

Monte Carlo

14 de setembro de 2021 – Brasília



- Eu gosto de pensar no método de monte carlo como uma forma de você fazer o papel de Deus ou Natureza e ter controle sobre os processos geradores de dados.

No mundo real, quando enfrentamos um problema empírico precisamos (1) coletar dados que estão disponíveis em alguma base ou (2) fazemos experimentos para gerar esses dados:

- Um exemplo do primeiro caso é você acessar a base do Banco Central de bancos e coletar dados contábeis de bancos. Usando esses dados contábeis, você pode por exemplo estudar os determinantes de lucratividade dos bancos.
- Um exemplo do segundo caso é quando uma empresa farmacêutica deseja testar um novo medicamento e considerando uma determinada amostra aplica esse novo medicamento em parte da amostra e, a partir daí, infere a respeito do potencial do novo medicamento. Em ambos os casos, você não tem acesso ao processo gerador dos dados. Você tem apenas acesso as realizações de eventos que ocorreram em um determinado momento e pode fazer hipóteses a respeito das variáveis que você coletou (ou da base ou do experimento).

No caso do método de monte carlo, você controla o processo gerador de dados de algumas variáveis aleatórias e usa esses processos para estudar ou entender mais sobre algum fenômeno. Isso pode ser interessante para ganhar intuição sobre algum problema que pode depois ser resolvido analiticamente ou realmente usar as amostras coletadas para resolver o problema de interesse (quando não houver solução analítica).

Exemplo

Considere por exemplo o problema simples de calcular a relação entre as áreas de um quadrado e um círculo que está inscrito nesse quadrado e tangencia os quatro lados dele. Assuma por exemplo que o círculo tem raio $r = 1$ e, conseqüentemente, o quadrado tem lado $l = 2$. A solução desse problema é óbvia, pois sabemos que a relação entre as áreas é dada por

$$\frac{\pi r^2}{l^2} = \frac{\pi}{4}.$$

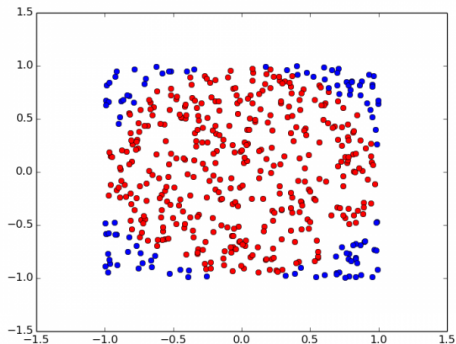
Será que podemos resolver esse problema usando o Método de Monte Carlo?

Sim! Note que podemos aproximar a solução respondendo uma pergunta relacionada: Dado que eu sorteio um vetor de variáveis aleatórias (x, y) uniformemente no intervalo $[-1, 1] \times [-1, 1]$, qual a probabilidade de ele cair dentro do círculo?

Monte Carlo

Relação entre círculo e quadrado

Métodos Computacionais
Professor Daniel O. Cajueiro
Economia (UnB)



```
import random
import matplotlib.pyplot as plt
def quadradoCirculo(n):
    nc=0
    plt.axis([-1.5, 1.5, -1.5, 1.5])
    plt.hold(True)
    for i in range(0,n):
        x=random.uniform(-1.0,1.0)
        y=random.uniform(-1.0,1.0)
        if(x**2+y**2<1.0):
            plt.plot([x],[y], 'ro')
            nc=nc+1.0
        else:
            plt.plot([x],[y], 'bo')
    return nc/n
```

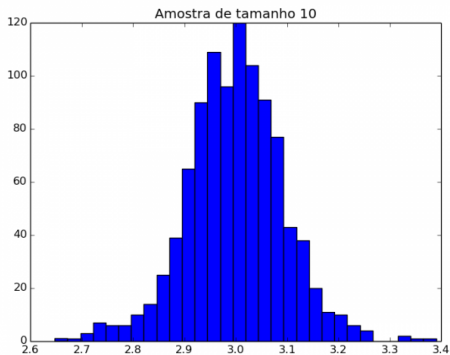


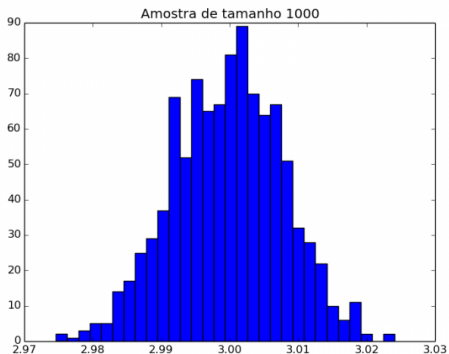
```
if __name__ == '__main__':  
    n=500  
    print(quadradoCirculo(n))
```

Exemplo

Suponha que você quer analisar o que ocorre com a distribuição do estimador de mínimos quadrados do coeficiente angular da reta dada por $y = \alpha + \beta x + u$ quando aumentamos o tamanho da amostra.

