

Xenaro García Suárez e Francisco Rodríguez Álvarez

Introducción aos métodos estatísticos e numéricos



Introducción aos métodos estatísticos e numéricos.

ÍNDICE

INTRODUCCIÓN.....	5
INTRODUCCIÓN Á PROBABILIDADE.....	6
PROBABILIDADE CONDICIONADA.....	16
CADEAS DE MARKOV.....	32
DISTRIBUCIÓN DE PROBABILIDADE.....	45
MOSTRAXE.....	78
ESTIMACIÓN.....	91
PROBAS DE HIPÓTESES.....	106
SERIES DE TEMPO.....	121
PROGRAMACIÓN LINEAR.....	138
ERROS. INTERPOLACIÓN. INTEGRACIÓN APROXIMADA.....	157
RESOLUCIÓN DE ECUACIONES.....	185

INTRODUCCIÓN

Na presentación da primeira edición deste manual (1997) dicíamos que un dos seus obxectivos era servir de apoio aos estudantes de segundo curso de bacharelato para a asignatura optativa *Métodos estatísticos e numéricos*. Mais non só ía dirixido a eles, senón tamén a todo aquel interesado nas ciencias da natureza ou nas sociais, por obriga ou curiosidade, que pretendese coñecer os rudimentos da análise de datos, da resolución efectiva de ecuacións e sistemas, e do pensamento probabilístico e estatístico, coñecementos indispensábeis na actualidade para entender un pouco mellor o noso planeta e o Universo.

Seguimos a crer que unha primeira aproximación aos métodos de resolución de certos supostos numéricos e probabilísticos, que pode ser a derradeira para moitos, pode facerse sen a necesidade de presentar en cada caso as demostracións dos procedementos propostos, e esa é a exposición que facemos no libro cando, poñamos por caso, se trata de acoutar un erro ou de resolver un problema de programación linear con máis de dúas incógnitas mediante a folla de cálculo, coa esperanza de que a comprobación na práctica da potencia dos métodos, sirva de acicate para que algúns se animen a profundar nas teorías que permiten un maior afinamento nas solucións.

Ao abordarmos esta segunda edición tivemos presente a nosa experiencia utilizando o manual durante máis dunha década. Actualizáronse algúns contidos, introducíronse outros e elimináronse certas cuestións que nos pareceron superfluas para un primeiro achegamento aos temas propostos.

No relativo a probabilidade e estatística preséntanse os exercicios resoltos indispensábeis, amosando determinadas estratexias que servirán de guía para outros moitos que se propoñen. Porén, pareceunos conveniente ser xenerosos na inclusión de exercicios resoltos de *Cálculo numérico*. Por unha banda, as referencias que existen a un nivel iniciático neste campo son moito máis escasas, e, ademais, até onde chegan as nosas pescudas, desde logo nada científicas, na impartición da asignatura *Métodos estatísticos e numéricos* faise moito máis fincapé na estatística que na análise numérica, tratándose esta última, polo xeral, cun nesgo demasiado teórico, ou canto menos pouco práctico en canto á resolución efectiva. Neste sentido, a utilización do computador, en concreto da folla de cálculo, parécenos imprescindible para executar dun xeito cómodo e eficaz os algoritmos presentados, e por iso a introducimos para amosar a súa potencia ao poñela ao servizo de moitas estratexias de resolución, que sen a súa axuda se farían tremendamente tediosas.

Queremos tamén deixar constancia de que para ilustrar moitos supostos se utilizaron datos reais, tirados fundamentalmente das páxinas do *IGE*, do *INE* e de *EUROSTAT*. En mans do lector está acudir no momento preciso a estas páxinas para dispor de datos actualizados; ao falarmos da taxa de paro debe terse presente a do último trimestre dispoñíbel, e o mesmo acontece coas piramides de idade, as taxas de fecundidade ou a esperanza de vida. O idóneo será contar cos últimos datos, comezando polos da contorna máis próxima. Neste sentido, *Internet*, é unha ferramenta que ofrece múltiples posibilidades á hora da aplicación dos modelos teóricos a parcelas da realidade.

1. INTRODUCCIÓN Á PROBABILIDADE

*“Eu coñezo o segredo dos dados,
e por iso son destro nos números.”
“Historia de Nala”. Mahabharata.*

Experimentos aleatorios

Determinadas situacións ás que estamos afeitos, como cubrir unha *primitiva*, xogar a cara ou cruz, facer predicións sobre a hora de saída dun voo, sobre o sexo do noso próximo irmán, fillo ou neto (antes de que o descubra a ecografía), sobre determinado proceso electoral ou sobre o tempo da fin de semana, teñen en común a imposibilidade de predicir cal vai ser o seu resultado preciso. Aos experimentos deste tipo, obxecto de estudo da teoría de probabilidade, chamáremoslle aleatorios.

A análise dun experimento aleatorio pódese abordar desde dous puntos de vista. O primeiro é observar o seu comportamento através dun período o suficientemente longo. Para isto cómpre que o experimento se poida repetir nas mesmas condicións un grande número de veces, de xeito que canto maior sexa este número máis información obteremos sobre os posíbeis resultados. O segundo é utilizar un modelo matemático que nos permita facer un estudo racional e sistemático. Na práctica, o máis frecuente é utilizar unha combinación de ambos métodos mediante a interacción de hipóteses teóricas xunto ás confirmacións experimentais.

Conxunto de resultados

A cada experimento aleatorio asóciase un conxunto formado por todos os resultados posíbeis que poida presentar. Este conxunto vai ser o marco de referencia e recibe o nome de *espazo das mostras* ou *conxunto de resultados*. Denotarémolo mediante a letra Ω . Se o experimento, poñamos por caso, é botar unha moeda ao ar, o conxunto de resultados será: $\Omega = \{cara, cruz\}$. Se é botar un dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Determinados experimentos, como escoller ao azar un día do ano ou un alumno de bacharelato de Galicia, darán lugar a conxuntos de resultados cun número elevado de elementos. Nesta caso: 365 e 32181, respectivamente.

Mesmo experimentos moi sinxelos poden xerar conxuntos de infinitos resultados. Se botamos unha moeda ao ar até que apareza cara, e contamos todas as tiradas, temos: $\Omega = \{1, 2, 3, \dots, n, \dots\}$, que é un conxunto infinito numerábel. Se o experimento consiste en contar o tempo de espera dun autobús que pasa nun intervalo

máximo de dez minutos, o conxunto de resultados sería: $\Omega = [0,10]$, que é un conxunto infinito non numerábel.

Un conxunto de resultados finito ou infinito numerábel chámase *discreto*. Un espazo infinito non numerábel diremos que é *continuo*.

Exercicios.

1. Descríbase o conxunto de resultados dos seguintes experimentos aleatorios:

- Botar unha moeda tres veces.
- Botar dous dados.
- Botar simultaneamente unha moeda e un dado.
- Cubrir unha quiniela de 15 partidos.
- Anotar o sexo da primeira persoa que pase.
- Escoller unha carta dunha baralla de 40.
- Tirar dúas cartas, unha a continuación de outra.
- Lanzar un dardo sobre unha diana de 15 cm de raio.
- Escoller ao chou un comité de tres estudantes dunha clase de 20.
- Escoller ao azar delegado e subdelegado nunha clase de 20.
- Comprar dous décimos de lotería.
- Sentar cinco persoas nun banco.
- Botar un dado até que apareza un 5.

2. Para calcular o número de resultados dalgúns dos experimentos do exemplo anterior, utilízanse técnicas para contar como diagramas de árbore ou coñecementos de combinatoria. Facendo uso delas cálculese o número de resultados posibles doutros experimentos aleatorios do mesmo tipo.

Acontecementos

Un subconxunto do conxunto de resultados recibe o nome de *acontecemento*. Se nun experimento aleatorio se dá un resultado r que pertence ao acontecemento A , dicimos que ocorre A . Utilizamos o símbolo de pertenza, $\in$, para indicar esta situación, e escribiremos $r \in A$.

Se consideramos o exemplo anterior do lanzamento do dado, podemos atopar, poñamos por caso, os acontecementos:

$$\begin{array}{ll} A = \text{"Aparece número par"} & A = \{2,4,6\} \\ B = \text{"Aparece o número 5"} & B = \{5\} \\ C = \text{"Aparece un número menor que 5"} & C = \{1,2,3,4\} \end{array}$$

Se o conxunto de resultados do experimento aleatorio ten n elementos entón terá 2^n subconxuntos, e xa que logo haberá 2^n acontecementos.

Un acontecemento A está *incluído* noutro B se sempre que ocorra A ocorre B ou, doutro xeito, todos os resultados de A sono tamén de B . Os acontecementos:

$D = \text{"Aparece número maior ou igual que 3"} \quad D = \{3, 4, 5, 6\}$

$E = \text{"Aparece número múltiplo de 3"} \quad E = \{3, 6\}$

verifican $E \subset D$. Utilizamos o signo de inclusión, $\subset$, para escribir esta relación.

Os acontecementos que constan dun único resultado chámanse *elementais*. O acontecemento $\emptyset$, que non ten ningún resultado, chámase *imposíbel* posto que non ocorrerá nunca. O propio Ω , como subconxunto de si mesmo, chamarase acontecemento *seguro*, xa que sempre ocorrerá.

Exercicios.

2. Escribíanse enunciados que correspondan a acontecementos dos experimentos aleatorios do exercicio 1 e áchense os resultados de que consta cada un deles.
3. Comproba para $n = 2$ e $n = 3$, que se Ω ten n elementos o seu número de subconxuntos é 2^n .

Operacións con acontecementos

As operacións que se realizan entre conxuntos teñen a súa interpretación con acontecementos. Así, definimos a *reunión* de A e B como o acontecemento que ocorre cando ben ocorre A ou ben ocorre B . Este novo acontecemento está formado por todos os resultados de A e todos os de B , e désígnase por $A \cup B$.

Se $A = \{4, 5, 6\}$ e $B = \{2, 4, 6\}$ entón $A \cup B = \{2, 4, 5, 6\}$.

A *intersección* de dous acontecementos A e B é un novo acontecemento que ocorre cando ambos os dous ocorren simultaneamente. Désígnase por $A \cap B$, e está formado polos elementos comúns a A e a B .

Se A e B son os acontecementos anteriores entón $A \cap B = \{4, 6\}$.

Dous acontecementos que teñan como intersección o acontecemento imposíbel chámanse *incompatíbeis*. Os acontecementos $C = \{1, 3, 5\}$ e $B = \{2, 4, 6\}$ verifican $C \cap B = \emptyset$; é dicir, son incompatíbeis.

O acontecemento *contrario* de A é o que ocorre se non ocorre A . Constará de todos os elementos do espazo das mostras que non pertencen a A . Utilizando os acontecementos anteriores, temos: $A^c = \{1, 2, 3\}$ e $B^c = \{1, 3, 5\}$.

Exercicios

4. Exprésense mediante un diagrama de Venn os acontecementos que resultan de realizar as operacións do parágrafo anterior.
5. Sexan A , B e C tres acontecementos asociados a un experimento aleatorio. Descríbanse mediante operacións con eles os acontecementos que a continuación se relacionan, facendo o correspondente diagrama de Venn.
 - Polo menos ocorre un dos tres
 - Ocorren A e B pero non C .
 - Unicamente ocorre A .

- Ocorre unicamente un dos tres acontecementos.
- Ocorren os tres acontecementos.
- Ocorren exactamente dous acontecementos.
- Ocorren menos de dous.

Álgebra de acontecementos

Ao abordar o estudo dun experimento aleatorio, como xa vimos, é preciso en primeiro lugar definir o conxunto de resultados e posteriormente os acontecementos que imos estudar. Se o espazo das mostras é finito, o conxunto de acontecementos estará formado por todos os subconxuntos del. Se o espazo é continuo, ás veces cómpre reducir os acontecementos a determinados subconxuntos do conxunto de resultados. Utilizaremos entón unha familia de acontecementos que teña as seguintes propiedades:

1. O acontecemento imposible pertence á familia.
2. Se un acontecemento pertence á familia o seu contrario tamén.
3. Se unha sucesión de acontecementos pertence á familia entón a súa unión tamén pertence.

Unha familia de acontecementos que verifique estas tres propiedades dise que é un *álgebra de acontecementos*. No caso finito a álgebra coincide co conxunto de todos os subconxuntos do espazo das mostras, e denótase por $\mathcal{P}(\Omega)$.

Probabilidade

Na linguaxe ordinaria a probabilidade ven sendo a medida do grao de incerteza dun acontecemento. Esta medida faise moitas veces en termos cualitativos, e así escoitamos con frecuencia comentarios do tipo: “seguro que hoxe chove”, “case seguro que aprobo o exame”, “e moi difícil que o equipo do que son seareiro gañe a liga”.

En ocasións, poñamos por caso, nos medios escritos e audiovisuais as predicións aparecen con máis precisión: “A probabilidade de que os nados na actualidade nun determinado país vivan máis de 75 anos é do 80%”, “a probabilidade de acertar a *primitiva* é practicamente nula”, “a probabilidade de padecer cancro de pulmón incrementase nun 50% nos fumadores”.

Ao tentar facer un estudo de determinado experimento aleatorio precisamos, ademais dun conxunto de resultados e dunha familia de acontecementos, unha estimación do grao de seguridade de que determinado acontecemento ocorra ao realizar o experimento. O propio carácter aleatorio do experimento semella indicarnos que non existen regras que garantan o éxito seguro da estimación, porén, podemos establecer unha serie de condicións que faciliten a nosa tarefa:

- A estimación debe realizarse mediante unha medida numérica.
- O número resultante desta medida debe expresar unha fracción ou unha porcentaxe da seguridade.
- De acordo co anterior ao acontecemento seguro débesele asignar o 100% ou o número 1.
- Da mesma maneira ao acontecemento imposible asignaráselle o cero.

- Os restantes acontecementos asociados ao experimento terán asignado un número comprendido entre cero e un. A designación deste número farase a partir da experiencia ou pola observación de determinadas propiedades.
- Se dous ou máis acontecementos son incompatíbeis, o número asignado a súa reunión debe ser a suma dos adxudicados a cada un deles.

Xeitos de asignar probabilidade

Tomando como base as regras enunciadas no apartado anterior, apareceron ao longo da historia distintas teorías que resumen os esforzos encamiñados a dar unha definición precisa e consistente de probabilidade. Aparecen así diversas escolas con diferentes puntos de vista sobre o concepto e a súa aplicación.

Pierre Laplace escribiu en 1812 o tratado *Teoría analítica das probabilidades*, un a modo de compendio de todos os coñecementos sobre o tema até daquela, case todos eles relacionados exclusivamente cos xogos de azar.

Laplace define a probabilidade dun acontecemento como o cociente entre o número de casos favorábeis, é dicir, o número de veces que o acontecemento pode aparecer nunha proba determinada, e o número de casos posíbeis do experimento. Isto presupón que todos os acontecementos do experimento deben ser igualmente probábeis, permitindo o cálculo efectivo de probabilidades en problemas sinxelos, mais amosándose insuficiente ao intentar adxudicar probabilidades en situacións máis complexas.

En 1919, Richard von Mises sistematiza resultados de investigacións anteriores e define a probabilidade dun acontecemento como o límite da súa frecuencia relativa. Esta definición é empírica e *a posteriori*, xa que para adxudicar probabilidades hai que realizar o experimento moitas veces. En contraposición, a concepción de Laplace era teórica e *a priori*. Aínda que a formulación de von Mises ten certa utilidade, adoece de problemas formais debido á falta de rigor lóxico na formulación do proceso infinito de paso ao límite.

O matemático ruso Andrei Kolmogorov deu en 1933 unha definición formal de probabilidade que recolle en tres axiomas as regras mínimas necesarias para poder introducir con precisión e rigor unha medida do grao de posibilidade de realización dun acontecemento en calquera tipo de experimento aleatorio; dá unha definición axiomática en termos estritamente matemáticos, eliminando así as ambigüidades das teorías anteriores.

Exercicios

6. Discútase a posibilidade de asignar probabilidades, *a priori* ou *a posteriori*, a cada un dos seguintes experimentos aleatorios:
 - Botar unha moeda ao ar.
 - Botar unha chincheta ao ar.
 - Dunha caixa que contén bólas brancas e negras escoller unha.
 - Botar un dado trucado con dobre posibilidade de saír número par.
 - Chamar o ascensor dende o terceiro andar dun edificio de 10; sobe ou baixa?
7. Discútanse as seguintes cuestións:

- Unha muller embarazada ten catro fillas. Que é máis probábel, que teña un neno ou unha nena?
- Un xogador de ruleta observa que nas últimas cen xogadas saíu número par. Terá vantaxe se aposta a impar?
- Un director de empresa ten que escoller de entre 100 empregados un grupo para unha determinada sección. Que é máis probábel, ser elixido se o grupo é de 40 ou se é de 60?
- Se o 90% dos accidentes se producen a velocidade menor que 100 km/h, será menos probábel ter un accidente se circulamos a 150 km/h?

Espazos equiprobábeis

Se nun espazo das mostras todos os acontecementos elementais teñen a mesma probabilidade, díse que é un espazo equiprobábel. No caso dun conxunto de resultados finito con n elementos, a probabilidade de cada acontecemento elemental será $1/n$. Se un acontecemento ten un número k de elementos, daquela a súa probabilidade será k/n . Isto non deixa de ser unha aplicación da regra de *Laplace* comentada no parágrafo anterior. Cando falamos de elixir un resultado dun experimento aleatorio “ao azar” referímonos a unha experiencia que se realiza nun espazo equiprobábel. Os xogos de azar máis comúns corresponden a espazos equiprobábeis; así, a probabilidade de sacar cruz ao botar unha moeda será $1/2$, e a de que apareza un 5 ao botar un dado será $1/6$.

Exercicios

8. Calcúlese a probabilidade de que:
 - Se botamos un dado apareza número par.
 - Ao escoller unha carta dunha baralla de 40 apareza o as de ouros.
 - Ao escoller unha carta dunha baralla de 40 apareza unha espada.
 - Ao escoller dúas cartas aparezan dous ouros.
 - Ao escoller ao azar tres alumnos nunha clase de 20 mozas e 10 mozos:
 - As tres sexan mulleres.
 - Os tres sexan homes
 - Dous sexan homes e unha muller.
 - Gañar o primeiro premio da lotaría, mercando un décimo.
 - Elixir ao azar un número do 1 ao 50 e que resulte múltiplo de tres.
 - Botar dous dados e que sumen sete.

