Departamento de Ciências e Engenharia de Biossistemas
Agrupamento de Matemática

**Matemática II**

# Alguns Modelos Probabilísticos

2013/2014

(MJ Martins, F. Valente e M. Mesquita)

## Modelos probabilísticos

- As v.a.s associadas a grande parte das experiências podem ser representadas por famílias de distribuições estatísticas – modelos probabilísticos – cujas propriedades são função de poucos parâmetros.
- Os parâmetros determinam a localização, a variabilidade e a forma das distribuições.
- Pela sua importância, ir-se-ão considerar as
  - Distribuições discretas:
    - uniforme discreta, binomial e Poisson;
  - Distribuições contínuas:
    - uniforme, normal.



## Modelos probabilísticos no R

- O *software* R disponibiliza "tabelas estatísticas" para um conjunto de distribuições.
- Para cada distribuição estatística é possível calcular
  1. função massa de probabilidade ou função densidade
  2. função distribuição cumulativa
  3. quantis
  4. números pseudo-aleatórios

Nota: Os números pseudo-aleatórios são muitas vezes utilizados para obter dados simulados que seguem uma dada distribuição.



## Modelos probabilísticos no R

| Distribution | R name | additional arguments |
|---|---|---|
| beta | beta | shape1, shape2, ncp |
| **binomial** | **binom** | **size, prob** |
| Cauchy | cauchy | location, scale |
| chi-squared | chisq | df, ncp |
| exponential | exp | rate |
| F | f | df1, df2, ncp |
| gamma | gamma | shape, scale |
| geometric | geom | prob |
| hypergeometric | hyper | m, n, k |
| log-normal | lnorm | meanlog, sdlog |
| logistic | logis | location, scale |
| neg. binomial | nbinom | size, prob |
| **normal** | **norm** | **mean, sd** |
| **Poisson** | **pois** | **lambda** |
| Student's t | t | df, ncp |
| **uniform** | **unif** | **min, max** |
| Weibull | weibull | shape, scale |
| Wilcoxon | wilcox | m, n |



## Modelos probabilísticos no R

- Ao nome R da distribuição acrescentam-se os prefixos:
  - `d` para função massa de probabilidade ou função densidade
  - `p` para função distribuição cumulativa
  - `q` para quantis
  - `r` para números pseudo-aleatórios

As funções `pxxx` e `qxxx` têm o argumentos lógico: `lower.tail` – se for `TRUE` (por omissão), as probabilidades são P[$X \leq x$], caso contrário são P[$X > x$].



## Distribuição uniforme discreta

- É a distribuição discreta mais simples. Caracteriza-se por todos os valores possíveis terem igual probabilidade.
- É usada como primeiro modelo para uma quantidade que toma $n$ valores distintos, $x_1, \ldots, x_n$, em que $x_1 \leqslant \ldots \leqslant x_n$

$$X \cap UD(x_1, \ldots, x_n)$$

$$P[X = x_i] = \frac{1}{n}, \quad i = 1, \ldots, n$$

## Distribuição uniforme discreta

- Valor médio: $E[X]=\mu=\frac{1}{n}\sum_{i=1}^{n} x_i$
- Variância: $Var(X)=\frac{1}{n}\sum_{i=1}^{n} x_i^2-\mu^2$
- Exemplo: Considere a experiência aleatória que consiste em lançar um dado de seis faces. Seja $X$ a v.a. que representa o nº pintas da face voltada para cima. $X \cap UD(1:6)$
  - pode tomar os valores $x=1,2,3,4,5,6$
  - com probabilidade $P[X=x]=\frac{1}{6}$



## Distribuição binomial

- É a distribuição discreta mais importante, muito aplicada em amostragem.
- **Sucessão de provas de Bernoulli independentes** é uma sequência de experiências aleatórias em que:
  - cada prova tem dois resultados possíveis, sucesso ou insucesso,
  - a probabilidade de sucesso $p$ mantém-se de prova para prova,
  - o resultado de cada prova é independente dos resultados das provas anteriores.
- A v.a. $X$ que representa o número de sucessos numa sucessão de $n$ provas de Bernoulli independentes, tem distribuição binomial com parâmetros $n$ e $p$.



