

Estimação da área debaixo de um gráfico através de Monte Carlo

Rennisson Davi D. Alves - 13687175

Abril de 2023

1 Introdução

No presente trabalho, discutiremos os 4 métodos de integração via Algoritmo de Monte Carlo, e a partir da análise de seus parâmetros, tentaremos estimar um N (amostra populacional) para cada um, que seja capaz de acertar com 95% de precisão o valor da integral de uma dada $f(x)$, dentro do intervalo $[0,1]$.

Os 4 métodos são os seguintes: Crude Monte Carlos, Hit or Miss, Importance Sampling e Control Variate.

No enunciado, foi definido como $f(x)$ a função: $e^{-Ax} \cos Bx$, onde $A = 0.386617636$ e $B = 0.47950399848$ (nossos respectivos RG e CPF). Tal função tem o grafico a seguir:

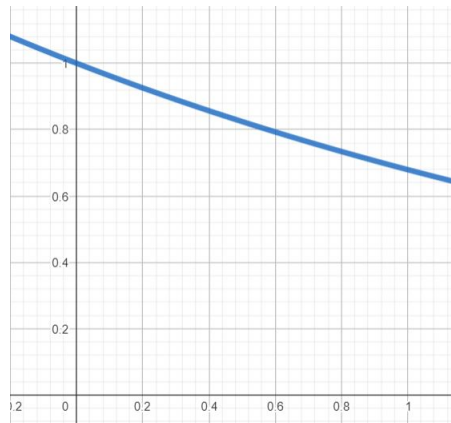


Figure 1: $f(x) = e^{-Ax} \cos Bx$

Por fim, como ferramentas foram utilizadas a linguagem de programação Python, suas bibliotecas Numpy e Scipy. As imagens dos gráficos foram feitas a partir do GeoGebra e algumas contas foram feitas com auxílio do WolframAlpha.

2 Crude Monte Carlo

É um resultado bastante famoso da estatística que a soma de variáveis aleatórias converge para a distribuição normal com média μ e variância $\frac{\sigma}{n}$. Isso foi visto no primeiro EP. Sendo assim, podemos estimar o tamanho da amostra pela seguinte relação:

$$n = \frac{z_{\alpha}^2 \sigma^2}{(0.0005\gamma)^2} \quad (1)$$

Foi dado que $\sigma^2 = \frac{1}{n} \int_0^1 (f(x) - \gamma)^2 dx$. Desenvolvendo, temos $\sigma^2 = \frac{1}{n} (\int_0^1 f^2(x) dx - 2\gamma \int_0^1 f(x) dx + \int_0^1 \gamma^2 dx)$. Vale notar que $\int_0^1 f(x) dx = \gamma$ e $\int_0^1 \gamma^2 dx = \gamma^2$.

Desse modo, $\sigma^2 = \frac{1}{n} (F - 2\gamma^2 + \gamma^2) = \frac{1}{n} (F - \gamma^2)$, onde $F = \int_0^1 f^2(x) dx$.

Pela figura 1 mostrada na introdução, é fácil ver que nossa função é positiva, decrescente e sua concavidade está virada para cima. Fazendo a análise através das derivadas, chegamos na mesma conclusão.

Também, a imagem da função $f(x)$ está no intervalo $[0,1]$. Portanto, neste caso, $0 \leq f^2(x) \leq f(x) \leq 1$. E quando $x = 0$, $f^2(0) = f(0) = 1$. Desse modo, $f^2(x)$ tem as mesmas propriedades de $f(x)$: positiva, decrescente e com concavidade para cima.