Probabilidade e frecuencia relativa

Na maioría dos casos nos que é preciso aplicar a teoría de probabilidade ás máis diversas situacións, non é doado asignar probabilidades *a priori*. Realízase entón unha sucesión de probas repetidas do experimento aleatorio e estúdase o comportamento do

acontecemento ao que se quere asignar probabilidade, calculando a súa frecuencia relativa: cociente entre o número de veces que aparece e o número de probas realizadas. Conforme aumentamos o número de probas a frecuencia relativa tende a se estabilizar arredor dun número, que tomamos como probabilidade do acontecemento en cuestión. Isto é o que se coñece como *Lei dos grandes números*; un resultado fundamental da teoría de probabilidades que se demostra unha vez reinterpretados os conceptos na teoría formal que abordaremos no seguinte apartado.

Exercicios

9. Bótese ao ar unha moeda 10, 50, 100, 200,... veces e calcúlese a frecuencia relativa do acontecemento “aparece cara”. Compróbase que as frecuencias relativas se aproximan a $\frac{1}{2}$.
10. Mediante repeticións sucesivas calcúlese a probabilidade de que ao botar unha chincheta ao ar quede coa punta cara arriba.

Definición axiomática de probabilidade

Unha aplicación, P , definida nunha álgebra de acontecementos con valores nos números reais, dirase que é unha probabilidade se verifica:

Axioma 1. $P(A) \geq 0$, para todo elemento A da álgebra de acontecementos.

Axioma 2. $P(\Omega) = 1$

Axioma 3. Se temos unha sucesión de acontecementos $A_1, A_2, \dots, A_n, \dots$, incompatíbeis dous a dous, a probabilidade da suma da unión é igual á suma das probabilidades de cada un deles.

$$P(A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots$$

Como consecuencia dos axiomas, dedúcese:

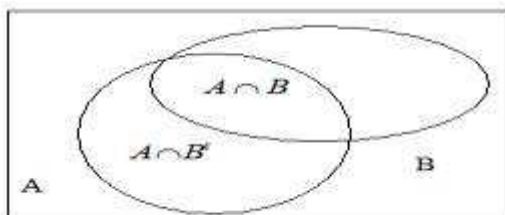
-Teorema 1. A probabilidade do acontecemento imposible é cero.

$$P(\emptyset) = 0,$$

xa que $P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset)$ [polo axioma 1], de onde $P(\emptyset) = 0$

-Teorema 2. Calquera que sexan os acontecementos A e B , verifícase

$$P(A) = P(A \cap B) + P(A \cap B^c)$$



Teñamos en conta que $A = (A \cap B) \cup (A \cap B^c)$ e que os dous acontecementos reunidos son incompatíbeis, e entón aplicamos o axioma 3.

-Teorema 3. Se o acontecemento A está incluído no acontecemento B , entón a $P(A)$ menor ou igual que $P(B)$.

$$A \subset B \Rightarrow P(A) \leq P(B)$$

-Teorema 4. As probabilidades dun acontecemento e o seu contrario suman un.

$$P(A) + P(A^c) = 1$$

- Teorema 5. A probabilidade da reunión de dous acontecementos é a suma das súas probabilidades menos a probabilidade da intersección.

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

Exercicios

11. Demóstranse os teoremas anteriores.
12. O conxunto de resultados dun experimento aleatorio é : $\Omega = \{a, b, c\}$.
Estúdese se as seguintes aplicacións son probabilidades.
 - a) $P(A) = 1/2, P(B) = 1/3, P(C) = 1/5$.
 - b) $P(A) = 1/3, P(B) = 2/3, P(C) = 0$.
 - c) $P(A) = 2/5, P(B) = 1/5, P(C) = 2/5$.
13. Sexan A e B dous acontecementos que verifican: $P(A) = 1/2, P(B) = 1/3, P(A \cap B) = 1/4$. Calcula a probabilidade dos seguintes acontecementos:
 $P(A^c), P(B^c), P(A \cup B), P(A \cap B^c), P(A^c \cap B^c)$
14. Un dado ten forma de tetraedro regular coas caras marcadas con números do 1 ao 4. Calcúlese a probabilidade de que ao botalo apareza: a) número impar, b) o número 3.
15. Botamos ao ar unha moeda tres veces seguidas. Calcula a probabilidade de que: a) aparezan tres caras, b) apareza polo menos unha cara, c) aparezan exactamente dúas caras.
16. O 2% dos artigos producidos por unha fábrica son defectuosos. Se escollemos aleatoriamente 10 artigos dun grupo de 1000, calcúlese a probabilidade de que: a) dous sexan defectuosos, b) ningún sexa defectuoso.
17. Un inspector de facenda investiga ao azar 4 declaracións da renda dun grupo de 40 entre as que se atopan 10 fraudulentas. Calcúlese a probabilidade de que : a) as catro declaracións investigadas sexan correctas, b) se de seguido investiga outras catro, as catro sexan fraudulentas.
18. O 60% dos preguntados nunha enquisa teñen menos de 20 anos. O 40% son fumadores. O 20% son fumadores e teñen menos de 20 anos. Se escollemos aleatoriamente un individuo entre os enquisados, calcúlese a probabilidade de que: a) teña 20 anos ou máis e fume, b) teña 20 anos ou máis e non fume.
19. Sácanse á venda 100 sobres sorpresa dos cales 20 teñan premio. Calcúlese a probabilidade de sacar premio se: a) se compra un único sobre, b) se compran dous sobres, c) se compran tres.

20. Tres persoas participan nun concurso de saltos. A ten o dobre de probabilidade de gañar que B, e este ten o triplo que C. Cálculase a probabilidade de que: a) gañe C, b) que non gañe B.
21. 10 fichas numeradas do 1 ao 10 esténdense aliñadas ao azar sobre unha mesa. Cálculase a probabilidade de que a) resulten ordenadas de menor a maior, b) o número 10 quede situado no último lugar.
22. Un dado ten dúas caras brancas, dúas negras, unha verde e unha azul. Se botamos o dado, calcúlese a probabilidade de que a) apareza a cara negra, b) apareza cara branca ou negra.
23. Unha tartaruga circula sobre os lados numerados do 1 ao 5 dun pentágono regular. Se nun momento dado se para, calcúlese a probabilidade de que sexa no número 3.
24. Métense dous obxectos ao azar en cinco caixas numeradas. Cálculase a probabilidade de que a caixa número dous quede baleira se: a) se mete un só obxecto en cada caixa, b) se se mete máis dun obxecto en cada caixa.
25. Unha diana ten forma de círculo con 20 cm de raio. Se sinalamos ao azar un punto calquera dela, calcúlese a probabilidade de que quede máis preto do centro que do exterior.
26. Unha persoa participa nun concurso no que ten que asociar tres fotos cos nomes de tres personaxes famosos. Se responde ao azar, cal é a probabilidade de que asocie correctamente cada foto con cada nome?
27. Un dado está trucado de xeito que o número 5 ten triplo de probabilidade de saír que calquera dos restantes. Cálculase a probabilidade de que ao botar o dado non apareza o 5.
28. Nunha empresa traballan 20 mulleres e 30 homes. Escóllese un comité de 5 persoas aleatoriamente. Cálculase a probabilidade de que: a) as cinco sexan mulleres?, b) as cinco sexan homes?, c) dous sexan homes e tres mulleres?, d) a persoa "T" forme parte do comité?
29. A un congreso asisten 200 persoas das cales 85 falan francés, 55 inglés e 55 alemán. Dos asistentes, 15 falan francés e inglés, 30 francés e alemán, e 20 inglés e alemán. Por último, 10 persoas falan os tres idiomas. Cálculase a probabilidade de que, escollido ao azar un asistente ao congreso, non fale ningún dos tres idiomas.
30. Cálculase a probabilidade de que, sinalando ao azar dous puntos do interior dun segmento, as tres partes resultantes formen un triángulo.
31. Un alumno ten que realizar un exame sobre 10 temas pero só sabe 6. Se o exame consta de 2 temas elixidos por sorteo, calcúlese a probabilidade de que conteste polo menos un deles.
32. No departamento de control de calidade dunha empresa estímase que de cada 1000 unidades fabricadas, 2 resultan defectuosas. Se escollemos ás toas dous artigos dun lote de 200, calcúlese a probabilidade de que: a) ningún dos dous sexa defectuoso, b) polo menos un dos dous sexa defectuoso.
33. Unha rapaza ten dúas avoas. Para visitar á primeira ten que coller un autobús que chega á parada ás horas en punto. Se quere ver á segunda, ten que coller un autobús que pasa pola mesma parada ás 8 e cuarto. Se sae da casa aleatoriamente e colle o primeiro autobús que pasa, cal das dúas avoas é a máis visitada?
34. Tiramos ao chou dúas cartas dunha baralla de corenta. Cálculase a probabilidade de que: a) escollamos dous ouros, b) escollamos unha espada e unha copa, c) unha das dúas cartas escollidas sexa un as, d) non se escolla ningunha figura.

34. Escóllemos aleatoriamente tres bolas dun conxunto de 30 das que 10 son negras. Calcúlese a probabilidade de que: a) as tres sexan negras, b) unha polo menos sexa negra, c) unha exactamente sexa negra.
35. Dous amigos deciden verse nunha cafetería entre as cinco e as seis para ir ao cine. Ningún dos dous esperará máis de 10 minutos polo outro. Calcúlese a probabilidade de que se atopen.
36. Para cubrir un boleto de lotaría primitiva dunha aposta hai que sinalar seis números dun total de 49. Calcúlese a probabilidade de ter 6, 5, 4 ou 3 acertos.
37. Formamos aleatoriamente números de tres cifras distintas escollidas entre 1, 2, 3, 4, 5 e 6. Calcúlese a probabilidade de que o número escollido sexa par.
38. Sete libros fican colocados nun andel. Se os retiramos e volvemos colocalos ao azar, cal é a probabilidade de que resulten colocados na posición na que estaban?
39. O 30% dos habitantes dunha cidade mercan o xornal A, o 20% o B e o 10% os dous. Se escollemos un habitante ao chou, cal é a probabilidade de que non merque ningún xornal?
40. Dous coches chegan pola mesma rúa a unha praza que ten tres saídas. Calcúlese a probabilidade de que os dous coches: a) saian pola mesma rúa, b) saian por rúas distintas.
41. Temos un cadrado inscrito nun círculo de raio r . Se escollemos un punto do círculo ao azar, calcúlese a probabilidade de que quede dentro do cadrado.
42. O adestrador dun equipo de fútbol distribúe ao chou as camnietas dos seus xogadores. Calcúlese a probabilidade de que tres xogadores leven a camiseta correcta.

2. PROBABILIDADE CONDICIONADA

Murmurábame e movíame a comparar esta alteración indefiníbel que se asemellaba en min a un cambio bastante brusco de certas probabilidades mentais.

Paul Valéry, *Monsieur Teste*

Exemplo 1

Un xogador desexa saber que probabilidade ten de sacar unha suma maior ou igual a 8 ao botar dous dados. Ao analizarmos o experimento, decatámonos doadamente de que en 15 resultados, dos 36 posíbeis, a suma é maior ou igual que 8.

Temos: $A = \text{"sacar suma} \geq 8" \Rightarrow P(A) = 15/36 = 0,41$.

Supoñamos agora que o xogador sabe que a suma dos dous dados é impar. Daquela os resultados posíbeis no canto de 36 serán 18, e destes unicamente 6 presentan unha suma maior ou igual que 8. Con esta nova información, temos que $P(A) = 6/18 = 0,33$; inferior á calculada ao principio.

Vemos que a información adicional sobre un determinado experimento aleatorio modifica substancialmente a súa probabilidade, e neste caso mesmo dun xeito aparentemente paradoxal, pois o feito de posuír maior información diminúe a probabilidade de que este ocorra.

Dicimos que a probabilidade calculada en segundo lugar está condicionada polo acontecemento “a suma dos dados é impar”. Se denotamos por B este acontecemento, a probabilidade en cuestión adóitase escribir da forma $P(A/B)$.

A información que asegura a ocorrencia do acontecemento B limita o conxunto de resultados Ω ao conxunto B , de xeito que debemos de excluír de A todos os resultados que non están en B , quedándonos co acontecemento intersección de A e B .

A probabilidade de que ocorra A sabendo que ocorreu B , virá dada pola expresión:

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

que chamamos *probabilidade de A condicionada por B*. Cómpre observar que esta probabilidade ten sentido sempre que $P(B) > 0$.

Notemos que a probabilidade concionada cumpre os axiomas da probabilidade xa que, evidentemente, $P(A/B) \geq 0$ e $P(\Omega/B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.

Ademais, se $A_1, A_2, \dots, A_n$ son incompatíbeis, entón:

$$\begin{aligned} P((A_1 \cup A_2 \cup \dots \cup A_n)/B) &= \frac{P((A_1 \cup A_2 \cup \dots \cup A_n) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup \dots \cup (A_n \cap B))}{P(B)} = \\ &= \frac{P(A_1 \cap B)}{P(B)} + \dots + \frac{P(A_n \cap B)}{P(B)} = P(A_1/B) + \dots + P(A_n/B) \end{aligned}$$

Exemplo 2

No primeiro curso de Económicas o 50% do alumnado aproba Estatística e o 40% Macroeconomía. A probabilidade de aprobar *Macro* tendo aprobada estatística é do 80%. Calcúlese a porcentaxe de alumnos que aproban as dúas asignaturas. Próbese que a metade dos alumnos suspende as dúas asignaturas.

Sexan A = “aprobar estatística” e B = “aprobar macro”; temos:

$$P(A) = 0,5, P(B) = 0,4, P(B/A) = 0,8$$

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B \cap A)}{0,5} = 0,8 \Rightarrow P(B \cap A) = 0,8 \cdot 0,5 = 0,4$$

A probabilidade de que un alumno suspenda as dúas asignaturas será:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,5$$

e a de que un alumno suspenda as dúas

$P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - 0,5 = 0,5$. Así pois, o 50% dos alumnos suspenden as dúas asignaturas

Exercicio 1

Calcúlese a probabilidade de sacar exactamente dúas caras no lanzamento de tres moedas se: a) se sabe que nunha das moedas sae cara. b) se sabe que na primeira das moedas sae cruz.

Regra do produto

Se na igualdade que define a probabilidade condicionada desexamos $P(A \cap B)$, obtemos: $P(A \cap B) = P(B) \cdot P(A/B)$

Esta igualdade permítenos calcular a probabilidade da intersección de dous acontecementos como o produto da probabilidade do primeiro pola do segundo condicionada ao primeiro. Na práctica resulta de gran utilidade como podemos comprobar nos seguintes exemplos:

Exemplo 3

Extraemos dúas cartas dunha baralla, unha a continuación da outra sen devolver a primeira. Calculemos a probabilidade de que a primeira extraída sexa unha copa e a segunda unha espada.

Sexan os acontecementos:

A = “Sacar copa na primeira extracción”

B = “Sacar espada na segunda extracción”.

Temos $P(A) = 10/40$ e $P(B/A) = 10/39$, posto que despois da primeira extracción temos 39 cartas das cales 10 son espadas. Utilizando a regra do produto:

$$P(A \cap B) = P(A) \cdot P(B/A) = \frac{10}{40} \cdot \frac{10}{39}$$

Exemplo 4

Nunha clase o 40% dos alumnos son homes e o 75% destes non queren ouvir falar de fútbol. Calculemos a probabilidade de que ao escoller un alumno da clase ao azar sexa home e futbolero.

Para iso consideremos os acontecementos:

A = “escoller home”

B = “escoller futbolero”

Entón: $P(A) = 0,4$

$P(B/A) = 0,25$

$P(A \cap B) = P(A) \cdot P(B/A) = 0,4 \cdot 0,25 = 0,10$

A regra do produto pode xeralizarse para un número calquera de acontecementos; no caso de tres, teríamos:

$$P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P(C/A \cap B)$$

Exemplo 5

Calculemos a probabilidade de que ao escoller tres persoas ao azar unha detrás doutra dun grupo formado por 10 mulleres e 15 homes, non saia escollido ningún home:

A probabilidade de non escoller home a primeira vez será $10/25$.

A probabilidade de non escoller home a segunda vez, sabendo que non escollemos home na primeira, será $9/24$, posto que o grupo, despois da primeira elección, está formado por 9 mulleres e 15 homes.

Analogamente, a probabilidade de non escoller home a terceira vez será: $8/23$.

Polo tanto, a probabilidade pedida será o produto das tres anteriores:

$$P(\text{“non escoller ningún home”}) = \frac{10}{25} \cdot \frac{9}{24} \cdot \frac{8}{23}$$

En xeral, a probabilidade de que ocorran simultaneamente n acontecementos ven dada por:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2/A_1) \cdot P(A_3/A_1 \cap A_2) \cdots P(A_n/A_1 \cap \dots \cap A_{n-1})$$

Exercicios

2. Escollemos tres cartas unha detrás doutra dunha baralla de 40. Calcúlese a probabilidade de que: a) As cartas escollidas sexan tres ouros. b) As tres sexan sotras. c) Non apareza ningunha espada.

3. Dun lote de 100 artigos dos cales 12 son defectuosos realízanse 5 extraccións consecutivas. Calcúlese a probabilidade de que o primeiro artigo defectuoso apareza na quinta extracción.

Acontecementos independentes

Exemplo 6

Supoñamos que facemos dúas extraccións consecutivas dunha caixa que contén 3 bólas brancas e 4 negras. Podemos realizar o experimento de dúas maneiras, devolvendo a bóla despois da primeira extracción ou deixándoa fóra da caixa. A probabilidade dun acontecemento deste experimento será distinta segundo a modalidade que sigamos. Así, ao calcularmos a probabilidade do acontecemento “sacar dúas bólas negras”, teríamos:

$$\text{Modalidade sen devolución: } P(\text{“sacar dúas bólas negras”}) = \frac{4}{7} \cdot \frac{3}{6}$$

$$\text{Modalidade con devolución: } P(\text{“sacar dúas bólas negras”}) = \frac{4}{7} \cdot \frac{4}{7}$$

No segundo caso, a probabilidade do acontecemento sacar negra na segunda extracción non se ve alterada polo que ocorre na primeira, xa que a caixa conserva a súa composición orixinal, daquela, a probabilidade do acontecemento na segunda extracción é a mesma que na primeira.

