ESTATÍSTICA

Michelle Hanne Soares de Andrade
michellehanne.andrade@gmail.com

Variáveis Aleatórias

Variáveis Aleatórias e sua Distribuições de Probabilidade

- **Exemplo de um experimento**
  - Suponha que joguemos para o alto uma moeda dez vezes e contemos o número de vezes em que dê cara.
  - Antes de atirarmos a moeda, sabemos que o número de caras que aparecerá será um inteiro entre 0 e 10, e, portanto, os resultados do experimento são bem definidos.
  - $X = 1$ se der cara, e $X = 0$ se der coroa.

O resultado do número de caras obtidas nas 10 jogadas é uma Variável Aleatória.

Variáveis Aleatórias e sua Distribuições de Probabilidade

- Uma **variável aleatória** é aquela que assume valores numéricos e tem um resultado que é determinado por um experimento.

Variáveis Aleatórias e sua Distribuições de Probabilidade

O espaço amostral de um experimento aleatório lista todas as saídas possíveis do fenômeno (numéricas ou não).

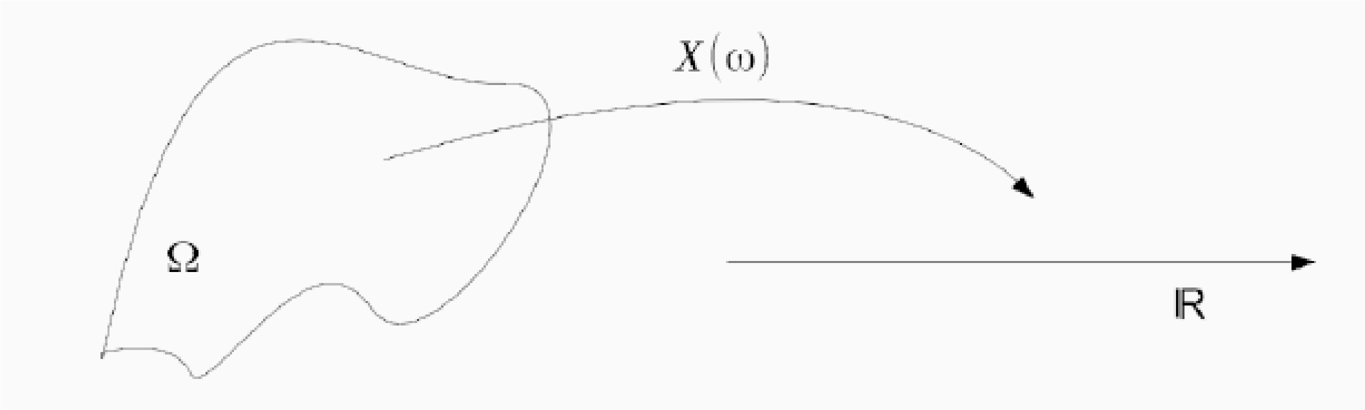
Em alguns casos, tal descrição é suficiente, contudo em outros casos, é mais interessante associarmos um número a cada uma dessas saídas.

Como o resultado do experimento é aleatório, temos que o valor a ele associado também é aleatório.

Por este motivo, a função que associa valores às saídas de um experimento é denominada uma variável aleatória.

Variáveis Aleatórias

Definição: Uma Variável Aleatória é uma função que associa valores reais aos elementos do espaço amostral de um experimento aleatório.



Variáveis Aleatórias

- **Notação:** variáveis aleatórias (v.a.'s) são usualmente denotadas por letras maiúsculas, tais como *X*. Após uma realização do experimento aleatório, o valor observado é representado por uma letra minúscula, tal como *x*.
- Uma variável aleatória que somente pode assumir os valores zero e um é chamada **variável aleatória de Bernoulli** (ou **binária**).

Variável Aleatória Discreta

- Uma **variável aleatória discreta** é a que somente assume um número finito ou infinito enumerável de valores.
- **Exemplo:** número de arranhões em uma superfície, número equivocado de bits transmitidos em uma sequência, etc.

- ***Uma variável aleatória de Bernoulli é o exemplo mais simples de uma variável aleatória discreta.***

Variável Aleatória Discreta

- **Exemplo 1 Moeda:** se ela for "justa", então, $P(X = 1) = 1/2$. Como a soma das probabilidades deve ser igual à unidade, $P(X = 0) = ½$.
- **Exemplo 2 Empresa Aérea:** decidir quantas reservas serão aceitas para um vôo com 100 lugares disponíveis.
  - Esse problema pode ser analisado no contexto de diversas variáveis aleatórias de Bernoulli da seguinte maneira: para um passageiro selecionado aleatoriamente, defina uma variável aleatória de Bernoulli como $X = 1$ se a pessoa aparecer para embarque, e $X = 0$ se não aparecer.

