Estatística Básica

Prof. Dr. Walmes Marques Zeviani
Prof. Dr. Paulo Justiniano Ribeiro Jr

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

Laboratório de
Estatística e Geoinformação
www.leg.ufpr.br

Estrutura do curso

O curso é divido em 3 partes:

Probabilidade noções de cálculo de probabilidade, variáveis aleatórias e propriedades, modelos para distribuição de variáveis aleatórias;

Estatística descritiva métodos para amostragem, organização, tratamento, análise, apresentação e interpretação de dados. Emprego de estatísticas descritivas e representações gráficas.

Inferência estatística ferramentas para fazer inferências baseado em dados amostrais. Métodos de estimação pontual e intervalar, testes de hipótese e predição.

Introdução à Estatística

O que estatística?

O que estatística?

A estatística utiliza-se das **teorias probabilísticas** *para explicar a frequência da ocorrência de eventos*

Serve para **modelar** *a aleatoriedade e a incerteza de forma a estimar ou possibilitar a* **previsão de fenômenos** *futuros.*

É uma ciência que se dedica à **coleta**, **organização**, **análise**, **apresentação** *e* **interpretação** *de dados.*

Auxilia em **tirar conclusões** *sobre as características das fontes de onde dados foram retirados para melhor* **compreendê-los**.

É uma tecnologia quantitativa para a ciência que permite avaliar as **incertezas** *e os seus efeitos no planejamento e interpretação de experiências e de observações de* **fenômenos** *da natureza e da sociedade.*

O que estatística?

Indispensável para a **tomada de decisões** *sob condições de incerteza, sob o menor* **risco** *possível.*

A estatística tem sido utilizada na pesquisa científica, para a **otimização** *de recursos econômicos, para o aumento da qualidade e produtividade, na otimização em análise de decisões, em questões judiciais, previsões e em muitas outras áreas.*

Por que estudar estatística?

- Impossibilidade de estudar a população e ter que tirar conclusões a partir de partes da mesma;
- Aumento da capacidade de registro de dados que precisam ser compreendidos;
- Expansão do conhecimento científico, das áreas de pesquisa e dos instrumentos de investigação;
- Necessidade de compreensão dos fenômenos naturais e sociais, de otimização de recursos, planejamento de atividades, redução de riscos, de previsão de resultados para correta tomada de decisão;

Por que estudar estatística?

A Estatística pode ser pensada como a **ciência de aprendizagem a partir de dados**.

Vivemos na **era da informação** *e a Estatística possui as ferramentas necessárias para melhor compreender a informação.*

Probabilidades

Objetivos

- Entender e descrever espaços amostrais e eventos para experimentos aleatórios;
- Interpretar probabilidades e usar probabilidades de resultados para obter probabilidades de eventos;
- Calcular probabilidades e eventos conjuntos, como união e intersecção de eventos individuais;
- Calcular e interpretar probabilidades condicionais;
- Determinar a independência de eventos;
- Compreender e aplicar o teorema de Bayes;

Definições

Experimento aleatório: (ou fenômeno aleatório) é a situação na qual observamos um sistema que produz resultados que não poder sem previstos, muito embora seja repetido toda vez da mesma maneira. Exemplo: lançar um dado/moeda, pesar um fruto a.a., etc.

Espaço amostral (Ω): Conjunto de todos os resultados possíveis de um experimento aleatório. O Ω pode conter um número finito ou infinito de pontos. Exempo: {cara,coroa}, {1,2,3,4,5,6}, $\mathbb{N}$, $\mathbb{R}$.

Ponto amostral (ω): são elementos de um espaço amotral. Exemplo: ω_1 = cara, ω_2 = coroa.

Evento: é um subconjunto de pontos do espaço amostral de um experimentos aleatório. Exemplo: A = "sair face par", B = "sair face menor que 3".

Exemplos

Experimento: lançar o dado e observar o resultado da face.

Espaço amostral: $\Omega = \{1,2,3,4,5,6\}$.

Pontos amostrais: $\omega_1 = 1,\ \omega_2 = 2,\ \ldots,\ \omega_6 = 6$.

Eventos: A = "sair face par", $B = \{\omega : \omega \leq 4\}$.

Exemplos

Experimento: retirar uma carta de um baralho de 54 cartas.

Espaço amostral: $\Omega = \{\clubsuit A, \clubsuit 2, \ldots, \heartsuit A, \ldots, \spadesuit A, \ldots, \diamondsuit J, \diamondsuit Q, \diamondsuit K\}$.

Pontos amostrais: $\omega_1 = \clubsuit A$, $\omega_2 = \clubsuit 2$, ..., $\omega_{54} = \diamondsuit K$.

Eventos: A = "sair um ais", B = "sair uma letra", C = "sair carta de $\clubsuit$".

Exemplos

Experimento: medir a circunferência de uma fruta a.a.

Espaço amostral: $\Omega = \mathbb{R}^+$.

Pontos amostrais: espaço amostral é infinito.

Eventos: A = "circunferência menor que 10cm", B $= \{x : x \geq 10\text{cm}\}$.

Operações com eventos

União: é o evento que consiste da união de **todos** os pontos amostrais dos eventos que a compõem. Denotamos a união do evento A com B por $A \cup B$.

Interseção: é o evento composto pelos pontos amostrais **comuns** aos eventos que a compõem. Denotamos a interseção de A com B por $A \cap B$.

Complemento: é o conjunto de pontos do espaço amostral que não estão no evento. Denotamos o complemento do evento A por A^c.

Exemplo

Considere o lancamento de um dado e os eventos $A = \{1,2,3,4\}$, $B = \{\omega : \omega \leq 3\}$, $C =$ "face par", $D =$ "face primo".