En certos experimentos aleatorios o feito de que ocorra determinado acontecemento non afecta á probabilidade de que ocorra outro. Neste caso a probabilidade deste segundo acontecemento condicionada polo primeiro resulta ser igual á probabilidade do propio acontecemento. De ambos os dous acontecementos dise que son *independentes*.

Diremos, polo tanto, que dous acontecementos A e B son independentes se: $P(A/B) = P(A)$. Neste caso a regra do produto transfórmase en:

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

puidendo considerarse esta última expresión como unha definición alternativa de independencia.

En xeral, diremos que os acontecementos $A_1, A_2, \dots, A_n$ son independentes se se verifica:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

Exemplo 7

Calculemos a probabilidade de sacar cinco cruces en cinco lanzamentos consecutivos dunha moeda. Dado que o resultado de cada lanzamento non depende dos anteriores, os cinco acontecementos son independentes, e daquela:

$$P(\text{"sacar cinco cruces"}) = 0,5 \cdot 0,5 \cdot 0,5 \cdot 0,5 \cdot 0,5$$

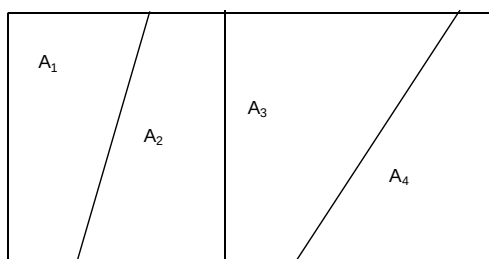
Exercicios

4. Un tirador de arco acerta na diana por termo medio tres de cada cinco frechas. Calcúlese a probabilidade de que ao facer tres lanzamentos acerte nos tres.

5. Unha muller que ten catro nenas está esperando o quinto fillo. Calcúlese a probabilidade de que a) sexa neno. b) sexa nena.

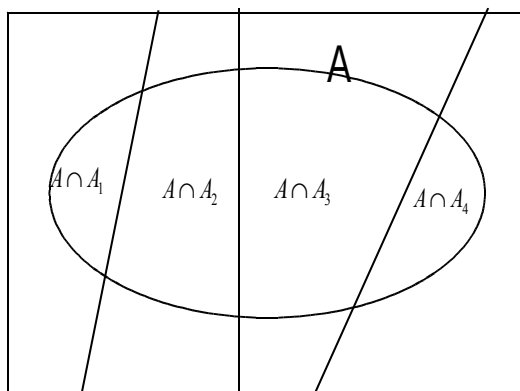
Teorema das probabilidades totais.

Supoñamos que nun experimento aleatorio temos un conxunto de acontecementos dous a dous incompatíbeis e tales que a súa reunión é todo o espazo das mostras. De tal conxunto diremos que é un partición do espazo das mostras.



A figura amosa unha división do conxunto de resultados Ω mediante unha partición formada por catro acontecementos. Dado un acontecemento A , como se ilustra na seguinte figura, podemos escribir:

$$A = (A \cap A_1) \cup (A \cap A_2) \cup (A \cap A_3) \cup (A \cap A_4)$$



Como os acontecementos que forman a partición son incompatíbeis, as súas interseccións con A tamén o serán, e polo tanto teremos:

$$P(A) = P(A \cap A_1) + P(A \cap A_2) + P(A \cap A_3) + P(A \cap A_4)$$

Aplicando a regra do produto a cada sumando:

$$P(A) = P(A_1)P(A/A_1) + P(A_2)P(A/A_2) + P(A_3)P(A/A_3) + P(A_4)P(A/A_4)$$

En ocasións tan só se poden asignar as probabilidades do segundo membro, e daquela a expresión resulta indispensable para calcular $P(A)$.

Exemplo 8

Un instituto ten 600 alumnos dos que 200 están no primeiro ciclo do ESO, 180 no segundo ciclo e 220 en bacharelato. A porcentaxe dos que len dous ou máis libros ao mes –á marxe das lecturas obrigatorias!- é o 5% para o primeiro ciclo, o 3% para o segundo e o 4% para bacharelato. Calcúlese a probabilidade de que escollendo un alumno ao azar lea dous ou máis libros ao mes.

Consideremos os seguintes acontecementos e as súas correspondentes probabilidades:

A = “escoller un alumno do primeiro ciclo do ESO” $P(A) = 200/600 = 0,33$
 B = “escoller un alumno do segundo ciclo do ESO” $P(B) = 180/600 = 0,30$
 C = “escoller un alumno de bacharelato” $P(C) = 220/600 = 0,36$

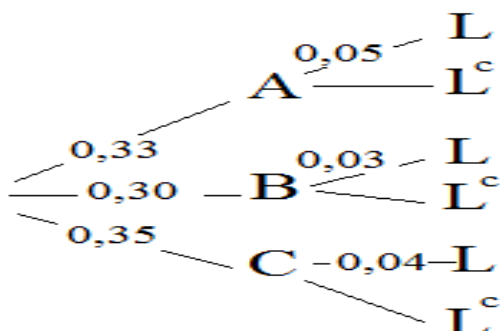
Os acontecementos son incompatíbeis dous a dous e ademais a súa reunión é o total dos alumnos do instituto.

Se L é o acontecemento “escoller un alumno lector”, o que é doado e escribir as probabilidades de L condicionado a A , B e C .

Como $P(L/A) = P(\text{“escoller a un alumno lector sabendo que é de 1º ciclo ESO”}) = 0,05$, e analogamente: $P(L/B) = 0,03$ e $P(L/C) = 0,04$, obtemos:

$$P(L) = P(A)P(L/A) + P(B)P(L/B) + P(C)P(L/C) = 0,33 \cdot 0,05 + 0,30 \cdot 0,03 + 0,36 \cdot 0,04 = 0,04.$$

Os cálculos anteriores pódense esquematizar mediante un diagrama de árbore da seguinte maneira:



Observamos que a probabilidade buscada é a suma dos produtos das probabilidades que aparecen en cada unha das pólas da árbore que conducen ao acontecemento L.

Podemos enunciar o teorema na súa forma máis xeral para un número n de acontecementos.

Teorema das probabilidades totais. Se $A_1, A_2, \dots, A_n$ forman unha partición do conxunto de resultados, e A é un acontecemento calquera, verificase:

$$P(A) = P(A_1)P(A / A_1) + P(A_2)P(A / A_2) + \dots + P(A_n)P(A / A_n)$$

Regra de Bayes

Utilizando á definición de probabilidade condicionada e o teorema das probabilidades totais, podemos responder algunhas cuestións máis respecto á probabilidade de certos acontecementos.

Exemplo 9

Se no suposto do exemplo 8, escollemos un alumno ao chou e resultou ser lector de dous o máis libros, cal é a probabilidade de que sexa do primeiro ciclo do ESO?

Seguindo a notación do exemplo anterior, a probabilidade pedida é:

$$P(A / L) = \frac{P(A \cap L)}{P(L)}$$

cuxo numerador, utilizando a regra do produto, se transforma en

$$P(A \cap L) = P(A)P(L / A).$$

Daquela

$$P(A / L) = \frac{P(A \cap L)}{P(L)} = \frac{P(A)P(L / A)}{P(L)}$$

onde o denominador xa está calculado anteriormente utilizando o teorema das probabilidades totais, co que:

$$P(A / L) = \frac{0,33 \cdot 0,05}{0,04} = 0,4125$$

A fórmula utilizada neste exemplo é debida ao matemático inglés Thomas Bayes (1702-1761), e pode xeralizarse para n acontecementos

Teorema de Bayes. Se $A_1, A_2, \dots, A_n$ forman unha partición do conxunto de resultados, e A é un acontecemento calquera, verifícase:

$$P(A_i / A) = \frac{P(A_i)P(A / A_i)}{P(A_1)P(A / A_1) + P(A_2)P(A / A_2) + \dots + P(A_n)P(A / A_n)}$$

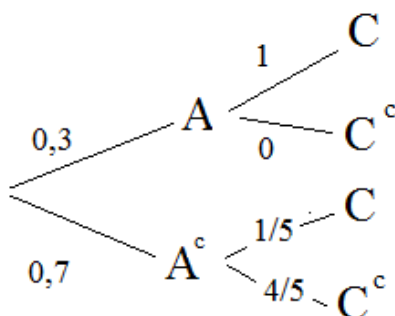
para calquera acontecemento A_i con $i=1,2,\dots,n$.

Na interpretación tradicional da fórmula considéranse as probabilidades condicionadas como efectos asociados a unha causa, en experimentos repetíbeis, que teñen unha confirmación empírica. Porén, os estatísticos chamados bayesianos admiten as probabilidades subxectivas, servindo entón o teorema para indicar como debemos modificar a adxudicación de probabilidades subxectivas unha vez recibida a información adicional da que nos dota un experimento.

Exemplo 10

Un test formula preguntas con cinco posíbeis respostas cada unha. Unha persoa ten probabilidade 0,3 de coñecer unha resposta; en caso contrario contesta ao azar. Se a persoa contesta correctamente, cal é a probabilidade de que coñeza a resposta?

Sexan os acontecementos A = “a persoa coñece a resposta”, C = “a persoa dá a resposta correcta”.



Se tomamos como partición A e A^c , temos:

$$P(A / C) = \frac{P(A)P(C / A)}{P(A)P(C / A) + P(A^c)P(C / A^c)} = \frac{0,3 \cdot 1}{0,3 \cdot 1 + 0,7 \cdot 0,2} = 0,68.$$

Exercicios

6. Unha urna A contén 5 bólas brancas e tres negras. Outra B contén 6 bólas brancas e tres negras. Escóllese unha urna ao azar e despois unha bóla desa urna. Calcúlese a probabilidade de que a bóla escollida sexa negra. Se realizamos o experimento e a bóla escollida foi negra, cal é a probabilidade de que procedese da urna A?

7. O 70% dos clientes dunha compañía de seguros ten máis de 25 anos. Un 5% de clientes deste grupo ten un accidente ao longo do ano. No caso dos clientes de 25 anos ou menos a porcentaxe é do 20%. Se escollemos un asegurado ao chou, cal é a probabilidade de que teña un accidente ese ano? Se unha persoa tivo un accidente, cal é a probabilidade de que teña 25 anos ou menos?

Algúns problemas clásicos

O problema de Galileo

Un xogador pediulle a Galileo que explicase por que aparentemente era máis favorábel apostar á suma 10 que a suma 9 no lanzamento de tres dados, se ambas sumas podían aparecer do mesmo número de maneiras.

En efecto:

Suma 9	1 2 6	1 3 5	1 4 4	2 2 5	2 3 4	3 3 3
Suma 10	1 3 6	1 4 5	2 2 6	2 3 5	2 4 4	3 3 4

Teñamos en conta que o 1, o 2 e o 6 poden aparecer en 1 2 6 (1 no primeiro dado, 2 no segundo e 6 no terceiro) mais tamén en 1 6 2, 2 6 1, etc. Ou sexa poden aparecer de todos os xeitos posíbeis en que se poden ordenar que son $3! = 6$ veces.

Analogamente 1, 4 e 4 poden aparecer tres veces: 1 4 4, 4 1 4 e 4 4 1; e 3 3 3 unha soa vez.

Botando as contas:

Casos favorábeis a suma 9 = $6+6+3+3+6+1 = 25$.

Casos favorábeis a suma 10 = $6+6+3+6+3+3 = 27$.

Como o número de casos posíbeis é $6^3 = 216$, temos que: $P(\text{"suma 9"}) = 25/216$ mentres que $P(\text{"suma 10"}) = 27/216$. Xa que logo ten máis probabilidade de éxito apostar á suma 10.

Dous problemas do cabaleiro de Meré

1) O cabaleiro de Meré, xogador impenitente, observara que ao botar catro veces un dado era máis probábel que aparecese polo menos un seis a que non aparecese ningún. Baseándose neste resultado, creu que ao botar dous dados 24 veces, sería máis doado obter un dobre seis que o acontecemento contrario. Como as súas observacións non avalaban a hipótese consultou o problema con Pascal, quen llo trasladou a Fermat. Este conseguiu dar a razón teórica de por que a observación do de Meré era certa.

En ambos casos o experimento consiste na repetición de probas independentes. Chamando A ao acontecemento “aparece polo menos un 6 ao botar 4 veces un dado”, temos que $P(A^c) = \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}$, e daquela: $P(A) = 1 - \frac{5^4}{6^4} = 0,5178 > \frac{1}{2}$

Ao botar dous dados, temos un conxunto de resultados de 36 elementos; polo tanto a probabilidade de non sacar 6 dobre nunha tirada é $35/36$. Se repetimos o

experimento 24 veces, a probabilidade de non sacar ningún 6 dobre será:
 $P(B^c) = \frac{35^{24}}{36^{24}} = 0,51$, e a probabilidade de sacar polo menos un 6 dobre nas 24 tiradas
 $P(B) = 1 - 0,51 = 0,49 < 1/2$.

2) En certa ocasión o empedernido xogador, fixo unha aposta co seu par, que consistía en apostar certa cantidade de diñeiro ao primeiro que gañase tres partidas dun xogo de baralla. Despois de que o noso cabaleiro gañase a primeira, o xogo foi suspendido por causas maiores, e outra volta foi Fermat quen tivo que ditar sentenza.

Podemos calcular todos os desenvolvementos posíbeis dun xogo entre dous xogadores que finaliza ao gañar un deles tres partidas (por exemplo mediante un diagrama de árbore)

caso	probabi- lidade	caso	probabi- lidade
AAA	1/8	BBB	1/8
AABA	1/16	BBAB	1/16
AABBA	1/32	BBAAB	1/32
AABBB	1/32	BBAAA	1/32
ABAA	1/16	BABB	1/16
ABABA	1/32	BABAB	1/32
ABABB	1/32	BABAA	1/32
ABBAA	1/32	BAABB	1/32
ABBAB	1/32	BAABA	1/32
ABBB	1/16	BAAA	1/16

Cada arranxo de letras dános un desenlace do xogo, onde A representa a partida gañada polo primeiro xogador, e B a gañada polo segundo. As probabilidades calcúlanse como produto de acontecementos independentes con probabilidade $\frac{1}{2}$.

Temos en total 20 acontecementos en 10 dos cales gaña o primeiro, e nos outros dez o segundo. Se sumamos a probabilidade de cada grupo, naturalmente obtemos $\frac{1}{2}$. O xogo é equitativo.

Se aceptamos a información adicional de que o primeiro xogador gañou a primeira partida, a táboa anterior fica reducida á metade esquerda, da que sumando obtemos que a probabilidade de que gañe A é $11/16$ e a de que gañe B, $5/16$.

Así pois, o reparto do fondo ao ser interrompida a partida debe facerse en proporción 11 a 5 favorábel ao primeiro xogador.

O problema do aniversario

Nun conxunto de n persoas trátase de calcular a probabilidade de que polo menos dúas estean de aniversario o mesmo día.

A probabilidade que temos que calcular é a do acontecemento contrario a:

A= “non hai ningunha coincidencia”

Sexa A_1 = “a primeira persoa escollida naceu un día calquera do ano”, daquela:

$$P(A_1) = \frac{365}{365}.$$

Sexa A_2 = “a segunda persoa escollida naceu un día distinto á primeira”. Temos:

$$P(A_2) = \frac{364}{365}.$$

Sexa A_3 = “a terceira persoa escollida naceu un día distinto á primeira e á segunda”.

Temos:

$$P(A_3) = \frac{363}{365}$$

Analogamente, teríamos:

$$P(A_n) = \frac{365 - (n - 1)}{365}.$$

Daquela, a probabilidade de que todos nacesen en días distintos, sería:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = \frac{365}{365} \times \frac{364}{365} \times \dots \times \frac{365 - (n - 1)}{365}$$

E a probabilidade pedida é a do acontecemento contrario:

$$P(A^c) = 1 - \frac{365}{365} \times \frac{364}{365} \times \dots \times \frac{365 - (n - 1)}{365}.$$

Para $n = 23$, temos: $P(A^c) = 0,50729723$, co cal, a partir de 23 persoas é máis probábel que haxa coincidencias que que non as haxa.

O problema das cartas e os sobres

Un axente electoral ten n sobres con nomes e enderezos xunto con n cartas persoais. Se introduce ao azar as cartas nos sobres, cal é a probabilidade de que polo menos unha carta coincida co seu sobre?

Este problema foi proposto por Pierre de Montmort (1678, 1719) e aparece baixo múltiples enunciados (devolver ao azar os gabáns dun roupeiro e que coincidan cos seus números de identificación, extraer bólas numeradas dunha urna e que coincida o seu número coa orde de extracción,...) e presenta a particularidade de que a partir de $n = 10$ a probabilidade de que o acontecemento se produza permanece practicamente invariábel.

Supoñamos os sobres numerados do 1 ao n . A primeira carta pode ser introducida de n maneiras diferentes, a segunda de $n-1$, etc. O número de resultados posíbeis deste experimento será $n!$. O acontecemento $A_i =$ “a carta i -ésima vai ao sobre i -ésimo” pode ocorrer de $(n-1)!$ maneiras, polo tanto $P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$. Como era de agardar: $P(A_1) + P(A_2) + \dots + P(A_n) = 1$

A probabilidade de que dúas cartas i e k vaian aos sobres correctos será:

$$P(A_i \cap A_k) = \frac{(n-2)!}{n!},$$

e o número total de interseccións $A_i \cap A_k : \binom{n}{2} = \frac{n!}{2! \cdot (n-2)!}$.

Xa que logo, a suma das probabilidades correspondentes a todas as interseccións de orde dous é:

$$\sum_{i \neq k} P(A_i \cap A_k) = \frac{n!}{2! \cdot (n-2)!} \frac{(n-2)!}{n!} = \frac{1}{2!}$$

Analogamente:

$$\sum_{i, k, j} P(A_i \cap A_k \cap A_j) = \frac{n!}{3! \cdot (n-3)!} \frac{(n-3)!}{n!} = \frac{1}{3!}$$

.....