Com o que foi feito até aqui, podemos definir que $f(x) < r_1$ e $f^2(x) < r_2$, onde r_1 e r_2 são as retas a seguir:

$$\begin{aligned} r_1 : (f(1) - f(0))x + f(0) &\Rightarrow r_1 : [(0.60273 - 1)x + 1] \\ r_2 : (f^2(1) - f^2(0))x + f^2(0) &\Rightarrow r_2 : [(0.36328 - 1)x + 1] \end{aligned}$$

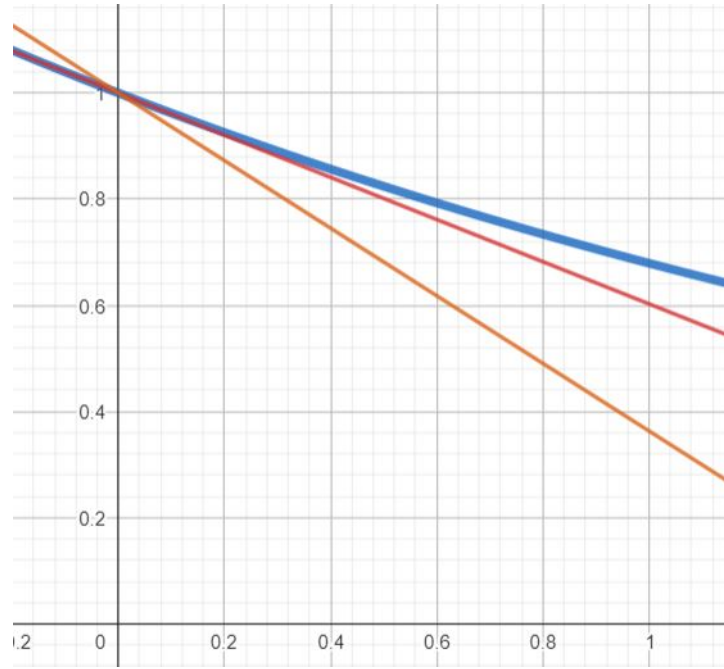


Figure 2: Em vermelho, a reta r_1 . Em laranja, a reta r_2 . Em azul, a função $f(x)$

Agora voltando ao core do nosso problema, para encontrar nosso $n = \frac{1}{n} (F - \gamma^2)$, vamos utilizar nossos resultados até aqui para chegar em F e γ^2 .

$$\begin{aligned} F &< \int_0^1 (f^2(1) - 1)x - 1 dx = \frac{(e^{-a} \cos(b))^2 + 1}{2} \\ \gamma &> f(1)(1 - 0) = e^{-a} \cos(b) \end{aligned}$$

Com estes resultados, temos

$$n = \frac{z_\alpha^2 (F - \gamma^2)}{(0.0005\gamma)^2} = \frac{1.96^2 \left(\frac{(e^{-a} \cos(b))^2 + 1}{2} - (e^{-a} \cos(b))^2 \right)}{(0.0005e^{-a} \cos(b))^2}$$

Substituindo $a = 0.386617636$ e $b = 0.47950399848$, obtemos $n = 13465687.63372519$. Logo, utilizaremos $n = 13465688$ como tamanho da nossa amostra para os testes a serem realizados.

3 Hit or Miss Monte Carlo

Esse método é bem similar ao método usado no EP1, por isso vamos continuar utilizando $n = \frac{z_\alpha^2 \sigma^2}{(0.0005\gamma)^2}$. Foi dado que $\sigma^2 = \frac{\gamma(1-\gamma)}{n}$. Assim, a estimação do tamanho da amostra fica $n = \frac{z_\alpha^2(1-\gamma)}{0.0005^2\gamma}$.

Aproveitando os resultados da seção anterior, sabemos que $f(1) = e^{-a} \cos(b) \leq \gamma$. Com isso então, segue que $n = \frac{1.96^2(1-e^{-a} \cos(b))}{0.0005^2 e^{-a} \cos(b)}$. Novamente fazendo $a = 0.386617636$ e $b = 0.47950399848$, obtemos $n = 10128001.95486523$.

Logo, $n = 10128002$ é o tamanho da amostra que vamos utilizar nos testes a serem realizados.

