Variáveis aleatórias discretas

Wagner H. Bonat
Elias T. Krainski
Fernando P. Mayer

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

14/03/2018

Sumário

Variáveis aleatórias

Em probabilidade, uma função X que associa a cada evento do espaço amostral um número real $X(\omega) \in \mathbb{R}$, é denominada uma **variável aleatória** (VA).

Uma variável aleatória pode ser classificada como **discreta** ou **contínua**, dependendo do domínio dos valores de X.

Exemplo: o número de alunos em uma sala é uma variável aleatória (discreta), denotada por X (maiúsculo). Uma observação dessa variável é denotada pela respectiva letra minúscula, *e.g.*, $x = 50$ alunos.

Em geral, denotamos a probabilidade de uma V.A. X assumir determinado valor x como

$$P[X] \quad \text{ou} \quad P[X = x]$$

Variáveis aleatórias

Dada a realização de um experimento aleatório qualquer, com um certo espaço de probabilidade, desejamos estudar a **estrutura probabilística** de quantidades associadas à esse experimento.

Note que antes da realização de um experimento, **não sabemos seu resultado**, entretanto seu espaço de probabilidade pode ser previamente estabelecido.

Dessa forma, podemos atribuir probabilidades aos *eventos* desse espaço amostral, dando origem ao conceito de **variável aleatória**.

Distribuições de probabilidade

Existem diversos *modelos probabilísticos* que procuram descrever vários tipos de variáveis aleatórias: são as **distribuições de probabilidade de variáveis aleatórias** (discretas ou contínuas).

A distribuição de probabilidades de uma VA X é, portanto, uma descrição das probabilidades associadas com os possíveis valores de X. Os valores que X assume determinam o **suporte** (S) da VA.

- **Variáveis discretas** $\rightarrow$ suporte em um conjunto de valores enumeráveis (finitos ou infinitos)
- **Variáveis contínuas** $\rightarrow$ suporte em um conjunto não enumerável de valores

Distribuições de probabilidade

Denomina-se de **distribuição de probabilidade** de alguma variável aleatória, a **regra** geral que define a

- **função de probabilidade** (fp) (V.A.s discretas), ou a
- **função densidade de probabilidade** (fdp) (V.A.s contínuas)

para a variável de interesse.

Existem muitas distribuições de probabilidade, mas algumas merecem destaque por sua importância prática.

Estas distribuições também são chamadas de **modelos probabilísticos**.

Sumário

Definição

A **função de probabilidade** (fp) da VA discreta X, que assume os valores $x_1, x_2, \ldots, x_n, \ldots$, é a função que atribui probabilidades a cada um dos possíveis valores: $\{[x_i, p(x_i)], i = 1, 2, ...\}$, ou seja,

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

com as seguintes propriedades:

i. A probabilidade de cada valor deve estar entre 0 e 1

$$0 \leq p(x_i) \leq 1, \quad \forall\, i = 1, 2, ...$$

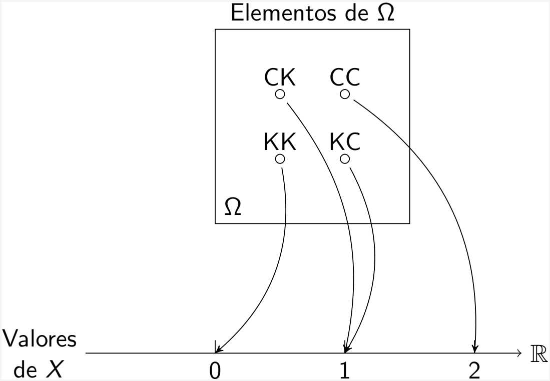
ii. A soma de todas as probabilidades é igual a 1

$$\sum_i p(x_i) = 1$$

Exemplo

Experimento

Lançamento de duas moedas. X = número de resultados cara (C)



Exemplo

Podemos montar uma tabela de distribuição de frequência para a variável aleatória X = número de resultados cara (C)

X	Frequência (f_i)	Frequência relativa (fr_i)
0	1	1/4
1	2	2/4
2	1	1/4
Total	4	1

Assim podemos associar a cada valor de X sua **probabilidade** correspondente, como resultado das **frequências relativas**

$$P[X = 0] = 1/4$$
$$P[X = 1] = 2/4 = 1/2$$
$$P[X = 2] = 1/4$$

Exemplo

Dessa forma, a **distribuição de probabilidade** da variável aleatória $X =$ número de resultados cara (C) é a tabela

X	$P[X = x_i] = p(x_i)$
0	$1/4$
1	$1/2$
2	$1/4$
Total	1

Repare que as propriedades da função de probabilidade estão satisfeitas:

i. As probabilidades $p(x_i)$ estão entre 0 e 1
ii. A soma de todas as probabilidades $p(x_i)$ é 1

Exemplo 3.1 (livro)

Com dados do último censo, a assistente social de um Centro de Saúde constatou que para as famílias da região, 20% não tem filhos, 30% tem um filho, 35% tem dois, e as restantes se dividem igualmente entre três, quatro ou cinco filhos.

Descreva a função de probabilidade da VA N definida como número de filhos.

Exemplo 3.2 (livro)

Na construção de um certo prédio, as fundações devem atingir 15 metros de profundidade e, para cada 5 metros de estacas colocadas, o operador anota se houve alteração no ritmo de perfuração previamente estabelecido. Essa alteração é resultado de mudanças para mais ou para menos, na resistência do subsolo.