- União:
$$A \cup B = \{1,2,3,4\},$$
$$A \cup C = \{1,2,3,4,6\},$$
$$A \cup D = \{1,2,3,4,5\}.$$
- Intersecção:
$$A \cap B = \{1,2,3\},$$
$$A \cup C = \{2,4\},$$
$$A \cup D = \{1,2,3\}.$$
- Complemento:
$$A^c = \{5,6\},$$
$$B^c = \{\omega : \omega > 3\},$$
$$D^c = \{4,6\}.$$

Tipos de eventos

Disjuntos: (mutuamente exclusivos) são eventos que possuem intersecção nula, ou seja, $A \cap B = \{\varnothing\}$. Denotamos a união do evento A com B por $A \cup B$.

Complementares: são eventos que a união é o espaço amostral, ou seja, $A \cup B = \Omega$. Denotamos a interseção de A com B por $A \cap B$.

Exaustivos: (disjuntos e complementares) são eventos que atendem ambas propriedades.

Axiomas da probabilidade

Probabilidade de um evento: quando o espaço amostral é discreto, a probabilidade de um evento A, denotada por $P(A)$, é a soma das probabilidades do elementos do espaço amostral (ω) que compõem A.

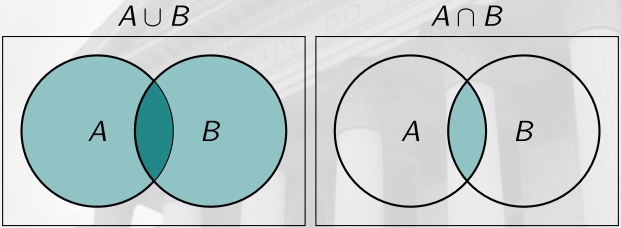
Axiomas de probabilidade

1. $P(\Omega) = 1$.
2. $0 \leq P(A) \leq 1$.

Regra da adição

Probabilidade da união entre eventos é dada por

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1}$$



Probabilidade condicional

Permite avaliar a probabilidade quando informações adicionais se tornam disponíveis ou quando certas condições são assumidas. A probabilidade de um evento B dado a realização do evento A é representado por

$$P(B|A). \tag{2}$$

Por exemplo, qual a probabilidade de sair um número par ao lançar um dado? Qual a probabilidade de sair um número par sabendo que o resultado é (ou condicional ao resultado ser) menor ou igual a 3?

Fórmula da probabilidade condicional

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ assumindo que } P(B) \neq 0. \tag{3}$$

Exemplo

Um lote 400 de produtos eletrônicos fabricados é classificado quando à apresentar defeito da superfície e defeito de funcionamento. A classificação é resumida na seguinte tabela

	Falha na superfície		
Falha no funcionamento	sim (evento S)	não	total
sim (evento F)	10	18	28
não	30	342	372
total	40	360	400

Qual a probabilidade de retirar um elemento com defeito funcional dado que apresenta defeito na superfície?

$$P(F|S) = 10/40 = 0.25. \tag{4}$$

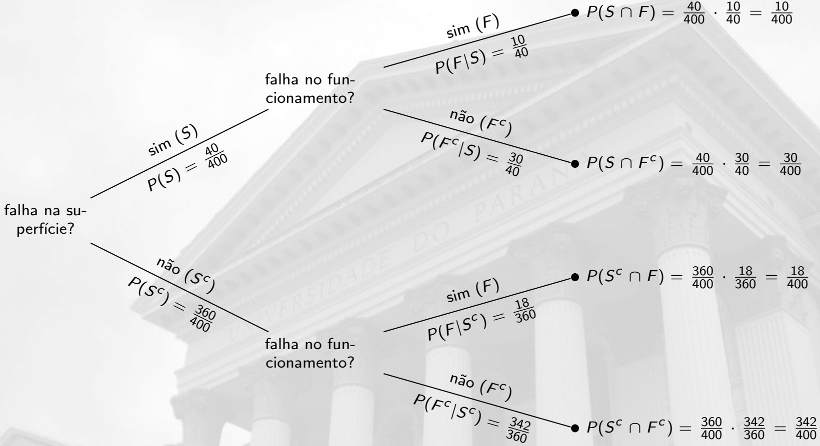
Aplicando a fórmula da probabilidade condicional verificamos que

$$P(F|S) = \frac{P(F \cap S)}{P(S)} = \frac{10/400}{40/400} = 10/40 = 0.25. \tag{5}$$

Árvore de probabilidades

É um **diagrama** para visulizar probabilidades condicionais. O diagrama inicia da esquerda e cada ramo representa um possível resultado para o primeiro evento que é não condicional, ou marginal. Em seguida, cada ramo é dividido para representar os resultados do segundo evento dado os possíveis resultados do primeiro evento, ou seja, uma probabilidade condicional. Seguindo essa regra, o diagrama pode representar probabilidades condicionais para muitos eventos. Em cada nó do diagrama a soma das probabilidades é 1.

Exemplo


falha na superfície?
sim (S)
$P(S) = \frac{40}{400}$
não (S^c)
$P(S^c) = \frac{360}{400}$
falha no funcionamento?
sim (F)
$P(F|S) = \frac{10}{40}$
não (F^c)
$P(F^c|S) = \frac{30}{40}$
falha no funcionamento?
sim (F)
$P(F|S^c) = \frac{18}{360}$
não (F^c)
$P(F^c|S^c) = \frac{342}{360}$
$P(S \cap F) = \frac{40}{400} \cdot \frac{10}{40} = \frac{10}{400}$
$P(S \cap F^c) = \frac{40}{400} \cdot \frac{30}{40} = \frac{30}{400}$
$P(S^c \cap F) = \frac{360}{400} \cdot \frac{18}{360} = \frac{18}{400}$
$P(S^c \cap F^c) = \frac{360}{400} \cdot \frac{342}{360} = \frac{342}{400}$

Regra da multiplicação

A regra da multiplicação é uma expressão derivada do conceito de probabilidade condicional. Uma vez que

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{6}$$

temos que

$$P(A \cap B) = P(B|A)P(A). \tag{7}$$

Com isso podemos obter a probabilidade de uma intersecção pelo produto de uma probabilidade marginal com uma probabilidade condicional. Isso é o que fazemos nos ultimos ramos do diagrama de árvore de probabilidade.

