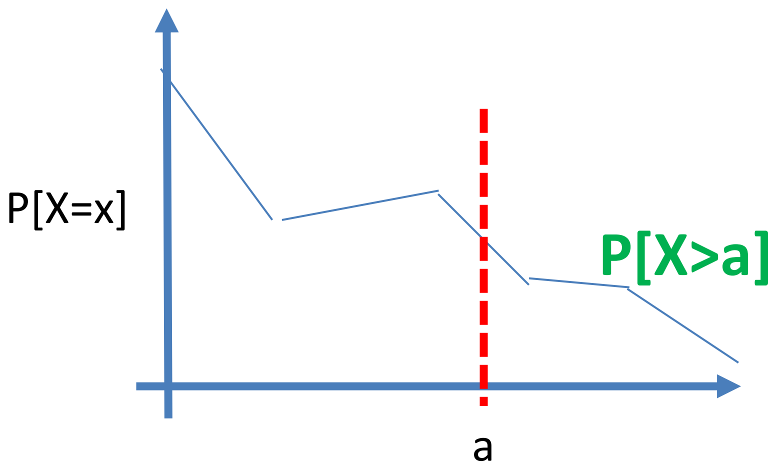
Variáveis Aleatórias em Situações Limite

Soma de Variáveis Aleatórias
Desigualdades de Markov e Chebyshev
Lei dos Grandes Números
Teorema do Limite Central

Motivação

- Vimos que E[*X*] dá informação sobre o valor médio de uma variável aleatória *X*
  - Muito útil para muitos problemas
- No entanto, em diversas situações interessa-nos saber a probabilidade para valores distantes de E[*X*]
  - Ou seja, o que acontece na "cauda" (tail) da distribuição



Motivação

- Também nos interessam situações limite

- Exemplo:
  - **o que acontece quando *n* tende para infinito** ao valor esperado da Média de *n* medições

- Outro exemplo, muito importante:
  - Ao fazermos centenas de milhar ou milhões de experiências nas nossas simulações (teoria frequencista) estamos de facto a garantir boas estimativas das probabilidades ?

Soma de variáveis aleatórias

- Vimos anteriormente um exemplo de soma de variáveis de Bernoulli
  – Na apresentação da Distribuição Binomial

- O que acontece se somarmos outros tipos de variáveis?

Soma de variáveis aleatórias

- Se somarmos duas variáveis aleatórias X_1 e X_2 quais as características da variável aleatória $S = X_1 + X_2$?
  - Em termos de momentos ?
    - Em especial média e variância
  - Em termos de função de distribuição ?
- E no caso geral $S_n = X_1 + X_2 + \cdots + X_n$?

Média da soma de *n* variáveis

- Sejam $X_1, X_2, \dots, X_n$, n variáveis aleatórias e $S_n = X_1 + X_2 + \cdots + X_n$ a sua soma
- Teorema: **O valor esperado da soma de *n* variáveis é igual à soma das médias**
- Demonstração

$$
\begin{aligned}
E[S_n] &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \Big(\sum_{j=1}^{n} x_j\Big) f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n = \\
&= \sum_{j=1}^{n} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_j f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n = \\
&= \sum_{j=1}^{n} \int_{-\infty}^{+\infty} x_j f_{X_j}(x_j) dx_j = \sum_{j=1}^{n} E[X_j]
\end{aligned}
$$

Variância da soma de *n* variáveis

- Considerando da mesma forma $S_n = X_1 + X_2 + \cdots + X_n$:
- Teorema: **A variância da soma de *n* variáveis aleatórias é dada pela soma de todas as variâncias e covariâncias**

$$Var(S_n) = \sum_{i=1}^{n} Var(X_i) + \sum_{\substack{j=1 \\ j \neq k}}^{n} \sum_{k=1}^{n} Cov(X_j, X_k)$$

- Demonstração:

$$\begin{aligned} Var\left(S_n\right) &= E\left[\sum_{j=1}^{n} \left(X_j - E\left[X_j\right]\right) \sum_{k=1}^{n} \left(X_k - E\left[X_k\right]\right)\right] = \\ &= \sum_{j=1}^{n} \sum_{k=1}^{n} E\left[\left(X_j - E\left[X_j\right]\right)\right] E\left[\left(X_k - E\left[X_k\right]\right)\right] \end{aligned}$$

Variância da soma de *n* variáveis

- Se as variáveis **são independentes**, $Cov\left(X_j, X_k\right) = 0$, para todo o $j \neq k$, pelo que:

- $\boldsymbol{Var(S_n) = \sum_{i=1}^{n} Var(X_i)}$
  - **Variância da soma igual a soma das variâncias**

- Se para além de independentes forem **identicamente distribuídas (IID)**
  e tivermos $E[X_i] = \mu$ e $Var(X_i) = \sigma^2, i = 1,2, \dots, n$
  a média e variância da soma são dadas por:

- $\mathrm{E}[S_n] = n\,\mu$ e $Var(S_n) = n\,\sigma^2$

Função de distribuição da soma de 2 variáveis aleatórias independentes

- Caso discreto (2 v.a. Discretas X e Y)
- Fazendo $\mathrm{Z} = X + Y$
- $p_Z(z) = P(X + Y = z)$

$$=\sum_{\{(x,y)|x+y=z\}} P(X = x, Y = y)$$

$$= \sum_x P(X = x, Y = z - x)$$

$$= \sum_x p_X(x) \;\; p_Y(z - x) \qquad \text{; devido à indep.}$$

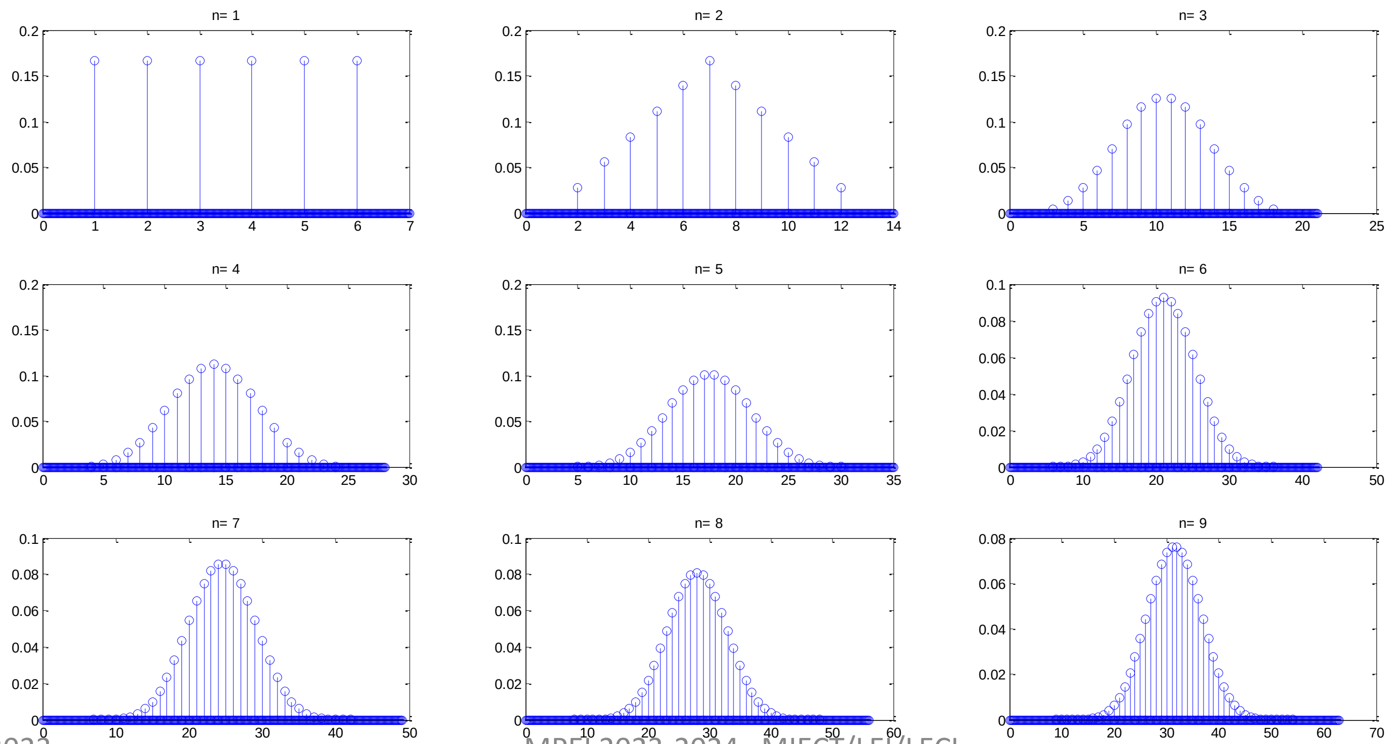
$$= p_X(x) * p_Y(z)$$

demoConvolucao.m

- Que é a **convolução** discreta de p_X e p_Y

Exemplo (em Matlab)

