Estatística EC

Lic Ciências do Ambiente

Maria Joana Soares & Ricardo Severino

2012/2103

Introdução

O que é a Estatística?

- A Estatística pode ser definida como o conjunto de instrumentos, procedimentos e técnicas que permitem, de forma adequada, recolher, organizar, explorar, descrever, analisar e interpretar dados.
- A Estatística visa auxiliar os especialistas no domínio em que se enquadram os fenómenos em estudo na compreensão desses mesmos fenómenos. Neste sentido, a Estatística é um método e não uma teoria.
- Por outro lado, a Estatística pode também ser vista como uma teoria matemática, com os seus termos primitivos, axiomas, definições, teoremas e respetivas demonstrações.

Conceitos básicos

- Chama-se população ao conjunto de todos os objetos, indivíduos, etc. que têm em comum uma ou mais características sobre as quais temos interesse em efetuar um estudo estatístico.
- A cada elemento da população dá-se o nome de unidade estatística.
- Cada uma das características em estudo é chamada variável.
- Uma observação é o valor que a variável assume numa determinada unidade estatística.

Exemplos

Como exemplos de populações, têm-se:

1. o conjunto de todos cidadãos portugueses;
2. o conjunto de todos os livros existentes na biblioteca da UM;
3. o conjunto de todos os parafusos produzidos por uma certa fábrica;
4. o conjunto de todos os possíveis lançamentos de um dado.

Em relação com cada uma das populações anteriores, as variáveis a estudar poderiam ser, por exemplo (respetivamente):

1. sexo, cor dos olhos, altura, peso;
2. ano de publicação, tipo de encadernação (*hardback*, *paperback*), número de exemplares existentes ;
3. tipo de parafuso (de cabeça chata, de cabeça redonda), qualidade do parafuso (defeituoso, não defeituoso);
4. número da face obtida (1 a 6), paridade do número da face obtida (par/ímpar).

Amostra/Dado Estatístico

- Dada uma população relativa a um certo estudo estatístico, chama-se amostra a um subconjunto finito dessa população.
- A cada observação da variável (ou variáveis) em estudo respeitante a cada unidade estatística pertencente à amostra, chamamos dado estatístico.

Nota: Por vezes, chama-se população ao conjunto de potenciais valores que a variável em estudo assumiria na população, por exemplo, fala-se na população das alturas ou dos pesos dos alunos inscritos na Universidade do Minho no ano letivo 2010/2011; neste caso, a amostra será um conjunto de dados estatísticos.

Amostragem aleatória simples

É fundamental que a amostra selecionada seja representativa da população. Um dos processos de garantir essa representatividade é fazer amostragem aleatória simples, em que todos os elementos da população têm as mesmas hipóteses de ser incluídos na amostra.
Uma amostra mal recolhida, também chamada viciada, enviesada ou tendenciosa, levará naturalmente a conclusões e previsões distorcidas.

Fases da análise estatística

As análises estatísticas são, essencialmente, compostas por três fases:

1. recolha de dados – amostragem ou planeamento de experiências;
2. tratamento inicial de dados, que inclui a sua ordenação, resumo (cálculo de algumas das suas características), apresentação (em tabelas, gráficos, etc.) e exploração desses dados – estatística descritiva e análise exploratória de dados;
3. indução, a partir do que se verifica numa amostra, para a população de que esta foi extraída – inferência estatística.

Neste curso introdutório, debruçar-nos-emos apenas sobre alguns aspetos das fases 2. e 3. da análise estatística.
Antes, porém, faremos uma revisão dos conceitos básicos de teoria das probabilidades, indispensável à compreensão dos capítulos seguintes.

PARTE I

DISTRIBUIÇÕES DE PROBABILIDADE

Probabilidade

Experiência aleatória

Chama-se experiência aleatória a uma experiência que satisfaça os três requisitos seguintes:

1. pode ser repetida em condições análogas;
2. conhecemos, à partida, todos os seus possíveis resultados;
3. o resultado que obtemos em cada realização da experiência é incerto.

Exemplo

Lançamento de um dado (com as faces numeradas de 1 a 6) e observação de qual a face que fica virada para cima.

Em oposição às experiências aleatórias, temos as experiências determinísticas, que, quando realizadas nas mesmas condições, conduzem aos mesmos resultados.

Espaço amostral/acontecimentos

- Ao conjunto de todos os resultados possíveis de uma experiência aleatória, chamamos espaço amostral ou espaço de resultados associado a essa experiência; este conjunto será denotado por Ω.
- Aos subconjuntos do espaço amostral Ω chamamos acontecimentos. Estes conjuntos são, geralmente, designados por letras maiúsculas do início do alfabeto latino, $A, B, C, \ldots$.
- O conjunto Ω (sendo um subconjunto de si próprio) é um acontecimento, chamado acontecimento universal.
- Ao conjunto vazio $\emptyset$ (que é um subconjunto de Ω) chamamos acontecimento nulo.
- Um subconjunto de Ω formado apenas por um elemento $\{\omega\}$ é chamado acontecimento elementar.
- Ao efetuar uma realização da experiência, dizemos que que um determinado acontecimento A ocorreu ou se realizou, se o resultado da experiência foi um dos elementos de A.

Exemplo

No caso do exemplo anterior do lançamento de um dado, poderemos considerar

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

como o espaço amostral.

Um acontecimento seria, por exemplo, o conjunto

$$A = \{1, 3, 5\}$$

(que poderíamos descrever como "saída de face com um número ímpar").

Este acontecimento ocorreria se, ao lançar o dado, saísse, por exemplo, o número 3.

Definição axiomática de probabilidade (Kolmogorov)

Definição

Chama-se medida de probabilidade (ou, abreviadamente, probabilidade) num espaço amostral Ω a uma função que a cada acontecimento $A \subseteq \Omega$ associa um valor real, $P(A)$ - ao qual chamamos probabilidade de A - e que satisfaz os seguintes axiomas:

P1 $P(A) \geq 0$ (não negatividade)

P2 $P(\Omega) = 1$

P3 Se $A_1, A_2, \ldots$ são acontecimentos disjuntos dois a dois, então $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \cdots$ (aditividade)

Nota: Quando Ω é um conjunto não numerável, existem alguns subconjuntos "patológicos" de Ω aos quais não é possível associar uma probabilidade; estes conjuntos são ditos não-probabilizáveis; neste curso, quando nos referirmos a um acontecimento, assumimos sempre que se trata de um subconjunto de Ω que é probabilizável.

Propriedades de uma medida de probabilidade

Da definição axiomática, resultam as seguintes propriedades importantes, válidas para quaisquer $A, B \subseteq \Omega$:

Prop P1 $P(\bar{A}) = 1 - P(A)$
(onde $\bar{A}$ designa o complementar de A em Ω, isto é, $\bar{A} = \{\omega \in \Omega : \omega \notin A\}$)

Prop P2 $P(\emptyset) = 0$

Prop P3 $P(A) \leq 1$

Prop P4 Se $A \subseteq B$, então $P(A) \leq P(B)$

Prop P5 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Definição clássica de probabilidade

No caso de Ω ser finito, com n elementos, um caso particular de medida de probabilidade consiste em atribuir a cada acontecimento elementar a probabilidade $\frac{1}{n}$ e a cada acontecimento A a probabilidade

$$P(A) = \frac{\text{número de elementos de } A}{n}.$$

Dizemos, então, que estamos no caso de resultados elementares igualmente prováveis.
Por exemplo, no caso do lançamento do dado, esta atribuição de probabilidade corresponde ao caso de o dado ser equilibrado, para o qual a probabilidade de sair cada face é $\frac{1}{6}$.
Este caso corresponde à primeira definição formal de probabilidade (definição clássica de probabilidade), introduzida por Pierre Simon de Laplace, em 1812.

Probabilidade condicionada. Independência

Após a realização de uma dada experiência aleatória, suponhamos que ocorreu um determinado acontecimento B, tal que $P(B) > 0$. Um dado acontecimento A terá agora uma nova probabilidade associada (eventualmente diferente da que tinha inicialmente) chamada probabilidade de A condicionada a B ou probabilidade de A dado B que denotaremos por $P(A|B)$, e que é definida por

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

Por exemplo, no caso do lançamento do dado, suponhamos que sabemos que ocorreu o acontecimento "saída de face far" ($B = \{2, 4, 6\}$) e que pretendemos saber a probabilidade de ter saído a face com o número 2, isto é $A = \{2\}$. Então

$$P(A|B) = \frac{P(\{2\} \cap \{2,4,6\})}{P(\{2,4,6\})} = \frac{P(\{2\})}{P(\{2,4,6\})} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}.$$

Independência

Dois acontecimentos A e B são ditos independentes se

$$P(A \cap B) = P(A)P(B)$$

Note-se que, se A e B são independentes (e $P(A) > 0$ e $P(B > 0)$) tem-se

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A),$$
$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B)P(A)}{P(A)} = P(B)$$

o que justifica a designação de independentes para os acontecimentos.
A noção de independência generaliza-se para vários acontecimentos. Por exemplo, diremos que três acontecimentos A_1, A_2, A_3 são independentes se se verificar

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$
$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$
$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

(Dizer que três acontecimentos são independentes não é equivalente a dizer que eles são independentes dois a dois).

