Inferência Estatística

Introdução

E.F.T[1]

[1]EACH-USP
Universidade de São Paulo

ACH2053

Outline

1. Estimadores de Bayes
   - Conceitos Gerais
   - Funções de Perda
   - Diferentes Funções de Perda
   - Estimadores de Máxima Verossimilhança

Outline

Natureza de um problema de estimação
utilidade

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição com $f(x|\theta)$, com θ desconhecido e que $\theta \in \Omega$. Suponha finalmente que o valor de θ será estimado a partir dos valores observados da amostra.

Um `estimador de um parâmetro` θ baseado nas variáveis $X_1, X_2, ..., X_n$ é uma função a valores reais $\delta(X_1, X_2, ..., X_n)$ que especifica o valor estimado de θ para cada possível conjunto de valores de $X_1, X_2, ..., X_n$.

Se os valores observados de $X_1, X_2, ..., X_n$ resultam em $x_1, ..., x_n$ então o valor estimado de θ será $\delta(x_1, ..., x_n)$. Como θ pertence ao intervalo Ω, é razoável pensar que $\delta(X_1, X_2, ..., X_n)$ deve pertencer a Ω.

Natureza de um problema de estimação

utilidade

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição com $f(x|\theta)$, com θ desconhecido e que $\theta \in \Omega$. Suponha finalmente que o valor de θ será estimado a partir dos valores observados da amostra.

Um `estimador de um parâmetro` θ baseado nas variáveis $X_1, X_2, ..., X_n$ é uma função a valores reais $\delta(X_1, X_2, ..., X_n)$ que especifica o valor estimado de θ para cada possível conjunto de valores de $X_1, X_2, ..., X_n$.

Se os valores observados de $X_1, X_2, ..., X_n$ resultam em $x_1, ..., x_n$ então o valor estimado de θ será $\delta(x_1, ..., x_n)$. Como θ pertence ao intervalo Ω, é razoável pensar que $\delta(X_1, X_2, ..., X_n)$ deve pertencer a Ω.

Natureza de um problema de estimação
utilidade

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição com $f(x|\theta)$, com θ desconhecido e que $\theta \in \Omega$. Suponha finalmente que o valor de θ será estimado a partir dos valores observados da amostra.

Um `estimador de um parâmetro` θ baseado nas variáveis $X_1, X_2, ..., X_n$ é uma função a valores reais $\delta(X_1, X_2, ..., X_n)$ que especifica o valor estimado de θ para cada possível conjunto de valores de $X_1, X_2, ..., X_n$.

Se os valores observados de $X_1, X_2, ..., X_n$ resultam em $x_1, ..., x_n$ então o valor estimado de θ será $\delta(x_1, ..., x_n)$. Como θ pertence ao intervalo Ω, é razoável pensar que $\delta(X_1, X_2, ..., X_n)$ deve pertencer a Ω.

Natureza de um problema de estimação

Distinção entre estimador e estimativa

A distinção entre os termos *estimador* e *estimativa* resulta de que um *estimador* $\delta(X_1, X_2, ..., X_n)$ é uma função das variáveis aleatórias $X_1, X_2, ..., X_n$, sendo portanto uma v.a. e sua distribuição pode ser derivada da conjunta de $X_1, X_2, ..., X_n$.
Por outro lado, uma *estimativa* é um valor específico $\delta(x_1, ..., x_n)$ do estimador, que é determinado usando valores específicos obervados $x_1, ..., x_n$.
Na notação vetorial, representaremos $\mathbf{X} = (X_1, X_2, ..., X_n)$ e $\mathbf{x} = (x_1, ..., x_n)$, assim, $\delta(\mathbf{X})$ (ou simplesmente δ) denotará um estimador e $\delta(\mathbf{x})$ uma estimativa.

Natureza de um problema de estimação

Distinção entre estimador e estimativa

A distinção entre os termos *estimador* e *estimativa* resulta de que um *estimador* $\delta(X_1, X_2, ..., X_n)$ é uma função das variáveis aleatórias $X_1, X_2, ..., X_n$, sendo portanto uma v.a. e sua distribuição pode ser derivada da conjunta de $X_1, X_2, ..., X_n$. Por outro lado, uma *estimativa* é um valor específico $\delta(x_1, ..., x_n)$ do estimador, que é determinado usando valores específicos obervados $x_1, ..., x_n$.