Variável Aleatória Contínua

- Caso a v.a. assuma valores ao longo de algum subintervalo de $\mathbb{R}$, temos uma variável contínua.
- Exemplo: corrente elétrica, temperatura, voltagem, peso, etc.

Variável Aleatória Discreta

Variável Aleatória Discreta

- **Exemplo Empresa Aérea:** se $\theta = 0{,}75$, existirá 75% de probabilidade de que um passageiro apareça para o embarque após ter feito a reserva, e 25% de probabilidade de que o passageiro não apareça.

$$P(X = 1) = \theta$$
$$P(X = 0) = 1 - \theta$$

Variável Aleatória Discreta

- De forma mais geral, qualquer variável aleatória discreta é completamente descrita listando seus possíveis valores e a probabilidade associada que ela assume para cada valor. Se *X* assumir os *k* possíveis valores {*x*1, ..., *xk*}, as probabilidades *p*1, *p*2, ..., *pk* serão definidas por:

$$p_j = P(X = x_j), j = 1, 2, \ldots, k,$$

"A probabilidade de *X* assumir o valor *xj* é igual a *pj*".

onde cada *pj* estará entre zero e um e *p*1+ *p*2+ ... *pk* = 1.

Variável Aleatória Discreta

- Como as variáveis aleatórias de Bernoulli são tão frequentes, temos uma notação especial para elas: $X \sim Bernoulli(\theta)$ é lida como "X tem uma distribuição de Bernoulli com probabilidade de sucesso igual a θ."

Variável Aleatória Discreta

- Seja X uma variável aleatória discreta com valores possíveis $x_1, x_2, ..., x_n$.
- A função de probabilidade de X, denotada por f, é uma função tal que

  1. $f(x_i) \geq 0, \quad \forall i$
  2. $\sum_{i=1}^{n} f(x_i) = 1$
  3. $f(x_i) = P(X = x_i), \quad \forall i$

- A função de probabilidade de uma v.a. discreta pode ser representada informalmente por meio tabelas.

Função de Densidade de Probabilidade (fdp)

- A **função densidade de probabilidade** (fdp) de X resume as informações relativas aos possíveis resultados de X e as probabilidades correspondentes:

$$f(x_j) = p_j, j = \ 1, 2, \dots, k,$$

- com $f(x) = 0$ de qualquer x não igual a xj para algum j.

Função de Densidade de Probabilidade (fdp)

- **Em outras palavras, para qualquer número real *x*, *f*(*x*) será a probabilidade que a variável aleatória *X* assumirá para o valor particular de *x*.**
- Quando lidamos com mais de uma variável aleatória, algumas vezes é útil subscrever a fdp em questão: f_x é a fdp de *X*, f_x é a fdp de *Y*, e assim por diante.

Função de Densidade de Probabilidade (fdp)

- **Exemplo:** Lotes de três peças são retirados de uma linha de produção para controle de qualidade. Suponha que a probabilidade de encontrar uma peça defeituosa é de 10% e defina a v.a. Y como sendo o número de peças defeituosas observadas.

Função de Densidade de Probabilidade (fdp)

Pontos amostrais	Y	f
$\overline{DDD}$	0	$(0,9)^3$
$D\overline{DD}$, $\overline{D}D\overline{D}$, $\overline{DD}D$	1	$3(0,9)^2(0,1)$
$\overline{D}DD$, $D\overline{D}D$, $DD\overline{D}$	2	$3(0,9)(0,1)^2$
DDD	3	$(0,1)^3$

- Em resumo,

$$Y(\overline{DDD}) = 0,\ Y(D\overline{DD}) = Y(\overline{D}D\overline{D}) = Y(\overline{DD}D) = 1,$$

$$Y(\overline{D}DD) = Y(D\overline{D}D) = Y(DD\overline{D}) = 2,\ Y(DDD) = 3$$

e

$$f(0) = (0,9)^3,\ f(1) = 3(0,9)^2(0,1),\ f(2) = 3(0,9)(0,1)^2 \text{ e } f(3) = (0,1)^3.$$

Variável Aleatória Contínua

- Uma variável X será uma **variável aleatória contínua** se assumir qualquer valor real com probabilidade zero.
- A ideia é que uma variável aleatória contínua X pode assumir tantos valores possíveis que não podemos enumerá-los ou compará-los com os inteiros positivos, de modo que a consistência lógica garante que X pode assumir cada valor com probabilidade zero.

