Motivação e componentes de distribuições de probabilidade

Prof. Walmes M. Zeviani

Departamento de Estatística
Universidade Federal do Paraná

Conteúdo

Neste vídeo

- Motivação para distribuições de probabilidade.
- Componentes de uma distribuição.
- As principais distribuições.

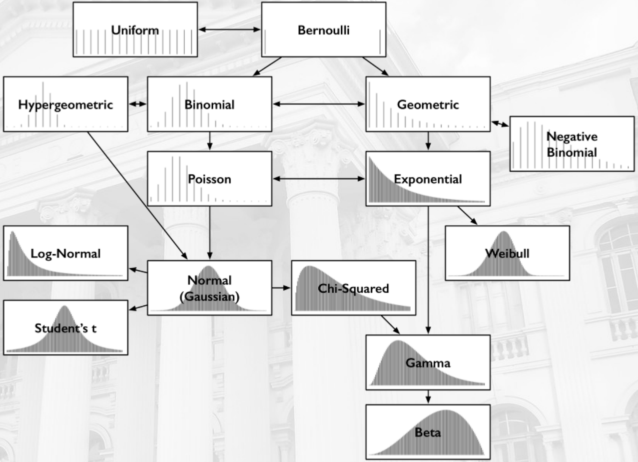

Figura 1. Modelos de distribuição de probabilidades. Fonte: https://bit.ly/2JkhFDr.

Revisão de variáveis aleatórias

Variável aleatória

Uma **variável aleatória** é uma variável que assume valores númericos associados aos resultados de um experimento aleatório, onde um (e apenas um) valor numérico é associado com cada ponto do espaço amostral.

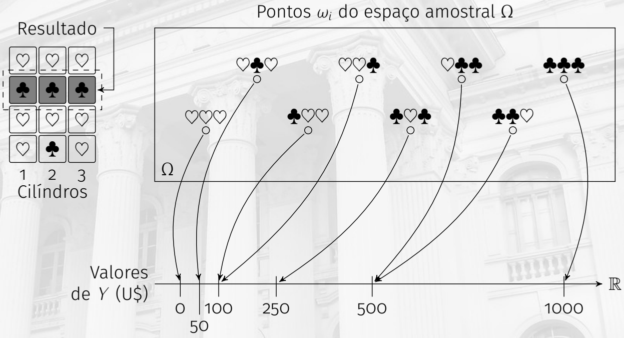

Figura 2. Exemplo de definição de variável aleatória.

Tipos de variáveis aleatórias

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

Discretas

- Variáveis aleatórias discretas assumem um **contável** número de valores dentro de um intervalo.
- Portanto, o **suporte é discreto**.
- Aqui vamos indicar suporte, conjunto de valores que a variável assume, por $y \in R_y$.

Contínuas

- Variáveis aleatórias contínuas assumem um **incontável** número de valores dentro de um intervalo.
- Portanto, o **suporte é contínuo**.
- Dessa forma $y \in R_y \subseteq \mathbb{R}$.

Função de probabilidade

A **função de probabilidade** de uma variável aleatória discreta é um gráfico, tabela ou fórmula que especifica a probabilidade associada com **cada possível valor** que a variável pode assumir.

Propriedades:

1. $p(y) \geq 0$ para todos os valores de y do suporte de Y, ou seja, $y \in R_y$.
2. $\sum_{\{y \in R_y\}} p(y) = 1$.

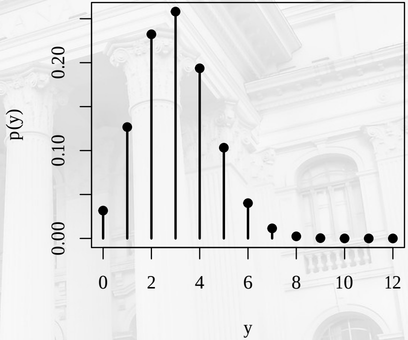

Figura 3. Gráfico da distribuição de uma variável aleatória discreta.