Regra multiplicativa. Fórmula de Bayes

Seja B tal que $P(B > 0)$. De $P(A|B) = \frac{P(A\cap B)}{P(B)}$ obtém-se, de imediato, a seguinte regra multiplicativa:

$$P(A \cap B) = P(A|B)P(B)$$

Se A e B são tais que $P(A) > 0$ e $P(B) > 0$, temos, então, a chamada fórmula de Bayes:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Variáveis aleatórias

Ao efetuar experiências aleatórias, os resultados podem ser descritos de várias formas. Por exemplo:

1. Na experiência do lançamento de uma moeda, os resultados podem ser descritos por palavras, como "saída de cara" ou "saída de escudo", ou usando, o símbolo C para "saída de cara", e E para "saída de escudo", pelos dois símbolos C e E.
2. Ao retirar uma carta de um baralho, os diversos resultados podem ser descritos por uma mistura de números e palavras (e.g. "sair o 10 de espadas").
3. No lançamento de um dado com as faces numeradas, os resultados podem ser descritos simplesmente pelos números $1, 2, 3, 4, 5, 6$.

Do ponto de vista matemático, seria mais simples se todos os resultados das experiências estivessem associados a números reais.

Variáveis aleatórias

Tal pode ser feito, através da atribuição de um código numérico para os diversos resultados da experiência.

1. No caso do lançamento da moeda, bastaria associar o código 0 para "saída de cara" ou C e 1 para "saída de escudo" ou E.
2. No caso do baralho, bastaria numerar as diversas cartas do baralho de 1 a 52 e os diversos resultados passariam a estar associados a cada um desses números.
3. No caso do dado, temos já números associados ao diversos resultados da experiência.

O que se pretende, então, é definir uma função de Ω (espaço amostral original) em $\mathbb{R}$.
Tal função é chamada uma variável aleatória.

Variável aleatória

Definição

Chama-se variável aleatória (v.a.) a toda a função X do espaço amostral Ω em $\mathbb{R}$.

Exemplo

No caso do lançamento da moeda, em que o espaço amostral era $\Omega = \{C, E\}$, uma v.a. X poderia ser

$$\begin{aligned} X : \{C, E\} &\to \mathbb{R} \\ C &\mapsto 1 \\ E &\mapsto 0 \end{aligned}$$

Esta v.a. toma apenas os valores 0 e 1. O conjunto da valores que a v.a. X assume, ou seja, $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ é o contradomínio de X, também designado por suporte de X.
As variáveis aleatórias são vulgarmente designadas por letras maiúsculas do final do alfabeto: X, Y, etc.

Probabilidades associadas a variáveis aleatórias

Até este momento, definimos probabilidade para acontecimentos, isto é, para subconjuntos do espaço amostral Ω.
Com a introdução de v.a.'s vamos estar interessados em fazer afirmações envolvendo probabilidades sobre (valores de) variáveis aleatórias, por exemplo, dizer que "a probabilidade de a v.a. X assumir valores (ou "estar") entre a e b é 0.5".
Tal é feito, transformando as afirmações acerca dos valores de X em afirmações acerca de subconjuntos de Ω, do seguinte modo.

Definição

Dado $B \subseteq \mathbb{R}$, definimos a probabilidade de X assumir valores em B, e escrevemos $P(X \in B)$, como sendo a probabilidade do conjunto formado pelos elementos de Ω cuja imagem por X está em B i.e.

$$P(X \in B) = P\Big(\{\omega \in \Omega : X(\omega) \in B\}\Big).$$

Nota: Escreveremos também $a \leq X < b$ em vez de $X \in [a, b)$, $X = a$ em vez de $X \in \{a\}$, etc.

Exemplo

Consideremos a experiência aleatória que consiste no lançamento de duas moedas equilibradas e seja X a v.a. que representa o número de caras que ocorrem.

- $\Omega = \{CC, CE, EC, EE\}$
- $$\begin{aligned} X(CC) &= 2 \\ X(CE) = X(EC) &= 1 \\ X(EE) &= 0 \end{aligned}$$

Então, tem-se, por exemplo

$$\begin{aligned} P(X = 2) &= P\Big(\{\omega \in \Omega : X(\omega) = 2\}\Big) = P\Big(\{CC\}\Big) = \tfrac{1}{4} \\ P(X = 1) &= P\Big(\{\omega \in \Omega : X(\omega) = 1\}\Big) = P\Big(\{CE, EC\}\Big) = \tfrac{2}{4} = \tfrac{1}{2} \\ P(X \geq 1) &= P(X = 1 \text{ ou } X = 2) = P(\{CE, CE, CC\}) \\ &= \tfrac{3}{4} = P(X = 1) + P(X = 2). \end{aligned}$$

Variáveis Aleatórias Discretas

Variáveis aleatórias discretas

Definição

Uma v.a. diz-se discreta se o seu contradomínio for um subconjunto finito ou infinito numerável de $\mathbb{R}$, i.e., for da forma

$$X(\Omega) = \{x_1, \ldots, x_n\} \quad \text{ou} \quad X(\Omega) = \{x_1, x_2, x_3, \ldots\}.$$

As variáveis aleatórias discretas ocorrem em muitos problemas, especialmente em casos em que estamos interessados em contar o número de vezes que qualquer coisa acontece: por exemplo, o número de parafusos defeituosos (numa caixa de 1000 parafusos), o número de bits com erro na transmissão de uma mensagem, o número de pessoas numa fila de espera, etc.

Função massa de probabilidade

Quando temos uma v.a. discreta, interessa-nos não só conhecer os valores que ela pode tomar (i.e., o seu contradomínio ou suporte), mas também a probabilidade com que assume cada um desses valores.

Definição

Designa-se por função massa de probabilidade (f.m.p.) (ou densidade discreta) de uma v.a. discreta X a função f_X que associa a cada elemento do contradomínio de X a respetiva probabilidade, i.e.

$$f_X(x_k) = P(X = x_k), \quad x_k \in X(\Omega).$$

É usual usarmos a notação p_k para designar a probabilidade de X assumir o valor x_k, isto é, $p_k = P(X = x_k)$.

É frequente explicitar-se a f.m.p. de uma v.a. discreta como uma tabela onde na primeira linha se indicam os pontos que constituem o suporte da variável e na segunda linha a probabilidade de cada um desses pontos. A f.m.p. pode também ser descrita indicando apenas $\{(x_k, p_k)\}$.

Exemplo

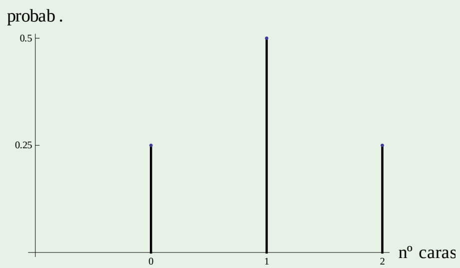
No caso anterior do lançamento de duas moedas equilibradas, em que X é a v.a. associada ao "número total de caras obtido nos dois lançamentos", será a seguinte a sua f.m.p.:

x_k	0	1	2
p_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

As f.m.p. podem também representar-se graficamente por diagramas de linhas.

Exemplo

No caso da v.a. do exemplo anterior, a respetiva f.m.p. poderia ser dada graficamente do seguinte modo:



Propriedades de uma f.m.p.

Uma função $\{(x_k, p_k)\}$ será uma f.m.p. se e só se satisfizer:

FMP1 $\quad p_k \geq 0$ (p_k não negativos)

FMP2 $\quad \sum_k p_k = 1$ (com soma unitária)

Sendo $\{(x_k, p_k)\}$ a f.m.p. de uma certa v.a. X, tem-se

$$P(X \in B) = \sum_{x_k \in B} P(X = x_k) = \sum_{k:\, x_k \in B} p_k.$$

Exemplo

Relativamente à f.m.p. do exemplo anterior, tem-se
$P(0 \leq X < 2) = P(X \in [0,2)) = P(X = 0) + P(X = 1) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.$

No Mathematica, o comando para a f.m.p é PDF (Probability Density Function).

