MAT 461 – Tópicos de Matemática II
Aula 8: Resumo de Probabilidade

Edson de Faria

Departamento de Matemática
IME-USP

28 de Agosto, 2013

- Nesta aula apresentaremos as desigualdades de Markov e Chebyshev, que possuem várias aplicações em Probabilidade (algumas das mais simples serão apresentadas através de exemplos) .
- Apresentaremos também dois dos resultados mais importantes em Probabilidade: o Teorema Central do Limite e a Lei dos Grandes Números.

Desigualdade de Markov

- Considere uma variável aleatória X com esperança $E[X]$ finita. Suponha que X assume apenas valores não-negativos, isto é, $X \geq 0$.
- Suponha que desejamos calcular $P[X \geq \delta]$, onde $\delta > 0$ é dado.
- Em geral, precisamos de mais informações para calcular o valor exato dessa probabilidade.
- A desigualdade de Markov é uma ferramenta que nos permite estimar a probabilidade acima sem informações adicionais sobre a distribuição de X.

Proposição (Desigualdade de Markov)

Seja X uma variável aleatória satisfazendo $X \geq 0$ e $E[X] < \infty$. Então, para todo $\delta > 0$, temos:

$$P[X \geq \delta] \leq \frac{E[X]}{\delta}$$

- Podemos verificar este fato sem dificuldade. Suponhamos que X é discreta, assumindo valores $x_1, x_2, \ldots$ (não-negativos).
- Então, pela definição de esperança:

$$E[X] \;=\; \sum_{x_i} x_i\, P[X = x_i] \;\geq\; \sum_{x_i \geq \delta} x_i\, P[X = x_i]\;.$$

- Portanto:

$$E[X] \;\geq\; \delta \sum_{x_i \geq \delta} P[X = x_i] \;=\; \delta\, P[X \geq \delta]\;,$$

e disto se segue imediatamente a desigualdade de Markov.
- A verificação é inteiramente análoga para variáveis aleatórias contínuas.

Exemplo 1

Exemplo (1)

Uma fábrica de televisores produz em média 150 *unidades por dia. O que podemos dizer sobre a probabilidade de que a fábrica produza, num dado dia,* 280 *ou mais unidades?*

Solução:

- Denotemos por X a quantidade de televisores produzidos pela fábrica num dado dia. Sabemos que $X \geq 0$, e que $E[X] = 150$.
- Não sabemos calcular o valor exato da probabilidade solicitada: isto requer conhecermos em detalhes a distribuição de probabilidade de X.
- Podemos, no entanto, calcular um limitante superior para tal probabilidade, mediante a desigualdade de Markov.
- Basta aplicarmos a desigualdade com $\delta = 280$:

$$P[X \geq 180] \leq \frac{150}{280} \simeq 0.536 .$$

Uma conseqüência importante da desigualdade de Markov é a desigualdade de Chebyshev.

Proposição (Desigualdade de Chebyshev)

Seja X uma variável aleatória com esperança $E[X] = \mu$ e variância $Var(X) = \sigma^2$. Então, para todo $\delta > 0$ temos:

$$P[|X - \mu| \geq \delta] \ \leq \ \frac{\sigma^2}{\delta^2}$$

- A verificação desta desigualdade não é difícil. Considere a variável aleatória $Y = (X - \mu)^2$.
- Note que Y é não-negativa e que a esperança de Y é igual à variância de X, pois:

$$E[Y] \ = \ E[(X - \mu)^2] \ = \ \mathrm{Var}(X) \ = \ \sigma^2 \ .$$

- Portanto, temos:

$$P[|X-\mu| \geq \delta] \;=\; P[|X-\mu|^2 \geq \delta^2] \;=\; P[Y \geq \delta^2] \;\leq\; \frac{\sigma^2}{\delta^2}\,,$$

onde, no último passo, aplicamos a desigualdade de Markov.

