Introdução aos Processos Estocásticos - Teoremas

Eduardo M. A. M. Mendes

DELT - UFMG
Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal de Minas Gerais
emmendes@cpdee.ufmg.br

Motivação

1) Lei dos Grandes Números - Justifica a interpretação de "frequência relativa" relativa da probabilidade.
   Exemplo: Lança-se uma moeda 1.000.000 vezes e conta-se o número de caras.

$$\frac{\text{número de caras}}{1.000.000} = 0,4999$$

   Logo

$$\begin{aligned} P[\text{caras}] &= 0,5 \\ &\approx \text{frequência relativa} \\ &= \text{frequência relativa quando } N \to \infty \end{aligned}$$

2) Teorema do Limite Central - justifica a suposição de distribuição gaussiana dos resultados.
Exemplo: Uma pessoa escolhida aleatoreamente $\rightarrow$ medida do peso.

$$P = \text{Peso Genético} + \text{Stress no Trabalho} + \text{Dieta} + \text{Educação} + \dots$$

ou seja, uma variedade de fatores que podem influenciar.

- Em princípio temos a necessidade de ter a modelagem das várias VAs.

O TLC diz

$$X_1 + X_2 + \cdots + X_N \sim \text{Gaussiana quando } N \rightarrow \infty$$

para X_i IID.

Leis dos Grandes Números

Definição: Se um evento de probabilidade p é observado repetidamente em ocasiões independentes, a proporção da freqüência observada deste evento em relação ao número total de repetições converge em direção a p à medida que o número de repetições se torna arbitrariamente grande.

- Lei Fraca dos Grandes Números - Dada uma variável aleatória X, a sua média amostral converge em probabilidade para o seu valor esperado.
- Lei Forte dos Grandes Números - Dada uma variável aleatória X, a sua média amostral converge quase certamente para o seu valor esperado.

Obs.: Se a variável aleatória não tem média, a Lei dos Grandes Números não se aplica.

Leis dos Grandes Números (cont.)

- Considere o lançamento de uma moeda não viciada várias vezes. Os resultados são

$$X_1, X_2, X_3, \ldots, X_N, \ldots$$

$$\begin{aligned} X_i &= 1 \quad \text{se cara} \\ &= 0 \quad \text{se coroa} \end{aligned}$$

- Assuma que X_i é IID, ou seja:
  1) Independentes - o lançamento de uma moeda não "depende" do lançamento anterior, ou seja, $P(X_i|X_{i-1}) = \frac{P(X_i \cap X_{i-1})}{P(X_{i-1})} = P(X_i)$.
  2) Identicamente Distribuídos - A mesma moeda é usada e lançada de uma mesma maneira toda vez. A lei de distribuição é mantida.

Leis dos Grandes Números (cont.)

O resultado aleatório do experimento pode ser modelado por

$$X = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \end{bmatrix}^T$$

e a PMF

$$p_{X_i}[k] = p_X[k] = \begin{cases} \frac{1}{2} & k = 0 \\ \frac{1}{2} & k = 1 \end{cases}$$

- Seja

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

para $N = 2 \rightarrow \bar{X} = \frac{1}{2}(X_1 + X_2)$. $\bar{X} = 0, \frac{1}{2}, 1, \ldots$ possui uma PMF.

Leis dos Grandes Números (cont.)

Calculando o valor esperado da nova variável aleatória (N qualquer), temos:

$$
\begin{aligned}
E_{\underline{X}}[\bar{X}] &= E_{\underline{X}}\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} E_{\underline{X}} X_i \\
&= \frac{1}{N}\sum_{i=1}^{N} E_{X_i} X_i \\
&= \frac{1}{N}\sum_{i=1}^{N} \left(0 \times \frac{1}{2} + 1\frac{1}{2}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{1}{2} = \frac{1}{2}
\end{aligned}
$$

Leis dos Grandes Números (cont.)

$$
\begin{aligned}
Var(\bar{X}) &= Var\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} Var(\underbrace{X_i}_{\text{Independentes}})
\end{aligned}
$$

mas

$$
\begin{aligned}
Var(X_i) &= E_{X_i}(X_i^2) - E_{X_i}^2(X_i) \\
&= 0^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2} - \left(\frac{1}{2}\right)^2 \\
&= \frac{1}{2} - \frac{1}{4} = \frac{1}{4}
\end{aligned}
$$

logo

$$Var(\bar{X}) = \frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{4} = \frac{\frac{1}{4}}{N}$$

quando $N \to \infty \to Var(\bar{X}) = 0$. Lei dos Grandes Números - Veja probprob15.1.r.

