Variáveis Aleatórias Contínuas e Distribuição de Probabilidades - parte III

Marcos Oliveira Prates

2012/02

Objetivos

Ao final deste capítulo você deve ser capaz de:

- Calcular probabilidades aproximadas usando a distribuição normal.
- Aplicar essa aproximação par o caso de variáveis com as seguintes distribuições:
  - binomial;
  - hipergeométrica;
  - Poisson.

- Vimos na aula anterior que a normal pode aproximar a distribuição de diversas variáveis aleatórias.
- Esse resultado é conhecido como **Teorema Central do Limite**.

Teorema Central do Limite

- Seja X uma variável aleatória com média μ e desvio padrão σ.
- Retiramos uma amostra de tamanho n de X.
- Seja $\bar{X}$ a média dessa amostra.
- Temos que

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \approx N(0,1)$$

quando o n aumenta.

Exemplo:

- Seja X a duração das conversar telefônicas em uma cidade.
- Não sabemos a distribuição de X, mas sabemos que

$$E(X) = 3 \quad Var(X) = 9\,.$$

- Coletamos uma amostra de 50 chamadas.
- Qual a probabilidade de que a média dessa amostra não ultrapasse 4 minutos?

$$P\left(\bar{X} < 4\right) = P\left(\frac{\bar{X} - 3}{\sqrt{9/50}} < \frac{4-3}{\sqrt{9/50}}\right) \approx P(Z < 2,36) = 0,9909.$$

Aproximação da Normal para Binomial

- Em muitos sistemas físicos o modelo binomial é o mais apropriado.
- Em alguns casos o tamanho da amostra n fica muito grande.
- Então fica difícil calcular probabilidades envolvendo a binomial.
- Vimos que uma binomial é uma soma de Bernoullis.
- Portanto uma Binomial pode ser aproximada por uma Normal.
- Desde que o n seja suficientemente grande.

- A área de cada barra é a probabilidade binomial de x.
- As áreas das barras podem ser aproximadas pelas áreas abaixo da função densidade da normal.

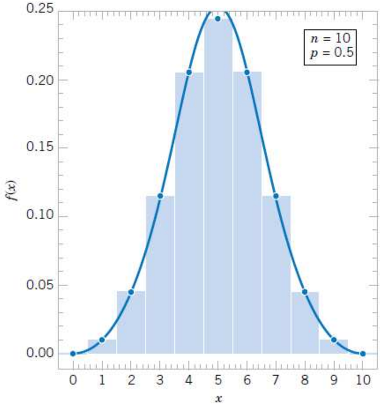
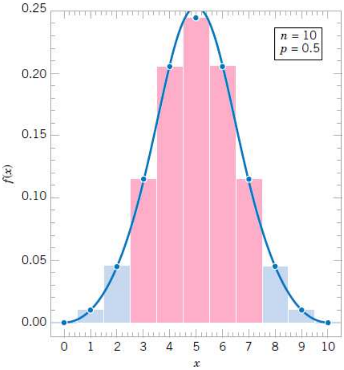
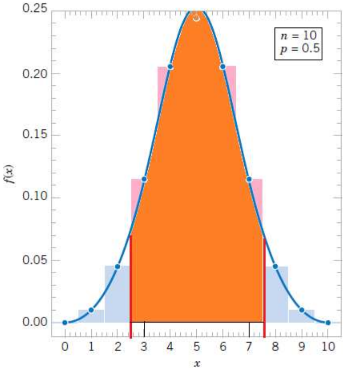


Figure 4-19 Normal approximation to the binomial distribution.

- Se X tem distribuição Binomial(10; 0,5).
- Se queremos calcular $P(3 \leq X \leq 7)$.





- Isso é chamado **correção de continuidade**.

- A correção de continuidade é necessária pois estamos aproximando uma variável discreta por uma contínua.
- Por exemplo:

$$P(X \leq 4) = P(X < 4,5)$$

$$P(X < 4) = P(X < 3,5)$$

$$P(X > 4) = P(X > 4,5)$$

$$P(X \geq 4) = P(X > 3,5)$$

(desenhar no quadro)

Exemplo:

- Considere um canal digital de comunicação.
- O número de **bits** transmitidos com erro podem ser modelados por uma Binomial.
- A probabilidade de um *bit* ser transmitido com erro é 1×10^{-5}.
- 16 milhões de *bits* são transmitidos.
- Qual a probabilidade de ter 150 ou menos erros?

