

Probabilidade e Variáveis Aleatórias: noções básicas

1 - O que é probabilidade?

Probabilidade tem como escopo **quantificar**, sob o ponto de vista **médio macroscópico** fenômenos, naturais ou não, ocorrendo de forma **sequencial** ou **simultânea**

Exemplos de fenômenos *aleatórios*:

- a temperatura (média, máxima, mínima) durante o ano em uma cidade;
- o ruído produzido por um aparelho de rádio mal sintonizado;
- o número de chamadas telefônicas em uma mesma conexão durante determinado intervalo;
- o número de pessoas diariamente contaminadas pelo covid19;
- o resultado de jogos de azar;
- o número de peças rejeitadas pelo controle de qualidade de um processo industrial;
- *et cetera ...*

Se esses (e outros) fenômenos são **realmente**
aleatórios, **por quê** se estuda probabilidades?

Observou-se que, nos fenômenos acima e em outros da mesma natureza, à medida que o número de observações aumentava, a média convergia para um valor constante, permanecendo a mesma para qualquer subsequência especificada antes que o experimento que originou esse valor tivesse sido executado.

Por exemplo:

no *experimento* de lançamento de uma moeda **não viciada**, sabemos de antemão, pois o “histórico” nos permite dizer, que a porcentagem de ocorrência de uma determinada face é 50% ou algo em torno disso. Além disso, se as saídas (faces) fossem contabilizadas somente a cada 3 lançamentos (poderia ser qualquer número), novamente iríamos obter algo em torno de 0,5 para cada uma das faces.

- A teoria de probabilidades procura descrever e prever esses valores em termos da probabilidade de *eventos*, *conjuntos de saídas* de determinado *experimento*.

2- Definições de Probabilidade

- Há três (pelo menos) definições para probabilidade
 1. definição clássica
 - foi a teoria mais aceita durante séculos;
 - atualmente, ainda é utilizada como hipótese inicial de probabilidades;
 - baseia-se não em *experimentos*, mas em *conhecimento prévio*.

A *probabilidade* de um *evento* A é um *número* $P(A)$ *determinado a priori*, associado a esse evento *sem necessidade* de experimentos, e é igual ao quociente

$$P(A) = \frac{n_A}{n},$$

onde n é o número *possível* de ocorrências e n_A é o número de saídas *favoráveis* ao evento A , desde que todos os *eventos* sejam *igualmente prováveis*.

- Problemas com a definição clássica
 - aplica-se apenas a uma classe de problemas em que as saídas favoráveis são igualmente prováveis. Por exemplo, no caso de um dado viciado em que a probabilidade de uma das faces fosse $1/5$ em vez de $1/6$, a definição clássica não forneceria resultados corretos;
 - assumir *a priori* que as faces de um dado são equiprováveis é algo natural, com base na simetria que o cubo exhibe e, também, na história de lançamentos de dados não viciados que, efetivamente, apresentam $1/6$ de probabilidade de ocorrência para cada uma das faces;

Observe o que escreve Papoulis, citando o “princípio da razão insuficiente”, postulado por **Jacob Bernoulli** e presente na obra “*Ars Conjectandi*”, publicada após sua morte, por seu sobrinho **Niklaus Bernoulli** em 1713:

Notes 1 The classical definition was introduced as a consequence of the *principle of insufficient reason*³: “In the absence of any prior knowledge, we *must* assume that the events A_i have equal probabilities.” This conclusion is based on the subjective interpretation of probability as a *measure of our state of knowledge* about the events A_i . Indeed, if it were not true that the events A_i have the same probability, then changing their indices we would obtain different probabilities without a change in the state of our knowledge.

- Problemas com a definição clássica
 - porém, pensemos no caso da cobrança de um pênalti no futebol. Assumir que $P(\text{gol}) = P(\text{não gol}) = 0,5$ seria a opção natural. Entretanto, ela não decorre de raciocínio lógico, pois não leva em consideração a preparação dos atletas ou as condições do gramado, por exemplo. Sem mais informações sobre o histórico de situações desse tipo envolvendo os mesmos atletas, não é possível afirmar que os dois eventos sejam equiprováveis.
 - caso o número de saídas seja um contínuo, algum parâmetro deve ser utilizado para a determinação da probabilidade. O *paradoxo de Bertrand* exemplifica um desses casos:

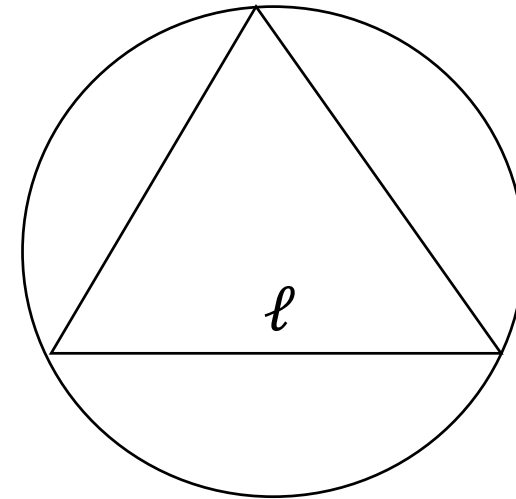
- ***Paradoxo de Bertrand***

Bertrand, Joseph. *Calcul des probabilités*: Paris, Gautier-Villars, 1889

Em uma circunferência de raio r , traça-se ao acaso uma corda. Qual a probabilidade de que o comprimento da corda seja inferior ao do lado ℓ do triângulo equilátero inscrito na circunferência?

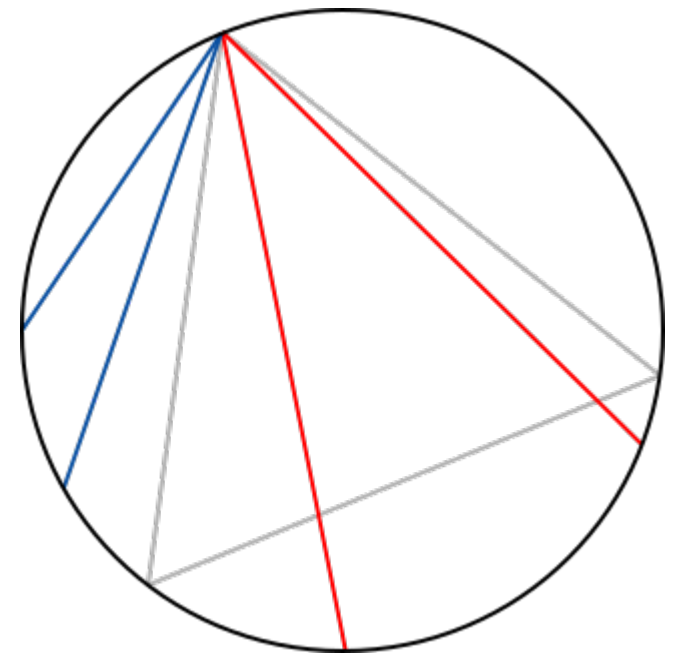
Da geometria básica, sabe-se que o lado de um triângulo inscrito em uma circunferência de raio r vale $r\sqrt{3}$

Vejamos as soluções possíveis:



1. considera-se a razão entre os comprimentos do arco definido pela corda aleatória e do arco referente ao lado do triângulo. Supondo o ponto inicial da corda como sendo um dos vértices do triângulo, caso o segundo ponto não ultrapasse o vértice adjacente ao primeiro, a corda será menor que o lado:

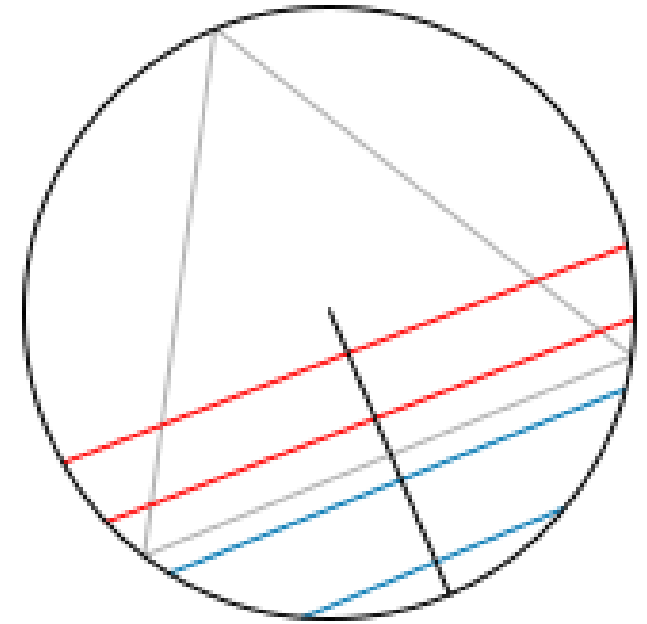
$$P(L < \ell) = \frac{\frac{2\pi r}{3}}{2\pi r} = \frac{1}{3}$$



