Modelos Teóricos Discretos

Introdução às Probabilidades e Estatística

Sumário

1. Modelo Binomial

Modelo binomial

Muitos experimentos resultam em respostas dicotómicas – isto é, respostas para as quais existem duas alternativas possíveis, como Sim-Não, Aprovado-Reprovado, Defeituoso-Não Defeituoso ou Masculino-Feminino.

Um exemplo simples de tal experimento é o experimento do lançamento de uma moeda. Uma moeda é lançada várias vezes, digamos, 10 vezes. Cada lançamento resulta num de dois resultados: Cara ou Coroa. Em última análise, estamos interessados na distribuição de probabilidade de X, o número de caras observadas.

Muitos outros experimentos equivalem a lançar uma moeda (equilibrada ou desequilibrada) um número fixo n de vezes e observar o número x de vezes que um dos dois resultados possíveis ocorre. Variáveis aleatórias que possuem estas características são chamadas de **variáveis aleatórias binomiais**.

Distribuição de Bernoulli

Definição (Prova de Bernoulli)

Uma experiência aleatória diz-se uma **prova de Bernoulli** se possuir apenas dois resultados possíveis

- um sucesso A, que ocorre com probabilidade p $(0 \leq p \leq 1)$;
- um insucesso $\overline{A}$, que ocorre com probabilidade $1 - p$.

Exemplos:

- Lançar uma moeda e observar se o resultado é "cara" ou "coroa".
- Testar um componente eletrónico para verificar se ele funciona ou não.
- Realizar um teste de diagnóstico para uma doença e verificar se o resultado é positivo (sucesso) ou negativo (insucesso).
- Aplicar uma corrente elétrica a uma célula e observar se ela gera uma resposta bioelétrica.
- Testar um sensor biomédico e verificar se ele deteta corretamente uma molécula-alvo, como a glicose.

Distribuição de Bernoulli (cont.)

Definição (Distribuição de Bernoulli)

A variável aleatória discreta

$$X = \text{“número de sucessos numa prova de Bernoulli”}$$

diz-se com **distribuição de Bernoulli com parâmetro p** e possui função massa de probabilidade dada por

$$P(X = x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & \text{c.c.} \end{cases} = \begin{cases} p^x (1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{c.c.} \end{cases}$$

Notação

- $X \sim \text{Bernoulli}(p)$
- $p = P(\text{sucesso})$
- $E(X) = p$
- $V(X) = p(1 - p)$

Distribuição Binomial

Definição (Distribuição binomial)

A variável aleatória discreta

$X =$ "número de sucessos num conjunto de n provas de Bernoulli independentes com probabilidade de sucesso comum e igual a p"

diz-se com **distribuição binomial de parâmetros (n, p)** e possui f.m.p. dada por

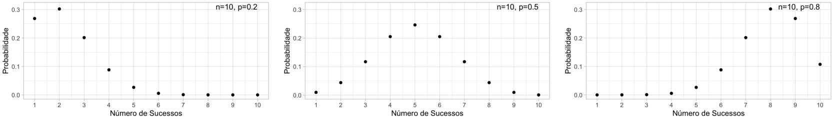
$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, 2, \ldots, n \\ 0, & \text{c.c.} \end{cases}$$

onde $\binom{n}{x} = C_x^n = \frac{n!}{(n-x)!x!}$.

Notação

- $X \sim \text{Binomial}(n, p)$
- $p = P(\text{sucesso})$
- $E(X) = np$
- $V(X) = np(1-p)$

Abaixo temos o gráfico da f.m.p. das v.a. $X \sim \text{Binomial}(10, p)$, para $p = 0.2,\ 0.5,\ 0.8$.



Distribuição Binomial (cont.)

