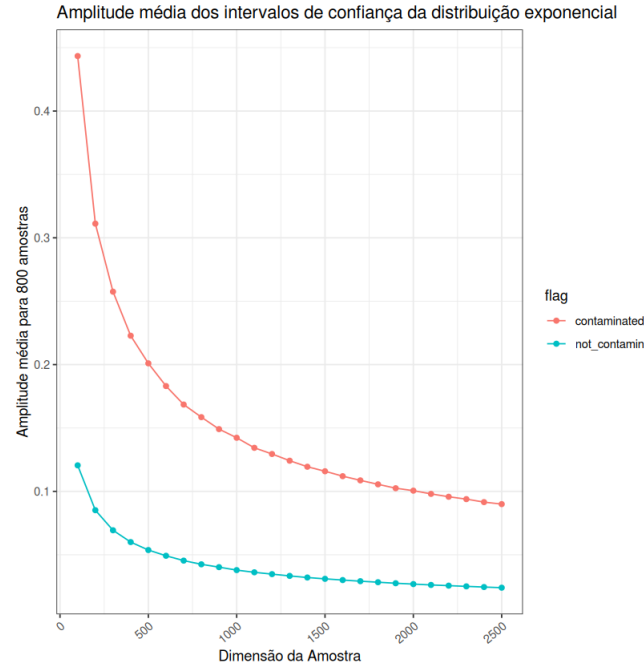


Exercício 10 - Projeto Computacional PE 2022

Diogo Gaspar, 99207

Consideremos como premissas que foram fixadas uma semente em 301 e um conjunto de tamanhos de amostras $\{100, 200, \dots, 2500\}$. O objetivo deste exercício passa por gerar 800 amostras com distribuição exponencial de valor esperado $\frac{1}{\lambda} = \frac{1}{3.13}$ para cada tamanho supra-mencionado. De seguida, substituir 25% das observações de cada amostra por outras geradas de uma população que modela a distribuição dos *outliers*, tal que $\lambda_c = 0.53$. Para cada amostra, construir um intervalo de confiança para o inverso do valor esperado, com nível de confiança $1 - \alpha = 0.94$. Por fim, para cada tamanho de amostra (contaminada e não contaminada), calcular a média da amplitude de todos os intervalos de confiança obtidos. Para tal, recorreu-se ao seguinte trecho de código R (utilizando as bibliotecas **ggplot2**, **lattice**, **plyr**, **tidyr** e **Rmisc**):

```
1  set.seed(301)
2  m <- 800
3  lambda_not_contaminated <- 3.13
4  lambda_contaminated <- 0.53
5  theoric_confidence_interval <- 0.94
6  dimensions <- seq(100, 2500, 100)
7  epsilon <- 0.25
8
9  not_contaminated <- c()
10 contaminated <- c()
11 for (n in dimensions) {
12   mean_widths <- calculate_mean_widths(n)
13   not_contaminated <- c(not_contaminated, mean_widths[1])
14   contaminated <- c(contaminated, mean_widths[2])
15 }
16
17 calculate_mean_widths <- function(n) {
18   not_contaminated <- c()
19   contaminated <- c()
20   for (i in 1:m) {
21     contaminated_amount <- floor(n * epsilon)
22     nc_exp <- rexp(n, rate=lambda_not_contaminated)
23     c_exp <- rexp(contaminated_amount, rate=lambda_contaminated)
24     c_exp <- c(c_exp[0:contaminated_amount], nc_exp[contaminated_amount:n])
25
26     nc_exp_confidence_interval <- CI(nc_exp, ci=theoric_confidence_interval)
27     c_exp_confidence_interval <- CI(c_exp, ci=theoric_confidence_interval)
28     not_contaminated <- c(not_contaminated, c(abs(nc_exp_confidence_interval[["upper"]] - nc_exp_
confidence_interval[["lower"]]))))
29     contaminated <- c(contaminated, c(abs(c_exp_confidence_interval[["upper"]] - c_exp_confidence_
interval[["lower"]]))))
30   }
31   return(c(mean(not_contaminated), mean(contaminated)))
32 }
33
34 df = data.frame(dimensions, not_contaminated, contaminated)
35 df <- pivot_longer(df, "not_contaminated":"contaminated")
36 df <- rename(df, c(name="flag", value="mean_widths"))
37
38 ggplot(df, aes(x = dimensions, y = mean_widths, colour = flag)) +
39   geom_line() +
40   geom_point() +
41   labs(x = "Dimensão da Amostra", y = "Amplitude média para 800 amostras") +
42   ggtitle("Amplitude média dos intervalos de confiança da distribuição exponencial") +
43   theme_bw() +
44   theme(axis.text.x = element_text(angle = 40, hjust=1))
```



Note-se que ambas as curvas, para amostras não contaminadas e contaminadas, seguem destinos semelhantes: começam relativamente elevadas, eventualmente acabando por começar a estabilizar próximo de 2500. Mais, note-se que amostras sem indivíduos contaminados apresentam amplitude média razoavelmente maior, levando portanto à conclusão de que amostras sem indivíduos contaminados têm maior grau de confiança.