Nos dois casos, medidas corretivas serão necessárias, encarecendo o custo da obra. Com base em avaliações geológicas, admite-se que a probabilidade de ocorrência de alterações é de 0.1 para cada 5 metros.

O custo básico inicial é de 100 UPC (unidade padrão de construção) e será acrescida de $50k$, com k representando o número de alterações observadas. Como se comporta a va custo das obras de fundações?

Função de distribuição de probabilidade

Em muitas situações, é útil calcularmos a probabilidade **acumulada** até um certo valor.

Definimos a **função de distribuição** ou **função acumulada de probablidade** de uma VA X pela expressão:

$$F(x) = P(X \leq x)$$

para qualquer número real x.

Exemplo 3.5 (livro)

Uma população de 1000 crianças foi analisada num estudo para determinar a efetividade de uma vacina contra um tipo de alergia.

No estudo, as crianças recebiam uma dose da vacina e, após um mês, passavam por um novo teste. Caso ainda tivessem tido alguma reação alérgica, recebiam outra dose da vacina. Ao fim de 5 doses todas as crianças foram consideradas imunizadas.

Os resultados completados estão na tabela a seguir.

	1	2	3	4	5
Freq.	245	288	256	145	66

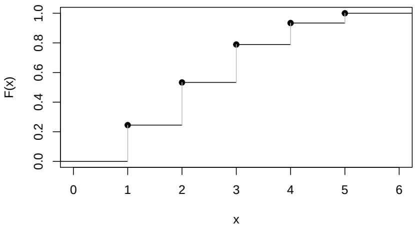
Para uma criança sorteado ao acaso qual a probabilidade dela ter recebido 2 doses? E até 2 doses?

Exemplo 3.5 (livro)

Tabela de frequência:

	n_i	f_i	f_{ac}
1	245	0.245	0.245
2	288	0.288	0.533
3	256	0.256	0.789
4	145	0.145	0.934
5	66	0.066	1.000
Sum	1000	1.000	

Grafico de $F(x)$:



Exemplo 3.5 (livro)

Assim,

$$F(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = 0,533$$

Note que podemos escrever

$$F(x) = P(X \leq x) = 0,533 \quad \text{para} \quad 2 \leq x < 3$$

E os valores completos da função de distribuição são:

$$F(x) = \begin{cases} 0 & \text{se } x < 1 \\ 0,245 & \text{se } 1 \leq x < 2 \\ 0,533 & \text{se } 2 \leq x < 3 \\ 0,789 & \text{se } 3 \leq x < 4 \\ 0,934 & \text{se } 4 \leq x < 5 \\ 1 & \text{se } x \geq 5 \end{cases}$$

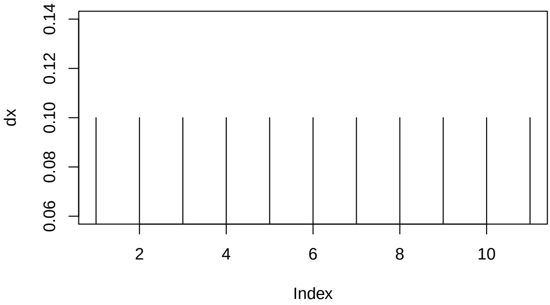
Sumário

Modelo Uniforme Discreto

Seja X uma VA 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.

Sua função de probabilidade é dada por

$$P(X = x_j) = 1/k, \forall j = 1, 2, \ldots, k.$$



Exemplo 3.7 (livro)

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?

Modelo Bernoulli

Dizemos que uma variável X segue o modelo Bernoulli se atribui 0 ou 1 à ocorrência de fracasso ou sucesso, respectivamente.

Sendo, p a probabilidade de sucesso, $0 \leq p \leq 1$, sua função de probabilidade é dada por

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

Exemplo 3.8 (livro): Sabe-se que a eficiência de uma vacina é de 80%. Um grupo de três indivíduos é sorteado, dentre a população vacinada, e submetido a testes para averiguar se a imunização foi efetiva, evento representado por I. Determine o comportamento da VA número de indivíduos imunizados neste grupo.

Modelo Binomial

Considere a repetição de n ensaios de Bernoulli independentes e todos com a mesma probabilidade de sucesso p.

A VA que conta o número total de sucessos é denominada Binomial com parâmetros n e p e sua função de probabilidade é dada por

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

onde

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

Notação: $X \sim b(n, p)$.

Calculando probabilidades baseadas na Binomial.

```
# [1] 0.141894
```

```
#  [1] 0.002 0.017 0.064 0.142 0.213 0.227 0.177 0.101 0.042 0.012 0.002 0.000
# [13] 0.000
```

Exemplo 3.9 (livro)

- O escore de um teste internacional de proficiência na língua inglesa varia de 0 a 700 pontos, com mais pontos indicando um melhor desempenho. Informações, coletadas durante vários anos, permite estabelecer o seguinte modelo para o desempenho no teste:

Pontos	$[0, 200)$	$[200, 300)$	$[300, 400)$	$[400, 500)$	$[500, 600)$	$[600, 7$
p_i	0.06	0.15	0.16	0.25	0.28	0.10

Várias universidades americanas, exigem um escore mínimo de 600 pontos para aceitar candidatos de países de língua não inglesa. De um grande grupo de estudantes brasileiros que prestaram o último exame, escolhemos ao acaso 20 deles. Qual é a probabilidade de no máximo 3 atenderem ao requisito mínimo?

Sumário

Sumário

Sumário

Sumário

Sumário

Sumário

Sumário

Sumário

Sumário

Sumário

Sumário