1. Introdução à Inferência Estatística – Estimação Pontual

Objectivos

Com o estudo deste capítulo o leitor deverá ser capaz de:

- Ter presente a noção de variável aleatória e suas propriedades;
- Descrever as propriedades da média e da variância;
- Conhecer as principais distribuições discretas e contínuas e saber relacioná-las;
- Compreender o que é a Inferência Estatística e quais os seus objectivos;
- Ter a noção de estimação pontual e reconhecer as qualidades de um estimador;
- Determinar o melhor estimador para um parâmetro com base no método da máxima verosimilhança.

Resumo

Pretende-se no início do capítulo, documentar o leitor com alguns conceitos e propriedades necessárias à abordagem dos temas principais a estudar - A Inferência Estatística e o Planeamento de Experiências.
É introduzida a noção de Inferência Estatística bem como os seus principais objectivos, e é estudada a Estimação Pontual. São apresentadas qualidades desejáveis num estimador, bem como o método da máxima verosimilhança, para a obtenção do melhor estimador pontual de um parâmetro. Finalmente são apresentados alguns exercícios resolvidos e propostos de modo a proporcionar ao leitor o contacto com possíveis aplicações práticas.

1. Revisão de conceitos básicos em Estatística

1.1 *Variáveis Aleatórias; Propriedades da média e da variância*

Uma **variável aleatória** não é mais do que uma função definida num determinado Universo. Na notação usual as variáveis aleatórias representam-se por letras maiúsculas e os valores observados das variáveis aleatórias por letras minúsculas. Uma variável aleatória pode ser do **tipo discreto**, se o conjunto de todos os possíveis valores que a variável possa assumir for finito ou uma infinidade numerável, ou do **tipo contínuo**, se a variável estiver definida num intervalo.

A estrutura probabilística de uma variável aleatória é descrita pela sua **distribuição de probabilidade**. Dada uma **variável aleatória discreta X**, que assume um conjunto discreto de valores, $X_1, X_2, ..., X_n$ com probabilidades $p_1, p_2, ..., p_n$, em que $p_1 + p_2 + ... + p_n = 1$, representa-se por p(X=x) ou simplesmente **p(x) a função de probabilidade** respectiva. Dada uma **variável aleatória contínua X**, assumindo valores num intervalo, então representa-se por **f(x) a função densidade de probabilidade** correspondente. As propriedades destas funções são resumidas em seguida.

Assim tem-se:

(i) Caso discreto

$$0 \leq p(x_i) \leq 1, \quad \forall_i, i = 1, ..., n$$

$$\sum_{i=1}^{n} p(x_i) = 1$$

(ii) Caso contínuo

$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Recorde-se ainda que:

$$P(a \leq x \leq b) = \int_{a}^{b} f(x)dx$$

Uma das **medidas de tendência central** mais importante é a **Média** ou **Esperança Matemática**. Representa-se por μ a média de uma distribuição de probabilidade e define-se como:

$$E(X) = \mu = \begin{cases} \sum_{i=1}^{n} x_i p_i, & \textit{Caso discreto} \\ \int_{-\infty}^{+\infty} x f(x)dx, & \textit{Caso contínuo} \end{cases}$$

Uma das **medidas de dispersão** mais importantes é sem dúvida a **Variância**, usualmente representada por σ^2 e definida como:

$$\sigma^2 = E\left[(X-\mu)^2\right]$$

ou simplesmente:

$$\sigma^2 = V(X) = E(X^2) - \left[E(X)\right]^2$$

Apresentam-se em seguida **as propriedades mais relevantes da média e da variância**. Seja X uma variável aleatória de média μ e variância σ^2, e k uma constante.

Relativamente à **esperança matemática** tem-se:

$$E(k) = k$$
$$E(X) = \mu$$
$$E(kX) = kE(X) = k\mu$$

Relativamente à **variância** tem-se:

$$V(k) = 0$$
$$V(X) = \sigma^2$$
$$V(kX) = k^2V(X) = k^2\sigma^2$$

Sejam agora duas variáveis aleatórias, por exemplo X_1 e X_2, com médias μ_1 e μ_2, e com variâncias σ_1^2 e σ_2^2 respectivamente.

Relativamente à **soma algébrica de variáveis aleatórias** pode demonstrar-se que:

$$E(X_1 \pm X_2) = E(X_1) \pm E(X_2) = \mu_1 \pm \mu_2$$

Este resultado é extensivo a n variáveis:

"A Esperança Matemática da soma algébrica de n variáveis aleatórias é a soma algébrica das Esperanças Matemáticas dessas variáveis".

