Estatística de Ordem Superior

Otto Tavares Nascimento

LPS-COPPE/UFRJ

07 de Outubro, 2020

Introdução

Motivação: Introduzir conceitos e ferramentas de Estatística de Ordem Superior.

- Definição de Estatística de Ordem Superior
- O que fazer quando não estamos trabalhando com uma distribuição Gaussiana?
- Função Característica, Momentos e Cumulantes
- Cumulantes e Medidas de Não Gaussianidade

Estatística de Ordem Superior

Estatística de Ordem superior é a parte da estatística que busca revelar mais informação de funções de densidade de probabilidade do que análises baseadas em primeira (média) e segunda (variância) ordens são capazes de fazer.

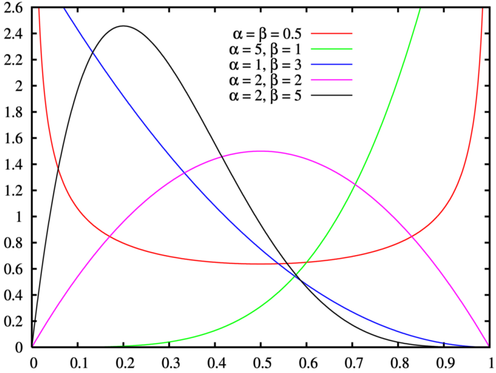

Figure 1: Distribuição Beta. Fonte : Wikimedia.

Distribuição Gaussiana é totalmente descrita por sua média e por sua variância.

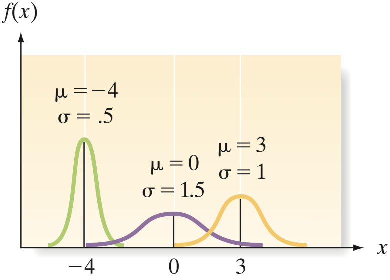

Figure 2: Distribuição Normal ou Gaussiana. Fonte : McClave, Benson, Sincich. (2015)

Análise Fatorial e Análise de Componentes Principais (PCA) são ferramentas baseadas em estatística de segunda ordem. Portanto, tem enorme poder quando os dados são descritos por uma normal.

O que fazer quando não estamos trabalhando com uma distribuição Gaussiana?

- Conceitos de estatística de ordem superior (HOS) são utilizados para acessar maior nível de informação da distribuição das variáveis aleatórias em questão.
- Idealmente, toda a informação contida na função de densidade de probabilidade é desejada para caracterizar sua distribuição. Busca-se no estudo da primeira e da segunda equações características as ferramentas para identificar e analisar distribuições na sua forma mais geral.
- Essa busca nos leva a introduzir os conceitos de momentos e de cumulantes.

Função Característica, Momentos e Cumulantes

- Assume-se $\mathrm{x} \in \mathrm{R}$ uma variável aleatória dotada de uma função densidade de probabilidade $\mathrm{p}(\mathrm{x})$ com média zero.
- A *primeira função característica* $\varphi(\omega)$ de x é definida como a transformada de Fourier contínua da função densidade de probabilidade $\mathrm{p}(\mathrm{x})$ (com $\jmath = \sqrt{-1}$ e ω a variável transformada correspondente a x):

$$\varphi(\omega) = \mathrm{E}\left\{exp(\jmath\omega\mathrm{x})\right\} = \int_{-\infty}^{\infty} exp(\jmath\omega\mathrm{x})\mathrm{p}_{\mathrm{x}}(\mathrm{x})dx \qquad (1)$$

- Expandir a função característica $\varphi(\omega)$ pela série de Taylor retorna:

$$\varphi(\omega) = \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{\mathrm{x}^k(\jmath\omega)^k}{k!}\mathrm{p}_{\mathrm{x}}(\mathrm{x})dx = \sum_{k=0}^{\infty} \mathrm{E}\left\{\mathrm{x}^k\right\} \frac{(\jmath\omega)^k}{k!} \qquad (2)$$

A primeira função característica - Função Geradora de Momentos

- Os coeficientes da expansão de Taylor feita na função característica são os momentos $\mathrm{E}\left\{\mathrm{x}^{k}\right\}$ de x, assumindo que eles existem.
- O k-ésimo momento é obtido através da derivada:

$$\mathrm{m}_k = (-\jmath)^k \frac{\partial^k \varphi(\omega)}{\partial \omega^k}\Big|_{\omega=0} \tag{3}$$

- A *primeira função característica* também é conhecida como função geradora de momentos.