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = \frac{1}{n!}$$

Co cal podemos calcular a probabilidade pedida:

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n) &= P(A_1) + P(A_2) + \dots + P(A_n) - \sum_{i \neq k} P(A_i \cap A_k) + \sum_{i, k, j} P(A_i \cap A_k \cap A_j) + \\ &+ \dots \pm P(A_1 \cap A_2 \cap \dots \cap A_n) = 1 - \frac{1}{2} + \frac{1}{3!} - \dots \pm \frac{1}{n!} \end{aligned}$$

A probabilidade, como podemos comprobar, apenas varía para valores de $n > 10$.

Para $n = 5$ obtemos unha probabilidade de 0,375.

Para $n = 10$, $p = 0,36787919$

Para $n = 30$, $p = 0,367879441171442$

Para $n = 100$, $p = 0,367879441171442$.

No resultado deste problema, como acontece *nos lugares máis insospeitados da matemática*, aparece o número e xa que:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{1}{2} + \frac{1}{3!} + \dots \pm \frac{1}{n} \right) = \frac{1}{e}$$

As caixas de Bertrand

En 1889 o matemático francés Edouard Bertrand (1822, 1900) presentou no seu libro *Calcul des probabilités* o seguinte problema, que aparentemente leva a unha solución paradoxal:

Hai tres caixas iguais pero con distintos contidos. A primeira contén dúas moedas de ouro, a segunda dúas moedas de prata e a terceira unha moeda de ouro e outra de prata. O xogador escolle unha caixa e extrae unha moeda. Cal é a probabilidade de que a outra moeda sexa do mesmo metal?

O conxunto de resultados da proba é :

$$\Omega = \{OO, PP, OP\}$$

e, polo tanto, a probabilidade de escoller unha caixa coas dúas moedas do mesmo metal, é $2/3$. Porén, o problema tamén se presta a outra interpretación. Se, poñamos por caso, a moeda extraída é de ouro, desbotamos a posibilidade de que a caixa escollida sexa a que contén dúas moedas de prata. Polo tanto, será ou ben a OO ou a OP, de xeito que a probabilidade de que a moeda que permanece no interior da caixa sexa do mesmo metal tería que ser $1/2$.

Este aparente paradoxo é utilizado por algúns feirantes e presentadores de concursos televisivos para inclinar ao participante cara á escolla menos favorábel. A versión máis extendida coñécese no mundo anglosaxón como o problema de Monty Hall (sobre el existen moitísimas páxinas web). Nesta interpretación o participante no concurso televisivo debe elixir unha porta de entre tres (todas pechadas), consistindo o premio en levar o que se atopa detrás da elixida. Sábese con certeza que tras dunha delas ocúltase un automóbil, e tras as outras dúas hai senllos calacús. Unha vez que o concursante elixi unha porta e lle comunique ao público e ao presentador a súa elección, Monty (o presentador), que sabe o que hai detrás das portas, abrirá unha das outras dúas e mostrará que detrás hai un calacú. Neste momento dáselle a opción ao concursante de cambiar de porta (ten dúas opcións). Debe o concursante manter a súa elección orixinal ou escoller a outra porta?

Esta pregunta xerou un intenso debate xa que a resposta como vimos semella contraditoria.

A urna de Polya

Temos unha urna que contén b bólas brancas e n bólas negras. Extraemos unha bóla ao azar e devolvémola á urna xunto con k bolas da mesma cor. Calcúlese a probabilidade de sacar p bólas brancas seguidas de q bolas negras.

Este experimento aleatorio é un modelo para o estudo teórico dos chamados *esquemas de contaxio*. En epidemioloxía utilízase para predicir os efectos que certa enfermidade contaxiosa pode ter sobre unha poboación. Se as bólas brancas representan individuos enfermos e as negras sans, a localización dun enfermo modifica a composición da poboación aumentando en k individuos contaxiados. Xa que logo, as taxas esperadas de individuos enfermos poden ser calculadas deste xeito.

A probabilidade de bóla branca na primeira extracción, será:

$$P(B_1) = \frac{b}{b+n}$$

A probabilidade de sacar dúas brancas nas dúas primeiras extraccións:

$$P(B_1 \cap B_2) = P(B_1)P(B_2 / B_1) = \frac{b}{b+n} \cdot \frac{b+k}{b+n+k}$$

Se facemos un número p de extraccións, a probabilidade de que sexan todas brancas, utilizando a regra do produto, será:

$$P(B_1 \cap B_2 \cap \dots \cap B_p) = \frac{b}{b+n} \cdot \frac{b+k}{b+n+k} \cdot \frac{b+2k}{b+n+2k} \cdots \frac{b+(p-1)k}{b+n+(p-1)k}$$

A probabilidade de que a bóla $p-1$ sexa negra é entón:

$$P(B_1 \cap B_2 \cap \dots \cap B_p \cap N_{p+1}) = \frac{b}{b+n} \cdot \frac{b+k}{b+n+k} \cdot \frac{b+2k}{b+n+2k} \cdots \frac{b+(p-1)k}{b+n+(p-1)k} \cdot \frac{n}{b+n+pk}$$

e se continuamos o proceso, a probabilidade de extraer q negras será:

$$\begin{aligned} P(B_1 \cap B_2 \cap \dots \cap B_p \cap N_{p+1} \cap \dots \cap N_{p+q}) = \\ = \frac{b}{b+n} \cdot \frac{b+k}{b+n+k} \cdot \frac{b+2k}{b+n+2k} \cdots \frac{b+(p-1)k}{b+n+(p-1)k} \cdot \frac{n}{b+n+pk} \cdots \frac{n+k(q-1)}{b+n+k(p+q-1)} \end{aligned}$$

Exercicios

8. Botamos dous dados ao ar. Se aparecen resultados distintos, cal é a probabilidade de que a suma sexa par?

9. O 40% dos habitantes dunha cidade ten 20 anos ou menos. Os homes son o 25%, e os homes de 20 anos ou menos, o 15%. Calcular a probabilidade de que escollido ao azar un habitante da cidade:

- a) Sexa muller maior de 20 anos.
- b) Se resulta home, que sexa maior de 20 anos.
- c) Se ten 20 anos ou menos, que sexa home.

10. Unha urna contén 5 bólas brancas, 3 vermellas e 6 azuis. Outra contén 2 brancas, 5 vermellas e 4 azuis. Escollemos unha urna ao azar e despois unha bóla desa urna. Calcúlese a probabilidade de:

- a) Escoller unha bóla vermella
- b) Non escoller bóla azul.

11. O total de artigos elaborados por unha fábrica repártense entre tres máquinas A, B e C, que producen respectivamente o 20, 30 e 50 por cento. As porcentaxes de artigos defectuosos de cada máquina son o 1, 1,5 e 2 por cento. Se escollemos un artigo ao azar, calcúlese a probabilidade de que:

- a) Sexa defectuoso.
- b) Se o artigo escollido é defectuoso, fora producido pola máquina A.
- c) Se non é defectuoso, fora producido por B.

12. A probabilidade de que unha alarma contra incendios funcione correctamente é 0,7. Se funciona, a probabilidade de que os bombeiros cheguen a tempo é 0,8, e se non funciona 0,3. Calcúlese:

- a) Probabilidade de que en caso de incendio os bombeiros cheguen a tempo.
- b) Se os bombeiros chegan a tempo, probabilidade de que soara a alarma.

13. Un xogador de fútbol consegue goles no 75% dos seus lanzamentos de penalti. Se realiza unha serie de tres, cal é a probabilidade de que polo menos meta un gol?

14. Nunha enquisa electoral obtense que o partido X levará o 45% dos votos. Se escollemos a cinco persoas calcúlese a probabilidade de que:

- a) Todas voten ao partido X.
- b) Polo menos unha non vote ao partido X.

15. Unha empresa de reparacións domésticas ten no cadro de persoal 20 pintores dos que 7 son mulleres, 6 fontaneiros e 6 fontaneiras, e 9 albaneis dos que son mulleres 3. Se escollemos un empregado ao chou, calcúlese a probabilidade de que:

- a) Sexa home e fontaneiro
- b) Sabendo que é muller, que sexa albanel.
- c) Sabendo que é pintor, que sexa home.

16. O 2% das pezas de recambio de automóbil que chegan a unha compañía distribuidora son defectuosas. Na distribuidora son sometidas a unha revisión que ten un erro do 5%. Se escollemos unha peza ao azar, calcúlese a probabilidade de que sexa defectuosa se na inspección foi declarada defectuosa.

17. Unha urna contén 8 bólas brancas e 5 negras. Extraemos unha bóla e introducimos 3 bólas de distinta cor á extraída. Se a continuación sacamos outra bóla, calcúlese a probabilidade de que:

- a) A bóla extraída sexa de cor negra.
- b) Se a segunda bóla é branca, que a primeira fose negra.

18. Temos dúas urnas. A primeira contén 2 bólas brancas e 1 negra. A segunda, 1 bóla branca e tres negras. Escollemos ao azar unha bóla da primeira e metémola na segunda. A continuación escollemos unha bóla da segunda urna que resultou ser negra. Cal é a probabilidade de que a bóla elixida fose branca?

19. Unha urna contén 4 bólas brancas e tres negras e noutra hai 3 brancas e 4 negras. Extraemos unha bóla da primeira e metémola na segunda. A continuación extraemos unha bóla da segunda e introducímola na primeira. Cal é a probabilidade de que a primeira teña a súa composición inicial?
20. Nunha tómbola hai 2 bonecas, tres balóns e 4 peluches. Numéranse todos os obxectos do 1 ao 9 e escóllense tres premios ao azar. O primeiro resulta ser un balón, o segundo resérvase como sorpresa e o terceiro é unha boneca. Calcúlese a probabilidade de que o segundo sexa un peluche.
21. As máquinas A, B e C fabrican peza similares. As producións respectivas por hora son de 300, 500 e 600 pezas. Sendo as porcentaxes de defectuosas do 2%, 3,5% e 2,5% respectivamente. Da produción total das tres máquinas ao final da xornada, escollemos unha peza aleatoriamente. Cal é a probabilidade de que sexa defectuosa?
22. Nun concurso televisivo os participantes deben someterse a tres probas independentes non eliminatorias, tendo cada unha delas o dobre de dificultade que a anterior (a probabilidade de superala é a metade). Sábese que o 0,8% dos participantes superan as tres probas. Cal é a probabilidade de superar dúas probas? Se un concursante superou a última, cal é a probabilidade de que superase as dúas anteriores?
23. O 0,5% das unidades producidas nunha fábrica son defectuosas cando o proceso está baixo control. Se o proceso está fóra de control prodúcese un 30% de unidades defectuosas. A probabilidade de que o proceso estea baixo control é do 92%. Se escollemos aleatoriamente unha unidade e resultou defectuosa, cal é a probabilidade de que o proceso estea baixo control?
24. O 60% dos que visitan uns Grandes Almacéns non teñen intención de mercar nada. O 30% dos que teñen intención de mercar non merca nada. Que é máis probábel, mercar sen ter intención ou non mercar ténendoa?
25. Nunha gaiola temos 4 cotorras e 3 loros. Nunha segunda gaiola, 5 loros e 10 cotorras. Un paxaro da primeira xaula cólase na segunda, e nesta o experimentador mete a man a cegas e saca un loro. Cal é a probabilidade de que o animal que pasou da primeira a segunda gaiola fose unha cotorra?
26. Dúas empresas A e B compiten no mesmo mercado e ambas as dúas estudan a posibilidade de abrir unha sucursal nunha nova zona. Se A non abre, B tampouco o fará con probabilidade 0,9, pero se A abre, B farao cunha probabilidade tamén de 0,9. Calcúlese a probabilidade de que se instalen as dúas ou ningunha.

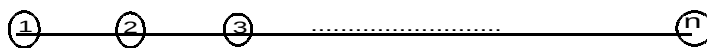
3. CADEAS DE MARKOV

*Ao elixir o presente, o pasado e o porvir
para situar o seu interrogante, o suxeito
proxectaba sobre a súa natureza íntima
unha luz preciosa, que se vía completada
pola elección dun acontecemento propicio
ou desfavorábel.*

Raymond Roussel. *Locus Solus*

Percorridos aleatorios

Supoñamos unha partícula que se move ao azar entre un número determinado de puntos dunha liña recta.



Cando a partícula está situada nun punto, móvese unha unidade cara á dereita con probabilidade p e cara á esquerda con probabilidade $q = 1 - p$. Se se atopa nun dos puntos extremos entón permanece indefinidamente atrapada nese punto.

Con relación a este proceso poderíamos facernos algunha pregunta. Por exemplo, sería interesante saber cantos pasos dará por termo medio unha partícula que comeza o seu percorrido nun certo punto antes de ser absorbida por un dos extremos. Ou, situada a partícula nun punto determinado, estimar as probabilidades de acadar un dos extremos, para deste xeito poder predicir a súa traxectoria máis probábel. Tamén poderíamos interesarnos por cal sería a probabilidade de pasar dun punto a outro nun certo número de pasos ou, coñecido o punto de saída, saber en que estado se atoparía o proceso despois dun número p de pasos.

Existen moitas variantes do movemento anterior. Ás veces considérase que cando a partícula chega a un dos puntos extremos é automaticamente desplazada ao punto contiguo, ou ben noutros casos salta con probabilidade 1 a un punto intermedio determinado, para a continuación repetir indefinidamente o proceso. Nestes casos é tamén interesante considerar a posibilidade de que o movemento chegue a estabilizarse, é dicir, que dalgunha maneira se faga estacionario.

Este proceso é unha descrición simplificada do fenómeno coñecido en Física como camiño aleatorio. Aparece, por exemplo, cando se estuda o movemento das moléculas dun gas dentro dun recipiente. O estudo deste fenómeno pode ser abordado mediante o modelo matemático que se estudará a continuación, debido ao matemático ruso Andrei Andreevitch **Markov** (1856-1922), que na súa versión máis desenvolvida atopa múltiples aplicacións, non só na Física senón tamén, poñamos por caso, na Psicoloxía, a Economía ou na Epidemioloxía.

Procesos estocásticos finitos

Moitos experimentos de ciencias, economía, enxeñería, etc., consisten en arranxos finitos de probas cada unha das cales presenta un número finito de acontecementos con probabilidades dadas. Procesos deste tipo reciben o nome de **procesos estocásticos**.

Exemplo 1

Escollemos cinco estudantes ao azar nunha clase formada por 30 homes e 10 mulleres e anotamos o seu sexo.

Este experimento consta de cinco probas iguais en cada unha das cales se poden dar dous acontecementos: escoller muller ou escoller home. Con axuda dun diagrama de árbore o o teorema da multiplicación podemos calcular as probabilidades de cada un destes dous acontecementos en cada un dos cinco pasos. A probabilidade de que ocorra un acontecemento nunha proba dependerá do que ocorra en todas as anteriores.

Exemplo 2

Disparamos cinco veces sobre unha diana con probabilidade de acertar cada vez igual a 0,3.

Temos outra vez un arranxo de cinco probas das que cada unha delas presenta dous acontecementos posíbeis: acertar ou errar. O cálculo das probabilidades destes acontecementos vese agora facilitado polo feito de que as probas son independentes e, polo tanto, a probabilidade de que ocorra un acontecemento determinado nunha das probas é independente do resultado das anteriores.

Exemplo 3

Durante os cinco días laborábeis unha persoa repártese entre dúas ocupacións: ou vai de pesca ou traballa no campo. Nunca pesca dous días seguidos, pero se un día traballa no campo é igualmente probábel que o día seguinte realice calquera das dúas tarefas. O luns tira unha moeda para decidir a que se vai dedicar.

Unha vez máis o experimento consta de cinco probas con dous acontecementos posíbeis en cada unha delas. A diferenza dos experimentos anteriores, a probabilidade de que en cada unha das probas ocorra un acontecemento depende **unicamente** do resultado da proba anterior. Procesos con esta característica serán o obxecto de estudo deste capítulo.

Exercicios

1. En cada un dos exemplos anteriores calcúlense as probabilidades dos seguintes acontecementos: No exemplo 1, probabilidade da serie HHMMM (H escoller home, M escoller muller). No exemplo 2, probabilidade de AAEEA (A acerto, E erro). No exemplo 3, probabilidade de PCCCP (P pesca, C traballa no campo)

2. Calcúlense as probabilidades seguintes referidas aos exemplos anteriores.

- a) Probabilidade de escoller unha muller na terceira elección.
- b) Probabilidade de acertar na diana no cuarto disparo.
- c) Probabilidade de ir pescar no segundo día.

Espazo de estados dun proceso

Asociado a un proceso estocástico temos o conxunto formado por todos os resultados posíbeis de cada unha das probas. Este conxunto recibe o nome de **Espazo de estados** do proceso, que designaremos:

$$E = \{a_1, a_2, \dots, a_n\}$$

O espazo de estados do proceso 1 sería: $E = \{H, M\}$

O espazo de estados do exemplo 2 sería: $E = \{A, E\}$

O espazo de estados do exemplo 3 sería: $E = \{P, C\}$

Cando o resultado da proba m é a_i dicimos que o proceso está no estado a_i no paso m -ésimo.

Cadeas finitas de Markov

Un proceso estocástico finito como o do exemplo 3 no que se verifica que o resultado dunha proba determinada depende como máximo da proba inmediatamente anterior, e non de ningún dos resultados previos, recibe o nome de **Cadea finita de Markov**.

Así, asociados a unha cadea de Markov temos:

- a) Un espazo de estados : $E = \{a_1, a_2, \dots, a_n\}$
- b) A probabilidade de que o estado a_j ocorra inmediatamente despois que o estado a_i , para cada dous estados a_i, a_j , á que chamaremos probabilidade de transición do estado a_i ao estado a_j e designaremos por p_{ij} .
- c) Se ordenamos todas as probabilidades de transición p_{ij} nunha matriz n por n , obtemos a **matriz de transición** do proceso:

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{pmatrix}$$

Nesta matriz cada liña contén as probabilidades de transición dun estado determinado a todos os demais, é dicir, as probabilidades de todos os resultados posíbeis da próxima proba. A suma de todas estas probabilidades debe ser 1 e, polo tanto, a suma

dos elementos de cada fila da matriz de transición vale 1. Matrices con esta característica dícense estocásticas.