A **covariância** é uma medida de associação linear entre duas variáveis, definida como:

$$\boxed{\mathrm{Cov}(X_1, X_2) = \mathrm{E}[(X_1 - \mu_1)(X_2 - \mu_2)]}$$

Relativamente à **soma e diferença de variâncias**, pode demonstrar-se que:

$$\boxed{\begin{aligned} \mathrm{V}(X_1 + X_2) &= \mathrm{V}(X_1) + \mathrm{V}(X_2) + 2\mathrm{Cov}(X_1, X_2) \\ \mathrm{V}(X_1 - X_2) &= \mathrm{V}(X_1) + \mathrm{V}(X_2) - 2\mathrm{Cov}(X_1, X_2) \end{aligned}}$$

Convém recordar que a *covariância entre duas variáveis aleatórias independentes é nula.*

Admitindo que $\mathbf{X_1, X_2}$ **são independentes**, tem-se:

$$\boxed{\mathrm{V}(X_1 \pm X_2) = \mathrm{V}(X_1) + \mathrm{V}(X_2) = \sigma_1^2 + \sigma_2^2}$$

e

$$\boxed{\mathrm{E}(X_1 X_2) = \mathrm{E}(X_1)\mathrm{E}(X_2) = \mu_1 \mu_2}$$

Com base numa amostra de dimensão n, designa-se por **soma de quadrados** e representa-se por SQ, a seguinte expressão:

$$\boxed{\mathrm{SQ} = \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

O **número de graus de liberdade** de uma soma de quadrados é igual ao número de elementos independentes nessa soma de quadrados. Por

exemplo SQ consiste numa soma de quadrados dos n elementos seguintes: $x_1 - \overline{x}, x_2 - \overline{x}, ..., x_n - \overline{x}$. Uma vez que $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$, estes elementos não são todos independentes. Apenas se pode afirmar que n-1 deles são independentes entre si, o que implica que SQ terá n-1 graus de liberdade.

Serão agora revistas as principais distribuições estatísticas usadas quer na inferência estatística quer no planeamento de experiências.

1.2 *Revisão das principais distribuições discretas*

Distribuição binomial

Uma das principais distribuições discretas conhecidas é a distribuição binomial. Esta distribuição é apropriada sempre que ao realizar repetições independentes de uma dada experiência, com reposição, o resultado de interesse se possa manifestar através duma dicotomia (cara-coroa, votar-não votar, etc).

Seja X uma variável aleatória e considere-se um acontecimento com interesse cuja probabilidade de ocorrência é p.

A **função de probabilidade de uma distribuição binomial** é dada por:

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0,...,n$$

em que n representa o número de repetições da experiência, p a probabilidade de ocorrência do acontecimento com interesse, q=1-p e k=0,...,n representa o número de ocorrências do acontecimento com interesse que se pretende calcular.

Verifica-se que, no caso da distribuição binomial:

$$E(X)= np \quad ; \quad V(X) = npq$$

Distribuição de Poisson

A distribuição de Poisson ou dos acontecimentos raros é usada em situações binomiais, em que n é grande e p pequeno ($n \geq 50$ e $np \leq 5$) daí a designação "acontecimentos raros".

Seja X uma variável aleatória discreta assumindo os valores 0,...,n.

X tem **uma distribuição de Poisson de parâmetro λ positivo**, representando-se $X \approx Pois(\lambda)$, se a função probabilidade for dada por:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0,...,n \; ; \quad \lambda > 0; \quad \lambda = np$$

Verifica-se que, no caso da distribuição de Poisson:

$$E(X)= np = \lambda \quad ; \quad V(X) = np = \lambda$$

1.3 *Revisão das principais distribuições contínuas*

Distribuição normal

Uma das principais distribuições contínuas é a distribuição normal, quer pelas qualidades que apresenta, quer pelas inúmeras aplicações nomeadamente no âmbito da Inferência Estatística e do Planeamento de Experiências.

Seja X uma variável aleatória contínua de média populacional μ e variância σ^2.

A **função densidade de probabilidade de uma distribuição normal** é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

A notação usual para representar uma variável aleatória X com média μ e variância σ^2 é $X \approx N(\mu, \sigma^2)$. O gráfico da função densidade de uma distribuição normal tem a forma de sino e é simétrico em relação à média.

O cálculo das probabilidades com base nesta função densidade torna-se penoso, uma vez que, para além de exigir o desenvolvimento em séries, exige também a elaboração de uma tabela para cada combinação diferente de μ e σ^2. Um caso particular extremamente importante é a distribuição normal de média zero e variância 1, conhecida como **distribuição normal reduzida** ou **distribuição normal standard**. Verifica-se que qualquer distribuição normal pode ser convertida numa normal reduzida, da seguinte forma:

$$z = \frac{x - \mu}{\sigma}$$

A variável aleatória Z diz-se normal reduzida ou standard, e representa-se por $Z \approx N(0,1)$. Esta distribuição encontra-se tabelada.

