# Relatório trabalho prático 3

César A. Galvão 19/0011572 Gabriela Carneiro 18/0120816

21 de August de 2022

## Contents

# 1 Introdução

O método de variáveis antitéticas é uma técnica para redução de variância usado no método de Monte Carlo. A simplicidade do método sugere que ele pode ser uma ferramenta poderosa para a redução da variância, mas poucas aplicações bem sucedidas de sua implementação foram reportadas. A ideia é obter duas estimativas correlacionadas negativamente para a média $\mu = \frac{1}{2}(\mu_2 + \mu_1)$, notando que a estimativa para a variância é reduzida por um termo de covariância.[1]

Utilizando uma amostra aleatória, um estimador para um para a quantidade de interesse $< g >= \int_0^1 g(x)p(x)dx$ é convencionalmente obtido. Introduzindo $sigma^2$, a variância de $g(x)$,

$$\sigma^2 \equiv \int_0^1 \{g(x) - < g >\}^2 \, p(x) \, dx = \int_0^1 g(x)^2 p(x) \, dx - < g >^2 \equiv < g^2 > - < g >^2 . \tag{1}$$

Um estimador para $sigma^2$, a variância amostral, pode ser obtida avaliando

$$s^2 \equiv \left( \frac{1}{N-1} \right) \sum_{i=1}^{N} (t_i \mu)^2, \tag{2}$$

e

$$s_\mu^2 = s^2/N \tag{3}$$

a variância da média.

Se $mu$ for calculada muitas vezes, a estimativa $mu_i$ iria coletivamente formar uma distribuição sobre a média das médias $\bar{\mu}$, sendo $s_\mu^2$ um estimador da variância $var(\bar{\mu})$ dessa segunda distribuição. Reconhecendo que a igualdade em (3) é obtida assumindo que as amostras $t_i$ são independentes, se algum de seus subconjuntos são correlacionados, então, na verdade

$$s_\mu^2 = \frac{1}{N} \left[ s^2 + 2 \sum_{p<q} Cov(t_p, t_q) \right] . \tag{4}$$

Se a correlação existe, $s_\mu^2$ conforme computado por (3) será uma estimativa pobre para a variável $var(\mu)$. Além disso, deve-se fazer uma distinção entre a redução da variância amostral e a redução da variância real, sem a qual, é possível observar reduções falsas ou paradoxais na variância. Dessa forma, três cenários podem ser observados:

- Redução de $\sigma^2$ e de $s_\mu^2$ (caso A), que é o objetivo;
- Redução de $\sigma^2$, mas não de $s_\mu^2$ (caso B);
- Redução de $s_\mu^2$, mas não de $\sigma^2$ (caso C).

[1] Milgram, M.S. (2001). On the Use of Antithetic Variates. In: Kling, A., Barão, F.J.C., Nakagawa, M., Távora, L., Vaz, P. (eds) Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-18211-2_29

# 2 Método

Foram utilizadas as funções de geração de amostras das variáveis aleatórias Uniforme e Normal – esta podendo ser gerada pelos métodos da rejeição e polar. Apenas a última foi utilizada para a geração da v.a. normal.

Para cada elemento de $Z = \{z_i \in (0, 0.01, 0.02, ..., 3.99)\}$, foi calculada a probabilidade acumulada da distribuição Normal pelos métodos de variável antitética e por estimador de regressão.

Para o primeiro método de estimação, utiliza-se $E(I) = p$, pois $I \sim Bernoulli(p)$. Calcula-se portanto $\sum\limits_{k=1}^{n} \frac{i_k}{n}$, caso em que a variância seria $\frac{\bar{i}(1-\bar{i})}{n}$.

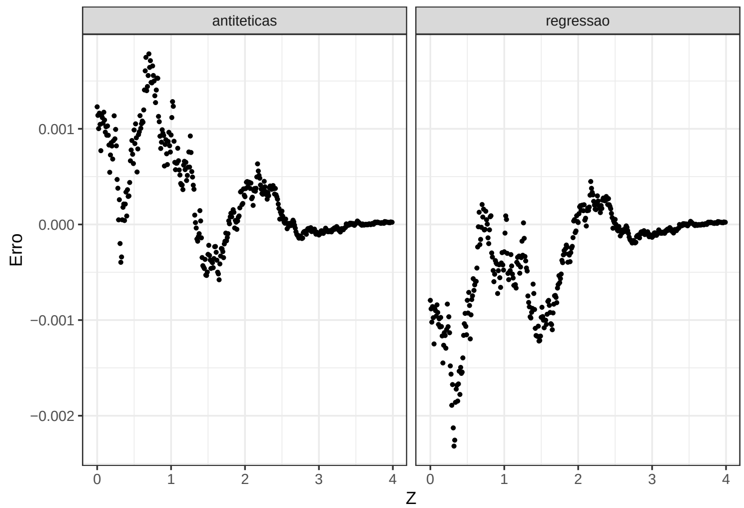
Para o segundo método, utiliza-se $I = aZ + b$, calcula-se $p$ por $\hat{a}E(Z) + \hat{b} = \hat{b}$. A variância é $\frac{\sigma^2}{n}$, em que $\sigma^2$ é a variância dos resíduos do modelo de regressão.

Em seguida, foram calculados os erros de estimação em relação à função implementada `pnorm()`. Por fim, são comparadas as variâncias entre os métodos.