Propriedade da distribuição binomial

Seja X o número de sucessos em n provas de Bernoulli independentes com probabilidade de sucesso comum p, isto é, $X \sim \text{Binomial}(n, p)$. Então, o número de insucessos nessas mesmas n provas de Bernoulli, Y, verifica

$$Y = n - X \sim \text{Binomial}(n, 1 - p)$$

Exemplo: Numa certa população, sabe-se que 30% das pessoas são a favor da exploração de recursos minerais em áreas protegidas. Se forem selecionadas 10 pessoas ao acaso e com reposição, qual é a probabilidade de encontrar 5 que são a favor da exploração de recursos minerais em áreas protegidas? Qual é a probabilidade de encontrar pelo menos 5 nas mesmas condições? Qual o número médio de pessoas a favor?

Resolução:

Seja $X-$ v.a. que indica o número de pessoas que são a favor da exploração de recursos minerais em áreas protegidas nas 10 selecionadas.

$X-$ número de sucessos em 10 provas de Bernoulli independentes com $p = 0.30$.

Logo

$$X \sim \text{Binomial}(n = 10, p = 0.30)$$

Queremos calcular $P(X = 5)$ e $P(X \geq 5)$.

Distribuição Binomial (cont.)

Exemplo 1 (cont.):

- $P(X = 5) = \binom{10}{5}(0.30)^5(0.70)^5 = 0.1029 \quad \text{(tabela)}$

Em R: $>$ dbinom(5,10,0.3) $\rightarrow$ [1] 0.1029193

Em Excel: $=$ BINOM.DIST(5,10,0.3,FALSE) $\rightarrow$ 0.102919345

Exemplo 1 (cont.):

- $$\begin{aligned} P(X \geq 5) &= \sum_{x=5}^{10} \binom{10}{x} (0.30)^x (0.70)^{10-x} = \cdots = 1 - P(X < 5) \\ &= 1 - P(X \leq 4) = \\ &= 1 - F_X(4) \\ &= 1 - 0.8497 \quad \text{(tabela)} \\ &= 0.1503 \end{aligned}$$

Em R: > 1-pbinom(4,10,0.3) $\rightarrow$ [1] 0.1502683

Em Excel: =1-BINOM.DIST(4;10;0.3;TRUE) $\rightarrow$ 0,150268333

- $E(X) = n \times p = 10 \times 0.30 = 3$ pessoas a favor.

2. Modelo Hipergeométrico

Distribuição Hipergeométrica

A distribuição hipergeométrica é usada quando se pretendem determinar probabilidades associadas à extração de n objetos sem reposição, de um total de N objetos, dos quais M verificam uma dada característica (sucesso).

Definição (Distribuição Hipergeométrica)

Considerem-se:

- $N =$ número total de elementos de uma população (dimensão da população);
- $M =$ número de elementos da população que possuem determinada característica (sucesso);
- $n =$ número de extrações ao acaso sem reposição.

Então, a v.a.

$X =$ "número de elementos com tal característica (sucesso), em n selecionados, ao acaso e sem reposição da população descrita acima"

diz-se com **distribuição hipergeométrica com parâmetros** (N, M, n) e possui f.m.p. dada por

$$P(X = x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & x = \max\{0, n - N + M\}, \ldots, \min\{n, M\} \\ 0, & c.c. \end{cases}$$

Distribuição Hipergeométrica (cont.)

Notação

- $X \sim \text{Hipergeométrica}(N, M, n)$
- $E(X) = n \times \frac{M}{N}$
- $V(X) = n \times \frac{M}{N} \times (1 - \frac{M}{N}) \times \frac{N-n}{N-1}$
- Note que a probabilidade de sucesso não é constante, pois as extrações são feitas sem reposição.
- **Aproximação da hipergeométrica à binomial**

  A função de distribuição hipergeométrica converge para a função de distribuição binomial quando n é muito inferior a N, na prática, quando $n < 0.1N$.

Ciências ULisboa

Exemplo: Um engenheiro biomédico está a analisar um lote de 100 amostras de tecido humano cultivado em laboratório. Sabe-se que 30 dessas amostras apresentam uma resposta positiva a um novo tratamento experimental. O engenheiro seleciona 10 amostras ao acaso, sem reposição.

(a) Qual é a probabilidade de que exatamente 4 das 10 amostras selecionadas apresentem uma resposta positiva ao tratamento?

