

Métodos Probabilísticos para Engenharia Informática

2016-2017

Aula 2

Probabilidades

Probabilidades

Conceitos essenciais

Probabilidade

- Qual a hipótese de chover amanhã?
- Qual a possibilidade de eu chegar a horas à aula ?
- Qual a probabilidade de eu ganhar o Euromilhões (ou um de vocês) ?
- São questões que colocamos frequentemente...
- e estão relacionadas com o **incerto / não determinístico**

Aleatório

- Em termos qualitativos, “qualquer coisa” que não seja predizível com certeza absoluta
- Acontecimento (evento) **cujo resultado não possa ser determinado com certeza absoluta.**
 - Caso contrário é determinístico
- adj. Que repousa sobre um acontecimento **incerto**, fortuito: contrato aleatório.
Diz-se de uma grandeza que pode tomar certo número de valores, a cada um dos quais está ligada uma probabilidade.
 - De: [dicionário online de português](#)
- <http://www.priberam.pt/dlpo/aleat%C3%B3rio>

Então qual o interesse ?

- Qual o interesse em estudar algo que não se pode prever ?
- Na maioria das aplicações **existe algum tipo de regularidade** que se manifesta se o número de observações / experiências for elevado

Problema Exemplo 1

- Qual a probabilidade de acertar num PIN/**password** de 4 dígitos escolhendo um PIN completamente ao acaso?

E de 20 dígitos ?

Problema Exemplo 2

- Consideremos uma família com 2 filhos
- Supondo que a probabilidade de nascer um rapaz é a mesma de nascer uma rapariga, qual a probabilidade de uma destas famílias com 2 filhos ter pelo menos 1 rapaz ?
 - De “O Acaso”, página 23

Problema Exemplo 3

- Em qual destas apostas se tem mais possibilidades de ganhar:
 - “pelo menos um seis em 4 lançamentos de 1 dado”
 - “pelo menos um DUPLO 6 em 24 lançamentos de 2 dados”
- Problema Clássico
- Problema do Cavaleiro de Méré
 - Antoine Gombaud, denominado Chevalier de Méré (séc. XVII), foi um nobre e jogador francês que se dedicou ao cálculo matemático de jogos de azar

Problema Exemplo 4

- Qual a probabilidade de 80% ou mais dos alunos de MPEI deste ano letivo que não reprovem por por faltas ter sucesso na UC?
- **O que precisamos saber**
 - Quantos alunos considerar?
 - Qual a probabilidade de cada um passar ?
 - Talvez simplificar ?
 - O que é isso de probabilidade ?

Probabilidade

- Noção de probabilidade

“Medida do grau de certeza associado a um resultado proveniente de um fenómeno de acaso”

- Palavra usada pela primeira vez por Bernoulli (1654-1705)

Recordar ...

- Experiência aleatória
 - Procedimento que deve produzir um resultado
 - Mas mesmo que seja repetido nas mesmas condições não garante que o resultado seja idêntico

Exemplos de experiências aleatórias

- Escolher uma letra do alfabeto
- Escolher um aluno para dar exemplos de experiências aleatórias
- Lançar uma moeda 5 vezes e registrar o número de caras
- Escolher um número real entre 0 e 1
- Medir e registrar o intervalo de tempo entre 2 mensagens que chegam a um servidor de email

Espaço de amostragem

- Conjunto (S) de **todos os resultados possíveis** de uma experiência aleatória
 - Em geral representado por S (de Sample Space)
- Resultados têm de ser **mutuamente exclusivos** e **não divisíveis**
- S é **discreto** se for contável
 - i.e. se contiver um número finito de elementos ou se contiver um número infinito em que se pode estabelecer uma correspondência biunívoca com o conjunto dos inteiros
- S é **contínuo** se não for contável
- Elementos de S são designados por resultados

