Intro à Probabilidade

Eduardo Horta

Exemplos!

Aula 11

Exemplo: Testes de Hipóteses

Exemplo (Testes de Hipóteses)

- Consideremos o seguinte experimento: lançar uma moeda 10 vezes e anotar o número total de caras obtidas.
- Suponha que o experimento foi executado e que observamos a ocorrência do evento B = "8 caras nos 10 lançamentos''.
- **Teoria da Probabilidade**:
  "se $\mathbb{P}(\text{cara}) = p$, qual o valor de $\mathbb{P}(B)$?''
- **Estatística** (ou melhor, *inferência* Estatística):
  "Essa moeda é honesta?''

Exemplo (Testes de Hipóteses)

- Utilizando resultados vistos até aqui, podemos concluir[1] que

$$\mathbb{P}(B) = \binom{10}{8} p^8 (1-p)^2 \tag{1}$$

- No caso em que $p = 1/2$, obtemos

$$\mathbb{P}(B) = 0.04394531$$

- Podemos interpretar o resultado acima da seguinte maneira: *se a moeda efetivamente for equilibrada, então o evento B que acabamos de observar é relativamente raro – sua chance de ocorrência é menor do que 5%.*

[1] **Exercício**: justifique a fórmula dada na expressão (1).

Exemplo (Testes de Hipóteses)

- Isso poderia nos levar a concluir que a moeda não é equilibrada, pois observar a ocorrência do evento B ao realizar o experimento seria extremamente improvável se essa hipótese fosse verdadeira.
- Agora, num grande número de lançamentos de uma moeda equilibrada (digamos, n lançamentos), sempre vamos observar *algum* resultado dentre os possíveis, isto é, observaremos alguma *string* particular constituída de **caras** e **coroas**, sendo que cada uma dessas *strings* – quando tomada individualmente – possui probabilidade de ocorrência igual a 2^{-n} (extremamente pequena para valores moderados de n).

Exemplo (Testes de Hipóteses)

- Quer dizer, dependendo das características do experimento, é possível que qualquer resultado particular, quando tomado individualmente, tenha probabilidade extremamente pequena.
- Olhando por outro lado, em qualquer experimento sempre observamos a ocorrência do evento Ω = "um dos resultados possíveis de fato ocorreu' ' (o que é óbio!), e sabemos que $\mathbb{P}(\Omega) = 1$ independentemente do valor de p.

Exemplo (Testes de Hipóteses)

- Os comentários acima ilustram que, em qualquer experimento, uma vez que observamos o resultado ω podemos responder se qualquer evento $A \subseteq \Omega$ ocorreu ou não (isto é, se $\omega \in A$ ou se $\omega \notin A$), e que podemos ter ao mesmo tempo a ocorrência de eventos de baixa e de alta probabilidades.

Exemplo (Testes de Hipóteses)

- Vamos agora propor um evento E que, de alguma maneira, capture a noção de uma *ocorrência atípica*, ou ainda, de uma *observação extrema*.
- Tomemos o seguinte evento:
  E = *"ocorrência de 8 ou mais* ***caras*** *em 10 lançamentos"*.

Exemplo (Testes de Hipóteses)

- Nesse caso, temos que E é o evento "ocorrência de 8 **caras** ou ocorrência de 9 **caras** ou ocorrência de 10 **caras** nos 10 lançamentos''. Sendo assim, obtemos

$$\mathbb{P}(E) = \binom{10}{8} p^8 (1-p)^2 + \binom{10}{9} p^9 (1-p) + \binom{10}{10} p^{10}$$

- No caso em que $p = 1/2$, ficamos com $\mathbb{P}(E) = 0.0546875$.

Exemplo (Testes de Hipóteses)

- Podemos determinar, *antes de executar o experimento*, que vamos considerar como extremos quaisquer resultados pertencentes ao evento E, isto é, podemos decidir que consideraremos como uma ocorrência *atípica* a observação de 8 ou 9 ou 10 **caras** nos 10 lançamentos.
- Uma vez executado o experimento, consideramos seu resultado ω como sendo *extremo* ou *atípico* se $\omega \in E$.

Exemplo (Testes de Hipóteses)

- É importante notar aqui que, utilizando apenas a informação disponível obtida a partir do experimento, *é impossível decidir com total certeza se essa moeda é equilibrada ou não*.
- O que temos é *alguma evidência* de que ela não é equilibrada.

Exemplo: Smallpox in Boston, 1721

Exemplo (Smallpox in Boston, 1721)

- A tabela[2] abaixo representa uma amostra de 6224 indivíduos que foram expostos ao vírus da varíola em 1721 em Boston, MA.
- Na época, os médicos acreditavam que a inoculação, que envolve expor a pessoa a uma forma controlada do vírus, poderia reduzir a *likelihood* de morte.

		inoculado		
		`sim`	`não`	Total
desfecho	`viveu`	238	5136	5374
	`morreu`	6	844	850
	Total	244	5980	6224

Tabela 1: Tabela de contingência para o *data set* `smallpox`.

