Bioestatística F

Modelo Binomial

Enrico A. Colosimo

Departamento de Estatística
Universidade Federal de Minas Gerais
http://www.est.ufmg.br/~enricoc

2011

Variável aleatória discreta

Definição

Uma quantidade X, associada a cada possível resultado do espaço amostral, é denominada de **variável aleatória discreta**, se assume valores num conjunto enumerável, com certa probabilidade.

- **Exemplo:** Observa-se o sexo das crianças em famílias com três filhos. Denotamos m para o sexo masculino e f para o sexo feminino. Existem oito possibilidades para uma família de três filhos. Estas possibilidades são listadas no espaço amostral:

$$\Omega = \{(mmm), (mmf), (mfm), (fmm), (mff), (fmf), (ffm), (fff)\}.$$

Variável aleatória discreta

Definimos

- X : número de crianças do sexo masculino (m).
- A cada possível resultado do espaço amostral, X associa um valor numérico

Ω	mmm	mmf	mfm	fmm	mff	fmf	ffm	fff
X	3	2	2	2	1	1	1	0

- Note que X assume valores no conjunto $\{0, 1, 2, 3\}$ que é enumerável, portanto X é variável aleatória discreta.

Variável aleatória discreta

- **Pergunta:** com qual probabilidade X assume os valores $\{0, 1, 2, 3\}$?
- Cada resultado possível do espaço amostral tem probabilidade $\frac{1}{8}$ de acontecer, então:

$$P(X = 0) = P(fff) = \frac{1}{8}$$

- A probabilidade da variável aleatória X assumir o valor zero é a mesma probabilidade do evento (fff) ocorrer. Da mesma forma:

$$\begin{aligned}
P(X = 1) &= P(\{mff\} \cup \{fmf\} \cup \{ffm\}) \\
&= P(mff) + P(fmf) + P(ffm) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} \\
P(X = 2) &= P(\{mmf\} \cup \{mfm\} \cup \{fmm\}) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} \\
P(X = 3) &= P(mmm) = \frac{1}{8}
\end{aligned}$$

- **Resumindo:**

X	0	1	2	3
$P(X = x_i)$	1/8	3/8	3/8	1/8

Definição

Seja X uma variável aleatória discreta e $x_1, x_2, x_3, \ldots$, seus diferentes valores. A função que atribui a cada valor da variável aleatória sua probabilidade é denominada de **função discreta de probabilidade** ou, simplesmente **função de probabilidade**.

- A notação a ser utilizada é:

$$P(X = x_i) = p(x_i) = p_i, i = 1, 2, \ldots$$

ou ainda,

X	x_1	x_2	x_3	$\ldots$
p_i	p_1	p_2	p_3	$\ldots$

- Uma função de probabilidade satisfaz:
  1. $0 \leq p_i \leq 1$
  2. $\sum_i p_i = p_1 + p_2 + \ldots = 1$
- **No exemplo:**

$$\begin{aligned} p_1 + p_2 + p_3 + p_4 &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= 1/8 + 3/8 + 3/8 + 1/8 = 1. \end{aligned}$$

- Um dado é lançado duas vezes, de forma independente. O espaço amostral deste experimento é:

$$\Omega = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

- Seja X a variável *soma dos dois lançamentos*, ou seja,
  $X = \text{“face do primeiro lançamento”} + \text{“face do segundo lançamento”}$.

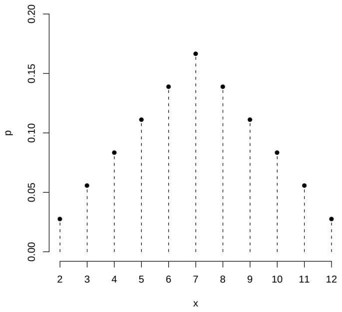
- Quando o evento $(1,1)$ ocorre, X associa a este resultado o valor 2. Da mesma forma temos:

$$\Omega = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\} \overset{X}{\Rightarrow} \left\{ \begin{array}{cccccc} 2, & 3, & 4, & 5, & 6, & 7, \\ 3, & 4, & 5, & 6, & 7, & 8, \\ 4, & 5, & 6, & 7, & 8, & 9, \\ 5, & 6, & 7, & 8, & 9, & 10, \\ 6, & 7, & 8, & 9, & 10, & 11, \\ 7, & 8, & 9, & 10, & 11, & 12 \end{array} \right\}$$

- X assume valores no conjunto $\{2,3,4,5,6,7,8,9,10,11,12\}$.
- A probabilidade dos possíveis resultados em Ω é $P[(1,1)] = P[(1,2)] = \ldots = P[(2,1)] = P[(2,2)] = \ldots = P[(6,6)] = 1/36$.
- $P[X = 2] = P[(1,1)] = 1/36$,
  $P[X = 3] = P[(1,2) \cup (2,1)] = 1/36 + 1/36 = 2/36$.