Fonte:
<https://commons.wikimedia.org/w/index.php?curid=2141731>

2. traça-se o raio da circunferência como a mediatriz de um dos lados e toma-se cordas perpendiculares a esse raio (e, portanto, paralelas ao referido lado). Como o lado do triângulo divide o raio em partes iguais, então, caso a corda não ultrapasse o lado que lhe é paralelo, seu comprimento será inferior ao desse lado:

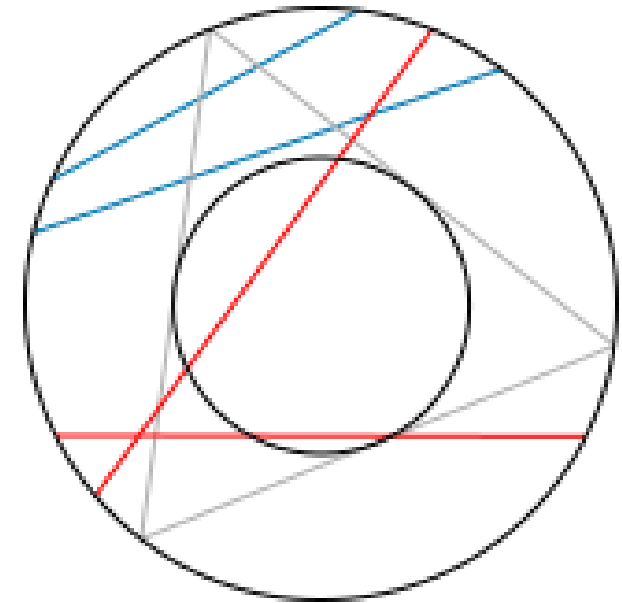
$$P(L < \ell) = \frac{r/2}{r} = \frac{1}{2}$$



Fonte:
<https://commons.wikimedia.org/w/index.php?curid=2141740>

3. traça-se a circunferência inscrita ao triângulo. As cordas que não interceptam a circunferência interna são menores que o lado do triângulo. A probabilidade procurada é, então a relação entre as áreas dos dois círculos:

$$P(L < \ell) = \frac{\pi \left(\frac{r}{2}\right)^2}{\pi r^2} = \frac{1}{4}$$



Fonte:
<https://commons.wikimedia.org/w/index.php?curid=2141745>

- *Qual das três soluções é a correta?*

As **três** são **corretas**! Por quê?

Pois *dependem de como é proposto o experimento* que gerou os eventos possíveis.

2. probabilidade como frequência relativa

- baseia-se na realização de experimentos

A probabilidade $P(A)$ de um evento A é

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n},$$

onde n_A é o número de saídas favoráveis, ocorrências do evento A e n é o número de experimentos realizados.

Dessa forma, *espera-se* que, *se* o experimento for executado n vezes e o evento A for obtido n_A vezes então, *com certo grau de confiabilidade*, a frequência relativa n_A/n de ocorrência do evento A será próxima de $P(A)$, ou seja,

$$P(A) \simeq \frac{n_A}{n}$$

desde que n seja suficientemente grande.

- Embora pareça razoável que a realização de experimentos forneça as probabilidades desejadas sem que qualquer informação a priori seja necessária, o que significa “numero de experimentos suficientemente grande”?
- esse é o problema: não importa quão elevados sejam os números n e n_A , serão **sempre finitos**.
- na prática, a expressão $P(A) \simeq \frac{n_A}{n}$ é utilizada apenas como hipótese inicial a ser posteriormente verificada, mas nunca como um resultado experimental.

3. definição axiomática

- deve-se a Andrey Nikolaevich Kolmogorov (1903-1987), no trabalho Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergeb. Math und ihrer Grenzgeb. vol. 2, (1933)

- baseia-se em **3 postulados**, listados abaixo:

1. a **probabilidade** $P(A)$ de um **evento** A é um **número** não negativo atribuído a esse evento, ou seja,

$$P(A) \geq 0$$

2. a **probabilidade** de um **evento certo** é igual a 1

$$P(S) = 1$$

3. se **dois eventos** A e B forem **mutuamente excludentes**, isto é, se $P(A \cap B) = 0$, então $P(A \cup B) = P(A) + P(B)$

3 – Conceitos envolvidos na definição axiomática

- a definição axiomática é a base de todo o desenvolvimento da *teoria de probabilidades* e da *teoria da medida*.
- nesse curso, recorreremos apenas a alguns pontos da teoria axiomática de probabilidades para a conceituação de variável aleatória, necessária para a compreensão dos processos estocásticos.

Conceitos envolvidos na teoria axiomática de probabilidade

- **conjunto:** é uma coleção de “objetos”, concretos ou abstratos.
Exemplo: conjunto de veículos que passou por uma praça de pedágio nas últimas 24 horas;
- **subconjunto:** é uma coleção de “objetos” que está contida na coleção maior;
Exemplo: conjunto de motocicletas que passou pela mesma praça de pedágio nas 24 horas;

Na teoria axiomática de probabilidade, *conjuntos* são denominados *eventos*

Conceitos envolvidos na teoria axiomática de probabilidade

- o interesse recai sobre o conjunto de todas as saídas de um experimento aleatório e sobre subconjuntos desse conjunto
- nomenclatura:

Ω - conjunto de todos os eventos, denominado espaço amostral do experimento aleatório;

$\{A\}$ - *subconjunto* do experimento, *evento* de interesse;

ζ – *saída* individual ou *realização*, resultado obtido no experimento.

Conceitos envolvidos na teoria axiomática de probabilidade

Campos sigma ou sigma-algebra

- considerando um experimento cujo espaço amostral seja Ω , caso Ω possua um número enumerável de elementos então, a cada subconjunto de Ω , pode-se **atribuir uma probabilidade** de maneira consistente com os axiomas enunciados anteriormente;
- a coleção de todos os subconjuntos de Ω é denominada maior *campo sigma* ou maior **σ —álgebra**
- a menor σ —álgebra é formada apenas pelo evento vazio (ϕ) e pelo evento certo, (Ω)
- caso o número de elementos do espaço amostral **não seja enumerável**, recai-se na *teoria da medida*, que não faz parte do escopo deste curso.

Portanto, um experimento fica definido pelo seu espaço de probabilidade, composto por:

Representação de um experimento

- espaço amostral Ω
- σ –álgebra (σ_a)
- função de probabilidade

$$\mathcal{F} \triangleq (\Omega, \sigma_a, P(\zeta))$$

Obs: pode haver diversas sigma-álgebra para um mesmo experimento

Exemplos

- 1) experimento (\mathcal{F}) : a face mostrada no lançamento de um dado não viciado

$$\Omega = \{\zeta_i\} = \{1, 2, 3, 4, 5, 6\} \quad \text{realizações}$$

$$\sigma - \text{álgebra} = \{\phi, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \Omega\} \quad \text{eventos}$$

$$\text{Função de probabilidade: } P(\phi) = 0; P(\Omega) = 1; P(\{\zeta_i\}) = \frac{1}{6} \quad \text{probabilidades}$$

Exemplos

2) experimento (\mathcal{F}) : no lançamento de um dado não viciado, os eventos de interesse são $A_1 = \{\text{saída de 3 ou 4}\}$ e $A_2 = \{\text{saída de 1}\}$

A σ – álgebra é composta por ϕ, Ω, A_1, A_2 e todas as possíveis uniões, intersecções e complementos dos eventos acima. Assim

$$\sigma - \text{álgebra} = \left\{ \begin{array}{l} \Omega = \{\zeta_i\} = \{1, 2, 3, 4, 5, 6\} \\ \phi, \Omega, \{1\}, \{3 \cup 4\}, \{\{1\} \cup \{3,4\}\} = \{1 \cup 3 \cup 4\}, \\ \overline{\{1 \cup 3 \cup 4\}} = \{2 \cup 5 \cup 6\}, \\ \overline{\{1\}} = \{2 \cup 3 \cup 4 \cup 5 \cup 6\}, \overline{\{3,4\}} = \{1 \cup 2 \cup 5 \cup 6\} \end{array} \right\}$$

Função de probabilidade:

$$P(\phi) = 0; P(\Omega) = 1; P(\{1\}) = \frac{1}{6}; P(\{3 \cup 4\}) = \frac{1}{3}; P(\{1 \cup 3 \cup 4\}) = \frac{1}{2};$$
$$P(\{2 \cup 3 \cup 4 \cup 5 \cup 6\}) = \frac{5}{6}; P(\{1 \cup 2 \cup 5 \cup 6\}) = \frac{2}{3}$$