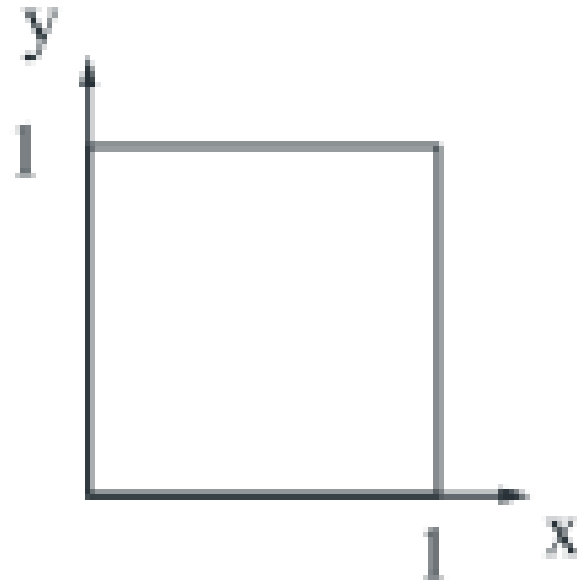
TPC

- Qual o espaço de amostragem de cada um dos problemas exemplo ?

Exemplo de espaço de amostragem contínuo

- Experiência aleatória: instante de chegada, em horas, de 2 alunos (x e y) a uma aula de 1 hora

- $S = \{(x, y) : 0 \leq x, y \leq 1\}$



Acontecimentos / eventos

- Os resultados das experiências não constituem necessariamente os únicos itens de interesse nos experimentos
 - Exemplo:
 - No caso da contagem de mensagens de email podemos estar interessados no facto de o número total exceder um determinado limiar ($n^o > L$)
 - Os itens de interesse são representados por subconjuntos de S
- **Acontecimento (evento)** A é um subconjunto de S
 - S é obviamente um subconjunto de S próprio e constitui o evento certo
 - O conjunto vazio, ϕ , também é subconjunto, o evento impossível
- A probabilidade é atribuída a eventos

Lei de probabilidade

- Regra que atribui probabilidade aos vários eventos
- Probabilidade: número associado a um evento que indica a “verosimilhança” de esse evento ocorrer quando se efectua o experimento
 - valor entre 0 e 1
 - 1 para acontecimento certo
 - 0 para acontecimento impossível

Cálculo de probabilidades

Como é que se definem/obtêm as probabilidades associadas a eventos ?

- Através de medição
- Através da construção de modelos probabilísticos

Como determinar a probabilidade?

- Probabilidades teóricas
- **Probabilidade empíricas**
- Probabilidades subjectivas
 - Exemplo:
 - Um Médico diz que tem 95 % de certeza de que determinada pessoa tem uma determinada doença
 - Uma casa de apostas estimou em 1/5 a probabilidade de Portugal ser campeão Europeu em 2016
 - E fomos Campeões 😊
 - Não nos interessam nesta UC

Diferentes abordagens

- Teoria clássica (de Laplace)
 - Probabilidades teóricas
- Frequencista
 - Probabilidades empíricas
- Teoria matemática

Noção clássica

Simon de Laplace (1749-1827)

- *“Pour étudier un phénomène, il faut réduire tous les événements du même type à un certain nombre de cas également possibles, et alors la probabilité d’un événement donné est une fraction, dont le numérateur représente le nombre de cas favorables à l’événement e dont le dénominateur représente par contre le nombre des cas possibles”*
 - pg 17 livro “O Acaso”
- Primeiro reduzir o fenómeno a um conjunto de resultados elementares, **“casos”, igualmente prováveis**

$$P(\text{evento}) = \frac{\text{número de casos favoráveis}}{\text{número de casos possíveis}}$$

Exemplo

- Lançamento de 1 DADO
 - Honesto
 - \Rightarrow qualquer face igualmente provável
- Probabilidade de obter certa face, ex: a 5 ?
- 6 resultados ou eventos elementares
 - Representáveis pelo conjunto $\{1,2,3,4,5,6\}$
- Ao evento “saída da face 5” apenas corresponde um caso favorável
 - $\rightarrow P(\text{“face 5”})=1/6$

