Inferência Estatística

Introdução

E.F.T[1]

[1]EACH-USP
Universidade de São Paulo

ACH2053

Outline

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.

Outline

Exemplos de EMV
casos de nota

- Em alguns problemas, para valores observados $\mathbf{x}$, um valor máximo para $f_n(\mathbf{x}|\theta)$ pode não ser obtido para qualquer ponto $\theta \in \Omega$. Neste caso o EMV não existe.
- Para certos $\mathbf{x}$ observados, o máximo para $f_n(\mathbf{x}|\theta)$ pode ser obtido em vários pontos de Ω , em tais casos, o EMV não é único e um destes pontos pode ser escolhido como a estimativa $\hat{\theta}$.
- Em outros casos, o EMV é existe e é único.

Exemplos de EMV

casos de nota

- Em alguns problemas, para valores observados **x**, um valor máximo para $f_n(\mathbf{x}|\theta)$ pode não ser obtido para qualquer ponto $\theta \in \Omega$. Neste caso o EMV não existe.
- Para certos **x** observados, o máximo para $f_n(\mathbf{x}|\theta)$ pode ser obtido em vários pontos de Ω , em tais casos, o EMV não é único e um destes pontos pode ser escolhido como a estimativa $\hat{\theta}$.
- Em outros casos, o EMV é existe e é único.

Exemplos de EMV

casos de nota

- Em alguns problemas, para valores observados $\mathbf{x}$, um valor máximo para $f_n(\mathbf{x}|\theta)$ pode não ser obtido para qualquer ponto $\theta \in \Omega$. Neste caso o EMV não existe.
- Para certos $\mathbf{x}$ observados, o máximo para $f_n(\mathbf{x}|\theta)$ pode ser obtido em vários pontos de Ω , em tais casos, o EMV não é único e um destes pontos pode ser escolhido como a estimativa $\hat{\theta}$.
- Em outros casos, o EMV é existe e é único.

Exemplos de EMV

Amostragem de uma Bernoulli

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição de Bernoulli com θ desconhecido ($0 \leq \theta \leq 1$). Para valores observados $x_1, ..., x_n$, onde cada x_i é 0 ou 1, a função de verossimilhança é:

$$f_n(\mathbf{x}|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \tag{1}$$

Exemplos de EMV

Amostragem de uma Bernoulli

O valor de θ que maximiza $f_n(\mathbf{x}|\theta)$ será o mesmo que maximiza $logf_n(\mathbf{x}|\theta)$. Portanto, será conveniente determinar o EMV encontrando o valor de θ que maximiza:

$$\begin{aligned} L(\theta) &= logf_n(\mathbf{x}|\theta) = \sum_{i=1}^{n}[x_i log(\theta) = (1-x_i)log(1-\theta)] \quad (2) \\ &= (\sum_{i=1}^{n} x_i)log(\theta) + (n - \sum_{i=1}^{n} x_i)log(1-\theta) \end{aligned}$$

Exemplos de EMV

Amostragem de uma Bernoulli

Se calcularmos a derivada $dL(\theta)/d\theta$, e igualarmos esta derivada a 0, resolvendo a equação para θ encontraremos que $\theta = \bar{x}_n$. Pode ser verificado que este valor maximiza $L(\theta)$, portanto, maximiza a função de verossimilhança. Segue-se que a EMV de θ é $\hat{\theta} = \bar{X}_n$.

Logo, se $X_1, X_2, ..., X_n$ são n ensaios de Bernoulli, então o EMV de probabilidade de sucesso (desconhecida) en qualquer ensaio dado é simplesmente a proporção de sucessos observados nos n ensaios.

Exemplos de EMV

Amostragem de uma Bernoulli

Se calcularmos a derivada $dL(\theta)/d\theta$, e igualarmos esta derivada a 0, resolvendo a equação para θ encontraremos que $\theta = \bar{x}_n$. Pode ser verificado que este valor maximiza $L(\theta)$, portanto, maximiza a função de verossimilhança. Segue-se que a EMV de θ é $\hat{\theta} = \bar{X}_n$.
Logo, se $X_1, X_2, ..., X_n$ são n ensaios de Bernoulli, então o EMV de probabilidade de sucesso (desconhecida) en qualquer ensaio dado é simplesmente a proporção de sucessos observados nos n ensaios.

Exemplos de EMV

Amostragem de uma Normal (com variância conhecida)

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição Normal com média μ desconhecida e variância σ^2 conhecida. Para valores observados $x_1, ..., x_n$, a função de verossimilhança é:

$$f_n(\mathbf{x}|\mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2] \tag{3}$$

Exemplos de EMV

Amostragem de uma Normal(com variância conhecida)

Pode-se ver que $f_n(\mathbf{x}|\mu)$ será maximizado pelo valor de μ que maximiza

$$Q(\mu) = \sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + n\mu^2$$

Se calcularmos a derivada $dQ(\mu)/d\mu$, igualamos a 0 e resolvemos esta equação para μ encontraremos que $\mu = \bar{x}_n$.