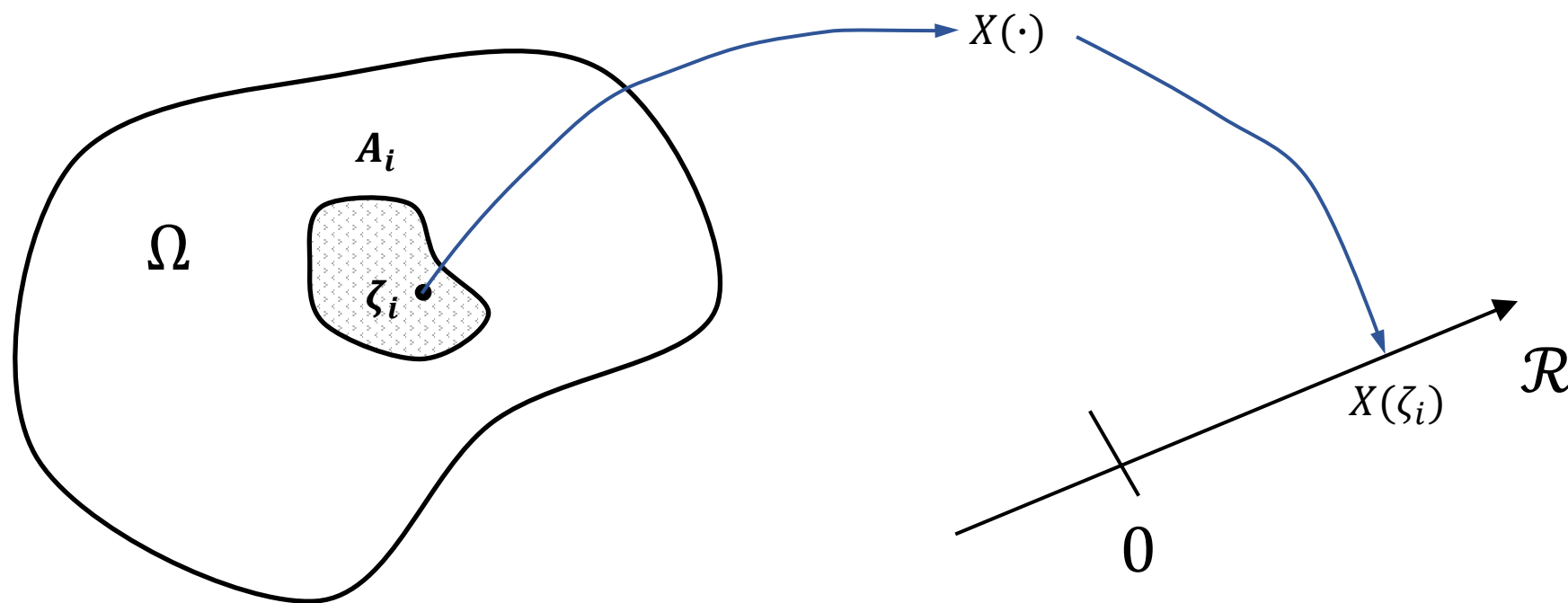
4 – Variáveis aleatórias

- Em Engenharia, é comum termos que lidar com grandezas físicas cuja caracterização depende da realização de medições;
- as medições utilizam sensores que emitem sinais que, convenientemente condicionados, fornecem a informação requerida;
- as grandezas físicas são caracterizadas por números reais, e não relacionadas a objetos, como, por exemplo, em um sorteio de um jogo de cartas ou o número de pontos na face de um dado;
- daí surge o conceito de variável aleatória;

*Dado um experimento definido por seu espaço amostral, sigma-álgebra e função de probabilidade, **variável aleatória** é uma função que mapeia cada evento individual da sigma-álgebra para a reta do conjunto dos números reais*

Variáveis aleatórias

*Dado um experimento definido por seu espaço amostral, sigma-álgebra e função de probabilidade, **variável aleatória** é uma função que mapeia cada evento individual da sigma-álgebra para a reta do conjunto dos números reais*



*As variáveis aleatórias podem ser **discretas** ou **contínuas***

Vejam os um exemplo de VA (ou V.A.) discreta

Experimento $\mathcal{F} \triangleq (\Omega, \sigma_a, P(A_i))$: soma das faces obtidas no lançamento simultâneo de 2 dados não viciados

$$\Omega = \{\zeta_i\} = \{1, 2, 3, 4, 5, 6\}$$
$$\sigma - \text{álgebra} = \left\{ \begin{array}{l} \phi, \Omega, \{11\}, \{12\}, \{21\}, \{13\}, \{31\}, \{22\}, \{14\}, \{41\}, \{23\}, \{32\}, \\ \{15\}, \{51\}, \{33\}, \{24\}, \{42\}, \{16\}, \{61\}, \{25\}, \{52\}, \{34\}, \\ \{43\}, \{44\}, \{26\}, \{62\}, \{35\}, \{53\}, \{45\}, \{54\}, \{36\}, \{63\}, \{55\}, \{64\}, \{46\}, \\ \{56\}, \{65\}, \{66\} \end{array} \right\}$$

Função de probabilidade:

$$P(\phi) = 0; P(\Omega) = 1; P(\{2\}) = \frac{1}{36}; P(\{3\}) = \frac{1}{18}; P(\{4\}) = \frac{1}{12}; P(\{5\}) = \frac{1}{9}; P(\{6\}) = \frac{5}{36}; P(\{7\}) = \frac{1}{6}; P(\{8\}) = \frac{5}{36};$$
$$P(\{9\}) = \frac{1}{9}; P(\{10\}) = \frac{1}{12}; P(\{11\}) = \frac{1}{18}; P(\{12\}) = \frac{1}{36}$$

Denominando A_i a cada evento possível, para obter uma VA associa-se a cada um deles um número $\in \mathbb{R}^1$, ou seja, uma *função mapeando os eventos para a reta real*

exemplo de VA discreta (continuação)

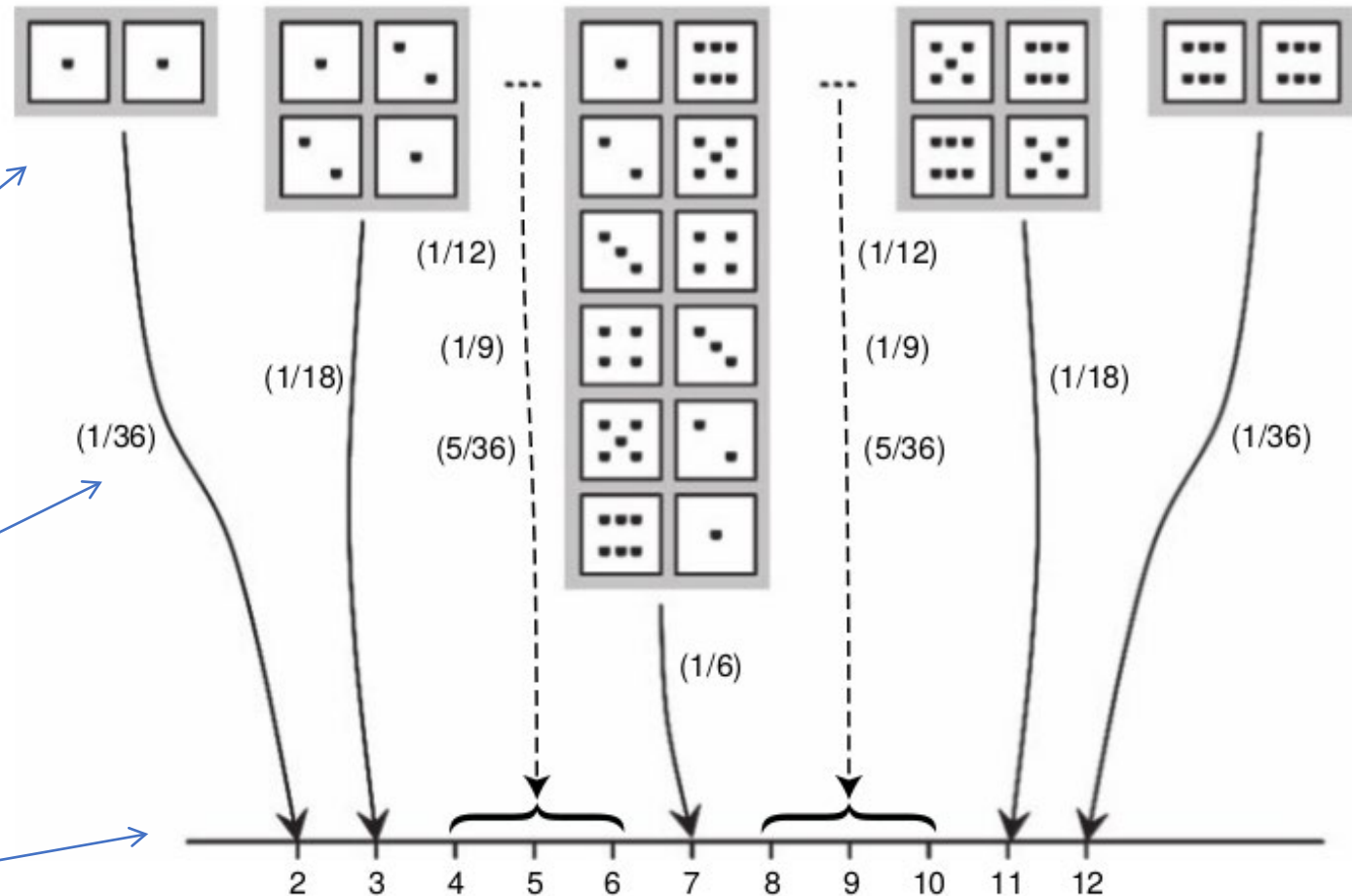
Experimento \mathcal{F} : soma das faces obtidas no lançamento simultâneo de 2 dados não viciados (continuação)