## Distribuição binomial

- $X \cap B(n, p)$ $n$ inteiro positivo e $0 < p < 1$
  - pode tomar os valores $x=0,1,2,\cdots,n$
  - com probabilidade $P[X=x]=\binom{n}{x} p^x (1-p)^{n-x}$ em que $\binom{n}{x}=\frac{n!}{x!(n-x)!}$
- Valor médio: $E[X]=np$
- Variância: $Var(X)=np(1-p)$
- Exemplo: Seja $X \cap B(4,\ 0.3)$
  - $P[X=2]=0.2646$ `dbinom(2,4,0.3)`
  - $P[X\leq 2]=0.9163$ `pbinom(2,4,0.3)`

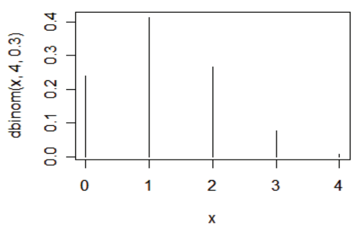


## Distribuição binomial

- Alguns gráficos da função massa de probabilidade:

```
>x<-0:4
>plot(x,dbinom(x,4,0.3),type="h")
```

$n$=4, $p$=0.3

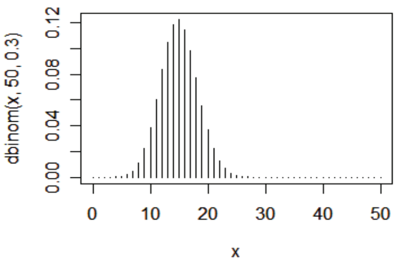





## Distribuição binomial

```
>x<-0:50
>plot(x,dbinom(x,50,0.3),type="h")
```

$n$=50, $p$=0.3






## Distribuição binomial (leitura de tabelas)

- Para calcular probabilidades de uma distribuição binomial pode recorrer-se também a tabelas.
- As tabelas disponíveis apresentam valores da <u>função distribuição cumulativa</u> para alguns valores de $n\ (2,\ 3,\dots,\ 20)$ e $p(0.05,\ 0.1,\dots,\ 0.5)$
- Para calcular probabilidades de uma distribuição com parâmetro $p>0.5$ utiliza-se a relação

$$X \cap B(n,\ p) \Leftrightarrow (n-X) \cap B(n,\ 1-p)$$

## Distribuição binomial (leitura de tabelas)

- Exemplo: Seja $X \cap B(4,\ 0.3)$

$P[X \leq 2] = 0.9163$

$P[X=2] = P[X \leq 2] - P[X \leq 1] = 0.9163 - 0.6517 = 0.2646$

| | | p | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| n | x | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.35 | 0.4 | 0.45 | 0.5 |
| 2 | 0 | 0.9025 | 0.8100 | 0.7225 | 0.6400 | 0.5625 | 0.4900 | 0.4225 | 0.3600 | 0.3025 | 0.2500 |
| | 1 | 0.9975 | 0.9900 | 0.9775 | 0.9600 | 0.9375 | 0.9100 | 0.8775 | 0.8400 | 0.7975 | 0.7500 |
| | 2 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 3 | 0 | 0.8574 | 0.7290 | 0.6141 | 0.5120 | 0.4219 | 0.3430 | 0.2746 | 0.2160 | 0.1664 | 0.1250 |
| | 1 | 0.9928 | 0.9720 | 0.9392 | 0.8960 | 0.8438 | 0.7840 | 0.7183 | 0.6480 | 0.5748 | 0.5000 |
| | 2 | 0.9999 | 0.9990 | 0.9966 | 0.9920 | 0.9844 | 0.9730 | 0.9571 | 0.9360 | 0.9089 | 0.8750 |
| | 3 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 4 | 0 | 0.8145 | 0.6561 | 0.5220 | 0.4096 | 0.3164 | 0.2401 | 0.1785 | 0.1296 | 0.0915 | 0.0625 |
| | 1 | 0.9860 | 0.9477 | 0.8905 | 0.8192 | 0.7383 | $\boxed{0.6517}$ | 0.5630 | 0.4752 | 0.3910 | 0.3125 |
| | 2 | 0.9995 | 0.9963 | 0.9880 | 0.9728 | 0.9492 | $\boxed{0.9163}$ | 0.8735 | 0.8208 | 0.7585 | 0.6875 |
| | 3 | 1.0000 | 0.9999 | 0.9995 | 0.9984 | 0.9961 | 0.9919 | 0.9850 | 0.9744 | 0.9590 | 0.9375 |
| | 4 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |



## Distribuição de Poisson

- Esta distribuição modela o número de ocorrências num intervalo de tempo ou numa região espacial.
- **Processo de Poisson** – refere-se ao número de ocorrências num intervalo de tempo ou numa região espacial que verifica as propriedades:
  - não tem memória; os números de ocorrências em intervalos/regiões disjuntos são independentes;
  - a probabilidade de ocorrência num intervalo/região *muito pequeno* é proporcional à amplitude/tamanho do intervalo/região;
  - a probabilidade de o número de ocorrências ser superior a um é nula em intervalos/regiões *muito pequenos*.



## Distribuição de Poisson

- Seja λ > 0 o nº médio de ocorrências num dado intervalo de tempo (ou numa região do espaço) num processo de Poisson. Então, a v.a. que representa o nº de ocorrências nesse intervalo de tempo (ou região), tem distribuição de Poisson com parâmetro λ.
- $X \cap P(\lambda),\ \lambda > 0$
  - pode tomar os valores $x = 0, 1, 2, \cdots,$
  - com probabilidade

$$P[X=x] = \frac{e^{-\lambda}\lambda^{x}}{x!}$$



## Distribuição de Poisson

- Valor médio: $E[X] = \lambda$
- Variância: $Var(X) = \lambda$
- Exemplo: Seja $X \cap P(3)$
  - $P[X=2] = 0.2240418$ `dpois(2,3)`
  - $P[X \leq 2] = 0.4231901$ `ppois(2,3)`
- **Propriedade**: se as v.a.s $X_i \quad i=1,\cdots,k$ são independentes e $X_i \cap P(\lambda_i)$ então

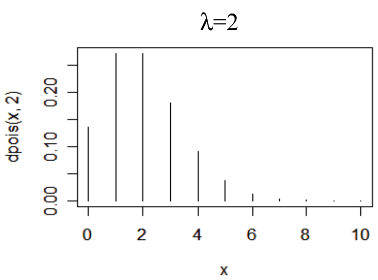
$$\sum_{i=1}^{k} X_i \cap P(\sum_{i=1}^{k} \lambda_i)$$



## Distribuição de Poisson

- Alguns gráficos da função massa de probabilidade:

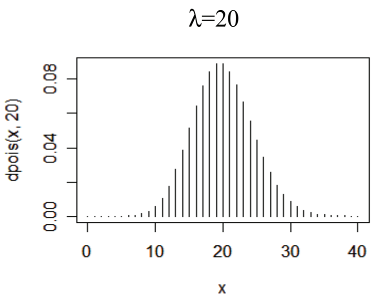
```
>x<-0:10
>plot(x,dpois(x,2),type="h")
```






## Distribuição de Poisson

```
>x<-0:40
>plot(x,dpois(x,20),type="h")
```



## Distribuição Poisson (leitura de tabelas)

- Existem também tabelas que facilitam o cálculo de probabilidades de uma distribuição de Poisson.
- As tabelas disponíveis apresentam valores da <u>função distribuição cumulativa</u> para alguns valores de λ.
- Exemplo: Seja $X \cap P(3)$