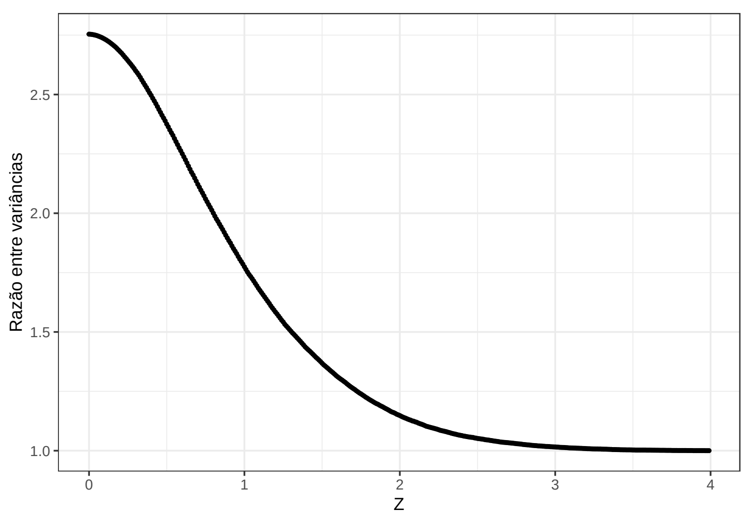
# 3 Resultados

## 3.1 Avaliação dos erros

Nota-se pelos gráficos a seguir que os erros para ambos os métodos seguem uma mesma tendência, se aproximando de zero na medida em que os valores de Z crescem.




Finalmente, compara-se as variâncias dos métodos de variáveis antitéticas e de regressão. Observa-se que quanto maior o valor de Z, mais próximo de 1 fica a razão entre as variâncias. Para pequenos valores de Z, portanto, a variância do método de variáveis antitéticas é maior.



# 4 Código

```
uniforme <- function(n){

  x <- c()

  a <- 16807
  m <- 2^31 -1

  # se tem o arquivo com seed, pega o seed k do arquivo
  # Se nao tiver arquivo, gera o arquivo e escreve o seed
  # seed começa com o relogio
  # Os que extrapolarem inserir no arquivo

  if (file.exists("../trabalho pratico 2/seeds.Rdata")){
    y <- readRDS("../trabalho pratico 2/seeds.Rdata")
  } else {
    y <- as.numeric(Sys.time())
  }

  for (i in 1:n){
    y <- (a*y)%%m
    x[i] <- y/m
  }

  #guardar o ultimo y num arquivo
  saveRDS(y, "../trabalho pratico 2/seeds.Rdata")

  return(x)
}

geranormal <- function(metodo){
  if (metodo == "rejeicao"){
    y <- geraexp(n = 1, lambda = 1)
    u <- uniforme(1)
    while(u > exp(-(y-1)^2)/2){
      u <- uniforme(1)
      y <- geraexp(n = 1, lambda = 1)
    }
    modZ <- abs(y)

    if(uniforme(1) <= 0.5){
      Z <- modZ
    } else {
      Z <- -modZ
    }
    return(Z)
  }

  if(metodo == "polar"){
    v1 <- (2*uniforme(1)-1)
    v2 <- (2*uniforme(1)-1)
```

```r
    u <- v1^2 + v2^2
    while (u > 1){
      v1 <- (2*uniforme(1)-1)
      v2 <- (2*uniforme(1)-1)
      u <- v1^2 + v2^2
    }
    x <- sqrt(-2*log(u)/u)*v1
    y <- sqrt(-2*log(u)/u)*v2
    return(list(x = x, y = y))
  }
}

intervalo <- seq(0, 3.99, 0.01)

prob_acum_z <- function(amostra, z, metodo){
  j <- as.numeric(amostra < z)
  if(metodo=="antitetica"){
    return(data.frame("z" = z,
                      "p" = mean(j), #media da va indicadora
                      "var_p"= mean(j)*(1-mean(j))/length(j))) #p(1-p)/n
  }
  if(metodo=="regressao"){
    reg <- lm(j ~ amostra)
    return(data.frame("z" = z,
                      "p" = reg$coefficients[[1]], #p = intercepto
                      "var_p"= var(reg$residuals)/length(j))) #sigma^2/n
  }
}

amostra <- c()
for(i in 1:100000){ amostra[i] <- geranormal("polar")$x}

antiteticas <- data.frame(z = 0, p = 0, var_p = 0)
for(i in 1:length(intervalo)){antiteticas[i,] <- prob_acum_z(amostra, intervalo[i], "antitetica")}

regressao <- data.frame(z = 0, p = 0, var_p = 0)
for(i in 1:length(intervalo)){regressao[i,] <- prob_acum_z(amostra, intervalo[i], "regressao")}

erros <- data.frame(
  intervalo,
  antiteticas = antiteticas$p - pnorm(intervalo),
  regressao = regressao$p - pnorm(intervalo)
) %>%
  pivot_longer(cols = -intervalo, names_to = "tipo", values_to = "erro")

graf <- ggplot(erros, aes(intervalo, erro))+
  geom_point(size = .8)+
  facet_grid(~tipo)+
  labs(x = "Z", y = "Erro")+
  theme_bw()

graf
```

```r
data.frame("z" = antiteticas$z, "razao" = antiteticas$var_p/regressao$var_p) %>%
  ggplot() +
  geom_point(aes(z, razao), size = .8) +
  labs(x = "Z", y = "Razão entre variâncias")+
  theme_bw()
```