- d) A matriz liña ou vector formado polas probabilidades do estado inicial do sistema. Chamaremos a este vector **vector de probabilidade inicial** e desinarémolo por :

$$P^{(0)} = (p_1^{(0)}, p_2^{(0)}, \dots, p_n^{(0)})$$

Obsérvese que a suma das probabilidades que expresan os elementos do vector debe ser igual a 1 e, polo tanto, o vector de probabilidade inicial é un vector estocástico.

Exemplo 4.

Calculemos os elementos correspondentes á cadea de Markov do exemplo 3.

O espazo de estados será o conxunto $E = \{P, C\}$ onde P corresponde ao acontecemento “ir pescar” e C ao acontecemento “traballar no campo”.

Temos catro probabilidades de transición :

Probabilidade de ir pescar se o día anterior se fixo o mesmo: $p_{11} = 0$

Probabilidade de traballar no campo se o día anterior foi pescar: $p_{12} = 1$

Probabilidade de ir pescar se o día anterior traballou no campo: $p_{21} = 1/2$

Probabilidade de traballar no campo se o día anterior se fixo o mesmo: $p_{22} = 1/2$

Estes valores quedan ordenados na matriz de transición:

$$P = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$$

Dado que o luns se tira unha moeda para saber se ten que ir pescar ou traballar no campo, o vector de probabilidade inicial será:

$$P^{(0)} = (1/2, 1/2)$$

Exercicios.

3. Nunha determinada bisbarra, se un día hai néboa o seguinte chove pero, se chove, o día seguinte é soleado. Tense observado que as probabilidades de que a un día con sol suceda un día neboento ou chuvioso son respectivamente 0,4 e 0,6. Escríbase a matriz de transición deste proceso.
4. Unha central de control de tráfico chequea a cotío un de tres puntos conflitivos A, B e C. A probabilidade de que un día determinado controle o mesmo punto que o día anterior é $1/2$ e as probabilidades de que controle un calquera dos outros dous puntos restantes son iguais. Escríbase a matriz de transición.
5. Estúdese se os procesos estocásticos dos exemplos 1 e 2 son procesos de Markov. En caso afirmativo escríbase a matriz de transición.
6. Realizamos tiradas repetidas cun dado normal e apuntamos despois de cada tirada o número máximo obtido desa tirada e de todas as anteriores. Trátase dun

proceso de Markov? En caso afirmativo escribíbase a matriz de transición do proceso.

Probabilidade de transición en k pasos

Á probabilidade de que un proceso pase do estado a_i ao estado a_j exactamente nun número k de pasos, chámasele **probabilidade de transición en k pasos**. Representámola por:

$$P_{ij}^{(k)}$$

As probabilidades de transición en k pasos poden ser ordenadas na matriz:

$$P^{(k)} = \begin{pmatrix} P_{11}^{(k)} & P_{12}^{(k)} & \cdots & P_{1n}^{(k)} \\ P_{21}^{(k)} & P_{22}^{(k)} & \cdots & P_{2n}^{(k)} \\ \cdots & \cdots & \cdots & \cdots \\ P_{n1}^{(k)} & P_{n2}^{(k)} & \cdots & P_{nn}^{(k)} \end{pmatrix}$$

Que recibe o nome de **matriz de transición de k pasos**.

O teorema seguinte caracteriza a matriz de transición de k pasos e proporcionanos un método para calcular as distintas probabilidades de transición.

Teorema 1.

Se P é a matriz de transición dunha cadea de Markov, entón a matriz de transición de k pasos e a k -ésima potencia de P .

$$P^{(k)} = P^k$$

Exemplo 4.

Consideremos o proceso de Markov do exemplo 3, a matriz de transición do proceso viña dada por:

$$P = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$$

Buscamos a probabilidade de que a persoa do enunciado se dedique a traballar no campo tres días despois de dedicarse a pescar. É dicir, a probabilidade de que o proceso pase do estado “ir pescar” ao estado “traballar no campo” exactamente en tres pasos.

Se calculamos a terceira potencia da matriz de transición P do proceso, obtemos a matriz:

$$P^3 = P \cdot P \cdot P = \begin{pmatrix} 1/4 & 3/4 \\ 3/8 & 3/8 \end{pmatrix}$$

Este cálculo pode facerse de xeito máis doado se utilizamos a función multiplicación de matrices da folla de cálculo. Os pasos a realizar serán:

Escribimos os elementos da matriz no rango A1-B2

Sinalamos o rango C1- D2.

Escollemos a función matriz multiplicación no grupo de funcións matemáticas.

No cadro de diálogo correspondente sinalamos A1-B2 como primeira matriz e, outra vez, A1-B2 como segunda.

Premendo control maiúsculas, ao se tratar dunha función matricial, aceptamos.

No rango C1-D2 aparecerán o elementos da matriz $P \cdot P = P^2$.

Repetindo o proceso para esta matriz e a primeira obteremos a terceira potencia de P.

Obsérvase que a probabilidade pedida corresponderá ao elemento da primeira fila, segunda columna da matriz.

$$p_{12}^{(3)} = \frac{3}{4}$$

Exercicios.

7. No proceso de Markov do exercicio 3, calcúlese a probabilidade de que se chove o martes tamén chova o xoves.
8. Un anuncio luminoso amosa a palabra **META**. As letras que o forman acéndense alternativamente segundo a seguinte regra: a iluminación dunha letra móvese cara á dereita con probabilidade $1/3$ e cara a esquerda con probabilidade $2/3$. Cando está acendida unha das letras dos extremos no paso seguinte ilumínase a letra central adxacente. Calcúlese a probabilidade de que, se está iluminada a letra T, se ilumine a letra E despois de tres pasos.
9. Unha persoa pode escoller entre tres liñas de autobús para se desprazar diariamente ao seu lugar de traballo. Se un día escolle a liña A, o día seguinte escolle a liña B e o día seguinte a liña C, pero cando escolle a liña C é igualmente probábel que o día seguinte escolla as liñas A ou B. Escríbase a matriz de transición do proceso e calculando despois a terceira potencia desta matriz, estímense as probabilidades seguintes:
 - a) Probabilidade de que escolla a liña B tres días despois de escoller a liña A.
 - b) Probabilidade de que tres días despois de escoller a liña C escolla outra vez esa liña.

Probabilidades totais

A probabilidade de que o proceso se atope no estado a_i despois de realizar k probas recibe o nome de probabilidade total ou absoluta. Denotaremos esta probabilidade por: $p_i^{(k)}$. A cada proba k corresponde un vector estocástico formado por todas as probabilidades totais desa proba.

$$P^{(k)} = (p_1^{(k)}, p_2^{(k)}, \dots, p_n^{(k)}),$$

que recibe o nome de **distribución de probabilidades do paso k**.

En particular, o vector de probabilidade inicial:

$$P^{(0)} = (p_1^{(0)}, p_2^{(0)}, \dots, p_n^{(0)})$$

corresponde á distribución de probabilidade de paso cero.

Podemos calcular todas as probabilidades absolutas correspondentes a unha proba **k** do sistema utilizando o seguinte teorema:

Teorema 2.

Consideremos unha cadea de Markov con matriz de transición P . Dado o vector de probabilidade inicial :

$$P^{(0)} = (p_1^{(0)}, p_2^{(0)}, \dots, p_n^{(0)})$$

temos que:

$$P^{(1)} = P^{(0)} \cdot P$$

$$P^{(2)} = P^{(1)} \cdot P = P^{(0)} \cdot P^2$$

.....

$$P^{(k)} = P^{(k-1)} \cdot P = P^{(0)} \cdot P^k$$

Así, as distribucións de probabilidade poden ser atopadas multiplicando o vector de probabilidade inicial pola potencia correspondente da matriz de transición.

Exemplo 5

Calculemos a distribución de probabilidade do exemplo 3 despois de catro días. Dado que a persoa tira inicialmente unha moeda o primeiro día para decidir a que se vai dedicar, o vector de probabilidade inicial será:

$$P^{(0)} = (1/2, 1/2)$$

Calculamos a matriz:

$$P^4 = \begin{pmatrix} \frac{3}{8} & \frac{5}{8} \\ \frac{5}{16} & \frac{11}{16} \end{pmatrix}$$

Polo tanto, a distribución de probabilidade despois de catro días será:

$$P^{(4)} = P^{(0)} \cdot P^4 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} \frac{3}{8} & \frac{5}{8} \\ \frac{5}{16} & \frac{11}{16} \end{pmatrix} = \begin{pmatrix} \frac{11}{32} & \frac{21}{32} \end{pmatrix}$$

Esta distribución indícanos que despois de catro días, a probabilidade de ir pescar é $11/32$ e a probabilidade de traballar no campo é $21/32$.

Exercicio

10. Calcúlese a distribución de probabilidades absolutas despois de catro pasos nos procesos de Markov descritos nos exercicios 8 e 9.

Cadeas de Markov regulares

Consideremos as cadeas de Markov dadas polas súas matrices de transición:

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad B = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ 0 & 1 \end{pmatrix}$$

Ambas as dúas corresponden a procesos de Markov de dous estados. Na cadea cuxa matriz é A, pódese pasar dun estado ao outro do sistema indefinidamente e a intervalos non regulares. Na cadea con matriz B, unha vez que o proceso entra no estado a_2 xa non é posíbel saír del. Dícese que este estado é un **estado absorbente**.

Analícemos o proceso con tres estados correspondente á matriz:

$$C = \begin{pmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{5} & 0 & \frac{4}{5} \\ \frac{3}{4} & \frac{1}{4} & 0 \end{pmatrix}$$

Neste caso, é posíbel pasar dun estado a outro indefinidamente sen que os pasos se repitan ciclicamente, análogamente a como sucedía no proceso da matriz A.

Consideremos por último o proceso dado pola matriz seguinte:

$$D = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Neste caso podemos tamén pasar dun estado a outro sen ningún tipo de atranco, sen embargo, os pasos por un estado determinado repítense a intervalos periódicos de amplitude 3.

As cadeas de Markov que teñen a propiedade de A e C , isto é, nas que é posíbel pasar dun estado a outro a través de todos os estados da cadea sen que este paso se realice de forma cíclica, reciben o nome de **cadeas de Markov regulares** e as matrices de transición correspondentes **matrices de transición regulares**.

Exercicios

11. Dadas as seguintes matrices estúdese se son matrices de transición regulares.

$$A = \begin{pmatrix} \frac{1}{7} & \frac{3}{7} & \frac{3}{7} \\ 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{4}{5} \end{pmatrix} \quad C = \begin{pmatrix} \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{pmatrix} \quad D = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & \frac{2}{5} & \frac{3}{5} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

12. Dísqutese a seguinte proposición: se unha matriz estocástica P corresponde a un proceso de Markov regular, é dicir, á unha matriz de transición regular, entón algunha potencia de P debe de ter todos os elementos distintos de cero.

Punto fixo dunha matriz de transición regular

Consideremos a matriz regular seguinte:

$$P = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Buscamos un vector de probabilidade $\mathbf{v}$ de dúas compoñentes tal que verifique :

$$\mathbf{v} \cdot \mathbf{P} = \mathbf{v}$$

Sexa $\mathbf{v} = (x, 1-x)$. Terá que verificarse a ecuación matricial:

$$(x \quad 1-x) \cdot \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (x \quad 1-x)$$

Multiplicndo e igualando elementos temos o sistema:

$$\begin{cases} \frac{1}{2} - \frac{1}{2}x = x \\ x + \frac{1}{2} - \frac{1}{2}x = 1-x \end{cases}$$

En calquera das dúas ecuacións atopamos $x = \frac{1}{3}$. Así, o vector buscado é:

$$v = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix},$$

que recibe o nome de **vector fixo de probabilidade da matriz regular P** .

Exercicio

13. Calcúlense os vectores fixos das matrices regulares do exercicio 12.

Distribución estacionaria

Consideremos a matriz regular:

$$P = \begin{pmatrix} 0 & 1 \\ \frac{2}{5} & \frac{3}{5} \end{pmatrix}$$

Mediante unha folla de cálculo, utilizando o método descrito anteriormente, podemos calcular as potencias sucesivas da matriz P . Para maior facilidade de cálculo escribimos os elementos fraccionarios en forma decimal.

$$P = \begin{pmatrix} 0 & 1 \\ 0,4 & 0,6 \end{pmatrix}$$

As potencias sucesivas son as seguintes:

$$P^2 = \begin{pmatrix} 0,4 & 0,6 \\ 0,24 & 0,76 \end{pmatrix} \quad P^3 = \begin{pmatrix} 0,24 & 0,76 \\ 0,304 & 0,696 \end{pmatrix} \dots\dots P^5 = \begin{pmatrix} 0,2784 & 0,7216 \\ 0,2886 & 0,7113 \end{pmatrix}$$

$$P^7 = \begin{pmatrix} 0,2845 & 0,7154 \\ 0,2861 & 0,7138 \end{pmatrix} \dots\dots P^{10} = \begin{pmatrix} 0,2857 & 0,7142 \\ 0,2856 & 0,7143 \end{pmatrix} \dots\dots P^{20} = \begin{pmatrix} 0,285714 & 0,714286 \\ 0,285714 & 0,714286 \end{pmatrix}$$

Observamos que conforme aumentan as potencias, os elementos das matrices estabilízanse nuns valores determinados, e que as dúas filas da matriz se aproximan até coincidir nas seis primeiras cifras.

Calculemos agora o vector fixo da matriz P :

$$(x \quad 1-x) \cdot \begin{pmatrix} 0 & 1 \\ \frac{2}{5} & \frac{3}{5} \end{pmatrix} = (x \quad 1-x)$$

Resolvendo a ecuación matricial atopamos:

$$v = \begin{pmatrix} \frac{2}{7} & \frac{5}{7} \end{pmatrix} = (0,28571428..., 0,71428571...)$$

Vemos agora que os valores aos que se aproximan as filas da matriz son precisamente as compoñentes do vector fixo de P .

Estes resultados poden ser xeneralizados no seguinte teorema:

Teorema 3. Se P é unha matriz de transición regular, as potencias sucesivas de P aproxímanse a unha matriz que ten todas as filas iguais e que coinciden coas compoñentes do vector fixo de P .

Exercicios

14. Compróbase o teorema anterior coas matrices regulares do exercicio 12.

Consideremos agora o vector de probabilidade inicial:

$$P^{(0)} = (p_1^{(0)}, p_2^{(0)}, \dots, p_n^{(0)})$$

Sabemos que a distribución absoluta correspondente ao paso k vén dada pola relación:

$$P^{(k)} = P^{(0)} \cdot P^k,$$

produto que despois dun número grande de pasos, se aproximará ao vector fixo v .

$$P^{(k)} = v$$

Ou, o que é o mesmo:

$$(p_1^{(k)}, p_2^{(k)}, \dots, p_n^{(k)}) = (v_1, v_2, \dots, v_n)$$

Así, podemos enunciar:

Teorema 4. Consideremos unha cadea de Markov regular. A probabilidade de que despois dun número grande de probas un estado determinado da cadea a_i suceda, é igual aproximadamente á compoñente i -ésima v_i do vector fixo da matriz de transición do proceso. Polo tanto, as probabilidades absolutas de que ocorra un determinado estado son independentes das condicións iniciais do sistema.

A distribución de probabilidade que define o vector fixo chámase **distribución de probabilidade estacionaria** da cadea de Markov.

Exemplo 5

Unha persoa ten constatado que se un día está arrefriado está un 70% seguro de que non o estará ao día seguinte, pero se certo día non está arrefriado ten un 60% de probabilidade de non o estar tampouco ao seguinte día. Ao longo dos anos, con que frecuencia estará arrefriado? A matriz de transición desta cadea de Markov é:

$$P = \begin{pmatrix} 0,3 & 0,7 \\ 0,4 & 0,6 \end{pmatrix}$$

Buscamos o vector fíxo:

$$(x \quad 1-x) \cdot \begin{pmatrix} 0,3 & 0,7 \\ 0,4 & 0,6 \end{pmatrix} = (x \quad 1-x)$$

e obtemos o valor do vector fíxo que nos dá a distribución estacionaria.

$$v = \begin{pmatrix} \frac{4}{11} & \frac{7}{11} \end{pmatrix}$$

Así, aproximadamente o 36% dos días a persoa estará arrefriada.

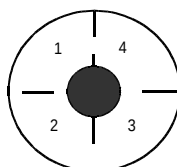
Exercicios

15. Atópanse as distribucións estacionarias correspondentes ás cadeas de Markov coas matrices de transición:

$$A = \begin{pmatrix} 0,5 & 0,5 \\ 0,25 & 0,75 \end{pmatrix} \quad B = \begin{pmatrix} 0,5 & 0,25 & 0,25 \\ 0 & 0 & 1 \\ 0,5 & 0,5 & 0 \end{pmatrix} \quad C = \begin{pmatrix} 0,125 & 0,375 & 0,5 \\ 0 & 0,5 & 0,5 \\ 0,75 & 0,25 & 0 \end{pmatrix}$$

16. Un estudante vai ao instituto andando ou en bus. Cando vai andando un día, a probabilidade de que ao día seguinte colla o bus é 0,6. A probabilidade de que vaia en bus dous días seguidos é 0,7. Aproximadamente con que frecuencia colle o bus ao longo de todo o curso?
17. Un supermercado realiza a experiencia seguinte en relación ás preferencias dos seus clientes. Obsérvase que o 80% das persoas que compran un día o produto A repite ao día seguinte. O 60% dos que non compran o produto A un día, mércano ao día seguinte. Se o 50% comprou o produto un día determinado, que podemos predicir para a compra do produto o segundo día? E para o terceiro? E para o milésimo?
18. Un estudo de audiencia demostra que se unha persoa ve certo *culebrón* de televisión unha tarde está un 30% segura de que tamén o verá a tarde seguinte, pero se non o ve unha tarde hai un 60% de posibilidades de que non o vexa tampouco a tarde seguinte. Que frecuencia de audiencia terá o *culebrón*?
19. Nunha cidade existen dous partidos políticos, un de dereitas e outro de esquerdas. Os alcaldes son elixidos por un período de catro anos e tense observado que a probabilidade de que a un alcalde de dereitas suceda outro do mesmo signo político é $\frac{3}{5}$ e que a un alcalde de esquerdas siga outro de esquerdas é $\frac{1}{2}$. Supoñamos que en 2000 houbo un alcalde de esquerdas. Cal é a probabilidade de que 2008 siga un alcalde de esquerdas ao fronte do municipio? Que ocorrerá no ano 2020?
20. Unha urna contén 2 bólas brancas e outra 3 bólas negras. Escollemos ao azar unha bóla de cada urna e depositámolas na outra. Consideremos o número de bólas negras na urna A para cada paso do experimento. Calcúlese a matriz de transición. Cal é a probabilidade de que a urna A conteña dúas bólas negras despois de tres pasos? Se repetimos a proba un número grande de veces, cal é a probabilidade de que haxa 2 bólas negras na urna A?

