

UNIVERSIDADE FEDERAL DE ALAGOAS

INSTITUTO DE COMPUTAÇÃO (IC)

LABORATÓRIO DE COMPUTAÇÃO CIENTÍFICA E ANÁLISE
NUMÉRICA (LACCAN)

**Relatório sobre Estimação por
Máxima Verossimilhança e Estimação
por Momentos**

Autor:

Marcos G. S. do
Nascimento

Orientador:

Alejandro C. Frery

03/07/2019

1 Introdução à Estimação de Parâmetros

Uma variável aleatória é caracterizada ou descrita pela sua distribuição de probabilidade. Esta, por sua vez, é descrita pelos seus parâmetros populacionais. Existe um grande interesse em conhecer os parâmetros populacionais da distribuição que se está trabalhando para conhecer como os dados estão se comportando. Como geralmente tais parâmetros não são conhecidos, é preciso desenvolver procedimentos para estimá-los.

As estimativas dos parâmetros populacionais da distribuição são realizadas a partir dos resultados (dados) de uma variável aleatória de uma amostra representativa extraída da população em questão. A esse procedimento damos o nome de estatística inferencial, visto que se está inferindo algo da população a partir de uma amostra populacional.

Existem duas formas de estimação de parâmetros: por ponto (pontual) ou por intervalo de confiança.

De forma breve, a estimativa pontual é um valor obtido a partir dos resultados (dados) de uma variável aleatória de uma amostra representativa da população. Por outro lado, a estimação de parâmetros por intervalo de confiança consiste em gerar um intervalo, centrado na estimativa pontual, no qual se admite que esteja o parâmetro da população.

2 Tipos de estimadores

Existem vários tipos de estimadores de parâmetros, dentre eles: Momentos, Log-momentos, Máxima Verossimilhança, Máxima Verossimilhança Iterada, Métodos Robustos e Estimação por Kernel.

Em particular, neste relatório será relatado exemplos de estimadores calculados por dois estimadores listados acima: pelo **Método dos Momentos** e pelo **Método da Máxima Verossimilhança**.

2.1 Estimação por Máxima Verossimilhança

A ideia básica deste método é bem intuitiva. Parece razoável que uma boa estimativa do parâmetro desconhecido θ seria o valor que maximiza a probabilidade de se obter os dados que são observados. É por conta dessa ideia que esse método de estimação é denominado de Máxima Verossimilhança.

Na prática, suponha que tenhamos uma amostra aleatória $X = (X_1, X_2, \dots, X_n)$ para a qual a função de densidade de probabilidade de cada X_i é $f(x_i; \theta)$

Então, a função de densidade de probabilidade conjunta da amostra aleatória X , denotada por $L(\theta)$ é dada por:

$$L(\theta) = P(X_1 = x_1 \dots X_n = x_n) = f(x_1; \theta) \dots f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (1)$$

A primeira igualdade é, naturalmente, apenas a definição da função de densidade de probabilidade conjunta. A segunda igualdade vem do fato de que temos uma amostra aleatória, o que implica, por definição, que cada X_i seja independente e igualmente distribuído. A última igualdade apenas usa a notação de produtório. Agora, à luz da ideia básica da estimação por máxima verossimilhança, uma maneira razoável de proceder é tratar a "função de verossimilhança" $L(\theta)$ como uma função de θ , e encontrar o valor do parâmetro que a maximiza.

A seguir estão listados alguns exemplos de estimadores de Máxima Verossimilhança (MV).

2.2 Exemplos de estimadores de MV

2.2.1 Distribuição normal

Vamos estimar a média da população considerando uma amostra de tamanho n . A função de densidade de probabilidade da Normal é dada por:

$$f(x_i, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[\frac{-(x_i - \mu)^2}{2\sigma^2} \right] \quad (2)$$

Portanto, a função de Verossimilhança é dada por:

$$L(\mu, \sigma) = \sigma^{-n} (2\pi)^{-\frac{n}{2}} \exp \left[\frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right] \quad (3)$$

Ao maximizar a função de verossimilhança em relação a μ temos que o melhor estimador para μ é dado por:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \quad (4)$$

2.2.2 Distribuição Uniforme

Considere uma amostra de tamanho n que segue uma distribuição $Uniforme(0, \theta)$ onde θ é desconhecido. Vamos encontrar o estimador de Máxima Verossimilhança para esta amostra.

