

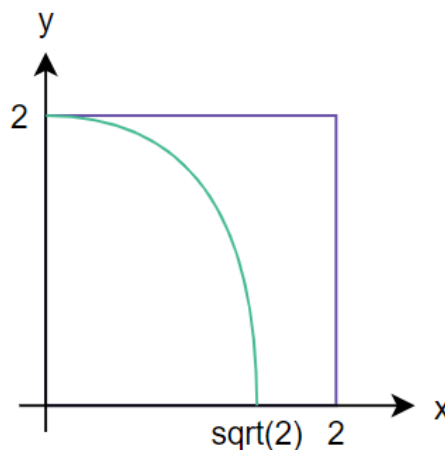
1 **Calculando $\sqrt{2}$:** Vimos em aula um algoritmo de Monte Carlo para calcular o valor de π utilizando a relação entre áreas. Inspirado nesta mesma ideia, construa um algoritmo de Monte Carlo para calcular o valor de $\sqrt{2}$:

1.1 Descreva a variável aleatória cujo valor esperado está relacionado com $\sqrt{2}$. Obtenha analiticamente o valor esperado da sua variável aleatória.

Considere a curva

$$y = 2 - x^2$$

e o quadrado de lado 2 representados na figura abaixo:



Além disso, seja A_q a área do quadrado e A_c a área embaixo da curva $y \leq 2 - x^2$. Desta forma, a área A_c é dada por:

$$\begin{aligned} A_c &= \int_0^{\sqrt{2}} 2 - x^2 \\ &= \left[2x - \frac{x^3}{3} \right]_0^{\sqrt{2}} \\ &= 2\sqrt{2} - \frac{2\sqrt{2}}{3} \\ &= \frac{4}{3}\sqrt{2} \end{aligned}$$

e a razão entre as áreas é dada por:

$$\frac{A_t}{A_c} = \frac{\sqrt{2}}{3}$$

Podemos definir 3 variáveis aleatórias: as variáveis X e Y , uniformes no intervalo $[0, 2]$, e a variável aleatória indicadora $g(x, y)$ definida como:

$$g(x, y) = \begin{cases} 1, & \text{se } y \leq 2 - x^2 \\ 0, & \text{caso contrário} \end{cases}$$

Pela formulação, a variável aleatória indicadora pode ser modelada como uma Bernoulli de parâmetro p igual à razão entre as áreas A_t e A_q :

$$p = \frac{A_t}{A_q} = \frac{\sqrt{2}}{4}$$

Como $g(x, y)$ é modelada como uma Bernoulli, seu valor esperado é dado por:

$$\mathbb{E}[g(x, y)] = p = \frac{\sqrt{2}}{4}$$

1.2 Calcule analiticamente a variância dessa variável aleatória.

A variância de uma indicadora modelada como uma Bernoulli é dada por:

$$\sigma^2 = p(1 - p)$$

Portanto, a variância de $g(x, y)$ é dada por:

$$\sigma^2 = \frac{\sqrt{2}}{4} \left(1 - \frac{\sqrt{2}}{4}\right) = 0.65$$

1.3 Implemente o método de Monte Carlo para gerar amostras da sua variável aleatória, calculando a média amostral M_n e utilizando-a para estimar $\sqrt{2}$. Sua função deve usar o algoritmo de Mersenne Twister para geração de números pseudo-aleatórios uniformes entre 0 e 1 (usar este algoritmo em todos os problemas da lista).

Dado que um experimento seja a geração de n amostras e a obtenção da média amostral, realizou-se o experimento 5 vezes para n amostras e a média e desvio padrão dos experimentos são apresentados na Figura 2.2.

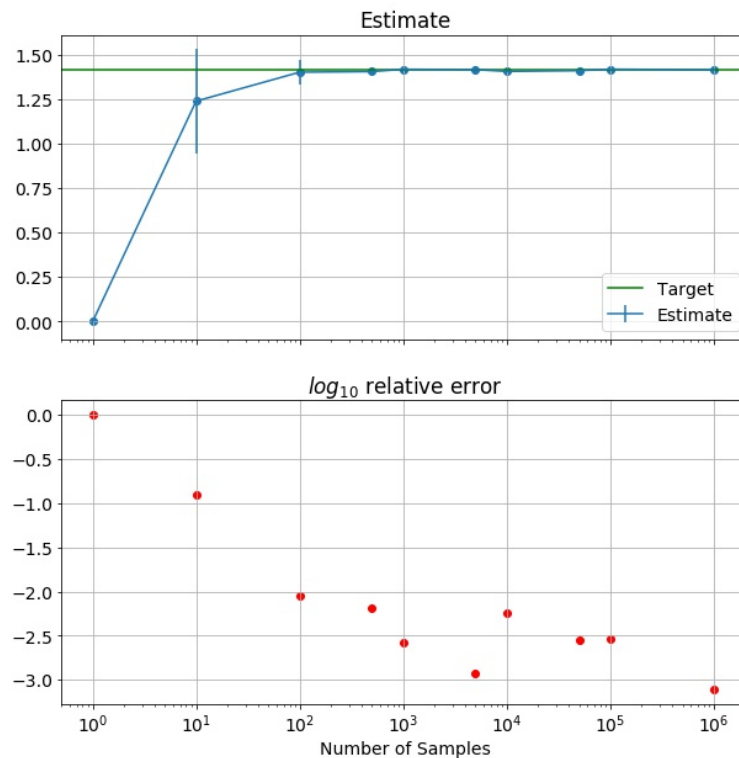


Figure 1: Média e desvio padrão para 5 realizações do experimento de retirada da média amostral utilizando o método Monte Carlo.