Em muitas técnicas estatísticas é frequentemente assumido que as variáveis são distribuídas de acordo com uma distribuição normal, por recurso ao **Teorema do Limite Central**, cujo enunciado se apresenta em seguida:

Teorema do Limite Central

Seja $X_1, X_2, ..., X_n$ uma sequência de n variáveis aleatórias independentes e identicamente distribuídas com média μ e variância σ^2, ambas finitas, e seja $\sum_{i=1}^{n} X_i$ a sua soma. Então ,

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}}$$

tem distribuição aproximadamente normal reduzida. Sendo $F_n(Z)$ a função de distribuição de Z_n e $\Phi(Z)$ a função de distribuição de uma normal reduzida, então $\lim_{n\to\infty}\left[\frac{F_n(Z)}{\Phi(Z)}\right] = 1$.

Em suma, com este teorema resulta o facto de se provar que a soma de n variáveis aleatórias independentes e identicamente distribuídas é aproximadamente normal.

A partir da distribuição normal pode deduzir-se uma outra, também bastante importante, conhecida por distribuição do Qui-Quadrado.

Distribuição do Qui-Quadrado

Considerem-se k variáveis aleatórias normais reduzidas, independentes e identicamente distribuídas, $Z_1, Z_2, ..., Z_k$. Abreviadamente escreva-se Z_i iid $\approx N(0,1)$, $i = 1,...,k$.

Então $X = Z_1^2 + Z_2^2 + ... + Z_k^2$, diz-se ser distribuída de acordo com uma **distribuição do Qui-Quadrado com k graus de liberdade**, ou abreviadamente:

$$X \approx \chi_k^2$$

Esta distribuição é assimétrica e prova-se ter média $\mu=k$ e variância $\sigma^2 = 2k$. Esta distribuição também se encontra tabelada.

Como exemplo de variável aleatória que segue uma distribuição do Qui-Quadrado tem-se:

$$\frac{SQ}{\sigma^2} \approx \chi_{n-1}^2$$

Uma vez que a variável χ^2 é obtida através da soma de variáveis independentes, então de acordo com o Teorema de Limite Central toda a distribuição do tipo χ^2 tende a aproximar-se da distribuição normal à medida que o número de graus de liberdade aumenta. A distribuição do Qui-Quadrado goza também da importante propriedade da **aditividade**. Assim, duas variáveis independentes com distribuições χ_n^2 e χ_p^2 terão distribuição χ_{n+p}^2.

$$\chi_n^2 + \chi_p^2 \approx \chi_{n+p}^2$$

Distribuição t-de-Student

Sejam Z e χ_k^2 variáveis aleatórias normais independentes e qui-quadrado, respectivamente. A variável aleatória T_k segue uma **distribuição t-de- Student com k graus de liberdade**, se:

$$T_k = \frac{Z}{\sqrt{\frac{\chi_k^2}{k}}}$$

Para a distribuição t-de-Student tem-se:

$$\mu = 0; \quad \sigma^2 = \frac{k}{k-2}, \; k>2.$$

Um exemplo de uma variável que segue a distribuição t-de-Student com n-1 graus de liberdade, será:

$$T = \frac{\overline{X} - \mu}{\hat{S}/\sqrt{n}} \approx t_{n-1}$$

Distribuição F

Sejam duas variáveis aleatórias independentes com **distribuição do Qui-Quadrado com** ν_1 **e** ν_2 **graus de liberdade**, respectivamente. Defina-se F, como se segue:

$$F_{\nu_1,\nu_2} = \frac{\chi_{\nu_1}^2 / \nu_1}{\chi_{\nu_2}^2 / \nu_2} = \frac{\nu_2 \chi_{\nu_1}^2}{\nu_1 \chi_{\nu_2}^2}$$

Para a distribuição F tem-se:

$$\mu = \frac{\nu_2}{\nu_2 - 2}, \nu_2 > 2 \; ; \quad \sigma^2 = \frac{\nu_2^2(2\nu_2 + 2\nu_1 - 4)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \nu_2 > 4$$

Seja $x_{11}, x_{12}, ..., x_{1n_1}$ uma amostra aleatória com n_1 observações e $x_{21}, x_{22}, ..., x_{2n_2}$ uma amostra aleatória com n_2 observações. Um exemplo de uma variável que segue a **distribuição F** será:

$$\frac{\hat{S}_1^2}{\hat{S}_2^2} \approx F_{n_1-1, n_2-1}$$

onde $\hat{S}_1^2$ e $\hat{S}_2^2$ representam as variâncias amostrais.