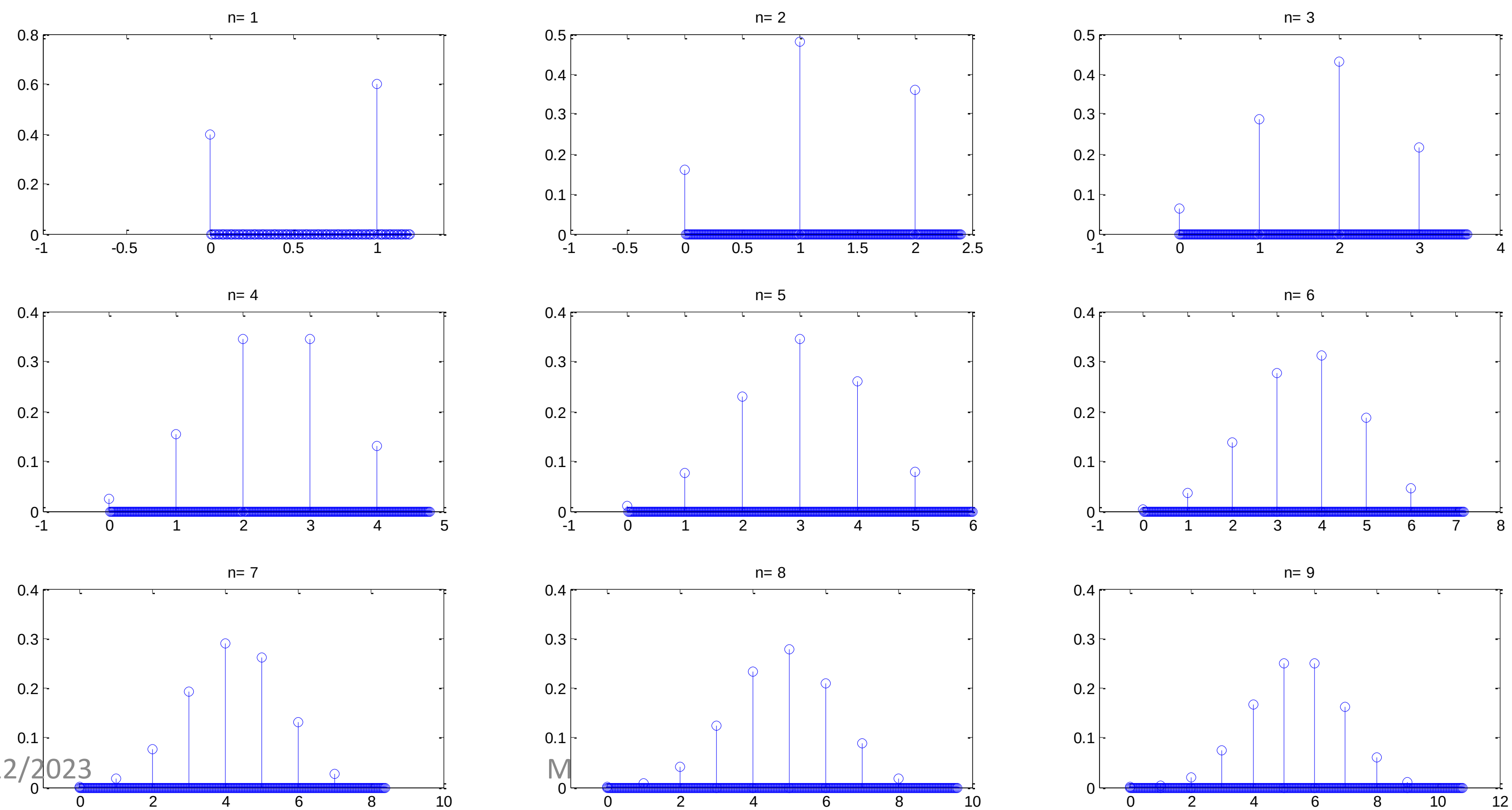
- Usando conv() e a f.m.p relativa à variável *X* correspondente ao lançamento de um dado honesto (*n*=1, 2, ..., 9)



Outro exemplo

demoConvolucaoMoeda.m

- Sendo *X* relativa ao número de caras num lançamento de moeda não honesta
  - com probabilidade de cara = 0,6



Média e variância da Média

- Se criarmos a v.a. relativa à média de *n* variáveis IID X_i, $\boldsymbol{M_n = \frac{S_n}{n}}$
- assumindo $E[X_i] = \mu$ e $Var(X_i) = \sigma^2$, teremos :

$$\boldsymbol{E[M_n]} = E\left[\frac{S_n}{n}\right] = \frac{\sum_i E[X_i]}{n} = E[X_i] = \boldsymbol{\mu}$$

- $\mathbf{Var}[\boldsymbol{M_n}] = Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\frac{\sum_i Var[X_i]}{1} = \frac{Var(X_i)}{n} = \boldsymbol{\frac{\sigma^2}{n}}$
- À medida que se aumenta o número de experiências vai diminuindo a variância da estimativa da média

Questões

1. Quão provável é termos um valor superior a um determinado valor?
   – Exemplo: o dobro do valor esperado

2. Quão próxima a média obtida com as amostras fica do valor médio (valor esperado) ?

3. Qual a distribuição da média para valores de *n* muito grandes ?

Desigualdades de Markov e Chebyshev

- Os dois teoremas que apresentaremos de seguida, sem muita preocupação com demonstrações, permitem estabelecer facilmente majorantes para probabilidades de certas classes de acontecimentos
  - partindo apenas do conhecimento da média e variância de uma variável aleatória

  - Mais informação, por exemplo, na secção 4.6 do livro de F. Vaz e A. Teixeira (Bibliografia)

Questão 1

- Probabilidade de termos valores superiores a um determinado valor ?

