2° Trabalho Prático - Estimação Pontual

Aluna: Stéfane Tame Monteiro Oliveira
N° USP: 10829970

Questão:

Uma determinada fábrica de computadores monta lotes com 5 computadores para venda. Deseja-se avaliar o número de computadores com defeitos em cada lote. Uma amostra de 100 lotes foi coletada e o número de computadores com defeitos em cada lote foi observado.

O conjunto de observações usado neste trabalho é do arquivo “Amostra 27 .txt”

1. Qual distribuição poderia explicar o comportamento destes dados?

O experimento tem uma amostra de 100 lotes, com 5 computadores cada. A cada computador, há chance de ter defeito ou não, isso implica em uma dicotomia, isto é, ou ter sucesso (computador defeituoso) ou ter fracasso (computador bom), que é característico da distribuição de Bernoulli. No entanto, como a cada lote há 5 computadores, conclui-se que temos a distribuição Binomial, que corresponde a 5 ensaios de Bernoulli.

Considere X_i a variável aleatória que corresponde ao i-ésimo lote, temos:

$X = (X_1, X_2, ..., X_{100})$, onde $X \sim Bin\,(n,\, p)$, onde n = 5 é a quantidade de computadores de cada lote e p é a probabilidade do computador ter defeito, no qual é o parâmetro desconhecido.

$$P(X = x) = \binom{5}{x} p^{x} . (1 - p)^{5-x},\ x = 0,...,5\ e\ 0 < p < 1$$

2. Construa a função de verossimilhança.

A função de verossimilhança se dá pelo produtório da distribuição, onde N = 100, pois é a quantidade de lotes na amostra X.

$$\mathcal{L}(p,\, x) = \prod_{i=1}^{100} \binom{5}{xi} p^{xi} . (1 - p)^{5-xi}$$

3. Avalie esta função considerando diferentes valores do(s) parâmetro(s), fornecendo uma estimativa para cada parâmetro desconhecido.

Utilizando R, temos o seguinte código, considerando a amostra 27:

```
# TRABALHO 2 - ESTIMAÇÃO PONTUAL

rm(list = ls(all = TRUE))
setwd('/home/stefane/Documentos/2021.1/Inferencia/Trabalho2')
library(readxl)
```

```
dados <- read.delim("Amostra 27 .txt", header = FALSE)
View(dados)

N = 100 #quantidade de lotes, ou seja, tamanho da amostra
n = 5 #quantidade de tentativas, ou seja, a quantidade de computadores em cada lote
x <- dados[,1] #x recebe meus dados

p = seq(0.01,1,len=100)
#vetor p -> diferentes valores para p (probabilidade do computador ter defeito)
vero=c()
log_vero=c()

for(i in 1:length(p)){
  #calcula as funções de verossimilhança e log-verossimilhança para cada p
  vero[i]=prod(dbinom(x, n, p[i]))
  log_vero[i]=log(vero[i])
}

vero
log_vero

par(mfrow=c(1,2))
x11()
plot(p,vero,type='l',col='red',lwd=2,ylab='Verossimilhança',xlab=expression(p))
x11()
plot(p,log_vero,type='l',col='blue',lwd=2,ylab='ln(Verossimilhança)',
xlab=expression(p))

EMV_p1=p[which(vero==max(vero),arr.ind=T)]
EMV_p1

EMV_p2=p[which(log_vero==max(log_vero),arr.ind=T)]
EMV_p2

EMV=mean(x)/n #estimador de p
EMV
```

E assim encontramos o resultado no vetor *vero*.