$P[X \leq 2] = 0.423$

$P[X = 2] = P[X \leq 2] - P[X \leq 1] = 0.423 - 0.199 = 0.224$

| | $\lambda = E[X]$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x$ | 2.2 | 2.4 | 2.6 | 2.8 | 3.0 | 3.2 | 3.4 | 3.6 | 3.8 | 4.0 |
| 0 | 0.111 | 0.091 | 0.074 | 0.061 | 0.050 | 0.041 | 0.033 | 0.027 | 0.022 | 0.018 |
| 1 | 0.355 | 0.308 | 0.267 | 0.231 | $\boxed{0.199}$ | 0.171 | 0.147 | 0.126 | 0.107 | 0.092 |
| 2 | 0.623 | 0.570 | 0.518 | 0.469 | $\boxed{0.423}$ | 0.380 | 0.340 | 0.303 | 0.269 | 0.238 |
| 3 | 0.819 | 0.779 | 0.736 | 0.692 | 0.647 | 0.603 | 0.558 | 0.515 | 0.473 | 0.433 |
| 4 | 0.928 | 0.904 | 0.877 | 0.848 | 0.815 | 0.781 | 0.744 | 0.706 | 0.668 | 0.629 |
| 5 | 0.975 | 0.964 | 0.951 | 0.935 | 0.916 | 0.895 | 0.871 | 0.844 | 0.816 | 0.785 |

## Distribuição uniforme

- É a distribuição contínua mais simples.
- É usada para representar uma quantidade que varia aleatoriamente no intervalo [*a*,*b*] e cuja probabilidade de tomar valores num subintervalo de [*a*,*b*] é proporcional ao seu comprimento.
- $X \cap U(a,b) \qquad a, b \in \mathbb{R} \quad a < b$
  - pode tomar valores no intervalo [*a*,*b*]
  - função densidade

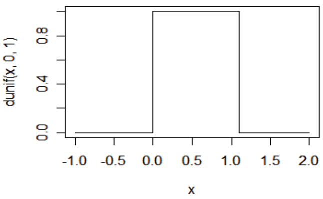
$$f(x) = \frac{1}{b-a}, \quad x \in [a,b]$$



## Distribuição uniforme

- Valor médio: $E[X] = (a+b)/2$
- Variância: $Var(X) = (b-a)^2/12$
- Exemplo: Seja $X \cap U(0,\,1)$
  - função densidade

```
> x<-seq(-1,2,0.1)
> plot(x,dunif(x,0,1),type="s")
```





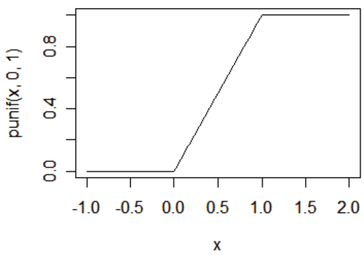

## Distribuição uniforme

- função distribuição cumulativa

```
> plot(x,punif(x,0,1),type="l")
```

ou

```
> curve(punif(x,0,1),from=-1, to=2)
```






## Distribuição normal

- É a distribuição contínua mais importante:
  - é um modelo adequado para representar muitos fenómenos do mundo real, nomeadamente características biométricas, variação de temperaturas, velocidades, volumes, erros, etc;
  - é muito usada em inferência estatística. Tal deve-se ao facto de, mesmo que a distribuição da população não seja normal, a distribuição das médias amostrais, sob certas condições, ser aproximadamente normal (Teorema Limite Central);
  - algumas variáveis aleatórias (como a binomial e a Poisson) podem ser aproximadas por uma v.a. normal.