Suponha que u é uma variável aleatória normal padrão e $\alpha = 2$ e $\beta = 3$. Vamos fazer o papel de Deus e gerar os valores de y e usar os vetores $[x, y]$ para estimar os coeficientes α e β (embora estejamos aqui particularmente interessados nas estimativas de β).





```
import random
import matplotlib.pyplot as plt
from scipy import stats
import numpy as np

def geracaoDados(n):
    alpha=2
    beta=3
    x=np.random.normal(0,4,n)
    u=np.random.normal(0,1,n)
    y=alpha*np.ones(n)+beta*x+u
    return x,y
```

```
def estimacao(x,y):  
    beta, alpha, r_value, p_value, std_err =  
    stats.linregress(x,y)  
    return alpha,beta
```

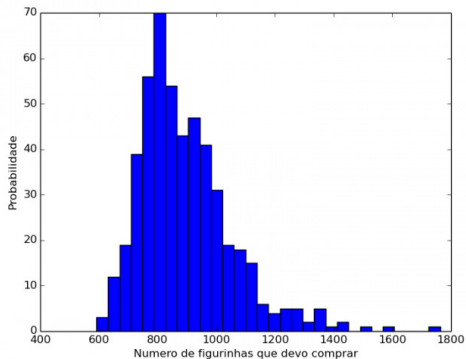
```
if __name__ == '__main__':  
    numeroAmostras=1000  
    n=1000 # numero de observacoes  
           # em cada amostra (parametro)  
    vetor=np.empty([numeroAmostras])  
    for i in range(0,numeroAmostras):  
        [x,y]=geracaoDados(n)  
        [alpha,beta]=estimacao(x,y)  
        vetor[i]=beta  
    matplotlib.pyplot.hist(vetor,bins=30)  
    plt.title('Amostra de tamanho 1000')
```

Exemplo

Vamos considerar o álbum de figurinhas da Copa de 2014 da Panini.

- 1) O álbum tem 640 figurinhas diferentes
- 2) Cada pacote vem com 5 figurinhas
- 3) Supõe-se que as figurinhas dentro de cada pacote e os pacotes são independentes



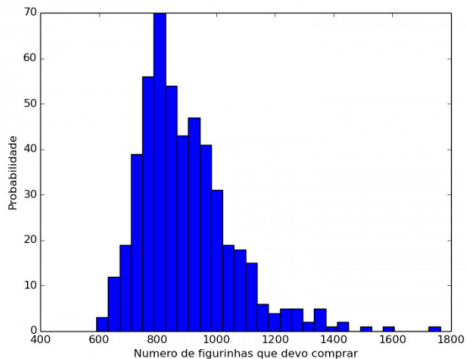


```
import numpy as np
import matplotlib.pyplot as plt

def geraPacote(numeroFigurinhasPacote,numeroFigurinhasAlbum):
    pacote=np.random.randint(0,numeroFigurinhasAlbum,
                             numeroFigurinhasPacote)

    return pacote
```

```
if __name__ == '__main__':  
    # Parametros  
    numeroFigurinhasPacote=5  
    numeroFigurinhasAlbum=640  
    numeroAmostras=500  
    simulations=np.empty([numeroAmostras])  
    album=np.zeros([numeroFigurinhasAlbum],dtype=bool)  
    # Variaveis  
    porcentagemPreenchida=0.95  
    preenchimento=int(porcentagemPreenchida*numeroFigurinhasAlbum)  
    #Codigo Principal  
    media=0  
    for i in range(numeroAmostras):  
        print("Amostra", i)  
        album=np.zeros([numeroFigurinhasAlbum],dtype=bool)  
        numeroFigurinhasPreenchidas=0  
        somaNumeroDePacotesUsados=0  
        while (numeroFigurinhasPreenchidas<preenchimento):  
            pacote=geraPacote(numeroFigurinhasPacote,numeroFigurinhasAlbum)  
            somaNumeroDePacotesUsados=somaNumeroDePacotesUsados+1  
            for j in range(numeroFigurinhasPacote):  
                if (not album[pacote[j]]):  
                    album[pacote[j]]=True  
                    numeroFigurinhasPreenchidas=numeroFigurinhasPreenchidas+1  
                    #print("Numero Figurinhas Preenchidas", numeroFigurinhasPreenchidas)  
            simulations[i]=somaNumeroDePacotesUsados  
        media=media+somaNumeroDePacotesUsados  
    media=media/numeroAmostras  
    print("Media de 100 simulacoes:", media)  
    matplotlib.pyplot.hist(simulations,bins=30)
```