21. Unha familia planifica cada ano as súas vacacións de tal xeito que se un ano vai á montaña ao ano seguinte vai ao mar, e ao segundo ano descansa na casa. Pero, ao ano seguinte é igualmente probábel que se traslade ao mar ou á montaña. En 2010 quedarán na casa. Onde é máis probábel que pasen as súas vacacións no ano 2015? Á larga, con que frecuencia pasarán as súas vacacións no mar?
22. No diagrama seguinte aparece un labirinto no que é posíbel atravesar cada unha das portas con igual probabilidade.



Unha persoa situada inicialmente no compartimento 1 móvese de un a outro sen interrupción. Escríbase a matriz de transición da cadea de Markov correspondente, e calcúlese a probabilidade de que despois de 4 pasos estea no compartimento número 3.

23. Se Pau Gasol encesta, a continuación faino Kobe Bryant, pero cando encesta Bryant, Gasol ten o dobre de probabilidade que Bryant de encestar a continuación. Quen ten mellor promedio?
24. Nunha cadea de interruptores eléctricos a probabilidade de que un deles estea aberto é $\frac{2}{5}$ se o precedente tamén o está, e $\frac{1}{3}$ se o interruptor precedente está pechado. Se abrimos ou pechamos o primeiro interruptor da cadea ao azar, calcúlese a probabilidade de que o cuarto interruptor estea aberto.
25. Supoñamos dúas ruletas con seis números cada unha delas. A ruleta A ten 4 números pares e 2 impares e a ruleta B, 1 un número par e 5 impares. Botamos ao ar unha moeda corrente; se sae cara facemos xirar a ruleta A e se sae cruz a B. Se aparece número par volvemos a xirar a ruleta A e se sae impar a B. O proceso repítese indefinidamente. Calcúlese:

- A probabilidade de que no segundo ensaio fagamos xirar a ruleta B.
- A probabilidade de que apareza número par no cuarto ensaio.
- A frecuencia con que aparecerá número impar tras moitos ensaios.
- Supoñamos que a moeda que botamos inicialmente está trucada de xeito que non coñecemos as probabilidades de sacar cara e cruz. Estímese a frecuencia de que saia número impar se o proceso se repite indefinidamente.

4.DISTRIBUCIÓNS DE PROBABILIDADE

No canto de estudar os posíbeis modelos que xeran aleatoriedade que non sexa do tipo de curva de campá, a miña teoría consiste en facer o contrario: coñecer a curva de campá tan fondamente como o dea feito para determinar onde se pode aplicar e onde non. [...] A ubicuidade da campá de Gauss non é unha propiedade do mundo, senón un problema da nosa mente que xorde do xeito en que contemplamos aquel.

O cisne negro. Nassim Nicholas Taleb.

Variábel aleatoria

Para adxudicar probabilidades á maioría dos experimentos aleatorios cómpre asociar a cada elemento do conxunto de resultados, Ω , un número real. Esta asociación realízase do xeito habitual en matemáticas; mediante unha aplicación.

Unha aplicación que fai corresponder a cada elemento de Ω un número real, chámase **variábel aleatoria**.

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \mathfrak{R} \\ r & \longrightarrow & X(r) \end{array}$$

En rigor, para que unha aplicación como a anterior sexa variábel aleatoria debe cumprirse que a imaxe inversa de cada intervalo da recta real sexa un acontecemento do espazo das mostras de Ω . Isto que se cumpre para calquera función se Ω é discreto, non sempre acontece se o conxunto de resultados é infinito non numerábel.

Se o conxunto imaxe $X(\Omega)$ é finito ou infinito numerábel a variábel aleatoria dise *discreta*, se é infinito non numerábel, *continua*.

Para indicar a probabilidade do acontecemento elemental cuxa imaxe é a , utilízase a notación: $P(X=a)$.

Analogamente, poñamos por caso, por $P(X \leq a)$ referirémonos a $P(\{r / X(r) \leq a\})$, e por $P(a \leq X \leq b)$ a $P(\{r / a \leq X(r) \leq b\})$.

Exemplo 1

Consideremos o experimento aleatorio que consiste en botar un dado. Sobre o conxunto de resultados $\Omega = \{1,2,3,4,5,6\}$ podemos definir a variábel aleatoria: $X(r) = 2r$. O seu conxunto imaxe será $\{2,4,6,8,10,12\}$, daquela:

$$P(X \leq 8) = P(\{1,2,3,4\}) = 4/6 \text{ e } P(4 \leq X \leq 9) = P(\{2,3,4\}) = 3/6.$$

Sobre o mesmo espazo das mostras podemos definir outra variábel aleatoria, por exemplo $Y(r) = \begin{cases} 0, & \text{se } r \text{ par} \\ 1, & \text{se } r \text{ impar} \end{cases}$

Neste caso o conxunto imaxe será $\{0,1\}$ e, por exemplo, $P(Y = 0) = P(\{2,4,6\}) = 3/6$ e $P(0 \leq X \leq 1) = P(\Omega) = 1$.

Exemplo 2

Extraemos ao azar tres bólas dunha urna que contén cinco bólas vermellas e tres azuis. Consideremos a variábel aleatoria $X = \text{"número de bólas azuis en cada extracción"}$.

Este experimento presenta $\binom{8}{3} = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} = 56$ resultados posíbeis

Non aparecerá bola azul en $\binom{5}{3} = 10$ casos.

Aparecerá unha bola azul en $\binom{3}{1} \binom{5}{2} = 30$ casos.

Aparecerán dúas bolas azuis en $\binom{3}{2} \binom{5}{1} = 15$ casos.

Aparecerán tres bolas azuis en $\binom{3}{3} = 1$ caso.

O conxunto imaxe da variábel aleatoria sería $X(\Omega) = \{0,1,2,3\}$ e, por exemplo,

$$P(X = 3) = 1/56,$$

$$P(X \leq 1) = 10/56 + 30/56 = 40/56,$$

$$P(1 \leq X \leq 2) = 30/56 + 15/56 = 45/56$$

Exercicio 1

Dunha baralla, facéndolle corresponder a cada carta a puntuación do tute (as 11, tres 10, rei 4, cabalo 3, sota 2, as restantes cero), tiramos unha carta ao azar.

Defínase formalmente a correspondente variábel aleatoria e calcúlense as probabilidades:

$$P(X = 3), P(X \leq 9), P(3 \leq X \leq 10), P(X \geq 8)$$

Función de probabilidade

Se X é unha variábel aleatoria finita con conxunto imaxe $X(\Omega) = \{a_1, a_2, \dots, a_n\}$, podemos definir a función $f(a_i) = P(X = a_i)$, para cada a_i . Esta función recibe o nome de *función de probabilidade*.

Se X é a variábel aleatoria do exemplo 1, a súa función de probabilidade viría dada por:

a_i	2	4	6	8	10	12
$f(a_i)$	1/6	1/6	1/6	1/6	1/6	1/6

Para a variábel Y do mesmo exemplo teríamos:

a_i	2	4
$f(a_i)$	1/2	1/2

As funcións de probabilidade no caso finito adóitanse representar mediante taboas como as anteriores.

A variábel aleatoria do exemplo 2 tería a seguinte función de probabilidade:

a_i	0	1	2	3
$f(a_i)$	10/56	30/56	15/56	1/56

Obsérvese que tal como foi definida a función de probabilidade debe verificar:

$$1) f(a_i) \geq 0 \text{ e } 2) \sum_{i=1}^n f(a_i) = 1$$

Se o conxunto de valores de X é infinito numerábel a definición anterior pode extenderse mudando a segunda propiedade por $\sum_{i=1}^{\infty} f(a_i) = 1$, é dicir: $\lim_{n \rightarrow \infty} \sum_{i=1}^n f(a_i) = 1$.

Exercicios .

1. Descríbase mediante unha táboa a función de probabilidade da variábel aleatoria do exercicio 1.
2. Botamos unha moeda até que apareza cruz. Sexa X = “número de lanzamentos”. Escribase a función de probabilidade de X e compróbase que cumpre as propiedades anteriores.
- 3.

Función de distribución discreta

Sexa X unha variábel aleatoria discreta tal que $X(\Omega) = \{a_1, a_2, \dots, a_n\}$. Definimos unha nova función $F: \mathfrak{R} \rightarrow \mathfrak{R}$ tal que $F(x) = P(a_i \leq x) = \sum_{a_i \leq x} f(a_i)$, que recibe o nome de *función de distribución acumulativa de X* .

Supoñamos que as imaxes de X estan ordenadas: $a_1 \leq a_2 \leq \dots \leq a_n$, entón:

$$\begin{aligned} x \in (-\infty, a_1) &\Rightarrow F(x) = 0 \\ x \in [a_1, a_2) &\Rightarrow F(x) = f(a_1) \end{aligned}$$

$$x \in [a_2, a_3) \Rightarrow F(x) = f(a_1) + f(a_2)$$

....

$$x \in [a_{n-1}, a_n) \Rightarrow F(x) = f(a_1) + f(a_2) + \dots + f(a_{n-1})$$

$$x \in [a_n, +\infty) \Rightarrow F(x) = f(a_1) + f(a_2) + \dots + f(a_{n-1}) + f(a_n) = 1$$

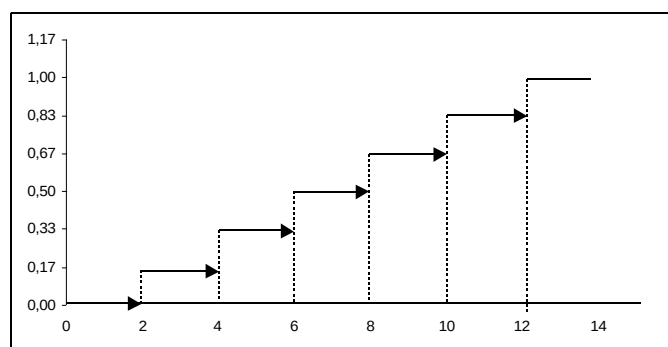
A función de distribución resulta ser unha función en escaleira tal que:

- 1) É monótona crecente xa que $f(a_i) \geq 0$, e polo tanto $F(x) \leq F(y)$ se $x \leq y$.
- 2) $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow \infty} F(x) = 1$

Exemplo 4

Debuxemos a gráfica da función de distribución da variábel aleatoria do exemplo 1.

a_i	2	4	6	8	10	12
$f(a_i)$	1/6	1/6	1/6	1/6	1/6	1/6



Exercicios

4. Unha caixa contén tres bólas brancas, dúas azuis e unha negra. Extraemos ao azar unha bóla de cada caixa. Sexa X unha variábel aleatoria que toma o valor 1 se a bóla é branca, 2 se é azul e 3 se é negra. Escribíbase as funcións de probabilidade e distribución de X.

5. Sexa X unha variábel aleatoria que toma os valores 0, 1 e 3, e que ten como función de probabilidade $f(a_i) = \frac{k}{a_i + 3}$

- a) Déterminese o valor de k .
- b) Escribíbase a función de probabilidade
- c) Debúxese a función de distribución.

Esperanza matemática

Se X é unha variábel aleatoria que toma os valores $a_1, a_2, \dots, a_n$, definimos o seu *valor esperado* ou *esperanza matemática* por:

$$E(x) = a_1 f(a_1) + a_2 f(a_2) + \dots + a_{n-1} f(a_{n-1}) + a_n f(a_n)$$

No caso de que variábel discreta tomase valores nun conxunto infinito numerábel, teríamos:

$$E(x) = \lim_{n \rightarrow \infty} \sum_{i=1}^n a_i f(a_i).$$

Este número dános o promedio dos valores que toma X ponderados polos valores da función de probabilidade en cada un deles. A esperanza tamén se chama media teórica, xa que non deixa de ser unha extensión da media frecuencial, mais neste caso calculada a priori. En xeral, para designar a esperanza utilízase a letra grega μ . En ocasións a esperanza, o mesmo que ocorría coa media frecuencial, non corresponde a un valor da variábel. Compróbo no seguinte exercicio.

Exercicio 5

Calcúlese a esperanza da variábel aleatoria correspondente ao experimento botar un dado e adxudicarlle a puntuación obtida.

Exemplo 5

Calcúlese o número esperado de caras ao botar tres moedas.

Sobre o espazo das mostras, onde 1 representa saír cara e 0 saír cruz, definimos a variábel aleatoria X = “número de caras”. A táboa seguinte danos a súa función de probabilidade:

a_i	0	1	2	3
$f(a_i)$	1/8	3/8	3/8	1/8

A esperanza será:

$$E(x) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1,125$$

Exemplo 6

Nun posto dunha feira estímase que as ganancias poden ser de 100 euros se o día é de choiva e de 600 no caso contrario. Se a probabilidade de que chova é 0,6, calcúlese a ganancia esperada.

A ganancia esperada será $E(X) = 0,6 \cdot 100 + 0,4 \cdot 600 = 300$ €.

Varianza

A maior ou menor dispersión dos valores de X respecto á esperanza ven dada pola *varianza*, definida por:

$$V(x) = E((x - \mu)^2) = \sum_{i=1}^n (a_i - \mu)^2 f(a_i)$$

No caso de tomar a variábel discreta valores nun conxunto infinito numerábel, teríamos:

$$V(x) = \lim_{n \rightarrow \infty} \sum_{i=1}^n (a_i - \mu)^2 f(a_i).$$

A raíz cadrada da varianza chámase *desvío típico*, ou *desvío padrón*, e designase habitualmente pola letra grega σ . Daquela e común representar a varianza por σ^2 .

Cando non se dispoña dunha folla de cálculo, hai algúns teoremas que facilitan o cálculo da varianza. Enunciaremos a continuación algúns deles:

Teorema 1.

$$V(x) = E(x^2) - E(X)^2 = \sum_{i=1}^n a_i^2 f(a_i) - \left[\sum_{i=1}^n a_i f(a_i) \right]^2$$

Exemplo 7

Cos datos do exemplo 5 teríamos:

$$E(X^2) = 0 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8} = 3$$

$$E(X)^2 = 1,125^2 = 1,2656$$

$$\text{Logo, } V(X) = E(X^2) - E(X)^2 = 3 - 1,2656 = 1,7344$$

Teorema 2.

Se X é unha variábel aleatoria e b un número real, verifícase:

$$a) E(X+b) = E(X) + b$$

$$b) V(X+b) = V(X)$$

$$c) E(bX) = bE(X)$$

$$d) V(bX) = b^2 V(X)$$

Exercicios

7. Dada a variábel aleatoria X con función de probabilidade:

a_i	1	3	5	7
$f(a_i)$	0,2	0,5	0,1	0,2

Calcúlese a esperanza, a varianza e o desvío típico de X .

8. Demóstranse os teoremas 1 e 2.

9. Unha moeda esta amañada de xeito que a probabilidade de saír cara e dobre da de saír cruz. Botamos a moeda até que aparezan catro cruces ou unha cara. Calcúlese o número esperado de lanzamentos da moeda.

Distribución binomial

Consideremos un experimento aleatorio que pode representar dous resultados excluíntes E (éxito) e F (fracaso) de maneira que $P(E) = p$ e $P(F) = q = 1-p$.

Se realizamos n probas repetidas e independentes deste experimento, en cada unha delas obteremos o resultado E ou o F . Defínese así un espazo das mostras onde cada resultado virá dado por unha n -tupla ordenada con compoñentes E ou F . Por exemplo, un dos posíbeis resultados sería $EEFFF...FE$.

Sobre este espazo podemos considerar a variábel aleatoria X que fai corresponder a cada un dos seus elementos o número de éxitos obtidos. Temos así unha variábel aleatoria discreta que recibe o nome de *binomial*.

Se un resultado contén un número k de éxitos e, polo tanto, un número $n-k$ de fracasos, dado que as probas son independentes, a súa probabilidade será:

$$P(\underbrace{EE \dots}_{k \text{ veces}} \underbrace{EF \dots}_{n-k \text{ veces}} F) = \underbrace{pp \dots}_{k \text{ veces}} \underbrace{pq \dots}_{n-k \text{ veces}} q = p^k q^{n-k}$$

Imos calcular a función de probabilidade desta variábel aleatoria que toma os valores $0, 1, 2, \dots, n$.

$$f(0) = P(X = 0) = P(\underbrace{FF \dots}_{n \text{ veces}} F) = q^n$$

A variábel aleatoria tomará o valor 1 para todos os acontecementos que conteñan un éxito, e como hai n formas de colocar o resultado E nunha n -tupla, teremos:

$$f(1) = P(X = 1) = npq^{n-1}$$

Para $X = 2$ o resultado presentará dous éxitos e $n-2$ fracasos, e tendo en conta o número de maneiras que se poden escoller dous lugares entre n , obtense:

$$f(2) = P(X = 2) = \binom{n}{2} p^2 q^{n-2}$$

En xeral, para $X = k$, k éxitos e $n-k$ fracasos, obtense a expresión xenérica para a función de probabilidade:

$$f(k) = P(X = k) = \binom{n}{k} p^k q^{n-k}$$

Exemplo 8

Escollidas seis persoas ao azar, calcúlese a probabilidade de que:

- Haxa tres homes e tres mulleres
- Haxa polo menos 5 homes
- Haxa polo menos unha muller

Podemos representar o experimento aleatorio como a repetición dunha proba de xeito independente seis veces, cos resultados posíbeis: muller ou home, con probabilidade 0,5 para cada caso. Tomando como variábel aleatoria o número de mulleres, e tendo en conta que se trata dunha variábel aleatoria binomial con $n = 6$ e $p =$

$$0,5, \text{ temos: } P(X = 3) = \binom{6}{3} \cdot 0,5^3 \cdot 0,5^3 = 0,3125$$

$$P(X = 0) + P(X = 1) = \binom{6}{0} \cdot 0,5^0 \cdot 0,5^6 + \binom{6}{1} \cdot 0,5 \cdot 0,5^5 = 0,109375$$

Teñamos en conta que $P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{6}{0} \cdot 0,5^6 = 0,984375$

Para valores de n suficientemente grandes, ou valores de p con decimais abondo, os cálculos das probabilidades binomiais fanse excesivamente laboriosos. Para evitalos existen táboas que facilitan estes cálculos, mais na actualidade adoitan facerse utilizando a computadora xa que cunha simple folla de cálculo podemos construír a táboa que queiramos.