Fórmula para regra da multiplicação

$$P(A \cap B) = P(B|A) \cdot P(A) = P(A|B) \cdot P(B). \tag{8}$$

Exemplo

Estágios de produção

A primeira etapa de um processo de produção gera peças dentro das especificações com probabilidade 0.9. Essas peças são submetidas a segunda etapa segunda etapa de produção onde uma peça atende as especificações com probabilidade 0.95. Qual a probabilidade de ambos estágios atenderem as especificações?

Sejam A e B os eventos em que em ambos os estágio atendam às especificações. A probabilidade requerida é

$$P(A \cup B) = P(B|A) \cdot P(A) = 0.95 \cdot 0.90 = 0.855. \quad (9)$$

Regra da probabilidade total (2 eventos)

Para quaisquer eventos A e B nós temos que

$$A = (A \cap B) \cup (A \cap B^c). \tag{10}$$

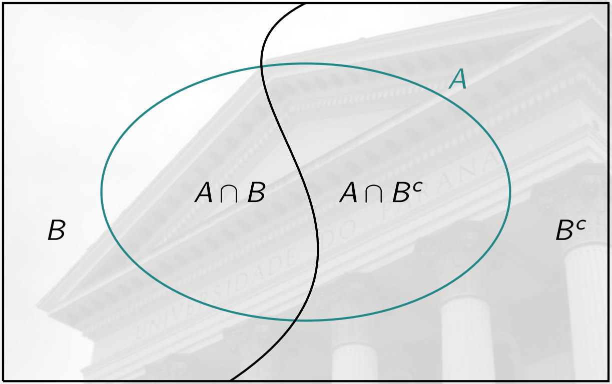
Como B e B^c são disjuntos por definição, ou seja, $B \cap B^c = \{\varnothing\}$, temos que

$$P(A) = P((A \cap B) \cup (A \cap B^c)) = P(A \cap B) + P(A \cap B^c). \tag{11}$$

Vimos anteriormente que as intersecções podem ser escritas em termos de probabilidades condicionais, assim

$$P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c) \tag{12}$$

Regra da probabilidade total (2 eventos)



Regra da probabilidade total (multiplos eventos)

Sendo os eventos B_1, B_2, B_3, ..., B_n, mutuamente exlusivos 2 a 2, temos

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n). \tag{13}$$

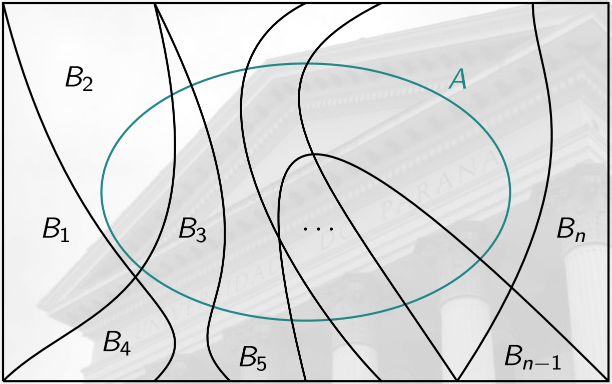
Escrevendo como probabilidade condicional, temos

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_n)P(B_n). \tag{14}$$

Usando o operador somatório, temos

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i). \tag{15}$$

Regra da probabilidade total (multiplos eventos)



Exemplo

Contaminação de semicondutores

Um *chip* falha com probabilidade de 0.10 se sujeito à alta contaminação. A falha ocorre com probabilidade de 0.005 se não sujeito a alta contaminação. O risco de um *chip* passar por alta contaminação é 0.20. Qual a probabilidade de um produto que usa um desses *chips* falhar?

Seja F o evento "o *chip* falhar" e H "ocorrer alta contaminação". Temos

$$P(F|H) = 0.10 \qquad P(F|H^c) = 0.005 \qquad P(H) = 0.2 \qquad P(H^c) = 0.8 \tag{16}$$

Assim, da equação para a probabilidade total, temos que

$$P(F) = P(F|H)P(H) + P(F|H^c)P(H^c) = 0.10 \cdot 0.20 + 0.005 \cdot 0.8 = 0.024, \tag{17}$$

que pode ser interpretada precisamente como a média ponderada das duas probabilidades de falha.

Independência

Os eventos A e B são **eventos independentes** se a ocorrência de B não altera a probabilidade de ocorrência de A, ou seja, eventos A e B são independentes se

$$P(B|A) = P(B) \text{ e também que } P(A|B) = P(A). \tag{18}$$

Com isso e a regra da probabilidade total temos que

$$P(B \cap A) = P(B|A) \cdot P(A) = P(B) \cdot P(A). \tag{19}$$

Isso significa que se dois eventos são idependentes, a probabilidade da ocorrência simultânea $P(B \cap A)$ é o produto das probabilidades marginais, $P(A)$ e $P(B)$.

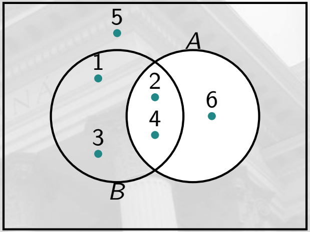
Exemplo

Lançamento de um dado

Considere o lançamento de um dado justo e os seguintes eventos

$$A = \text{“resultado é um número par”}$$
$$B = \text{“resultado é um número menor ou igual a 4”}$$

Os eventos A e B são independentes?