Segue-se então que o EMV de μ é $\hat{\mu} = \bar{X}_n$.

Pode ser visto acima que o valor de $\hat{\mu}$ não é afetado pelo valor da variância σ^2.

Exemplos de EMV

Amostragem de uma Normal(com variância conhecida)

Pode-se ver que $f_n(\mathbf{x}|\mu)$ será maximizado pelo valor de μ que maximiza

$$Q(\mu) = \sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + n\mu^2$$

Se calcularmos a derivada $dQ(\mu)/d\mu$, igualamos a 0 e resolvemos esta equação para μ encontraremos que $\mu = \bar{x}_n$. Segue-se então que o EMV de μ é $\hat{\mu} = \bar{X}_n$.

Pode ser visto acima que o valor de $\hat{\mu}$ não é afetado pelo valor da variância σ^2.

Exemplos de EMV

Amostragem de uma Normal(com variância conhecida)

Pode-se ver que $f_n(\mathbf{x}|\mu)$ será maximizado pelo valor de μ que maximiza

$$Q(\mu) = \sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + n\mu^2$$

Se calcularmos a derivada $dQ(\mu)/d\mu$, igualamos a 0 e resolvemos esta equação para μ encontraremos que $\mu = \bar{x}_n$.
Segue-se então que o EMV de μ é $\hat{\mu} = \bar{X}_n$.
Pode ser visto acima que o valor de $\hat{\mu}$ não é afetado pelo valor da variância σ^2.

Exemplos de EMV

Amostragem de uma Normal (com variância desconhecida)

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição Normal com média μ desconhecida e variância σ^2 desconhecida. Para valores observados $x_1, ..., x_n$, a função de verossimilhança é:

$$f_n(\mathbf{x}|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2] \tag{4}$$

Esta função será maximizada para possíveis valores de $-\infty < \mu < \infty$ e $\sigma^2 > 0$. Em lugar de maximizar a $f_n(\mathbf{x}|\mu, \sigma^2)$ é mais fácil maximizar $log f_n(\mathbf{x}|\mu, \sigma^2)$.

Exemplos de EMV

Amostragem de uma Normal (com variância desconhecida)

Temos:

$$L(\mu, \sigma^2) = log f_n(\mathbf{x}|\mu, \sigma^2) = -\frac{n}{2} log(2\pi) - \frac{n}{2} log\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \quad (5)$$

Devemos encontrar os valores de μ e de σ^2 para os quais $L(\mu, \sigma^2)$ é máximo encontrando os valores de μ e de σ^2 que satisfazem as equações:

$$\frac{\partial L(\mu, \sigma^2)}{\partial \mu} = 0 \quad (6)$$

$$\frac{\partial L(\mu, \sigma^2)}{\partial \sigma^2} = 0 \quad (7)$$

Exemplos de EMV

Amostragem de uma Normal (com variância desconhecida)

Da primeira equação obtemos:

$$\frac{\partial L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = \frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - n\mu)$$

encontramos aqui que $\mu = \bar{x}_i$. Também:

$$\frac{\partial L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$$

substituindo μ por $\bar{x}_n$ encontramos que

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \tag{8}$$

Exemplos de EMV

Amostragem de uma Normal (com variância desconhecida)

Observamos portanto que os EMV de μ e de σ^2 (que maximizam a função $L(\mu, \sigma^2)$) são

$$\hat{\mu} = \bar{X}_n$$

e

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

Exemplos de EMV
Amostragem de uma Uniforme

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição Uniforme no intervalo $(0, \theta)$ com θ desconhecido($\theta > 0$). A fdp de cada observação tem a forma:

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{para} \quad 0 \leq x \leq \theta \\ 0 & \text{c.c.} \end{cases} \tag{9}$$

Exemplos de EMV

Amostragem de uma Uniforme

portanto, a distribuição conjunta $f_n(\mathbf{x}|\theta)$ de $X_1, X_2, ..., X_n$ tem a forma:

$$f_n(\mathbf{x}|\theta) = \begin{cases} \frac{1}{\theta^n} & \text{para} \quad 0 \leq x_i \leq \theta \quad (i = 1, ..., n) \\ 0 & \text{c.c.} \end{cases} \tag{10}$$

Pode se mostrar que o EMV de θ deve ser um valor de θ para a qual $\theta \geq x_i$ para $i = 1, ..., n$ e maximiza $1/\theta^n$ entre tais valores. Como $1/\theta^n$ é uma função decrescente em θ, a estimativa será o menor valor de θ tal que $\theta \geq x_i$ para $(i = 1, ..., n)$. Como este valor é $\theta = max(x_1, ..., x_n)$ o EMV de θ é $\hat{\theta} = max(x_1, ..., x_n)$.

