

Material Auxiliar - Unidade 2
Convergência Estocástica e Resultados Limite
Explicações Detalhadas e Didáticas

Curso de Inferência Estatística

Outubro 2025

Sumário

1	Introdução	2
2	Notação $O(\cdot)$ e $o(\cdot)$ - Big O e Little o	2
2.1	Motivação e Intuição	2
2.2	Definições Formais	2
2.3	Propriedades Importantes	2
2.4	Aplicação em Séries de Taylor	3
3	Convergência em Probabilidade	3
3.1	Intuição e Definição	3
3.2	Interpretação Prática	3
3.3	Métodos para Provar Convergência em Probabilidade	3
3.4	Propriedades Algébricas	4
4	Convergência em Distribuição	4
4.1	Definição e Diferenças	4
4.2	Diferenças entre Convergências	4
4.3	Método da Função Geradora de Momentos	5
5	Lei Fraca dos Grandes Números	5
5.1	Versões e Interpretação	5
5.2	Interpretação Prática	5
5.3	Aplicações	5
6	Teorema Central do Limite	6
6.1	Enunciado e Importância	6
6.2	Por Que é Tão Importante?	6
6.3	Interpretação Geométrica	6
6.4	Versões Padronizadas	6
7	Teorema de Slutsky	7
7.1	Enunciado e Utilidade	7
7.2	Por Que é Útil?	7
7.3	Aplicação em Testes de Hipóteses	7

8	Teorema de Mann-Wald (Método Delta)	7
8.1	Enunciado	7
8.2	Interpretação	8
8.3	Ideia da Prova	8
8.4	Aplicações Práticas	8
9	Resumo e Conexões	8
9.1	Hierarquia das Convergências	8
9.2	Teoremas Principais e Suas Relações	9
9.3	Estratégia de Resolução de Problemas	9

1 Introdução

Este material auxiliar complementa as notas de aula da Unidade 2, fornecendo explicações mais detalhadas e didáticas dos principais conceitos abordados. O objetivo é facilitar a compreensão dos teoremas de convergência e suas aplicações práticas.

2 Notação $O(\cdot)$ e $o(\cdot)$ - Big O e Little o

2.1 Motivação e Intuição

A notação $O(\cdot)$ e $o(\cdot)$ é fundamental para descrever o comportamento assintótico de sequências e funções. Intuitivamente:

- $a_n = O(b_n)$: " a_n cresce *no máximo* tão rápido quanto b_n "
- $a_n = o(b_n)$: " a_n cresce *mais devagar* que b_n "

2.2 Definições Formais

Definição 2.1 (Big O para sequências). Sejam $\{a_n, n \geq 1\}$ e $\{b_n, n \geq 1\}$ sequências de números reais. Dizemos que

$$a_n = O(b_n) \quad \text{se e somente se} \quad \exists k > 0, n_0 \in \mathbb{N} : \left| \frac{a_n}{b_n} \right| \leq k, \quad \forall n \geq n_0$$

Isto é, a razão $|a_n/b_n|$ é limitada para n suficientemente grande.

Definição 2.2 (Little o para sequências). Dizemos que

$$a_n = o(b_n) \quad \text{se e somente se} \quad \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$$

Isto é, a_n é desprezível comparado a b_n quando n é grande.

Exemplo 2.1 (Comparações Comuns). 1. $10n^2 + n = O(n^2)$ porque $\frac{10n^2 + n}{n^2} = 10 + \frac{1}{n} \leq 11$ para $n \geq 1$

2. $n = o(n^2)$ porque $\lim_{n \rightarrow \infty} \frac{n}{n^2} = \lim_{n \rightarrow \infty} \frac{1}{n} = 0$

3. $\log(n) = o(n)$ porque $\lim_{n \rightarrow \infty} \frac{\log(n)}{n} = 0$

4. $n^{1/2} = O(n)$ mas $n \neq O(n^{1/2})$

2.3 Propriedades Importantes

Observação 2.1 (Álgebra de O e o). 1. Se $a_n = o(b_n)$, então $a_n = O(b_n)$ (mas a recíproca é falsa)