Na notação vetorial, representaremos $\mathbf{X} = (X_1, X_2, ..., X_n)$ e $\mathbf{x} = (x_1, ..., x_n)$, assim, $\delta(\mathbf{X})$ (ou simplesmente δ) denotará um estimador e $\delta(\mathbf{x})$ uma estimativa.

Natureza de um problema de estimação

Distinção entre estimador e estimativa

A distinção entre os termos *estimador* e *estimativa* resulta de que um *estimador* $\delta(X_1, X_2, ..., X_n)$ é uma função das variáveis aleatórias $X_1, X_2, ..., X_n$, sendo portanto uma v.a. e sua distribuição pode ser derivada da conjunta de $X_1, X_2, ..., X_n$.
Por outro lado, uma *estimativa* é um valor específico $\delta(x_1, ..., x_n)$ do estimador, que é determinado usando valores específicos obervados $x_1, ..., x_n$.
Na notação vetorial, representaremos $\mathbf{X} = (X_1, X_2, ..., X_n)$ e $\mathbf{x} = (x_1, ..., x_n)$, assim, $\delta(\mathbf{X})$ (ou simplesmente δ) denotará um estimador e $\delta(\mathbf{x})$ uma estimativa.

Outline

1. Estimadores de Bayes
   - Conceitos Gerais
   - Funções de Perda
   - Diferentes Funções de Perda
   - Estimadores de Máxima Verossimilhança

Funções de Perda

Um bom estimador δ de um parâmetro θ é aquele que com alta probabilidade, o valor do erro definido por $\delta(X) - \theta$ será próximo de 0.

Devemos assumir que para cada possível valor de θ em Ω e cada possível estimativa $a \in \Omega$ existe um número $L(\theta, a)$ que mede a perda ou o custo quando o valor verdadeiro é θ e sua estimativa é a. De forma geral, a mayor distância entre θ e a, maior será o valor de $L(\theta, a)$.

Funções de Perda

Um bom estimador δ de um parâmetro θ é aquele que com alta probabilidade, o valor do erro definido por $\delta(X) - \theta$ será próximo de 0.
Devemos assumir que para cada possível valor de θ em Ω e cada possível estimativa $a \in \Omega$ existe um número $L(\theta, a)$ que mede a perda ou o custo quando o valor verdadeiro é θ e sua estimativa é a. De forma geral, a mayor distância entre θ e a, maior será o valor de $L(\theta, a)$.

Funções de Perda

Considere um problema em que deve ser estimado θ sem se observar valores na a.a. Se escolhermos uma particular estimativa a, então a sua perda esperada será:

$$E[L(\theta, a)] = \int_{\Omega} L(\theta, a)\xi(\theta)d(\theta) \tag{1}$$

Assumimos também que se escolherá uma estimativa para a qual a perda esperada será mínima.

Em problemas de estimação, uma função L para a qual o valor de $E[L(\theta, a)]$ será minimizada é chamada *função de perda*

Funções de Perda

Considere um problema em que deve ser estimado θ sem se observar valores na a.a. Se escolhermos uma particular estimativa a, então a sua perda esperada será:

$$E[L(\theta, a)] = \int_{\Omega} L(\theta, a)\xi(\theta)d(\theta) \qquad (1)$$

Assumimos também que se escolherá uma estimativa para a qual a perda esperada será mínima.
Em problemas de estimação, uma função L para a qual o valor de $E[L(\theta, a)]$ será minimizada é chamada *função de perda*

Definição de um estimador de Bayes

Supondo que o valor de **x** do vetor aleatório **X** pode ser observado antes de estimar θ, e seja $\xi(\theta|\mathbf{x})$ a posteriori de θ em Ω. Para qualquer estimativa *a* que possa ser usada, sua perda esperada neste caso, será:

$$E[L(\theta, a)|\mathbf{x}] = \int_{\Omega} L(\theta, a)\xi(\theta|\mathbf{x})d(\theta) \tag{2}$$

Desta forma, deveria-se escolher uma estimativa *a* para a qual o valor esperado é mínimo.