Exemplos de EMV

Não Existência de um EMV

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição Uniforme no intervalo $(0, \theta)$ com θ desconhecido($\theta > 0$). Suponha que em lugar de escrever a fdp como na equação 9 escrevemos da seguinte forma :

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{para} \quad 0 < x < \theta \\ 0 & \text{c.c.} \end{cases} \tag{11}$$

A única diferença das equações 9 e 11 é que na segunda a desigualdade é estrita. Ambas podem ser usadas como fdp da distribuição uniforme.

Exemplos de EMV

Não Existência de um EMV

portanto, se a equação 11 é usada, o EMV de θ será o valor de θ para a qual $\theta > x_i$ para $(i = 1, ..., n)$ e que maximiza $1/\theta^n$ entre todos os valores. Note que os possíveis valores de θ não incluem o valor de $\theta = max(x_1, ..., x_n)$ desde que θ deve ser estritamente maior que cada valor observado x_i $(i = 1, ..., n)$. Como θ pode ser escolhido como um valor arbitrário próximo de $max(x_1, ..., x_n)$ mas não igual a este valor, então segue-se que o EMV de θ não existe.

Este exemplo mostra uma dificuldade dos EMV, pois deveria ser irrelevante o uso das equações 9 ou da 11.

Exemplos de EMV

Não unicidade dos EMV

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição Uniforme no intervalo $(\theta, \theta + 1)$ com θ desconhecido $(-\infty < \theta < \infty)$. A fdp conjunta tem a forma:

$$f_n(\mathbf{x}|\theta) = \begin{cases} 1 & \text{para} \quad \theta \leq x_i \leq \theta + 1 (i = 1, ..., n) \\ 0 & \text{c.c.} \end{cases} \tag{12}$$

A condição $\theta \leq x_i$ para $(i = 1, ..., n)$ é equivalente à $\theta \leq min(x_1, ..., x_n)$. Similarmente, $x_i \leq \theta + 1$ para $(i = 1, ..., n)$ é equivalente a $\theta \geq max(x_1, ..., x_n) - 1$.

Exemplos de EMV

Não unicidade dos EMV

Escrevemos então $f_n(\mathbf{x}|\theta)$ da seguinte forma:

$$f_n(\mathbf{x}|\theta) = \begin{cases} 1 & \text{para} \quad max(x_1, ..., x_n) - 1 \leq x_i \leq min(x_1, ..., x_n) \\ 0 & \text{c.c.} \end{cases} \tag{13}$$

Desta forma, é possível selecionar como EMV de θ qualquer valor de θ no intervalo $max(x_1, ..., x_n) - 1 \leq \theta \leq min(x_1, ..., x_n)$.

Outline

1. Estimadores Maxima Verossimilhança
   - Estimadores de Máxima Verossimilhança
   - Casos dignos de nota
   - Propriedades dos Estimadores de Maxima Verossimilhança

Invariância

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição $f(x|\theta)$ com θ desconhecido, e seja $\hat{\theta}$ o EMV de θ. Para valores observados $x_1, ..., x_n$ a função de verossimilhança $f_n(\mathbf{x}|\theta)$ é maximizado quando $\hat{\theta} = \theta$.

Suponha que mudamos o parâmetro tal que expressaremos a função de probabilidade (densidade) em termos de um parâmetro $\tau = g(\theta)$, onde g é uma injeção de θ. Denotemos por $\theta = h(\tau)$ a função inversa. Então a função de distribuição de cada valor observado será $f[x|h(\tau)]$ e a função de verossimilhança será $f_n[\mathbf{x}|h(\tau)]$.

Invariância

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição $f(x|\theta)$ com θ desconhecido, e seja $\hat{\theta}$ o EMV de θ. Para valores observados $x_1, ..., x_n$ a função de verossimilhança $f_n(\mathbf{x}|\theta)$ é maximizado quando $\hat{\theta} = \theta$.
Suponha que mudamos o parâmetro tal que expressaremos a função de probabilidade (densidade) em termos de um parâmetro $\tau = g(\theta)$, onde g é uma injeção de θ. Denotemos por $\theta = h(\tau)$ a função inversa. Então a função de distribuição de cada valor observado será $f[x|h(\tau)]$ e a função de verossimilhança será $f_n[\mathbf{x}|h(\tau)]$.