Temos que $P(A) = 1/2$, $P(A|B) = P(A \cap B)/P(B) = \frac{2/6}{4/6} = 1/2$. Da mesma forma, $P(B) = 2/3$, $P(B|A) = P(B \cap A)/P(A) = \frac{2/6}{3/6} = 2/3$. Portando, os eventos A e B são independentes. Saber que A ocorreu não muda a probabilidade de B ocorrer e vice-versa.

Independência de múltiplos eventos

Se existem k eventos independentes, a intersecção desses eventos é dado por

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_k) = \prod_{i=1}^{k} P(A_i). \quad (20)$$

Exemplo - circuito em série

O circuito abaixo só funciona se houver um caminho de dispositivos em funcionamento. A probabilidade de cada dispositivo é mostrada no diagrama. Suponha que os dispositivos falhem de forma independente. Qual a probabilidade do circuito funcionar?

$$\longrightarrow \boxed{0.95} \longrightarrow \boxed{0.85} \longrightarrow$$

Sejam os eventos

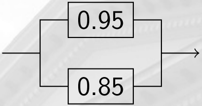
$$A = \text{"dispositivo da esquerda funciona"} \; B = \text{"dispositivo da direita funciona"}.$$

Então para o dispositivo funcionar tem-se que

$$P(A \text{ e } B) = P(A \cap B) = P(A) \cdot P(B) = 0.95 \cdot 0.85 = 0.8075. \tag{21}$$

Exemplo - circuito em paralelo

O circuito abaixo só funciona se houver um caminho de dispositivos em funcionamento. A probabilidade de cada dispositivo é mostrada no diagrama. Suponha que os dispositivos falhem de forma independente. Qual a probabilidade do circuito funcionar?



Sejam os eventos

$$A = \text{“dispositivo da esquerda funciona”} \quad B = \text{“dispositivo da direita funciona”}.$$

Então para o dispositivo funcionar tem-se que

$$\begin{aligned} P(A \text{ ou } B) &= P(A \cup B) = P(A) + P(B) - P(A \cap B) && (22)\\ &= 0.95 + 0.85 - 0.95 \cdot 0.85 = 0.9925, \text{ou ainda} && (23)\\ P(A \cup B) &= 1 - P(A^c \cap B^c) && (24)\\ &= 1 - [(1 - 0.95) \cdot (1 - 0.85)] = 0.9925. && (25) \end{aligned}$$

Teorema de Bayes

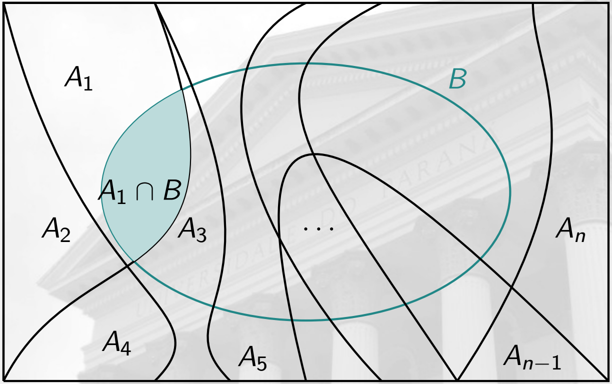
Se A_1, A_2, A_3, ..., A_n forem eventos mutuamente excludentes e exaustivos e B for qualquer evento, então

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(B|A_1) \cdot P(A_1)}{\sum_{i=1}^{n} P(A_i|B) \cdot P(A_i)}, \quad (26)$$

sendo que o mesmo vale para A_2, A_3, etc.

Rev. T. Bayes

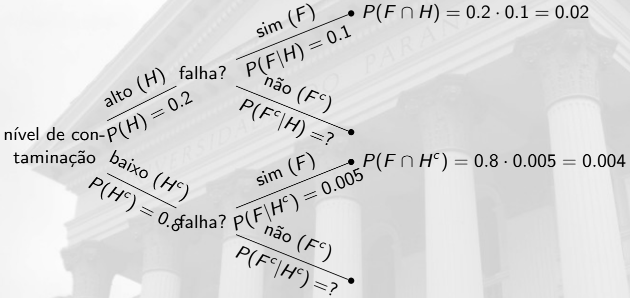
Teorema de Bayes

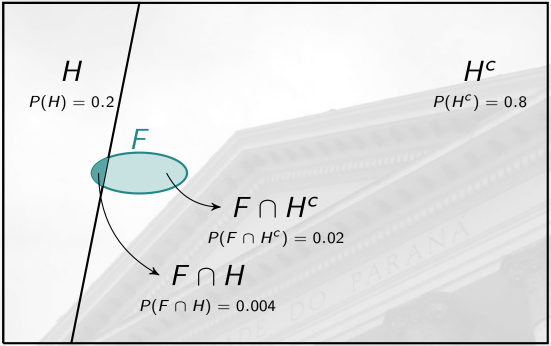


Exemplo

Semicondutores

Dado que um semicondutor falhou, qual a probabilidade de ter passado por alta contaminação durante a produção? Informações são dadas na árvore de probabilidades abaixo.





$$P(F) = P(F \cap H) + P(F \cap H^c) = 0.02 + 0.004, \text{ pela regra da prob. total} \tag{27}$$