Variante do problema

- E se 2 faces tivessem o 5 marcado ?
- Espaço de amostragem ?
 - $S=\{1,2,3,4,5\}$? \Rightarrow casos possíveis =5
 - $S=\{1,2,3,4,\mathbf{5},\mathbf{5}\}$
- $P(\text{"sair 5"})=2/6$

Exemplo de aplicação (em Java)

- Probabilidade de termos “0123” numa sequência de 4 dígitos
- Como fazer ? Sugestões ?
- Relembro que precisamos contar todos os casos possíveis

Resolvendo...

- Ideia 1 : 4 ciclos for ...
 - Limitação do código... se quisermos 5 etc
- Ideia 2 ...
 - Usar recursividade ...

Possível solução

```
void comb(String example, String alphabet, int len, List<String> list)
{

    if (len == 0) { // new combination available, store it ...
        list.add(example); // store it in a list
        return;
    }
    else {
        for (int i=0;i<alphabet.length(); i++){ // all alphabet

            // recursive cal
            comb(example+alphabet.charAt(i), alphabet, len-1, list);

        }
    }
}
```

Exemplo de utilização

```
public static void main(String[] args) {  
  
    String alphabet = "0123456789";  
  
    final int MAX=7;  
  
    for (int n=1;n<=MAX;n++) {  
        int possible=comb(alphabet,n);  
  
        System.out.printf("t Prob=%.8f\n",1/(double)possible);  
    }  
}
```

- [PINs.java]

Regras básicas (OU)

- $P(\text{"sair face maior que 4"}) ?$
 $= P(\text{"sair face 5 ou face 6"}) = P(\{5,6\}) = 2/6$
 $= P(\{5\}) + P(\{6\})$
- $P(\text{"face par"}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = 1/2$
- $P(\text{"qualquer face"}) = 6 \times 1/6 = 1$

... $P(A \cup B) = P(A) + P(B)$

Sempre ???

Regras básicas

- $P(\text{"face menor ou igual a 4"})$
 $= 1 - P(\text{"face maior que 4"})$
 $= 1 - 2/6 = 4/6$

Regra do complemento

$$P(\bar{A}) = 1 - P(A)$$

Regras básicas (E)

- $P(\text{"face par E face menor ou igual a 4"}) =$
 $= P(\text{"face par"}) \times P(\text{"face menor ou igual a 4"})$
 $= \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$

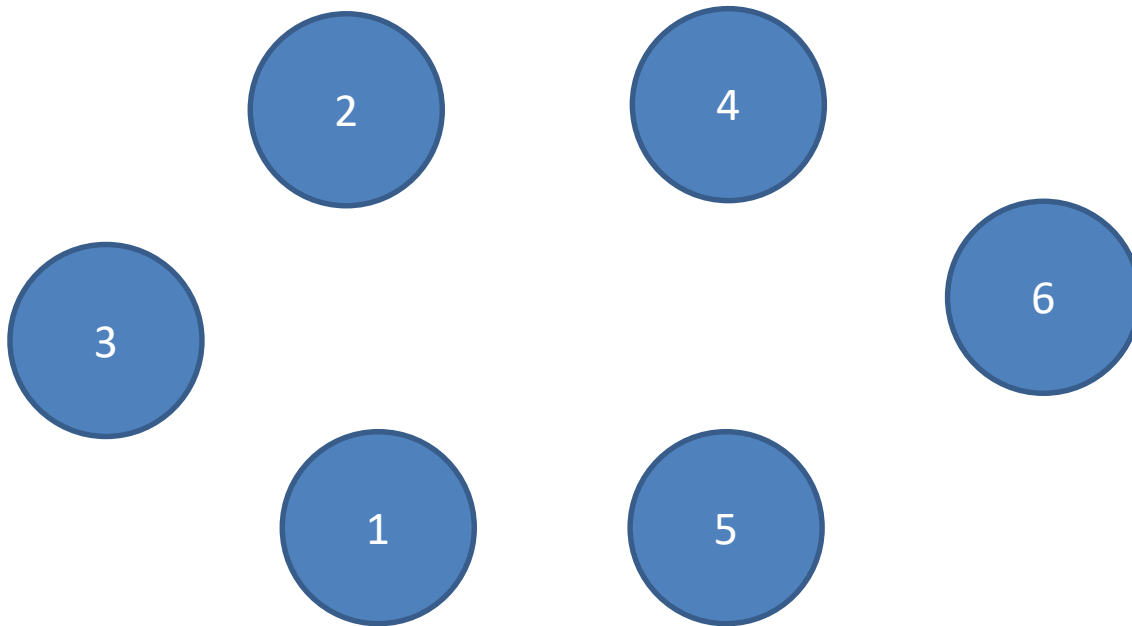
De facto existem 2 possibilidades em 6 , {2,4}