Observe que ao aumentar o número n de amostras, a estimativa do valor esperado se aproxima do seu valor teórico com uma variância decrescente com o número de amostras n . Isso corrobora com o cálculo teórico de que a variância das médias amostrais σ_n^2 diminui com o número de amostras conforme:

$$\sigma_n^2 = \frac{\sigma^2}{n}$$

onde σ^2 é a variância da variável aleatória.

1.4 Seja \hat{e}_n o valor do estimador após n amostras. Trace um gráfico do erro relativo do estimador, ou seja $|\hat{e}_n - \sqrt{2}|/\sqrt{2}$ em função de n , para $n = 1, \dots, 10^6$ (utilize escala log-log no gráfico). O que você pode concluir?

Ainda na Figura 2.2, note que o erro relativo entre o que se estima e seu valor teórico de referência também reduz com o número de amostras. Para um total de 10^6 amostras, observou-se que o erro relativo \hat{e} chega a valores inferiores a 10^{-3} .

2 Transformada Inversa: Utilize o método da transformada inversa para gerar amostras de uma V.A. X com as seguintes densidades:

2.1 Distribuição exponencial com parâmetro $\lambda > 0$, cuja função densidade é dada por $f_X(x) = \lambda e^{-\lambda x}$, para $x \geq 0$.

A CDF da distribuição exponencial é dada por:

$$\begin{aligned} F_X(i) &= P[X \leq i] \\ &= \int_0^i \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^i e^{-\lambda x} dx \\ &= \lambda \left[\frac{1}{-\lambda} e^{-\lambda x} \right]_0^i \\ &= 1 - e^{-\lambda i} \end{aligned}$$

O método da transformada inversa é aplicado resolvendo-se a equação $F_X(y) = x$:

$$\begin{aligned} x &= 1 - e^{-\lambda y} \\ y &= \frac{\log(1 - x)}{\lambda} \end{aligned}$$

Assim

$$f_X^{-1}(x) = \frac{\log(1 - x)}{\lambda}$$

Tomando-se 100k amostras $x \sim U[0, 1]$ para diferentes valores de λ , obtemos a *complementary cumulative distribution function* (CCDF) das amostras de $f_X(x)$ na Figura 2:

2.2 Distribuição de Pareto com parâmetros $x_0 > 0$ e $\alpha > 0$, cuja função densidade é dada por $f_X(x) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}$

A CDF da distribuição de pareto é dada por:

$$\begin{aligned} F_X(i) &= P[X \leq i] \\ &= \int_{x_0}^i \frac{\alpha x_0^\alpha}{x^{\alpha+1}} dx \\ &= \alpha x_0^\alpha \int_{x_0}^i x^{-\alpha-1} dx \\ &= x_0^\alpha \left[-x_0^{-\alpha} \right]_{x_0}^i \\ &= 1 - \left(\frac{x_0}{i} \right)^\alpha \end{aligned}$$

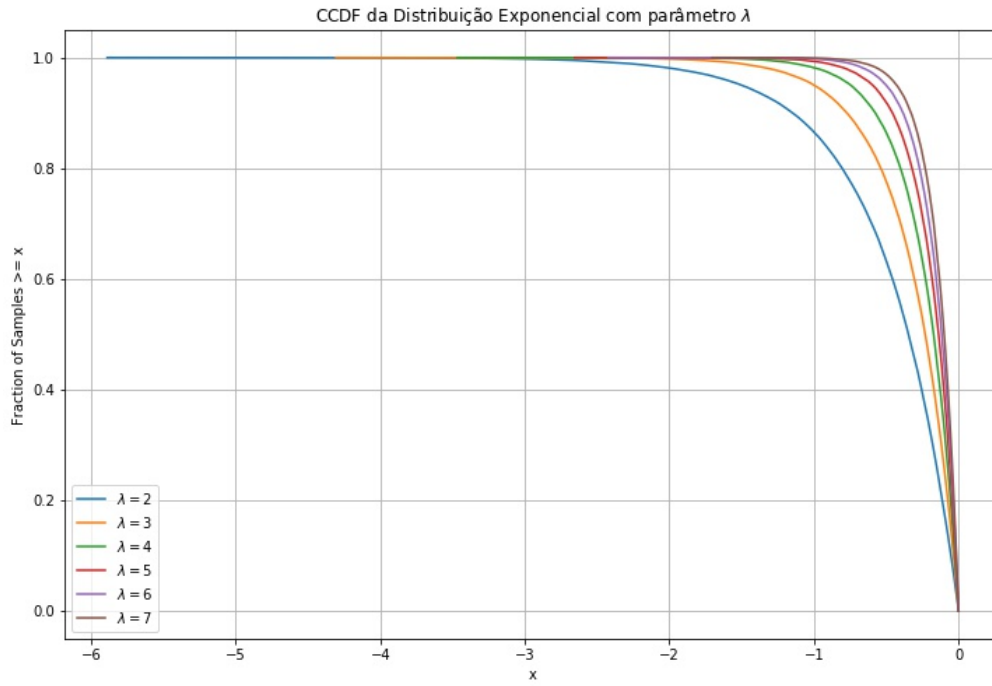


Figure 2: CCDF da distribuição exponencial com parâmetro λ

Analogamente, resolvemos a equação $F_X(y) = x$:

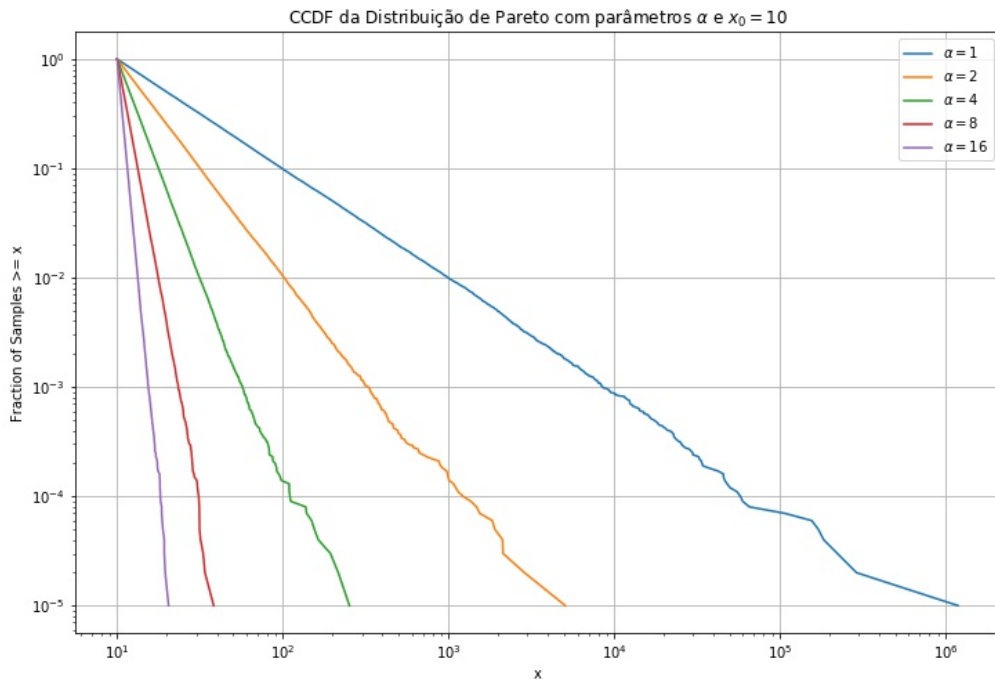
$$x = 1 - \left(\frac{x_0}{y}\right)^\alpha$$

$$y = \frac{x_0}{(1-x)^{1/\alpha}}$$

Assim,

$$F_X^{-1}(x) = \frac{x_0}{(1-x)^{1/\alpha}}$$

Tomando-se 100k amostras $x \sim U[0, 1]$ para diferentes valores de α e utilizando $x_0 = 10$, obtemos a CCDF das amostras de $f_X(x)$ na Figura 2.2:



3 Contando domínios na Web: Quantos domínios web existem dentro da UFRJ? Mais precisamente, quantos domínios existem dentro do padrão de nomes $http://www.[a-z](k).ufrj.br$, onde $[a-z](k)$ é qualquer sequência de caracteres de comprimento k ou menor? Construa um algoritmo de Monte Carlo para estimar este número.

3.1 Descreva a variável aleatória cujo valor esperado está relacionado com a medida de interesse. Obtenha analiticamente o valor esperado da sua variável aleatória.

Seja W_k o conjunto de sequência de caracteres de comprimento k ou menos. Podemos definir uma variável aleatória $g(w)$ tal que:

$$g(w) = \begin{cases} 1, & \text{se } w \text{ é domínio} \\ 0, & \text{caso contrário} \end{cases}$$

Considerando que todos os domínios tenham a mesma probabilidade de existir:

$$f_w(W) = \frac{1}{|W|}$$

Desta forma, o valor esperado desta variável é dado por:

$$\begin{aligned} \mathbb{E}[g(w)] &= \sum_{i=1}^{|W|} f_w(i) \cdot g(i) \\ &= \sum_{i=1}^{|W|} \frac{1}{|W|} \cdot 1 \\ &= \frac{1}{|W|} \sum_{i=1}^{|W|} 1 \\ &= \frac{1}{|W|} S_k \end{aligned}$$

onde S_k é a soma de todos os domínios possíveis com até k caracteres. Portanto,

$$S_k = \mathbb{E}[g(w)]|W|$$

No entanto, o valor esperado $\mathbb{E}[g(w)]$ pode ser estimado por amostragem através de:

$$M_n = \frac{1}{n} \sum_{i=1}^n g(w_i)$$

Finalmente,

$$S_k \approx M_n |W|$$

3.2 Calcule a variância dessa variável aleatória.

Como $g(w)$ pode ser vista como uma Bernoulli de parâmetro $p = 1/|W|$, sua variância é dada por:

$$\text{Var}(g(w)) = p(1-p) = \frac{1}{|W|} \left(1 - \frac{1}{|W|}\right)$$

A variância do estimador M_n é dada por:

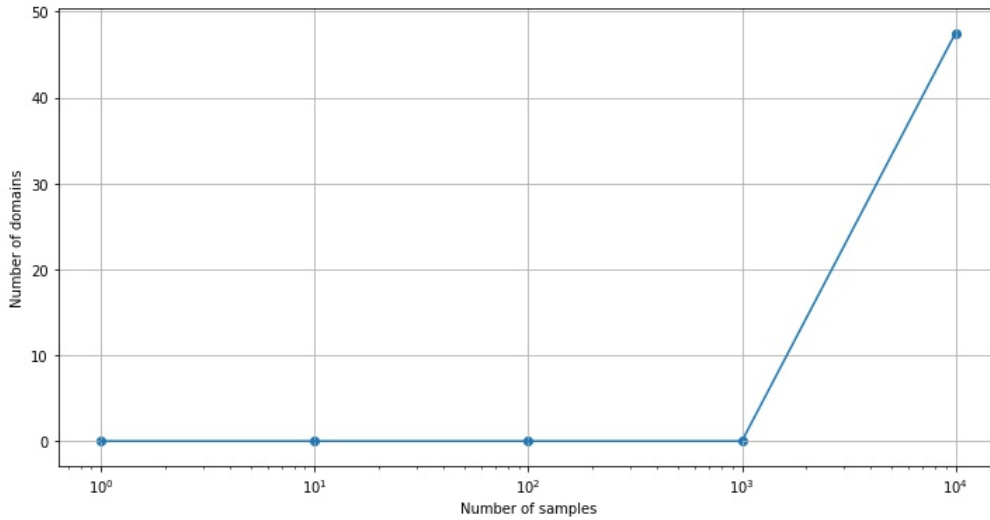
$$\begin{aligned}
Var(M_n) &= Var\left(\frac{1}{n} \sum_{i=1}^n g(w_i)\right) \\
&= \frac{1}{n} \sum_{i=1}^n Var(g(w_i)) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{1}{|W|} \left(1 - \frac{1}{|W|}\right) \\
&= \frac{1}{|W|} \left(1 - \frac{1}{|W|}\right) \frac{1}{n} \cdot n \\
&= \frac{|W| - 1}{|W|^2}
\end{aligned}$$

3.3 Implemente o método de Monte Carlo para gerar amostras da sua variável aleatória. Ou seja, você deve consultar o domínio gerado para determinar se o mesmo existe (utilize uma biblioteca *www* para isto).

A implementação do método encontra-se disponível neste [github](#).

3.4 Assuma que $k = 4$. Seja \hat{w}_n o valor do estimador do número de domínios após n amostras. Trace um gráfico de \hat{w}_n em função de n para $n = 1, \dots, 10^4$ (ou mais, se conseguir). O que você pode dizer sobre a convergência de \hat{w}_n ?

A Figura 3.4 apresenta o resultado do número de domínios encontrados para diferentes valores de n . Podemos perceber que o número de domínios encontrados para $k = 4$ é consideravelmente baixo mesmo para um número de amostras como $n = 10^4$. Para as potências de 10 inferiores a esta, nenhum domínio pôde ser encontrado. Desta forma, não se pode afirmar que o algoritmo converge para $n \leq 10^4$.



4 Gerando amostras Normais: Seja Z uma variável aleatória com distribuição Normal com média 0 e variância 1. Em particular, a função densidade de Z é dada por $f_Z(x) = 1/\sqrt{2\pi} \cdot e^{-x^2/2}$ com $-\infty < x < \infty$. Repare que Z assume valores positivos e negativos, mas com caudas que possuem a mesma probabilidade. Ou seja, $P[Z \geq z] = P[Z \leq -z]$, para todo $z \geq 0$. Construa um gerador de números aleatórios para Z . Dica: Utilize o método de amostragem por rejeição!

Para aplicarmos o método da amostragem por rejeição, precisamos de uma função $f_Y(x)$ tal que $\lambda f_Y(x) \geq f_Z(x)$. Como $f_Z(x)$ é simétrica em relação à origem, podemos amostrar apenas valores positivos de x e espelhar a distribuição. Assim, a distribuição $\lambda f_Y(x)$ deve estar definida para $x \in [0, \infty)$.

Neste caso, podemos fazer com que $c f_Y(x)$ tenha uma distribuição geométrica tal que o parâmetro λ escolhido satisfaça a inequação.

Assim, seja $\lambda f_Y(x) = \lambda e^{-\lambda x}$:

$$\begin{aligned} c f_Y(x) &\geq f_Z(x) \\ \lambda e^{-\lambda x} &\geq 1/\sqrt{2\pi} \cdot e^{-x^2/2} \\ e^{\ln \lambda - \lambda x} &\geq e^{-\ln \sqrt{2\pi} - x^2/2} \\ \ln \lambda - \lambda x &\geq -\ln \sqrt{2\pi} - x^2/2 \\ \frac{1}{2}x^2 - \lambda x + \ln \lambda \sqrt{2\pi} &\geq 0 \\ ax^2 + bx + c &\geq 0 \end{aligned}$$

A inequação acima possui um polinômio do 2o grau cuja parábola tem a concavidade para cima. Ou seja, caso o determinante $\Delta = b^2 - 4ac$ seja menor ou igual a zero, a parábola possui apenas 1 ou nenhuma raiz e a inequação é satisfeita. Assim,

$$\begin{aligned} \Delta &= b^2 - 4ac \\ &= \lambda^2 - 2 \ln \lambda \sqrt{2\pi} \end{aligned}$$

O polinômio possui 1 ou nenhuma raiz quando Δ é menor ou igual a zero, o que acontece para $\lambda \in [0.4, 1.7]$. Escolhendo $\lambda = 1$, temos:

$$f_Y(x) = e^{-x}$$

Vimos no exercício 2.1 que, a partir de uma variável aleatória uniforme X , podemos gerar amostras da distribuição exponencial através da transformada inversa:

$$y \sim F_Y^{-1}(x) = -\log(1 - x)$$

Assim, o método da amostragem por rejeição nos permite criar amostras de $f_Z(x)$. Tomando-se um número aleatório $u \sim U[0, 1]$, a amostra $z = y$ é aceita se $u < f_Z(x)$ e rejeitada caso contrário.