2. Se $a_n = O(b_n)$ e $c_n = O(d_n)$, então:

- $a_n \cdot c_n = O(b_n \cdot d_n)$
- $a_n + c_n = O(\max\{|b_n|, |d_n|\})$

3. $O(1)$ significa limitado: $|a_n| \leq k$ para algum $k > 0$ e n grande

4. $o(1)$ significa que $a_n \rightarrow 0$

2.4 Aplicação em Séries de Taylor

A notação $O(\cdot)$ é essencial para expressar aproximações via série de Taylor:

Exemplo 2.2 (Série de Taylor). Para uma função $F(x)$ derivável até ordem n em torno de x_0 :

$$F(x) = \sum_{k=0}^n \frac{F^{(k)}(x_0)}{k!} (x - x_0)^k + o((x - x_0)^n)$$

quando $x \rightarrow x_0$.

Por exemplo:

- $e^x = 1 + x + \frac{x^2}{2} + O(x^3)$ quando $x \rightarrow 0$
- $\log(1 + x) = x - \frac{x^2}{2} + O(x^3)$ quando $x \rightarrow 0$

3 Convergência em Probabilidade

3.1 Intuição e Definição

A convergência em probabilidade expressa a ideia de que, à medida que n cresce, a probabilidade de U_n estar "longe" de u torna-se arbitrariamente pequena.

Definição 3.1 (Convergência em Probabilidade). Uma sequência de variáveis aleatórias $\{U_n, n \geq 1\}$ converge em probabilidade para um número u se

$$P(|U_n - u| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0, \quad \forall \varepsilon > 0$$

Notação: $U_n \xrightarrow{P} u$

3.2 Interpretação Prática

Pense em U_n como uma estimativa de u baseada em n observações. Convergência em probabilidade significa que:

- Com n grande, é *altamente improvável* que U_n esteja longe de u
- Para qualquer margem de erro $\varepsilon > 0$ que você escolha, a probabilidade de erro pode ser tornada arbitrariamente pequena aumentando n

Exemplo 3.1 (Média Amostral). Se X_1, X_2, \dots são v.a.'s i.i.d. com $E[X_i] = \mu$ e $\text{Var}(X_i) = \sigma^2 < \infty$, então

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu$$

Isso significa que a média amostral converge para a média populacional.

3.3 Métodos para Provar Convergência em Probabilidade

1. **Desigualdade de Chebyshev:** Se $E[U_n] \rightarrow u$ e $\text{Var}(U_n) \rightarrow 0$, então $U_n \xrightarrow{P} u$
2. **Convergência de momentos:** Se $E[|U_n - u|^r] \rightarrow 0$ para algum $r > 0$, então $U_n \xrightarrow{P} u$
3. **Função geradora de momentos:** Se $M_{U_n}(t) \rightarrow e^{tu}$ para todo t , então $U_n \xrightarrow{P} u$

3.4 Propriedades Algébricas

Observação 3.1 (Álgebra da Convergência em Probabilidade). Se $U_n \xrightarrow{P} u$ e $V_n \xrightarrow{P} v$, então:

1. $U_n + V_n \xrightarrow{P} u + v$
2. $U_n \cdot V_n \xrightarrow{P} u \cdot v$
3. $\frac{U_n}{V_n} \xrightarrow{P} \frac{u}{v}$ (se $P(V_n = 0) = 0$ e $v \neq 0$)
4. Se $g(\cdot)$ é contínua, então $g(U_n) \xrightarrow{P} g(u)$

4 Convergência em Distribuição

4.1 Definição e Diferenças

A convergência em distribuição é um conceito mais fraco que convergência em probabilidade.

Definição 4.1 (Convergência em Distribuição). Uma sequência $\{U_n, n \geq 1\}$ com f.d.a. $F_n(u)$ converge em distribuição para uma v.a. U com f.d.a. $F(u)$ se

$$F_n(u) \xrightarrow{n \rightarrow \infty} F(u)$$

em todos os pontos de continuidade de $F(\cdot)$.

Notação: $U_n \xrightarrow{D} U$

4.2 Diferenças entre Convergências

- **Convergência em Probabilidade \Rightarrow Convergência em Distribuição**
- **Convergência em Distribuição \nRightarrow Convergência em Probabilidade** (em geral)
- **Exceção:** Se $U_n \xrightarrow{D} c$ (constante), então $U_n \xrightarrow{P} c$

Exemplo 4.1 (Distinção Importante). Considere $X \sim N(0, 1)$ e defina $U_n = X$ para todo n . Então:

- $U_n \xrightarrow{D} X$ (trivialmente, pois $F_n = F$ para todo n)
- $U_n \not\xrightarrow{P} X$ (não faz sentido: $U_n - X = 0$ sempre!)