Aplicação das regras (OU novamente)

- $P(\text{"face par OU face menor ou igual a 4"}) = ?$
- Se fizermos $P(\text{"face par"}) + P(\text{"face menos ou igual a 4"})$ dá $7/6 > 1$!!
- Qual o erro ?

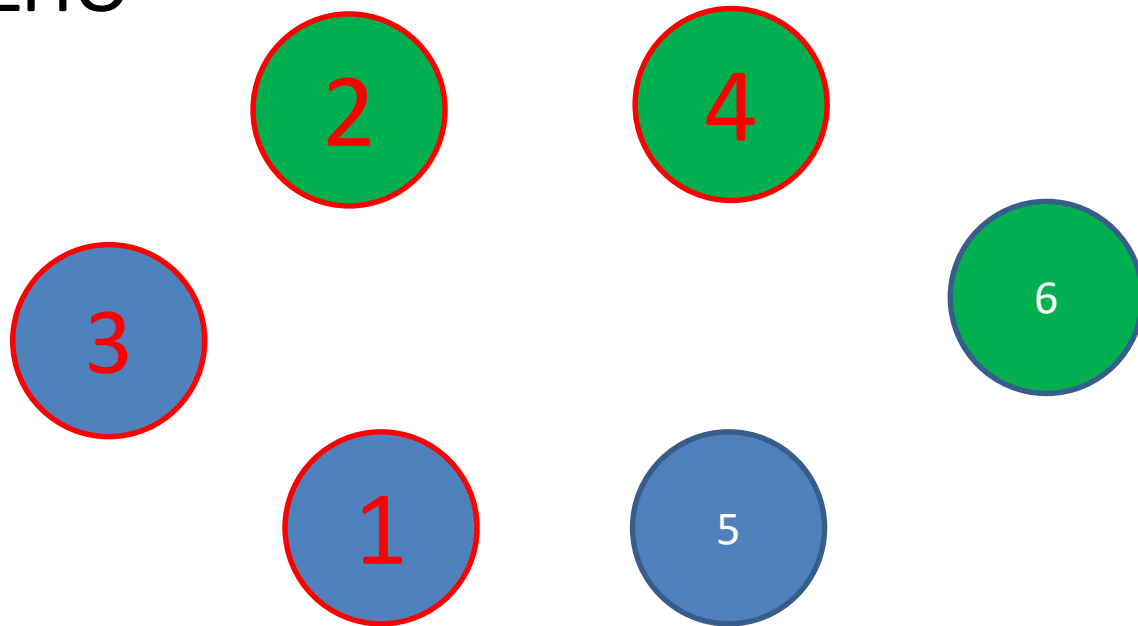
Analisemos

S



Eventos

- A=“face par” VERDE
- B=“face menor ou igual a 4” limite e texto a VERMELHO



...