- Exemplo:
- A média das classificações numa turma é 15,2.
- Será que conseguimos determinar um limite superior para probabilidade de um dos alunos ter nota igual ou superior a 17 ?

Desigualdade de Markov

- Seja X uma variável aleatória **não negativa**
- Pela Desigualdade de Markov:

$$P(X \geq a) \leq \frac{E[X]}{a}, \ \forall a > 0$$

- Esta desigualdade dá-nos um limite superior para a probabilidade de a função X ser maior ou igual a um determinado valor
- Qual o valor de P com a = E[X] ?
  - E a < E(X) ?
  - E a > E(X) ?

Desigualdade de Markov

- Demonstração:
- $E[X] = ?$
- $= \cancel{\int_0^a x f_X(x) dx} + \int_a^\infty x f_X(x) dx \geq$
- $\geq \int_a^\infty x f_X(x) dx \geq \int_a^\infty a f_X(x) dx =$
- $\geq \; a \, P[X \geq a]$
- Logo: $P[X \geq a] \leq \frac{E[X]}{a}$

Exemplo 1

- A média da altura de uma população é 1,65 m.
- Qual o limite superior de probabilidade de um indivíduo ultrapassar os 2 metros ?

- $P(X \geq 2) \leq \frac{1,65}{2} = 0,825$

(Nota: Limite não muito útil ou significativo !)

Exemplo 2

- A média das classificações numa turma é 15,2.
- Qual o limite superior de probabilidade de um dos alunos ter nota igual ou superior a 17 ?

- $P(X \geq 17) \leq \frac{15,2}{17} = 0.8941$

- E superior a 19?

- $P(X \geq 19) \leq \frac{15,2}{19} = 0.8$

Questão 2

- Quão provável é a **diferença entre a variável e o seu valor esperado** ser superior/inferior a um determinado valor ?

- Isto é $P(|X - E[X]| \geq a) = ?$

  Ou $P(|X - E[X]| < \ a) = ?$

- Exemplo: Probabilidade de os valores diferirem da média mais que 2 desvios padrão ?

Desigualdade de Chebyshev

- Pela Desigualdade de Chebyshev temos:
- $P(|X - E[X]| \geq a) \leq \frac{Var(X)}{a^2}$

- Ou, em alternativa:
- $P(|X - E[X]| < a) \geq 1 - \frac{Var(X)}{a^2}$

Desigualdade de Chebyshev

- Demonstração:

- Define-se $D^2 = (X - E[X])^2$
  - É óbvio que $D^2 \geq 0$ e $D^2 \geq a^2 \Leftrightarrow |D| \geq a$
- Aplicando a Desigualdade de Markov

- $P(|D| \geq a) = P(D^2 \geq a^2)$
- $P(D^2 \geq a^2) \leq \frac{E\left[(X-E[X])^2\right]}{a^2}$
- $P(D^2 \geq a^2) \leq \frac{Var(X)}{a^2}$; $P\left((X - E[X])^2 \geq a^2\right) \leq \frac{Var(X)}{a^2}$

- Assume-se E[*X*] e Var(*X*) finitos

Desigualdade de Chebyshev

- Se expressarmos a em função do desvio padrão, fazendo $a = h\sigma$, teremos:

$$P(|X - E[X]| \geq h\sigma) \leq \frac{\sigma^2}{(h\sigma)^2} = \frac{1}{h^2}$$

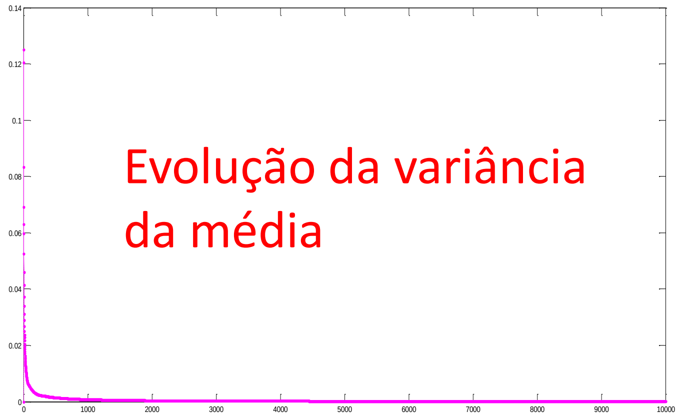
- Ou seja: a probabilidade de obter um valor que dista da média de h desvios padrão ou mais é menor ou igual a $\frac{1}{h^2}$
  - Exemplos:
    - h=1 => P <= 1
    - **h=2 => P <= ¼** Valores com pouca precisão

Questão 3

- Ao fazermos centenas de milhar ou milhões de experiências nas nossas simulações (teoria frequencista) estamos de facto a garantir boas estimativas das probabilidades ?