Definição de um estimador de Bayes

Para cada valor possível de $\mathbf{x}$ do vetor aleatório $\mathbf{X}$, seja $\delta^*(\mathbf{x})$ o valor da estimativa a para a qual o valor esperado em 2 es mínimo. Então, a função $\delta^*(\mathbf{X})$ para a qual os valores são especificados será um estimador de θ. Este estimador é chamado de *estimador de Bayes de* θ. Para cada $\mathbf{x}$ de $\mathbf{X}$, o valor $\delta^*(\mathbf{x})$ é escolhido de forma a que

$$E[L(\theta, \delta^*(\mathbf{x}))|\mathbf{x}] = \min_{a \in \Omega} E[L(\theta, a)|\mathbf{x}] \qquad (3)$$

Outline

Função de Perda Erro Quadrático

Esta função é definida como

$$L(\theta, a) = (\theta - a)^2 \tag{4}$$

quando esta função é usada, a estimativa de Bayes $\delta^*(\mathbf{x})$ para qualquer valor observado de $\mathbf{x}$ será o valor de a para a qual $E[(\theta - a)^2|\mathbf{x}]$ é mínimo.
Pode ser mostrado que para qualquer distribuição de θ, o valor esperado de $(\theta - a)^2$ será mínimo quando o valor de a escolhido é igual à média da distribuição de θ. Assim, quando o valor esperado de $(\theta - a)^2$ é calculado com respeito à posteriori de θ, o a mínimo será igual à média $E(\theta|\mathbf{x})$ da posteriori. Isto é: $\delta^*(\mathbf{X}) = E(\theta|\mathbf{X})$.

Função de Perda Erro Quadrático

Esta função é definida como

$$L(\theta, a) = (\theta - a)^2 \tag{4}$$

quando esta função é usada, a estimativa de Bayes $\delta^*(\mathbf{x})$ para qualquer valor observado de $\mathbf{x}$ será o valor de a para a qual $E[(\theta - a)^2|\mathbf{x}]$ é mínimo.
Pode ser mostrado que para qualquer distribuição de θ, o valor esperado de $(\theta - a)^2$ será mínimo quando o valor de a escolhido é igual à média da distribuição de θ. Assim, quando o valor esperado de $(\theta - a)^2$ é calculado com respeito à posteriori de θ, o a mínimo será igual à média $E(\theta|\mathbf{x})$ da posteriori. Isto é: $\delta^*(\mathbf{X}) = E(\theta|\mathbf{X})$.

Função de Perda Erro Absoluto

Esta função é definida como:

$$L(\theta, a) = |\theta - a| \tag{5}$$

Para qualquer valor observado de **x**, o estimador de Bayes será o valor de a para a qual o valor esperado $E(|\theta - a||\mathbf{x})$ é minimo. É possível mostrar que para qualquer distribuição de probabilidade de θ, o valor esperado de $|\theta - a|$ será mínimo quando o a escolhido é igual à mediana da distribuição de θ. Da mesma forma, quando o valor esperado é calculado com respeito da distribuição a posteriori de θ, este valor será mínimo quando o a escolhido for igual à mediana da distribuição a posteriori de θ.

Função de Perda Erro Absoluto

Esta função é definida como:

$$L(\theta, a) = |\theta - a| \tag{5}$$

Para qualquer valor observado de **x**, o estimador de Bayes será o valor de *a* para a qual o valor esperado $E(|\theta - a||\mathbf{x})$ é minimo. É possível mostrar que para qualquer distribuição de probabilidade de θ, o valor esperado de $|\theta - a|$ será mínimo quando o *a* escolhido é igual à mediana da distribuição de θ. Da mesma forma, quando o valor esperado é calculado com respeito da distribuição a posteriori de θ, este valor será mínimo quando o *a* escolhido for igual à mediana da distribuição a posteriori de θ.

Outras Funções de Perda

Em alguns problemas pode ser apropriado usar funções de perda com a forma

$$L(\theta, a) = |\theta - a|^k \tag{6}$$

onde k é algum inteiro diferente de 1 ou 2.

Outra função de perda poderia ser por exemplo:

$$L(\theta, a) = \lambda(\theta)(\theta - a)^2 \tag{7}$$

onde $\lambda(\theta)$ é alguma função positiva em θ.