O diagrama ao lado é a expressão gráfica das operações descritas:

realizações ζ_i

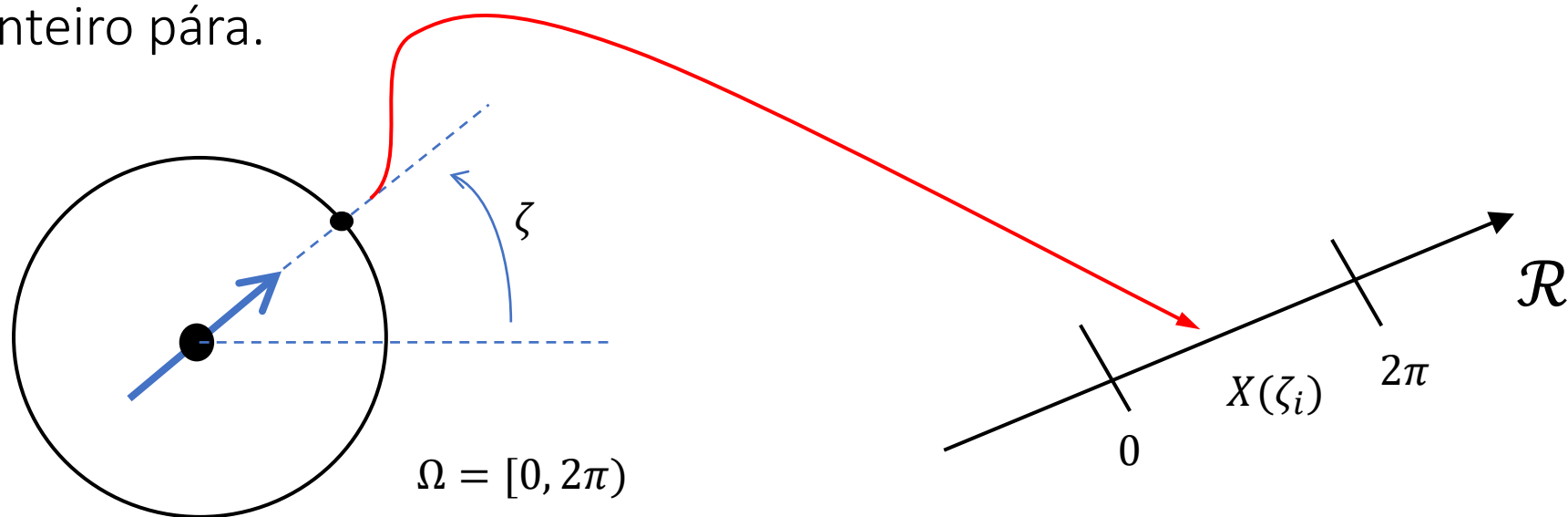
eventos A_i e suas probabilidades

função da VA soma das realizações dos 2 dados



Vejamos um exemplo de VA contínua

Experimento $\mathcal{F} \triangleq (\Omega, \sigma - \acute{a} = ?, P(?))$: girar um ponteiro sobre um disco que não possui marcações em sua borda. Saída do experimento: a posição angular onde o ponteiro pára.



$$P(X(\zeta_i)) = 0, \forall \{\zeta_i\} \subset \Omega$$

No entanto, sabemos que qualquer saída estará em algum local no intervalo contínuo $[0, 2\pi)$; então, como obter a probabilidade de qualquer evento de interesse?

VA contínua: função densidade de probabilidade

Deve-se pensar não em probabilidade $P(\zeta_i)$, mas em uma função *densidade de probabilidade* $f_X(\zeta)$

Assim, é possível obter a probabilidade propriamente dita de que uma saída esteja em um determinado intervalo de interesse integrando a função densidade de probabilidade nesse intervalo:

$$P(\zeta_1 < X(\zeta) < \zeta_2) = \int_{\zeta_1}^{\zeta_2} f_X(\zeta) d\zeta$$

VA contínua: função densidade de probabilidade

No exemplo do disco, supondo-o apoiado sobre uma superfície perfeitamente horizontal, é válido assumir por hipótese que qualquer realização (saída) no intervalo $[0, 2\pi)$ seja equiprovável:

$$f_X(\zeta) = \frac{1}{2\pi}, \quad 0 \leq \zeta < 2\pi$$

função densidade de probabilidade

A probabilidade de $f_X(\zeta)$ estar no intervalo $[\zeta_1, \zeta_2]$ é:

$$P(\zeta_1 < X(\zeta) < \zeta_2) = \int_{\zeta_1}^{\zeta_2} f_X(\zeta) d\zeta = \int_{\zeta_1}^{\zeta_2} \frac{1}{2\pi} d\zeta$$

$$P(\zeta_1 < X(\zeta) < \zeta_2) = \frac{1}{2\pi} (\zeta_2 - \zeta_1)$$

De maneira consistente,

$$P(0 \leq X(\zeta) < 2\pi) = \int_0^{2\pi} f_X(\zeta) d\zeta = 1$$

VA contínua: função distribuição de probabilidade ou de probabilidade cumulativa

De forma genérica, $F_X(x) = \int_0^x f_X(\zeta) d\zeta$, x no espaço da variável aleatória

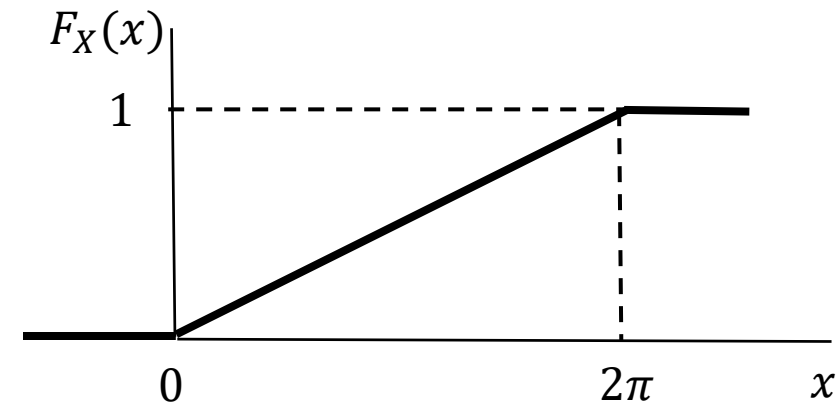
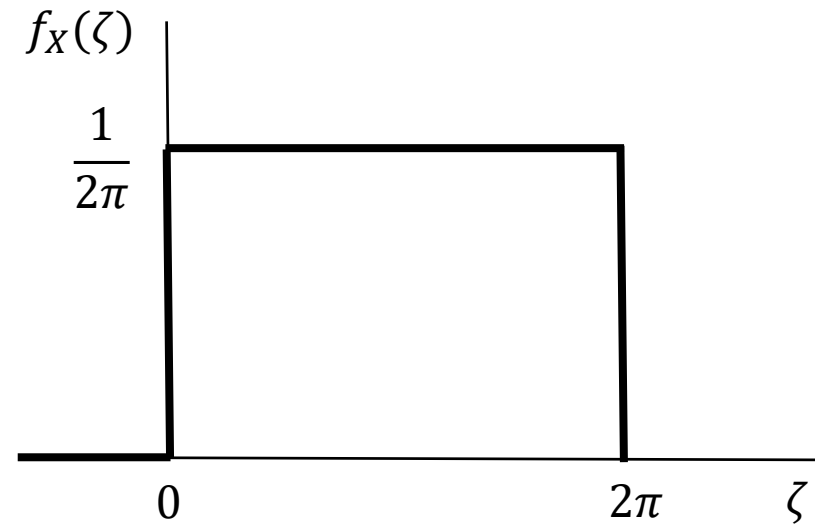
Definição:

$F_X(x) \triangleq$ *função distribuição de probabilidade ou
função de probabilidade cumulativa*

A função densidade de probabilidade e a função distribuição de probabilidade são maneiras alternativas de caracterizar uma variável aleatória contínua.