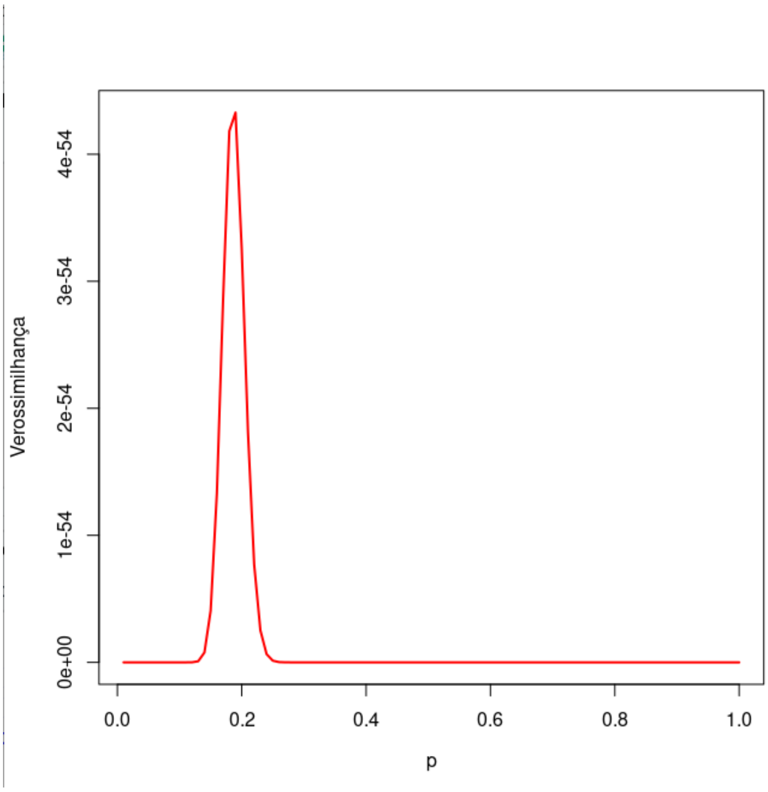

Figura 1: Gráfico da função de verossimilhança

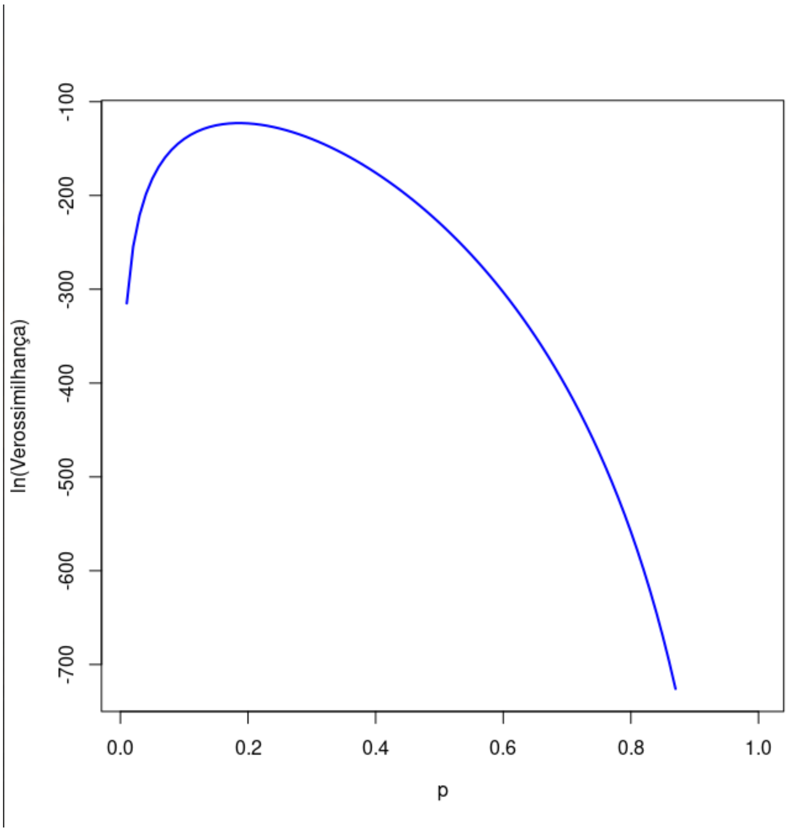

Figura 2: Gráfico da função de log-verossimilhança

4. Considerando o método de máxima verossimilhança, encontre a(s) estimativa(s) para o(s) parâmetro(s) desconhecido(s).

Temos a função de verossimilhança:

$\mathcal{L}(p,\ x)\ =\ \{(\,^{5}_{\sum xi})\}^{100}\,p^{\sum xi}.\,(1\ -\ p)^{\sum(5-xi)}$, onde o somatório é de i = 1, ..., 100.

Calculemos a função log-verossimilhança:

$\ell(p,\ x)\ = 100\,.\ \ln\{(\,^{5}_{\sum xi})\}\ +\ \sum xi\,.\ \ln\,(p)\ \ +\sum(5\ -\ xi)\,.\ \ln(1\ -\ p)$

O EMV é obtido por:

$(\ell(p,\ x))'\ =\ 0$, ou seja,

$(\ell(p,\ x))'\ = \sum xi\,.\ p^{-1} + \sum(5\ -\ xi)\,.\ (p\ -\ 1)^{-1}\ \ = 0$

$p\ =\ \sum xi\,.\ (\sum 5)^{-1} = \sum xi\,.\ (100\ *\ 5)^{-1}$

$p\ =\ \dot{x}\,.\ (5)^{-1}$

Aplicando o ẋ considerando a amostra 27, temos:

$p\ =\ 0,93/5\ =\ 0,186$

No final do código da questão c), há cálculos computacionais em que mostram o resultado para o EMV, considerando a amostra 27. As variáveis EMV_p1 e EMV_p2 são os valores de p, onde a função de verossimilhança e log-verossimilhança assumem seu máximo, respectivamente.