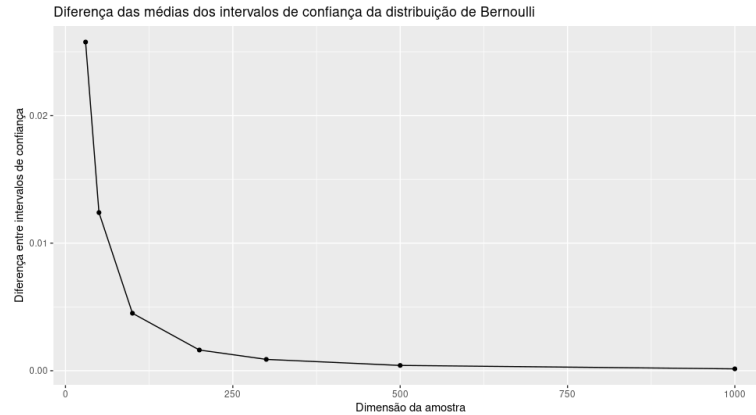


# Exercício 9 - Projeto Computacional PE 2022/2023

Consideremos como premissas que foram fixadas uma semente em 301 e um conjunto de tamanhos de amostras 30, 50, 100, 200, 300, 500 e 1000 respetivamente. O objetivo deste exercício passa por gerar 2500 amostras dos tamanhos respetivos com distribuição de Bernoulli e um parâmetro  $p = 0.3$ . De seguida, para cada amostra, calcular a diferença de tamanhos dos intervalos de confiança construídos por dois métodos: um usando equações de segundo grau onde se usa a média amostral das amostras e um valor esperado  $\frac{1}{\phi} = \frac{1+\gamma}{2}$  para cada tamanho supra-mencionado e um método usando a variável fulcral. Ambos os métodos tem que ser construídos com um nível de confiança aproximado de  $\gamma = 0.98$ . Com isto feito, calcula-se a média das 2500 diferenças de intervalo de confiança geradas. Para tal, recorreu-se ao seguinte trecho de código R (utilizando as biblioteca ggplot2):

```
1  set.seed(1532)
2
3  n_values <- c(30, 50, 100, 200, 300, 500, 1000)
4  k <- 2500
5  gamma <- 0.98
6  prob <- 0.3
7
8  calculate_ci_length_method1 <- function(x_bar, n, z) {
9    a <- ((z^2)/n) + 1
10   b <- -((z^2)/n) - 2*x_bar
11   c <- x_bar^2
12   discriminant <- b^2 - 4 * a * c
13   p1 <- (-b + sqrt(discriminant)) / (2 * a)
14   p2 <- (-b - sqrt(discriminant)) / (2 * a)
15   limit1 = Re(p1)
16   limit2 = Re(p2)
17   ci_length <- limit1 - limit2
18   return(ci_length)
19 }
20
21 calculate_ci_length_method2 <- function(x_bar, n, z) {
22   p1 <- x_bar - (z * sqrt((x_bar*(1-x_bar))/n))
23   p2 <- x_bar + (z * sqrt((x_bar*(1-x_bar))/n))
24   ci_length <- p2 - p1
25   return(ci_length)
26 }
27
28 diff_means <- sapply(n_values, function(n) {
29   z <- qnorm(0.5 + gamma/2)
30   samples <- matrix(rbinom(k * n, size = 1, prob = prob), nrow = k)
31   xBars <- colMeans(samples)
32   ci_lengths_method1 <- calculate_ci_length_method1(xBars, n, z)
33   ci_lengths_method2 <- calculate_ci_length_method2(xBars, n, z)
34   m1 <- mean(ci_lengths_method1)
35   m2 <- mean(ci_lengths_method2)
36   mean_diff <- abs(m1 - m2)
37   return(mean_diff)
38 })
39
40 data <- data.frame(n = n_values, diff_means = diff_means)
41
42 ggplot(data, aes(x = n, y = diff_means)) +
43   geom_line() +
44   geom_point() +
45   labs(x = "Dimensão da amostra", y = "Diferença entre intervalos de confiança") +
46   ggtitle("Diferença das médias dos intervalos de confiança da distribuição de Bernoulli")
```



Note-se que à medida que o tamanho da amostra aumenta, a diferença entre as medias dos intervalos de confiança torna-se rapidamente mais pequena à medida que nos aproximamos de 1000. Podemos, portanto, retirar deste gráfico que quanto maior o tamanho da população, mais podemos **confiar** na utilização de ambos os metodos de descobrir o intervalo de confiança já que a diferença entre eles reduz.