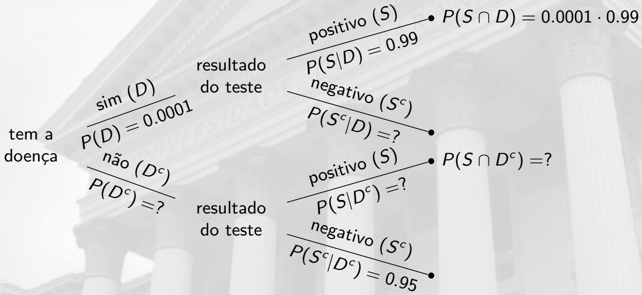
E assim

$$P(H|F) = \frac{P(F \cap H)}{P(F)} = \frac{0.02}{0.02 + 0.004} = \frac{0.02}{0.024} = 0.833\bar{3}. \tag{28}$$

Exemplo

Diagnóstico médico

A probabilidade do teste identificar corretamente alguém com a doença, dando positivo, é de 0.99. De identificar alguém sem a doença é de 0.95. A incidência da doença na população é de 0.0001. Se você fez o teste e deu positivo, qual a probabilidade de você ter a doença?



Usando propriedades de eventos disjuntos e complementares obtemos as probabilidades que faltam no diagrama de árvore de probabilidades. Assim, podemos obter o que se pede

$$P(D|S) = \frac{P(D \cap S)}{P(S)} = \frac{P(D \cap S)}{P(D \cap S) + P(D^c \cap S)} \tag{29}$$

$$= \frac{0.0001 \cdot 0.99}{0.0001 \cdot 0.99 + 0.9999 \cdot 0.05} = \frac{9.9 \cdot 10^{-5}}{0.050094} = 0.001976. \tag{30}$$

Com isso entendemos que, apesar do teste ter alta capacidade de identificar a doença, quando positivo a sua probabilidade de estar doente (quase 0.2%) ainda é muito baixa devido a baixíssima incidência da doença na população (0.01%).

Variáveis aleatórias

Objetivos

- Fazer correspondência entre valores de uma v.a. e eventos do espaço amostral;
- Determinar probabilidades a partir de funções de probabilidade, densidade de probabilidade e probabilidade acumulada;
- Calcular médias e variâncias para v.a.;
- Compreender as suposições dos modelos para distribuição de v.a.;
- Selecionar uma distribuição apropriada para representar probabilidades em uma aplicação específica;

Definição

Variável aleatória (v.a.)

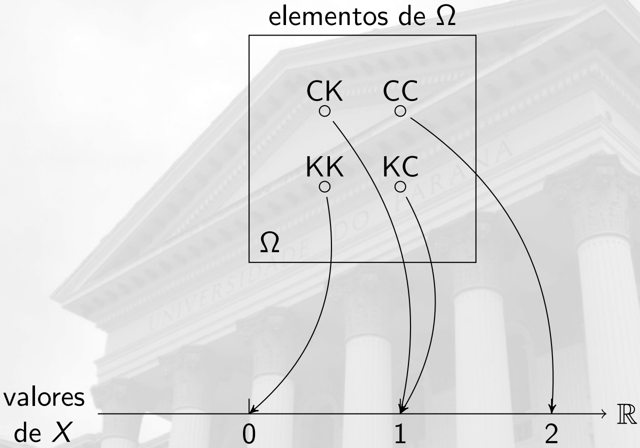
Uma **variável aleatória** é uma função que confere um número real a eventos (ou cada resultado) no espaço amostral de um experimento aleatório.

Notação

Uma variável aleatória é denotada por uma letra maiúscula, tal como X. Depois do experimento ser conduzido, o valor medido/observado da v.a. é denotado por uma letra minúscula, tal como $x = 70$ gramas.

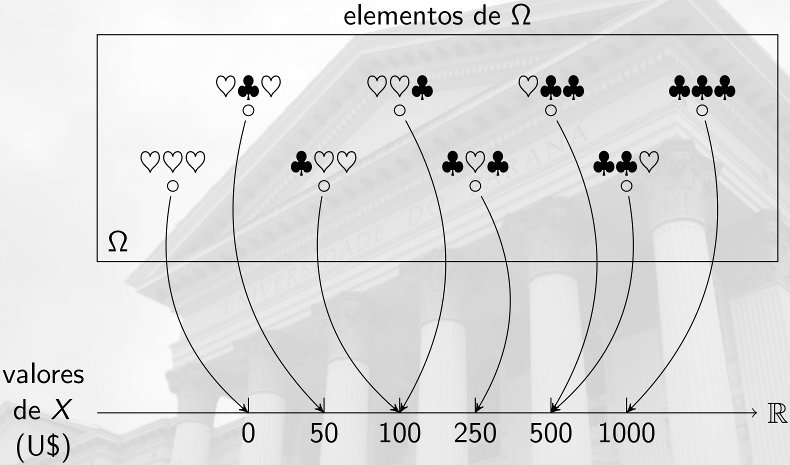
Lançamento de duas moedas

Experimento: lanças 2 moedas, v.a. X: número de resultados cara;



Jogo de caça-níquel

Experimento: girar os cilindros (3 cilindros de 4 cédulas com 3♡ e 1♣ em cada); v.a. X: prêmio do resultado (U$);



Tipos de variáveis aleatórias

Discretas

Uma v.a. **discreta** apresenta um conjunto **contável** (finito ou infinito) de valores que pode assumir.

Ex: número de caras ao lançar 3 moedas, prêmio de uma máquina caça-níquel, número de votos recebidos, aprovação no vestibular, número de leitões por gestação, número de acidentes de trânsito por ano, número de acesso diário ao bebedouro, grau de uma multa de trânsito, grau de uma queimadura na pele.

Contínuas

Uma v.a. **contínua** apresenta um conjunto **infinito** de valores que pode assumir dentro de um intervalo limitado ou aberto.