4 Importance Sampling

Para esse método, escolhi por inspeção visual $g(x) = \beta(0.35, 1)$ já que a f.d.p. da Distribuição Beta tem o domínio de 0 a 1 e para os parâmetros descritos ela aparenta proporcionalidade já que é decrescente e de concavidade voltada pra cima assim como nossa $f(x)$, como vimos na seção anterior. Segue um esboço de como é a f.d.p. da função Beta em questão:

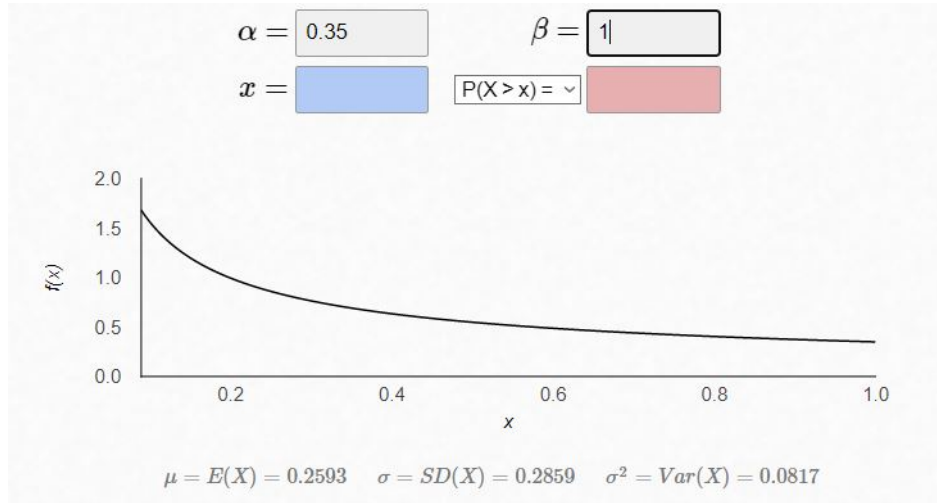


Figure 3: $\beta(0.35, 1)$

Foi-nos dado que $Var[\hat{\gamma}_s] = \frac{\int_0^1 (\frac{f(x)}{g(x)} - \gamma)^2 g(x) dx}{n}$. Agora vamos desenvolver um pouco a variância em questão:

$$\sigma_s^2 = \int_0^1 \frac{f^2(x)}{g(x)} - 2\gamma f(x) + \gamma^2 g(x) dx$$

Note que $\int_0^1 f(x) = \gamma$ e $\int_0^1 g(x) = 1$. Além disso, vamos chamar $\int_0^1 \frac{f^2(x)}{g(x)} = Y$. Logo

$$\sigma_s^2 = \frac{(Y - \gamma^2)}{n}$$

Vamos também desenvolver Y:

$$\frac{f^2(x)}{g(x)} = (e^{-ax} \cos(bx))^2 \sqrt{x} \cdot \beta(0.35, 1) = (e^{-ax} \cos(bx))^2 2.86\sqrt{x}$$

Sabemos da seção anterior que $f^2(x)$ é positiva, decrescente de máximo igual a 1, ou seja, $0 < f^2(x) < f(x) < 1$. Note também que $2.86\sqrt{x}$ é uma metade de parábola sobre o eixo das abscissas, como segue o esboço:

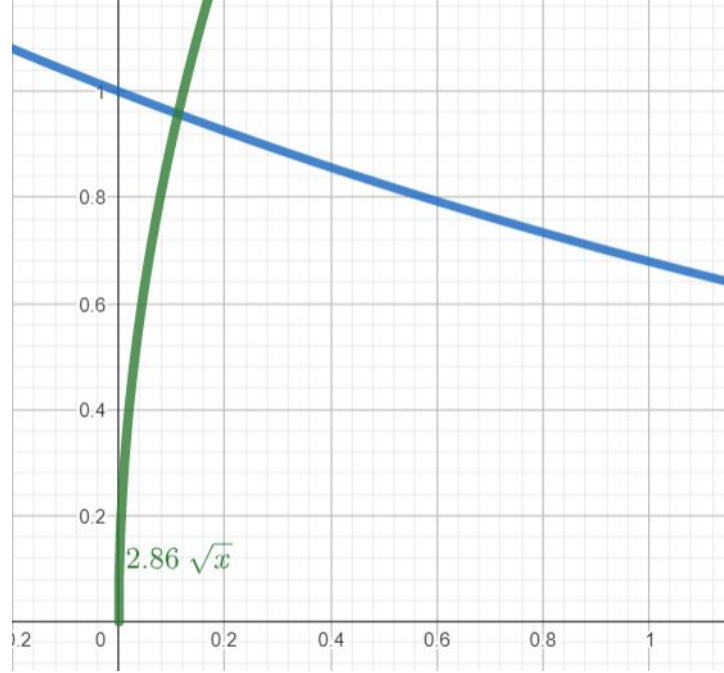


Figure 4: $g(x) = 2.86\sqrt{x}$

Com isso podemos calcular a área debaixo de $2.86\sqrt{x}$

$$\int_0^1 2.86\sqrt{x} dx = \frac{2.86 \cdot 2}{3} = 1.90667$$

e portanto, $Y \leq 1.90667$. Junto com o que sabemos da seção prévia, em que $f(1) \leq \gamma$. Ocorre para um tamanho de amostra n o que segue:

$$n = \frac{1.96^2(Y - \gamma^2)}{(0.0005\gamma)^2}$$

Substituindo os devidos valores, então:

$$n = \frac{1.96^2(1.90667 - f^2(1))}{(0.0005f(1))^2} = \frac{1.96^2(1.90667 - 0.36329)}{0.0005^2 \cdot 0.36329} = 65281715.52203473$$

Portanto, $n = 65281716$ é o tamanho da amostra que será utilizada nos testes a serem realizados.

5 Control Variate

Para este método, vamos assumir a função $\varphi(x)$ como sendo a reta r_1 descrita na seção 2, já que por inspeção visual julgamos que ela tem uma boa mimetização para nossa $f(x)$. Dessa forma então:

$$\gamma' = \int_0^1 \varphi(x) dx = f(1) + \frac{1-f(1)}{2} = \frac{e^{-a} \cos(b)+1}{2}$$

Seja agora

$$Var[\hat{\gamma}] = \frac{Var[f(x_i)] + Var[\varphi_i] - 2\rho(f(x_i), \varphi(x_i)) \sqrt{Var[f(x_i)]Var[\varphi(x_i)]}}{n}$$

vamos desenvolver cada parcela da expressão afim de facilitar a obtenção do tamanho da amostra (n). Vale antes citar um resultado importante para seguirmos, assuma $h(x)$ uma função, então ocorre:

$$X \text{ Unif}(0, 1) \Rightarrow E_x(h(x)) = \int_0^1 h(x) dx$$

Agora, tomando F e γ da seção anterior:

$$Var[f(x_i)] = E(f^2(x_i)) - E^2(f(x_i)) = \int_0^1 f^2(x) dx - (\int_0^1 f(x))^2 dx = F - \gamma^2$$

Analogamente:

$$Var[\varphi(x_i)] = \int_0^1 \varphi^2(x) dx - (\int_0^1 \varphi(x))^2 dx$$

Mas como $\varphi^2(x) < \varphi(x) \Rightarrow \int_0^1 \varphi^2(x) dx \leq \int_0^1 \varphi(x) dx$, então vamos optar por utilizar $Var[\varphi(x_i)] = \gamma' - \gamma'^2$