Função densidade de probabilidade

A **função de densidade de probabilidade** de uma variável aleatória contínua é uma fórmula que especifica a probabilidade associada com a ocorrência de **valores dentro de um intervalo** $(a, b) \in R_y$.

Propriedades:

1. $f(y) \geq 0$ para todos os valores de y do suporte de Y, ou seja, $\{y \in R_y\}$.
2. $\int_{\{y \in R_y\}} f(y)\,\mathrm{d}y = 1$.
3. $\mathrm{P}(a < Y \leq b) = \int_a^b f(y)\,\mathrm{d}y$.

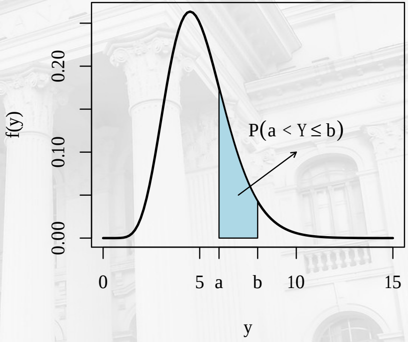

Figura 4. Gráfico da função de densidade de uma variável aleatória contínua.

Função de distribuição acumulada

A função de distribuição acumulada, $F(y)$, de uma variável aleatória determina

$$F(y) = \mathrm{P}(Y \leq y).$$

Conforme o suporte, é assim definida

- Discreta

$$F(y) = \sum_{\{u \in R_y : u \leq y\}} p(u).$$

- Contínua

$$F(y) = \int_{\{u \in R_y : u \leq y\}} f(u)\,\mathrm{d}u = \int_{-\infty}^{y} f(u)\,\mathrm{d}u.$$

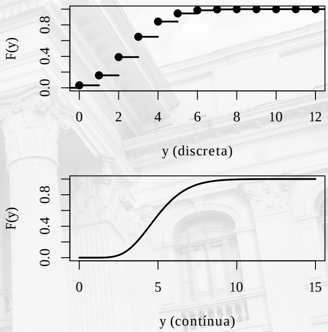

Figura 5. Gráfico da função de distribuição acumulada de uma variável aleatória discreta (topo) e contínua (base).

Média e variância

Discretas

- Valor esperado (média)

$$\mu = \mathrm{E}(Y) = \sum_{\{y \in R_y\}} y \cdot p(y).$$

- Variância

$$\sigma^2 = \mathrm{V}(Y) = \sum_{\{y \in R_y\}} (y - \mu)^2 \cdot p(y).$$

Contínuas

- Valor esperado (média)

$$\mu = \mathrm{E}(Y) = \int_{\{y \in R_y\}} y \cdot f(y)\,\mathrm{d}y.$$

- Variância

$$\sigma^2 = \mathrm{V}(Y) = \int_{\{y \in R_y\}} (y - \mu)^2 \cdot f(y)\,\mathrm{d}y.$$

Distribuição de probabilidade para várias variáveis

- Distribuições de probabilidade podem considerar $k \geq 2$ variáveis aleatórias **simultaneamente**.
- Função de probabilidade ou densidade de probabilidade **conjunta**.
- Função de distribuição **marginal** de cada variável aleatória.
- Função de distribuição **condicional** de uma variável aleatória.
- **Independência** entre variáveis aleatórias.

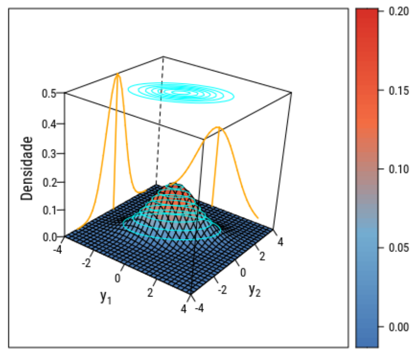

Figura 6. Exemplo da função de densidade conjunta de duas variáveis aleatórias contínuas.