A disposición sería a seguinte:

	A	B	C	D
1	Nº n de probas	Valor de p	Nº k de éxitos	P(X = k)
2	10	0,4	3	0,214991

Na cela D2 pegamos a función :

$$= \text{COMBINAT}(A2; C2) * B2^C2 * (1 - B2)^(A2 - C2)$$

Que é a expresión da función de probabilidade de X utilizando as prestacións da folla de cálculo.

No exemplo temos a probabilidade de acadar 3 éxitos nunha distribución binomial de 10 probas con probabilidade de éxito 0,4.

Introducindo os valores correspondentes nas celas A2, B2 e C2, podemos calcular as probabilidades que desexemos.

Existe outra función da folla de cálculo que nos proporciona o resultado anterior directamente: DISTR.BINOM.

No cadro de diálogo que aparece escribimos o número de éxitos, nº de probas (ensaos) e probabilidade de éxito, sinalando “falso” na caixa de “Acumulado”. Porén, se poñemos “verdadeiro”, obteremos o valor da suma das probabilidades de éxito dende cero até o valor sinalado ao principio; é dicir, o valor da función de distribución acumulativa de X:

$$F(k) = P(X \leq k) = \sum_{i=0}^k \binom{n}{i} p^i q^{n-i}$$

Calculemos por exemplo, a probabilidade de que de 30 nacementos 16 sexan nenas, sendo a probabilidade de nacer nena 0,48339259 e a de nacer neno 0,51660741.

Poñemos:

$$N = 30 ; k = 16 ; p = 0,48339259$$

$$= \text{COMBINAT}(30;16)* 0,48339259 ^{16}*0,51660741^{14}= 0,1246467.$$

Calculemos agora a probabilidade de que de 30 nacementos 16 como máximo sexan nenas, é dicir, $p(X \leq 16)$. Podemos repetir o proceso anterior:

$$P(X = 0) + P(X = 1) ++P(X = 16)$$

ou ben utilizar a función:

$$= \text{DISTR.BINOM}(16;30;0,48339259;\text{VERDADERO}) = 0,767299$$

Esperanza e Varianza da distribución Binomial.

Sabemos que a esperanza dunha variábel aleatoria finita ven dada por:

$$\mu = E(X) = \sum_0^n x_i f(x_i),$$

onde $f(x_i)$ é a función de probabilidade : $f(x_i) = P(X = x_i)$

Defínese a **varianza** da variábel por:

$$\sigma^2 = V(X) = \sum (x_i - \mu)^2 f(x_i)$$

O desvío típico da variábel defínese por :

$$\sigma = \sqrt{V(X)} = \sqrt{\sum (x_i - \mu)^2 f(x_i)}$$

Na seguinte folla realizamos os cálculos:

	A	B	C	D
1	x_i	$f(x_i)=P(X=x_i)$	$x_i \cdot f(x_i)$	$(x_i - \mu)^2 \cdot f(x_i)$
2	0	0,000470185	0	0,016926659
3	1	0,00470185	0,00470185	0,117546246
4	2	0,021941966	0,043883932	0,351071455
5	3	0,063387902	0,190163705	0,570491115
6	4	0,126775803	0,507103213	0,507103213
7	5	0,185937845	0,929689224	0,185937845
8	6	0,206597605	1,239585632	0
9	7	0,177083662	1,239585632	0,177083662
10	8	0,118055774	0,944446196	0,472223098
11	9	0,061214105	0,550926947	0,550926947
12	10	0,024485642	0,244856421	0,391770274
13	11	0,007419892	0,081618807	0,185497289
14	12	0,001648865	0,019786377	0,059359132
15	13	0,000253672	0,00329773	0,012429904
16	14	2,41592E-05	0,000338229	0,001546188
17	15	1,07374E-06	1,61061E-05	8,69731E-05
18	Σ	1	$\mu = 6$	$\sigma^2 = 3,6$

Na cela B2 pegamos DISTR.BINOM(A2;15;0,4:Falso) e extendemos, utilizando o comando ENCHER CARA ABAIXO, até B17 para obter os valores da función de probabilidade. Se sumamos todos estos valores na cela B18 nos debe dar 1.

Na cela C2 escribimos a fórmula $=A2*B2$ e extendemos outra vez até C17. Na cela C18 sumamos dende C2 até C17 e obtemos o valor da esperanza..

En D2 escribimos $= (A2 - 6)^2*B2$ e unha vez máis extendemos até D17. Sumando en D18 obtemos o valor da varianza.

Na taboa comprobamos que:

$$E(x) = 15 \cdot 0,4 = 6 ; \quad V(X) = 15 \cdot 0,4 \cdot 0,6 = 3,6 ; \quad \sigma = \sqrt{15 \cdot 0,4 \cdot 0,6} = 1,8973666$$

que coinciden cos valores teóricos en calquera distribución binomial, na que sempre se cumpre: $E(x) = np$, $V(x) = npq$, $\sigma = \sqrt{npq}$.

No exemplo 8 observamos:

$$E(x) = np = 6 \cdot 0,5 = 3 \quad \text{e} \quad \sigma = \sqrt{npq} = \sqrt{6 \cdot 0,5 \cdot 0,5} = 1,2247$$

Exercicios

10. Nunha poboación na que o 47% ten grupo sanguíneo A, escóllense ao azar 5 persoas, cal é a probabilidade de que:
 - a) Ningunha teña ese grupo sanguíneo?
 - b) Tres o teñan?
 - c) Polo menos dúas o teñan?
11. Botamos dous dados 100 veces, calcúlese a frecuencia esperada de que sumen 5.

Distribución de Poisson

A variábel aleatoria de Poisson é unha variábel discreta infinita que está asociada a experimentos nos que os “éxitos” buscados se expresan por unidade de área, tempo, etc. Por exemplo:

- nº de defectos dun pavimento por m^2 .
- nº de avións que toman terra nun aeroporto por día, hora, minuto,
- nº de erratas por páxina
- nº de chamadas telefónicas por hora, minuto, ...
- nº de chegadas de trasatlánticos a un porto por día, mes, etc, etc.

A función que nos dá a probabilidade de que ocorran x “éxitos” por unidade de tempo, área, etc, ven dada por:

$$f(k) = P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{con } k=1, 2, 3, \dots, \text{ sendo } \lambda \text{ unha constante maior}$$

que cero.

Existen táboas para a función de probabilidade da distribución de Poisson, de todas formas é doado construír unha utilizando as funcións da folla de cálculo dun xeito análogo a como fixemos no caso da binomial.

	A	B	C
1	Valor de λ	Nº k de éxitos	P(X=k)
2	0,8	5	0,001226968

Na cela A2 introducimos o valor do parámetro λ da distribución. En B2 o número de éxitos . En C2 pegamos a fórmula :

$$= \text{EXP}(-A2)*A2^B2 / \text{FACT} (B2)$$

que nos dará a probabilidade de obter k éxitos.

No caso anterior puxemos $\lambda = 0,8$ e $x = 5$, obtendo $P(X = 5) = 0,001226968$.

Esperanza e varianza da distribución de Poisson

Partimos da seguinte táboa:

	A	B	C	D	E
1	Valor de λ	k	f(k)=P(X = k)	k · f(k)	(k- μ) ² · f(k)
2	0,8	0	0,449328964	0	0,287570537
3	0,8	1	0,359463171	0,359463171	0,014378527
4	0,8	2	0,143785269	0,287570537	0,207050787
5	0,8	3	0,038342738	0,115028215	0,185578853
6	0,8	4	0,007668548	0,030674191	0,078525928
7	0,8	5	0,001226968	0,006134838	0,021643709
8	0,8	6	0,000163596	0,000981574	0,004423627
9	0,8	7	1,86966E-05	0,000130877	0,000718699
10	0,8	8	1,86966E-06	1,49573E-05	9,69234E-05
11	0,8	9	1,66192E-07	1,49573E-06	1,11748E-05
12	0,8	10	1,32954E-08	1,32954E-07	1,12532E-06
13	0,8	11	9,66938E-10	1,06363E-08	1,006E-07
14	0,8	12	6,44625E-11	7,7355E-10	8,08618E-09
15	0,8	13	3,96692E-12	5,157E-11	5,90437E-10
16	0,8	14	2,26681E-13	3,17354E-12	3,9497E-11
17	0,8	15	1,20897E-14	1,81345E-13	2,43776E-12
18	0,8	16	6,04484E-16	9,67174E-15	1,3966E-13
19	0,8	17	2,84463E-17	4,83587E-16	7,46545E-15
20			1	0,8	0,8

Na cela A1 escribimos o valor $\lambda = 0,8$.

Na cela B1 escribimos 0 e na cela B2 a fórmula: $= B1+1$, de xeito que se extendemos cara abaixo temos a sucesión dos numeros naturais.

Na cela C1 pegamos a fórmula da función de probabilidade : $f(k) = e^{-\lambda} \frac{\lambda^k}{k!}$,
que traducida á folla de cálculo sería :

$$=EXP(-A1)*A1^B1/Fact(B1)$$

Dado que se trata dunha variábel discreta infinita deberíamos estender esta columna indefinidamente, mais observamos que a partir da cela 19 os valores son case nulos e a suma de todos eles é 1. A computadora prescinde do resto e redondea a 1.

Na cela D1 escribimos a fórmula: = B1*C1, para calcular os produtos dos valores da variábel pola súa función de probabilidade.

Na cela E20 sumamos todos os resultados da columna para obter a media:

$$= SUMA(D2:D19)$$

Na cela E1 Escribimos: = (B1- \$D\$20)^2*C1, expresión da varianza. É importante observar que utilizamos o signo \$ para que o valor de D20, que corresponde á media, non varie ao estender a fórmula.

En E20 calculamos a suma de todos os valores da columna para obter a varianza.

Observamos que tanto o valor da media como o da varianza é 0,8, coincidente co valor do parámetro λ da distribución. Pódese deducir, que unha variábel de Poisson sempre cumpre:

$$\mu = E(X) = \lambda \qquad \sigma^2 = V(X) = \lambda.$$

Aproximación da distribución binomial pola de Poisson.

Se realizamos un número grande de probas dunha distribución binomial con probabilidade de éxito próxima a cero, as probabilidades binomiais e de Poisson aproxímanse moito.

Demóstrase que:

$$\lim_{n \rightarrow \infty} \binom{n}{x} p^x q^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}, \text{ sendo } \lambda = np$$

Comprobemos o anterior utilizando a folla de cálculo. Comparamos os valores das funcións de probabilidade dunha binomial con $n = 20$ e $p = 0,05$ coa correspondente de Poisson $\lambda = np = 20 \cdot 0,05 = 1$. Resulta a seguinte táboa:

	A	B	C
1	X_i	Binomial(20;0,05)	Poisson $\lambda = 1$
2	0	0,358486	0,367879
3	1	0,377354	0,367879
4	2	0,188677	0,183940
5	3	0,059582	0,061313
6	4	0,013328	0,015328
7	5	0,002245	0,003066
8	6	0,000295	0,000511
9	7	0,000031	0,000073
10	8	0,000003	0

En B2 pegamos a función: =DISTR.BINOM(A2;20;0,05;FALSO)

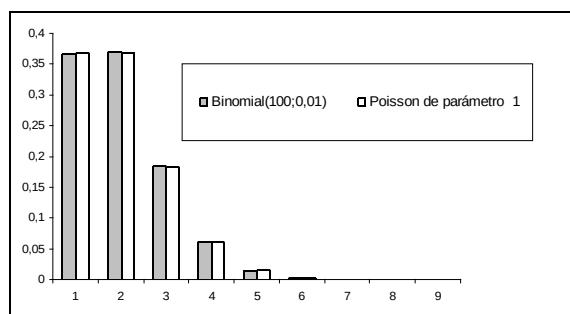
En C2 poñemos: = EXP(-1)*1^A2/FACT(A2) = EXP(-1)/FACT(A2)

Se na distribución binomial n ten un valor cada máis grande, e p cada máis pequeno, a aproximación é aínda mellor. Poñamos $n = 100$, $p = 0,01$ e, xa que logo, $\lambda = 100 \cdot 0,01 = 1$.

Na táboa seguinte se observa esta mellora da aproximación.

	A	B	C
1	x	Binomial(100;0,01)	Poisson $\lambda = 1$
2	0	0,366032	0,367879
3	1	0,369730	0,367879
4	2	0,184865	0,183940
5	3	0,060999	0,061313
6	4	0,014942	0,015328
7	5	0,002898	0,003066
8	6	0,000463	0,000511
9	7	0,000063	0,000073
10	8	0,000007	0,000009

e facendo a representación gráfica:



Exemplo 9

Sábase que o 2% de determinada poboación non ten teléfono móbil. Se escollemos ao azar 100 persoas da referida poboación, cal é a probabilidade de que 3 non teñan móbil?

Trátase dunha variábel aleatoria binomial con $n = 100$ e $p = 0,02$. Daquela:

$$P(X = 3) = \binom{100}{3} \cdot 0,02^3 \cdot 0,98^{97} = 0,18227594.$$

Tamén poderíamos utilizar a distribución de Poisson con $\lambda = 100 \cdot 0,02 = 2$. Entón:

$$P(X = 3) = e^{-2} \frac{2^3}{3!} = 0,13533528 \cdot \frac{8}{6} = 0,18044704$$

Exercicios

12. Supoñamos que un libro de 327 páxinas contén 25 erros tipográficos. Se estes están aleatoriamente distribuídos polo libro, calcúlese a probabilidade de que :
- Unha páxina estea libre de erros.
 - 40 páxinas escollidas non teñan erros.

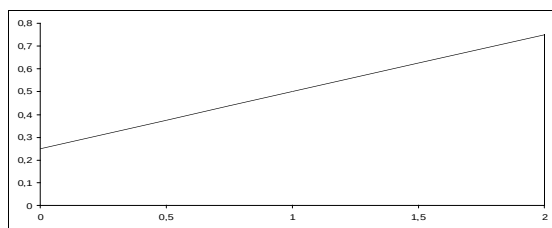
13. As faltas de asistencia a clase de certo grupo de alumnos distribúense aleatoriamente cunha media de unha clase cada 30. Cal é a probabilidade de que un alumno escollido ao azar no faltase durante unha semana (32 clases)?

Variábeis aleatorias continuas

Consideremos a función:

$$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{4}(x+1) & \text{se } 0 \leq x < 2 \\ 0 & \text{se } x > 2 \end{cases}$$

Sen máis que botar unha ollada á figura podemos comprobar que a área pechada por $f(x)$, entre $x = 0$, $x = 2$ e o eixo de abscisas, é igual a 1

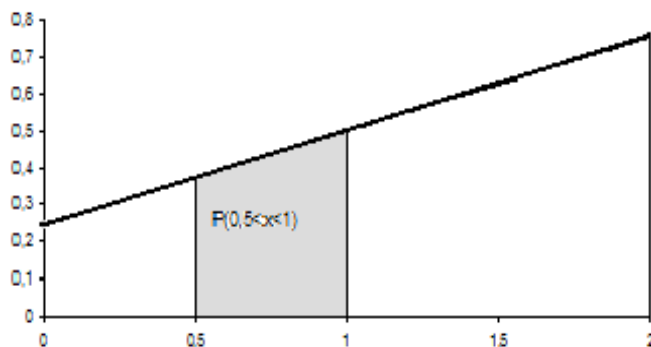


xa que é a do trapezo de bases $f(0)$ e $f(2)$, e altura 2:

$$\frac{f(0) + f(2)}{2} \times 2 = \frac{1}{4} + \frac{3}{4} = 1.$$

Unha función como a anterior que sexa non negativa e que peche co eixe de abscisas unha área igual a 1, diremos que é a función de densidade dunha variábel aleatoria continua, que neste caso toma valores distintos de cero no intervalo $[0,2]$.

Se quixésemos, poñamos por caso, calcular $P(0,5 < X < 1)$ teríamos que achar a área do recinto:



que é:

$$\frac{(f(0,5) + f(1))}{2} \times 0,5 = \frac{\frac{1,5}{4} + \frac{2}{4}}{2} \times 0,5 = 0,21875.$$

En xeral, para obter $P(X \leq x)$, debemos calcular a área entre 0 e x ,

ou sexa,
$$\frac{(f(0) + f(x))}{2} \times x = \frac{\frac{1}{4} + \frac{1}{4}(x+1)}{2} \times x = \frac{1}{4}x + \frac{1}{8}x^2.$$

A este mesmo resultado poderíamos ter chegado mediante o cálculo elemental de primitivas.

A función:

$$F(x) = P(X \leq x) = \begin{cases} \frac{1}{4}x + \frac{1}{8}x^2 & \text{se } x \in [0,2] \\ 0 & \text{noutro caso} \end{cases}$$

chámase función de distribución e permítenos calcular a probabilidade de calquera intervalo.

Por exemplo:

$$P(0,5 \leq X \leq 1) = F(1) - F(0,5) = \frac{1}{4} \cdot 1 + \frac{1}{8} \cdot 1^2 - \left(\frac{1}{4} \cdot 0,5 + \frac{1}{8} \cdot 0,5^2 \right) = 0,21875$$

En xeral, se $f(x)$ é unha función definida positiva, tal que a área pechada por ela e o eixe de abscisas, é 1 (en notación formal: $\int_{-\infty}^{\infty} f(x)dx = 1$), diremos que é unha función de densidade.

Defínese a función de distribución por:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt \quad (\text{a área pechada pola función de densidade até } x).$$

Esta función é monótona non decrecente e cumpre $\lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} \int_{-\infty}^x f(t)dt = 1$.

Tense ademais $P(a \leq x \leq b) = F(b) - F(a)$

Nunha distribución continua as probabilidades obtéñense como áreas, daquela a probabilidade de que a variábel tome un valor concreto será 0, ou sexa $P(X = a) = 0$, e xa que logo $P(X \leq a) = P(X < a)$. Como xeralización do concepto de esperanza (media teórica) e varianza no caso discreto, definimos para variábeis continuas:

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx = \mu,$$

ou sexa, a área pechada pola función $xf(x)$ e o eixe de abscisas, e as ordenadas correspondentes aos extremos do intervalo onde é distinta de cero.