Ex: peso de um fruto, teor de açucar da cana-de-açúcar, área foliar coberta por fungo, pH do solo, precipitação diária, concentração de uma substância, diâmetro do colmo, pureza de um metal, tempo para conclusão de uma tarefa, instante de chegada de um e-mail, retorno financeiro de um investimento.

Variáveis aleatórias discretas

Distribuição de probabilidades

A **distribuição de probabilidades** de uma v.a. X é uma descrição das probabilidades associadas com os possíveis valores de X. Os valores que X compoem o que se chama de **suporte** da v.a.. Para uma v.a. discreta, a distribuição é frequentemente especificada por apenas uma lista de valores possíveis, juntamente com a probabilidade de cada um. Em alguns casos, é conveniente/possível expressar a probabilidade por uma fórmula (modelo).

Exemplo - jogo de caça-níquel

X: prêmio pago pela máquina em uma jogada.

ω	x	$P(\omega)$	$\Pr(X = x)$
$\clubsuit\clubsuit\clubsuit$	1000	$(1/4)^3$	$1/64$
$\clubsuit\clubsuit\heartsuit$, $\heartsuit\clubsuit\clubsuit$	500	$2 \cdot (1/4 \cdot 1/4 \cdot 3/4)$	$6/64$
$\clubsuit\heartsuit\clubsuit$	250	$1/4 \cdot 1/4 \cdot 3/4$	$3/64$
$\heartsuit\heartsuit\clubsuit$, $\clubsuit\heartsuit\heartsuit$	100	$2 \cdot (1/4 \cdot 3/4 \cdot 3/4)$	$18/64$
$\heartsuit\clubsuit\heartsuit$	50	$1/4 \cdot 3/4 \cdot 3/4$	$9/64$
$\heartsuit\heartsuit\heartsuit$	0	$(3/4)^3$	$27/64$

Os valores da coluna x e $\Pr(X = x)$ representam a distribuição de probabilidades da v.a. X pois associam uma probabilidade a cada valor que X assume.

Propriedades de uma d.p.

- Ser positiva para todos os valores de X:

$$0 \leq \Pr(X = x) \leq 1, \forall\ x;$$

- A soma das probabilidades deve ser 1:

$$\sum_{\forall x} \Pr(X = x) = 1.$$

Gráfico de uma distribuição de probabilidades

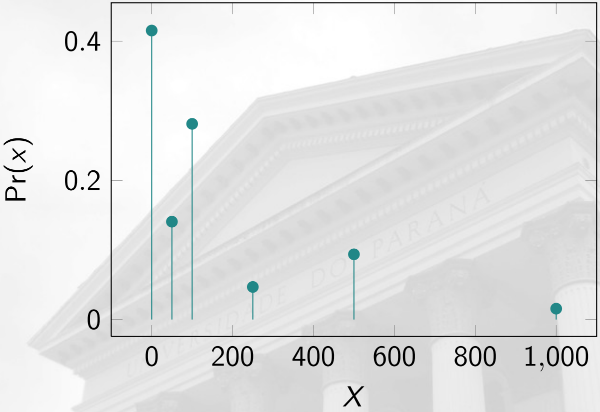


Figura 1: Probabilidades em função dos valores que a v.a. X (prêmio pago pela máquina em uma jogada) assume.

Distribuição de probabilidades acumulada

A função de distribuição acumulada de uma v.a. discreta X, denotada por $F(x)$, é

$$F(x) = \Pr(X \leq x) = \sum_{\forall x_i \leq x} \Pr(X = x). \tag{31}$$

$F(x)$ satisfaz as seguintes propriedades:

- tem imagem no intervalo [0,1] e domínio no $\mathbb{R}$:

$$0 \leq F(x) \leq 1, \forall\, x \in \mathbb{R};$$

- é não decrescente:

$$\text{se } x_1 < x_2, \text{ então } F(x_1) \leq F(x_2).$$

Gráfico de uma d.p. acumulada

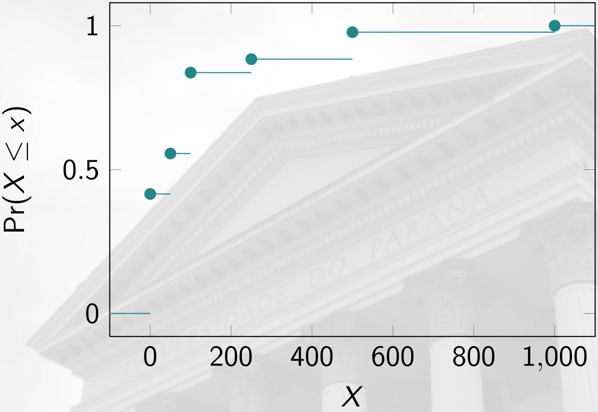


Figura 2: Probabilidades acumuladas em função dos valores que a v.a. X (prêmio pago pela máquina em uma jogada) assume.

Valor esperado (média)

O valor esperado de uma v.a. é uma medida do centro da distribuição de probabilidades. Ela representa o valor médio da v.a. quando ela é observada infinitamente, por isso chamado de média da v.a.. A **média** ou **valor esperado** de uma v.a. discreta X, denotado como μ ou $E(X)$, é

$$\mu = E(X) = \sum_{\forall x} x \cdot \Pr(X = x). \tag{32}$$

A $E(X)$ é, portanto, a média ponderada dos valores que X assume, com os pesos iguais às probabilidades.

Valor esperado (média)

Valor esperado do prêmio por jogada do caça-níquel

$$
\begin{aligned}
\mu = E(X) &= \sum_{\forall x} x \cdot \Pr(X = x) \\
&= 0 \cdot (27/64) + 50 \cdot (9/64) + 100 \cdot (18/64) \\
&+ 250 \cdot (3/64) + 500 \cdot (6/64) + 1000 \cdot (1/64) \\
&= 109.37
\end{aligned}
$$

Isso significa que se jogarmos nesse caça-níquel infinitamente, na média de todas as jogadas, o valor recebido como prêmio será igual a 109.37.