Outras Funções de Perda

Em alguns problemas pode ser apropriado usar funções de perda com a forma

$$L(\theta, a) = |\theta - a|^k \tag{6}$$

onde k é algum inteiro diferente de 1 ou 2.
Outra função de perda poderia ser por exemplo:

$$L(\theta, a) = \lambda(\theta)(\theta - a)^2 \tag{7}$$

onde $\lambda(\theta)$ é alguma função positiva em θ.

Estimadores de Bayes para grandes amostras

Suponha que a proporção θ de ítens defeituosos em um grande carregamento é desconhecida e que a priori de θ é uniforme em $(0, 1)$. Suponha que na estimativa de θ a função de erro quadrático será usada. Suponha finalmente que em uma a.a. de 100 ítens do carregamento, exatamente 10 ítens são encontrados defeituosos.

Como a distribuição Uniforme é uma Beta com parâmetros $\alpha = 1$ e $\beta = 1$ e como $n = 100$ e $y = \sum x_i = 10$ segue-se que a estimativa de Bayes é $\delta(\mathbf{x}) = 11/102 = 0,108$.

Estimadores de Bayes para grandes amostras

Suponha que a proporção θ de ítens defeituosos em um grande carregamento é desconhecida e que a priori de θ é uniforme em $(0,1)$. Suponha que na estimativa de θ a função de erro quadrático será usada. Suponha finalmente que em uma a.a. de 100 ítens do carregamento, exatamente 10 ítens são encontrados defeituosos.
Como a distribuição Uniforme é uma Beta com parâmetros $\alpha = 1$ e $\beta = 1$ e como $n = 100$ e $y = \sum x_i = 10$ segue-se que a estimativa de Bayes é $\delta(\mathbf{x}) = 11/102 = 0,108$.

Estimadores de Bayes para grandes amostras

Suponha agora, que em lugar de uma distribuição uniforme, usamos a priori: $\xi(\theta) = 2(1-\theta)$ para $0 < \theta < 1$, e os resultados da a.a. são os mesmos de acima. Como $\xi(\theta)$ é uma beta com parâmetros $\alpha = 1$ e $\beta = 2$, então a estimativa de Bayes é $\delta(\mathbf{x}) = 11/103 = 0,107$.

As duas prioris são diferentes, a primeira com média $1/2$ e a segunda com média $1/3$. Mas, como o número de observações é grande ($n = 100$), as estimativas de Bayes são similares, além de que estão próximos da estimativa de proporções de itens defeituosos na amostra: $\bar{x}_n = 0,1$.

Estimadores de Bayes para grandes amostras

Suponha agora, que em lugar de uma distribuição uniforme, usamos a priori: $\xi(\theta) = 2(1-\theta)$ para $0 < \theta < 1$, e os resultados da a.a. são os mesmos de acima. Como $\xi(\theta)$ é uma beta com parâmetros $\alpha = 1$ e $\beta = 2$, então a estimativa de Bayes é $\delta(\mathbf{x}) = 11/103 = 0,107$.
As duas prioris são diferentes, a primeira com média $1/2$ e a segunda com média $1/3$. Mas, como o número de observações é grande ($n = 100$), as estimativas de Bayes são similares, além de que estão próximos da estimativa de proporções de itens defeituosos na amostra: $\bar{x}_n = 0,1$.

Consistência dos Estimadores de Bayes

Uma sequência de estimadores que converge ao valor desconhecido do parâmetro sendo estimado, quando $n \to \infty$ é chamado de um *sequência consistente de estimadores* .

Consistência dos Estimadores de Bayes

Uma sequência de estimadores que converge ao valor desconhecido do parâmetro sendo estimado, quando $n \to \infty$ é chamado de um *sequência consistente de estimadores* .

Outline

Limitações dos Estimadores de Bayes

para aplicar a teoría dos estimadores Bayesianos, é necessário:

- especificar a função de perda,
- determinar uma priori para o parâmetro.

ou também, θ pode ser um vetor, para a qual haveria de se especificar uma distribuição a priori multivariada.

Limitações dos Estimadores de Bayes

para aplicar a teoría dos estimadores Bayesianos, é necessário:

- especificar a função de perda,
- determinar uma priori para o parâmetro.

ou também, θ pode ser um vetor, para a qual haveria de se especificar uma distribuição a priori multivariada.

Limitações dos Estimadores de Bayes

para aplicar a teoría dos estimadores Bayesianos, é necessário:

- especificar a função de perda,
- determinar uma priori para o parâmetro.

ou também, θ pode ser um vetor, para a qual haveria de se especificar uma distribuição a priori multivariada.