Agora considere $U_n = (-1)^n X$:

- $U_n \xrightarrow{D} X$ (ambos têm distribuição $N(0, 1)$)
- $U_n \not\xrightarrow{P} X$ (pois $|U_n - X|$ não vai para zero)

4.3 Método da Função Geradora de Momentos

Um método poderoso para provar convergência em distribuição:

Observação 4.1 (Teorema de Continuidade de Lévy). *Se $M_{U_n}(t) \rightarrow M_U(t)$ para todo t em uma vizinhança de zero, então $U_n \xrightarrow{D} U$.*

Este método é frequentemente usado nas provas do TCL.

5 Lei Fraca dos Grandes Números

5.1 Versões e Interpretação

A Lei Fraca dos Grandes Números (LFGN) é um dos resultados fundamentais da probabilidade.

Observação 5.1 (LFGN - Versão Simples). *Se X_1, \dots, X_n são v.a.'s i.i.d. com $E[X_i] = \mu < \infty$ e $\text{Var}(X_i) = \sigma^2 < \infty$, então*

$$\bar{X}_n \xrightarrow{P} \mu$$

Observação 5.2 (LFGN de Khinchin). *A condição de variância finita pode ser relaxada: basta $E[X_i] = \mu < \infty$.*

5.2 Interpretação Prática

- A média amostral é um estimador *consistente* da média populacional
- Quanto maior a amostra, mais confiável é a estimativa
- Justifica a "Lei dos Grandes Números" empírica: frequências relativas convergem para probabilidades

Exemplo 5.1 (Lançamento de Moedas). Se $X_i = 1$ (cara) ou $X_i = 0$ (coroa) com $P(X_i = 1) = p$, então

$$\frac{\text{número de caras em } n \text{ lançamentos}}{n} = \bar{X}_n \xrightarrow{P} p$$

5.3 Aplicações

1. **Estimação de parâmetros:** \bar{X}_n estima μ , S_n^2 estima σ^2
2. **Simulação Monte Carlo:** Aproximar $E[g(X)]$ por $\frac{1}{n} \sum_{i=1}^n g(X_i)$
3. **Testes de hipóteses:** Proporções amostrais convergem para proporções populacionais

6 Teorema Central do Limite

6.1 Enunciado e Importância

O Teorema Central do Limite (TCL) é possivelmente o teorema mais importante da estatística.

Observação 6.1 (TCL - Versão Clássica). *Se X_1, \dots, X_n são v.a.'s i.i.d. com $E[X_i] = \mu$ e $\text{Var}(X_i) = \sigma^2 < \infty$, então*

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0, 1)$$

6.2 Por Que é Tão Importante?

1. **Universalidade:** Funciona para *qualquer* distribuição com variância finita
2. **Base para inferência:** Justifica o uso da distribuição normal em intervalos de confiança e testes
3. **Aproximação prática:** Com n moderadamente grande ($n \geq 30$), \bar{X}_n tem distribuição aproximadamente normal

6.3 Interpretação Geométrica

O TCL diz que:

- A distribuição de \bar{X}_n fica mais concentrada em torno de μ (taxa $1/\sqrt{n}$)
- A forma da distribuição de $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ converge para a curva normal
- Não importa a distribuição original dos X_i !

Exemplo 6.1 (Distribuição Uniforme). Se $X_i \sim U(0, 1)$ (distribuição uniforme), então $E[X_i] = 1/2$ e $\text{Var}(X_i) = 1/12$.

$$\frac{\bar{X}_n - 1/2}{\sqrt{1/(12n)}} = \sqrt{12n} \left(\bar{X}_n - \frac{1}{2} \right) \xrightarrow{D} N(0, 1)$$

Embora X_i seja uniforme (nada parecido com normal), \bar{X}_n tem distribuição aproximadamente $N(1/2, 1/(12n))$ para n grande.