Temos 3 com fundo verde $\Rightarrow P(A) = \frac{1}{2}$

Temos 4 com vermelho $\Rightarrow P(B) = \frac{2}{3}$

... mas temos 2 casos com verde e vermelho

– No mínimo perigoso 😊

- Estávamos a contar 2 vezes a intersecção

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 $= \frac{1}{2} + \frac{2}{3} - \frac{2}{6} = \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$

Testar as regras

(num problema exemplo)

- Considere uma família com 2 filhos e que a probabilidade de nascer rapaz é igual à de nascer uma rapariga.
- Designando o nascimento de um filho por M e uma filha por F, qual a probabilidade de MF ?
- Probabilidade de pelo menos 1 rapaz numa família com 2 filhos ?

Resolução

- Pelo menos 1 rapaz \Rightarrow MF ou FM ou MM
- MF é a intersecção (“e”) de M no primeiro e F no segundo $= \frac{1}{2} * \frac{1}{2}$
- Similar para MM e FM
- $P(MF) + P(MM) + P(FM) = \frac{3}{4}$
 - Devido à união (“ou”)

Problema do Cavaleiro de Méré

- Aplicação de teoria Clássica
- Criar lista de todas as possibilidades (S)
 - 4 lançamentos
 - 1111
 - 1112
 - 1113 ...
 - 24 lançamentos ...
- Contar casos favoráveis
- Calcular probabilidade
- Sugestão de TPC (Java?)

Problema do Cavaleiro de Méré

- $P(\text{"sair pelo menos um seis em 4 lançamentos de 1 dado"})$ vs $P(\text{"sair DUPLO 6 em 24 lançamentos de 2 dados"})$
- Melhor usar a regra do complemento..
- $P(\text{"nenhum 6 em 4 lançamentos"}) =$
- $P(\text{"não 6 na primeira E não 6 na segunda E ..."})$
 $= P(\text{"não 6 na primeira"}) \times P(\text{"não 6 na segunda"})$
- ...
- $= 5/6 \times 5/6 \dots = (5/6)^4$

- P(“sair pelo menos um seis em 4 lançamentos de 1 dado”)

$$= 1 - \left(\frac{5}{6}\right)^4$$

$$= 0,51775$$

- P (“sair pelo menos um DUPLO 6 em 24 lançamentos de 2 dados”) =

$$= 1 - \left(\frac{35}{36}\right)^{24}$$

$$= 0,49141$$

Não esquecer

- Estas regras e definição clássica ASSUMEM dados honestos, moedas honestas, igual probabilidade de nascer rapaz e rapariga, **equiprobabilidade para os eventos elementares**
- Uma questão que surge naturalmente é se na prática tais valores são ou não razoáveis ?

Noção frequencista

- Noção introduzida por De Moivre (1718)
- Repete-se a experiência um certo número de vezes (N).
- Seja k o número de vezes que ocorre o acontecimento que nos interessa (ex: “sair face 5 num dado”)
- Determina-se $f=k/N$, ou seja a frequência relativa de ocorrência
- Usa-se esta frequência como uma medida empírica de probabilidade

Frequência relativa

- Definição:
 - Se uma experiência for repetida N vezes nas mesmas condições a frequência relativa do evento A é

$$f(A) = \frac{\# \text{ ocorrências do evento } A}{N}$$

- Se a frequência relativa convergir quando N aumenta, então o limite da frequência relativa é a probabilidade de A

$$p(A) = \lim_{N \rightarrow \infty} \frac{\# \text{ ocorrências de } A}{N}$$

Frequência relativa (cont.)

