Proposta de resolução dos Exercicios 8, 9 e 10 - Projeto Computacional PE 21/22

M. Rosário Oliveira

2022

Pergunta 8 (exemplo)

Usando o R e fixando a semente em 728, gere $m = 1200$ amostras de dimensão $n = 1412$ de uma população X, com distribuição Exponencial de valor esperado $1/\lambda = 1/1.04$, i.e. $X \sim Exp(\lambda = 1.04)$.

Para cada uma das amostras geradas, construa um intervalo de confiança aproximado para λ. Considere o nível de confiança $1 - \alpha = 0.90$.

Indique a média da amplitude dos m intervalos de confiança obtidos utilizando 6 casas decimais.

Código para confirmação (Resposta: valor numérico)

```
set.seed(728)
n<-1412
m<-1200
alpha<-0.1
ci.level <-1-alpha
lambda<-1.04

 ###Geracao das observacoes e calculo da amplitude média do IC aproximado
  ci.comprim<-rep(0,m)
    for(i in 1:m)
  {
    x<-rexp(n,lambda)
    ci.comprim[i]<-2*qnorm(1-alpha/2)/(sqrt(n)*mean(x))
  }
  ci.comprim.mean<-round(mean(ci.comprim),6)
  ci.comprim.mean
```

```
[1] 0.091064
```

Pergunta 9 (exemplo)

Usando o R e fixando a semente em 346, gere $m = 1250$ amostras de dimensão n, onde $n \in \{100, 200, 300, \ldots, 5000\}$, de uma população X, com distribuição Exponencial de valor esperado $1/\lambda = 1/0.93$, i.e. $X \sim Exp(\lambda = 0.93)$.

Para cada uma das amostras geradas, construa um intervalo de confiança aproximado para λ. Considere o nível de confiança $1 - \alpha = 0.98$.

Para cada valor de n, calcule a Média da Amplitude dos m intervalos de confiança obtidos, MA(n).

Construa um gráfico colocando no eixo dos xx a dimensão da amostra, n, e no eixo dos yy o valor de MA(n).

Submeta um ficheiro com extensão pdf, com uma única página, que inclua:

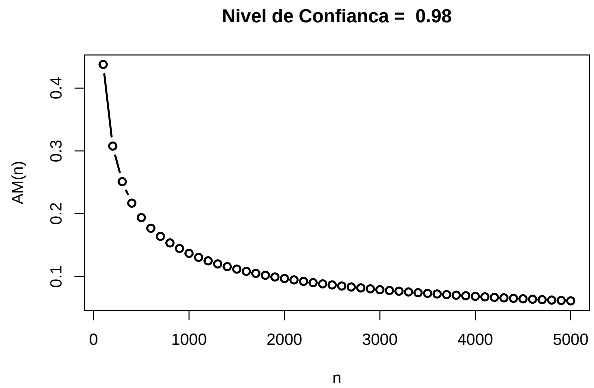
- Valores dos parâmetros: semente, m, λ e $(1-\alpha)$.
- O código em R.
- O gráfico produzido.
- Comentários sobre o gráfico obtido na alínea anterior.

Exemplo de código (há muitas alternativas para fazer este código...)

Cotação:

- Código da geração - 0.9 valor
- Código dos gráficos e qualidade dos gráficos (legendas, título, apresentação geral), verificar se calculam bem a amplitude do intervalos de confiança $2\Phi^{-1}(1-\alpha/2)/(\sqrt{n}\,\bar{x})$ - 0.9 valor
- Comentário - ex. "Quando dimensão da amostra (n) aumenta, a amplitude do intervalo de confiança diminuiu, aproximando-se de zero" - 0.2 valor

```
set.seed(346)
m<-1250
alpha<-0.02
ci.level <-1-alpha
lambda<-0.93

###Geracao das observações e cálculo da amplitude média do IC aproximado
jump<-100
n.values<-seq(100,5000,jump)
k<-length(n.values)###Sample size
ci.comprim.mean.vect<-c()

for (j in 1:k)
{
  ci.comprim<-rep(0,m)

  for(i in 1:m)
  {
    x<-rexp(n.values[j],lambda)
    ci.comprim[i]<-2*qnorm(1-alpha/2)/(sqrt(n.values[j])*mean(x))
  }
  ci.comprim.mean.vect<-c(ci.comprim.mean.vect,mean(ci.comprim))

}
plot(n.values,ci.comprim.mean.vect,lwd=2,type="b",xlab="n",ylab="AM(n)",
     main=paste("Nivel de Confianca = ",ci.level))
```



Pergunta 10 (exemplo)

Usando o R e fixando a semente em 159, gere $m = 1000$ amostras de dimensão n, onde $n \in \{100, 200, 300, \ldots, 5000\}$, de uma população X, com distribuição Exponencial de valor esperado $1/\lambda = 1/2.76$, i.e. $X \sim Exp(\lambda = 2.76)$.

Considere as amostras geradas anteriormente e substitua quaisquer $\epsilon \times 100\% = 20\%$ das observaç~{o}es de cada amostra por outras geradas de uma população que modela a distribuição dos outliers, X_c, tal que $X_c \sim Exp(\lambda_c = 0.04)$.

Para cada uma das amostras geradas sem contaminação (respetivamente, com contaminação), construa um intervalo de confiança aproximado para o inverso do valor esperado. Considere o nível de confiança $(1 - \alpha) = 0.92$.

Para cada valor de n, calcule a Média da Amplitude dos m intervalos de confiança: $\mathrm{MA}(n)$, no caso das amostras geradas sem contaminação e $\mathrm{MA}^c(n)$, no caso de haver contaminação.

Construa um gráfico colocando no eixo dos xx a dimensão da amostra, n, e no eixo dos yy os valores de $\mathrm{MA}(n)$ e $\mathrm{MA}^c(n)$.