6.4 Versões Padronizadas

- **σ conhecido:** $Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0, 1)$
- **σ desconhecido:** $T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{D} N(0, 1)$
(onde S_n é o desvio padrão amostral)

7 Teorema de Slutsky

7.1 Enunciado e Utilidade

O Teorema de Slutsky permite combinar convergências de tipos diferentes.

Observação 7.1 (Teorema de Slutsky). *Se $U_n \xrightarrow{D} U$ e $V_n \xrightarrow{P} c$ (constante), então:*

1. $U_n + V_n \xrightarrow{D} U + c$
2. $U_n \cdot V_n \xrightarrow{D} c \cdot U$
3. $\frac{U_n}{V_n} \xrightarrow{D} \frac{U}{c}$ (se $c \neq 0$)

7.2 Por Que é Útil?

O teorema de Slutsky é crucial quando:

- Temos uma convergência em distribuição mas precisamos fazer operações algébricas
- Queremos substituir parâmetros desconhecidos por estimadores consistentes
- Provamos distribuições assintóticas de estatísticas de teste

Exemplo 7.1 (Substituição do Desvio Padrão). Pelo TCL: $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0, 1)$

Como $S_n \xrightarrow{P} \sigma$, pelo Slutsky:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \cdot \frac{\sigma}{S_n} \xrightarrow{D} N(0, 1) \cdot 1 = N(0, 1)$$

Isso justifica usar S_n quando σ é desconhecido!

7.3 Aplicação em Testes de Hipóteses

O teorema de Slutsky permite construir estatísticas de teste quando parâmetros são desconhecidos, substituindo-os por estimadores consistentes sem alterar a distribuição assintótica.

8 Teorema de Mann-Wald (Método Delta)

8.1 Enunciado

O Método Delta é uma ferramenta para encontrar a distribuição assintótica de transformações de estimadores.

Observação 8.1 (Teorema de Mann-Wald). *Se $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2(\theta))$ e $g(\cdot)$ é uma função diferenciável com $g'(\theta) \neq 0$, então*

$$\sqrt{n}[g(T_n) - g(\theta)] \xrightarrow{D} N(0, \sigma^2(\theta) \cdot [g'(\theta)]^2)$$

8.2 Interpretação

O método delta diz que:

- Se T_n é aproximadamente normal com taxa $1/\sqrt{n}$
- Então $g(T_n)$ também é aproximadamente normal com taxa $1/\sqrt{n}$
- A variância é "inflada" por $[g'(\theta)]^2$

8.3 Ideia da Prova

A prova usa aproximação de Taylor de primeira ordem:

$$g(T_n) \approx g(\theta) + g'(\theta)(T_n - \theta)$$

Multiplicando por \sqrt{n} :

$$\sqrt{n}[g(T_n) - g(\theta)] \approx g'(\theta) \cdot \sqrt{n}(T_n - \theta)$$

Como $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2)$, o resultado segue.

Exemplo 8.1 (Transformação Logarítmica). Se \bar{X}_n estima $\mu > 0$ e queremos estimar $\log(\mu)$, tome $g(x) = \log(x)$.

Como $g'(x) = 1/x$, temos:

$$\sqrt{n} [\log(\bar{X}_n) - \log(\mu)] \xrightarrow{D} N\left(0, \frac{\sigma^2}{\mu^2}\right)$$

Exemplo 8.2 (Transformação de Variância). Para estimar a variância σ^2 , usamos S_n^2 . Se queremos estimar o desvio padrão $\sigma = \sqrt{\sigma^2}$, usamos $g(x) = \sqrt{x}$ com $g'(x) = \frac{1}{2\sqrt{x}}$.

8.4 Aplicações Práticas

1. Transformações estabilizadoras de variância
2. Intervalos de confiança para funções de parâmetros
3. Testes de hipóteses sobre transformações
4. Modelos não-lineares

9 Teorema Central do Limite para Variância Amostral

9.1 Motivação

Enquanto o TCL clássico trata da distribuição assintótica de \bar{X}_n , é natural perguntar: qual a distribuição assintótica de S_n^2 (a variância amostral)?