Função de distribuição

Definição

Dada uma v.a. X, a função de distribuição de X, denotada por F_X, é a função dada por

$$F_X(x) = P(X \leq x),\ x \in \mathbb{R}.$$

Se a v.a. discreta X tem f.m.p. $\{(x_k, p_k)\}$, a respetiva função de distribuição é dada por

$$F_X(x) = \sum_{x_k \leq x} P(X = x_k) = \sum_{k:\, x_k \leq x} p_k.$$

Exemplo

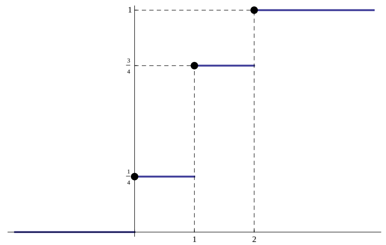
No caso do exemplo anterior, em que a f.m.p. f_X é dada por

x_k	0	1	2
p_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

tem-se

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{4}, & 0 \leq x < 1, \\ \frac{3}{4}, & 1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

A representação gráfica da função de distribuição anterior é a seguinte:



A função F_X (e qualquer função de distribuição de uma v.a. discreta) é uma função:

- em escada
- não decrescente
- com descontinuidades nos pontos x_k (pontos do suporte de X)
- contínua à direita (e com limite finito à esquerda) em cada um dos pontos x_k

Além disso, tem-se

- $$\lim_{x \to -\infty} F_X(x) = 0 \quad \text{e} \quad \lim_{x \to +\infty} F_X(x) = 1.$$

Vimos que, a partir da f.m.p. de uma v.a. discreta, podemos construir a sua função de distribuição.
Reciprocamente, se tivermos uma v.a. discreta da qual conheçamos a função de distribuição, podemos facilmente reconstruir a respetiva f.m.p.:

- o suporte da v.a. é constituído pelos pontos x_k que são pontos de descontinuidade de F;
- em cada x_k, o valor $f_X(x_k) = P(X = x_k)$ é dado pelo valor do salto de F nesse ponto, i.e.

$$f_X(x_k) = F(x_k) - F(x_k^-).$$

Dar a distribuição de probabilidade de uma v.a. discreta X é dar a sua função massa de probabilidade ou a sua função de distribuição.

No Mathematica, o comando para a função de distribuição é CDF (Cummulative Distribution Function).

Caraterísticas teóricas de uma v.a. discreta

Interessam-nos agora o cálculo de algumas medidas de localização, dispersão e forma que caracterizam as distribuições de probabilidade de uma v.a. discreta X.
No que se segue, supomos que X é ou uma v.a. discreta com distribuição de probabilidade dada por uma f.m.p. $\{(x_k, p_k)\}$.

Definição

O valor médio de X (ou valor esperado de X ou esperança de X), designado por μ_X (por vezes, apenas por μ) ou por $E(X)$, é dado

$$\mu_X = E(X) = \sum_k x_k p_k$$

Nota: O valor médio é dado pela expressão acima, apenas quando a série envolvida for absolutamente convergente, i.e. tivermos $\sum_k |x_k| p_k < \infty$; caso contrário, dizemos que o valor médio não existe. Nas definições seguintes que envolvam séries, será sempre usado o mesmo princípio.

Exemplo

Seja X a v.a. correspondente ao número de caras obtido no lançamento de duas moedas equilibradas, cuja f.m.p. é dada por

x_k	0	1	2
p_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

O valor médio desta v.a. é dado por:

$$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1.$$

Note-se que o valor médio ou valor esperado de X, **nem sempre é** um dos valores que X pode assumir.

Propriedades do valor médio

1. Se g for uma função real e $Y = g(X)$ for uma v.a., então

$$E(Y) = E(g(X)) = \sum_k g(x_k) p_k$$

Exemplo

Retomando a v.a. discreta X anterior de f.m.p. dada por

x_k	0	1	2
p_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

tem-se, por exemplo,

$$E(X^2) = \sum_k x_k^2 p_k = = 0^2 \times \frac{1}{4} + 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} = \frac{3}{2}.$$

Propriedades do valor médio

2. $$E(a + bX) = a + bE(X)$$

Dem:

$$E(a + bX) = \sum_k (a + bx_k)p_k = \sum_k (ap_k + bx_k p_k) = \sum_k ap_k + \sum_k bx_k p_k$$
$$= a \sum_k p_k + b \sum_k x_k p_k = a + bE(X)$$

3. $$E(a) = a$$

Dem: Basta tomar $b = 0$ na fórmula anterior.

Variância

Definição

Se X é uma v.a. discreta com valor médio μ, a variância de X, denotada por $\mathrm{var}(X)$ ou σ_X^2, ou por vezes, apenas por σ^2, é definida por

$$\mathrm{var}(X) = E\Big((X-\mu)^2\Big) = \sum_k (x_k - \mu)^2 p_k$$

Exemplo

No caso da v.a. discreta X que temos vindo a considerar, cuja f.m.p. é

x_k	0	1	2
p_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

tem-se, como vimos, $E(X) = 1$, vindo, portanto

$$\mathrm{var}(X) = \sum_k (x_k - 1)^2 p_k = (0-1)^2 \frac{1}{4} + (1-1)^2 \frac{1}{2} + (2-1)^2 \frac{1}{4} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Propriedades da variância

1. $$\mathrm{var}(X) = E(X^2) - (E(X))^2$$

Dem:

$$\mathrm{var}(X) = E\Big((X-\mu)^2\Big) = E(X^2 - 2\mu X + \mu^2)$$
$$= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2$$
$$= E(X^2) - \mu^2 = E(X^2) - (E(X))^2$$

Exemplo

No caso da v.a. X anterior, já tínhamos calculado $E(X^2) = \frac{3}{2}$. Como $E(X) = 1$, tem-se

$$\mathrm{var}(X) = \frac{3}{2} - 1 = \frac{1}{2}.$$

Propriedades da variância

2. $$\mathrm{var}(a + bX) = b^2\,\mathrm{var}(X)$$

3. $$\mathrm{var}(bX) = b^2\,\mathrm{var}(X)$$

4. $$\mathrm{var}(a) = 0$$

Desvio padrão. Estandardização de uma v.a.

Definição

A quantidade $\sqrt{\mathrm{var}(X)} = \sigma$ é chamada desvio padrão da v.a. X.

O resultado seguinte, cuja demonstração (muito simples) deixamos ao cuidado dos alunos, será usado com frequência.

Se X é uma v.a. com valor médio $E(X) = \mu$ e variância $\mathrm{var}(X) = \sigma^2$ (desvio padrão σ), então a v.a.

$$Y = \frac{X - \mu}{\sigma}$$

tem valor médio nulo e variância igual a 1 (logo, o seu desvio padrão também é igual a 1), i.e. tem-se

$$E(Y) = 0 \quad \text{e} \quad \mathrm{var}(Y) = 1.$$

Coeficiente de assimetria. Moda

Definição

Sendo X uma v.a. discreta com valor médio μ e desvio padrão σ, a quantidade

$$\beta_1 = \frac{E\Big((X-\mu)^3\Big)}{\sigma^3} = \frac{1}{\sigma^3}\sum_k (x_k - \mu)^3\, p_k$$

é chamada coeficiente de assimetria da v.a. X.

Definição

Chama-se moda de uma v.a. discreta X (de f.m.p. $\{(x_k, p_k)\}$) ao valor x_k ao qual corresponde o maior p_k.

Nota: Pode haver mais de uma moda (dizemos que a distribuição é unimodal se tiver uma só moda, bimodal se tiver duas modas, etc, dizendo-se plurimodal se tiver várias modas).

Modelos de Probabilidade Discretos

Existem certas distribuições de probabilidade discretas que aparecem com muita frequência em aplicações. Vamos agora estudar algumas dessas distribuições (modelos probabilísticos).

Definição

Chama-se experiência de Bernoulli (ou prova de Bernoulli) a uma experiência aleatória com apenas dois resultados possíveis - um a que chamamos sucesso (S) e outro que consideramos insucesso (NS).

O espaço amostral para uma experiência de Bernoulli é $\Omega = \{S, NS\}$.

Exemplo

1. Lançar uma moeda e considerar saída de cara como sucesso e saída de escudo como insucesso.
2. Verificar a qualidade de um artigo numa linha de produção: sucesso - artigo não defeituoso; insucesso - artigo defeituoso.

Distribuição de Bernoulli

Se X for a v.a. tal que $X(S) = 1$ e $X(NS) = 0$, então o suporte dessa v.a. é $X(\Omega) = \{0, 1\}$ e a sua f.m.p. é dada por $P(X = 0) = 1 - p$ e $P(X = 1) = p$ onde $p\,(0 < p < 1)$ é a probabilidade de sucesso (e $1 - p$ é a probabilidade de insucesso).