Motivação sobre os modelos probabilísticos

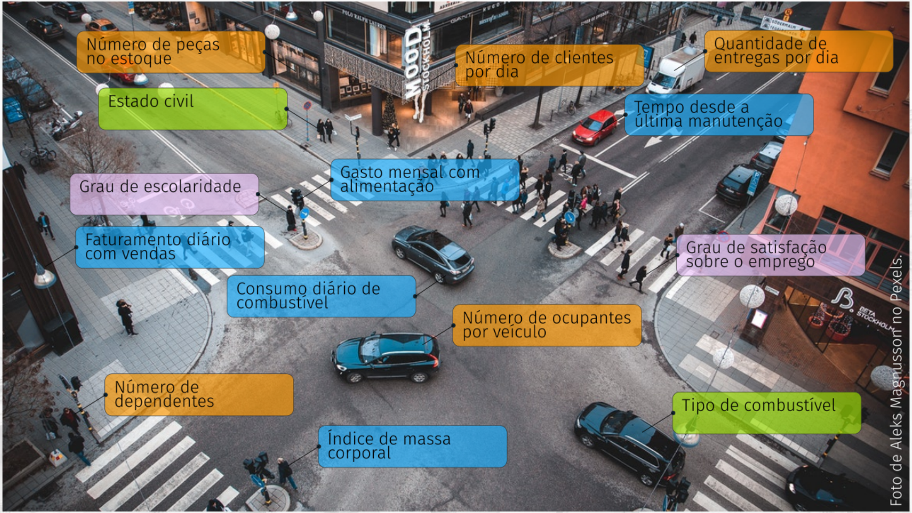

Figura 7. Variáveis aleatórias de uma típica cena cotidiana.

Obter a distribuição de probabilidades

Considere o experimento aleatório de retirar $r = 3$ bolas sem reposição de uma caixa com $m = 4$ bolas brancas e $n = 3$ bolas pretas. Determine a distribuição de probabilidade de Y sendo o número de bolas brancas retiradas da caixa.

Tabela 1. Distribuição de probabilidades da variável aleatória Y.

y	Abordagem 1	Abordagem 2	$p(y)$
0	$(3/7) \cdot (2/6) \cdot (1/5)$	$1/\binom{7}{3}$	0.0286
1	$\vdots$	$\vdots$	$\vdots$
2	$\vdots$	$\vdots$	$\vdots$
3	$\cdots$	$\binom{4}{3}/\binom{7}{3}$	0.1143

Figura 8. Bolas brancas e pretas.

Generalizar a função de probabilidades

Tabela 2. Distribuição de probabilidades da variável aleatória Y.

y	Expressão	$p(y)$
0	$\binom{4}{0} \cdot \binom{3}{3} / \binom{7}{3}$	0.0286
1	$\binom{4}{1} \cdot \binom{3}{2} / \binom{7}{3}$	0.3429
2	$\binom{4}{2} \cdot \binom{3}{1} / \binom{7}{3}$	0.5142
3	$\binom{4}{3} \cdot \binom{3}{0} / \binom{7}{3}$	0.1143
Soma		1

Então,temos uma **fórmula geral** para determinar as probabilidades

$$p(y) = \underbrace{\binom{m}{y}}_{(1)} \cdot \underbrace{\binom{n}{r-y}}_{(2)} \bigg/ \underbrace{\binom{m+n}{r}}_{(3)}$$

em que m, n e r foram os **parâmetros** acima mencionados.

Significado dos termos: (1) formas de arranjar bolas brancas; (2) formas de arranjar bolas pretas; e (3) número de arranjos possíveis com todas as bolas.