Observação 9.1 (TCL para S_n^2). Se X_1, \dots, X_n são v.a.'s i.i.d. com $E[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$, e $\mu_4 = E[(X_i - \mu)^4] < \infty$, então

$$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow[n \rightarrow \infty]{d} N(0, \mu_4 - \sigma^4)$$

9.2 Interpretação

- A variância assintótica é $\mu_4 - \sigma^4$, que depende do quarto momento central
- Para distribuições simétricas, μ_4 mede o "peso nas caudas"
- Distribuições com caudas pesadas têm μ_4 maior, logo maior variabilidade em S_n^2

9.3 Comparação com Normalidade

Para $X_i \sim N(\mu, \sigma^2)$:

- $\mu_4 = 3\sigma^4$, logo a variância assintótica é $3\sigma^4 - \sigma^4 = 2\sigma^4$
- Isto coincide com a variância exata de $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$

9.4 Ideia da Prova

A prova combina o TCL clássico com o Teorema de Slutsky:

1. Defina $W_n = \frac{n-1}{n}S_n^2$ e $Y_i = (X_i - \mu)^2$
2. Mostre que $W_n = \bar{Y}_n - (\bar{X}_n - \mu)^2$
3. Aplique TCL a \bar{Y}_n : $\sqrt{n}(\bar{Y}_n - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4)$
4. Note que $\sqrt{n}(\bar{X}_n - \mu)^2 \xrightarrow{P} 0$ (é de ordem $O_P(1/\sqrt{n})$)
5. Use Slutsky para concluir sobre W_n , depois relate a S_n^2

9.5 Aplicação Prática

Este teorema permite construir intervalos de confiança assintóticos para σ^2 :

$$IC_{1-\alpha}(\sigma^2) = S_n^2 \pm z_{\alpha/2} \cdot \frac{\sqrt{\mu_4 - \sigma^4}}{\sqrt{n}}$$

onde μ_4 pode ser estimado por $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^4$.

10 Estimadores Consistentes

10.1 Definição e Intuição

Consistência é uma propriedade fundamental de estimadores que garante convergência para o parâmetro verdadeiro.

Definição 10.1 (Estimador Consistente). Um estimador $T_n = T_n(X_1, \dots, X_n)$ de $\tau(\theta)$ é **consistente** (no sentido fraco) se

$$T_n \xrightarrow[n \rightarrow \infty]{P} \tau(\theta), \quad \forall \theta \in \Theta$$

10.2 Interpretação Prática

Um estimador consistente significa que:

- Com amostras grandes, a probabilidade de erro significativo torna-se desprezível
- É um requisito mínimo para que um estimador seja considerado "bom"
- Diferente de não-viesamento (propriedade de amostra finita), consistência é assintótica

10.3 Relação entre Viés, Variância e Consistência

Observação 10.1 (Consistência via EQM). Se $EQM_\theta[T_n] = E_\theta[(T_n - \tau(\theta))^2] \rightarrow 0$, então T_n é consistente.

Como $EQM = Viés^2 + Variância$, temos consistência quando:

$$B_\theta^2[T_n] \rightarrow 0 \quad e \quad Var_\theta[T_n] \rightarrow 0$$

Exemplo 10.1 (Estimadores Clássicos). 1. \bar{X}_n é consistente para μ (pela LFGN)

2. $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ é consistente para σ^2

3. O EMV é geralmente consistente sob condições de regularidade

10.4 Exemplo Detalhado: Máximo da Uniforme

Exemplo 10.2 (Consistência de $X_{n:n}$ para $U(0, \theta)$). Se $X_1, \dots, X_n \sim U(0, \theta)$, o EMV de θ é $\hat{\theta} = X_{n:n} = \max\{X_1, \dots, X_n\}$.

A f.d.a. de $X_{n:n}$ é:

$$F_{X_{n:n}}(t) = \begin{cases} 0, & t < 0 \\ (t/\theta)^n, & 0 \leq t \leq \theta \\ 1, & t > \theta \end{cases}$$

Para $\varepsilon > 0$ com $\varepsilon < \theta$:

$$P(|X_{n:n} - \theta| < \varepsilon) = P(\theta - \varepsilon < X_{n:n} < \theta) \tag{1}$$

$$= 1 - (\theta - \varepsilon)^n / \theta^n \tag{2}$$

$$= 1 - (1 - \varepsilon/\theta)^n \rightarrow 1 \tag{3}$$

quando $n \rightarrow \infty$.

Logo, $X_{n:n} \xrightarrow{P} \theta$.