Os métodos mais comuns são:

- O método de amostragem natural
- O método de amostragem por inversão
- O método de amostragem por rejeição

Como extrair amostras de uma distribuição?

O método de amostragem natural

Métodos Computacionais
Professor Daniel O. Cajueiro
Economia (UnB)

A idéia aqui é simplesmente encontrar uma população que possui uma determinada distribuição e tomar amostras dessa população. Para isso garanta que cada elemento da população tenha a mesma probabilidade de ser selecionada.

Seja $F(x) = P(X \leq x)$ uma função de distribuição acumulada contínua e estritamente crescente com inversa F^{-1} dada por:

$$F^{-1}(u) = \{x \mid F(x) \geq u\} \quad (0 < u < 1)$$

Se U é uma variável aleatória uniforme no intervalo $[0, 1]$ então $F^{-1}(U)$ tem distribuição F :

Prova de Corretude:

$$\begin{aligned} P(F^{-1}(U) \leq x) \\ &= P(U \leq F(x)) && \text{(aplicando } F, \text{ em ambos os lados)} \\ &= F(x) && \text{(visto que } P(U \leq y) = y) \end{aligned}$$

Ele consiste basicamente de dois passos:

Passo 1: Você gera uma variável aleatória usando uma distribuição auxiliar, que você tem amostras de variáveis aleatórias disponíveis (por exemplo, a distribuição uniforme)

Passo 2: Você testa se essa variável aleatória poderia ter vindo da distribuição de interesse. Em caso positivo, você aceita. Em caso contrário, você rejeita.

Exemplo

Considere que você deseja gerar uma amostra da variável aleatória conhecida como triangular, cuja função de distribuição seja dada por:

$$f(x) = x \text{ se } x \in [0, 1]$$

$$f(x) = 2 - x \text{ se } x \in [1, 2]$$

Suponha que você tem disponível amostras de variáveis aleatórias com distribuição uniforme. Como você pode usar a distribuição uniforme para gerar a distribuição de uma variável aleatória triangular?

Exemplo

(Continuação)

Usando o método de amostragem por rejeição, você poderia fazer o seguinte:

- 1) Gerar uma variável uniforme (x, y) , onde x tem distribuição uniforme no intervalo $[0, 2]$ e y tem distribuição uniforme no intervalo $[0, 1]$.
- 2) Aceitar apenas os pontos x com a probabilidade exatamente igual aquela dada pela função de distribuição $f(x)$. Isso é feito aceitando as variáveis aleatórias que tenham $y \leq f(x)$. Veja as regiões definidas na figura abaixo.

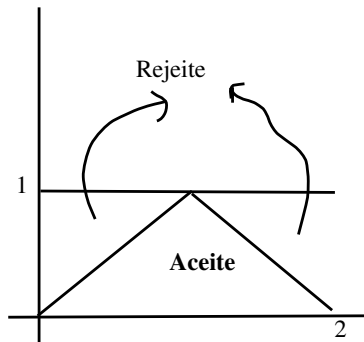
Como extrair amostras de uma distribuição?

O método de amostragem por rejeição

Métodos Computacionais
Professor Daniel O. Cajueiro
Economia (UnB)

Exemplo

(Continuação)



Como extrair amostras de uma distribuição?

O método de amostragem por rejeição

Métodos Computacionais
Professor Daniel O. Cajueiro
Economia (UnB)

```
import matplotlib.pyplot as plt
import random
import numpy as np
def triangularDistribution(n):
    sample=np.empty([n])
    i=0
    while(i<n):
        x=np.random.uniform(0,2)
        y=np.random.uniform(0,1)
        if(x<1):
            if(y<=x):
                sample[i]=x
                i=i+1
        else:
            if(y<=2-x):
                sample[i]=x
                i=i+1
    return sample
```

Como extrair amostras de uma distribuição?