- Seja $\bar{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$

$$\bar{X}_N \to \frac{1}{2} = E_X[X] \quad \text{quando } N \to \infty$$

Em geral, a Lei dos Grandes Números diz que, para VAs IID,

$$\frac{1}{N}\sum_{i=1}^{N} X_i \to E_X[X] \quad \text{quando } N \to \infty$$

Leis dos Grandes Números (cont.)

e

$$\hat{E}(g(X)) = \frac{1}{N}\sum_{i=1}^{N} g(X_i)$$

Teorema: Se $X_1, X_2, \ldots, X_N$ são IID com média $E_X(x)$ e variância σ^2 então

$$\begin{aligned} \lim_{N\to\infty} \bar{X}_N &= E_X(x) \\ &\text{ou} \\ \lim_{N\to\infty} P\left[|\bar{X}_N - E_X(x)| > \epsilon\right] &= 0 \end{aligned}$$

para qualquer $\epsilon > 0$ pequeno, ou seja, convergência em probabilidade

Prova 1: Considere

$$P\left[|\bar{X}_N - E_X(x)| > \epsilon\right] = P\left[|\underbrace{\bar{X}_N}_{Y} - \underbrace{E_{\bar{X}}(\bar{X}_N)}_{E(Y)}| > \epsilon\right]$$

Usando a desigualdade de Chebyshev

$$P\left[|Y - E(Y)| > \epsilon\right] \leq \frac{Var(Y)}{\epsilon^2}$$

mas $Var(Y) = Var(\bar{X}_N) = \frac{\sigma^2}{N}$.
Logo

$$\begin{aligned} P\left[|\bar{X}_N - E_X(x)| > \epsilon\right] &\leq \frac{\frac{\sigma^2}{N}}{\epsilon^2} = \frac{\sigma^2}{N\epsilon^2} \\ \lim_{N\to 0} P\left[|\bar{X}_N - E_X(x)| > \epsilon\right] &\leq \lim_{N\to\infty} \frac{\sigma^2}{N\epsilon^2} = 0 \end{aligned}$$

Leis dos Grandes Números (cont.)

Prova 2: Usando a expansão de séries de Taylor para funções complexas, sabemos que a função característica de uma variável aleatória X com média finita pode ser escrita como

$$\phi_X(\omega) = 1 + i\omega E(X) + o(\omega), \quad t \to 0$$

Repare que todas as variáveis X_1, X_2, X_3, ... têm a mesma função característica. Usando propriedades da função característica podemos escrever

$$\phi_{\frac{1}{N}X}(\omega) = \phi_X\left(\frac{\omega}{N}\right) \quad \text{e} \quad \phi_{X+Y}(\omega) = \phi_X(\omega)\phi_Y(\omega),$$

se X e Y independentes, temos

$$\phi_{\bar{X}}(\omega) = \left[\phi_X\left(\frac{\omega}{N}\right)\right]^N = \left[1 + iE(X)\frac{\omega}{N} + o\left(\frac{\omega}{N}\right)\right]^N \to e^{i\omega E(X)},$$

quando $\quad N \to \infty$.

Leis dos Grandes Números (cont.)