Definição

Se X é uma v.a. cuja f.m.p. é $\begin{array}{c|cc} x_k & 0 & 1 \\ \hline p_k & 1-p & p \end{array}$ dizemos que X é uma variável aleatória de Bernoulli (ou modelo de Bernoulli) com parâmetro p e escrevemos $X \frown Ber(p)$.

Facilmente se verifica que, se $X \frown Ber(p)$, então

$$\mu_X = p \quad \text{e} \quad \sigma_X^2 = p(1-p).$$

No Mathematica, o comando associado à distribuição de Bernoulli é BernoulliDistribution.

Modelo binomial

Considere o seguinte tipo de experiência aleatória:

1. A experiência consiste na repetição de n (n fixo) provas de Bernoulli (cada prova com apenas dois resultados possíveis: sucesso, insucesso).
2. As provas são independentes, ou seja, o resultado obtido numa prova não afecta o resultado das restates provas.
3. A probabilidade de sucesso em todas as provas é constante (constante que denotaremos por p, $0 < p < 1$).

Se for X a v.a. que representa o número de sucessos obtidos na experiência anterior, é fácil de verificar que X tem como suporte o conjunto $\{0, 1, \ldots, n\}$, sendo a sua f.m.p. dada por

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \; k = 0, 1, 2, \ldots, n$$

onde

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Modelo binomial

Definição

Diz-se que uma v.a. X tem uma distribuição binomial de parâmetros n e p ($n \in \mathbb{N}, 0 < p < 1$) e escreve-se $X \frown Bi(n, p)$ se a f.m.p. de X for dada por

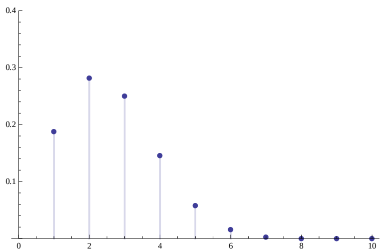
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \; k = 0, 1, 2, \ldots, n.$$

Nota: O caso $n = 1$ corresponde a uma distribuição de Bernoulli $Ber(p)$.

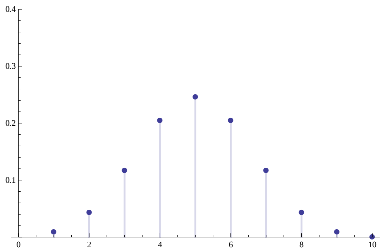
Se $X \frown Bi(n, p)$, então pode mostrar-se que:

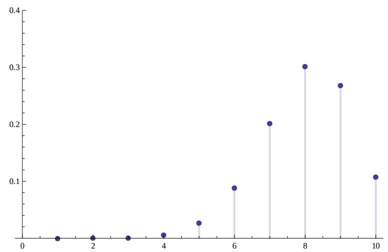
- $\mu_X = np$
- $\sigma_X^2 = np(1-p)$
- $\beta_1 = \dfrac{1-2p}{\sqrt{np(1-p)}}$
- $\lfloor p(n+1) \rfloor$ é moda de X (onde $\lfloor a \rfloor$ designa o maior inteiro não superior a a).

Nas figuras seguintes apresentam-se gráficos da f.m.p. do modelo binomial para diferentes valores de p, para $n = 10$.



Função massa de probabilidade do modelo $Bi(10, 0.25)$





Função massa de probabilidade do modelo $Bi(10, 0.8)$

No Mathematica, para trabalhar com a distribuição binomial, deve usar a função BinomialDistribution.

Amostragem aleatória numa população dicotómica

Dizemos que uma dada população é dicotómica se cada elemento da população tem ou não tem uma certa caraterística.

Exemplo

- numa população humana: fumador/não fumador, portador/não portador de um certo vírus
- numa população constituída por parafusos: defeituoso/não defeituoso

Suponhamos que temos uma população dicotómica e seja p a proporção de elementos com a caraterística em estudo. Se recolhermos uma amostra aleatória de tamanho n dessa população – i.e., se fizermos n extrações ao acaso, com reposição, de elementos dessa população – então, a v.a. que corresponde ao número de elementos da amostra que têm a caraterística em causa tem uma distribuição $Bi(n, p)$.

Nota: Quando o número de elementos da população N é muito grande (e o tamanho da amostra, n, é pequeno em comparação com N) fazer amostragem com ou sem reposição é praticamente equivalente e podemos ainda aplicar o modelo binomial se usarmos amostragem sem reposição.

Distribuição de Poisson

Definição

Uma v.a. X com contradomínio $\{0, 1, 2, \ldots\}$ diz-se uma variável de Poisson com parâmetro λ ($\lambda > 0$), se a sua f.m.p. for dada por

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \; k = 0, 1, 2, \ldots.$$

Se X tem uma distribuição de Poisson com parâmetro λ, escrevemos

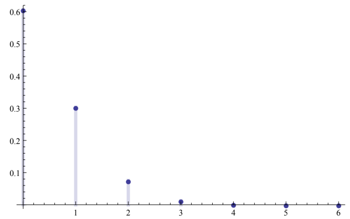
$$X \frown Poi(\lambda).$$

Características teóricas do modelo de Poisson

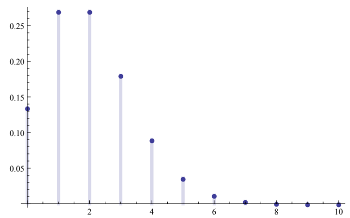
Se $X \frown Poi(\lambda)$, tem-se:

- $\mu_X = \lambda$
- $\sigma_X^2 = \lambda$
- $\beta_1 = \frac{1}{\sqrt{\lambda}}$
- $\lfloor \lambda \rfloor$ é moda de X (e, se λ é inteiro, então λ e $\lambda - 1$ são modas de X).

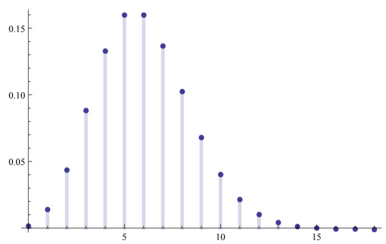
Nas figuras seguintes apresentam-se gráficos da f.m.p. da distribuição $Poi(\lambda)$ para diferentes valores de λ.



F.m.p. do modelo $Poi(0.5)$



F.m.p do modelo $Poi(2)$

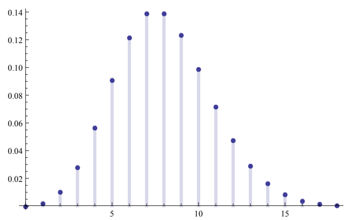


F.m.p. do modelo $Poi(6)$

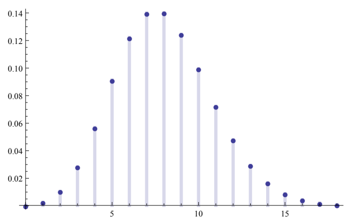
Modelo de Poisson e Modelo Binomial

O modelo binomial $Bi(n,p)$, quando n é grande e p é pequeno (i.e. quando o sucesso é um acontecimento "raro"), pode ser aproximado pelo modelo de Poisson com parâmetro $\lambda = np$. Este facto justifica o nome de lei dos acontecimentos raros atribuído geralmente ao modelo de Poisson. De facto, mostra-se que quando $n \to \infty$ com $\lambda = np$ mantido constante, se tem

$$\binom{n}{k} p^k (1-p)^{n-k} \to e^{-\lambda} \frac{\lambda^k}{k!}$$



F.m.p do modelo $Poi(8)$



F.m.p do modelo $Bi(1000, 0.008)$

Aplicação do modelo de Poisson

O modelo de Poisson é utilizado como modelo de contagem de fenómenos que ocorrem no tempo (ou no espaço) de forma aleatória (quando a probabilidade de ocorrência desses fenómenos é pequena). O parâmetro λ corresponde ao número médio de ocorrências numa unidade de tempo (ou espaço).

Exemplo

- número de partículas radioativas emitidas por unidade de tempo
- número de colónias de bactérias por unidade de volume de uma certa solução
- número de terramotos ocorridos (num determinado local) por ano

No Mathematica, a função associada ao modelo de Poisson é a função PoissonDistribution.

Outros modelos discretos

1. **Modelo uniforme em n pontos $X \frown U\{x_1, \ldots, x_n\}$**

$$X(\Omega) = \{x_1, \ldots, x_n\}$$
$$P(X = x_k) = \frac{1}{n}; \quad k = 1, 2, \ldots, n,$$

Exemplo

A v.a. que representa o número obtido no lançamento de um dado equilibrado, com as faces numeradas de 1 a 6, tem distribuição uniforme no conjunto $\{1, 2, \ldots, 6\}$.