Variável Aleatória Contínua

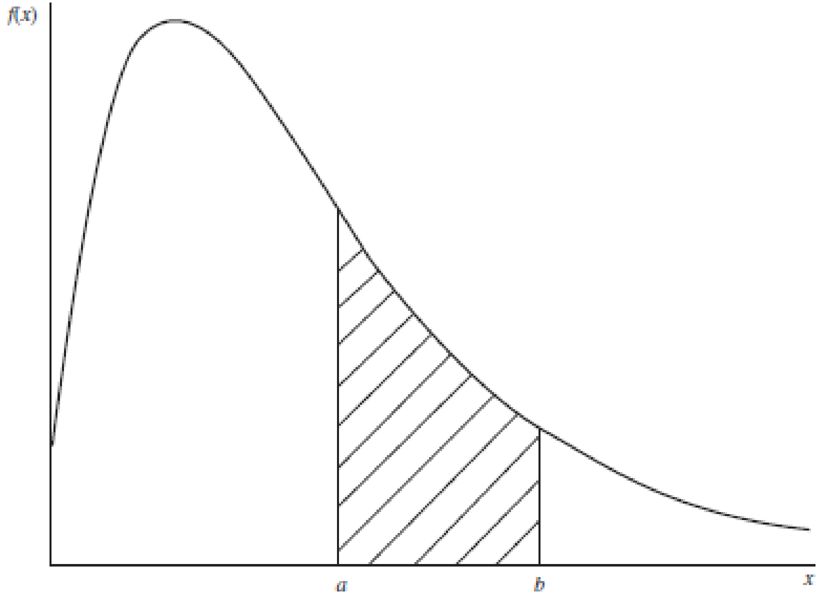
- Por exemplo: a medida mais refinada do preço de um bem é em termos de centavos. **Podemos imaginar relacionando todos os possíveis valores de preços ordenadamente, o que tecnicamente faz com que preço seja uma variável aleatória discreta.** Porém, existem tantos valores possíveis de preços que o uso da mecânica das variáveis aleatórias discretas não é viável.

Função de Densidade de Probabilidade

- Podemos definir a fdp para as Variáveis Aleatórias Contínuas.
- Porém, não faz sentido discutir a probabilidade de que uma variável aleatória contínua assuma um valor em particular, usamos a fdp de uma variável aleatória contínua somente para computar eventos envolvendo uma diversidade de valores.

Função de Densidade de Probabilidade

- Por exemplo, se a e b forem constantes onde $a < b$, a probabilidade de X estar entre os números a e b, $P(a \leq X \leq b)$, será a *área* sob a fdp entre os pontos *a* e *b*.



integral da função *f* entre os pontos *a* e *b*.

Função de Distribuição Cumulativa

- Ao computar probabilidades para variáveis aleatórias contínuas, é mais fácil trabalhar com a **função de distribuição cumulativa** (**fdc**). Se X for qualquer variável aleatória, então, sua fdc será definida por qualquer numero x real pela equação:

$$F(x) \equiv P(X \leq x).$$

Função de Distribuição Cumulativa

- Para **variáveis aleatórias discretas**, será obtida somando a fdp para todos os valores *xj* tais que $xj \leq x$. Para uma variável aleatória contínua, *F*(*x*) será a área sob a fdp, *f*, à esquerda do ponto *x*.
- Como *F*(*x*) é simplesmente uma probabilidade, ela estará sempre entre 0 e 1. Além disso, se $x1 \leq x2$, então, $P(X \leq x1) \leq P(X \leq x2)$, isto é, $F(x1) \leq F(x2)$, Isso significa que uma fdc é uma função crescente (ou pelo menos não-decrescente) de *x*.

Função de Distribuição Cumulativa

- De uma forma em geral, para uma v.a. discreta *X*, a função de distribuição acumulada satisfaz:

1. $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$;
2. $0 \leq F(x) \leq 1$;
3. $x \leq y \Rightarrow F(x) \leq F(y)$.

- **Além disso, mesmo que a v.a. *X* assuma apenas uma quantidade enumerável de valores, sua distribuição acumulada *F* é definida para todos os reais.**

Função de Distribuição Cumulativa

- Determinar a função de distribuição acumulada da v.a. X dada por

X	$f(X)$
0	0,6561
1	0,2916
2	0,0486
3	0,0036

Solução:

$$
\begin{aligned}
x < 0: &\quad F(x) = P(X \leq x) = 0,0000 \\
0 \leq x < 1: &\quad F(x) = P(X \leq x) = 0,6561 \\
1 \leq x < 2: &\quad F(x) = P(X \leq x) = 0,9478 \\
2 \leq x < 3: &\quad F(x) = P(X \leq x) = 0,9964 \\
x \geq 3: &\quad F(x) = P(X \leq x) = 1,0000
\end{aligned}
\quad \Rightarrow \quad
F(x) = \begin{cases} 0,0000; & x < 0 \\ 0,6561; & 0 \leq x < 1 \\ 0,9478; & 1 \leq x < 2 \\ 0,9964; & 2 \leq x < 3 \\ 1,0000; & x \geq 3 \end{cases}
$$