Exemplo: (solução)

- Seja X o número de *bits* transmitidos com erro.
- X tem distribuição Binomial com $n = 16$ milhões e $p = 1 \times 10^{-5}$.
- A probabilidade requerida é

$$P(X \leq 150) = \sum_{x=0}^{150} \binom{16.000.000}{x} (10^{-5})^x (1-10^{-5})^{16.000.000-x}$$

- Essa probabilidade é muito difícil de se calcular.
- Podemos usar a aproximação pela normal.
- Vejamos como isso é feito.

Aproximação da Binomial pela Normal

- Seja X uma variável com distribuição Binomial(n,p).
- A variável
$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$
será aproximadamente uma variável normal padrão.
- Precisamos aplicar a **correção de continuidade**.
- Temos que
$$P(X \leq x) = P(X \leq x + 0,5) \approx P\left(Z \leq \frac{x + 0,5 - np}{\sqrt{np(1-p)}}\right)$$
$$P(X \geq x) = P(X \geq x - 0,5) \approx P\left(Z \geq \frac{x - 0,5 - np}{\sqrt{np(1-p)}}\right)$$
- A aproximação é boa para $np > 5$ e $n(1-p) > 5$.

- Lembre que para uma variável Binomial X

$$E(X) = np \qquad Var(X) = np(1-p)\,.$$

- A expressão

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

é uma fórmula para padronizar X.

- As probabilidades envolvendo X podem ser aproximadas usando a normal padrão.
- A aproximação é boa quando n é grande relativo a p.

Exemplo:

- Considere o exemplo da comunicação digital.
- X o número de *bits* transmitidos com erro.
- X tem distribuição Binomial com $n = 16$ milhões e $p = 1 \times 10^{-5}$.
- Podemos calcular probabilidade de $X \leq 150$ da seguinte forma

$$P(X \leq 150) = P(X \leq 150,5)$$
$$= P\left(\frac{X - 160}{\sqrt{160(1 - 10^{-5})}} \leq \frac{150,5 - 160}{\sqrt{160(1 - 10^{-5})}}\right)$$
$$P(Z \leq -0,75) = 0,227\,.$$

- Observe que

$$np = (16 \times 10^{6})(1^{-5}) = 160 \quad \text{e} \quad n(1-p) > 160\,.$$

- A aproximação deve ser boa nesse caso.

Exemplo:

- Considere o exemplo da transmissão de *bits*.
- 50 *bits* são transmitidos.
- A probabilidade de que ocorra um erro é $p = 0,1$.
- Qual a probabilidade de que 2 ou menos erros ocorram?

Exemplo: (solução)

- Usando a distribuição binomial

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$\binom{50}{0}(0,9)^{50} + \binom{50}{1}(0,9)^{49}(0,1) + \binom{50}{2}(0,9)^{48}(0,1)^2 = 0,112.$$

- Usando a distribuição Normal

$$P(X \leq 2) = P\left(\frac{X-5}{\sqrt{50(0,1)(0,9)}} \leq \frac{2,5-5}{\sqrt{50(0,1)(0,9)}}\right)$$

$$= P(Z \leq -1,18) = 0,119\,.$$

- Mesmo a amostra não sendo muito grande o resultado aproximado é bom.

Exemplo:

- Queremos agora calcular a probabilidade de $X \geq 9$

$$P(X \geq 9) = P(X \geq 8,5) \approx P\left(Z \geq \frac{8,5-5}{2,12}\right)$$

$$= P(Z \geq 1,65) = 0,05\,.$$

- Se quisermos calcular probabilidade de $X = 5$

$$P(X = 5) = P(4,5 \leq X \leq 5,5) = P\left(\frac{4,5-5}{2,12} \leq Z \leq \frac{5,5-5}{2,12}\right)$$

$$= P(-0,24 \leq Z \leq 0,24) = 0,19\,.$$

- A probabilidade exata é

$$P(X = 5) = 0,1849\,.$$

- A aproximação é boa.