No Mathematica, o comando asociado à distribuição uniforme em n pontos é a função DiscreteUniformDistribution.

1. **Modelo geométrico de parâmetro p $X \frown Geo(p)$**

$$X(\Omega) = \{0, 1, 2, 3, \ldots\}$$
$$P(X = k) = (1-p)^k p, \;\; k = 0, 1, 2, 3, \ldots$$

É a distribuição da v.a. X que representa o número de provas de Bernoulli (independentes e com probabilidade de sucesso p) realizadas até que ocorra o primeiro sucesso (número de tentativas falhadas).

No Mathematica, o comando associado à distribuição geométrica é GeometricDistribution.

3. **Modelo hipergeométrico de parâmetro m, n, N** $X \frown Hip(m, n, N)$

$$P(X = k) = \frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}}$$

Dada uma população de N elementos, dos quais n têm uma determinada caraterística, se retiramos dessa população uma amostra de tamanho m, sem reposição, a v.a. que representa o número de elementos da amostra com a caraterística em causa tem uma distribuição $Hip(m, n, N)$.

No Mathematica, o comando associado à distribuição hipergeométrica é HypergeometricDistribution.

Comandos do Mathematica para Modelos Discretos

Dist. Bernoulli	BernoulliDistribution
Dist. Binomial	BinomialDistribution
Dist. Geométrica	GeometricDistribution
Dist. Hipergeométrica	HypergeometricDistribution
Dist. Poisson	PoissonDistribution
Dist. Uniforme	DiscreteUniformDistribution
Valor médio	Mean
Variância	Variance
Desvio Padrão	StandardDeviation
Coeficiente de Assimetria	Skewness
Função massa de probabilidade	PDF
Função de distribuição	CDF
Cálculo de probabilidades	Probability

Variáveis Aleatórias Contínuas

Variáveis aleatórias não discretas

Consideremos, por exemplo, a v.a. X que representa a dose de um determinado medicamento que deve ser dada a um certo doente até que este reaja positivamente.
Neste caso, os valores possíveis desta v.a. são todos os valores do intervalo $]0, \infty[$ de $\mathbb{R}$, o qual é um conjunto não numerável. Trata-se de uma v.a. não discreta.
Naturalmente, neste caso não é possível definir a f.m.p da maneira que fizemos para o caso de uma variável discreta.
No entanto, poderá existir uma função f que desempenhe uma papel "semelhante" ao da f.m.p para o caso discreto.
Tal função, chamada função densidade de probabilidade (f.d.p.), deverá satisfazer as seguintes propriedades:

FDP1 $\quad f_X(x) \geq 0, \forall x$

FDP2 $\quad \displaystyle\int_{-\infty}^{\infty} f_X(x)dx = 1$

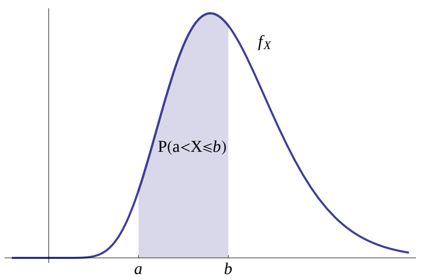
Variável aleatória contínua

Diremos que X é uma v.a. contínua se existir uma função densidade de probabilidade f_X tal que, para qualquer subconjunto (probabilizável) $B \subseteq \mathbb{R}$ se tenha

$$P(X \in B) = \int_B f_X(x)dx$$

Por exemplo, se $B =]a, b]$, ter-se-á

$$P(X \in B) = P(a < X \leq b) = \int_a^b f_X(x)dx$$



Tem-se, naturalmente, atendendo às propriedades do integral

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b),$$

o que não é válido no caso de X ser uma v.a. discreta.
Note-se que, no caso de uma v.a. contínua com f.d.p. f_X, tem-se

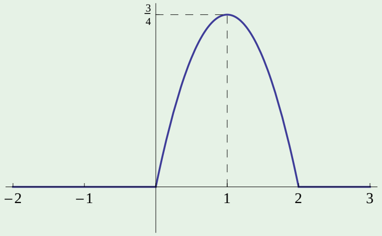
$$P(X = a) = \int_a^a f_X(x)dx = 0,$$

ou seja, a probabilidade de uma v.a. contínua assumir um valor particular a é zero.

Exemplo

Seja X a v.a. contínua cuja f.d.p. é dada por

$$f_X(x) = \begin{cases} \frac{3}{4}(2x - x^2), & 0 < x < 2, \\ 0, & \text{restantes valores de } x \end{cases}$$



Exemplo

Note-se que temos, de facto, $f_X(x) \geq 0$ para todo o $x \in \mathbb{R}$ (uma vez que, para $0 < x < 2$, se tem $\frac{3}{4}(2x - x^2) > 0$) e também

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^2 \frac{3}{4}(2x - x^2)dx$$

$$= \frac{3}{4}\left[x^2 - \frac{x^3}{3}\right]_{x=0}^{x=2} = 1.$$

Neste caso, tem-se, por exemplo,

$$P(X > 1) = \int_1^{\infty} f_X(x)dx = \frac{3}{4}\int_1^2 (2x - x^2)dx = \frac{1}{2}.$$

Função de distribuição

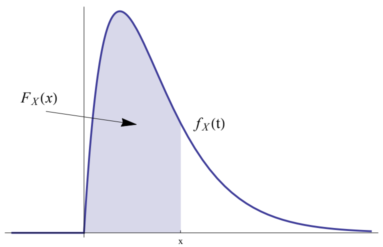
Tal como para o caso discreto, também no caso de uma v.a. contínua X é possível definir a sua função de distribuição, F_X, a qual é dada por

$$F_X(x) = P(X \leq x), \ x \in \mathbb{R}.$$

Se X tem f.d.p. f_X, então

$$\boxed{F_X(x) = \int_{-\infty}^{x} f_X(t)dt}$$

Graficamente, o valor de $F_X(x)$ é a área da figura compreendida entre o eixo das abcissas e a curva $f_X(t)$, para $t \leq x$.



A equação

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

mostra que F_X é uma primitiva de f_X. Assim sendo, tem-se

$$P(a < x < b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a).$$

Naturalmente, tem-se também

$$P(a \leq x < b) = P(a < x \leq b) = P(a \leq X \leq b) = F_X(b) - F_X(a).$$

Exemplo

No caso da v.a. do exemplo anterior, tem-se a seguinte função de distribuição:

- para $x < 0$, $F_X(x) = \int_0^x f_X(t)dt = \int_0^x 0dt = 0$
- para $0 \leq x \leq 2$,

$$F_X(x) = \frac{3}{4}\int_0^x (2t - t^2)dt$$
$$= \frac{3}{4}\left[t^2 - \frac{t^3}{3}\right]_{t=0}^{t=x} = \frac{3}{4}(x^2 - \frac{x^3}{3})$$

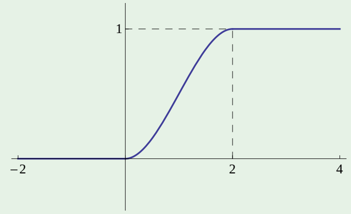
- para $x > 2$,

$$F_X(x) = \frac{3}{4}\int_0^2 (2t - t^2)dt = 1$$

Exemplo

A função de distribuição desta v.a. é, assim, dada por

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{3}{4}(x^2 - \frac{x^3}{3}), & 0 \leq x \leq 2 \\ 1, & x > 2 \end{cases}$$



A função de distribuição F_X de uma v.a. contínua X é (sempre) uma função contínua, não decrescente e que satisfaz

$$\lim_{x \to -\infty} F_X(x) = 0 \quad \text{e} \quad \lim_{x \to +\infty} F_X(x) = 1.$$

Caraterísticas teóricas de v.a.'s contínuas

À semelhança do que fizemos no caso discreto, vamos definir agora algumas caraterísticas para as v.a.'s contínuas. Em tudo quanto se segue, X designa uma v.a. contínua com função densidade de probabilidade f_X.

Definição

O valor médio de X, ou valor esperado de X, ou esperança de X, designado por μ_X (por vezes, apenas por μ) ou $E(X)$, é definido do seguinte modo:

$$\mu_X = \mu = E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$$

Nota: O valor médio é dado pela expressão indicada, apenas quando o integral for absolutamente convergente, i.e. tivermos $\int_{-\infty}^{\infty} |x| f_X(x)dx < \infty$. Nas definições seguintes que envolvam integrais será sempre usado o mesmo princípio, isto é, a caraterística em causa estará definida se e só se o integral envolvido na sua definição for absolutamente convergente.