O método de amostragem por rejeição

Métodos Computacionais
Professor Daniel O. Cajueiro
Economia (UnB)

```
if __name__ == '__main__':  
    n=5000 #numero de amostras  
    sample=triangularDistribution(n)  
  
    fig = plt.figure()  
    ax = fig.add_subplot(111)  
    ax=plt.hist(sample,bins=50)  
    plt.title('Distribuicao Triangular')
```

Como extrair amostras de uma distribuição?

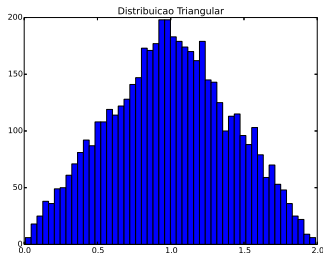
O método de amostragem por rejeição

Métodos Computacionais
Professor Daniel O. Cajueiro
Economia (UnB)

Exemplo

(Continuação)

Usando esse código, geramos uma amostra de variáveis aleatórias triangulares apresentada no seguinte histograma:



1. Marcelo recebeu 250000 reais de herança e deseja ficar um período sem trabalhar. Ele por levar uma vida minimalista e já possuir imóvel tem gastos mensais que podem ser bem modelados por uma distribuição normal com média 3000 e variância 1000. Considere também que ele tem acesso a um investimento que dá líquido 1% ao mês. (a) Por quantos meses em média ele consegue viver sem trabalhar? (b) Plot a distribuição do número de meses que ele pode ficar sem trabalhar. (c) Considere agora que a variância da sua distribuição de gastos aumenta para 2000. Como isso afeta os resultados? Sugira uma métrica mais robusta que a média para informar o número de meses que ele pode ficar sem trabalhar. Observação: Esse exercício pode ser implementado de duas formas. Usando um loop while e usando uma recursão. Faça uma função para cada uma das formas em python.

2. A maioria das linguagens computacionais já possui geradores de números aleatórios. Como esses números são gerados? Discuta a geração desses números tendo como base o Capítulo 7 [Random numbers] do livro Numerical Recipes in Fortran 77. Implemente pelo menos 2 exemplos de geradores de números aleatórios em python e discuta as diferenças (vantagens e desvantagens) entre eles.

3. O Capítulo 4 do livro “Monte Carlo Methods – Malvin H. Kalos and Paula A. Whitlock” explora diferentes aspectos do método de Monte Carlo. Discuta cada uma das técnicas apresentadas nas seções abaixo e implemente as ideias usando exemplos das seções desse livro. Explore todos os detalhes das implementações dos algoritmos apresentados. Use e abuse de figuras.

- a) 4.1
- b) 4.2
- c) 4.3
- d) 4.4
- e) 4.5

4. (★) Você pode resolver qualquer exercício apresentado no Capítulo 3 "Random Numbers" do Volume 2 The Art of Computer Programming que tenha score maior ou igual que 25 e satisfaça as condições abaixo. A solução inclui:

- a) Enunciar cuidadosamente o problema em português;
- b) Referenciar o problema, isto é, "Esse problema xxx apresentado no Livro Knuth no Capítulo XXX e na página XXX".
- c) Se depender de alguma teoria não estudada explicitamente em sala de aula, fazer cuidadosamente a conexão com a teoria em sala de aula apresentando os detalhes que permitam que seus colegas entendam a solução do seu problema.
- d) Mesmo quando não é pedido no problema, você deve implementar a solução ou algo relacionado com a solução na linguagem de sua escolha. Quando for o caso use e abuse de figuras geradas por você (e apresente os códigos usados para isso).

5. (★) Você pode resolver qualquer exercício apresentado no Capítulo 2 ou 3 do livro Monte Carlo Statistical Methods by Christian Robert and George Casella . A solução inclui:
- a) Enunciar cuidadosamente o problema em português;
 - b) Referenciar o problema, isto é, "Esse problema xxx apresentado no Livro do "Robert and Casella" no Capítulo XXX e na página XXX".
 - c) Se depender de alguma teoria não estudada explicitamente em sala de aula, fazer cuidadosamente a conexão com a teoria em sala de aula apresentando os detalhes que permitam que seus colegas entendam a solução do seu problema.
 - d) Mesmo quando não é pedido no problema, você deve implementar a solução ou algo relacionado com a solução na linguagem de sua escolha. Quando for o caso use e abuse de figuras geradas por você (e apresente os códigos usados para isso).

Fim

14 de setembro de 2021 – Brasília