Voltando à média de *n* variáveis aleatórias ...

- Como vimos, **a variância da média das estimativas** tende para 0 à medida que *n* aumenta



- O que se pode interpretar como a probabilidade da média das amostras se aproximar do valor médio ser cada vez maior
  - aproximando-se de 1

Voltando à média de *n* variáveis aleatórias ...

- Qual a probabilidade da média das amostras se aproximar do valor médio (a menos de ϵ) ?
  - Ou seja: $P(|M_n - E[M_n]| < \epsilon)$

- Recorrendo à Desigualdade de Chebyshev temos:
- $P(|M_n - E[M_n]| \geq \epsilon) \leq \frac{Var(M_n)}{\epsilon^2}$
- $P(|M_n - E[M_n]| \geq \epsilon) \leq \frac{\frac{\sigma^2}{n}}{\epsilon^2}$
- $P(|M_n - E[M_n]| < \epsilon) \geq 1 - \frac{\sigma^2}{n\,\epsilon^2}$

Lei fraca dos grandes números

- Passando ao limite a última expressão teremos:

$$\lim_{n\to\infty} P(|M_n - E[M_n]| < \epsilon) = 1$$

- Resultado que é conhecido por **Lei Fraca dos Grandes Números**

Leis dos grandes números (LGN)

- Existe um segundo enunciado (fora dos objectivos de MPEI), a **lei forte dos grandes números**, que afirma:

$$P(\lim_{n\to\infty} M_n = \mu) = 1$$

- A **Lei Fraca dos Grandes Números** afirma que para um valor de *n* suficientemente elevado a média das amostras estará muito próxima do valor esperado

- Enquanto que a **lei forte** garante que é certo que o limite para que tende a média (das amostras) é o valor esperado

L. G. N. e definição frequencista

- Consideremos uma sequência de experiências aleatórias independentes
- e seja I_j uma variável aleatória indicadora da ocorrência do evento A na experiência de ordem j

  [1 significa que A ocorreu]

- O número total de ocorrências de A nas *n* experiências será:

$$N_n = I_1 + I_2 + \cdots + I_n$$

L. G. N. e definição frequencista

- Como a frequência relativa de A é

$$f_A(n) = \frac{(I_1 + I_2 + \cdots + I_n)}{n}$$

- f_A é a média das variáveis aleatórias I_i
  - E estimativa de $p(A)$

- Então (pelas duas leis dos grandes números):

$$\lim_{n \to \infty} P(|f_A(n) - p(A)| < \epsilon) = 1$$

e

$$P[\lim_{n \to \infty} f_A(n) = p(A)] = 1$$

- Permitindo-nos dizer que **a frequência relativa é uma boa estimativa da probabilidade**

Um pouco de História (para terminar esta parte)

- 1713: Lei fraca descrita por Jacob **Bernoulli**
- 1835: **Poisson** chama-lhe “La Loi des Grands Nombres”
  - Lei dos Grandes Números em Francês
- 1909: Émile Borel desenvolve a Lei forte para variáveis de Bernoulli
- 1928: Andrei Nikolaevich **Kolmogorov** prova a Lei forte no caso geral

Qual a distribuição de M_n para valores de *n* muito grandes ?

Questão

- Já vimos o comportamento limite da média de uma sequência de variáveis aleatórias

- Conseguimos avançar mais e dizer alguma coisa quanto à distribuição ?