- $0 \leq f(A) \leq 1$
- Numa experiência com k resultados possíveis em N experiências:
 - $S = \{A_1, A_2, A_3, \dots, A_k\}$
 - o resultado A_i ocorre N_i vezes
 - Cada um dos resultados possíveis terá uma frequência relativa de $f(A_i) = N_i/N$
 - $\sum_{i=1}^k f(A_i) = 1$

Exemplos/Demos (em Java)

- Probabilidade de Cara no lançamento de uma moeda
 - CoinWithGraphics
- Probabilidade de uma das faces no lançamento de um Dado honesto
 - Dice ...
- Probabilidades das várias somas possíveis no lançamento de n Dados
 - Ndices ..

Exemplos com cartas (Java)

- $P(\text{"Rei de Ouros (numa carta)"}) ?$
 - $P(\text{"5 Ouros em 5 cartas"}) ?$
 - Etc
-
- Cards.java

Código exemplo

```
static double sim(int nCards, String pattern, int nOcurrances, int nreps, boolean debug) {

    // the deck
    List<String> deck = new ArrayList<>();
    for (String c: card)
        for (String t: type)
            deck.add(c+" de "+t);

    int count = 0; // count of number of favorable outcomes

    // REPEAT EXPERIMENT OF OBTAINING A HAND
    for (int rep=1; rep<=nreps; rep++) {

        // deal hand

        List<String> hand= sample(deck, nCards);

        if (debug) System.out.println("HAND="+hand);

        // CHECK IF FAVORABLE OUTCOME
        int numFav=0;
        for (String h: hand)
            if (h.contains(pattern))
                numFav++;

        // if pattern occurs the number of times we want
        if (numFav==nOcurrances)
            count = count + 1;
    }

    return (count/(double)nreps);
}
```


Exemplo em Matlab

- Probabilidade de **sair 2 caras em 3 lançamentos**
- Como se simula a experiência de lançar uma moeda ?
- Como se simula 3 lançamentos ?
- Como se repete “muitas” vezes ?
- Como contar as ocorrências do evento ?

Simular lançamentos ...

% simular 1 lançamento

l= rand() >0.5 % assumiremos que 1 = “cara”

% simular os 3 lançamentos

l3= rand (3, 1) > 0.5 % ou l3 = rand(1,3) > 0.5

% repetir N vezes

N= 1e6 % mas comecem com valor pequeno

lancamentos= rand(3,N)>0.5; % importante o “;”

ocorrências ... freq. relativa

% contar num ocorrências de “2 caras”

% contar num caras (1s) em cada experiência

% (que se encontra numa coluna da matriz lançamentos)

numCarasNaExperiencia= sum (lançamentos);

% contar vezes em que esse número de caras é 2

numOcorrencias = sum (numCarasNaExperiencia ==2)

% calcular freq relativa

$f = \text{numOcorrencias} / N$

% usar como estimativa da probabilidade

$p_A = f$

Variação com N

% variação da frequência relativa em função de N

N= 1e5

lancamentos = rand(3,N) > 0.5;

sucessos= sum(lancamentos)==2; % 1 = sucesso

fabsol = **cumsum**(sucessos);

frel = fabsol ./ (1:N);

plot(1:N, frel);

Problema do Cavaleiro de Méré

- $P(\text{"sair pelo menos um seis em 4 lançamentos de 1 dado"})$ vs $P(\text{"sair pelo menos um DUPLO 6 em 24 lançamentos de 2 dados"})$
- Simulação **em Matlab** [cavaleiro.m]
- TPC: Simulação em Java

Parte do código – 2 dados

```
L=24
```

```
dado1=floor(rand(L,N) * 6)+1 ;
```

```
dado2=floor(rand(L,N) * 6)+1 ;
```

```
seis= (dado1==6 & dado2==6);
```

```
% contar o num de seis (somando ao longo das colunas
```

```
contagem=sum(seis);
```

```
% determinar as experiencias com pelo menos 1 seis
```

```
expsFavoraveis=contagem>=1;
```

```
% contar casos favoráveis (equivale a somar)
```

```
fav=sum(expsFavoraveis)
```

```
prob2=fav/N
```