Repare que o limite $e^{i\omega E(X)}$ é a função característica de uma variável aleatória constante $E(X)$ e assim

$$\bar{X} \xrightarrow{D} E(X) \quad \text{para } n \to \infty$$

Mas como a variável é constante, a convergência em distribuição é equivalente à convergência em probabilidade, logo

$$\bar{X} \xrightarrow{P} E(X)$$

Exemplo: Determinar se um sinal $S = A$ (constante) está presente em meio à contaminação por ruído.

$$X_i = s + w_i = A + w_i \quad i = 1, 2, \ldots$$

para W sendo IID com $E_w(w) = 0$ e $Var(W) = \sigma^2 < \infty$.

$$\bar{X}_N \to E_X(x) = A \quad \text{quando } N \to \infty$$

Leis dos Grandes Números (cont.)

Quando não há sinal

$$\bar{X}_N \rightarrow E_X(x) = E_W(w) = 0$$

Teorema do Limite Central

A Lei dos Grandes Números dá informação sobre a largura e localização da PDF/PMF de $\bar{X}_N$.

$$\begin{aligned} \text{largura} &\rightarrow 0 \\ \text{localização} &\rightarrow E_X(x) \end{aligned}$$

E sobre a PDF quando $N \rightarrow \infty$?

Exemplo:

$$X_i \quad \text{é IID e } X_i \sim U\left(-\frac{1}{2}, \frac{1}{2}\right)$$

Teorema do Limite Central (cont.)

com $E(x) = \frac{a+b}{2}$ e $Var(x) = \frac{(b-a)^2}{12}$. A PDF de X_1 e X_2 é obtida pela convolução da PDF de X_1 com a PDF de X_2

$$
\begin{aligned}
p_{S_2} &= p_X(x) * p_X(x) \\
p_{S_3} &= p_X(x) * p_X(x) * p_X(x)
\end{aligned}
$$

Podemos, então, calcular

$$
\begin{aligned}
E(S_3) &= E\left(\sum_{i=1}^{3} X_i\right) = 0 \quad X_i \sim U\left(-\frac{1}{2}, \frac{1}{2}\right) \\
Var(S_3) &= Var\left(\sum_{i=1}^{3} X_i\right) = 3 Var(X_i) = 3\frac{1}{12}
\end{aligned}
$$

que é uma boa aproximação para a Gaussiana (repare que a média é zero).

Teorema do Limite Central (cont.)

Exemplo: Quando $E(X) \neq 0$. Considere

$$S_N = \sum_{i=1}^{N} X_i \quad \sim U(0,1)$$

logo

$$\begin{aligned} E(X) &= \frac{a+b}{2} = \frac{1}{2} \neq 0 \\ Var(X) &= \frac{(b-a)^2}{12} = \frac{1}{12} \end{aligned}$$

Teorema do Limite Central (cont.)

O valor esperado é

$$
\begin{aligned}
E(S_N) = E\left(\sum_{i=1}^{N} X_i\right) & \\
&= \sum_{i=1}^{N} E(X_i) = NE(X) \\
&= N \times \frac{1}{2} = \frac{N}{2} \quad \rightarrow \text{“anda” com o valor de } N. \\
Var(S_N) &= NVar(X) = \frac{N}{12}
\end{aligned}
$$

Repare que

$$N \uparrow \rightarrow \quad E(S_N) \uparrow \text{ e } Var(S_N) \uparrow$$

É necessário normalizar, ou seja, $E(\bullet) = 0$ e $Var(\bullet) = 1$

Teorema do Limite Central (cont.)

$$
\begin{aligned}
Z_N &= \frac{S_N - E(S_N)}{\sqrt{Var(S_N)}} \\
&= \frac{S_N - NE_X(X)}{\sqrt{NVar(X)}}
\end{aligned}
$$

A PDF da soma normalizada de um grande número de VAs contínuas IID convergirá para uma PDF Gaussiana.

$$
N \to \infty \quad \to \quad Z_N \sim N(0,1)
$$

Teorema: Se $X_1, X_2, \ldots, X_N$ são VAs contínuas IID com média $E_X(X)$ e variância $Var(X)$ então para $N \to \infty$.

$$
\frac{\sum_{i=1}^{N} X_i - NE_X(X)}{\sqrt{NVar(X)}} \to N(0,1)
$$

Teorema do Limite Central (cont.)

