N = 1000$, $\pi^* = 3.084$, erro = 1.867



$N = 10000$, $\pi^* = 3.160$, erro = 0.570



$N = 100000$, $\pi^* = 3.137$, erro = 0.152



Por que precisamos de amostragem eficiente?



- ▶ Se a distribuição *a posteriori* não for uma distribuição conhecida, ainda podemos avaliá-la em uma **grade densa**.
- ▶ Avaliar em uma grade **não é tão simples** quanto parece à primeira vista, por várias razões. Devemos garantir que a grade **cubra toda a extensão** da distribuição *a posteriori*.
- ▶ Se a distribuição tiver suporte infinito, poderíamos usar uma grade **apenas se cobrirmos todas as regiões de alta densidade de probabilidade**. No entanto, se a expressão para a distribuição *a posteriori* for complicada, não será fácil **identificar** essas regiões.
- ▶ A distribuição *a posteriori* também pode ser altamente “pontagudo”, caso em que precisamos usar uma amostragem muito **refinada** em torno desta região.
- ▶ A *a posteriori* pode até ser **multimodal**, então apenas a amostragem em torno do primeiro máximo encontrado será **insuficiente**.
- ▶ Poderíamos tentar descobrir a faixa de amostragem e densidade ideais por **tentativa e erro**: visualize a *a posteriori* e estude se quantidades como a média ou a moda são estáveis com mudanças na amostragem.
- ▶ Mas esta não é uma opção realista se estivermos processando um **grande número de conjuntos de dados**.

Por que precisamos de amostragem eficiente?



- ▶ Para problemas com **três ou mais parâmetros** a situação é ainda mais complicada.
- ▶ Todas as questões anteriores se aplicam, mas agora *a posteriori* existe em um espaço de parâmetros dimensional mais alto, onde são possíveis **superfícies mais complexas** que são **impossíveis de se visualizar** completamente.
- ▶ Também é **muito mais demorado** determinar *a posteriori* avaliando-a em uma grade de parâmetros regular, mas agora multidimensional.
- ▶ Se precisarmos de N pontos de grade para amostrar um parâmetro com densidade suficiente, para dois parâmetros precisaremos de N^2 pontos de grade.
- ▶ Para dimensões d , precisaremos de N^d pontos de grade.
- ▶ Mesmo com $d = 5$ e $N = 100$ (bastante modesto), já precisaríamos de 10^{10} pontos de grade, um número **inviável** de cálculos da função de verossimilhança (que normalmente é mais demorado para calcular do que a priori).

Por que precisamos de amostragem eficiente?



- ▶ Poderíamos evitar a ineficiência de uma grade amostrando **diretamente** da função densidade de probabilidade *a posteriori*.
- ▶ Gostaríamos de poder extrair amostras de uma função densidade de probabilidade arbitrária—vamos chamá-la de $g(\theta)$ —de tal forma que a distribuição de frequência das amostras seja igual a $g(\theta)$ no **limite de um grande número de amostras**.
- ▶ Vamos supor que podemos calcular $g(\theta)$ sem exigir que ela seja normalizada. Como podemos amostrar a partir dela?
- ▶ Isso não é tão simples, porque para amostrar eficientemente precisaríamos extrair preferencialmente de regiões de **alta probabilidade relativa**.
- ▶ Mas não é óbvio como poderíamos **localizar essas regiões sem avaliar explicitamente a função** em todos os lugares, o que não é melhor do que avaliar em uma grade.
- ▶ Existem **maneiras eficientes de amostrar** uma distribuição de probabilidade arbitrária usando métodos de Monte Carlo.

Por que precisamos de amostragem eficiente?



- ▶ Podemos então usar um conjunto de amostras $\{\theta\}$ para representar essa distribuição.
- ▶ Este é o caso mesmo quando $g(\theta)$ **não é normalizado**: redimensionar a distribuição por um fator constante **não mudará** a frequência relativa com que as amostras são extraídas.
- ▶ Assim, uma vez que tenhamos esse conjunto de amostras, podemos aproximar essencialmente qualquer **propriedade** de $g(\theta)$ —como média, variância e intervalos de confiança—diretamente do conjunto de amostras.
- ▶ Também é simples usar tais amostras para **aproximar integrais** (veja o [slide 9](#) novamente).
- ▶ A **integração de Monte Carlo** é uma técnica numérica que se baseia em amostragem aleatória para aproximar um resultado, aplicando esse processo para a estimação numérica de integrais.
- ▶ De maneira geral, um problema de inferência estatística pode ser formulado como a estimação de uma integral do tipo

$$\theta = \int_D h(x) dx$$

Se a integral possui forma fechada, então não há necessidade de qualquer método de aproximação.