VA contínua: função distribuição de probabilidade ou de probabilidade cumulativa

No caso do exemplo do disco,



5 – Média e variância de uma variável aleatória

O *valor esperado ou média* de uma VA X contínua é, por definição, dado por

$$E[X] \triangleq \int_{-\infty}^{+\infty} x f_X(x) dx$$

Esse número também é denotado por η_X ou η

O *valor esperado ou média* de uma VA X discreta é dado por

$$E[X] = \sum_i p_i x_i, \quad p_i = P(X = x_i)$$

5 – Média e variância de uma variável aleatória

Considere uma VA X contínua cuja média seja η_X . A VA $(X - \eta_X)$ denota o desvio de X em relação ao seu valor esperado. Para evitar cancelamentos, define-se

$$\sigma^2 \triangleq E[(X - \eta_X)^2] > 0$$

Perceba que $h(X) = (X - \eta_X)^2$ também é uma VA. Calculemos seu valor médio pela expressão anteriormente apresentada:

$$\sigma^2 = E[h(X)] = \int_{-\infty}^{+\infty} h(X) f_X(x) dx = \int_{-\infty}^{+\infty} (X - \eta_X)^2 f_X(x) dx > 0$$

σ^2 é a variância da VA X ; $\sigma = \sqrt{E[(X - \eta_X)^2]}$ é o desvio-padrão da VA X

5 – Média e variância de uma variável aleatória

De $\sigma^2 \triangleq E[(X - \eta_X)^2]$,

$$\begin{aligned}\sigma^2 &= E[X^2 - 2X\eta_X + \eta_X^2] = E[X^2] - 2\eta_X E[X] + \eta_X^2 = E[X^2] - 2E[X]E[X] + E[X]^2 \\ &\Rightarrow \sigma^2 = E[X^2] - E[X]^2\end{aligned}$$

Como $\sigma^2 > 0 \forall X$,

$$\sigma^2 = E[X^2] - E[X]^2 > 0 \Rightarrow E[X^2] > E[X]^2$$

A média e a variância de uma VA são também conhecidos como *momentos* de ordens 1 e 2 respectivamente.

6 – Função característica de uma variável aleatória

A média e a variância de uma VA são também conhecidas como *momentos* de ordens 1 e 2 respectivamente. A função característica associada a uma VA contínua X cuja função densidade de probabilidade é $f_X(x)$ é definida como

$$\psi_X(\omega) \triangleq E[e^{i\omega x}] = \int_{-\infty}^{\infty} f_X(x) e^{i\omega x} dx$$

Nota-se que a função característica de uma VA é exatamente a transformada de Fourier da função densidade de probabilidade $f_X(x)$ trocando-se o sinal da exponencial complexa.

A função característica é particularmente interessante para o cálculo de *momentos de quaisquer ordens*. Explica-se:

6 – Função característica de uma variável aleatória

os momentos de ordem k de uma VA X cuja função densidade de probabilidade é $f_X(x)$ são definidos como

$$E[X^k] \triangleq \int_{-\infty}^{\infty} x^k f_X(x) e^{i\omega x} dx, \quad k \in \mathcal{N}$$

Derivando a função característica com respeito ω e fazendo $\omega = 0$ obtém-se:

$$\left[\frac{d\psi_X}{d\omega} \right]_{\omega=0} = \left[\int_{-\infty}^{\infty} i x f_X(x) e^{i\omega x} dx \right]_{\omega=0} = \int_{-\infty}^{\infty} i x f_X(x) dx$$

$$\left[\frac{d^2\psi_X}{d\omega^2} \right]_{\omega=0} = \left[\int_{-\infty}^{\infty} i^2 x^2 f_X(x) e^{i\omega x} dx \right]_{\omega=0} = \int_{-\infty}^{\infty} i^2 x^2 f_X(x) dx, \quad etc.$$

6 – Função característica de uma variável aleatória

Nota-se que

$$E[X^k] = \frac{1}{i^k} \left[\frac{d\psi_X}{d\omega} \right]_{\omega=0} \quad k \in \mathcal{N}$$

Dessa forma, com o auxílio das tabelas de transformadas de Fourier, pode-se calcular momentos de quaisquer ordem *sem efetuar as respectivas integrações*

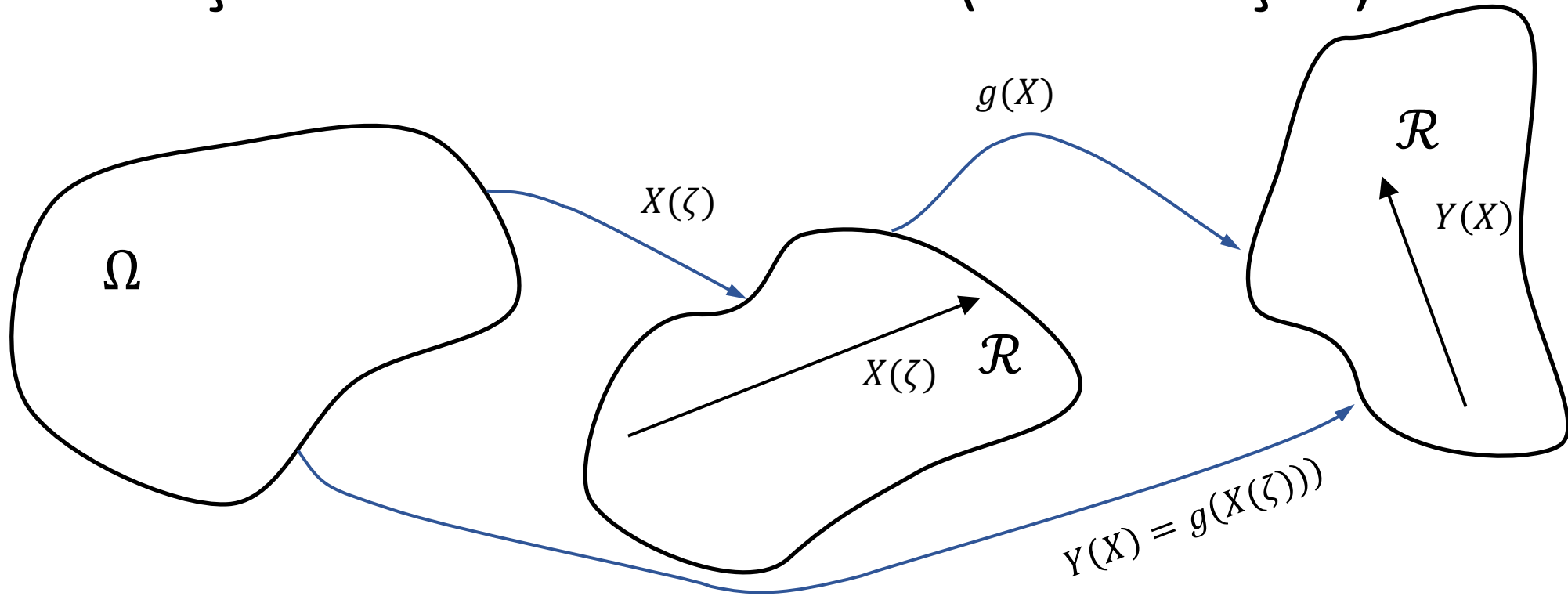
7 – Função de variável aleatória (FVA)

- considere uma variável aleatória $X(\zeta) \in \mathcal{R}$ para todo $\zeta \in \Omega$;
- adote $g(X)$ uma função da variável aleatória X ;
- a expressão $Y = g(X)$ é uma nova VA assim definida:

Para um ζ , $X(\zeta)$ é um número e $g(X)$ é um outro número, especificado em termos de $X(\zeta)$ e de $g(X)$. Esse número é o valor $Y = g(X)$ assumido pela VA Y .

Portanto, a função da VA X é uma função composta $Y = g(X) = g(X(\zeta))$ cujo domínio é o espaço amostral Ω , conjunto de todos os eventos ζ .

7 – Função de variável aleatória (continuação)



Portanto, a função da VA X é uma função composta $Y = g(X) = g(X(\zeta))$ cujo domínio é o espaço amostral Ω , conjunto de todos os eventos ζ .

7 – Função de variável aleatória (continuação)

A *função distribuição de probabilidade* da VA Y representa a probabilidade do evento $\{Y \leq y\}$ correspondente a todas as saídas ζ tais que

$$Y(\zeta) = g(X(\zeta)) \leq y.$$

Assim,

$$F_Y = P(\{Y \leq y\}) = P(\{g(X)\}) \leq y$$

(vale ressaltar que X está sendo considerada contínua)

7 – Função de variável aleatória (continuação)

Exemplo

A função Y da VA X é $Y = X^2$. Obter a função distribuição de probabilidade $F_Y(y)$.