## Distribuição normal

- $X \cap N(\mu, \sigma) \qquad \mu \in \mathbb{R}, \ \sigma > 0$
  - pode tomar qualquer valor real
  - função densidade

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}$$

- tem forma de sino e um máximo para $x = \mu$
- é simétrica em relação à recta vertical $x = \mu$
- tem pontos de inflexão para $x = \mu \pm \sigma$
- $P[\mu - \sigma < X < \mu + \sigma] = 0.68$
- $P[\mu - 2\sigma < X < \mu + 2\sigma] \simeq 0.95$
- $P[\mu - 3\sigma < X < \mu + 3\sigma] = 0.997$

## Distribuição normal

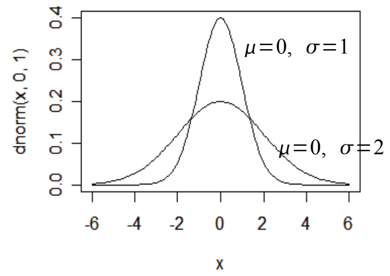
- Valor médio: $E[X]=\mu$
- Variância: $Var(X)=\sigma^2$
- Exemplo: Seja $X \cap N(2,\ 4)$
  - $P[X\leq 2]=0.5$ `pnorm(2,2,4)`
  - $P[X>2.5]=0.4502618$ `1-pnorm(2.5,2,4)`



## Distribuição normal

- Características da função densidade – variação com $\sigma$

```
> curve(dnorm(x,0,1), from=-6, to=6)
> curve(dnorm(x,0,2), from=-6, to=6, add=T)
```

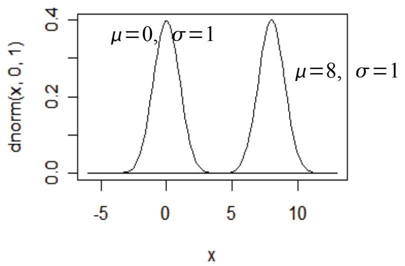





## Distribuição normal

- Características da função densidade – variação com $\mu$

```
> curve(dnorm(x), from=-6, to=13)
> curve(dnorm(x,8,1), from=-6, to=13, add=T)
```






## Distribuição normal reduzida

- $Z\cap N(\mu=0,\sigma=1)$ tem distribuição normal reduzida
- Se $X\cap N(\mu,\sigma)$ então $Z=\dfrac{X-\mu}{\sigma}\cap N(0,1)$
- O cálculo de probabilidades envolvendo $X\cap N(\mu,\sigma)$ pode ser reduzido ao cálculo com a v.a. $Z\cap N(0,1)$, por exemplo:

$$P[\mu-3\sigma< X<\mu+3\sigma]=$$
$$P\left[-3<\frac{X-\mu}{\sigma}<3\right]=\Phi(3)-\Phi(-3)=0.997 \qquad (\Phi(z)=P[Z\leq z])$$



## Distribuição normal reduzida (leitura de tabelas)

- O cálculo de probabilidades da distribuição normal só pode ser efectuado através de métodos numéricos.
- Existem tabelas com os valores da <u>função distribuição cumulativa</u> da normal reduzida, $\Phi(z)$.
- As tabelas disponíveis apresentam probabilidades para valores positivos da variável ($z\geq 0$).

  Para valores negativos, utiliza-se a relação

$$\Phi(-z)=1-\Phi(z),$$

consequência da simetria da distribuição normal.



## Distribuição normal reduzida (leitura de tabelas)

- Exemplo: Seja $Z\cap N(0,\ 1)$

$P[Z\leq 0.23]=\Phi(0.23)=0.59095$

$P[Z\leq -0.23]=\Phi(-0.23)=1-\Phi(0.23)=1-0.59095=0.40905$

| | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.50000 | 0.50399 | 0.50798 | 0.51197 | 0.51595 | 0.51994 | 0.52392 | 0.52790 | 0.53188 | 0.53586 |
| 0.1 | 0.53983 | 0.54380 | 0.54776 | 0.55172 | 0.55567 | 0.55962 | 0.56356 | 0.56749 | 0.57142 | 0.57535 |
| 0.2 | 0.57926 | 0.58317 | 0.58706 | $\boxed{0.59095}$ | 0.59483 | 0.59871 | 0.60257 | 0.60642 | 0.61026 | 0.61409 |
| 0.3 | 0.61791 | 0.62172 | 0.62552 | 0.62930 | 0.63307 | 0.63683 | 0.64058 | 0.64431 | 0.64803 | 0.65173 |
| 0.4 | 0.65542 | 0.65910 | 0.66276 | 0.66640 | 0.67003 | 0.67364 | 0.67724 | 0.68082 | 0.68439 | 0.68793 |