Situações práticas análogas

1. Probabilidade de ter 3 (y) peças defeituosas em uma amostra de tamanho 10 (r) peças de um lote com 1000 ($m + n$) peças que possui 100 (m) defeituosas:

$$p(3) = \binom{100}{3} \cdot \binom{1000-100}{10-3} / \binom{1000}{10} = 0.0569.$$

2. Probabilidade de adivinhar quais são as 4 ($y = r = m$) caixas que possuem um presente em 8 ($m + n$) caixas:

$$p(4) = \binom{4}{4} \cdot \binom{4}{0} / \binom{8}{4} = 0.0143.$$

3. Probabilidade de selecionar 2 ($y = r$) alunos numa turma de 50 ($m + n$) e eles serem do grupo de 5 (m) alunos selecionados na aula anterior:

$$p(2) = \binom{5}{2} \cdot \binom{45}{0} / \binom{50}{2} = 0.0082.$$

Estimar quantidades desconhecidas

Biólogos foram a campo, capturaram e marcaram 21 macacos (m) em uma reserva ecológica. Após 1 mês, eles retornaram e fizeram outra captura igual à primeira. Dos 49 (r) macacos capturados, 5 (y) apresentam a marca. Qual o tamanho da população de macacos ($m + n$)?

$$m = 21, \quad r = 49, \quad y = 5, \quad n = ?$$

E se n fosse... qual a probabilidade de observar este resultado $y = 5$?

Figura 9. Um filhote de primata. Foto de Pixabay no Pexels.

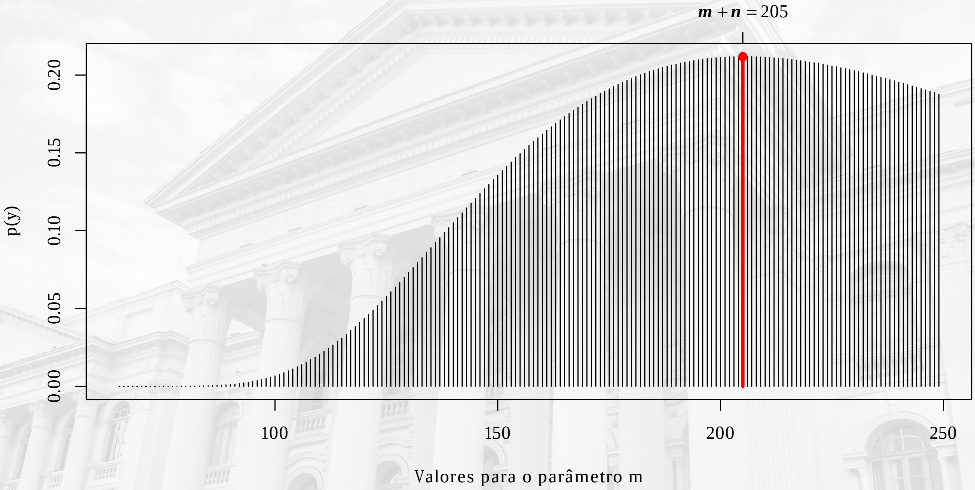

Figura 10. Gráfico da probabilidade de $P(Y = 5)$ em função de valores para o parâmetro m.

Questões importantes

Inferência

- Porque não $m + n = 200$ ao invés de $m + n = 205$?
- Se fosse outra amostra, o $m + n$ estimado teria outro valor?
- De que tamanho é o erro de estimação?
- Por que não regra de três?

$$
\begin{aligned}
5 &\text{ ——————— } 21\\
49 &\text{ ——————— } m + n
\end{aligned}
$$

$$m + n = 21 \cdot 49/5 = 205.8.$$

Suposições

- E se a chance de capturar os macacos não for igual?
- E se os macacos marcados se tornarem mais dóceis ou mais ariscos?
- E se os macacos andarem em bandos?
- Usar armadilhas é que tipo de amostragem: probabilística ou não probabilística?

Testar hipóteses

- Aconteceu com R. A. Fisher e Muriel Bristol.
- A Muriel declarou saber **discriminar a bebida** segundo a ordem em que chá e leite eram adicionados à xícara.
- **Experimento** proposto:
  - 8 xícaras ($m + n$) iguais.
  - 4 de cada tipo ($m = n$) servidas aleatoriamente.
- Resposta: número de xícaras corretamente classificadas como "**leite após chá**" (y) das 4 xícaras indicadas (r).