2. Introdução e objectivos da Inferência Estatística

No âmbito da **inferência estatística** o objectivo não é apenas a caracterização e estudo da amostra, mas sim tentar a partir dela caracterizar toda a população. *Pretende-se a partir da amostra inferir resultados extensivos a toda a população donde esta foi extraída.* De um modo geral a inferência a partir da amostra para a população pode ser feita de duas formas: por **estimação** ou por **testes de hipóteses**, como é sintetizado de seguida.

Estimação - Num problema de estimação o conjunto de todas as possíveis acções que podem ser tomadas é conhecido por **espaço de decisão**. Os resultados obtidos a partir da amostra são usados para inferir acerca da população donde a amostra foi extraída; a estimação de parâmetros engloba dois processos distintos:

(i) **Estimação pontual** - a partir da amostra procura-se obter um único valor de certo parâmetro populacional;

(ii) **Estimação por intervalos** - a partir da amostra procura-se construir um intervalo de valores, digamos $\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2$, com uma certa probabilidade de conter o verdadeiro valor do parâmetro populacional θ.

Testes de hipóteses – Os resultados obtidos a partir da amostra são usados para testar valores de certos parâmetros da população ou a natureza da própria população. Assim consideram-se dois tipos de testes de hipóteses:

(i) **Paramétricos** - são formuladas hipóteses relativamente ao valor de um parâmetro populacional,

(ii) **Não paramétricos** - são formuladas hipótese relativamente à natureza da distribuição da população.

3. Estimação pontual

3.1 *Estimadores pontuais e suas propriedades*

Na notação usual as variáveis aleatórias representam-se por letras maiúsculas e os valores observados das variáveis aleatórias, estimativas, por letras minúsculas, tal como já foi referido anteriormente. Assim, define-se o **estimador pontual de um parâmetro** θ como uma estatística $\hat{\Theta}$ cujos valores concretos, representados por $\hat{\theta}$, são **estimativas** deste parâmetro. Por exemplo a média amostral $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ é um estimador pontual para a média populacional, μ, e $\hat{s}^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$ é um estimador pontual para a variância populacional σ^2.

Não se pode esperar que um estimador estime um parâmetro populacional

sem erros, apesar de se esperar que assuma um valor muito próximo do seu verdadeiro valor. Note-se que para o mesmo parâmetro podem ainda ser considerados vários estimadores. A preocupação na escolha do melhor estimador entre vários conduziu ao estabelecimento de alguns critérios acerca das qualidades ou **propriedades desejáveis dos estimadores**.

Propriedade 1: Não enviesamento

Considere-se θ um parâmetro associado a uma dada população e $\hat{\theta}$ um estimador desse parâmetro. O **enviesamento** do estimador $\hat{\theta}$ é a diferença entre o valor esperado do estimador, $E(\hat{\theta})$, e o valor do próprio parâmetro, θ.

$$\text{Enviesamento}(\hat{\theta}) = E(\hat{\theta}) - \theta = \mu_{\hat{\theta}} - \theta$$

Um estimador diz-se **não enviesado** ou **centrado** quando o seu enviesamento for nulo, ou seja $E(\hat{\theta}) - \theta = 0$.

Para um **estimador não enviesado** ou **centrado**, tem-se portanto:

$$E(\hat{\theta}) = \theta$$

Um estimador diz-se enviesado se $E(\hat{\theta}) \neq \theta$.

Propriedade 2: Eficiência

Quando se pretende comparar dois estimadores não enviesados, digamos $\hat{\theta}_1$ e $\hat{\theta}_2$, com base em amostras de igual dimensão n, e com variâncias σ_1^2 e σ_2^2 respectivamente, **o estimador mais eficiente é o de menor variância**. Assim, se $\sigma_1^2 < \sigma_2^2$ então $\hat{\theta}_1$ será um estimador mais eficiente que $\hat{\theta}_2$.

Definição: Se $\hat{\theta}$ for um estimador não enviesado, diz-se que $\hat{\theta}$ é um **estimador não enviesado de variância mínima de** θ, se para todos os

estimadores θ^*, tais que $E(\theta^*)=\theta$, se tiver $Var[\hat{\theta}] < Var[\theta^*]$ para todo θ.