- Se np ou $n(1-p)$ forem pequenos:
  - a aproximação é distorcida.
  - Dois casos são ilustrados abaixo:

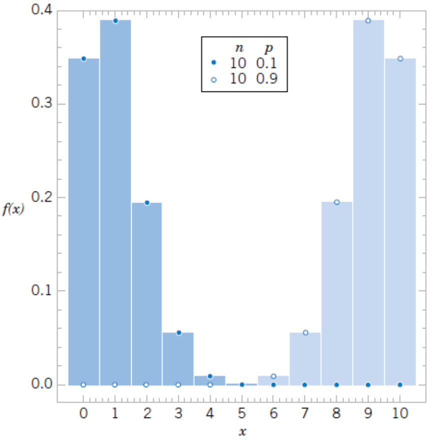


Figure 4-20 Binomial distribution is not symmetrical if p is near 0 or 1.

- A distribuição Binomial é uma boa aproximação para Hipergeométrica.
- Desde que o tamanho da amostra n seja pequeno com relação ao total N.
- Uma regra prática é aproximar a Hipergeométrica pela Binomial se
$$\frac{n}{N} < 0,1 \, .$$
- Relembre que
$$p = \frac{K}{N}$$
proporção de sucessos na população.

- Então podemos aproximar a hipergeométrica pela Normal desde que

$$\frac{n}{N} < 0,1 \qquad np > 5 \qquad n(1-p) > 5 \,.$$

- Veja o esquema abaixo.

hypergeometric distribution	$\approx$ $\frac{n}{N} < 0.1$	binomial distribution	$\approx$ $np > 5$ $n(1-p) > 5$	normal distribution

Figure 4-21 Conditions for approximating hypergeometric and binomial probabilities.

Aproximação da Normal para Poisson

- A distribuição de Poisson pode ser vista como limite de uma Binomial.
- Quando o número de tentativas aumenta e a probabilidade diminui.
- Então a normal também pode ser usada para aproximar a distribuição de Poisson.

Aproximação da Poisson pela Normal

- Seja X uma variável com distribuição Poisson(λ).
- Temos que
$$E(X) = \lambda \qquad Var(X) = \lambda\,.$$
- Então
$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$
é aproximadamente uma variável Normal Padrão.

Exemplo:

- Considere que o número de partículas em uma superfície segue uma distribuição Poisson.
- Suponha que esperamos observar 1000 partículas por m^2.
- Analisamos um metro quadrado da superfície.
- Qual a probabilidade de observarmos 950 ou menos partículas?

Exemplo: (solução)

- Podemos calcular a probabilidade de maneira exata

$$P(X \leq 950) = \sum_{x=0}^{950} \frac{e^{-1000} 1000^{x}}{x!} .$$

- Essa soma é computacionalmente difícil de ser calculada.
- Podemos aproximar a probabilidade por

$$\begin{aligned} P(X \leq 950) &\approx P\left(Z \leq \frac{950 + 0,5 - 1000}{\sqrt{1000}}\right) \\ &\approx P(Z \leq -1,58) = 0,057 . \end{aligned} \qquad (1)$$

Exemplo:

- Estamos analisando as vendas em um supermercado.
- Sabemos que a procura diária por arroz em kg é uma variável aleatória.
- Temos ainda que o valor esperado dessa variável é 40 kg e sua variância 25 kg^2 .
- Foram encomendados 14500 kg para suprir o próximo ano.
- Qual a probabilidade do stock de arroz cobrir a demanda nesse período?
- Considere que o ano tem 364 dias.

Exemplo: (solução)

- Seja
$$X_i = \text{procura de arroz no dia } i$$
para $i = 1, 2, \ldots, 364$.
- Temos que
$$E(X_i) = 40kg \quad Var(X_i) = 25kg^2$$
- A probabilidade requerida é
$$P(\sum_{i=1}^{364} X_i \leq 14500)\,.$$
- Pelo TCL temos que
$$\sum_{i=1}^{364} X_i \sim N\left((364)(40); (364)(25)\right)\,.$$

- Temos que

$$P(\sum_{i=1}^{364} X_i \leq 14500) \approx P\left(\frac{\sum_{i=1}^{364} X_i - 14560}{\sqrt{364(25)}} \leq \frac{14500 + 0,5 - 14560}{\sqrt{364(25)}}\right)$$

$$\approx P(Z < -0,63) = 1 - 0,7357 = 0,2643\,.$$

- Conclusão: é recomendável aumentar o estoque.