Assumindo correlação alta tal que $\rho = 0.9$, além de tudo que já foi citado, temos então que

$$Var[\hat{\gamma}] = \frac{F - \gamma^2 + \gamma' - \gamma'^2 - 2 \cdot 0.9 \sqrt{(F - \gamma^2)(\gamma' - \gamma'^2)}}{n}$$

Assumindo os valores desejados para os parâmetros em questão (considerando as desigualdades desenvolvidas na seção 2.1 e o que fora citado inicialmente nesta mesma seção) onde substituímos $f(1)$ quando há γ , $\frac{f^2(1)+1}{2}$ quando há F e $\frac{f(1)+1}{2}$ quando há γ' , obtemos:

$$n = \frac{1.96^2 Var[\hat{\gamma}]}{(0.0005\gamma)^2} = \frac{1.96^2 \frac{0.07233}{n}}{(0.0005f(1))^2} = 3059461.42916106$$

Logo, $n = 3059462$ é o tamanho da amostra que vamos utilizar nos testes a serem realizados.

6 Implementação, experimentos e conclusões

Todos os experimentos realizados com os N obtidos através da inspeção analítica foram bem sucedidos. Cada um deles atingiu em mais de 95% dos casos o intervalo de confiança requerido.

Como todos o N são bem altos, foi necessária a utilização da biblioteca *numpy* para otimização do código. Com ela, o tempo de execução melhorou muito. A melhora de performance compensa a dificuldade de leitura do código, apesar de não ser muito difícil a sua leitura.

Implementei cada método como se fosse uma classe. Cada classe tem seu metodo construtor, que varia de acordo com sua lei de formação e estimação. Há também o método *estimativa()*, que vai retornar o valor da estimativa do método para a integral de $f(x)$.

A variável *f_values*, comum em todas as classes implementadas, guarda o valor da $f(x)$ enunciada em todos os pontos sorteados randomicamente (guardados no array *points*) dentro do intervalo especificado, portanto ela é um array de tamanho n.

Para padronização dos testes e para facilitar a averiguação dos resultados, foi configurar uma seed fixa no início do programa, que no caso é o NUSP 13687175.

A seguir estão colocadas os resultados dos experimentos para cada um dos métodos. Como dito anteriormente, todos eles foram bem-sucedidos, atingindo com precisão acima de 95% o intervalo de confiança requerido.

CRUDE			Hit or Miss		
Sample		Estimativa	Sample		Estimativa

13465688		0.8008788478315866	10128002		0.8006826025508289
13465688		0.8009218503594411	10128002		0.8007146918019961
13465688		0.8009570310358601	10128002		0.8007799564020623
13465688		0.8009292519046286	10128002		0.8009697272966573
13465688		0.8009291644653034	10128002		0.800900611986451
13465688		0.8009263533942478	10128002		0.8009945100721741
13465688		0.8009650802962974	10128002		0.8008699050414879
13465688		0.8008938455172597	10128002		0.8008088860961915
13465688		0.800960583675736	10128002		0.8008561807156042

Figure 5: Testes dos métodos Crude e Hit or Miss

IMPORTANCE SAMPLING			CONTROL VARIATE		
Sample		Estimativa	Sample		Estimativa

65281716		0.8009519277028223	3059462		0.8009310664356926
65281716		0.8009454432232516	3059462		0.8009304784136317
65281716		0.800999393410472	3059462		0.8009304766108872
65281716		0.8010442804039266	3059462		0.800930274471663
65281716		0.8009271581669657	3059462		0.8009320989304807
65281716		0.8009659115456615	3059462		0.8009303607360677
65281716		0.8008631108133429	3059462		0.8009308901300582
65281716		0.8010336151368975	3059462		0.8009310922011245
65281716		0.8008807135475341	3059462		0.8009302346636975

Figure 6: Teste dos métodos Importance Sampling e Control Variate

Para comparação, a integral definida da nossa $f(x)$ é 0.800931. Veja que todas as estimativas se aproximaram muito do valor real.