Exemplo: Dado $X_i \sim N(0,1)$, examine PDF de $Y = \sum_{i=1}^{N} X_i^2$ quando $N \to \infty$. A verdadeira PDF é $Y \sim \chi_N^2$.

Para aplicar TLC precisamos verificar

a) Independência - $X_1, X_2, \ldots, X_N$ são independentes logo $X_1^2, X_2^2, \ldots, X_N^2$ são independentes.

b) Identicamente distribuidos - $X_1, X_2, \ldots, X_N$ têm a mesma PDF logo $X_1^2, X_2^2, \ldots, X_N^2$ têm a mesma PDF.

Teorema do Limite Central (cont.)

$$\frac{\sum_{i=1}^{N} X_i^2 - NE(X^2)}{\sqrt{NVar(X^2)}} = ?$$

com

$$X^2 \sim \chi_1^2 \rightarrow \begin{cases} E_X(X^2) & = & 1 \\ Var(X^2) & = & 2 \end{cases}$$

Normalizando temos

$$Z_N = \frac{\sum_{i=1}^{N} X_i^2 - N}{\sqrt{2N}} \rightarrow N(0,1)$$

e

$$\underbrace{Z_N}_{N(0,1)} \sqrt{2N} + N = \sum_{i=1}^{N} X_i^2$$

Teorema do Limite Central (cont.)

Logo

$$\sum_{i=1}^{N} X_i^2 \sim N(N, 2N)$$

Esboço da Prova do TLC

Seja $Z_N = \frac{S_N - NE_X(X)}{\sqrt{NVar(X)}}$. Para $Z_N \to N(0,1)$, vamos usar a função característica

$$\phi_{Z_N}(\omega) \to \phi_Z(\omega) = e^{-\frac{1}{2}\omega^2}$$

Portanto

$$\begin{aligned}
\phi_{Z_N}(\omega) &= E_{Z_N}\left[e^{\jmath\omega Z_N}\right] \\
&= E_X\left[e^{\jmath\omega \frac{\sum_{i=1}^{N} X_i - NE_X(X_i)}{\sqrt{NVar(X_i)}}}\right] \\
&= \prod_{i=1}^{N} E_{X_i}\left[e^{\jmath\omega \frac{X_i - E_X(X_i)}{\sqrt{NVar(X_i)}}}\right] \\
&= \left[E_X\left[e^{\jmath\omega \frac{X - E_X(X)}{\sqrt{NVar(X)}}}\right]\right]^N
\end{aligned}$$

Esboço da Prova do TLC (cont.)

Olhando para $E_X\left[e^{\jmath\omega\frac{X-E_X(X)}{\sqrt{NVar(X)}}}\right]$, temos

$$
\begin{aligned}
E_X\left[e^{\jmath\omega\frac{X-E_X(X)}{\sqrt{NVar(X)}}}\right] \underbrace{=}_{\text{séries}} & E_X\left[\sum_{k=0}^{\infty}\frac{(\jmath\omega)^k}{k!}\left(\frac{X-E(X)}{\sqrt{NVar(X)}}\right)^k\right] \\
= & \sum_{k=0}^{\infty}\frac{(\jmath\omega)^k}{k!}E_X\left[\left(\frac{X-E(X)}{\sqrt{NVar(X)}}\right)^k\right] \\
= & 1+\jmath\omega E_X\left[\frac{X-E(X)}{\sqrt{NVar(X)}}\right]+ \\
& \frac{1}{2}(\jmath\omega)^2 E_X\left[\left(\frac{X-E(X)}{\sqrt{NVar(X)}}\right)^2\right]+\dots
\end{aligned}
$$

Esboço da Prova do TLC (cont.)

Mas

$$
\begin{aligned}
E_X\left[\frac{X-E(X)}{\sqrt{NVar(X)}}\right] &= 0\\
E_X\left[\left(\frac{X-E(X)}{\sqrt{NVar(X)}}\right)^2\right] &= \frac{E_X\left[(X-E(X))^2\right]}{NVar(X)} = \frac{1}{N}
\end{aligned}
$$

Desconsiderando os termos de alta-ordem

$$
\begin{aligned}
\phi_{Z_N}(\omega) &= \left(1+\frac{1}{2}(\jmath\omega)^2\frac{1}{N}\right)^N\\
&= \left(1-\frac{1}{2}\frac{\omega^2}{N}\right)^N \rightarrow e^{-\frac{1}{2}\omega^2} \text{ quando } N\rightarrow\infty
\end{aligned}
$$

$= \phi_Z(\omega)$, $Z \sim N(0,1)$.