Por que precisamos de amostragem eficiente?



- ▶ Caso não seja possível resolver a integral de forma analítica, e se D for de uma ou duas dimensões, então existem diversos métodos de **quadratura** para aproximar o valor dessa integral.
- ▶ No entanto, quando a dimensão em D for alta, a integração de Monte Carlo é uma alternativa mais viável (que pode ser usada em problemas de baixa dimensão também).
- ▶ Se a função h for decomposta de forma a ter um componente que é uma densidade de probabilidade, ou seja,

$$h(x) = g(x)f(x)$$

onde $\int_D f(x)dx = 1$ e $f(x) \geq 0$, então a integral θ pode ser vista como a esperança da variável aleatória $Y = g(X)$, onde X tem distribuição $f(x)$, isto é,

$$\theta = \mathbb{E}[g(X)] = \int_D g(x)f(x)dx.$$

- ▶ Com uma amostra aleatória x_1, \dots, x_m da distribuição $f(x)$ da variável aleatória X , então uma estimativa **não viesada** de θ é a média amostral

$$\hat{\theta} = \frac{1}{m} \sum_{i=1}^m g(x_i)$$

Aproximação de Monte Carlo

Integração simples

- Considere o problema de estimar

$$\theta = \int_0^1 g(x)dx.$$

- Se X_1, \dots, X_m é uma amostra aleatória de $U(0, 1)$, então

$$\hat{\theta} = \frac{1}{m} \sum_{i=1}^m g(x_i)$$

- Pode-se mostrar que $\hat{\theta}$ converge para $E[\hat{\theta}] = \theta$ quando $m \rightarrow \infty$ com probabilidade 1, pela **Lei Forte dos Grandes Números**.

Lei Forte dos Grandes Números

Sejam X_1, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas com $E[X] = \mu$ e $\text{Var}[X] = \sigma^2$, e definimos $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Então, para $\epsilon > 0$,

$$P \left(\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| < \epsilon \right) = 1$$

isto é, converge quase certamente para μ .

Aproximação de Monte Carlo

Integração simples



- ▶ Por exemplo, obtenha uma estimativa de

$$\theta = \int_0^1 e^{-x} dx$$

- ▶ Aproximação de Monte Carlo (*código em R*):

```
## Obtem m valores da U(0,1)
m <- 10000
x <- runif(m)
## Calcula g(x)
theta.hat <- exp(-x)
## Calcula a média
(m.theta.hat <- sum(theta.hat)/m)
## Saida: 0.6313544
```

- ▶ Solução “analítica” e numérica:

```
## Solução analítica
(theta <- 1 - exp(-1))
## Saida: 0.6321206
```

```
## Integração numérica no R
integrate(function(x) exp(-x), lower = 0, upper = 1)
## Saida: 0.6321206
```

Aproximação de Monte Carlo

Integração



- Um caso mais geral é estimar a integral do tipo

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

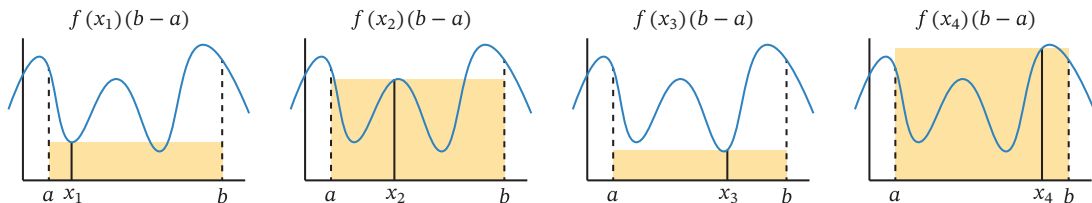
- Nesse caso, temos que substituir a $U(0,1)$ por alguma outra densidade com suporte no intervalo dos limites de integração. Por exemplo, se $X \sim U(a,b)$, então $f(x) = \frac{1}{b-a}$ e


$$\begin{aligned}\theta &= \frac{1}{f(x)} \int_a^b g(x)f(x)dx \\ &= (b-a) \int_a^b g(x) \frac{1}{b-a} dx \\ &= (b-a)E[g(X)]\end{aligned}$$