Exemplo

Retomando a X a v.a. contínua considerada anteriormente, com f.d.p. dada por

$$f_X(x) = \begin{cases} \frac{3}{4}(2x - x^2), & 0 < x < 2, \\ 0, & \text{restantes valores de } x \end{cases}$$

tem-se

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \frac{3}{4}\int_0^2 (2x^2 - x^3)dx = \frac{3}{4}\left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_{x=0}^{x=2} = 1.$$

Valor médio de uma função de X

Sendo g uma função real, tem-se

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

Exemplo

No caso da v.a. contínua que temos vindo a considerar tem-se, por exemplo

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\ &= \frac{3}{4} \int_0^2 x^2 (2x - x^2) dx \\ &= \frac{3}{4} \left[\frac{x^4}{2} - \frac{x^5}{5} \right]_{x=0}^{x=2} = \frac{6}{5}. \end{aligned}$$

Propriedades do valor médio

As propriedades do valor médio que referimos para v.a.'s discretas são também válidas no caso de v.a.'s contínuas, i.e., temos

$$E(a + bX) = a + bE(X)$$

de onde se obtém, em particular

$$E(bX) = bE(X) \quad \text{e} \quad E(a) = a$$

Variância e desvio padrão

Definição

Se X é uma v.a. com valor médio μ_X, a variância de X, denotada por $\text{var}(X)$ ou σ_X^2, ou por vezes, apenas por σ^2, é definida como

$$\text{var}(X) = \sigma_X^2 = E\Big((X - \mu_X)^2\Big) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

Tem-se, também, tal como no caso discreto

$$\text{var}(X) = E(X^2) - (E(X))^2$$

e

$$\text{var}(a + bX) = b^2 \,\text{var}(X)$$

Definição

A quantidade $\sqrt{\text{var}(X)} = \sigma$ é chamada desvio padrão da v.a. X.

Estandardização de uma v.a.

Tal como no caso discreto, se X for uma v.a. com valor médio $E(X) = \mu$ e variância $\text{var}(X) = \sigma^2$ (desvio padrão σ), então a v.a.

$$Y = \frac{X - \mu}{\sigma}$$

terá valor médio nulo e variância igual a 1, i.e. tem-se

$$E(Y) = 0 \quad \text{e} \quad \text{var}(Y) = 1.$$

Coeficiente de assimetria

Definição

Sendo X uma v.a. contínua (com f.d.p. f_X), com valor médio μ e desvio padrão σ, a quantidade

$$\beta_1 = \frac{E\Big((X-\mu)^3\Big)}{\sigma^3} = \frac{1}{\sigma^3}\int_{-\infty}^{\infty} (x_k-\mu)^3\, f_X(x)dx$$

é chamada coeficiente de assimetria da v.a. X.

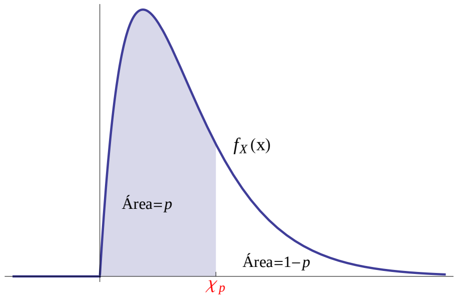
Quantis teóricos

Definição

Sendo $0 < p < 1$, chama-se quantil (de probabilidade) p, e denota-se por χ_p, o valor dado por

$$\chi_p = \inf\{x \in \mathbb{R} : F_X(x) \geq p\}.$$

Se desenharmos o gráfico da f.d.p. f_X, $p \times 100\%$ da área compreendida entre a curva f_X e o eixo dos xx estará para a esquerda de χ_p e $(1-p) \times 100\%$ para a sua direita.



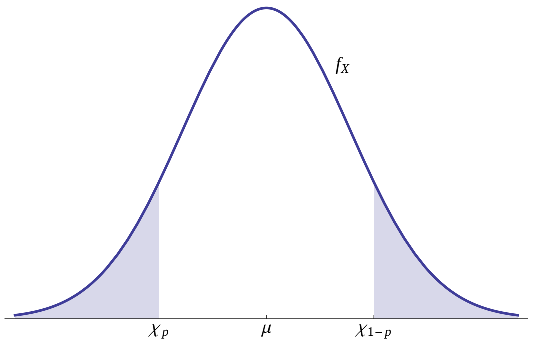
Mediana, quartis

- O quantil de probabilidade $p = 1/2$ é chamado mediana de X.
- Os quantis de probabilidades $p = 1/4, 1/2, 3/4$ são chamados quartis.
- De modo análogo se definem os decis ($p = i/10; i = 1, \ldots, 10$), percentis($p = i/100; i = 1, \ldots, 100$), etc.

Nota Definimos apenas os quantis para o caso de v.a's contínuas, por ser para este tipo de variáveis que faremos grande uso desse conceito; no entanto, a definição apresentada é válida também para o caso discreto.

Quantis de distribuições simétricas

Convém referir que, no caso de termos uma distribuição simétrica (com valor médio μ), os quantis de probabilidade p e $1-p$, χ_p e χ_{1-p}, estarão situados simetricamente em relação a μ.



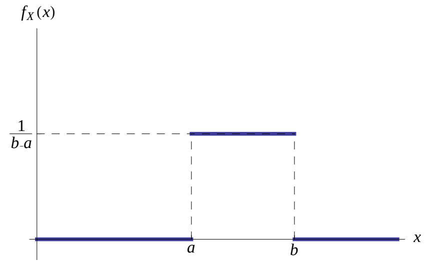
Modelo uniforme

Vamos apresentar agora alguns modelos de distribuições contínuas importantes em aplicações.

Definição

Diz-se que uma v.a. X tem distribuição uniforme no intervalo $[a, b] \subset \mathbb{R}$ e escreve-se, abreviadamente, $X \frown U[a, b]$, se a sua f.d.p. for

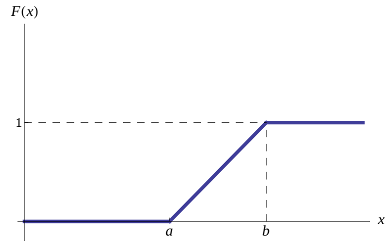
$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & \text{se } a \leq x \leq b, \\ 0, & \text{outros valores de } x. \end{cases}$$



Modelo Uniforme

A função de distribuição de $X \frown U[a, b]$ é dada por

$$F_X(x) = \begin{cases} 0, & \text{se } x < a \\ \dfrac{x-a}{b-a}, & \text{se } a \leq x < b \\ 1, & \text{se } x \geq b \end{cases}$$



Características teóricas da distribuição uniforme

Se $X \frown U[a, b]$, tem-se:

- $\mu_X = \dfrac{a+b}{2}$
- $\sigma_X^2 = \dfrac{(b-a)^2}{12}$
- $\chi_{1/2} = \dfrac{a+b}{2}$
- $\beta_1 = 0$

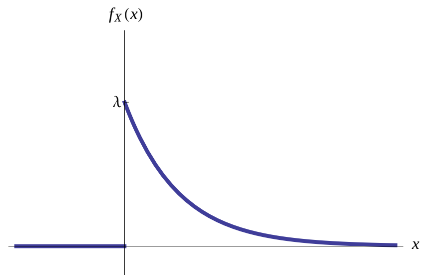
No Mathematica, o comando para a distribuição uniforme é UniformDistribution.

Distribuição exponencial

Definição

Diz-se que uma v.a. X tem uma distribuição exponencial com parâmetro λ $(\lambda > 0)$ e escreve-se $X \frown Exp(\lambda)$, se a sua função densidade de probabilidade for dada por

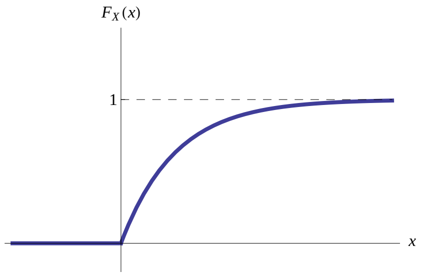
$$f_X(x) = \begin{cases} 0, & \text{se } x < 0, \\ \lambda e^{-\lambda x}, & \text{se } x \geq 0 \end{cases}$$



Distribuição exponencial

A função de distribuição de uma v.a. $X \frown Exp(\lambda)$ é dada por

$$F_X(x) = \begin{cases} 0, & \text{se } x < 0, \\ 1 - e^{-\lambda x}, & \text{se } x \geq 0, \end{cases}$$



Características teóricas da distribuição exponencial

Se $X \frown Exp(\lambda)$, $\lambda > 0$, tem-se:

- $\mu_X = \dfrac{1}{\lambda}$
- $\sigma_X^2 = \frac{1}{\lambda^2}$
- $\chi_{1/2} = \frac{\log 2}{\lambda}$
- $\beta_1 = 2$ (a distribuição tem assimetria positiva, para qualquer λ)

Nota: O valor médio e o desvio padrão de uma v.a. $X \frown Exp(\lambda)$ são iguais (ambos valem $\frac{1}{\lambda}$).