(b) Qual é a probabilidade de que pelo menos 2 das 10 amostras selecionadas apresentem uma resposta positiva ao tratamento?

(c) Calcule a média e a variância do número de amostras que apresentam uma resposta positiva ao tratamento entre as 10 amostras selecionadas.

Respostas: **(a)** 0.2075, **(b)** 0.8644, **(c)** $E(X) = 3$, $V(X) = 1.9090$

Ciências ULisboa

3. Modelo Binomial Negativo

Distribuição Binomial Negativa

Definição 1 (Distribuição Binomial Negativa)

A variável aleatória discreta

$X =$ "número de provas de Bernoulli (independentes e com probabilidade de sucesso comum igual a p) até obter o k-ésimo sucesso"

diz-se com **distribuição binomial negativa de parâmetros k e p** e possui f.m.p. dada por

$$P(X=x) = \begin{cases} \binom{x-1}{k-1}(1-p)^{x-k}p^k, & x = k, k+1, \ldots; k = 1, 2, \ldots \\ 0, & c.c. \end{cases}$$

Notação

- $X \sim \text{Bn}(k, p)$
- $E(X) = \frac{k}{p}$
- $V(X) = \frac{k(1-p)}{p^2}$

Distribuição Binomial Negativa (cont.)

A variável aleatória discreta com distribuição binomial negativa pode ser definida de outro modo...

Definição 2 (Distribuição Binomial Negativa)

A variável aleatória discreta

$$Y = X - k = \text{“número de insucessos até obter o k-ésimo sucesso"}$$

diz-se com **distribuição binomial negativa de parâmetros k e p** e possui f.m.p. dada por

$$P(Y=y) = \begin{cases} \binom{y+k-1}{k-1}(1-p)^y p^k, & y = 0,1,2,\ldots; k = 1,2,\ldots \\ 0, & c.c. \end{cases}$$

Notação

- $Y \sim \text{Bn}(k,p)$
- $E(Y) = \frac{k(1-p)}{p}$
- $V(Y) = \frac{k(1-p)}{p^2} = V(X)$

Exemplo: Um engenheiro biomédico está a desenvolver um novo teste de diagnóstico para identificar uma proteína específica em amostras de tecido humano. A probabilidade de encontrar essa proteína numa amostra é de 0.25. O engenheiro precisa identificar 7 amostras que contenham a proteína para validar o teste.

(a) Qual é a probabilidade de que o engenheiro precise analisar exatamente 20 amostras de tecido para identificar 7 amostras com a proteína?

(b) Qual é a probabilidade de que o engenheiro precise analisar no máximo 8 amostras de tecido para identificar 7 amostras com a proteína?

(c) Qual é a probabilidade de que o engenheiro encontre a 7ª amostra com a proteína na 18ª amostra analisada?

Respostas: **(a)** 0.03934, **(b)** 0.000381, **(c)** 0.031903

4. Modelo Geométrico

Distribuição Geométrica

Definição 1 (Distribuição Geométrica)

A variável aleatória discreta

$X =$ "número de provas de Bernoulli (independentes e com probabilidade de sucesso comum igual a p) realizadas até à ocorrência do primeiro sucesso"

diz-se com **distribuição geométrica com parâmetro p** e possui f.m.p. dada por

$$P(X = x) = \begin{cases} (1-p)^{x-1}p, & x = 1, 2, 3, \ldots \\ 0, & c.c. \end{cases}$$

Notação

- $X \sim \text{Geométrica}(p)$
- $p = P(sucesso)$
- $E(X) = \frac{1}{p}$
- $V(X) = \frac{1-p}{p^2}$
- **Propriedade da falta de memória**

$$P(X > i + j | X > j) = P(X > i), \quad \forall i, j = 1, 2, \cdots$$

A variável aleatória discreta com distribuição geométrica pode ser defina de outro modo... (O R trabalha com essa definição).