[2] Fonte: https://www.openintro.org

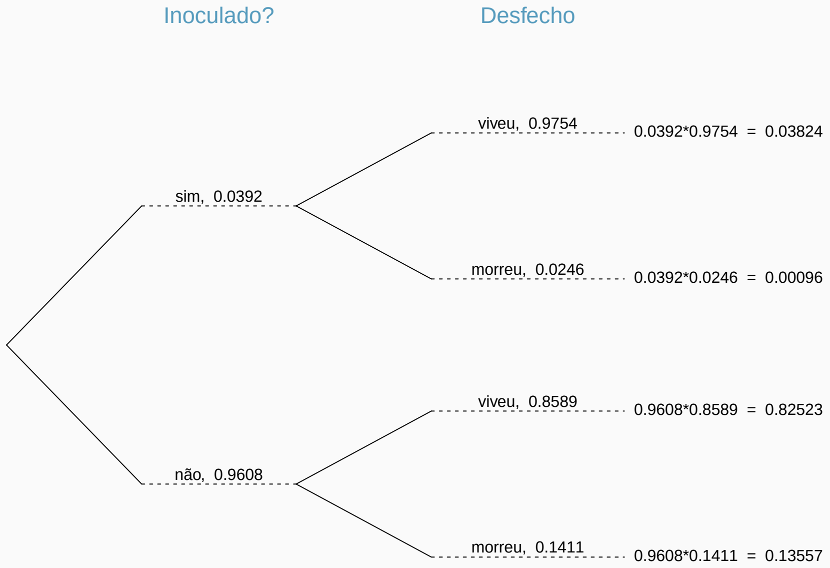
Exemplo (Smallpox in Boston, 1721)

		inoculado		
		`sim`	`não`	Total
desfecho	`viveu`	0.0382	0.8252	0.8634
	`morreu`	0.0010	0.1356	0.1366
	Total	0.0392	0.9608	1.0000

Tabela 2: Tabela de proporções para o *data set* `smallpox`.

- **Exercício**: Escreva, em notação formal, a probabilidade de morte de uma pessoa selecionada ao acaso, dentre as que não foram inoculadas. Qual o valor numérico dessa probabilidade?

Exemplo (Smallpox in Boston, 1721)



Exemplo (Smallpox in Boston, 1721)

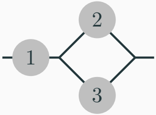
- **Importante**: Os cidadãos de Boston decidiam por conta própria se desejavam receber ou não a inoculação do vírus.
- Com base na informação acima, devemos considerar esse um estudo observacional ou experimental?
- Quais fatores de confusão potencialmente influenciam a decisão de receber ou não a inoculação? E o desfecho?

Aula 12

Exemplo: Confiabilidade de Sistemas

Exemplo (Confiabilidade de Sistemas)

- Vamos considerar o sistema representado pela figura abaixo



- O sistema é constituído de três componentes, denotados (muito apropriadamente) por 1, 2 e 3, respectivamente.
- Uma vez iniciado o sistema, cada componente pode funcionar ou falhar, sendo que o componente i funciona com probabilidade $p_i \in (0, 1]$, $i = 1, 2, 3$,

Exemplo (Confiabilidade de Sistemas)

- Os componentes funcionam independentemente; isto é se denotarmos por E_i o evento “componente i funciona’’, $i = 1, 2, 3$, então nosso modelo probabilístico para descrever o funcionamento do sistema é o seguinte:

$$\begin{aligned}
&\mathbb{P}(E_i) = p_i, && i = 1, 2, 3;\\
&\mathbb{P}(E_i \cap E_j) = p_i p_j, && \text{se } i \neq j;\\
&\mathbb{P}(E_1 \cap E_2 \cap E_3) = p_1 p_2 p_3.
\end{aligned}$$

Exemplo (Confiabilidade de Sistemas)

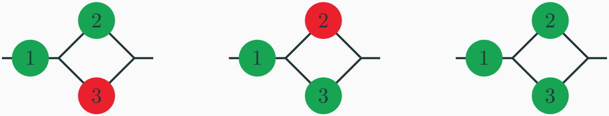
- O sistema **funciona** se, e somente se, for possível percorrer um caminho, da esquerda para a direita, sem passar por nenhuma componente que tenha falhado.
- Isto é: o sistema funciona se, e somente se, ocorrer um dos eventos $E_1 \cap E_2$, ou $E_1 \cap E_3$ ou $E_1 \cap E_2 \cap E_3$: denotando por E o evento "o sistema funciona' ', então

$$\begin{aligned} E &= (E_1 \cap E_2) \cup (E_1 \cap E_3) \cup (E_1 \cap E_2 \cap E_3) \\ &\overset{*}{=} (E_1 \cap E_2) \cup (E_1 \cap E_3) \\ &\overset{*}{=} E_1 \cap (E_2 \cup E_3) \end{aligned}$$

- **Exercício**: Justifique as igualdades marcadas com $*$ acima.

