Material Auxiliar - Unidade 2

Convergência Estocástica e Resultados Limite

Explicações Detalhadas e Didáticas

Curso de Inferência Estatística

Outubro 2025

Sumário

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(·) e o(·) - 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

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

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 $\lim_{n\to\infty} \frac{n}{n^2} = \lim_{n\to\infty} \frac{1}{n} = 0$
3. $\log(n) = o(n)$ porque $\lim_{n\to\infty} \frac{\log(n)}{n} = 0$
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 $a_n \to 0$*

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:

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

quando $x \to x_0$.

Por exemplo:

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

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

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

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 $X_1, X_2, \ldots$ 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 $E[U_n] \to u$ e $\text{Var}(U_n) \to 0$, então $U_n \xrightarrow{P} u$
2. **Convergência de momentos:** Se $E[|U_n - u|^r] \to 0$ para algum $r > 0$, então $U_n \xrightarrow{P} u$
3. **Função geradora de momentos:** Se $M_{U_n}(t) \to e^{tu}$ 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

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

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** $\not\Rightarrow$ **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 $M_{U_n}(t) \to M_U(t)$ 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 $Var(X_i) = \sigma^2 < \infty$, 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 $X_1, \ldots, X_n$ são v.a.'s i.i.d. com $E[X_i] = \mu$ e $Var(X_i) = \sigma^2 < \infty$, 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 $\mathrm{Var}(X_i) = 1/12$.

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

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*

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

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:

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

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 Resumo e Conexões

9.1 Hierarquia das Convergências

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

9.2 Teoremas Principais e Suas Relações

1. **LFGN:** $\bar{X}_n \xrightarrow{P} \mu$ (onde está o valor)
2. **TCL:** $\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \xrightarrow{D} N(0,1)$ (quão rápido chega lá e qual a forma da distribuição)
3. **Slutsky:** Permite combinações algébricas
4. **Método Delta:** Estende para transformações não-lineares

9.3 Estratégia de Resolução de Problemas

1. Identificar se T_n é uma média ou função de médias
2. Aplicar LFGN ou TCL conforme apropriado
3. Se há parâmetros desconhecidos, usar Slutsky para substituí-los
4. Se há transformação não-linear, usar Método Delta
5. Verificar condições (i.i.d., momentos finitos, etc.)