## Propriedades

**1) Transformação afim de uma normal**:

se $X \cap N(\mu, \sigma)$ então $a+bX \cap N(a+b\mu, |b|\sigma)$ $(a, b \in \mathbb{R})$

**2) Estabilidade da soma de normais**:

Se as v.a.s $X_1, X_2, \cdots, X_n$ têm distribuição normal e são **independentes**, então a sua soma ainda tem distribuição normal. Mais concretamente:

$$\left.\begin{array}{l} X_i \cap N(\mu_i, \sigma_i), \;\; i=1,2,\ldots,n \\ X_1, X_2, \ldots, X_n \text{ v.a. independentes} \end{array}\right\} \Rightarrow \sum_{i=1}^{n} X_i \cap N\left(\sum_i \mu_i, \sqrt{\sum_i \sigma_i^2}\right)$$



## Propriedades (cont.)

Se as v.a.s $X_1, X_2, \cdots, X_n$ são independentes e identicamente distribuídas (i.i.d.) com $X_i \cap N(\mu, \sigma), \forall i$, pelas propriedades anteriores tem-se

**3) Estabilidade da soma de normais i.i.d.**:

$$\sum_{i=1}^{n} X_i \cap N\left(n\mu, \sigma\sqrt{n}\right)$$

**4) Média (amostral) de *n* normais i.i.d.**:

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \;\cap\; N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$



## Teorema Limite Central

- A soma de “muitas” v.a.s independentes e identicamente distribuídas, com variância finita, tem distribuição **aproximadamente** normal.

- Se as v.a.s $X_1, X_2, \cdots, X_n$ são i.i.d. com $E[X_i]=\mu$ e $Var[X_i]=\sigma^2$ então

$$\sum_{i=1}^{n} X_i \sim N(n\mu, \sqrt{n}\sigma) \quad \text{para } n \text{ elevado } (n \geq 30)$$

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \quad \text{para } n \text{ elevado } (n \geq 30)$$



## Teorema Limite Central (cont.)

- Consequências do TLC
  - Aproximação da binomial pela normal

    $X \cap B(n, p) \rightarrow X \sim N(np, \sqrt{np(1-p)})$ para *n* elevado $(np>5 \text{ e } n(1-p)>5)$

  - Aproximação da Poisson pela normal

    $X \cap P(\lambda) \xrightarrow[(\lambda>12)]{} X \sim N(\lambda, \sqrt{\lambda}), \;\; \lambda$ elevado $(\lambda > 12)$

- **Correcção de continuidade:** Na aproximação de uma distribuição discreta por uma contínua, há que ter em conta que $P[X_{\text{continua}}=x]=0, \;\forall x$.

$$P[X_{\text{discreta}}=x] \longrightarrow P[x-0.5<X_{\text{continua}}<x+0.5]$$



## Bibliografia


- Neves, M. (2009) Introdução à Estatística e Probabilidade. Apontamentos de apoio às aulas.
  - Teoria da Probabilidade: *www.isa.utl.pt/dm/estat/estat/seb2.pdf*
- Pestana, D.D. e Velosa, S.F. (2002) Introdução à Probabilidade e à Estatística, vol.I, Fundação Caloustre Gulbenkian, Lisboa.
- An Introduction to R, cap. 8 (Probability distributions), cap. 11 (Statistical models in R). Manual do R disponível através do Help ou em http://cran.r-project.org/doc/manuals/R-intro.html