Temos então que $f(x) = \frac{1}{\theta}$ se $0 \leq x \leq \theta$ ou 0, caso contrário.

A função de Verossimilhança é dada por:

$$L(x_1, x_2, \dots, x_n; \theta) = \frac{1}{\theta^n} \quad (5)$$

Lembrando que o caso acima é válido para $x_1, x_2, \dots, x_n \leq \theta$, valendo 0, caso contrário.

Note que $\frac{1}{\theta}$ é uma função decrescente. Então para minimizá-la nós escolhendo o menor valor possível para θ . Considerando o tamanho da amostra, nós precisamos ter $\theta \geq x_i$.

Considerando o dito acima, temos que o melhor estimador é dado por:

$$\hat{\theta} = \max(X_1, X_2, \dots, X_n) \quad (6)$$

2.2.3 Distribuição Exponencial

Considerando a forma como os estimadores acima foram calculados, temos que para a distribuição exponencial o processo é análogo.

A função de Verossimilhança é dada por:

$$L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^n x_i} \quad (7)$$

Maximizando essa função, temos:

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n X_i} \quad (8)$$

2.2.4 Distribuição Binomial

Agora estamos no contexto de uma **distribuição discreta**, mas em ambos os casos de termos uma Distribuição Contínua ou Discreta, o estimador de Máxima Verossimilhança para o parâmetro desconhecido θ é o valor discreto ou contínuo que maximiza a função de verossimilhança, independente do tipo de distribuição.

Então, de forma análoga ao que já foi calculado anteriormente, considere uma amostra de tamanho n e que $X_i \sim \text{Binomial}(m, \theta)$. Temos que a função de verossimilhança é calculada como:

$$L(x_1, x_2, \dots, x_n; \theta) = f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n; \theta) \quad (9)$$

$$= \prod_{i=1}^n f_{X_i}(x_i; \theta) \quad (10)$$

$$= \prod_{i=1}^n \binom{m}{x_i} \theta^{x_i} (1 - \theta)^{m-x_i} \quad (11)$$

$$= \left[\prod_{i=1}^n \binom{m}{x_i} \right] \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{mn - \sum_{i=1}^n x_i}. \quad (12)$$

Note que na equação 12 o primeiro termo não depende do parâmetro θ , então podemos reescrever a equação da seguinte forma:

$$L(x_1, x_2, \dots, x_n; \theta) = c \quad \theta^s (1 - \theta)^{mn-s}, \quad (13)$$

onde c não depende de θ e $s = \sum_{k=1}^n x_i$.

Então, derivando esta equação e igualando a derivada a 0 para encontrar o ponto de máximo, temos que:

$$\hat{\theta} = \frac{1}{mn} \sum_{k=1}^n x_i. \quad (14)$$

2.2.5 Distribuição Poisson

Novamente, suponha que $X = (X_1, X_2, \dots, X_n)$ são observações iid que segue uma distribuição Poisson com o parâmetro λ desconhecido. A função de verossimilhança é dada a seguir:

$$L(x_1, x_2, \dots, x_n; \lambda) = \prod_{i=1}^n f(x_i; \lambda) \quad (15)$$

$$= \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \quad (16)$$

$$= \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{x_1! x_2! \dots x_n!} \quad (17)$$

$$(18)$$

Diferenciando o log dessa função em relação a λ , isto é, diferenciando a função log-Verossimilhança de Poisson, temos:

$$l(x_1, x_2, \dots, x_n; \lambda) = \sum_{i=1}^n x_i \log \lambda - n\lambda \quad (19)$$

Ignorando os termos constantes que não dependem do parâmetro em questão, obtemos que o estimador de Máxima Verossimilhança é dado por:

$$\hat{\lambda} = \sum_{i=1}^n \frac{x_i}{n} \quad (20)$$

2.3 Estimação pelo Método dos Momentos

Uma outra forma de encontrar estimadores de parâmetros populacionais, como a média e a variância por exemplo, é através do método dos momentos. Este método é baseado nos momentos teóricos e amostrais das variáveis aleatórias envolvidas. É mais simples de calcular em relação ao método anterior de Máxima Verossimilhança, mas pode gerar, com mais frequência, estimadores menos precisos.