Definição: A **eficiência** de um estimador pode ser medida através do **erro quadrático médio:**

$$\text{Eficiência}_{\hat{\theta}} = E[(\hat{\theta}-\theta)^2] = \sigma^2_{\hat{\theta}} + (\text{enviesamento}_{\hat{\theta}})^2$$

Para se comparar a eficiência de dois estimadores é usado o conceito de **eficiência relativa**. Sejam $\hat{\theta}_1$ e $\hat{\theta}_2$ estimadores de um mesmo parâmetro θ. A eficiência de $\hat{\theta}_1$ relativamente a $\hat{\theta}_2$ é dada por :

$$\text{Eficiência relativa}_{\hat{\theta}_1/\theta_2} = \frac{E[(\hat{\theta}_1-\theta)^2]}{E[(\hat{\theta}_2-\theta)^2]}$$

Definição: Diz-se que $\hat{\theta}$ é o **melhor estimador linear não enviesado** de θ (BLUE-"Best Linear Umbased Estimator"), se obedecer às seguintes condições:

(i) $E(\hat{\theta})=\theta$

(ii) $\hat{\theta}=\sum_{i=1}^{n} a_i X_i$; $\hat{\theta}$ é uma função linear da amostra;

(iii) $V(\hat{\theta}) \leq V(\theta^*)$, onde θ^* é qualquer outro estimador de θ que satisfaça (i) e (ii).

Propriedade 3: Consistência

Seja $\hat{\theta}$ um estimador do parâmetro θ e n a dimensão da amostra. Diz-se que $\hat{\theta}$ é um **estimador consistente** quando para $\forall_{\varepsilon>0}$ se verifica a condição:

$$\lim_{n\to\infty} \text{Prob}\left[\left|\hat{\theta}-\theta\right|<\varepsilon\right]=1$$

Trata-se portanto da convergência em probabilidade do estimador para o parâmetro.

Como se viu, é relativamente simples verificar se um estimador é ou não enviesado; é também bastante fácil comparar as variâncias de dois estimadores não enviesados. No entanto não é muito simples a aplicação do critério anterior para a verificação da consistência de um estimador. Na prática recorre-se a:

Teorema:

Seja $\hat{\theta}$ um estimador do parâmetro θ e n a dimensão da amostra.

Se forem verificadas as condições,

(i) $E(\hat{\theta})=\theta$, não enviesamento,

e

(ii) $\lim_{n\to\infty} \text{Var}(\hat{\theta})=0$

então $\hat{\theta}$ é um estimador consistente de θ.

3.2 *O Método da Máxima Verosimilhança*

Estudámos alguns critérios que permitem a escolha de estimadores com as melhores qualidades. Dado um determinado estimador para um parâmetro desconhecido é já possível averiguar se este é ou não enviesado, eficiente, sendo ainda possível a escolha do mais eficiente, ou consistente. No entanto não se dispõe ainda de um procedimento geral que conduza à

escolha dos melhores estimadores. Existem vários destes procedimentos dos quais se destacam o **método dos mínimos quadrados**, o **método dos momentos** e o **método da máxima verosimilhança** ao qual é dedicada a atenção neste ponto. A exposição deste método terá em conta o caso discreto, caso1, e o caso contínuo, caso 2.

Caso 1: Distribuições discretas

Considere-se uma população discreta representada pela variável aleatória X com função de probabilidade p(x), conhecida a menos de k parâmetros, $\theta_1, \theta_2, ..., \theta_k$. Seja $X_1, X_2, ..., X_n$ uma amostra aleatória da variável aleatória X e sejam $x_1, x_2, ..., x_n$ os valores observados dessa variável aleatória.

A probabilidade de ocorrência de uma amostra aleatória simples similar à que se obteve efectivamente a partir da população é conhecida por **função de verosimilhança**, que se define obviamente em função da amostra e de θ:

$$L(X_1, X_2, ..., X_n; \theta) = \text{Prob}(X_1 = x_1, ..., X_n = x_n | \theta_1, \theta_2, ..., \theta_k) \Leftrightarrow$$

$$L(X_1, X_2, ..., X_n; \theta) = \prod_{i=1}^{n} \text{Prob}(X_i = x_i | \theta_1, \theta_2, ..., \theta_k)$$

Os **estimadores de máxima verosimilhança dos parâmetros** $\theta_1, \theta_2, ..., \theta_k$ são os valores destes parâmetros que maximizam a função de verosimilhança. Em suma, os estimadores de máxima verosimilhança são os valores dos parâmetros que tornam máxima a probabilidade de ocorrência de uma amostra idêntica àquela que efectivamente ocorreu.

Assim, admitindo que a função de verosimilhança é diferenciável, os estimadores de máxima verosimilhança podem ser obtidos resolvendo o sistema de equações definido por:

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1,...,k$$

Normalmente as soluções obtidas desta forma são máximos dessas funções, uma vez que correspondem a pontos estacionários das funções verosimilhança. No entanto, é conveniente verificar se a segunda derivada em ordem ao respectivo parâmetro é negativa, para assim assegurar a existência de máximos.