Submeta um ficheiro com extensão pdf, com uma única página, que inclua:

- Valores dos parâmetros: m, λ, λ_c, ϵ e $(1 - \alpha)$.
- O código em R.
- O gráfico produzido.
- Comentários sobre o gráfico obtido na alínea anterior.

Exemplo de código (há muitas alternativas para fazer este código...)

Cotação:

A contaminação tem o efeito de alterar (neste caso diminuir) o valor esperado da população. No caso presente, seja $\bar{X}_0(n - n_0)$ a média de uma a.a. de X de dimensão $n - n_0 \simeq (1 - \epsilon)n$ e $\bar{X}_c(n_0)$ a média de uma a.a. de X_c de dimensão $n_0 \simeq \epsilon n$.

A média da amostra aleatória contaminada é:

$$\frac{1}{\hat{\lambda}_{CC}} = \frac{(n - n_0)\bar{X}_0(n - n_0) + n_0 \bar{X}_c(n_0)}{n} \xrightarrow{P} (1 - \epsilon)\frac{1}{\lambda} + \epsilon \frac{1}{\lambda_c}.$$

O comprimento do intervalo de confiança para λ é: $\Delta = \frac{2qnorm(1-\alpha/2)}{\bar{x}\sqrt{n}} = \frac{2qnorm(1-\alpha/2)}{\sqrt{n}} \hat{\lambda}_{MMV}$.

Se $\lambda > \lambda_c$ espera-se que $\hat{\lambda}_{MMV} > \hat{\lambda}_{CC}$ e as estimativas de Δ sem contaminação sejam maiores que as correspondentes com contaminação.

Espera-se que os alunos façam os seguintes comentários:

- Quando n aumenta a amplitude média dos intervalos diminuiu. - 0.2 valor
- A contaminação conduz a intervalos de confiança com amplitudes médias menores '{a}s dos intervalos obtidos sem contaminação, desde que $\lambda > \lambda_c$. - 0.2 valor
- Se a estimativa pontual de λ, quando há contaminação, subestima o verdadeiro valor de λ, então obtenho um intervalo de menor comprimento mas centrado numa estimativa pontual mais longe do verdadeiro valor de λ, sendo por isso um pior intervalo de confiança.- 0.2 valor

Cotação:

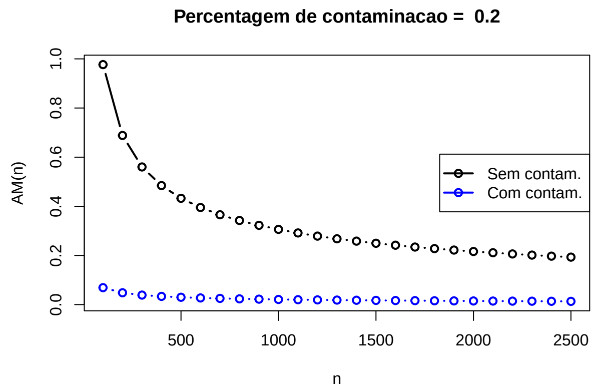
- Código da geração - 0.7 valor

- Código dos gráficos e qualidade dos gráficos (legendas, título, apresentação geral) - 0.7 valor

- Comentário - 0.6 valor - ver distribuição da pontuação acima sugerida.

```r
set.seed(159)
n<-1351
m<-1000
alpha<-0.08
ci.level <-1-alpha
lambda<-2.76
eps<-0.20
lambda.c<-0.04

###Geracao das observacoes - X \sim Exp(lambda) distribution e
### Calculo comprimento esperado para cada valor de n
jump<-100###100
n.values<-seq(100,2500,jump)
k<-length(n.values)
ci.comprim.mean.vect<-ci.comprim.mean.vect0<-c()

###
for (j in 1:k)
{
  ci.comprim<-ci.comprim0<-rep(0,m)
  x<-x0<-c()

  m.aux<-m0.aux<-c()

  for(i in 1:m)
  {
    n.bom<-round(n.values[j]*(1-eps))
    x0<-rexp(n.values[j],lambda)##Clean data
    x<-x0[1:n.bom]
    if(n.bom != n.values[j]) x<-c(x,rexp((n.values[j]-n.bom),lambda.c))##Contaminated data

    ci.comprim0[i]<-2*qnorm(1-alpha/2)/(sqrt(n.values[j])*mean(x0))##Clean data
    ci.comprim[i] <-2*qnorm(1-alpha/2)/(sqrt(n.values[j])*mean(x))##Contaminated data
  }
  ci.comprim.mean.vect0<-c(ci.comprim.mean.vect0,mean(ci.comprim0))
  ci.comprim.mean.vect<-c(ci.comprim.mean.vect,mean(ci.comprim))
}
```

```r
yylim<-c(min(ci.comprim.mean.vect0,ci.comprim.mean.vect),
        max(ci.comprim.mean.vect0,ci.comprim.mean.vect))
plot(n.values[1:k],ci.comprim.mean.vect0[1:k],lwd=2,type="b",xlab="n",ylab="AM(n)",
     main=paste("Percentagem de contaminacao = ",eps),ylim=yylim)
points(n.values[1:k],ci.comprim.mean.vect[1:k],lwd=2,type="b",col="blue")
legend(x="right",legend=c("Sem contam.","Com contam."),col=c("black","blue"),lwd=c(2,2),pch=c(1,1))
```




```r
##O EMV de lambda, na presença de contaminação é lambda.hat:
lambda.hat<-1/(1/lambda*(1-eps)+1/lambda.c*eps);lambda.hat
```

```
[1] 0.1890411
```

```r
c(lambda,lambda.c,lambda.hat)
```

```
[1] 2.7600000 0.0400000 0.1890411
```