Tomando-se 50k amostras $x \sim U[0, 1]$, obtemos o histograma da Figura 3.

Note que, pela modelagem, a média naturalmente é nula pois forçou-se que as amostras x e $-x$ estivessem presentes no histograma, para todos os x amostrados. Além disso, a variância apresentada se aproximou do valor teórico de 1.

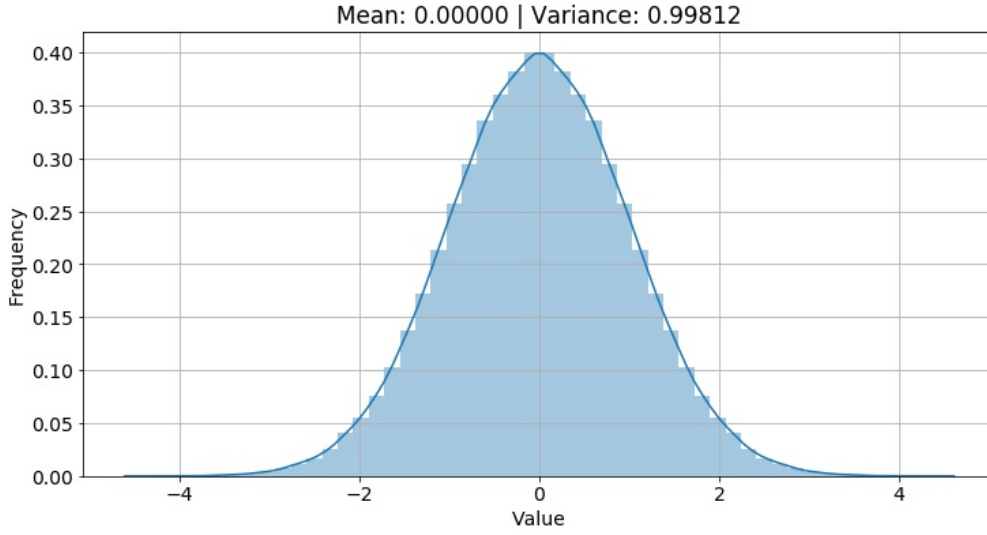


Figure 3: Histograma de 50k amostras para geração de uma normal com $\mu = 0$ e $\sigma^2 = 1$.

5 Estimando somas com Importance Sampling: Considere o problema visto em aula, de aplicar o método de Monte Carlo para estimar o valor de $G_N = \sum_{i=1}^N i \log i$. Use sua intuição para encontrar uma função de probabilidade proponente, $h(i)$, que tenha variância inferior ao melhor estimador visto em aula.

5.1 Assuma que $N = 10^3$. Calcule numericamente o segundo momento do seu estimador.

Seja $h(i) = i^m / K$ onde a constante de normalização K é dada por $K = \sum_{j=1}^N j^m$. Queremos escolher m de forma a minimizar o segundo momento

$$\mathbb{E}_h \left[\left(\frac{g(i)}{h(i)} \right)^2 \right]$$

Desenvolvendo o segundo momento, temos:

$$\begin{aligned} \mathbb{E}_h \left[\left(\frac{g(i)}{h(i)} \right)^2 \right] &= \sum_{i=1}^N \left(\frac{g(i)}{h(i)} \right)^2 h(i) \\ &= \sum_{i=1}^N \frac{i^2 \log^2 i \cdot K^2 i^m}{i^{2m}} \frac{1}{K} \\ &= K \sum_{i=1}^N \frac{\log^2 i}{i^{m-2}} \end{aligned}$$

Utilizando diversos valores de m , os valores do segundo momento são apresentados na Figura 4.

Vemos empiricamente que, para $m = 1.2$, obtemos um valor de 1.024×10^{13} para o segundo momento, o que é menor do que o valor 1.03×10^{13} obtido na aula para $m = 1$.

5.2 Implemente o método de Monte Carlo para estimar o valor de G_N . Trace um gráfico do erro relativo do estimador, em função de $n = 1, \dots, 10^7$ (calcule o valor exato da soma para determinar o erro relativo).

O valor exato da soma é dado por:

$$G_{N=10^3} = \sum_{i=1}^N i \log i = 3207332.341$$

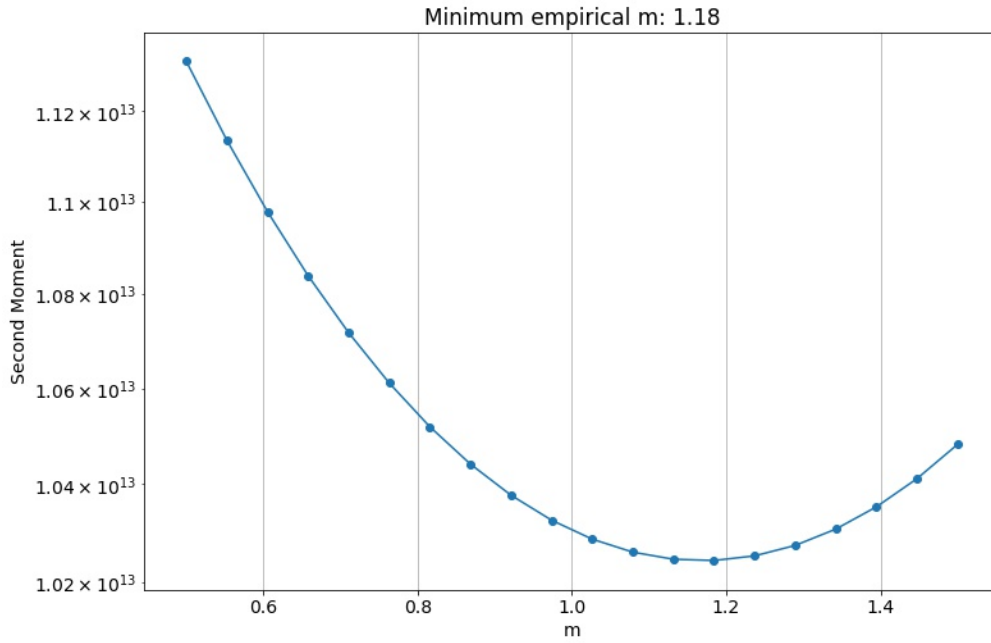


Figure 4: Varredura do parâmetro m para encontrar valores do segundo momento.

