Estatística não paramétrica

Aula 6

Manoel Santos-Neto
Atualização: 05 de setembro de 2023

O que você irá aprender nesta aula?

1. Teste de Lilliefors.

Introdução

O teste de Kolmogorov é adequado quando temos interesse em verificar se uma a.a. é oriunda de uma distribuição com FDA $F_0(x)$ "completamente especificada", i.e., funções de distribuições com todos os seus parâmetros conhecidos. E quando você tem interesse em testar se a amostra provém de uma família de distribuições, por exemplo, $N(\mu, \sigma^2), \Gamma(\alpha, \beta)$, mas sem conhecer seus parâmetros? Devemos estimar os parâmetros usando a própria amostra, e em geral, utiliza-se a mesma estatística de teste.

Pergunta:

Usaremos os mesmos quantis do caso anterior?

- Não, pois existe uma variabilidade inerente ao estimador, i.e, existe uma variabilidade maior do que na situação anterior.
- Existem tabelas especificas para algumas classes de distribuições.

Introdução

A primeira modificação do teste de Kolmogorov (e sem dúvida a mais famosa) foi realizada com o objetivo de se testar especificamente normalidade. Isto é, a hipótese nula assume que a amostra é retirada de uma família de distribuições normais, sem especificar seu parâmetros μ e σ^2. Este teste foi apresentado por Lilliefors (1967). Uma característica interessante deste teste, com relação ao teste de Komolgorov, é a necessidade do uso de métodos compultacionalmente intesivos para estimar os quantis da distribuição exata da estatística de teste.

Suposição:

$$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} F_X(x).$$

Interesse:

Testar $\mathcal{H}_0 : X_i \sim N(\mu, \sigma^2)$ contra $\mathcal{H}_1 : X_i \text{ não possui distribuição } N(\mu, \sigma^2)$.

Introdução

Dada uma amostra aleatória, obtemos os estimadores de máxima verossimilhaça de μ e σ^2, sob $\mathcal{H}_0$, isto é,

$$\widehat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{e} \quad \widehat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

obtemos os valores amostrais normalizados, i.e,

$$Z_i = \frac{X_i - \bar{X}}{S}, \quad i = 1, \ldots, n.$$

Observação 1:

Perceba que $\widehat{F}_n(x_i) = \widehat{F}_n(z_i), \forall i = 1, \ldots, n.$

Observação 2:

O teste é realizado usando as observações padronizadas, ao invés das observações originais.

Teste de Lilliefors

Estatística de Teste

É a diferença absoluta máxima entre a função de distribuição cumulativa empírica e a hipotética. Pode ser calculada como $D = \max\{D^+, D^-\}$ com

$$D^+ = \max_{i=1,...,n} \left\{ \frac{i}{n} - p_{(i)} \right\} \quad \text{e} \quad D^- = \max_{i=1,...,n} \left\{ p_{(i)} - \frac{(i-1)}{n} \right\},$$

em que $p_{(i)} = \Phi\left(z_i\right)$ e $\Phi(\cdot)$ é a função de distribuição acumulada da distribuição normal padrão. Para um teste de nível α, rejeitamos $\mathcal{H}_0$ se $D \geq d$, em que $\Pr_{\mathcal{H}_0}(D \geq d) = \alpha$.

Observação 3:

O teste acima é equivalente a testar se a amostra é oriunda de uma população normal com média $\bar{X}$ e variância S^2, através do teste de Kolmogorov. Perceba que a diferença é que no teste de Lilliefors padronizamos as observações, ao contrário do teste de Kolmogorov.

Teste de Lilliefors

Pergunta:

Distribuição da estatística de teste sob a hipótese nula pode ser obtida através de simulação. Podemos usar este procedimento para obter o $\text{valor-}p$. Como?

Exemplo:

Considere a seguinte amostra ordenada: 13.9, 17.7, 17.9, 18.3, 18.5, 18.9, 19.4, 19.8, 20.2, 20.6, 21.1, 21.3, 21.7, 21.9, 22.0, 22.2, 22.7, 22.8, 23.2, 23.3, 23.4, 23.8, 24.4, 24.9. Teste ao nível de 5%, se a amostra é oriunda de uma distribuição normal.

```
library(nortest)
dados <- c(13.9, 17.7, 17.9, 18.3, 18.5, 18.9, 19.4, 19.8, 20.2, 20.6, 21.1, 21.3, 21.7, 21.9, 22.0, 22.2, 22.7, 22.8, 23.2, 23.3, 23.4, 23.
lillie.test(dados)
```

```
##
##     Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  dados
## D = 0.1072, p-value = 0.6762
```