Exemplo (2)

No exemplo anterior, suponhamos que a variância do número de unidades fabricadas por dia é conhecida, digamos igual a 100. *Que podemos dizer sobre a probabilidade de que o número de unidades fabricadas num determinado dia fique entre* 120 *e* 180*?*

Solução:

- Como antes, denotamos por X o número de unidades produzidas num dia pela fábrica.
- Já sabíamos que $E[X] = 150$. Agora, temos a informação adicional de que $\mathrm{Var}(X) = 100$.
- Queremos estimar $P[120 < X < 180]$. Sabemos que

$$P[120 < X < 180] \;=\; 1 - P[|X - 150| \geq 30] \;.$$

- Para estimarmos esta última probabilidade, aplicamos a desigualdade de Chebyshev para X com $\mu = 150$, $\sigma^2 = 100$ e $\delta = 30$, obtendo:

$$P[|X - 150| \geq 30] \;\leq\; \frac{100}{30^2} = \frac{1}{9} \;.$$

- Portanto, temos:

$$P[120 < X < 180] \;\geq\; 1 - \frac{1}{9} \;=\; \frac{8}{9} \simeq\; 0.889 \;.$$

- Ou seja, podemos dizer que a probabilidade desejada é $\geq 88.9\%$.

- A desigualdade de Chebyshev é muitas vezes empregada para a análise de somas de variáveis aleatórias independentes e com mesma média e mesmo desvio padrão.
- Tal resultado pode ser enunciado como segue.

Proposição

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes e com mesma média e mesmo desvio padrão, digamos $E[X_i] = \mu$ e $Var(X_i) = \sigma^2$ $(i = 1, 2, \ldots, n)$. Considere a soma:

$$S_n \;=\; X_1 + X_2 + \cdots + X_n\ .$$

Então, para todo $\delta > 0$ temos:

$$P\left[\left|\frac{S_n}{n} - \mu\right| \geq \delta\right] \;\leq\; \frac{\sigma^2}{n\delta^2}$$

- A verificação da proposição acima é bastante simples. Escreva $X = \dfrac{S_n}{n}$.
- Em seguida, note que

$$E[X] \;=\; \frac{1}{n} E[S_n] \;=\; \frac{1}{n} \sum_{i=1}^{n} E[X_i] \;=\; \frac{1}{n} \cdot n\mu \;=\; \mu$$

bem como

$$\mathrm{Var}(X) \;=\; \frac{1}{n^2}\mathrm{Var}(S_n) \;=\; \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) \;=\; \frac{1}{n^2} \cdot n\sigma^2 \;=\; \frac{\sigma}{n}$$

- Assim, aplicando a desigualdade de Chebyshev para X com $E[X] = \mu$ e $\mathrm{Var}(X) = \sigma^2/n$, obtemos:

$$P[|X - \mu| \geq \delta] \;\leq\; \frac{(\sigma^2/n)}{\delta^2} \;=\; \frac{\sigma^2}{n\delta^2} \;,$$

e isso é exatamente o que queríamos mostrar.

Exemplo 3

Eis aqui uma aplicação típica deste último resultado.

Exemplo (3)

Uma pesquisa de rua será realizada em Votolândia para determinar a proporção p de votos que o candidato Mixuruca receberá nas próximas eleições para prefeito ($0 \leq p \leq 1$). Qual deve ser o número de pessoas entrevistadas para que estejamos pelo menos 95% seguros de que o valor de p tenha sido determinado com erro inferior a 0.05? Assuma que as decisões individuais de cada eleitor são independentes.

Solução:

- Digamos que o número de pessoas entrevistadas seja n.
- Denotemos por X_i ($i = 1, 2, \ldots, n$) a variável aleatória de Bernoulli que assume o valor 1, com probabilidade p, se a i-ésima pessoa entrevistada declara intenção de votar no candidato Mixuruca, e assume o valor 0, com probabilidade $1 - p$, caso contrário.

- Temos assim uma seqüência de variáveis de Bernoulli $X_1, X_2, \ldots, X_n$ com mesma esperança $\mu = E[X_i] = p$ e mesma variância $\sigma^2 = \mathrm{Var}(X_i) = p(1-p)$.
- A variável aleatória binomial $S_n = X_1 + X_2 + \cdots + X_n$ denota o número de pessoas, dentre as entrevistadas, que declaram seu voto ao candidato Mixuruca. Assim, S_n/n é uma aproximação do valor de p (que é desconhecido).
- Queremos que este valor aproximado – S_n/n – difira do valor real de p por no máximo 0.05, com probabilidade maior ou igual a 95%.
- Em outras palavras, queremos que n seja suficientemente grande para que:

$$P\left[\left|\frac{S_n}{n} - p\right| \le 0.05\right] \ \ge \ 0.95 \ .$$

- Como prosseguir? Chebyshev!