- Comecemos com alguns exemplos …

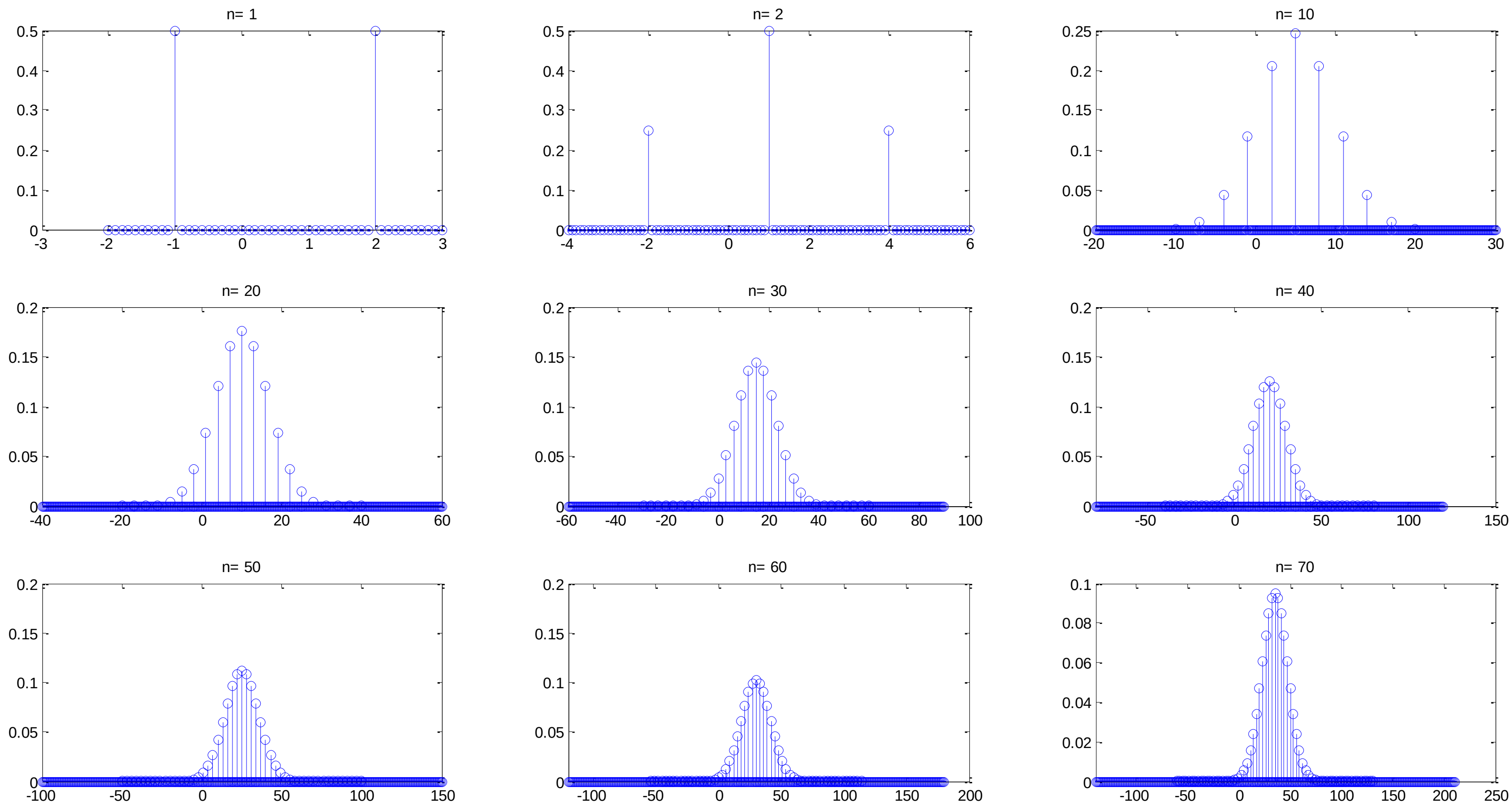
Exemplo 1

- Consideremos um jogo em lançamos uma moeda ao ar e **perdemos 1 Euro se sair CARA** e **ganhamos 2 Euros se sair COROA**
- A moeda é honesta e existe independência entre as jogadas

- Como se comporta a distribuição com as jogadas ?