Utilização do modelo exponencial

A distribuição exponencial está geralmente associada a variáveis que medem a quantidade de tempo que decorre até que um acontecimento específico (geralmente, "raro") ocorra: por exemplo, o tempo, contado a partir deste instante, que decorre até que ocorra um tremor de terra, ou até que recebamos uma chamada telefónica por engano ou o intervalo de tempo entre falhas consecutivas de um dado aparelho, etc.

No Mathematica, para trabalhar com a distribuição exponencial, deverá usar o comando ExponentialDistribution.

Distribuição normal ou Gaussiana

De todas as distribuições contínuas, a de maior importância é, sem dúvida, a distribuição normal ou Gaussiana.

Definição

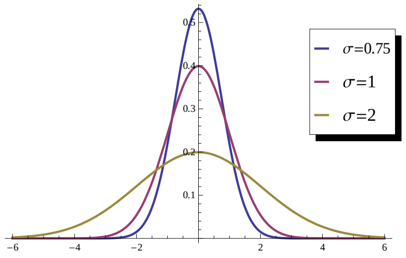
Diz-se que uma v.a. X tem uma distribuição normal (ou uma distribuição Gaussiana) com parâmetros μ e σ ($\sigma > 0$), e escreve-se abreviadamente $X \frown N(\mu, \sigma)$, se a sua f.d.p. for dada por

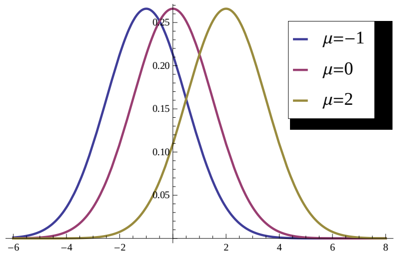
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

É fácil de mostrar que, se $X \frown N(\mu, \sigma)$, então

- $E(X) = \mu$
- $\text{var}(X) = \sigma^2$

Assim, os parâmetros μ e σ da v.a. $X \frown N(\mu, \sigma)$ são o seu valor médio e o seu desvio padrão, respectivamente.





A distribuição normal é simétrica e a sua f.d.p.tem a bem conhecida "forma de sino".

Estandardização da normal

Uma propriedade importante da distribuição normal é que, se X tiver uma distribuição normal, então a v.a. $Y = a + bX$ terá ainda uma distribuição normal.
Então, se tivermos uma v.a. $X \frown N(\mu, \sigma)$, se considerarmos a v.a. estandardizada

$$Z = \frac{X-\mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma}X$$

esta variável aleatória será ainda normal e, como sabemos, terá valor médio $\mu_Z = 0$ e desvio padrão $\sigma_Z = 1$, ou seja, ter-se-á $Z \frown N(0,1)$.
Mais geralmente, tem-se

$$X = \sigma Z + \mu \frown N(\mu, \sigma) \iff Z = \frac{X-\mu}{\sigma} \frown N(0,1).$$

Normal reduzida $Z \frown N(0,1)$

Quando uma v.a. tem uma distribuição $N(0,1)$ dizemos que é uma normal reduzida ou standard e reservamos, geralmente, a letra Z para a designar.
A f.d.p. da normal reduzida é usualmente denotada por ϕ e a sua expressão é dada por

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

A função de distribuição da normal reduzida Z é usualmente designada por Φ e é dada por

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

Nota: Note-se que esta função não tem uma expressão analítica, uma vez que a função $\phi(x)$ não é integrável analiticamente.

Características teóricas do modelo normal

Seja $X \frown N(\mu, \sigma)$. Tem-se, então:

- $\mu_X = \mu$;
- $\sigma_X^2 = \sigma^2$;
- $\chi_{1/2} = \mu$;
- $\beta_1 = 0$.

O comando do Mathematica para a distribuição normal é NormalDistribution.

Comandos do Mathematica para Modelos Contínuos

Dist. Uniforme	UniformDistribution
Dist. Exponencial	ExponentialDistribution
Dist. Normal	NormalDistribution
Valor médio	Mean
Mediana	Median
Quantil	Quantile
Quartis	Quartiles
Variância	Variance
Desvio Padrão	StandardDeviation
Coeficiente de Assimetria	Skewness
Função densidade de probabilidade	PDF
Função de distribuição	CDF
Cálculo de probabilidades	Probability

Complementos sobre Variáveis Aleatórias

Variáveis independentes

O conceito de acontecimentos independentes estende-se, de um modo natural, ao caso de variáveis aleatórias.

Definição

Duas variáveis X e Y (definidas no mesmo espaço amostral) são ditas independentes se, para quaisquer subconjuntos são A e B de $\mathbb{R}$ tivermos

$$P(X \in A, Y \in B) = P(X \in A)\, P(X \in B).$$

Mais geralmente, n variáveis aleatórias $X_1, X_2, \ldots, X_n$ são independentes se e só se para quaisquer subconjuntos $B_1, B_2, \ldots, B_n$ de $\mathbb{R}$ tivermos

$$P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = P(X_1 \in A_1)\, P(X_2 \in B_2) \ldots P(X_n \in B_n)$$

Valor médio e variância da soma de v.a.'s

Dadas duas v.a.'s X e Y (independentes ou não), pode provar-se que

$$E(X+Y) = E(X) + E(Y)$$

Relativamente à variância, o mesmo tipo de resultado não é, em geral, verdadeiro. Basta pensar, por exemplo, que

$$\mathrm{var}(X+X) = \mathrm{var}(2X) = 2^2\,\mathrm{var}(X) = 4\,\mathrm{var}(X) \neq \mathrm{var}(X) + \mathrm{var}(X).$$

No entanto, pode provar-se que, se X e Y forem v.a.'s independentes, então

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y).$$

Os resultados anteriores generalizam-se para n variáveis; por exemplo, tem-se

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n).$$

Variáveis i.i.d.

Consideremos uma dada experiência aleatória e seja X uma v.a. associada a essa experiência (por exemplo, o lançamento de um dado com as faces numeradas de 1 a 6 e a v.a. que corresponde ao número da face que fica voltada para cima).
Considerem-se n repetições, nas mesmas condições, dessa experiência e sejam $X_1, X_2, \ldots, X_n$ as v.a.'s associadas a cada uma das repetições.
Então essas variáveis aleatórias são, naturalmente, independentes e, além disso, todas elas têm a mesma distribuição de probabilidade (distribuição igual à da v.a. X) – dizemos então que as variáveis são independentes e identicamente distribuídas (i.i.d.) com X.

Amostragem aleatória

Suponhamos que temos uma população com uma v.a. X em estudo (por exemplo, a variável *altura* na população dos estudantes da Universidade do Minho).
Consideremos uma v.a. X_1 associada à altura de um indivíduo escolhido aleatoriamente dessa população; seja X_2 a v.a. associada à altura de novo indivíduo escolhido aleatoriamente da população, e assim sucessivamente, até termos n variáveis aleatórias correspondentes, respectivamente, às alturas de n indivíduos escolhidos aleatoriamente da população (indivíduos esses que constituem uma amostra casual da população).
Isto corresponde a escolher, ao acaso e com reposição, uma amostra da população e a considerar as variáveis aleatórias associadas a cada um dos elementos da amostra.
As variáveis $X_1, X_2, \ldots, X_n$ assim obtidas são independentes e igualmente distribuídas, sendo a distribuição de cada X_i igual à distribuição de X (i.e. estas variáveis são i.i.d. com X.)

Amostra aleatória teórica de uma v.a. X

Definição

Dada uma v.a. X com uma certa distribuição, se tivermos n v.a.'s $X_1, X_2, \ldots, X_n$ i.i.d. com X, dizemos que essas variáveis formam uma amostra aleatória (teórica) de dimensão n de X.

A uma amostra aleatória $X_1, X_2, \ldots, X_n$ de X associamos as seguintes variáveis aleatórias:

- v.a. soma $S_n = X_1 + X_2 + \cdots + X_n$
- v.a. média amostral $\overline{X} = \dfrac{S_n}{n} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$
- v.a. variância amostral $S^2 = \dfrac{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1}$
- v.a. desvio padrão amostral $S = \sqrt{\dfrac{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1}}$

Valor médio e variância de S_n e $\overline{X}$

Seja X uma v.a. tal que $E(X) = \mu$ e $\mathrm{var}(X) = \sigma^2$ e seja $X_1, \ldots, X_n$ uma amostra aleatória de X.
Nota: Em particular, as v.a.'s X_i são independentes, $E(X_i) = \mu$ e $\mathrm{var}(X_i) = \sigma^2$.