Solução: para obter $F_Y(y) = P(\{Y \leq y\})$, devemos impor que $X^2 \leq Y$.

Se $y \geq 0$, $x^2 \leq y$ para $-\sqrt{x} \leq y \leq \sqrt{x}$. Assim,

$$F_Y(y) = P(\{-\sqrt{x} \leq y \leq \sqrt{x}\}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \text{ (obs.: entenda-se que, aqui } \zeta \equiv x, y)$$

Se $y < 0$, não existe valor de $x \in \mathcal{R}$ tal que $X^2 \leq Y$. Portanto,

$$F_Y(y) = P(\{X^2 \leq Y\}) = P(\{\phi\}) = 0$$

7 – Função de variável aleatória (continuação)

Exemplo (continuação)

É possível também calcular a função densidade de probabilidade de Y , $f_Y(y)$, por derivação da função distribuição de probabilidade com respeito a y :

$$\begin{aligned}\frac{dF_Y(y)}{dy} &= \frac{dF_X(\sqrt{y})}{dy} \cdot \frac{d\sqrt{y}}{dy} - \frac{dF_X(-\sqrt{y})}{dy} \cdot \frac{-d\sqrt{y}}{dy} \Rightarrow \\ f_Y(y) &= f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}, \quad \forall y > 0\end{aligned}$$

$$\therefore f_Y(y) = \begin{cases} f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}, & \forall y > 0 \\ 0, & \forall y \leq 0 \end{cases}$$

8 – Média e variância de uma função de VA

O *valor esperado ou média* de uma função $Y = g(X)$ da VA contínua X é

$$E[Y] \triangleq \int_{-\infty}^{+\infty} y f_Y(y) dy$$

Aparentemente, para obter $E[Y]$ é necessário, antes, calcular $f_Y(y)$. No entanto, o mesmo resultado pode ser obtido a partir do conhecimento de $g(X)$ e $f_X(x)$:

$$E[Y] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

(para a demonstração, ver Papoulis)

8 – Média e variância de uma função de VA

Para a variância de $Y = g(X)$ apresenta-se, sem demonstrar, o resultado obtido por Papoulis:

Supondo que $E[X] = \eta_X$, a variância σ_Y^2 de $Y = g(X)$ é dada por

$$\sigma_Y^2 \approx |g'(\eta_x)|^2 \sigma_X^2$$

9 – Distribuições bivariadas (2 VAs)

São dadas duas funções aleatórias X e Y , ambas definidas a partir dos experimentos

$$\mathcal{F}_X \left(\Omega_X, \sigma_{a_X}, F_X(x) \right), \text{ e } \mathcal{F}_Y \left(\Omega_Y, \sigma_{a_Y}, F_Y(y) \right)$$

Por exemplo, considere o experimento de lançar duas moedas não-viciadas ao mesmo tempo, assim definido:

$\zeta_1 = \{CaCa\}$, $\zeta_2 = \{CaCo\}$, $\zeta_3 = \{CoCo\}$ são as saídas;

$\Omega = \{CaCa, CaCo, CoCo\}$ é o espaço amostral;

$\sigma_a = \{\phi, \Omega, \{\zeta_1\}, \{\zeta_2\}, \{\zeta_3\}, \{\zeta_1 \cup \zeta_2\}, \{\zeta_1 \cup \zeta_3\}, \{\zeta_2 \cup \zeta_3\}\}$ cujas respectivas

probabilidades são $0, 1, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{2}, \frac{3}{4}$

9 – Distribuições bivariadas (2 VAs)

Definem-se, agora, duas VAs,

$$\text{VA } X(\zeta) = \begin{cases} 0, & \text{se pelo menos uma } Ca \text{ sair} \\ 1, & \text{se nenhuma } Ca \text{ sair} \end{cases}$$

$$\text{VA } Y(\zeta) = \begin{cases} -1, & \text{se uma } Co \text{ e uma } Ca \text{ sair} \\ 1, & \text{se } CaCa \text{ ou } CoCo \text{ sair} \end{cases}$$

e a função distribuição de probabilidade conjunta é

$$P(X = 0 \cap Y = 1) = P(\{CaCa\}) = 1/4$$

$$P(X = 0 \cap Y = -1) = P(\{CaCo\}) = 1/2$$

$$P(X = 1 \cap Y = 1) = P(\{CoCo\}) = 1/4$$

$$P(X = 1 \cap Y = -1) = P(\{CoCa\}) = 1/2$$

$$P(X = 0 \cup Y = -1) = P(\{CaCo\}) = 1/2$$

$$P(X = 0 \cup Y = 1) = P(\{CaCa\}) = 1/4$$

A função distribuição de probabilidade para cada VA é:

$$P(X = 0) = \frac{3}{4}; \quad P(X = 1) = \frac{1}{4};$$

$$P(Y = -1) = \frac{1}{2} \quad P(Y = 1) = \frac{1}{2};$$

$$P(X = 1 \cup Y = -1) = P(\{CaCo\}) = 1/2$$

$$P(X = 1 \cup Y = 1) = P(\{CoCo\}) = 1/4$$

9 – Distribuições bivariadas (2 VAs) - formalização

Para duas VAs X e Y genéricas, ambas definidas a partir dos experimentos

$$\mathcal{F}_X \left(\Omega_X, \sigma_{a_X}, F_X(x) \right), \text{ e } \mathcal{F}_Y \left(\Omega_Y, \sigma_{a_Y}, F_Y(y) \right),$$

deseja-se obter sua função distribuição conjunta, isto é, a probabilidade de que um ponto genérico (X, Y) ocupe uma região D no plano. Cada uma das funções distribuição individualmente definem apenas estatísticas marginais de X e de Y , mas não a estatística conjunta,

$$F_{XY}(x, y) = P_X(X \leq x) \cap P_Y(Y \leq y) = P_{XY}(X \leq x, Y \leq y) = P_{XY}((x, y) \in D)$$

9 – Distribuições bivariadas (2 VAs) - formalização

Por definição, a função distribuição conjunta é uma probabilidade; então

$$F_{XY}(x, y) \geq 0, \forall x, y ;$$

$$P(\{x \leq \infty, y \leq \infty\}) \Rightarrow F_{XY}(\infty, \infty) = 1 \text{ (evento certo)}$$

$$P(\{x \leq -\infty, y \leq -\infty\}) \Rightarrow F_{XY}(-\infty, -\infty) = 0 \text{ (evento impossível)}$$

Como

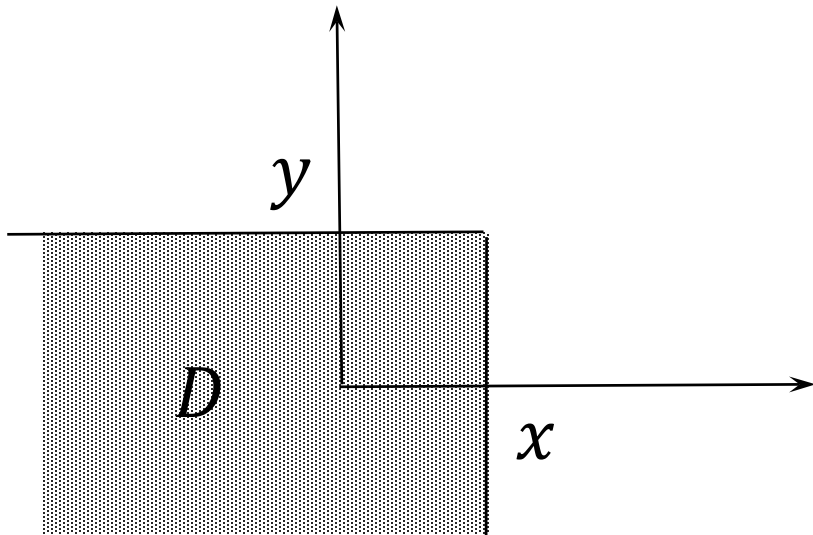
$$P(\{X \leq x, y \leq \infty\}) = P(\{X \leq x\}) \cap P(\{Y \leq \infty\}) = P(\{X \leq x\}) \cap \Omega = P(\{X \leq x\})$$

$$\Rightarrow F_{XY}(x, \infty) = F_X(x)$$

$$\Rightarrow F_{XY}(\infty, y) = F_Y(y)$$

9 – Distribuições bivariadas (2 VAs) - formalização

A distribuição de probabilidade conjunta, $F_{XY}(x, y) = P_{XY}((x, y) \in D)$, é passível de representação:



Como se observa na figura, a distribuição de probabilidade $F_{XY}(x, y)$ é contínua. Sendo também diferenciável, justifica-se a hipótese de que a distribuição conjunta seja o resultado da integração da sua função densidade de probabilidade conjunta, definida por

$$f_{XY}(x, y) \triangleq \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y)$$

9 – Distribuições bivariadas (2 VAs) - formalização

A dupla integração da função densidade de probabilidade conjunta, $f_{XY}(x, y)$, irá fornecer a função distribuição de probabilidade conjunta, $F_{XY}(x, y)$:

$$f_{XY}(x, y) \triangleq \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y)$$
$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(\alpha, \beta) d\alpha d\beta$$

A partir da densidade e da distribuição de probabilidade conjunta, é possível obter as respectivas estatísticas marginais como se segue

9 – Distribuições bivariadas (2 VAs) - formalização

$$F_{XY}(x, \infty) = F_X(x) \Rightarrow \int_{-\infty}^{\infty} f_{XY}(x, \beta) d\beta = f_X(x) ;$$

$$F_{XY}(\infty, y) = F_Y(y) \Rightarrow \int_{-\infty}^{\infty} f_{XY}(\alpha, y) d\alpha = f_Y(y) ;$$

Prova para as distribuições marginais $F_X(x)$ e $F_Y(y)$:

$$\{X \leq \infty\} = \{Y \leq \infty\} = \Omega. \text{ Então, } \{X \leq x\} = \{X \leq x, Y \leq \infty\} = P(\{X \leq x, Y \leq \infty\}) =$$

$$F_{XY}(x, \infty) = F_X(x)$$

$$\{Y \leq y\} = \{X \leq \infty, Y \leq y\} = P(\{X \leq \infty, Y \leq y\}) = F_{XY}(\infty, y) = F_Y(y)$$

9 – Distribuições bivariadas (2 VAs) - formalização

$$F_{XY}(x, \infty) = F_X(x) \Rightarrow \int_{-\infty}^{\infty} f_{XY}(x, \beta) d\beta = f_X(x) ;$$

$$F_{XY}(\infty, y) = F_Y(y) \Rightarrow \int_{-\infty}^{\infty} f_{XY}(\alpha, y) d\alpha = f_Y(y) ;$$

Prova para as densidades marginais $f_X(x)$ e $f_Y(y)$:

$$\frac{\partial}{\partial x} F_{XY}(x, y) = \frac{\partial}{\partial x} \left\{ \int_{-\infty}^x \int_{-\infty}^y f_{XY}(\alpha, \beta) d\alpha d\beta \right\} = \int_{-\infty}^y f_{XY}(x, \beta) d\beta$$

$$\frac{\partial}{\partial y} F_{XY}(x, y) = \frac{\partial}{\partial y} \left\{ \int_{-\infty}^x \int_{-\infty}^y f_{XY}(\alpha, \beta) d\alpha d\beta \right\} = \int_{-\infty}^x f_{XY}(\alpha, y) d\alpha$$

9 – Distribuições bivariadas (2 VAs) - formalização

$$F_{XY}(x, \infty) = F_X(x) \Rightarrow \int_{-\infty}^{\infty} f_{XY}(x, \beta) d\beta = f_X(x) ;$$

$$F_{XY}(\infty, y) = F_Y(y) \Rightarrow \int_{-\infty}^{\infty} f_{XY}(\alpha, y) d\alpha = f_Y(y) ;$$

Prova para as densidades *marginais* $f_X(x)$ e $f_Y(y)$ (continuação)

Levando o limite de integração de ambas as integrais até ∞ obtêm-se as expressões procuradas:

$$\begin{aligned} \int_{-\infty}^y f_{XY}(x, \beta) d\beta &\Rightarrow \int_{-\infty}^{\infty} f_{XY}(x, \beta) d\beta = f_X(x) = \frac{\partial}{\partial x} F_{XY}(x, \infty) \\ \int_{-\infty}^x f_{XY}(\alpha, y) d\alpha &\Rightarrow \int_{-\infty}^{\infty} f_{XY}(\alpha, y) d\alpha = f_Y(y) = \frac{\partial}{\partial y} F_{XY}(\infty, y) \end{aligned}$$

9 – Distribuições bivariadas (2 VAs) - formalização

A *covariância* entre as VAs escalares X e Y é o número definido por

$$C_{XY} = E[(X - \eta_X)(Y - \eta_Y)]$$

Desenvolvendo a expressão entre colchetes tem-se:

$$C_{XY} = E[XY] - E[X\eta_Y] - E[Y\eta_X] + E[\eta_X\eta_Y] \Rightarrow$$

$$C_{XY} = E[XY] - E[X]\eta_Y - E[Y]\eta_X + E[X]E[Y] \Rightarrow$$

$$C_{XY} = E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] \Rightarrow$$

$$C_{XY} = E[XY] - E[X]E[Y]$$

9 – Distribuições bivariadas (2 VAs) - formalização

O *coeficiente de correlação* entre as VAs escalares X e Y é o número definido por

$$\rho_{XY} \triangleq \frac{C_{XY}}{\sigma_X \sigma_Y}, \quad \text{com} \quad \begin{cases} |C_{XY}| \leq \sigma_X \sigma_Y \\ |\rho_{XY}| \leq 1 \end{cases}$$

Duas VAs são *não-correlacionadas* quando sua covariância for zero :

$$C_{XY} = 0 \Rightarrow \rho_{XY} = 0 \Rightarrow E[XY] = E[X]E[Y]$$

Duas VAs são *ortogonais* se

$$E[X]E[Y] = 0$$

9 – Distribuições bivariadas (2 VAs) - formalização

Se duas VAS forem independentes, isto é, se

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

serão também não-correlacionadas

$$C_{XY} = 0 \Rightarrow \rho_{XY} = 0 \Rightarrow E[XY] = E[X]E[Y]$$

A recíproca , entretanto, não é verdadeira. Considere o exemplo a seguir:

admite-se que uma VA Z possua distribuição uniforme no intervalo $[0,1]$. Assim,

$$f_Z(\zeta) = \begin{cases} 1 & \text{para } \zeta \in [0,1] \\ 0 & \text{para } \zeta \notin [0,1] \end{cases}$$

9 – Distribuições bivariadas (2 VAs) - formalização

Montam-se duas funções da VA Z : $\begin{cases} X_Z(\zeta) = \sin(2\pi\zeta) \\ Y_Z(\zeta) = \cos(2\pi\zeta) \end{cases}$

X, Y são não-correlacionadas pois $E[X] = E[Y] = E[XY] = 0$;

X, Y não são independentes. Para provar essa afirmação, considera-se momentos conjuntos de ordem superior:

$$\begin{aligned} 4a. \text{ ordem} &\rightarrow E[X^2 Y^2] = \frac{1}{8} \\ E[X^2] = E[Y^2] &= \frac{1}{2} \implies E[X^2] \cdot E[Y^2] = \frac{1}{4} \\ \therefore E[X^2 Y^2] &\neq E[X^2] \cdot E[Y^2] \end{aligned}$$

Se X e Y fossem independentes, momentos conjuntos de qualquer ordem deveriam ser iguais ao produto dos momentos individuais.

9 – Distribuições bivariadas (2 VAs) - formalização

- Pelo exemplo anterior, percebe-se que os dois primeiros momentos estatísticos não são suficientes para caracterizar univocamente a independência/correlação entre VAs genéricas;
- há, no entanto, uma *importante exceção* a esse fato: caso as VAs sejam normais, ou gaussianas, apenas os dois primeiros momentos, o valor esperado (ou média) e a correlação/covariância caracterizam completamente a distribuição conjunta;
- assim, se duas VAs X, Y forem gaussianas e não correlacionadas, serão também independentes e vice-versa;
- a distribuição gaussiana será abordada mais adiante.

9 – Distribuições bivariadas (2 VAs) - formalização

Considere duas VAs contínuas X, Y e uma função $Z = g(X, Y)$. A média (ou valor esperado) da distribuição conjunta é, por definição, dada por

$$E[Z] \triangleq \int_{-\infty}^{+\infty} z f_Z(z) dz = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(X, Y) f_{XY}(x, y) dx dy$$

Considere duas VAs discretas X, Y e uma função $Z = g(X, Y)$. A média (ou valor esperado) da distribuição conjunta é, por definição, dada por

$$E[Z] \triangleq \sum_j \sum_k g(X_j, Y_k) P_{jk},$$

com P_{jk} igual à probabilidade conjunta dos eventos X_j e Y_k .