Função de Distribuição Cumulativa

- Determinar a função de distribuição acumulada da v.a. X

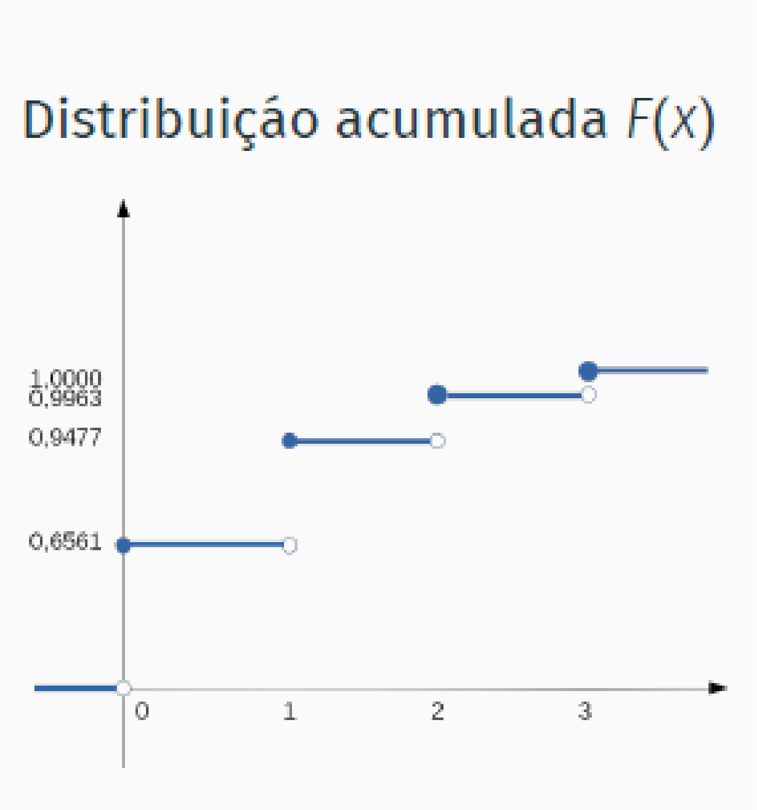
Solução:

$$
\begin{aligned}
x < 0: &\quad F(x) = P(X \leq x) = 0,0000 \\
0 \leq x < 1: &\quad F(x) = P(X \leq x) = 0,6561 \\
1 \leq x < 2: &\quad F(x) = P(X \leq x) = 0,9478 \\
2 \leq x < 3: &\quad F(x) = P(X \leq x) = 0,9964 \\
x \geq 3: &\quad F(x) = P(X \leq x) = 1,0000
\end{aligned}
\quad \Rightarrow \quad
F(x) = \begin{cases}
0,0000; & x < 0 \\
0,6561; & 0 \leq x < 1 \\
0,9478; & 1 \leq x < 2 \\
0,9964; & 2 \leq x < 3 \\
1,0000; & x \geq 3
\end{cases}
$$

Função de Distribuição Cumulativa

É importante notar que o tamanho do salto em um ponto *Xi* qualquer equivale à probabilidade de observarmos tal ponto.

$$P(X = x_i) = F(x_i) - \lim_{x \to x^-} F(x).$$



Função de Distribuição Cumulativa

As **funções de distribuições cumulativas** foram **tabuladas** para todas as distribuições contínuas importantes em probabilidade e estatística. A mais conhecida delas é a **distribuição normal**.

Distribuições Conjuntas, Distribuições Condicionais e Independência

- **Probabilidade Conjunta:** Probabilidade de que uma pessoa que faz uma reserva compareça para embarque *e* que seja uma pessoa que viaje a negócios.
- ***Probabilidade Condicional***: condicional à pessoa ser uma pessoa que viaje a negócios, qual é a probabilidade de que ela compareça para embarque?