Têm-se os seguintes resultados:

- $E(S_n) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = n\mu$
- $E(\overline{X}) = \dfrac{1}{n} E(S_n) = \mu$
- $\mathrm{var}(S_n) = \mathrm{var}(X_1 + \cdots + X_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n) = n\sigma^2$
- $\mathrm{var}(\overline{X}) = \dfrac{1}{n^2} \mathrm{var}(S_n) = \dfrac{\sigma^2}{n}$

Nota: O desvio padrão de S_n é $\sqrt{n}\sigma$ e o desvio padrão de $\overline{X}$ é $\dfrac{\sigma}{\sqrt{n}}$.

Soma de algumas variáveis aleatórias

O seguinte resultado carateriza a soma de v.a.'s de alguns tipos especiais. No que se segue, $X_1, \ldots, X_n$ são v.a.'s independentes. Tem-se:

- Se $X_i \frown Ber(p)$, então $S_n \frown Bi(n, p)$
- Se $X_i \frown Poi(\lambda)$, então $S_n \frown Poi(n\lambda)$
- Se $X_i \frown N(\mu, \sigma)$, então $S_n \frown N(n\mu, \sqrt{n}\sigma)$

Sendo X uma v.a. com distribuição normal, então, como já referimos, qualquer variável da forma $Y = a + bX$ também é normal. Tem-se, então o seguinte resultado para a média amostral $\overline{X} = \frac{1}{n} S_n$ relativa a uma amostra aleatória de uma distribuição normal:

$$X_i \frown N(\mu, \sigma) \Rightarrow \overline{X} \frown N(\mu, \sigma/\sqrt{n})$$

$$\Rightarrow Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \frown N(0, 1)$$

Teorema Limite Central

Acabámos de ver que a média $\overline{X}$ de uma amostra aleatória $X_1, \ldots, X_n$ da distribuição normal $N(\mu, \sigma)$ tem uma distribuição normal $N(\mu, \sigma/\sqrt{n})$ ou, equivalentemente, que a v.a. $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ tem distribuição $N(0, 1)$.
Apresentamos agora um teorema que contém um dos resultados mais interessantes da Teoria de Probabilidade, o qual pode ser visto como uma extensão destes resultados.

Teorema (Teorema Limite Central – TLC)

Sejam $X_1, \ldots, X_n$ v.a.'s i.i.d. com valor médio μ e variância finita σ^2 e sejam S_n e $\overline{X}$ a soma e a média dessas v.a.'s, respetivamente. Então, se n for suficientemente grande, a v.a.

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

tem aproximadamente uma distribuição $N(0, 1)$.

Por outras palavras, o TLC afirma que, para n suficientemente grande, S_n é aproximadamente $N(n\mu, \sigma\sqrt{n})$ e $\overline{X}$ aproximadamente $N(\mu, \sigma/\sqrt{n})$. Escrevemos, então

$$S_n \overset{\text{apox.}}{\frown} N(n\mu, \sigma\sqrt{n}), \quad \overline{X} \overset{\text{apox.}}{\frown} N(\mu, \sigma/\sqrt{n})$$

Notas

1. Mais precisamente, o teorema estabelece que, nas condições dadas, se F_n for a função de distribuição da v.a. $\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$, então $\lim_{n\to\infty} F_n(x) = \Phi(x)$, onde Φ é a função de distribuição da normal $N(0, 1)$.
2. Na prática, se as v.a.'s não forem muito assimétricas, basta que $n > 30$ para obtermos uma aproximação razoável.
3. O Teorema Limite Central (aqui apresentado na sua forma clássica) admite generalizações, no sentido de ser possível enfraquecer as suas hipóteses; por exemplo, sob certas condições, a convergência para a normal mantém-se, mesmo que as v.a.'s não sejam identicamente distribuídas nem mesmo independentes; é este um dos motivos pelos quais o modelo normal é tão preponderante na modelação de fenómenos, uma vez que muitos deles podem ser vistos como "soma" de contribuições aleatórias.

Aplicações do Teorema Limite Central

Aproximação da binomial pela normal

Sendo $X_i \frown Ber(p)$ (X_i i.i.d.), tem-se $S_n \frown Bi(n, p)$. Então, para n suficientemente grande, tem-se, aplicando o TLC,

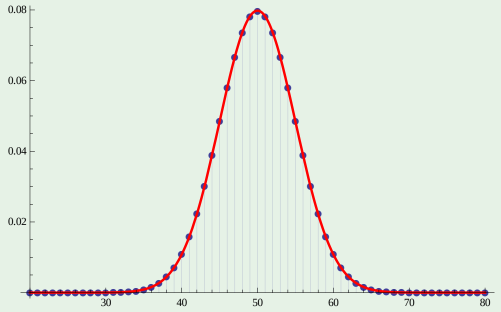
$$X \frown Bi(n, p) \Rightarrow X \overset{\text{apox.}}{\frown} N(np, \sqrt{np(1-p)})$$

$$\Rightarrow Z = \frac{X - np}{\sqrt{np(1-p)}} \overset{\text{apox.}}{\frown} N(0, 1)$$

uma vez que, se $X_i \frown Ber(p)$, se tem $E(X_i) = p$ e $\text{var}(X_i) = p(1-p)$.
Nota: Geralmente, a aproximação só é usada se $\min\{np, n(1-p)\} \geq 5$.

Exemplo

Na figura seguinte, apresentam-se, em sobreposição no mesmo gráfico, a f.m.p. de uma distribuição $Bi(100, 0.5)$ e a f.d.p. da $N(50, 5)$ (note-se que, para $n = 100$ e $p = 0.5$, vem $\sqrt{np(1-p)} = \sqrt{25} = 5$), onde é patente o ajuste.



f.m.p da $Bi \sim (100, 0.5)$ (azul) e f.d.p. da $N(50, 5)$ (vermelho).

Aplicações do Teorema Limite Central

Aproximação da Poisson pela normal

Sendo $X_i \frown Poi(\lambda/n)$ (X_i i.i.d.), tem-se $S_n \frown Poi(\lambda)$. Assim, temos

$$X \frown Poi(\lambda) \Rightarrow X \overset{\text{apox.}}{\frown} N(\lambda, \sqrt{\lambda})$$

$$\Rightarrow Z = \frac{X - \lambda}{\sqrt{\lambda}} \overset{\text{apox.}}{\frown} N(0, 1)$$

já que, se $X_i \frown Poi(\lambda/n)$, se tem $E(X_i) = \lambda/n$ e $\text{var}(X_i) = \lambda/n$.

Nota: Em geral, só se se usa a aproximação quando $\lambda > 5$.

De realçar que a aproximação é sempre feita por uma normal cuja média e desvio padrão são iguais aos da distribuição que estamos a aproximar.

Lei dos Grandes Números

Apresentamos agora outro resultado muito importante da Teoria da Probabilidade, conhecido por Lei dos Grandes Números.

Teorema (Lei dos Grandes Números - LGN)

Seja $X_1, X_2, \ldots$ uma sequência de v.a.'s i.i.d., cada uma delas com valor médio $E(X_i) = \mu$. Para $n \in \mathbb{N}$, seja $\overline{X} = \frac{X_1 + \cdots + X_n}{n}$ a média de $X_1, \ldots, X_n$. Então, para qualquer $\epsilon > 0$, tem-se

$$\lim_{n \to \infty} P(\mu - \epsilon < \overline{X} < \mu + \epsilon) = 1.$$

Dizemos que $\overline{X}$ converge em probabilidade para μ.

Aplicação da Lei dos Grandes Números

Consideremos uma dada experiência aleatória. Seja A uma acontecimento fixo (associado a essa experiência) e seja $p = P(A)$ a probabilidade de ocorrência desse acontecimento.Suponhamos que efetuamos uma sequência de repetições independentes dessa mesma experiência e consideremos as v.a.'s X_i definidas por:

$$X_i = \begin{cases} 1, & \text{se } A \text{ ocorre na } i\text{-ésima repetição} \\ 0, & \text{se } A \text{ não ocorre na } i\text{-ésima repetição} \end{cases}$$

Então, X_i são i.i.d. com distribuição $Ber(p)$, tendo-se $\mu = E(X_i) = p$.
Neste caso, a v.a. $\overline{X} = \frac{X_1 + \cdots + X_n}{n}$ corresponde à frequência relativa da ocorrência de A nas n repetições da experiência, $f_n(A)$ (ou seja, à proporção de ocorrências de A).
A LGN estabelece, portanto, que $f_n(A)$ converge (em probabilidade) para $p = P(A)$.
Assim sendo, faz sentido aproximar $P(A)$ pela frequência relativa da sua ocorrência, quando n for suficientemente grande.