Problema 4 - aprovações a MPEI

- **Simulação** em Matlab [simulMPEI.m]

```
p=aprov/total ;      % prob aluno passar em 2015 - 2016
n=50*2;              % alunos de MPEI
N=1e6;               % experiências
k=fix(0.8 * n);       % os 80 %
```

```
aprovados = rand(n,N) < p;  %% 1 indica aprovado
```

```
numOcorrencias =0;
for k1=k:n
    sucessos= sum(aprovados)==k1 ;
    fprintf('%d aprovados -> %8d , p=%.5f\n',k1,sum(sucessos),sum(sucessos)/N);
    numOcorrencias = numOcorrencias +sum(sucessos);
End
```

```
probSimulacao= numOcorrencias /N;
fprintf('prob de %d ou mais em %d passarem é de %.4f\n',k,n,probSimulacao);
```

Exemplos de resultados

Prob aprovação = 0.8962 (baseado edição anterior)

80 aprovados -> 1721 , p=0.00172
81 aprovados -> 3660 , p=0.00366
82 aprovados -> 7448 , p=0.00745
83 aprovados -> 13980 , p=0.01398
84 aprovados -> 24506 , p=0.02451
85 aprovados -> 39659 , p=0.03966
86 aprovados -> 59943 , p=0.05994
87 aprovados -> 83473 , p=0.08347
88 aprovados -> 106277 , p=0.10628
89 aprovados -> 124191 , p=0.12419
90 aprovados -> 130987 , p=0.13099
91 aprovados -> 124396 , p=0.12440
92 aprovados -> 104728 , p=0.10473
93 aprovados -> 78250 , p=0.07825
94 aprovados -> 50022 , p=0.05002
95 aprovados -> 27396 , p=0.02740
96 aprovados -> 12368 , p=0.01237
97 aprovados -> 4330 , p=0.00433
98 aprovados -> 1149 , p=0.00115
99 aprovados -> 202 , p=0.00020
100 aprovados -> 13 , p=0.00001

prob de 80 ou mais em 100 passarem é de 0.9987 MPEI MIECT/LEI 2016-2017

TOTAL INCLUINDO REPROVADOS POR FALTAS

Prob aprovação = 0.7600 (baseado edição anterior)

80 aprovados -> 63021 , p=0.06302
81 aprovados -> 49257 , p=0.04926
82 aprovados -> 35789 , p=0.03579
83 aprovados -> 24857 , p=0.02486
84 aprovados -> 15942 , p=0.01594
85 aprovados -> 9491 , p=0.00949
86 aprovados -> 5315 , p=0.00532
87 aprovados -> 2648 , p=0.00265
88 aprovados -> 1223 , p=0.00122
89 aprovados -> 540 , p=0.00054
90 aprovados -> 183 , p=0.00018
91 aprovados -> 70 , p=0.00007
92 aprovados -> 31 , p=0.00003
93 aprovados -> 7 , p=0.00001
94 aprovados -> 1 , p=0.00000
95 aprovados -> 1 , p=0.00000
96 aprovados -> 0 , p=0.00000
97 aprovados -> 0 , p=0.00000
98 aprovados -> 0 , p=0.00000
99 aprovados -> 0 , p=0.00000
100 aprovados -> 0 , p=0.00000

prob de 80 ou mais em 100 passarem é de 0.2084

Simple mas não perfeita

- Conceptualmente é extremamente simples e pode ser aplicada praticamente a todas as experiências
- Tem, no entanto, algumas desvantagens:
 - Em muitos casos requer considerável dispêndio de tempo
 - As experiências devem poder ser repetidas em condições idênticas
 - Quando o espaço amostral é infinito surgem questões de fiabilidade uma vez que só podemos efectuar um número finito de repetições da experiência
 - A própria obtenção dos valores coloca algumas questões:
 - Quantos ensaios se tem de efectuar para termos medidas fiáveis ?
 - Como se lida com medições sujeitas a erro ?