10.5 Tamanho Amostral Mínimo

Um aspecto prático interessante: dado $\varepsilon > 0$ e $\delta \in (0, 1)$, qual o tamanho amostral mínimo n_0 para garantir

$$P(|X_{n:n} - \theta| < \varepsilon) \geq 1 - \delta?$$

Solução: De $1 - (1 - \varepsilon/\theta)^n \geq 1 - \delta$, obtemos

$$n \geq \frac{\log \delta}{\log(1 - \varepsilon/\theta)}$$

Assim:

$$n_0 = \left\lceil \frac{\log \delta}{\log(1 - \varepsilon/\theta)} \right\rceil + 1$$

Exemplo 10.3 (Cálculo Numérico). Para $\theta = 1$, $\varepsilon = 0.1$, $\delta = 0.05$:

$$n_0 = \left\lceil \frac{\log(0.05)}{\log(0.9)} \right\rceil + 1 = \left\lceil \frac{-2.996}{-0.105} \right\rceil + 1 = 29$$

Com 29 observações, temos 95% de chance de $X_{n:n}$ estar a menos de 0.1 de $\theta = 1$.

11 Propriedades Assintóticas dos Estimadores de Máxima Verossimilhança

11.1 Introdução

Os EMVs possuem propriedades assintóticas excepcionais que justificam sua popularidade na prática estatística.

11.2 Eficiência Relativa Assintótica

Definição 11.1 (Eficiência Relativa Assintótica). Se dois estimadores $T_n^{(1)}$ e $T_n^{(2)}$ para $g(\theta)$ são assintoticamente normais:

$$\sqrt{n}(T_n^{(i)} - g(\theta)) \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_i^2(\theta)), \quad i = 1, 2$$

a eficiência relativa assintótica (ERA) de $T_n^{(2)}$ em relação a $T_n^{(1)}$ é:

$$\text{ERA} = \frac{\sigma_1^2(\theta)}{\sigma_2^2(\theta)}$$

11.3 Interpretação

- ERA > 1: $T_n^{(2)}$ é mais eficiente (menor variância assintótica)
- ERA = 1: Ambos são igualmente eficientes
- ERA < 1: $T_n^{(1)}$ é mais eficiente

11.4 Teorema Central do Limite para EMVs

Observação 11.1 (TCL para EMVs). *Sob condições de regularidade, se $\hat{\theta}_n$ é o EMV de θ , então:*

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{d} N(0, I_X^{-1}(\theta))$$

onde $I_X(\theta)$ é a informação de Fisher:

$$I_X(\theta) = E_\theta \left[\left(\frac{\partial \log f(X; \theta)}{\partial \theta} \right)^2 \right]$$

11.5 Condições de Regularidade

As condições necessárias incluem:

1. **(A1) Diferenciabilidade:** $f(x; \theta)$ é três vezes diferenciável em θ
2. **(A2) Troca de derivação e integração:** É válido trocar $\frac{\partial}{\partial \theta}$ com \int
3. **(A3) Informação de Fisher finita:** $0 < I_X(\theta) < \infty$
4. **(A4) Dominação local:** A terceira derivada é dominada por função integrável
5. **(A5) Existência de solução consistente:** A equação de verossimilhança tem solução $\hat{\theta}_n \xrightarrow{P} \theta$

11.6 Propriedades dos EMVs

Sob as condições de regularidade, os EMVs são:

1. **Consistentes:** $\hat{\theta}_n \xrightarrow{P} \theta$
2. **Assintoticamente não-viesados:** $\lim_{n \rightarrow \infty} E[\hat{\theta}_n] = \theta$
3. **Assintoticamente normais:** A distribuição converge para normal
4. **Assintoticamente eficientes:** Atingem o limite inferior de Cramér-Rao assintótico

11.7 Limite Inferior de Cramér-Rao Assintótico

Para qualquer estimador não-viesado T_n de θ :

$$\text{Var}(T_n) \geq \frac{1}{n \cdot I_X(\theta)}$$

Os EMVs atingem este limite assintoticamente!