A segunda função característica - Função geradora de cumulates

- Para definir a função geradora de cumulantes, introduz-se a *segunda função caracterísitca* $\phi(\omega)$ de x, que é dada pelo logaritmo natural da primeira função característica:

$$\phi(\omega) = \ln(\varphi(\omega)) = \ln(\mathrm{E}\{exp(\jmath\omega\mathrm{x})\}) \tag{4}$$

- Os cumulantes κ_k de x também são definidos através da expansão de Taylor da segunda função característica.

$$\phi(\omega) = \sum_{k=0}^{n} \kappa_k \frac{(\jmath\omega)^k}{k!} \tag{5}$$

- O k-ésimo cumulante é obtido através da derivada:

$$\kappa_k = (-\jmath)^k \frac{\partial^k \phi(\omega)}{\partial\omega^k}\Big|_{\omega=0} \tag{6}$$

Cumulantes e Medidas de não Gaussianidade

- Através de cumulantes de ordem superior é possível identificar e medir a divergência de um vetor de variáveis aleatórias não-gaussianas para um vetor gaussiano.
- O exemplo mais simples da utilização de cumulantes de ordem superior para acessar não-gaussianidade está na utilização da Assimetria e da Curtose. No entanto, seu uso deve ser feito com moderação, devido à alta sensibilidade dessas estatísticas a *outliers*.
- Informação Mútua e Negentropia são outras formas de medir não gaussianidade. É possível aproximar o valor da Negentropia e da Informação Mútua de uma determinada distribuição através do cálculo de cumulantes - Common (1993).
- Máxima não gaussianidade e descorrelação não linear são os princípios de Análise de Componentes Independentes. HOS é utilizada para identificar as distribuições.

Ilustrando a utilidade do cálculo de Cumulantes na identificação de distribuições.

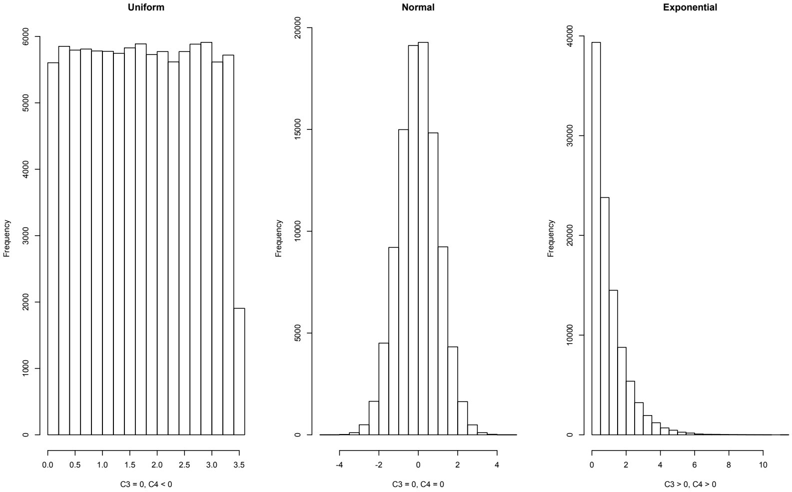

Figure 3: Dist. Uniforme (Esq.), Dist. Normal (Meio) e Dist. Exponencial (Dir.)

- OBRIGADO

Referências

- Higher-Order Statistics in Signal Processing - Nandi (1998)
- The advanced theory of statistics, vol. 1 - Kendall (1945)
- Introductory Probability and Statistical Applications - Meyer (1983)
- The elements of Statistical Learning - Hastie et. al (2009)
- Estatística Básica - Bussab, Morettin (2017)
- Statistics for Business and Economics - McClave, Benson,Sincich. (2015)
- Independent Component Analysis - Oja et. al (2001)
- Independent component analysis, A new concept? - Common (1992)