Exemplo (Confiabilidade de Sistemas)

- As 3 figuras abaixo ilustram os casos em que o sistema **funciona**.



Exemplo (Confiabilidade de Sistemas)

- O número $\mathbb{P}(E)$ é dito a **confiabilidade do sistema**, isto é, a confiabilidade de um sistema é a probabilidade desse sistema funcionar.
- **Exercício**: calcule a confiabilidade do sistema acima.

Exemplo: Interpretando probabilidades condicionais

Exemplo (Interpretando probabilidades condicionais[3])

- Suponha que, em uma população constituída de 1 milhão de indivíduos, 1% das pessoas estão contaminadas com um vírus.
- Um teste para detectar a doença tem probabilidade 10% de resultar em um **falso negativo**, quer dizer, a chance de o teste detectar corretamente que uma pessoa contaminada tem a doença é de 90%.
- O mesmo teste tem probabilidade 10% de resultar em um **falso positivo**.
- Suponha que uma pessoa foi selecionada ao acaso dessa população, e que o teste indicou **positivo**.
- Nessa situação, o indivíduo recém testado se pergunta: qual será a chance de ele ter a doença?

[3] Fonte: Williams, *Wheighing the Odds*

Exemplo (Interpretando probabilidades condicionais)

- **Resposta errada**: 90%. De fato, como essa pessoa foi selecionada *ao acaso* de uma população em que apenas 10K indivíduos estão contaminados, é muito mais provável que ele não tenha a doença.
- **Resposta certa**: suponha que todas as pessoas da população decidem fazer o teste. Então, das 990K pessoas sadias, em torno de $990\text{K} \times 10\% = 99\text{K}$ testarão positivo.
- Semelhantemente, das 10K pessoas contaminadas, em torno de $10\text{K} \times 90\% = 9\text{K}$ testarão positivo.
- Logo, da população de 1 milhão de pessoas, em torno de $99\text{K} + 9\text{K} = 108\text{K}$ testarão positivo, das quais somente 9K realmente têm a doença.

Exemplo (Interpretando probabilidades condicionais)

- Formalizando, ficamos assim: se o experimento consiste em sortear, ao acaso, uma pessoa da população e executar o teste, podemos definir os eventos
  1. D = "pessoa selecionada tem a doença";
  2. $S = D^{\complement}$ = "pessoa selecionada saudável";
  3. T_+ = "resultado do teste positivo";
- Sabemos que $\mathbb{P}(D) = 0.01$, que $\mathbb{P}(T_+ \,|\, D) = 0.9$, e que $\mathbb{P}(T_+ \,|\, S) = 0.1$.
- Pelo Teorema de Bayes,

$$\mathbb{P}(D \,|\, T_+) = \frac{\mathbb{P}(T_+ \,|\, D)\mathbb{P}(D)}{\mathbb{P}(T_+ \,|\, D)\mathbb{P}(D) + \mathbb{P}(T_+ \,|\, S)\mathbb{P}(S)} = \frac{1}{12}.$$

Exemplo: Expansões diádicas aleatórias

Exemplo (Expansões diádicas aleatórias)

- **Preliminares**: todo número real ω pertencente ao intervalo $[0, 1]$ pode ser representado na forma

$$\omega = 0.\omega_1\omega_2\omega_3 \cdots = \sum_{k=1}^{\infty} \frac{\omega_k}{10^k} \tag{2}$$

onde os *dígitos* ω_k pertencem ao conjunto $\{0, 1, 2 \ldots, 9\}$.

Exemplo (Expansões diádicas aleatórias)

- Por exemplo, o número $\pi/10$ pode ser representado como

$$\frac{\pi}{10} = 0.3141593\cdots = \frac{3}{10} + \frac{1}{100} + \frac{4}{1000} + \frac{1}{10^4} + \frac{5}{10^5} + \frac{9}{10^6} + \frac{3}{10^7} + \cdots$$

- Nesse caso, temos $\omega_1 = 3$, $\omega_2 = 1$, $\omega_3 = 4$, etc.

Exemplo (Expansões diádicas aleatórias)

- A expansão dada em (2) é a chamada **expansão decimal** de ω.
- Essa representação é essencialmente única, exceto nos caso em que a expansão é terminada por zeros ou noves.
- Por exemplo,

$$1 = 0.9999\ldots, \qquad 0.5 = 0.49999\ldots,$$

  etc.
- Por simplicidade, vamos considerar sempre a representação terminada em noves (exceto para o número 0).