Definição de um Estimador de Máxima Verossimilhança

Suponha que as v.a. $X_1, ..., X_n$ formam uma a.a. cuja distribuição é $f(x|\theta)$, onde θ é desconhecido e pertence ao espaço paramêtrico Ω. Para qualquer vetor observado $\mathbf{x} = (x_1, ..., x_n)$ na amostra, o valor da conjunta será denotados por $f_n(\mathbf{x}|\theta)$.

Quando $f_n(\mathbf{x}|\theta)$ é considerado uma função de θ para um vetor $\mathbf{x}$ dado, é chamado de *função de verossimilhança*.

Definição de um Estimador de Máxima Verossimilhança

Suponha que as v.a. $X_1, ..., X_n$ formam uma a.a. cuja distribuição é $f(x|\theta)$, onde θ é desconhecido e pertence ao espaço paramêtrico Ω. Para qualquer vetor observado $\mathbf{x} = (x_1, ..., x_n)$ na amostra, o valor da conjunta será denotados por $f_n(\mathbf{x}|\theta)$.
Quando $f_n(\mathbf{x}|\theta)$ é considerado uma função de θ para um vetor $\mathbf{x}$ dado, é chamado de *função de verossimilhança*.

Definição de um Estimador de Máxima Verossimilhança

Suponha que as v.a. $X_1, ..., X_n$ formam uma a.a. cuja distribuição é $f(x|\theta)$, onde θ é desconhecido e pertence ao espaço paramêtrico Ω. Para qualquer vetor observado $\mathbf{x} = (x_1, ..., x_n)$ na amostra, o valor da conjunta será denotados por $f_n(\mathbf{x}|\theta)$.
Quando $f_n(\mathbf{x}|\theta)$ é considerado uma função de θ para um vetor $\mathbf{x}$ dado, é chamado de *função de verossimilhança*.

Definição de um Estimador de Máxima Verossimilhança

Supondo que o vetor **x** vem de uma distribuição discreta. Se uma estimativa de θ deve ser selecionada, não considerariamos qualquer valor de $\theta \in \Omega$ com o qual seria impossível conseguir o atual valor de **x**.

Suponha que a probabilidade $f_n(\mathbf{x}|\theta)$ de obter o atual vetor observado é alta para determinado valor de $\theta = \theta_0$ e pequena para qualquer outro valor de $\theta \in \Omega$. Então, naturalmente estimariamos o valor de θ como θ_0.

Definição de um Estimador de Máxima Verossimilhança

Supondo que o vetor $\mathbf{x}$ vem de uma distribuição discreta. Se uma estimativa de θ deve ser selecionada, não considerariamos qualquer valor de $\theta \in \Omega$ com o qual seria impossível conseguir o atual valor de $\mathbf{x}$.
Suponha que a probabilidade $f_n(\mathbf{x}|\theta)$ de obter o atual vetor observado é alta para determinado valor de $\theta = \theta_0$ e pequena para qualquer outro valor de $\theta \in \Omega$. Então, naturalmente estimariamos o valor de θ como θ_0.

Definição de um Estimador de Máxima Verossimilhança

Supondo que o vetor **x** vem de uma distribuição discreta. Se uma estimativa de θ deve ser selecionada, não considerariamos qualquer valor de $\theta \in \Omega$ com o qual seria impossível conseguir o atual valor de **x**.
Suponha que a probabilidade $f_n(\mathbf{x}|\theta)$ de obter o atual vetor observado é alta para determinado valor de $\theta = \theta_0$ e pequena para qualquer outro valor de $\theta \in \Omega$. Então, naturalmente estimariamos o valor de θ como θ_0.

Definição de um Estimador de Máxima Verossimilhança

Para cada valor possível do vetor $\mathbf{x}$, seja $\delta(\mathbf{x}) \in \Omega$ o valor de $\theta \in \Omega$ para a qual a função de verossimilhança $f_n((\mathbf{x}|\theta)$ é um máximo, e seja $\hat{\theta} = \delta(\mathbf{X})$ o estimador de θ definido desta forma. O estimador $\hat{\theta}$ é chamado de *estimador de maxima verossimilhança* de θ, ou abreviadamente EMV de θ.