As **estimativas de máxima verosimilhança** que são obtidas a partir de amostras concretas correspondem a valores particulares dos correspondentes estimadores de máxima verosimilhança.

Por vezes na prática torna-se bastante mais simples recorrer à **maximização da transformada logarítmica da função de verosimilhança,**

$$\ln[\ L(X_i \mid\ \theta_1, \theta_2, ..., \theta_k)$$

do que à maximização da própria função. Esta simplificação de cálculos resulta do facto de a aplicação de logaritmos converter o produtório em somatório. Uma vez que L e ln (L) têm pontos estacionários comuns, o procedimento de obtenção de máximos conduz aos mesmos resultados.

Caso 2 : Distribuições contínuas

No caso das distribuições contínuas não faz sentido a maximização da probabilidade de ocorrência de uma amostra idêntica àquela que de facto ocorreu, uma vez que tal probabilidade é sempre nula. O objectivo será

maximizar a densidade de probabilidade definida para os valores observados da variável aleatória, de acordo com os valores dos parâmetros.

Considere-se uma população contínua representada pela variável aleatória X com função densidade de probabilidade $f(x|\theta_1,\theta_2,...,\theta_k)$, que depende portanto de k parâmetros desconhecidos, $\theta_1,\theta_2,...,\theta_k$.

Sejam $x_1,x_2,...,x_n$ os valores de uma amostra. **A função de verosimilhança da amostra de dimensão n**, é uma função de θ definida por:

$$L(\theta_1,\theta_2,...,\theta_k|\ x_1,x_2,...,x_n) = \prod_{i=1}^{n} f(x_i|\theta_1,\theta_2,...,\theta_k)$$

Os **estimadores de máxima verosimilhança** de $\theta_1,\theta_2,...,\theta_k$ são obtidos simultaneamente como os valores destes parâmetros que maximizam a função verosimilhança, ou seja resolvendo:

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1,...,k$$

Analogamente ao caso discreto, convém averiguar se a segunda derivada é negativa para assegurar a existência de máximos. De um modo geral é também conveniente a consideração da **maximização do logaritmo da função L** em vez da própria função, por razões já expostas.

Assim, a maximização de L corresponde a:

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \quad i = 1,...,k$$

Exercícios Resolvidos

1. Seja X uma variável aleatória com distribuição binomial de parâmetros n e p , e sejam $\hat{p}_1$ e $\hat{p}_2$ dois estimadores do parâmetro p, tais que:

$$\hat{p}_1 = \frac{X}{n} \quad \text{e} \quad \hat{p}_2 = \frac{X+2}{n+3}$$

1.1 Algum destes estimadores é enviesado? Justifique pormenorizadamente a sua resposta.

1.2 Calcule $V(\hat{p}_1)$ e $V(\hat{p}_2)$.

2. Seja X uma variável aleatória que representa o número de avarias de um computador central de uma certa unidade de trabalho. Sabe-se que X obedece a uma distribuição de Poisson de parâmetro λ desconhecido. Para este parâmetro foram indicados dois estimadores $\hat{\lambda}_1$ e $\hat{\lambda}_2$, tal que:

$$\hat{\lambda}_1 = \frac{X_1 + X_2 + \ldots + X_n}{n} \quad ; \quad \hat{\lambda}_2 = \frac{X_2 + X_{n-1}}{2}$$

Compare os estimadores propostos quanto ao enviesamento e à eficiência.

3. No departamento de Controlo de Qualidade de uma fábrica de impressoras foram analisados cinco lotes com 20 impressoras cada, registando-se as respectivas percentagens de impressoras com defeitos, 3.1%, 3.5%, 2.7%, 2.3% e 4.5%.

3.1 Considere a distribuição do número de impressoras defeituosas por lote e apresente o estimador de máxima verosimilhança da probabilidade de uma impressora ser defeituosa.

3.2 Com base na amostra dos cinco lotes calcule a estimativa de máxima verosimilhança.

Resolução:

1.1 $X \approx Bin(n,p)$ então $E(X)= np$ e $V(X)= npq= np(1-p)$

$\hat{p}_1$ e $\hat{p}_2$ são dois estimadores do parâmetro p, tais que:

$$\hat{p}_1 = \frac{X}{n} \quad \text{e} \quad \hat{p}_2 = \frac{X+2}{n+3}$$

Para averiguar se algum dos estimadores é enviesado é necessário a determinar:

$$E(\hat{p}_1)=E\left(\frac{x}{n}\right)=\frac{1}{n}E(X)=\frac{1}{n}np=p$$

$$E(\hat{p}_2)=E\left(\frac{x+2}{n+3}\right)=\frac{1}{n+3}E(X+2)=\frac{E(X)+E(2)}{n+3}=\frac{np+2}{n+3}$$

Conclusão: Uma vez que $E(\hat{p}_1) = p$, então $\hat{p}_1$ é estimador não enviesado de p; quanto a $\hat{p}_2$ verifica-se ser $E(\hat{p}_2)=\frac{np+2}{n+3} \neq p$, donde $\hat{p}_2$ é estimador enviesado de p.