- A função de probabilidade de X é dada por:

X	2	3	4	5	6	7	8	9	10	11	12
p_i	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36



- **Observações:**
  - p_i pertence ao intervalo $(0,1)$ para $i = 1, \ldots, 11$ (ex: $p_1 = 1/36 \in (0,1)$)
  - $\sum_{i=1}^{11} p_i = p_1 + p_2 + \ldots + p_{11} = 1/36 + 2/36 + \ldots + 1/36 = 1$
  - A função de probabilidade de X satisfaz às condições 1 e 2, logo é de fato uma função de probabilidade.
- **Pergunta:** Qual a probabilidade da soma dos resultados dos dois lançamentos ser menor do que 6?

$$
\begin{aligned}
P(X < 6) &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\
&= p_1 + p_2 + p_3 + p_4 \\
&= 1/36 + 2/36 + 3/36 + 4/36 = 10/36.
\end{aligned}
$$

Principais modelos discretos

- Algumas variáveis aleatórias aparecem com bastante frequência em situações práticas e justificam um estudo mais aprofundado.
- Em geral nesses casos, a função de probabilidade pode ser escrita de uma maneira mais compacta, isto é, existe uma **lei** para atribuir as probabilidades.
- Por exemplo, se uma variável aleatória W tem função de probabilidade dada por

W	1	2	3	4	5	6
p_i	1/21	2/21	3/21	4/21	5/21	6/21

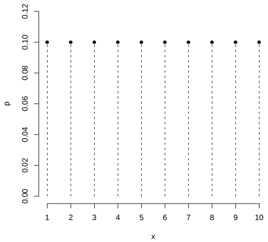
- Uma forma abreviada de apresentar a variável e sua função de probabilidade é $P(W = k) = k/21, k = 1, 2, \ldots, 6$.

Modelo Uniforme Discreto

Seja X uma variável aleatória cujos possíveis valores são representados por $x_1, x_2, \ldots, x_k$. Dizemos que X segue o modelo *Uniforme Discreto* se atribui a **mesma probabilidade** $1/k$ a cada um desses k valores, isto é, sua função de probabilidade é dada por

$$P(X = x_j) = 1/k, \quad \text{para} \quad j = 1, 2, \ldots, k.$$

- Para $k = 10$ o gráfico da função de probabilidade é



- **Observação:** A expressão na definição anterior, de fato, representa uma função de probabilidade
  1. $P(X = k) = 1/k \in [0, 1]$
  2. $\sum_k P(X = k) = \sum_k 1/k = k \times 1/k = 1$

- **Exemplo:** Uma rifa tem 100 bilhetes numerados de 1 a 100. Tenho 5 bilhetes consecutivos numerados de 21 a 25 e meu colega tem outros 5 bilhetes, com os números 1, 11, 29, 68 e 93. Quem tem maior possibilidade de ser sorteado?
- À primeira vista tem-se a impressão de que "espalhar" os números é a melhor maneira de ganhar o sorteio.
- Assumindo a honestidade da rifa, todos os números tem a mesma probabilidade de ocorrência, com $1/100$ para cada um.
- A variável aleatória em questão, o *número sorteado*, segue o modelo uniforme e, portanto, eu e meu colega temos a mesma probabilidade de ganhar a rifa.
- Neste sorteio, como no modelo uniforme em geral, a maior ou menor probabilidade de ganhar depende de quantos bilhetes se tem e não da particular escolha do número.

Principais modelos discretos

- Em muitas situações práticas a variável de interesse assume somente dois valores.
    - Em um levantamento do sistema de saúde, as pessoas são perguntadas se tem ou não um seguro de hospitalização.
    - Amostras de sangue são testadas para a presença ou ausência de um antígeno.
    - Ratos alimentados com uma substância cancerígena em potencial são examinados para tumores.
    - As pessoas são classificadas como tendo ou não tendo lábio leporino.
    - A injeção de um composto causar ou não arritmia cardíaca em cães.
    - Crianças recém-nascidas são classificadas como tendo ou não a síndrome de Down.

- Estas situações têm alternativas dicotômicas, que genericamente podem ser representadas por respostas do tipo *sucesso-fracasso*.
- Esses experimentos recebem o nome de *Ensaios de Bernoulli* e dão origem a uma variável aleatória com o mesmo nome.