Em suma, o método dos momentos envolve equacionar momentos da amostra com momentos teóricos. Então, vamos começar lembrando as definições dos momentos teóricos, bem como aprender as definições dos momentos das amostras.

A seguir estão algumas definições:

1. $E(X^k)$ é o k-ésimo momento teórico da distribuição (sobre a origem)
2. $E[(X - \mu)^k]$ é o k-ésimo momento teórico (sobre a média)
3. $M_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ é o k-ésimo momento amostral
4. $M_k^* = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$ é o k-ésimo momento amostral sobre a média

A ideia por trás desse método é bastante simples. Vejamos os seguintes passos da primeira forma de se calcular os estimadores por este método:

1. Equacione o primeiro momento da amostra sobre a origem $M_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ ao primeiro momento teórico $E(x)$
2. Equacione o segundo momento da amostra sobre a origem $M_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$ ao segundo momento teórico $E(x^2)$
3. Continuar igualando momentos amostrais sobre a origem, M_k , com os correspondentes momentos teóricos $E(X^k)$, $k = 3, 4, \dots$, até que se tenha tantas equações quanto se tem de parâmetros.
4. Por fim, resolva os parâmetros.

Os valores resultantes são chamados de estimadores de método de momentos. Parece razoável que esse método forneça boas estimativas, já que a distribuição empírica converge, em certo sentido, para a distribuição de probabilidade. Portanto, os momentos correspondentes devem ser aproximadamente iguais.

Em alguns casos, em vez de usar os momentos da amostra sobre a origem, é mais fácil usar os momentos da amostra sobre a média. Ao fazê-lo, fornecemos uma forma alternativa do método dos momentos que é análogo ao método anterior só que com a diferença de a partir do segundo momento as equações envolvem os momentos amostrais sobre a média ao invés da origem.

2.3.1 Distribuição de Bernoulli

suponha que $X = (X_1, X_2, \dots, X_n)$ são observações aleatórias iid que seguem uma distribuição Bernoulli com o parâmetro p desconhecido. Vamos calcular o estimador de momentos para p .

O primeiro momento teórico sobre a origem é:

$$E(x_i) = p \quad (21)$$

Temos apenas um parâmetro para o qual estamos tentando derivar o método do estimador de momentos. Portanto, precisamos apenas de uma equação. Igualando o primeiro momento teórico sobre a origem com o momento de amostra correspondente, obtemos:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i \quad (22)$$

2.3.2 Distribuição Normal

suponha que $X = (X_1, X_2, \dots, X_n)$ são observações aleatórias iid que seguem uma distribuição Bernoulli com os parâmetros média μ e variância σ^2 . Vamos calcular o estimador de momentos para ambos.

Os primeiros e segundos momentos teóricos sobre a origem são:

$$E(X_i) = \mu \quad (23)$$

$$E(X_i^2) = \sigma^2 + \mu^2 \quad (24)$$

Incidentalmente, esse segundo momento pode ser derivado da manipulação da fórmula de atalho para a variância. Nesse caso, temos dois parâmetros para os quais estamos tentando derivar o método de estimadores de momentos. Portanto, precisamos de duas equações aqui. Igualando o primeiro momento teórico sobre a origem com o momento de amostra correspondente, obtemos:

$$E(X) = \mu = \frac{1}{n} \sum_{i=1}^n X_i \quad (25)$$

E, igualando o segundo momento teórico sobre a origem com o momento da amostra correspondente, obtemos:

$$E(X^2) = \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \quad (26)$$

Agora, a primeira equação nos diz que o estimador de momentos para a média μ é a média da amostra:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \quad (27)$$

E, substituindo a média da amostra por μ na segunda equação e resolvendo para σ^2 , obtemos que o estimador de momentos para a variância é:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \quad (28)$$

que pode ser reescrito como:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (29)$$

Para este exemplo, os estimadores de método de momentos são os mesmos dos estimadores de máxima verossimilhança.

2.3.3 Distribuição Gama

Suponha que $X = (X_1, X_2, \dots, X_n)$ são observações aleatórias iid que seguem uma distribuição Gama com os parâmetros α (Forma) e β (Escala).