Distribuições Conjuntas e Independência

- Sejam *X* e *Y* variáveis aleatórias discretas. Então, (*X*,*Y*) têm uma **distribuição conjunta**, que é totalmente definida pela *função densidade de probabilidade conjunta* de (*X*,*Y*):

$$f_{X,Y}(x,y) = \mathrm{P}(X = x, Y = y),$$

- Em um caso, é fácil obter a fdp conjunta se forem dadas as fdps de *X* e *Y*. Em particular, as variáveis aleatórias *X* e *Y* são independentes se, e somente se,

$$f_{X,Y}(x,y) = f_X(x) f_y(y)$$

Distribuições Conjuntas e Independência

- Se X e Y forem discretas:

$$P(X = x, Y = y) = P(X = x)P(Y = y);$$

- em outras palavras, a probabilidade de que $X = x$ e $Y = y$ é o produto das duas probabilidades $P(X = x)$ e $P(Y = y)$

Distribuições Conjuntas e Independência

EXEMPLO B.1

(Arremessos de Lances Livres)

Considere um jogador de basquetebol fazendo dois lances livres. Seja X uma variável aleatória de Bernoulli igual a um se ele converter o primeiro arremesso, e zero, caso contrário. Seja Y uma variável aleatória de Bernoulli igual a um se ele converter o segundo arremesso. Suponha que ele seja um jogador que converte 80% dos arremessos, de forma que $P(X = 1) = P(Y = 1) = 0{,}8$. Qual é a probabilidade de o jogador converter os dois arremessos?

Se X e Y forem independentes, podemos facilmente responder a essa pergunta: $P(X = 1, Y = 1) = P(X = 1)(\ P(Y = 1) = (0{,}8)(0{,}8) = 0{,}64$. Portanto, existe 64% de probabilidade de converter ambos os lances livres. Se a probabilidade de converter o segundo arremesso depender de o primeiro arremesso ter sido convertido — isto é, X e Y não são independentes — então, esse cálculo simples não será válido.

Independência de Variáveis

Serão **variáveis aleatórias independentes** se, e somente se, suas fdps conjuntas forem o produto das fdps individuais para quaisquer ($x1$, $x2$, ..., xn). Essa definição de independência também é válida para variáveis aleatórias contínuas.

Independência de Variáveis

- Um fato útil sobre a questão da independência é que se ***X* e *Y* forem independentes** e definirmos novas variáveis aleatórias ***g*(*X*) e *h*(*Y*)** para quaisquer funções *g* e *h*, então, essas **novas variáveis aleatórias também serão independentes**.
- Normalmente, estamos interessados no número de sucessos em uma sequência de ensaios de Bernoulli *independentes*. Ex: Jogar a Moeda
  - Podemos ter uma Independência como aproximação. Ex: Reservas de um voo.

Distribuição Conjunto e Independência de Variáveis

- **Suponha que a variável de interesse principal é o número total de passageiros que comparecem para embarque** das *n* reservas - variável *X*.
  - Como cada Yi será igual à unidade quando uma pessoa comparece para embarque, podemos escrever $X = Yi + Y2 + ... + Yn$.
  - Agora, assumindo que cada *Yi* tem probabilidade de sucesso θ e que os *Yi* são independentes, é possível mostrar que *X* tem uma **distribuição binomial**.

Distribuição Conjunto e Independência de Variáveis

- Isto é, a função densidade de probabilidade de X é

$$f(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x},\ x = 0, 1, 2, \ldots, n,$$

$$X \sim Binomial(n, \theta).$$

Exemplo

- Se o voo tiver 100 lugares disponíveis, a empresa aérea estará interessada em $P(X > 100)$. Suponha, inicialmente, que *n* = 120, de modo que a companhia aérea aceitará 120 reservas, e que a probabilidade de que cada pessoa compareça para embarque seja $\theta = 0{,}85$.
- Então, $P(X > 100)$= P(*X* = 101)+ P(*X*= 102)+... +P(*X* = 120):
  - Se *n*=120, $\theta = 0{,}85$ e o valor apropriado de *x* (101 a 120), a probabilidade de que mais de 100 pessoas comparecerão para embarque é cerca de **0,659.**
  - Se n=110, a probabilidade de que mais de 100 pessoas comparecerão para embarque será de **apenas 0,024**

Exercício

1) Num experimento aleatório observa-se o gênero das crianças em famílias com três filhos (M: masculino e F: feminino).

O espaço amostral fica dado por $\Omega = \{(MMM),(MMF),(MFM),(FMM),(MFF),(FMF),(FFM),(FFF)\}$.

1)Quais são as possíveis variáveis Aleatórias desse espaço amostral?

2) Quais são os valores da variável aleatória “número de crianças do sexo masculino” e “número de crianças do sexo feminino”

Exercício

2) Joga-se dois dados equilibrados e observa-se as faces superiores. Considere a variável aleatória X: soma das faces superiores.

Como fica a função de probabilidade de X?