Observações sobre o TLC

O resultado do TLC pode ser escrito como

$$\lim_{N\to\infty} P\left(u_1 \leq \frac{x - NE(X)}{\sqrt{NVar(X)}} \leq u_2\right) = \int_{u_1}^{u_2} \frac{1}{\sqrt{2\pi}} e^{-x^2} dx,$$

para todo u_1 e u_2 finitos.

- Note que, para N finito, a distribuição da soma $X = X_1 + X_2 + \ldots + X_N$ pode ser bem diferente da Gaussiana no que se refere às caudas. Entretanto o peso dessas regiões não-gaussianas tendem a zero quando N tende a infinito.
- O TLC se preocupa mais com a região central que tem um peso finito para N grande.

As principais hipóteses que asseguram a validade do TLC Gaussiano são:

Observações sobre o TLC (cont.)

- Os X_i têm que ser variáveis aleatórias independentes, ou pelo menos não muito correlacionadas (a função de correlação deve ter um decaimento suficiente rápido quando $|i - j|$ se torna grande).
- As variáveis aleatórias X_i não precisam ser necessariamente *identicamente distribuídas*. O que deve acontecer é que a variância de todas essas distribuições não sejam muito diferentes tal que não haja dominância de uma destas variâncias sobre as outras.

$$X = a_1 X_1 + a_2 X_2 + \ldots + a_N X_N$$

onde a_i são coeficientes arbitrários.

- Formalmente o TLC só é aplicado quando N tende a infinito. Na prática N é finito e deve ser grande suficiente para que a parte central da distribuição seja parecida com a Gaussiana. O valor mínimo de N para que isso aconteça depende da distribuição de X_i, sua distância para a Gaussiana e de quanto a Gaussiana pode aproximar as caudas.

Observações sobre o TLC (cont.)

- o TLC não diz nada sobre as caudas da distribuição de X mas somente que a região central da distribuição poder ser bem descrita por uma Gaussiana. A região central é uma região com pelo menos $\sqrt{N}\sigma$ em torno da média de X. A largura da região que pode ser bem aproximada pela Gaussiana depende da distribuição de X.

TLC e Teoria da Informação

A quantidade $\mathcal{I}$ denominada Entropia (quantidade de informação faltante ou perdida) associada à função de distribuição de probabilidade P é definida como

$$\mathcal{I}(P) = -\int P(x)log(P(x))dx$$

A distribuição que maximiza $\mathcal{I}(P)$ para um dado valor de variância é obtida tomando a derivada funcional com respeito a $P(x)$

$$\frac{\partial}{\partial P(x)}\left[\mathcal{I}(P) - \xi\int x'^2 P(x')dx' - \xi'\int P(x')dx'\right] = 0$$

onde ξ é fixado pela condição $\int x^2 p(x)dx = \sigma^2$ e ξ' pela normalização de $P(x)$. A solução da igualdade acima é a Gaussiana.

TLC e Teoria da Informação (cont.)

- O valor númerico da Entropia para a Gaussiana é:

$$\mathcal{I}_G = \frac{1}{2} + \frac{1}{2}\log(2\pi) + \log(\sigma) \approx 1,419 + \log(\sigma)$$

- O valor númerico da Entropia para a Exponencial é:

$$\mathcal{I}_E = 1 + \frac{\log 2}{2} + \log(\sigma) \approx 1,346 + \log(\sigma)$$

Observe que a operação de convolução (soma de variáveis) é uma operação de queima de informação, pois todos os detalhes da distribuição elementar vão sendo perdidos até que a Gaussiana surja.

A Gaussiana é a lei da máxima entropia ou mínima informação.