11.8 Aplicação Prática

Exemplo 11.1 (Intervalo de Confiança via EMV). Se $\hat{\theta}_n$ é o EMV, um IC assintótico de nível $1 - \alpha$ para θ é:

$$IC_{1-\alpha}(\theta) = \hat{\theta}_n \pm z_{\alpha/2} \cdot \frac{1}{\sqrt{n \cdot I_X(\hat{\theta}_n)}}$$

onde $I_X(\hat{\theta}_n)$ é a informação de Fisher avaliada em $\hat{\theta}_n$.

Exemplo 11.2 (Distribuição Normal). Para $X_i \sim N(\mu, \sigma^2)$ com σ^2 conhecido:

- EMV: $\hat{\mu}_n = \bar{X}_n$
- Informação de Fisher: $I_X(\mu) = 1/\sigma^2$
- Distribuição assintótica: $\sqrt{n}(\bar{X}_n - \mu) \sim N(0, \sigma^2)$

Neste caso, a distribuição assintótica coincide com a exata!

11.9 Vantagens e Limitações

Vantagens:

- Propriedades ótimas assintoticamente
- Princípio unificado para construir estimadores
- Aproximação normal facilita inferência

Limitações:

- Requer condições de regularidade
- Propriedades são assintóticas (podem não valer para n pequeno)
- Computação pode ser complexa (requer otimização numérica)

12 Resumo e Conexões

12.1 Hierarquia das Convergências

Convergência quase certa \Rightarrow Convergência em Probabilidade \Rightarrow Convergência em Distribuição

12.2 Teoremas Principais e Suas Relações

1. **LFGN:** $\bar{X}_n \xrightarrow{P} \mu$ (onde está o valor)
2. **TCL clássico:** $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0, 1)$ (quão rápido chega lá e qual a forma da distribuição)
3. **TCL para S_n^2 :** $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4)$ (distribuição assintótica da variância amostral)
4. **Slutsky:** Permite combinações algébricas de convergências diferentes
5. **Método Delta:** Estende para transformações não-lineares
6. **Consistência:** Propriedade fundamental de estimadores ($T_n \xrightarrow{P} \theta$)
7. **TCL para EMVs:** Propriedades assintóticas ótimas dos estimadores de máxima verossimilhança

12.3 Estratégia de Resolução de Problemas

1. **Identificar o tipo de problema:**
 - Convergência pontual? \rightarrow Usar LFGN ou consistência
 - Distribuição assintótica? \rightarrow Usar TCL
 - Variância/segunda ordem? \rightarrow TCL para S_n^2
2. **Verificar condições:**

- Variáveis i.i.d.?
- Momentos necessários existem?
- Condições de regularidade satisfeitas?

3. Aplicar teoremas apropriados:

- Para médias: LFGN ou TCL clássico
- Para variâncias: TCL para S_n^2
- Para EMVs: TCL para EMVs

4. Lidar com complicações:

- Parâmetros desconhecidos? \rightarrow Slutsky
- Transformações não-lineares? \rightarrow Método Delta
- Múltiplas convergências? \rightarrow Álgebra de convergências

5. Construir inferência:

- Intervalos de confiança assintóticos
- Testes de hipóteses assintóticos
- Regiões de confiança

12.4 Quadro Sinótico: Quando Usar Cada Teorema

Objetivo	Teorema	Condições
Estimador consistente?	LFGN ou EQM $\rightarrow 0$	Momentos finitos
Distribuição de X_n ?	TCL clássico	$E[X_i^2] < \infty$
Distribuição de S_n^2 ?	TCL para S_n^2	$E[X_i^4] < \infty$
σ desconhecido?	Slutsky	$S_n \xrightarrow{P} \sigma$
Função $g(\bar{X}_n)$?	Método Delta	$g'(\mu) \neq 0$
EMV?	TCL para EMVs	Regularidade

12.5 Mensagem Final

Este material auxiliar complementa as notas de aula, fornecendo:

- Explicações detalhadas dos conceitos principais
- Interpretações práticas dos resultados teóricos
- Exemplos numéricos e aplicações
- Estratégias para resolução de problemas

Os teoremas de convergência formam a base da inferência estatística moderna. Compreendê-los profundamente permite:

- Justificar procedimentos estatísticos comuns
- Desenvolver novos métodos para problemas específicos

- Avaliar propriedades de estimadores e testes
- Construir aproximações úteis para cálculos práticos

Estude com atenção, pratique muito, e boa sorte!