Interpretação geométrica do valor esperado

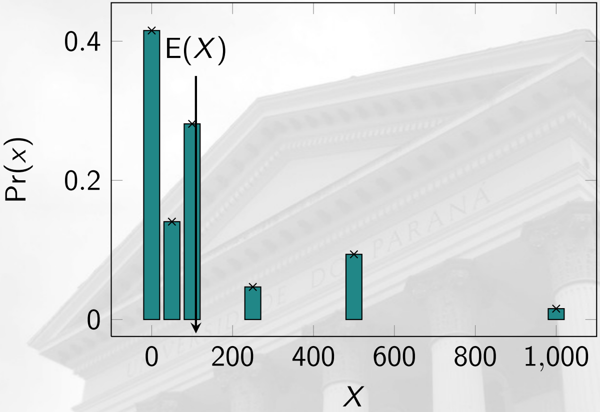


Figura 3: Interpretação geométrica do valor esperado que representa o ponto de equilíbrio (centro) da distribuição de probabilidades.

Variância

A variância de uma v.a. representa a dispersão da distribuição de probabilidades. Ela mede o quando as probabilidades estão próximas, ou concentradas, ao redor do valor central (μ). Representamos a variância de uma v.a. por σ^2 ou $V(X)$, onde

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_{\forall x} (x-\mu)^2 \cdot P(x) = \sum_{\forall x} x^2 \cdot P(x) - \mu^2. \quad (33)$$

O desvio-padrão de X é $\sigma = \sqrt{\sigma^2}$.

Variância

Variância do prêmio por jogada do caça-níquel

$$
\begin{aligned}
\sigma^2 = V(X) &= \sum_{\forall x} (x - \mu)^2 \cdot \Pr(X = x) \\
&= (0 - 109.37)^2 \cdot (27/64) + (50 - 109.37)^2 \cdot (9/64) \\
&+ (100 - 109.37)^2 \cdot (18/64) + \cdots + (1000 - 109.37)^2 \cdot (1/64) \\
&= 1122714.99
\end{aligned}
$$

O desvio-padrão é $\sigma = \sqrt{\sigma^2} = 1059.58$

Exemplo

Mensagens

O número de mensagens enviadas por hora, atráves de uma rede de computadores, é uma v.a. discreta e tem a seguinte distribuição de probabilidades

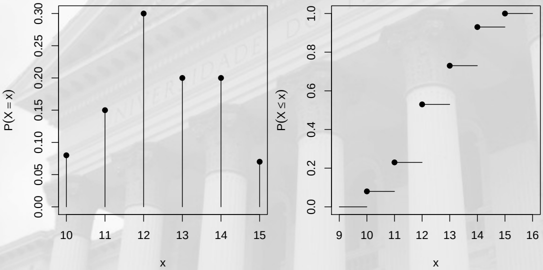
x = número de mensagens	10	11	12	13	14	15
$\Pr(X = x)$	0.08	0.15	0.30	0.20	0.20	0.07

Faça o gráfico da distribuição de probabilidades, da distribuição de probabilidades acumulada, determine a média e o desvio-padrão do número de mensagens enviadas por hora.

Exemplo

x	$\Pr(X = x)$	$\Pr(X \leq x)$	$x \cdot P(x)$	$(x-\mu)^2$	$(x-\mu)^2 \cdot P(x)$
10	0.08	0.08	0.80	6.25	0.50
11	0.15	0.23	1.65	2.25	0.34
12	0.30	0.53	3.60	0.25	0.07
13	0.20	0.73	2.60	0.25	0.05
14	0.20	0.93	2.80	2.25	0.45
15	0.07	1.00	1.05	6.25	0.44
soma			12.50		1.85

$\mu = 12.50$, $\sigma = \sqrt{\sigma^2} = \sqrt{1.85} = 1.36$.

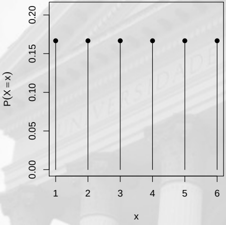


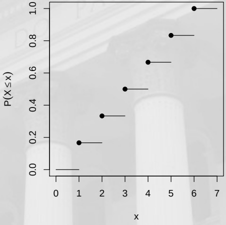
Distribuição Uniforme Discreta

Uma v.a. X tem distribuição uniforme discreta se cada um dos n valores em seu suporte, isto é, $x_1, x_2, \ldots, x_n$, tiver igual probabilidade. Então

$$\Pr(X = x) = p(x) = 1/n. \tag{34}$$

São casos dessa distribuição o resultado do lançamento um dado justo, o sorteio de um número no bingo, na loteria, etc. Representamos $X \sim UD(n)$.





Distribuição Bernoulli

Uma v.a. X tem distribuição Bernoulli se apresenta dois restultados possíveis, chamados frequentemente de **sucesso** (ou 1) e **fracasso** (ou 0). A probabilidade de sucesso é representada pelo parâmetro p, e a do fracasso é $1 - p$. Então

$$\Pr(X = x) = p(x) \begin{cases} p & \text{, se } x = 1 \\ 1 - p & \text{, se } x = 0. \end{cases} \tag{35}$$

O parâmetro p têm o seguinte espaço paramétrico $\Theta = \{p : 0 < p < 1\}$. São casos dessa distribuição o resultado do lançamento de uma moeda, o sexo de um bebê, o voto a proposta, o resultado de um teste de germinação, etc. Representamos $X \sim Ber(p)$ e temos que

- valor esperado: $\mu = E(X) = 0 \cdot (1 - p) + 1 \cdot (p) = p$
- variância: $\sigma^2 = V(X) = 0^2 \cdot (1 - p) + 1^2 \cdot (p) - p^2 = p(1 - p)$.