10 – A distribuição gaussiana e o Teorema do Limite Central

- a função densidade de probabilidade de uma VA X cuja distribuição é gaussiana possui a forma genérica dada por

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\eta)^2}{2\sigma^2}}, \quad \text{com } \eta \equiv E[x] \text{ e } \sigma^2 \equiv E[x^2]$$

- A distribuição gaussiana é importante pois a soma de diversas VAs independentes cujas funções distribuição de probabilidade (ou densidade de probabilidade) sejam quaisquer, resulta em uma outra VA cuja distribuição tende a uma gaussiana;
- curiosamente, diversos fenômenos naturais são constituídos pela soma de inúmeras contribuições de eventos aleatórios independentes e distintos;

10 – A distribuição gaussiana e o Teorema do Limite Central

Esse fato peculiar é expresso de maneira formal no *Teorema do Limite Central*: dada uma VA Y construída a partir da soma de n VAs X_j independentes, isto é, $Y = X_1 + X_2 + \cdots + X_n$, sua média e variância são, respectivamente,

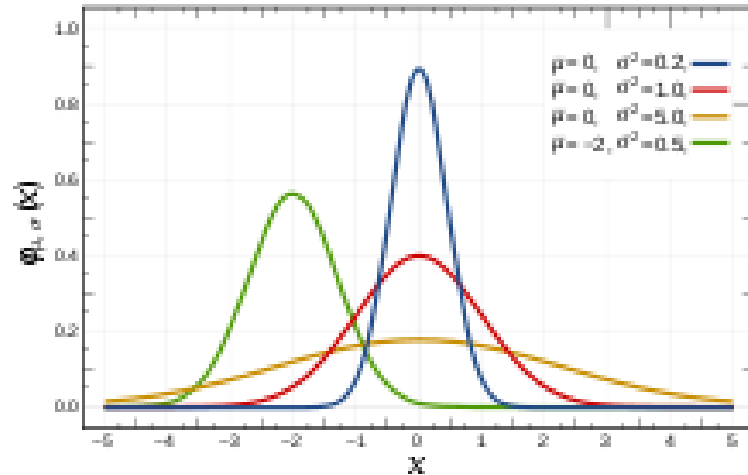
$$E[Y] = \eta_Y = E[X_1] + E[X_2] + \cdots + E[X_n] = \eta_{X_1} + \eta_{X_2} + \cdots + \eta_{X_n}$$

$$E[Y^2] = \sigma_Y^2 = E[X_1^2] + E[X_2^2] + \cdots + E[X_n^2] = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \cdots + \sigma_{X_n}^2$$

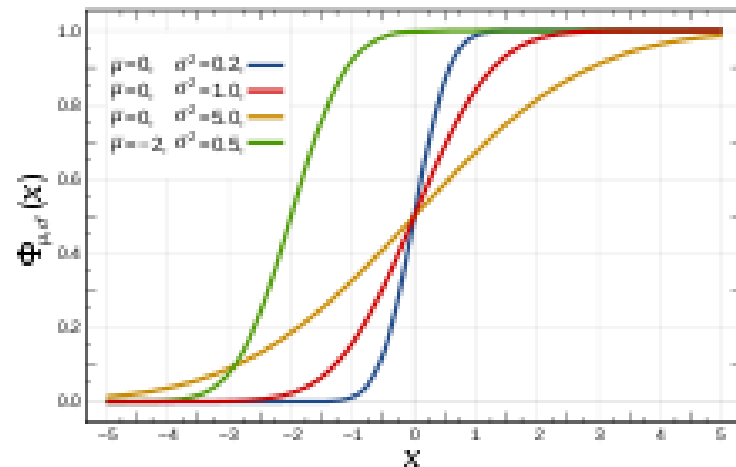
O *Teorema do Limite Central* afirma que, sob certas circunstâncias, as funções distribuição e densidade de probabilidade da VA Y tenderão à suas homóloga gaussiana com mesmas média e variância à medida que n aumenta:

$$\lim_{n \rightarrow \infty} F_Y(y) \simeq \mathcal{N}_Y(\eta, \sigma^2)$$

10 – A distribuição gaussiana e o Teorema do Limite Central



Qual o valor de n mínimo para que uma soma de distribuições quaisquer tender a uma gaussiana?



Não há uma regra definitiva para a escolha de n .
Dependendo das distribuições individuais, esse valor pode sofrer grandes variações. Um exemplo elucidado esse ponto.

10 – A distribuição gaussiana e o Teorema do Limite Central

Considere uma VA Y formada pela soma de VAs X_j , com $j = 1, 2, 3$ uniformemente distribuídas no intervalo $[0, 1]$ com estatística

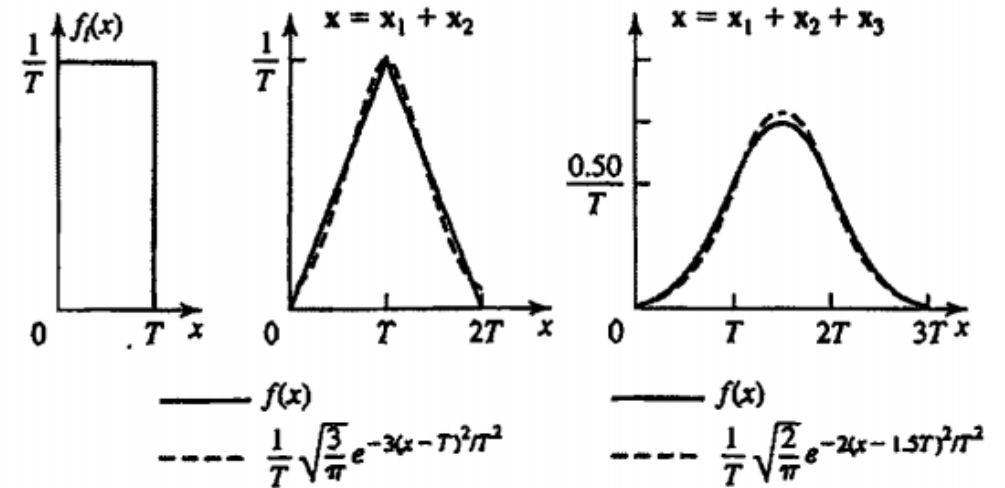
$$\eta_1 = \frac{T}{2}, \quad \sigma_1^2 = \frac{T^2}{12}$$

$$\eta_2 = T, \quad \sigma_2^2 = \frac{T^2}{6}$$

$$\eta_3 = \frac{3T}{2}, \quad \sigma_3^2 = \frac{T^2}{4}$$

$$f_{Y_2}(y) \simeq \frac{1}{T} \sqrt{\frac{3}{\pi}} e^{-\frac{3(y-T)^2}{T^2}}$$

$$f_{Y_3}(y) \simeq \frac{1}{T} \sqrt{\frac{2}{\pi}} e^{-\frac{2(y-1.5T)^2}{T^2}}$$



11 – Sequências de variáveis aleatórias

O *vetor aleatório* $\mathbf{X}^T = [X_1, X_2, \dots, X_n]$ é um vetor cujos componentes são VAs contínuas, discretas ou mistas. Basicamente, é uma *generalização* das distribuições bivariadas.

A probabilidade de que \mathbf{X} esteja em uma determinada região \mathcal{D} do espaço n -dimensional é dada por

$$P[\mathbf{X} \in \mathcal{D}] \triangleq \int_{\mathcal{D}} f_{\mathbf{X}}(\mathbf{X}) d\mathbf{X}, \text{ onde}$$
$$f_{\mathbf{X}}(\mathbf{X}) = f_{\mathbf{X}}(X_1, X_2, \dots, X_n) = \frac{\partial^n F_{\mathbf{X}}(X_1, X_2, \dots, X_n)}{\partial X_1, \partial X_2, \dots, \partial X_n}$$

é a *função densidade de probabilidade conjunta* (ou multivariável) das VAs X_j

e $F_{\mathbf{X}}(X_1, X_2, \dots, X_n) = P(\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\})$

é a distribuição conjunta de probabilidade

11 – Sequências de variáveis aleatórias

Em $F_{\mathbf{X}}(X_1, X_2, \dots, X_n)$, seja $X_k = \infty$. Então $F_{\mathbf{X}}(X_1, \dots, X_k = \infty, \dots, X_n)$ será a distribuição de probabilidade conjunta das $n - k$ VAs remanescentes.

De maneira semelhante, caso a função densidade de probabilidade conjunta $f_{\mathbf{X}}(X_1, X_2, \dots, X_n)$ seja integrada apenas com respeito a alguma(s) VAs, a função distribuição de probabilidade conjunta será correspondente às VAs remanescentes.

Exemplo: seja o *vetor aleatório* $\mathbf{X}^T = [X_1, X_2, X_3, X_4]$

$$F_{\mathbf{X}}(X_1, X_3) = F_{\mathbf{X}}(X_1, \infty, X_3, \infty)$$

$$f_{\mathbf{X}}(X_1, X_3) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{\mathbf{X}}(X_1, X_2, X_3, X_4) dX_2 dX_4$$

12 – Processos estocásticos



PRÓXIMA AULA