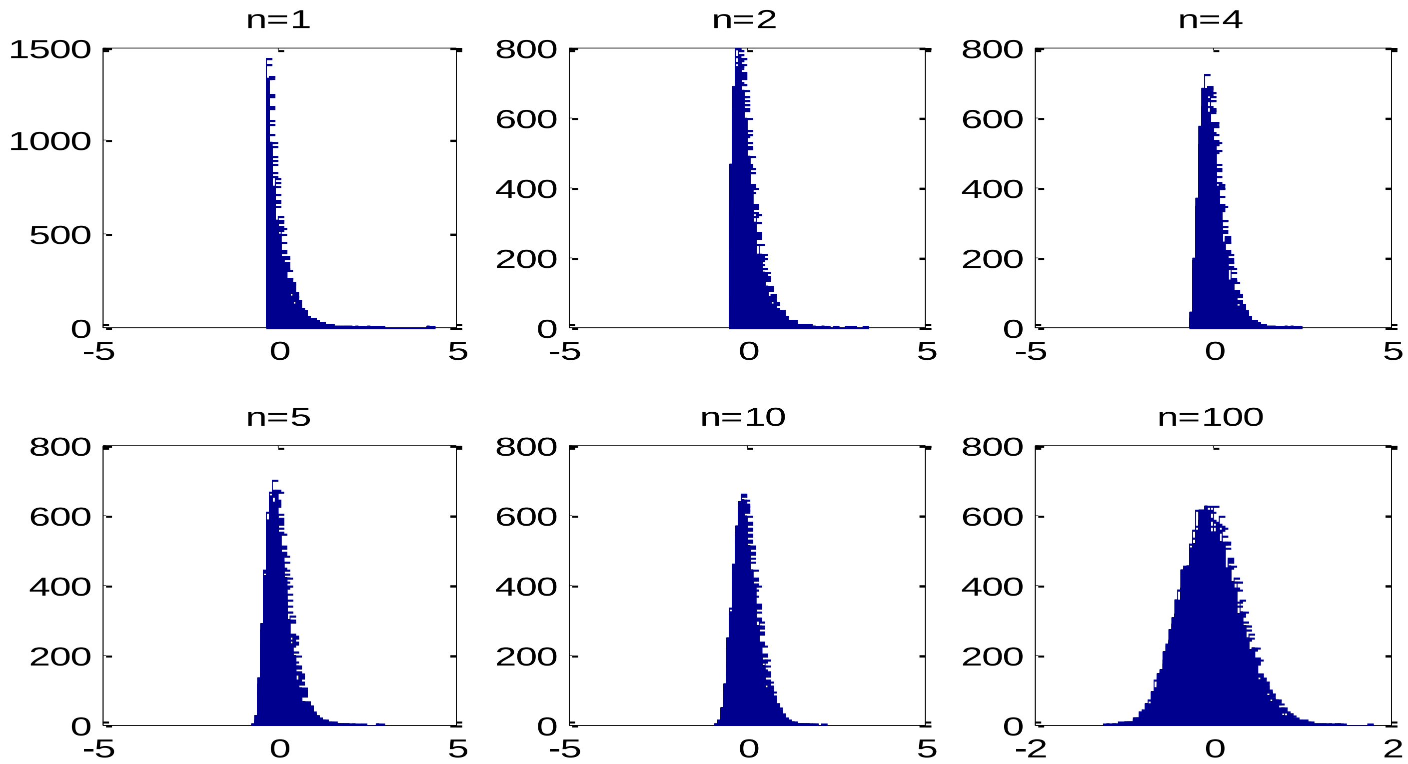
Continuando o jogo

- Recorrendo a simulação em Matlab…



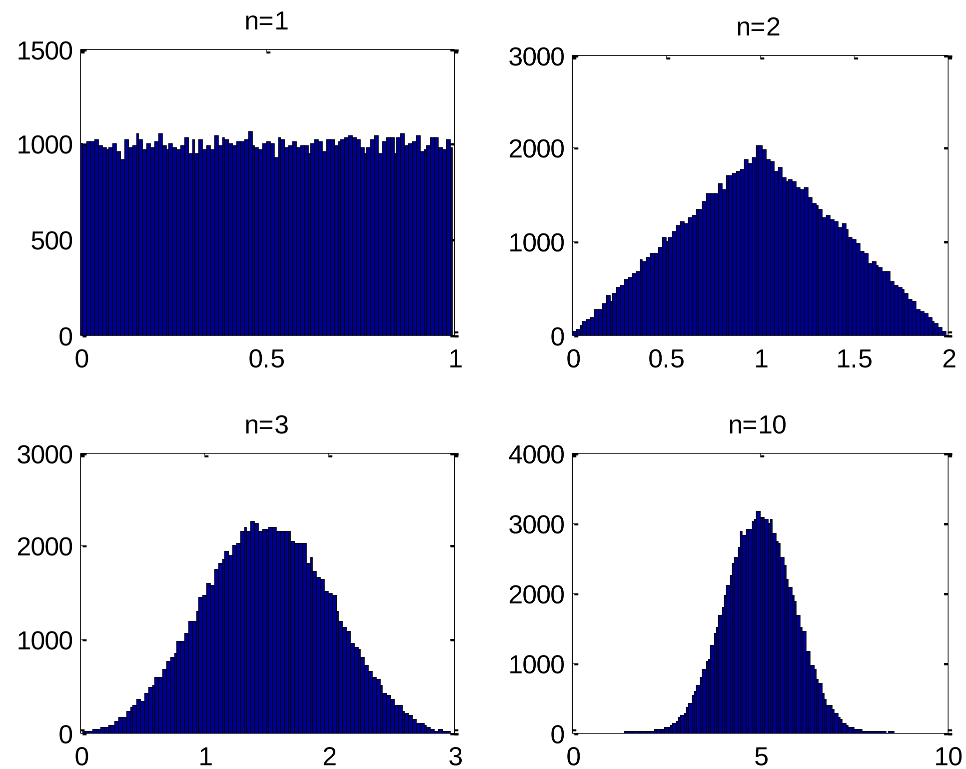
E se tivermos outras distribuições iniciais ?