1.2 Calcular $V(\hat{p}_1)$ e $V(\hat{p}_2)$;

$$V(\hat{p}_1) = V\left(\frac{X}{n}\right)=\frac{1}{n^2}V(X)=\frac{np(1-p)}{n^2}=\frac{p(1-p)}{n}$$

$$V(\hat{p}_2) = V\left(\frac{X+2}{n+3}\right)=\frac{1}{(n+3)^2}V(X)=\frac{np(1-p)}{(n+3)^2}$$

2. Quanto ao enviesamento há que calcular $E(\hat{\lambda}_1)$ e $E(\hat{\lambda}_2)$, com

$$\hat{\lambda}_1 = \frac{X_1+X_2+...+X_n}{n} \quad \text{e} \quad \hat{\lambda}_2 = \frac{X_2+X_{n-1}}{2}$$

Tem-se portanto:

$$E(\hat{\lambda}_1) = E\left(\frac{X_1+X_2+...+X_n}{n}\right)=\frac{1}{n}E\left[\sum_{i=1}^{n}X_i\right]=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}n\lambda=\lambda$$

$$E(\hat{\lambda}_2) = E\left(\frac{X_2 + X_{n-1}}{2}\right) = \frac{1}{2}[E(X_2) + E(X_{n-1})] = \frac{1}{2}(\lambda + \lambda) = \frac{2\lambda}{2} = \lambda$$

Conclusão: Uma vez que $E(\hat{\lambda}_1)=\lambda$ e $E(\hat{\lambda}_2)=\lambda$ então $\hat{\lambda}_1$ e $\hat{\lambda}_2$ são estimadores não enviesados de λ.

Quanto à eficiência é necessário calcular ambas as variâncias; como se sabe o estimador de menor variância é o mais eficiente. Assim, tem-se:

$$V(\hat{\lambda}_1) = V\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n^2} V\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2} n\lambda = \frac{\lambda}{n}$$

$$V(\hat{\lambda}_2) = V\left(\frac{X_2 + X_{n-1}}{2}\right) = \frac{1}{4}[V(X_2) + V(X_{n-1})] = \frac{1}{4}(\lambda + \lambda) = \frac{\lambda}{2}$$

Para $n>2$ verifica-se ser $\hat{\lambda}_1$ mais eficiente que $\hat{\lambda}_2$, pois $\frac{V(\hat{\lambda}_1)}{V(\hat{\lambda}_2)} < 1$.

Note-se que $\frac{V(\hat{\lambda}_1)}{V(\hat{\lambda}_2)} = \frac{\frac{\lambda}{n}}{\frac{\lambda}{2}} = \frac{2\lambda}{n\lambda} = \frac{2}{n} < 1$, para $n > 2$.

3.1 Trata-se de um problema que envolve 5 lotes de 20 impressoras cada. As impressoras podem apenas ser classificadas de acordo com uma dicotomia, defeituosa - não defeituosa, admitindo-se que as amostras são extraídas com reposição. Considerando X a variável aleatória representativa do número de impressoras defeituosas por lote, então X será distribuída de acordo com uma binomial de parâmetros 20 e p, ou seja, $X \approx \text{Bin}(20, p)$, em que p representa a probabilidade de uma impressora escolhida ao acaso ser defeituosa. Pode-se então escrever a função de probabilidade de X, como se segue:

$$P(X = x) = \binom{20}{x} p^x (1-p)^{20-x}$$

Para determinação do estimador da máxima verosimilhança primeiro estabelece-se a função de verosimilhança. Tem-se então:

$$L(x_1,x_2,x_3,x_4,x_5;p)=\binom{20}{x_1}p^{x_1}(1-p)^{20-x_1}\times\binom{20}{x_2}p^{x_2}(1-p)^{20-x_2}\times$$

$$\times\binom{20}{x_3}p^{x_3}(1-p)^{20-x_3}\times\binom{20}{x_4}p^{x_4}(1-p)^{20-x_4}\times\binom{20}{x_5}p^{x_5}(1-p)^{20-x_5}=$$

$$=\prod_{i=1}^{5}\binom{20}{x_i}p^{\sum_{i=1}^{5}x_i}\times(1\text{-}p)^{\sum_{i=1}^{5}(20-x_i)}$$