- Podemos aplicar a desigualdade de Chebyshev na forma dada pela última proposição com $\mu = p$, $\sigma^2 = p(1-p)$ e $\delta = 0.05$, obtendo:

$$P\left[\left|\frac{S_n}{n} - p\right| \geq 0.05\right] \leq \frac{p(1-p)}{n \cdot (0.05)^2} \leq \frac{1}{4n \cdot (0.05)^2}$$

  onde usamos o fato de que $p(1-p) \leq \frac{1}{4}$.
- A probabilidade do lado esquerdo dessa desigualdade é complementar à probabilidade desejada, e portanto deve ser inferior a $1 - 0.95 = 0.05$.
- Assim, basta que n seja grande o suficiente para que

$$\frac{1}{4n \cdot (0.05)^2} \leq 0.05 \ .$$

- Ou seja, basta tomar n tal que

$$n \geq \frac{1}{4 \cdot (0.05)^3} = 2000$$

- Em outras palavras, pelo menos 2000 pessoas devem ser entrevistadas.

- O assim chamado Teorema Central do Limite é, sem dúvida, o resultado mais importante da Teoria das Probabilidades.
- Este resultado lida com somas $S_n = X_1 + X_2 + \cdots + X_n$ de variáveis aleatórias independentes e identicamente distribuidas, e com as respectivas médias S_n/n.
- Intuitivamente, o Teorema Central do Limite afirma que, à medida em que n cresce, tais médias – normalizadas por um fator conveniente – se comportam como variáveis aleatórias normais.
- A primeira versão deste resultado foi obtida por P. Laplace e A. de Moivre no século XVIII, para somas (ou médias) de variáveis aleatórias binomiais.
- O resultado geral que temos em mente é o seguinte.

Teorema Central do Limite

Lembramos que a distribuição acumulada de uma variável aleatória normal padrão Z é dada pela primitiva da função Gaussiana:

$$P[Z \leq x] \;=\; \Phi(x) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt \;.$$

Teorema

Sejam $X_1, X_2, \ldots, X_i, \ldots$ variáveis aleatórias independentes e identicamente distribuidas, com $E[X_i] = \mu$ e $Var(X_i) = \sigma^2$ (para todo $i = 1, 2, \ldots$). Para cada $n = 1, 2, \ldots$, seja $S_n = X_1 + X_2 + \cdots + X_n$. Então, para quaisquer a e b reais com $a < b$, temos:

$$P\left[a < \frac{S_n - n\mu}{\sigma\sqrt{n}} < b\right] \xrightarrow{n\to\infty} \Phi(b) - \Phi(a) \;=\; \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-t^2/2}\, dt$$

- Observe que, se escrevemos:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} ,$$

então cada uma das variáveis aleatórias Z_n tem média 0, pois $E[S_n] = n\mu$, e desvio padrão 1, pois $\text{Var}(S_n) = n\sigma^2$.

- Ou seja, as variáveis aleatórias Z_n estão normalizadas.
- Assim, o que o TCL está dizendo é que, à medida em que n cresce, tais variáveis normalizadas "se aproximam" da variável normal padrão Z, num certo sentido probabilístico.
- Este fato está na raiz de inúmeras aplicações. De fato, boa parte da ciência empírica chamada Estatística tem como pilar justamente o TCL!

Exemplo 4 (= Exemplo 3 revisitado)

Como primeira ilustração do uso do TCL, vejamos novamente o exemplo da eleição considerado anteriormente.

Exemplo (4)

Uma pesquisa de rua será realizada em Votolândia para determinar a proporção p de votos que o candidato Mixuruca receberá nas próximas eleições para prefeito ($0 \leq p \leq 1$). Qual deve ser o número de pessoas entrevistadas para que estejamos pelo menos 95% seguros de que o valor de p tenha sido determinado com erro inferior a 0.05? Assuma que as decisões individuais de cada eleitor são independentes.