Invariância

Suponha que $X_1, X_2, ..., X_n$ formam uma a.a. de uma distribuição $f(x|\theta)$ com θ desconhecido, e seja $\hat{\theta}$ o EMV de θ. Para valores observados $x_1, ..., x_n$ a função de verossimilhança $f_n(\mathbf{x}|\theta)$ é maximizado quando $\hat{\theta} = \theta$.
Suponha que mudamos o parâmetro tal que expressaremos a função de probabilidade (densidade) em termos de um parâmetro $\tau = g(\theta)$, onde g é uma injeção de θ. Denotemos por $\theta = h(\tau)$ a função inversa. Então a função de distribuição de cada valor observado será $f[x|h(\tau)]$ e a função de verossimilhança será $f_n[\mathbf{x}|h(\tau)]$.

Invariância

Como $f_n(\mathbf{x}|\theta)$ é maximizado quando $\theta = \hat{\theta}$, segue-se que $f_n[\mathbf{x}|h(\tau)]$ será maximizado quando $h(\tau) = \hat{\theta}$. Portanto, o EMV $\hat{\tau}$ deve satisfazer a relação $h(\hat{\tau}) = \hat{\theta}$ ou, $\hat{\tau} = g(\hat{\theta})$

Propriedade da invariância:

Se $\hat{\theta}$ é o EMV de θ, então $g(\hat{\theta})$ é o EMV de $g(\theta)$.

Invariância

Como $f_n(\mathbf{x}|\theta)$ é maximizado quando $\theta = \hat{\theta}$, segue-se que $f_n[\mathbf{x}|h(\tau)]$ será maximizado quando $h(\tau) = \hat{\theta}$. Portanto, o EMV $\hat{\tau}$ deve satisfazer a relação $h(\hat{\tau}) = \hat{\theta}$ ou, $\hat{\tau} = g(\hat{\theta})$

Propriedade da invariância:

Se $\hat{\theta}$ é o EMV de θ, então $g(\hat{\theta})$ é o EMV de $g(\theta)$.

Consistência

Em muitos problemas, a sequência de EMV converge em probabilidade ao valor desconhecido de θ quando $n \to \infty$.

Foi visto que a sequência de estimadores de Bayes para θ é uma sequência consistente de estimadores, portanto, para uma priori dada e uma amostra suficientemente grande, o estimador de Bayes e o EMV de θ serão valores muito proximos um do outro e proximos também do valor desconhecido de θ.

No exemplo da a.a. extraida de uma Bernoulli, foi mostrado que se a priori de θ é uma Beta, então a diferença entre o estimador de Bayes de θ e a média amostral $\bar{X}_n$ converge a 0 quando $n \to \infty$.

Desta forma, as sequências de estimadores de Bayes e a de EMV, são sequências consistentes.

Consistência

Em muitos problemas, a sequência de EMV converge em probabilidade ao valor desconhecido de θ quando $n \to \infty$.
Foi visto que a sequência de estimadores de Bayes para θ é uma sequência consistente de estimadores, portanto, para uma priori dada e uma amostra suficientemente grande, o estimador de Bayes e o EMV de θ serão valores muito proximos um do outro e proximos também do valor desconhecido de θ.
No exemplo da a.a. extraida de uma Bernoulli, foi mostrado que se a priori de θ é uma Beta, então a diferença entre o estimador de Bayes de θ e a média amostral $\bar{X}_n$ converge a 0 quando $n \to \infty$.
Desta forma, as sequências de estimadores de Bayes e a de EMV, são sequências consistentes.

Consistência

Em muitos problemas, a sequência de EMV converge em probabilidade ao valor desconhecido de θ quando $n \to \infty$.

Foi visto que a sequência de estimadores de Bayes para θ é uma sequência consistente de estimadores, portanto, para uma priori dada e uma amostra suficientemente grande, o estimador de Bayes e o EMV de θ serão valores muito proximos um do outro e proximos também do valor desconhecido de θ.

No exemplo da a.a. extraida de uma Bernoulli, foi mostrado que se a priori de θ é uma Beta, então a diferença entre o estimador de Bayes de θ e a média amostral $\bar{X}_n$ converge a 0 quando $n \to \infty$.

Desta forma, as sequências de estimadores de Bayes e a de EMV, são sequências consistentes.

Dependência com os planos amostrais

Se o experimentador decide fixar o valor de n antes de tomar as observações ou preferir usar algum plano amostral, pode ser mostrado que a função de verossimilhança $L(\theta)$ baseado nos valores observados será:

$$L(\theta) = f(x_1|\theta)...f(x_n|\theta)$$

Segue então que o EMV de θ será o mesmo, independentemente do plano amostral usado. Em outras palavras, o valor de θ depende apenas dos valores observados $x_1, ..., x_n$ e não do plano que o experimentador decidiu usar.