Aplicando logaritmos vem:

$$\ln L=\ln\left[\prod_{i=1}^{5}\binom{20}{x_i}p^{\sum_{i=1}^{5}x_i}\times(1\text{-}p)^{\sum_{i=1}^{5}(20-x_i)}\right]=$$

$$=\ln\left(\prod_{i=1}^{5}\binom{20}{x_i}\right)+\left(\sum_{i=1}^{5}x_i\right)\ln p+\left[\sum_{i=1}^{5}(20-x_i)\right]\ln(1-p)$$

A derivada do logaritmo de L em ordem a p será:

$$\frac{\partial \ln L}{\partial p}=0+\sum_{i=1}^{5}x_i\frac{1}{p}-\sum_{i=1}^{5}(20-x_i)\frac{1}{1-p}$$

Determine-se a segunda derivada, a fim de assegurar tratar-se de um máximo:

$$\frac{\partial^2 \ln L}{\partial p^2}=\sum_{i=1}^{5}x_i\left(\frac{-1}{p^2}\right)-\sum_{i=1}^{5}(20-x_i)\left(\frac{-(-1)}{(1-p)^2}\right)=-\frac{\sum_{i=1}^{5}x_i}{p^2}-\frac{\sum_{i=1}^{5}(20-x_i)}{(1-p)^2}<0$$

Uma vez comprovado que a segunda derivada é negativa, pode

afirmar-se que o ponto estacionário da função de verosimilhança é um máximo.

Assim, igualando a primeira derivada a zero, tem-se:

$$\frac{\partial \ln L}{\partial p} = 0 \Leftrightarrow \sum_{i=1}^{5} x_i \frac{1}{p} - \sum_{i=1}^{5} (20 - x_i) \frac{1}{1-p} = 0 \Leftrightarrow$$

$$\Leftrightarrow (1-p)\sum_{i=1}^{5} x_i - p\sum_{i=1}^{5} (20 - x_i) = 0 \Leftrightarrow \sum_{i=1}^{5} x_i - p\sum_{i=1}^{5} 20 = 0 \Leftrightarrow$$

$$\Leftrightarrow -p\sum_{i=1}^{5} 20 = - \sum_{i=1}^{5} x_i \Leftrightarrow \hat{p} = \frac{\sum_{i=1}^{5} x_i}{5(20)} = \frac{\sum_{i=1}^{5} x_i / 5}{20} = \frac{\overline{x}}{20}$$

O estimador da máxima verosimilhança é dado por:

$$\hat{p} = \frac{\overline{x}}{20}$$

3.2 Para a amostra de 5 lotes tem-se:

$$\hat{p} = \frac{1}{20}\left(\frac{20(0.031) + 20(0.035) + 20(0.027) + 20(0.023 + 20(0.045)}{5}\right) \Leftrightarrow$$

$$\Leftrightarrow \hat{p} = \frac{1}{20}(0.644) = 0.0322$$

Exercícios Propostos

1. Considere uma variável aleatória X com média μ e variância σ^2, e todas as possíveis amostras de tamanho 3. Considere $\hat{\mu}_1$ e $\hat{\mu}_2$ estimadores de μ, tais que:

$$\hat{\mu}_1 = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3 \quad ; \quad \hat{\mu}_2 = \frac{2}{3}X_1 + \frac{1}{6}X_2 + \frac{1}{6}X_3$$

1.1 Qual é destes estimadores o de variância mínima?

1.2 Classifique $\hat{\mu}_1$ e $\hat{\mu}_2$ quanto ao enviesamento.

Solução: **1.1** $\hat{\mu}_1$ **1.2** São ambos estimadores não enviesados.

2. Mostre que:

$$\hat{S}^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

é um estimador não enviesado do parâmetro σ^2.

Solução: Verifica-se que $E(\hat{S}^2) = \sigma^2$.

3. Considere a variável aleatória X, contínua, com função densidade de probabilidade:

$$f(x) = (\lambda+1)(1-x)^{\lambda}, \quad 0 < x < 1$$

onde λ é um parâmetro desconhecido (λ>-1).

3.1 Apresente a função de verosimilhança $L(x,\lambda)$.

3.2 Deduza o estimador de máxima verosimilhança para o parâmetro λ, com base numa amostra de dimensão n.

Solução:

3.1 $L(x,\lambda) = (\lambda+1)^n \prod_{i=1}^{n}(1-x_i)^{\lambda}$; **3.2** $\hat{\lambda} = -1 - \frac{n}{\sum_{i=1}^{n}\ln(1-x_i)}$