- Como antes, modelamos o comportamento da i-ésima pessoa que responde à pesquisa por uma variável aleatória de Bernoulli X_i com

$$X_i = \begin{cases} 1 & \text{se } i \text{ vota em Mixuruca} \\ 0 & \text{caso contrário} \end{cases}$$

e $P[X_i = 1] = p$, $P[X_i = 0] = q = 1 - p$.
- Lembre-se que $\mu = E[X_i] = p$, e que $\sigma^2 = \mathrm{Var}(X_i) = pq$, para todo $i = 1, 2, \ldots, n$.
- A proporção de votos recebidos por Mixuruca na amostra de n pessoas pesquisadas é, também como antes:

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- Assim, trata-se de determinar n mínimo de tal modo que

$$P\left[\left|\frac{S_n}{n} - p\right| \leq 0.05\right] \geq 0.95 \ .$$

- Mas temos:

$$P\left[\left|\frac{S_n}{n}-p\right|\leq 0.05\right] \;=\; P\left[-0.05\leq\frac{S_n-np}{n}\leq 0.05\right]$$

$$=\; P\left[-0.05\sqrt{\frac{n}{pq}}\leq\frac{S_n-np}{\sqrt{npq}}\leq 0.05\sqrt{\frac{n}{pq}}\right]$$

- Portanto, queremos determinar n mínimo tal que:

$$P\left[-0.05\sqrt{\frac{n}{pq}}\leq\frac{S_n-np}{\sqrt{npq}}\leq 0.05\sqrt{\frac{n}{pq}}\right] \;\geq\; 0.95$$

- Pelo TCL, o lado esquerdo desta desigualdade é aproximadamente igual a:

$$\Phi\left(0.05\sqrt{\frac{n}{pq}}\right)-\Phi\left(-0.05\sqrt{\frac{n}{pq}}\right)$$

desde que n seja suficientemente grande.

- Por outro lado, perguntamos: qual é o valor de x tal que

$$\Phi(x) - \Phi(-x) \;=\; 0.95 \quad ?$$

- Consultando uma tabela com os valores da acumulada da normal padrão, obtemos $x \simeq 1.96$.
- Assim, basta escolhermos n de tal modo que

$$0.05\sqrt{\frac{n}{pq}} \;\geq\; 1.96 \;.$$

ou seja, tal que

$$n \;\geq\; \left(\frac{1.96}{0.05}\right)^2 pq \;.$$

- Como $pq \leq \dfrac{1}{4}$, basta portanto escolhermos n de tal modo que:

$$n \;\geq\; \frac{1}{4}\left(\frac{1.96}{0.05}\right)^2 \;\simeq\; 384.16 \;.$$

- O menor n que verifica esta última desigualdade é $n = 385$

Exemplo 5

Um outro tipo de aplicação simples é o seguinte.

Exemplo (5)

Uma moeda honesta é lançada 100 *vezes sucessivamente. Qual é a probabilidade de que ocorram entre* 55 *e* 65 *caras?*

Solução:

- Temos aqui a repetição de 100 experimentos de Bernoulli com probabilidade de sucesso $\frac{1}{2}$ ($X_i = 1$ se o resultado do i-ésimo lançamento resultar em cara, $X_i = 0$ se resultar em coroa, $i = 1, 2, \ldots, 100$).
- Assim, o número total de caras após 100 lançamentos é $S = X_1 + X_2 + \cdots + X_{100}$.
- Esta é uma variável binomial com parâmetros $n = 100$ e $p = \frac{1}{2}$.

- Logo, o valor exato da probabilidade solicitada é:

$$P[55 \leq S \leq 65] = \sum_{k=55}^{65} \binom{100}{k} \left(\frac{1}{2}\right)^{k} \left(\frac{1}{2}\right)^{100-k}$$

$$= \left(\frac{1}{2}\right)^{100} \sum_{k=55}^{65} \binom{100}{k} .$$

- O valor exato desta última expressão pode em tese ser obtido, mas de forma um pouco laboriosa laboriosa. É preferível utilizar o TCL: este é o ponto deste exemplo.
- Sabemos que $E[S] = 100 \times \frac{1}{2} = 50$ e que

$$\mathrm{Var}(S) = 100\mathrm{Var}(X_1) = 100 \times \frac{1}{4} = 25 .$$

- Assim, o desvio padrão de S é $\sigma = \sqrt{25} = 5$.

- O ponto agora é que a variável aleatória normalizada:

$$S^* = \frac{S - 50}{5}$$

é, pelo TCL, aproximadamente uma normal padrão.

- Em outras palavras, temos:

$$P\left[55 \leq S \leq 65\right] = P\left[\frac{55-50}{5} \leq \frac{S-50}{5} \leq \frac{65-50}{5}\right]$$

$$= P\left[1 \leq S^* \leq 3\right] \simeq \Phi(3) - \Phi(1) \simeq 0.157 \ .$$

- Ou seja, a probabilidade desejada é aproximadamente igual a 15.7%.