Distribuição Binomial

Uma v.a. X tem distribuição Binomial se é o **número de sucessos** obtido em n provas de Bernoulli se

- as tentativas forem independentes
- a probabilidade de sucesso (p) permanecer constante em todas as tentativas.

Assim X tem distribuição binomial com parâmetros n e p, $\Theta = \{0 < p < 1, n \in \mathbb{N}_*^+\}$. O suporte é o conjunto 0, 1, 2, ..., n. Representamos por $X \sim Bin(n{,}p)$ e a função de probabilidade de X é

$$\Pr(X = x) = p(x) = \binom{n}{x} p^x (1-p)^{n-x}. \tag{36}$$

- valor esperado: $\mu = E(X) = n \cdot p$
- variância: $\sigma^2 = V(X) = n \cdot p \cdot (1-p)$.

Distribuição Geométrica

Uma v.a. X tem distribuição Geométrica se é o **número tentativas** até que o **primeiro sucesso** seja obtido em uma série de provas independentes de Bernoulli (p constante). Assim

$$\Pr(X = x) = p(x) = (1 - p)^{x-1} \cdot p. \tag{37}$$

X tem distribuição geométrica com parâmetro p, $\Theta = \{0 < p < 1\}$. O suporte é o conjunto 0, 1, 2, ... Representamos por $X \sim Geo(p)$ onde

- valor esperado: $\mu = E(X) = 1/p$
- variância: $\sigma^2 = V(X) = (1 - p)/p^2$.

Distribuição Binomial Negativa

Uma v.a. X tem distribuição Binomial Negativa se é o **número tentativas** até que r **sucessos** sejam obtidos em uma série de provas independentes de Bernoulli (p constante). Assim

$$\Pr(X = x) = p(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r. \tag{38}$$

X tem distribuição binomial negativa com parâmetros p e r, $\Theta = \{0 < p < 1, r \in \mathbb{N}_*^+\}$. O suporte é o conjunto r, $r+1$, $r+2$, ... Representamos por $X \sim BN(p,r)$ onde

- valor esperado: $\mu = E(X) = r/p$
- variância: $\sigma^2 = V(X) = r(1-p)/p^2$.

Distribuição Hipergeométrica

Uma v.a. X tem distribuição Hipergeométrica se é o **número sucessos** obtidos em uma amostra de n elementos sem reposição de uma população com N elementos sendo que K deles são classificados como sucesso e $N - K$ como fracasso. Assim

$$\Pr(X = x) = p(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \cdot x. \tag{39}$$

X tem distribuição hipergeométrica com parâmetros n, N e K, $\Theta = \{N \in \mathbb{N}_*^+, n \in \mathbb{N}_*^+ < N, K \in \mathbb{N}_*^+ < N\}$. O suporte é o conjunto dos números inteiros entre $\max\{0, n - (N - K)\}, \ldots, \min\{K, n\}$. Representamos por $X \sim Hiper(n, N, K)$ onde

- valor esperado: $\mu = E(X) = nK/N$
- variância: $\sigma^2 = V(X) = \frac{K(N-K)n(N-n)}{N^2(N-1}$.

Distribuição de Poisson

Uma v.a. X tem distribuição de Poisson se é o **número de eventos** ocorridos em um intervalo de forma que

- o intervalo possa ser subdividido em subintervalos suficientemente pequenos
- a probabilidade de 2 eventos no mesmo subintervalo seja zero
- a probabilidade de 1 evento seja a mesma em qualquer subintervalo e proporcional ao comprimento desse subintervalo
- os eventos sejam independentes dos ocorridos em outros subintervalos.

Assim

$$\Pr(X = x) = p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}. \tag{40}$$

X tem distribuição de Poisson com parâmetro λ, $\Theta = \{\lambda > 0\}$. O suporte é o conjunto dos números inteiros positivos. Representamos por $X \sim Poi(\lambda)$ onde

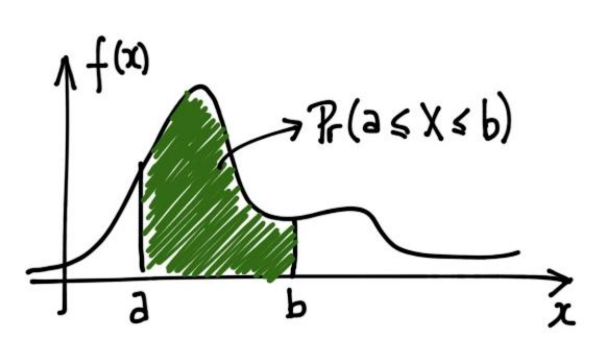
- valor esperado: $\mu = E(X) = \lambda$
- variância: $\sigma^2 = V(X) = \lambda$.

Funções densidade de probabilidade

Diferente das v.a. discretas, o **suporte** para uma variável aleatória contínua X contém infinos valores dentro de um intervalo x_a e x_b. Sendo assim, não é possível associar probabilidade à cada particular valor do suporte de X e portanto não existe função de probabilidade. Para o caso contínuo representa a distribuição de probabilidades por meio da **função densidade de probabilidade**.

Uma função densidade de probabilidade é uma função tal que

1. $f(x) \geq 0$;
2. $\int_{\forall x} f(x)\,\mathrm{d}x = 1$;
3. $\Pr(a \leq X < b) = \int_a^b f(x)\,\mathrm{d}x, b > a$.


f(x)
Pr(a ≤ X ≤ b)
a
b
x