Definição 2 (Distribuição Geométrica)

A variável aleatória discreta

$$Y = X - 1 = \text{“número de insucessos até obter o primeiro sucesso"}$$

diz-se com **distribuição geométrica com parâmetro p** e possui f.m.p. dada por

$$P(Y = y) = \begin{cases} (1-p)^y p, & y = 0, 1, 2, \ldots \\ 0, & c.c. \end{cases}$$

Notação

- $Y \sim \text{Geométrica}(p)$
- $p = P(sucesso)$
- $E(Y) = \frac{1-p}{p}$
- $V(Y) = \frac{1-p}{p^2} = V(X)$

- No caso das v.a. binomiais e hipergeométricas, não se consegue obter uma expressão explícita para a função de distribuição:
  - binomial: $F_X(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$;
  - hipergeométrica: $F_X(x) = P(X \leq x) = \sum_{k=\max\{0,n-N+M\}}^{x} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$.
- No caso da v.a. geométrica, consegue-se obter uma expressão explícita para a função de distribuição:
  - $F_X(x) = P(X \leq x) = \sum_{k=1}^{x} p(1-p)^{k-1} = p\frac{1-(1-p)^x}{1-(1-p)} = 1-(1-p)^x$
  - $F_Y(y) = P(Y \leq y) = \sum_{k=0}^{y} p(1-p)^k = p\frac{1-(1-p)^{y+1}}{1-(1-p)} = 1-(1-p)^{y+1}$

Ciências ULisboa

Exemplo: Seja X a v.a. que indica o número de lançamentos de um dado equilibrado até surgir a primeira face 2.

(a) Qual a probabilidade da face 2 surgir no terceiro lançamento?

(b) Qual o número esperado de lançamentos do dado até sair a face 2?

(c) Qual a probabilidade de serem necessários mais de 10 lançamentos sabendo que já houve 6 lançamentos do dado sem que a face 2 saísse?

Resolução:

(a)

- X = “número de lançamentos de um dado equilibrado até surgir a primeira face 2"
- $X \sim$ Geométrica $\left(p = \frac{1}{6}\right)$
- $P(X = 3) = \left(1 - \frac{1}{6}\right)^{3-1} \times \frac{1}{6} = \left(\frac{5}{6}\right)^{2} \times \frac{1}{6} = 0.11574$

Em R: $Y = X - 1 \Rightarrow P(X = 3) = P(Y = 2)$

dgeom(2,1/6) $\rightarrow$ [1] 0.1157407

Resolução (cont.):

(b) $E(X) = \frac{1}{p} = 6$

(c)

$$\begin{aligned} P(X > 10|X > 6) &= P(X > 4) \\ &= 1 - P(X \leq 4) \\ &= 1 - F_X(4) \\ &= 1 - \left(1 - \left(1 - \frac{1}{6}\right)^4\right) \\ &= 0.4822 \end{aligned}$$

Em R: $Y = X - 1 \Rightarrow 1 - P(X \leq 4) = 1 - P(Y \leq 3)$

1-pgeom(3,1/6) $\rightarrow$ [1] 0.4822531

Exercício: Um físico está a estudar a presença de uma partícula subatómica instável produzida numa reação nuclear rara. A probabilidade de detetar essa partícula em qualquer evento de colisão de partículas é de 0.15. Suponha que cada colisão é independente das anteriores.

(a) Qual é a probabilidade de que a primeira colisão em que a partícula seja detetada ocorra na terceira tentativa?

(b) Qual é a probabilidade de que o físico precise realizar pelo menos quatro colisões para detetar a partícula pela primeira vez?

(c) Calcule o número esperado de colisões necessárias até detetar a partícula pela primeira vez.

(d) Determine a variância do número de colisões necessárias para detetar a partícula pela primeira vez.

5. Modelo de Poisson

Distribuição de Poisson

O modelo de Poisson é adequado quando se pretende estudar o número de ocorrências de um certo acontecimento num determinado intervalo de tempo ou região do espaço. Suponhamos que se verificam as seguintes hipóteses:

1. a ocorrência do acontecimento num determinado intervalo é independente da ocorrência do acontecimento em qualquer outro intervalo distinto;
2. a probabilidade de exatamente uma ocorrência do acontecimento em qualquer intervalo de amplitude h arbitrariamente pequena é aproximadamente λh;
3. a probabilidade de duas ou mais ocorrências do acontecimento em qualquer intervalo de amplitude h arbitrariamente pequena é aproximadamente 0.