Modelo Bernoulli

Dizemos que uma variável X segue o modelo *Bernoulli* se atribui 0 ou 1 à ocorrência de fracasso ou sucesso, respectivamente. Com p representando a probabilidade de *sucesso*, $0 \leq p \leq 1$, sua função discreta de probabilidade é dada por

X	0	1
p_i	$1-p$	p

$$P(X = x) = p^x(1-p)^{1-x}, x = 0, 1.$$

Modelo Binomial

Considere a repetição de n ensaios de Bernoulli independentes e todos com a mesma probabilidade de sucesso p. A variável aleatória que conta o número total de sucessos é denominada Binomial com parâmetros n e p e sua função de probabilidade é dada por

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \ldots, n,$$

com $\binom{n}{k}$ representando o coeficiente binomial calculado por

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Notação: $X \sim b(n, p)$ indica que a variável aleatória X segue o modelo Binomial com parâmetros n e p.

- O coeficiente binomial $\binom{n}{k}$ é o número de diferentes maneiras que k objetos podem ser selecionados a partir de n objetos.
- **Exemplos:**
  - De 10 residentes, três são escolhidos para cobrir um serviço hospitalar em um feriado. Em quantas maneiras podem ser escolhidos os residentes?

$$\binom{10}{3} = \frac{10!}{3!7!} = \frac{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10}{(1 \times 2 \times 3)(1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7)} = 120$$

  - De oito pacientes consecutivos, quatro são designados ao tratamento A e de quatro ao tratamento B. De quantas maneiras podem ser feitas as designações de tratamentos? Pense em oito posições como oito objetos; precisamos escolher quatro para os pacientes do tratamento A.

$$\binom{8}{4} = \frac{8!}{4!4!} = \frac{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8}{(1 \times 2 \times 3 \times 4)(1 \times 2 \times 3 \times 4)} = 70$$

- **Exemplo:** Dez pacientes são tratados cirurgicamente. Para cada pessoa há uma chance de 70% da cirurgia ser bem sucedida (ou seja, $p = 0,7$). ($X \sim b(10; 0,7)$)
  - Qual é a probabilidade de apenas cinco cirurgias serem bem sucedidas?

$$\begin{aligned} P(X=5) &= \binom{10}{5} 0,7^5 (1-0,7)^{10-5} \\ &= \frac{10!}{5!5!} \times 0,7^5 \times 0,3^5 \\ &= 252 \times 0,16807 \times 0,00243 = 0,1029 \end{aligned}$$

  - Qual é a probabilidade de apenas cinco ou menos cirurgias serem bem sucedidas?

$$\begin{aligned} P(X \leq 5) &= P(X=0) + P(X=1) + P(X=2) \\ &+ P(X=3) + P(X=4) + P(X=5) \\ &= 0,0000 + 0,0001 + 0,0014 + 0,0090 + 0,0368 + 0,1029 \\ &= 0,1502 \end{aligned}$$

Quatro condições caracterizam dados binomiais

1. Uma resposta toma uma e apenas uma de duas possibilidades (variável dicotômica).
2. A resposta é observada um número conhecido de vezes. Cada observação é um *ensaio de Bernoulli*.
3. A probabilidade de *sucesso* é a mesma (constante) em cada ensaio.
4. A resposta de um ensaio não deve ser influenciada pela resposta dos outros ensaios (os ensaios são independentes).

Média e Variância

Valor Esperado (Média)

Dada a variável aleatória X, assumindo os valores $x_1, x_2, \ldots, x_n$, chamamos de valor médio, ou valor esperado, ou esperança matemática de X o valor

$$E(X) = \sum_{i=1}^{n} x_i p_i.$$

Notação: $E(X) = \mu$.

- No exemplo do lançamento dos dois dados, a média de X é

$$E(X) = 2 \times 1/36 + 3 \times 2/36 + \ldots + 12 \times 1/36 = 252/36 = 7$$

Média e Variância

Variância

Seja X uma variável aleatória com $P(X = x_i) = p_i, i = 1, 2, \ldots, k$ e média μ. A variância de X é a ponderação pelas respectivas probabilidades, dos desvios relativos a média, elevados ao quadrado, isto é,

$$Var(X) = \sum_{i=1}^{k} (x_i - \mu)^2 p_i.$$

Notação: $Var(X) = \sigma^2$.

- Extraindo a raiz quadrada da variância obtemos o *desvio-padrão* que é representado por σ.
- $Var(X) = E[(X - \mu)^2]$, ou seja a variância é a méda da variável aleatória $(X - \mu)^2$.
- Pode se mostrar que $Var(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$.

Média e Variância

- Para o modelo Uniforme $(1, k)$: $E(X) = \frac{k+1}{2}$, $Var(X) = \frac{k^2-1}{12}$
- Para o modelo Bernoulli (p): $E(X) = p$, $Var(X) = p(1-p)$
- Para o modelo Binomial (n, p): $E(X) = np$, $Var(X) = np(1-p)$