Tabela 3. Probabilidade de acerto casual das xícaras servidas com chá antes de leite.

y	$\mathrm{P}(Y = y)$	$F(y)$
0	0.0143	0.0143
1	0.2286	0.2429
2	0.5143	0.7571
3	0.2286	0.9857
4	0.0143	1.0000

- A chance de **acerto casual** é $< 2\%$.
- Qual a conclusão se ela acertar as 4 xícaras?

Aspectos importantes de modelos probabilísticos

- Modelos de distribuição de probabilidade fornecem **fórmulas gerais** para tratar várias situações similares.
- Permitem determinar **probabilidades** de eventos.
- Servem para fazer **estimação** de parâmetros.
- Servem para realizar **testes de hipótese**.
- São usados para **compreender/acomodar o efeito** de outras variáveis sobre o comportamento de uma variável.
- São empregados para fazer **previsão**.
- Fazem uso de **suposições** que podem ou não ser atendidas.
- Podem ser **flexibilizados** para incluir características específicas de cada fenômeno.

Componentes das distribuições de probabilidade

Os componentes

Parâmetro m.
Espaço paramétrico: $m \geq 1$.

Parâmetro n.
Espaço paramétrico: $n \geq 1$.

Função de probabilidade ou de densidade de probabilidade.

$$p(y; m, n, r) = \binom{m}{y} \cdot \binom{n}{r-y} \Big/ \binom{m+n}{r}.$$

Parâmetro r.
Espaço paramétrico: $r \geq 1$.

Variável aleatória Y.
Suporte: $R_y = \{y \in \mathbb{N} : \max(0, r-n) \leq y \leq \min(r, m)\}$.

Figura 11. Componentes de um modelo de distribuição de probabilidades.

Definições

Variável aleatória

- **Variável aleatória**: resultado numérico da observação de um fenômeno aleatório.
- **Suporte**: conjunto de valores que a variável aleatória Y pode assumir.
- Conforme o suporte, as variáveis aleatórias são:
  - Discretas.
  - Contínuas.

Parâmetro

- **Parâmetro**: variável que é parte da distribuição de probabilidades.
- **Espaço paramétrico**: conjunto de valores válidos para o parâmetro da distribuição.
- Conforme o espaço paramétrico, os parâmetros são:
  - Discretos.
  - Contínuos.
- A distribuição pode ter qualquer quantidade de parâmetros, até mesmo nenhum.

As principais distribuições de probabilidade

As principais distribuições de probabilidade

Discretas

1. Uniforme Discreta.
2. Bernoulli.
3. Binomial.
4. Poisson.
5. Geométrica.
6. Binomial negativa.
7. Hipergeométrica*.

Contínuas

1. Uniforme Contínua.
2. Exponencial.
3. Normal.
4. Lognormal.
5. Gama.
6. Weibull.
7. Beta.

*Foi o modelo probabilístico usado como fio condutor na motivação.

Considerações finais

Considerações finais

Revisão

- Motivação sobre os distribuições de probabilidade.
- Componentes de uma distribuição.
- As principais distribuições.

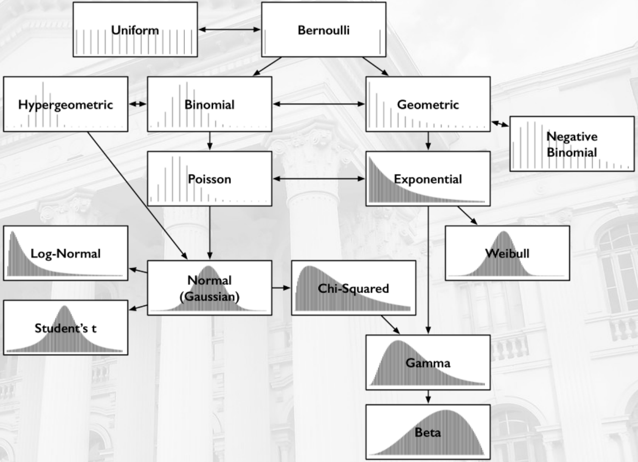

Figura 12. Modelos de distribuição de probabilidades. Fonte: https://bit.ly/2JkhFDr.