Exemplo (Expansões diádicas aleatórias)

- Uma representação alternativa é a conhecida *representação binária*[4] de um número $\omega \in [0, 1]$, dada por

$$\omega = 0.\omega_1\omega_2\omega_3 \cdots = \sum_{k=1}^{\infty} \frac{\omega_k}{2^k} \tag{3}$$

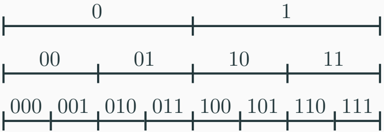
  onde os *dígitos* ω_k pertencem ao conjunto $\{0, 1\}$.
- Nessa representação, o número $\pi/10$ se escreve na forma

$$\frac{\pi}{10} = 0.0101000 = \frac{0}{2} + \frac{1}{4} + \frac{0}{8} + \frac{1}{16} + \frac{0}{32} + \frac{0}{64} + \frac{0}{128} + \cdots$$

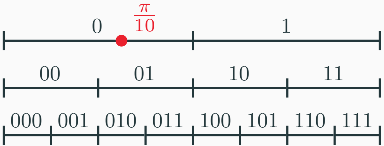
[4] Também dita a representação **diádica** de ω.

Exemplo (Expansões diádicas aleatórias)

Intervalos diádicos



Exemplo (Expansões diádicas aleatórias)



Exemplo (Expansões diádicas aleatórias)

- A pergunta agora é: *o que acontece se selecionamos os dígitos da expansão binária de forma aleatória*?
- Uma maneira de fazer isso é a seguinte: considere uma moeda perfeitamente equilibrada tendo o dígito 1 impresso em uma de suas faces (correspondendo a **cara**) e o dígito 0 na outra (correspondendo a **coroa**).
- Suponha que dispomos de *tempo infinito* e que vamos jogar a moeda infinitas vezes, com cada lançamento independente dos anteriores.
- A ideia é obter um número real no intervalo $[0, 1]$ formando a sua expansão diádica de acordo com os resultados sucessivos do lançamento da moeda.

Exemplo (Expansões diádicas aleatórias)

- Por exemplo, se nos primeiros 4 lançamentos observarmos **cara**, **cara**, **coroa**, **cara**, isto é, 1101, então sabemos que o número ω que estamos sorteando se situa no intervalo aberto à esquerda em $^{13}/_{16}$ e fechado à direita em $^{14}/_{16}$.

- Para determinar com exatidão o número que estamos sorteando, seria preciso continuar lançando a moeda indefinidamente (e assim obteríamos uma aproximação cada vez melhor do valor de ω).

Exemplo (Expansões diádicas aleatórias)

- Como qualquer string (concatenação) de n **caras** e **coroas** tem probabilidade 2^{-n}, vemos que a probabilidade do número sorteado se situar em um intervalo diádico qualquer é igual ao comprimento desse intervalo: por exemplo,

$$\mathbb{P}\Big(0, {}^{1}\!/_{2}\Big] = \frac{1}{2}, \qquad \mathbb{P}\Big({}^{24}\!/_{32}, {}^{25}\!/_{32}\Big] = \frac{1}{32},$$

  etc.

Exemplo (Expansões diádicas aleatórias)

- Tomando uniões de intervalos diádicos disjuntos, podemos calcular a probabilidade de alguns outros intervalos, por exemplo

$$\begin{aligned}\mathbb{P}\left(1/4, 3/4\right] &= \mathbb{P}\left\{\left(1/4, 2/4\right] \cup \left(2/4, 3/4\right]\right\} \\ &= \mathbb{P}\left(1/4, 2/4\right] + \mathbb{P}\left(2/4, 3/4\right] \\ &= \frac{1}{2}\end{aligned}$$

Exemplo (Expansões diádicas aleatórias)

- De fato, tomando intersecções e uniões de intervalos diádicos, dados $0 \leq a < b \leq 1$ podemos calcular a probabilidade de *qualquer* intervalo $(a, b] \subseteq [0, 1]$:

$$\mathbb{P}(a, b] = b - a.$$

- O fato de que vale a igualdade acima não é tão fácil de ser obtido – por ora vamos acreditar que esse seja o caso.

Exemplo (Expansões diádicas aleatórias)

- A conclusão é a seguinte: se utilizamos os resultados de sucessivos lançamentos de uma moeda equilibrada para determinar os dígitos de uma expansão diádica, então o número ω assim obtido será uniformemente distribuído no intervalo $[0, 1]$.
- Quer dizer, a probabilidade de ω pertencer a um intervalo qualquer é dada pelo comprimento do intervalo.