Definição (Distribuição de Poisson)

A variável aleatória discreta

$X =$ "número de ocorrências de um acontecimento por unidade de tempo ou de espaço (comprimento, área, volume, etc)"

diz-se com **distribuição de Poisson de parâmetro $\lambda > 0$** e possui f.m.p. dada por

$$P(X = x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & c.c. \end{cases}$$

Notação

- $X \sim \text{Poisson}(\lambda)$
- $E(X) = \lambda$
- $V(X) = \lambda$
- λ representa o número médio de ocorrências de um acontecimento por unidade de tempo ou espaço.

Distribuição de Poisson

- $X-$ "número de ocorrências no intervalo λt"
- $X \sim \text{Poisson}(\lambda t)$
- $$P(X = x) = \begin{cases} \frac{e^{-\lambda t}(\lambda t)^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & c.c. \end{cases}$$
- $E(X) = \lambda t$
- $V(X) = \lambda t$

Aditividade da distribuição de Poisson

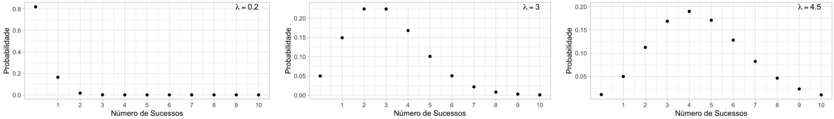
Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes, com $X_i \sim \text{Poisson}(\lambda_i)$ para $i = 1, 2, \ldots, n$ então

$$\sum_{i=1}^{n} X_i \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right).$$

Ciências ULisboa

Distribuição de Poisson (cont.)

Abaixo temos o gráfico das v.a. $X \sim \text{Poisson}(\lambda)$, para $\lambda = 0.2, 3, 4.5$



Exemplo: Geólogos estão a estudar a ocorrência de terramotos numa região específica. Eles observaram que, em média, ocorrem 3 terramotos por mês nessa região. O número de terramotos por mês pode ser modelado por uma distribuição de Poisson.

(a) Calcule a probabilidade de ocorrer exatamente 2 terramotos em um mês.

(b) Calcule a probabilidade de ocorrer mais de 4 terramotos em um mês.

(c) Suponha que a equipa de geólogos está a planear um sistema de alerta para terramotos. Eles querem saber a probabilidade de ocorrer pelo menos 1 terramoto num período de 2 semanas.

Distribuição de Poisson (cont.)

Resolução:

(a)

Variável aleatória

- $X =$ "número de terremotos por mês"
- $X \sim \text{Poisson}(\lambda)$
- $E(X) = 3 = \lambda$

Função de probabilidade de X

- $P(X = x) = \frac{3^x e^{-3}}{x!}, \quad x = 0, 1, 2, \ldots$

Probabilidade pedida

- $P(X = 2) = \frac{3^2 e^{-3}}{2!} = 0.2240$ (tabela)

Em R: > dpois(2,3) $\rightarrow$ [1] 0.2240418

Ciências ULisboa

Resolução:

(b)

Probabilidade pedida

$$\begin{aligned} P(X > 4) &= 1 - P(X \leq 4) \\ &= 1 - F_X(4) \\ &= 1 - 0.8153 \quad \text{(tabela)} \\ &= 0.1847 \end{aligned}$$

Em R: > 1-ppois(4,3) → [1] 0.1847368

Distribuição de Poisson (cont.)

Resolução:

(c)

- $\tilde{X}-$ "número de terramotos em 2 semanas"
- $\tilde{X} \sim \text{Poisson}(\lambda = 1.5)$
- $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \frac{1.5^0 e^{-1.5}}{0!} = 0.7768$

Em R: > 1-dpois(0,1.5) $\rightarrow$ [1] 0.7768698

Referências

- Manuel Cabral Morais (2020): Probabilidade e Estatística: Teoria, Exemplos & Exercícios. IST Press, 1a edição.