O primeiro momento do estimador é dado por:

$$\begin{aligned}\mathbb{E}_h \left[\frac{g(i)}{h(i)} \right] &= \sum_{i=1}^N \frac{g(i)}{h(i)} h(i) \\ &= \sum_{i=1}^N g(i) \\ &= G_N\end{aligned}$$

Assim, o estimador é não-enviesado e basta que apliquemos monte-carlo para estimar o valor esperado e, consequentemente, o valor da soma G_N . Para isso, precisamos de amostras de $h(i)$, o que pode ser feito através do método da transformada inversa.

Para a pdf $h(i) = x^m/K$, a CDF é dada por:

$$Y = H(i) = \frac{x^{m+1}}{(m+1)K}$$

Cuja inversa é dada por:

$$\begin{aligned}H(Y) &= i \\ \frac{Y^{m+1}}{(m+1)K} &= i \\ Y &= \left[(m+1)Ki \right]^{1/(m+1)}\end{aligned}$$

Para $n = 10^7$ amostras, temos a distribuição de $h(i)$ apresentada na Figura 5.

Após amostrarmos $y \sim h(i)$, aplicarmos $g(y)/h(y)$ e retirando-se a média do resultado, temos a evolução do erro relativo na Figura 6.

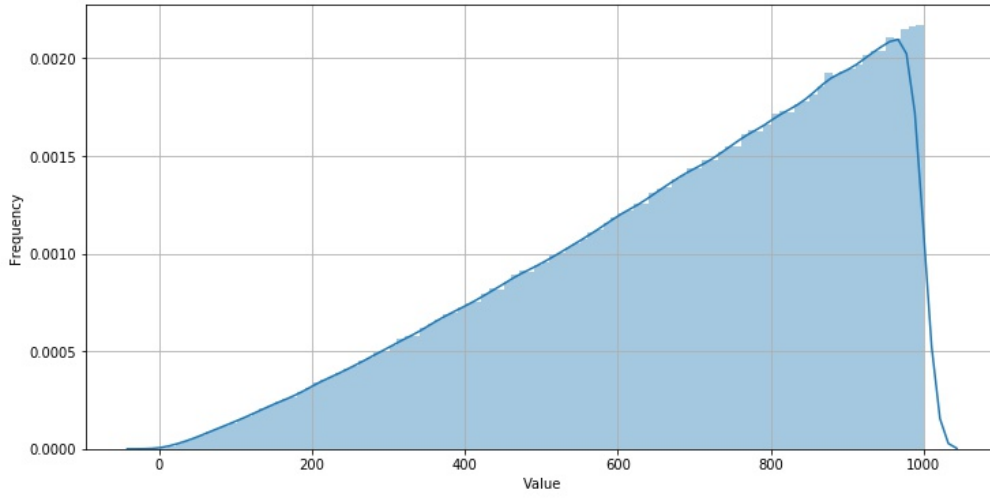


Figure 5: Distribuição de $h(i)$ para $n = 10^7$ amostras.

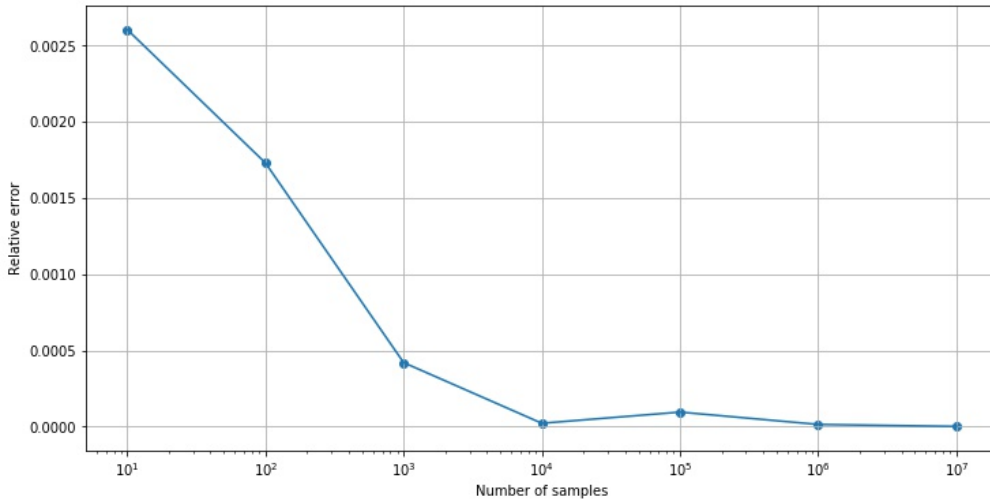


Figure 6: Evolução do erro relativo com o número de amostras.