- Exponencial: y=-log(rand(1,len))./lambda



Outro exemplo

- **Usando geração** de números aleatórios:
- Geradas 10 sequências de números aleatórios com distribuição uniforme no intervalo [0,1] e somadas …



Teorema do Limite Central

- Nos exemplos, para valores grandes de *n*, temos sempre uma distribuição com a forma da Gaussiana
- De facto demonstra-se que a soma de variáveis i.i.d. tende para uma distribuição normal quando o número de variáveis é grande
  - Teorema do Limite Central

- A média é, como já vimos, igual à das variáveis originais

De uma forma mais formal

- Sendo:
  - $X_1, X_2, \ldots$ variáveis aleatórias I.I.D.
  - X_i com distribuição *F*, $E[X_i] = \mu$ e $Var(X_i) = \sigma^2$
    - μ e σ^2 finitos
  - S_n a soma das *n* primeiras variáveis
  - $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ v.a. de média nula e variância unitária
- O Teorema do Limite Central afirma:

$$\lim_{n\to\infty} P(Z_n \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} \, dx$$

- Isto é, a função de distribuição de Z_n **tende para a distribuição de uma variável Normal normalizada $N(0,1)$**

Aplicando à média (M_n)

- Fazendo $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

- Pelo TLC temos

$$M_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{quando } n \to \infty$$

A distribuição da **média de *n* variáveis i.i.d. tende para a distribuição normal com parâmetros $\boldsymbol{\mu}$ e $\boldsymbol{\frac{\sigma^2}{n}}$**

Teorema do Limite Central

- O Teorema do Limite Central é a razão da importância da distribuição Normal/Gaussiana

  – É um **resultado extremamente importante e abre caminho a muitas aplicações**

- "Formulação qualitativa":

  Coisas que são o resultado da soma de muitos pequenos efeitos tendem a ser Gaussianas

Demos online

- Wolfram Demonstrations Project : **The Central Limit Theorem**

The central limit theorem states that the sampling **distribution of the sample mean approaches a normal distribution as the size of the sample grows**.

This means that the histogram of the means of many samples should approach a bell-shaped curve.

Each sample consists of 200 pseudorandom numbers between 0 and 100, inclusive.

- http://demonstrations.wolfram.com/TheCentralLimitTheorem/

Demos

- **Central Limit Theorem Applied to Samples of Different Sizes and Ranges**

- http://demonstrations.wolfram.com/CentralLimitTheoremAppliedToSamplesOfDifferentSizesAndRanges/

- This Demonstration shows the applicability of the central limit theorem (CLT) to the means of samples of random integer or real numbers having random ranges.
- It allows the user to generate such datasets and plot the histogram of their means.
- Superimposed on the histogram is the normal (Gaussian) distribution function that gives the theoretical distribution of these sample means.
- Also shown for comparison are the numeric values of the mean and standard deviation, both of the theoretical distribution and of the generated data.

Exemplo de aplicação do TLC

- Suponha que as despesas feitas por cada cliente de um restaurante são variáveis aleatórias I.I.D. com média 6.5 Euros e desvio padrão 2.5 Euros.

- Estime a probabilidade de os primeiros 100 clientes gastarem um total superior a 600 Euros

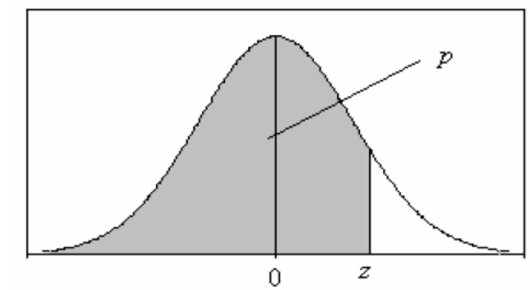
Resolução

- Consideremos $S_{100} = X_1 + X_2 + \cdots + X_{100}$
- Como $E[S_{100}] = 100\mu = 650$
- e $n\sigma^2 = 625$

- Teremos $Z_{100} = \frac{S_{100} - 650}{25}$
- Como pelo TLC Z_{100} segue um lei $N(0,1)$:
- $P(S_{100} > 600) = P\left(Z_{100} > \frac{600-650}{25}\right)$
- $= P\left(Z_{100} > \frac{600-650}{25}\right) = \boldsymbol{P(Z_{100} > -2)}$

Calc. probabilidades na N(0,1)

- $P(Z_{100} > -2)$?
- Como se obtém ?
- Existem valores tabelados de



$$P(Z \leq z) = \Phi(z)$$

  – Exemplo:

1,8	0,96407	0,96485
1,9	0,97128	0,97193
2,0	0,97725	0,97778
2,1	0,98214	0,98257
2,2	0,98610	0,98645

- $P(Z_{100} > -2) = 1 - \Phi(-2)$
- $= 1 - \left(1 - \Phi(2)\right) = \Phi(2) =$

=0,97725

Em Matlab

- Obter $\Phi(2)$

```
z=2
m=0
sigma=1
p = cdf('Normal',z,m,sigma)
    >>  0.9772
```

OU

```
p= normcdf(z,m,sigma)
```

Nota: usa Statistics Toolbox

Em Matlab

- Com ferramentas como o Matlab não é necessário efectuar a normalização
- Aplicando directamente a S_{100}:

```
s=600                  %  pq queremos P(S100 > s=600)
m=650                  % média de S100
sigma=25               % desvio padrão de S100

p = 1- cdf('Normal',600,m,sigma)
  >>> 0.9772
```

OU

```
p = 1- normcdf(600,m,sigma)
```