Utilizando o método da Máxima Verossimilhança é complicado diferenciar por conta da função gama maior dada por $\Gamma(\alpha)$. Então, ao invés de encontrar os estimadores de máxima verossimilhança, vamos encontrar o estimador de momentos para ambos os parâmetros mencionados.

O primeiro momento teórico sobre a origem é:

$$E(X_i) = \alpha\theta \quad (30)$$

O segundo momento teórico sobre a média é:

$$Var(X_i) = E(X_i - \mu) = \alpha\theta^2 \quad (31)$$

Como temos dois parâmetros para os quais estamos tentando derivar estimadores de momentos, precisamos de duas equações. Igualando o primeiro momento teórico sobre a origem com o momento de amostra correspondente, obtemos:

$$E(X) = \alpha\theta = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \quad (32)$$

E, igualando o segundo momento teórico sobre a média com o momento da amostra correspondente, obtemos:

$$Var(X) = \alpha\theta^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (33)$$

Agora, temos apenas que resolver os dois parâmetros. Vamos começar resolvendo α na primeira equação ($E(X)$). Fazendo isso, obtemos:

$$\alpha = \frac{\bar{X}}{\theta} \quad (34)$$

Agora, substituindo o valor α na segunda equação ($Var(x)$), temos:

$$\alpha\theta^2 = \left(\frac{\bar{X}}{\theta}\right) \theta^2 = \bar{X}\theta = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (35)$$

Resolvendo para θ nessa última equação, obtemos que o método do estimador de momento é dado por:

$$\hat{\theta} = \frac{1}{n\bar{X}} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (36)$$

Substituindo de volta esse valor de θ na equação que temos para α , obtemos que o método do estimador de momento para esse parâmetro é dado por:

$$\hat{\alpha} = \frac{\bar{X}}{\hat{\theta}_{MM}} = \frac{\bar{X}}{(1/n\bar{X}) \sum_{i=1}^n (X_i - \bar{X})^2} = \frac{n\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (37)$$

3 Conclusões

Em suma, o método da Máxima Verossimilhança apresenta os 5 passos básicos a seguir:

1. Encontrar a função de verossimilhança
2. Aplicar o logaritmo de base natural (\ln)
3. Derivar a função com relação ao parâmetro desconhecido e igualar a zero para encontrar o máximo
4. Verificar se o valor encontrado é um ponto de máximo

Neste método, tentamos encontrar os valores dos parâmetros que seriam mais prováveis de produzir os dados que de fato observamos. Em outras palavras, o princípio desse estimador é encontrar valores para os parâmetros da função de Verossimilhança que maximizem a probabilidade os valores amostrais da variável aleatória.

Prós do Método da Máxima Verossimilhança (MLE):

- Para amostras de tamanho grande (30 em diante) MLE é não enviesado (imparcial)
- Consistente
- Eficiente (produz o MSE mínimo do que outros métodos, incluindo o Método dos Momentos)
- Método mais usado em inferências estatísticas

Contras do Método da Máxima Verossimilhança (MLE):

- O MLE pode ser altamente tendencioso para pequenas amostras.
- Às vezes, o MLE não tem solução de forma fechada.
- O MLE pode ser sensível aos valores iniciais, o que pode não dar um ótimo global.

Por outro lado, o Método dos Momentos (MM) envolve equacionar momentos da amostra com momentos teóricos. Consiste em um dos métodos mais antigos e possui procedimento muito simples. É baseado no pressuposto de que os momentos da amostra devem fornecer BOAS ESTIMATIVAS dos momentos correspondentes da população.

Prós do Método dos Momentos (MM):

- Fácil de calcular e sempre funciona
- O método geralmente fornece estimadores quando outros métodos não o fazem ou quando os estimadores são difíceis de obter (como no caso da distribuição Gama).
- É um estimador Consistente

Contras do Método dos Momentos (MM):

- Eles geralmente não são os “melhores estimadores” disponíveis. Pelo melhor, queremos dizer mais eficiente, ou seja, atingir mínimo MSE.
- Às vezes pode ser sem sentido. Por exemplo na estimação do parâmetro θ de uma distribuição Uniforme $U(0, \theta)$.