6 Integração de Monte Carlo: Considere a função $f(x) = x^\alpha$ com $\alpha > 0$. Defina $g(\alpha, a, b) = \int_a^b f(x)dx$ com $0 \leq a < b$, como sendo a integral de $f(x)$ no intervalo $[a, b]$. Iremos calcular g usando Monte Carlo.

6.1 Determine analiticamente o valor de $g(\alpha, a, b)$. Dica: relembre Cálculo I.

$$\begin{aligned} g(\alpha, a, b) &= \int_a^b f(x)dx \\ &= \int_a^b x^\alpha dx \\ &= \frac{1}{\alpha + 1} x^{\alpha+1} \Big|_a^b \\ &= \frac{1}{\alpha + 1} (b^{\alpha+1} - a^{\alpha+1}) \end{aligned}$$

6.2 Descreva a variável aleatória cujo valor esperado está relacionado com $g(\alpha, a, b)$. Obtenha analiticamente o valor esperado da sua variável aleatória.

Seja X uma variável aleatória uniforme definida no intervalo $[a, b]$.

$$\begin{aligned}
\mathbb{E}_X[f(x)] &= \int_a^b f_X(x)f(x)dx \\
&= \int_a^b \frac{1}{b-a} f(x)dx \\
&= \frac{1}{b-a} \int_a^b f(x)dx \\
&= \frac{1}{b-a} g(\alpha, a, b)
\end{aligned}$$

Assim,

$$g(\alpha, a, b) = (b-a)\mathbb{E}_X[f(x)]$$

Podemos estimar $\mathbb{E}_X[f(x)]$ pelo método de monte carlo:

$$M_N = \frac{1}{N} \sum_{i=1}^N f(X)$$

Portanto,

$$g(\alpha, a, b) \approx (b-a)M_N$$

6.3 Implemente o método de Monte Carlo para gerar amostras da sua variável aleatória, calculando a média amostral M_n e utilizando-a para estimar $g(\alpha, a, b)$. Repare que α, a, b são parâmetros do seu programa.

A implementação do método encontra-se disponível neste [github](#).

6.4 Seja \hat{g}_n o valor do estimador após n amostras. Trace um gráfico do erro relativo do estimador, ou seja $|\hat{g}_n - g(\alpha, a, b)|/g(\alpha, a, b)$ em função de n , para $n = 1, \dots, 10^6$ (utilize escala log-log no gráfico). Utilize os seguintes valores para os parâmetros: $\alpha = \{1, 2, 3\}$, $a = 0$, $b = \{1, 2, 4\}$. O que você pode concluir em relação ao erro e os parâmetros?

O resultado experimental encontra-se na Figura 7. Nota-se que, de maneira geral, o erro diminui com o aumento do número de amostras, o que é natural para uma estimação utilizando o método de monte carlo.

No entanto, note que o tempo de convergência varia com os parâmetros α e b . Como fixou-se $a = 0$, o parâmetro b indica o intervalo no qual a função deve ser integrada. Desta forma, considerando uma amostragem pela variável uniforme definida no intervalo $[a, b]$, necessita-se de mais amostras conforme b aumenta para manter um erro relativo baixo.

Por outro lado, o parâmetro α indica onde estará a massa de valores a serem integrados. Quanto maior o α , mais próximo de b devem estar os valores amostrados, o que não pode ser garantido pela amostragem da distribuição uniforme.

Assim, para um α pequeno, a variação no parâmetro b gerou pouca diferença no erro relativo, ao passo que para α maior a variação de b necessita de alterações no número de amostras.