- De maneira geral, para calcular $\theta = \int_a^b g(x)dx$:
 1. Gere X_1, \dots, X_m de $U(a,b)$
 2. Calcule $\bar{g}(x) = \frac{1}{m} \sum_{i=1}^m g(x_i)$
 3. $\hat{\theta} = (b-a)\bar{g}(x)$

Aproximação de Monte Carlo

Integração



$$\frac{1}{4} \left(\boxed{f(x_1)(b-a)} + \boxed{f(x_2)(b-a)} + \boxed{f(x_3)(b-a)} + \boxed{f(x_4)(b-a)} \right) \approx \text{Area under } f(x)$$


Aproximação de Monte Carlo

Integração



- ▶ Por exemplo, obtenha uma estimativa de

$$\theta = \int_2^4 e^{-x} dx$$

- ▶ Aproximação de Monte Carlo (código em R):

```
## Obtem m valores da U(2,4)
m <- 10000
a <- 2; b <- 4
x <- runif(m, min = a, max = b)
## Calcula g(x)
theta.hat <- exp(-x)
## Calcula a média * (b - a)
(m.theta.hat <- (sum(theta.hat)/m) * (b - a))
## Saida: 0.1171233
```

- ▶ Solução “analítica” e numérica:

```
## Solução analítica
(theta <- exp(-2) - exp(-4))
## Saida: 0.1170196

## Integração numérica no R
integrate(function(x) exp(-x), lower = 2, upper = 4)
## Saida: 0.1170196
```


Aproximação de Monte Carlo

Probabilidade



- ▶ Vemos que, nos dois casos anteriores, estamos na verdade obtendo uma estimativa das probabilidades $P[0 < X < 1]$ e $P[2 < X < 4]$, respectivamente.
- ▶ Vejamos um caso em que a aproximação de Monte Carlo não fornece bons resultados:

$$P[X > 2] = \int_2^{\infty} e^{-x} dx = \frac{1}{e^2} \approx 0,1353352$$

- ▶ Poderíamos pensar em uma aproximação *grosseira*, supondo que vamos amostrar de $X \sim \mathcal{U}(2, \infty)$, onde podemos aproximar o valor de ∞ por um número grande. Por exemplo:

```
## Obtem m valores da U(2,Inf)
m <- 10000
a <- 2; b <- 1e5
x <- runif(m, min = a, max = b)
## Calcula g(x)
theta.hat <- exp(-x)
## Calcula a média * (b - a)
(m.theta.hat <- (sum(theta.hat)/m) * (b - a))
## Saida: 0.2770476
```

Aproximação de Monte Carlo

Mais um exemplo



- Considere que uma variável aleatória X tem distribuição Beta($a = 9,39, b = 33,67$), e queremos obter $P[0,3 < X < 0,5]$, ou seja,

$$P[0,3 < X < 0,5] = \int_{0,3}^{0,5} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^a (1-x)^b dx \approx 0,1018369$$

Pela integração simples de Monte Carlo, podemos facilmente obter esta estimativa:

```
a <- 9.39; b <- 33.67
m <- 10000
x <- runif(m, 0.3, 0.5)
## Calcula g(x)
theta.hat <- dbeta(x, a, b)
## Calcula a média
(m.theta.hat <- (sum(theta.hat)/m) * (0.5 - 0.3))
## Solução analítica
pbeta(0.5, a, b) - pbeta(0.3, a, b)
## Saída: 0.1014207
```

Aproximação de Monte Carlo

Mais um exemplo



- No caso da Beta, calcular probabilidades do tipo $P[X > 0,2]$, por exemplo, é mais fácil, pois sabemos que o domínio da Beta está no intervalo $(0, 1)$. Portanto, os limites de integração são definidos

$$P[X > 0,2] = \int_{0,2}^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^a (1-x)^b dx \approx 0,5876479$$

Assim, podemos fazer

```
m <- 10000
x <- runif(m, 0.2, 1)
## Calcula g(x)
theta.hat <- dbeta(x, a, b)
## Calcula a média
(m.theta.hat <- (sum(theta.hat)/m) * (1 - 0.2))
## Saida: 0.585035
```