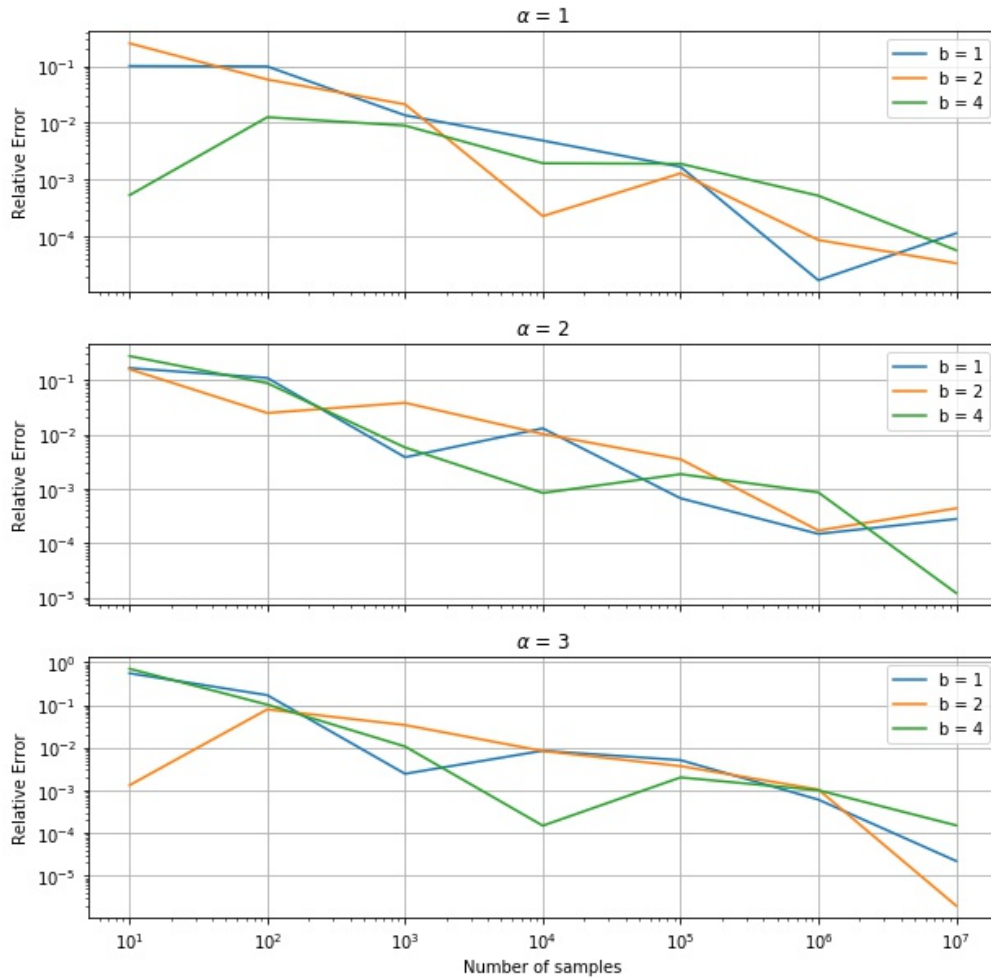


Figure 7: Evolução do erro relativo com o número de amostras.

7 Gerando subconjuntos: Considere $S_{k,n}$ um espaço amostral dado por todos os subconjunto de tamanho k dentre n objetos. Assuma que cada elemento deste espaço amostral tem a mesma probabilidade, dada por $1/S_{k,n}$.

7.1 Descreva um algoritmo eficiente para gerar amostras deste espaço. Dica: pense em permutação!

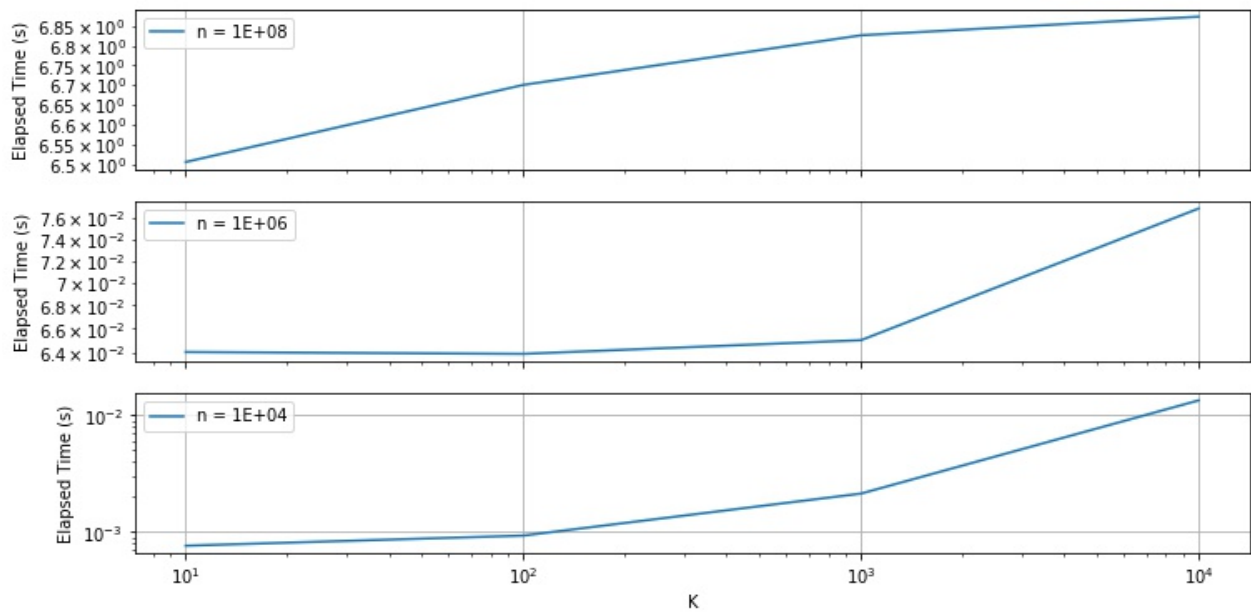
O algoritmo proposto funciona com base no *Knuth Shuffle*. Ao se ter como entrada um vetor de tamanho n , o algoritmo *Knuth Shuffle* retorna uma permutação do espaço amostral com o mesmo tamanho n .

No entanto, estamos interessados em um subconjunto do espaço amostral de tamanho k . Desta forma, ao invés de realizarmos n iterações como no algoritmo original, podemos fazer apenas k iterações e retornar um vetor com os k primeiros elementos.

7.2 Implemente seu algoritmo. Calcule o tempo médio de execução fazendo $r = 10^3$ rodadas independentes para os seguintes parâmetros $n = \{10^4, 10^6, 10^8\}$, $k = \{10^1, 10^2, 10^3, 10^4\}$. Trace um gráfico com valores de k no eixo-x e diferentes curvas com os valores de n . O que você pode concluir?

A Figura 7.2 apresenta o resultado do tempo médio de 10^3 rodadas para diferentes valores de k e n .

Note que, naturalmente, o tempo médio de execução do algoritmo cresce com o aumento de n e k . Entretanto, o impacto do parâmetro n no tempo médio de execução do algoritmo é consideravelmente maior. Isso se dá devido às permutações que são realizadas em um vetor de tamanho n , onde tanto o elemento de índice i quanto o índice $n - i$ devem ser acessados a cada iteração, o que é uma operação de custo $O(n)$. Por outro



lado, o momento de parada das iterações (isto é, quando o algoritmo chega na k -ésima iteração) depende apenas do número k , sendo uma operação $O(k)$.