A varianza,

$$V(x) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

S

erá a área pechada pola función $(x - \mu)^2 f(x)$, o eixe de abscisas e as ordenadas correspondentes aos extremos do intervalo onde é distinta de cero.

Como no caso discreto, defínese o desvío típico por $\sigma = \sqrt{V(X)}$.

Exemplo 10

Calcular a esperanza, varianza e desvío da variábel aleatoria que ten función de

$$\text{densidade: } f(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{4}(x+1) & \text{se } 0 \leq x < 2 \\ 0 & \text{se } x > 2 \end{cases}$$

Temos:

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^2 xf(x)dx = \int_0^2 x \frac{x+1}{4} dx = \frac{7}{6}$$

$$\begin{aligned}
 V(x) &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = E(X^2) - E(X)^2 = \\
 &= \int_0^2 x^2 f(x) dx - \left(\frac{7}{6}\right)^2 = \int_0^2 x^2 \frac{x+1}{4} dx - \left(\frac{7}{6}\right)^2 = 0,30555556 \\
 E \sigma(x) &= 0,5527708.
 \end{aligned}$$

Exercicios

Compróbase que a seguinte función é de densidade. Calcúlese $P(4 \leq X \leq 6)$, a esperanza e o desvío.

$$f(x) = \begin{cases} 0 & \text{se } x < 2 \\ \frac{1}{64}(14-x) & \text{se } 2 \leq x < 10 \\ 0 & \text{se } x > 10 \end{cases}$$

13. Dada a función $f(x) = \begin{cases} kx & \text{se } 0 \leq x < 5 \\ k(10-x) & \text{se } 5 \leq x \leq 10 \\ 0 & \text{se } x > 10 \end{cases}$

- Calcúlese o valor de k para que f(x) sexa de densidade.
- Calcúlese a función de distribución correspondente.
- Calcúlese a esperanza e a varianza.

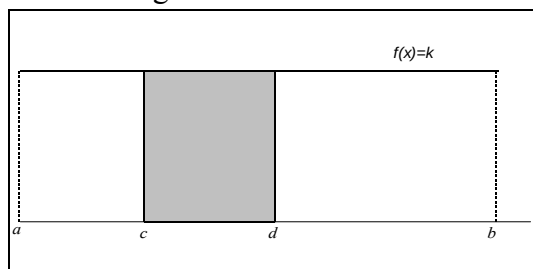
Distribución uniforme

Unha variábel aleatoria díse que ten distribución uniforme no intervalo $[a, b]$, se a súa función de densidade ven dada por:

$$f(x) = \begin{cases} k & \text{se } x \in [a, b] \\ 0 & \text{se } x \notin [a, b] \end{cases}$$

Dado que $\int_{-\infty}^{\infty} f(x) dx = \int_a^b k dx = k(b-a) = 1$, dedúcese $k = \frac{1}{b-a}$

A probabilidade de que a variábel tome un valor entre dous puntos do intervalo $[a, b]$, virá dada pola área do rectángulo:



Execicio 14

Representése a función de distribución dunha variábel aleatoria uniforme $[3,7]$. Calcúlese a súa esperanza e a súa varianza.

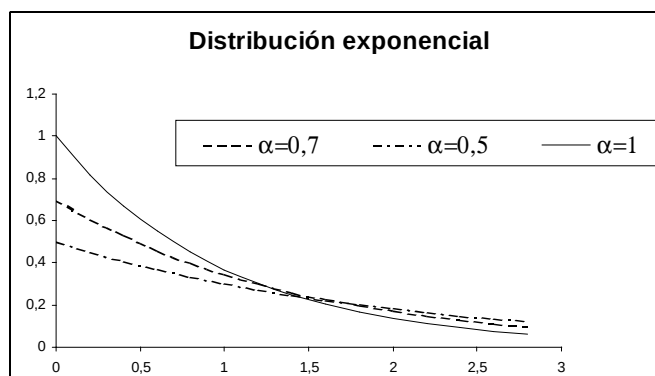
Distribución exponencial

Dise que unha variábel aleatoria continua X ten unha distribución exponencial con parámetro $\alpha > 0$ se a súa función de densidade de probabilidade é:

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

Na folla de cálculo pódese obter a gráfica da función de densidade para diversos valores de α . Mediante o seguinte esquema, extendendo até a fila 35ª obtemos as correspondentes a $\alpha = 0,1$; $0,5$ e 1 .

A	B	C	D
0	=0,1*EXP(-0,1*A1)	=0,5*EXP(-0,5*A1)	=EXP(-A1)
=A1+0,2	=0,1*EXP(-0,1*A2)	=0,5*EXP(-0,5*A2)	=EXP(-A2)



Demóstrase que $E(x) = \frac{1}{\alpha}$ e $\sigma_x^2 = \frac{1}{\alpha^2}$. Daquela o parámetro da distribución, se sabemos que o modelo exponencial é o axeitado para tratar unha determinada situación, pode estimarse mediante o inverso da media empírica.

Exemplo 11

O tempo que tarda un individuo en ser atendido nun banco é unha variábel aleatoria que se distribúe exponencial cunha media de 3 minutos. Cal é a probabilidade de que unha persoa sexa atendida en menos de 2?

O tempo en minutos para que un individuo sexa atendido nun banco, X , distribúese exponencial con media de 3 minutos, é dicir, o valor esperado é $\frac{1}{\alpha} = 3$, de onde $\alpha = \frac{1}{3}$ (parámetro da exponencial). A función de distribución de probabilidade acumulada será:

$$P(X \leq x) = F_X(x) = \int_0^x \alpha e^{-\alpha t} dt = \frac{\alpha e^{-\alpha t}}{-\alpha} \Big|_0^x = -e^{-\alpha x} + 1 = 1 - e^{-\alpha x},$$

$$\text{Se } \frac{1}{\alpha} = 3, \quad P(x \leq 2) = 1 - e^{-\frac{2}{3}} = 0,48658288.$$

Mais, se aínda non se estudou integración (ou malia que se estudase, se o que queremos é resolver problemas), utilizamos a a folla de cálculo para obter os valores da función de distribución:

$$P(X \leq 2) = \text{DISTR.EXP}(2; 1/3; \text{VERDADERO}) = 0,48658288.$$

(“VERDADERO” se se quere obter a probabilidade acumulada, como neste caso)

Exercicios

15. O tempo de duración da eficacia de certo reactivo químico é de 50 horas. Se o tempo de eficacia é unha variábel aleatoria que segue unha distribución exponencial, calcúlese:
 - a) A probabilidade de que a eficacia sexa de entre 80 e 100 horas.
 - b) A probabilidade de que a eficacia sexa menor de 30 horas.
16. Supóñase que o tempo en horas de traballo sen fallo con determinado programa informático segue unha distribución exponencial de $\alpha = 0,02$.
 - a) Calcúlese a probabilidade de que o programa funcione sen fallo nas primeira cen horas
 - b) Calcúlese a probabilidade de que o programa funcione sen fallo a partir das 50 horas.

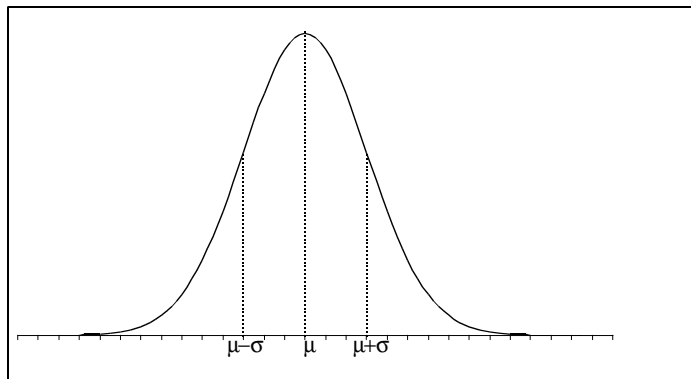
Distribución Normal

Unha variábel aleatoria continua dise que segue unha distribución normal de media μ e desvío σ , simbolicamente $X \rightarrow N(\mu, \sigma)$, se a súa función de densidade é:

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

Evidentemente a función é definida positiva e demóstrase que $\int_{-\infty}^{\infty} f(x) dx = 1$.

Ademais, a función é simétrica con respecto a $x=\mu$, co máximo neste valor, puntos de inflexión en $\mu+\sigma$ e $\mu-\sigma$ e ten forma de campá. A súa representación gráfica é a seguinte:



A función de distribución ven dada por

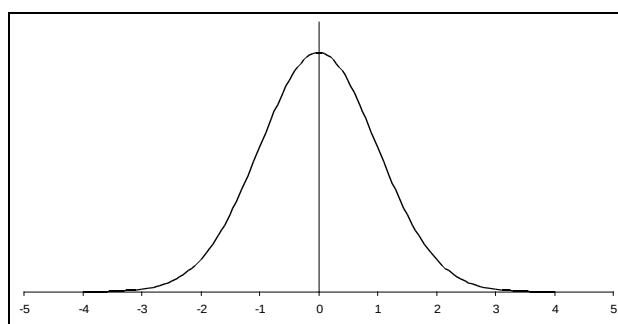
$$F(x) = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$

Son moitas as variábeis que seguen esta distribución, poñamos por caso: o peso, a altura, o sobrevalorado CI, etc. Pero as súas aplicacións máis sobranceiras baséanse en que a media das mostras, en certas condicións bastante xerais, tamén segue unha distribución normal. Disto daremos conta ao estudar a teoría da mostraxe.

A distribución normal considerase como a máis importante do cálculo de probabilidades e asúmese, polo xeral, que todas aquelas variábeis aleatorias asociadas á medidas de caracteres sometidos a pequenas influencias dun grande número de factores, admiten unha boa aproximación mediante ela. En certo sentido, máis do 95% do cálculo de probabilidades e da estatística está relacionado coa distribución normal; ora ben, tamén hai quen sostén que a teoría de probabilidades verdadeiramente interesante é aquela que fica fóra da *normalidade*, a que atinxe a fenómenos moi pouco probábeis e con distribucións raras; precisamente aqueles que se deron na formación do Universo, da Terra, do home,...

En condicións moi xerais, acéptase que os erros de medida seguen unha distribución normal con media cero e desvío 1. Daquela, a súa función de densidade será:

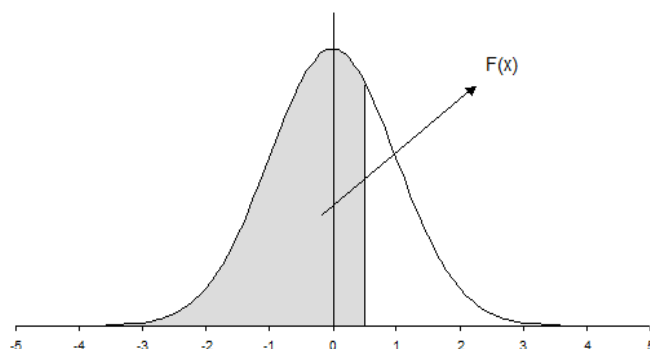
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$



Podemos comprobar, utilizando a folla de cálculo, que a área pechada pola esta curva, desde $-\infty$ a ∞ , é 1, aproximando por trapezoides de altura 0,1. Extendendo as seguintes celas até acadar o valor 4 na primeira columna, obtense como valor para a área 0,999935767. Desprezamos, sen que se produza demasiado erro –da orde das cenmilésimas–, os valores de x maiores que 4 e menores de -4 . [Ver regra dos trapezoides no capítulo 10]

-4	$= (1/(2*PI())^0,5)*EXP(-(A1^2/2))$	
=A1+0,1	$= (1/(2*PI())^0,5)*EXP(-(A2^2/2))$	$= 0,1*(B2+B1)/2$

Probabilidades baixo a curva normal



Como nas anteriores distribucións continuas, a función de distribución $F(x)$ dá a área acumulada até a abscisa x .

No caso da figura, onde a distribución é a $N(0,1)$, utilizando a folla de cálculo, obtemos: $F(0,563) = \text{DISTR.NORM}(0,563;0;1;\text{VERDADERO}) = 0,7132826$.

Existen **táboas** que dan o valor da función de distribución $N(0;1)$. Para calcular $F(0,563)$, o valor aproximado, desprezando as milésimas, estará na intersección da fila que comeza por 0,5, coa columna que comeza por 0,06; obtense $F(0,56) = 0,7123$. Son poucas as táboas que conteñen as abscisas negativas, daquela se queremos calcular, poñamos por caso, $F(-1,23)$, debemos ter en conta:

$F(-1,23) = P(X \leq -1,23) = P(X \geq 1,23)$, xa que a $N(0,1)$ é simétrica respecto ao eixo y .

Entón:

$$P(X \geq 1,23) = 1 - P(X \leq 1,23) = 1 - F(1,23) = 1 - 0,8906514 = 0,10934862.$$

Podemos calcular, xa que logo, utilizando a táboa, $F(x)$ para calquera valor de x , positivo ou negativo, e en consecuencia a probabilidade de calquera intervalo, que virá dada por $P(a \leq X \leq b) = F(b) - F(a)$.

Se queremos calcular os valores dunha normal calquera, utilizando a folla de cálculo obtémolos directamente. Por exemplo para $F(76)$ nunha $N(67,15)$, temos que escribir na correspondente cela: =DISTR.NORM(76;67;15;VERDADERO), onde 76 é o valor da variábel, 67 a media, e 15 o desvío. “VERDADERO” indica que estamos a traballar coa función de distribución; se poñemos “FALSO” obteremos a de densidade.

No caso de utilizar táboas, tense en conta que se X é $N(\mu, \sigma)$, entón $Z = \frac{X - \mu}{\sigma}$ segue unha distribución $N(0,1)$. Daquela, chamado F_x á función de distribución de X , e F_z á de z , temos:

$$F_x(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(z \leq \frac{x - \mu}{\sigma}\right) = F_z\left(\frac{x - \mu}{\sigma}\right)$$

Por exemplo, para calcular $F(34)$, para unha $N(30,7)$, temos:

$$F_x(34) = F_z\left(\frac{34 - 30}{7}\right) = F_z(0,57142857) = 0,71614546$$

E para calcular $F(28)$,

$$F_x(28) = F_z\left(\frac{28 - 30}{7}\right) = F_z(-0,28571429) = 1 - F_z(0,28571429) = 0,38754854$$

Exemplo 12

As alturas de 1000 nenos seguen unha distribución normal de media 1 metro e desvío 20 cm. Cal é a proporción de nenos que miden entre 120 e 130 cm.?

$$P(120 \leq X \leq 130) = F_z\left(\frac{130 - 100}{20}\right) - F_z\left(\frac{120 - 100}{20}\right) = F_z(1,5) - F_z(1) = 0,0919.$$

En moitas ocasións interesa facer buscas inversas nunha función de distribución normal. Se queremos, poñamos por caso, calcular un valor de x tal que $P(X \leq x) = 0,70$, se seguimos co exemplo anterior, teríamos:

$$x = \text{DISTR.NORM.INV}(0,70;100;20) = 110,5.$$

Este valor da abscisa chámase *percentil 70*. Por debaixo del está o 70% que menos mide e por riba o 30% que máis mide; adoita denotarse p_{70} .

Se tivéssemos que utilizar a táboa, partindo de $F_z\left(\frac{p_{70} - 100}{20}\right) = 0,70$, buscamos na parte central da táboa a probabilidade que máis se lle aproxima: 0,6985, que está na intersección da fila que comeza por 0,5 e da columna que comeza por 0,02. Xa que logo, o valor da correspondente abscisa da $N(0,1)$ é 0,52. Entón: $\frac{p_{70} - 100}{20} = 0,52$ e $p_{70} = 110,4$.

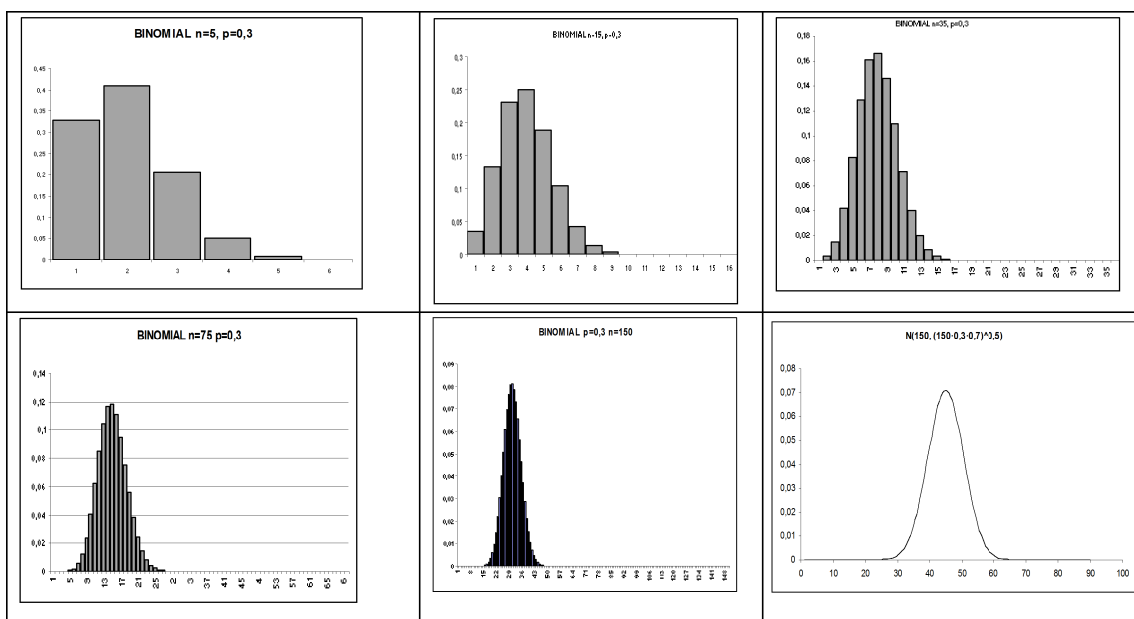
Exercicio 17

O consumo medio de auga por persoa e día é de 160 litros, cun desvío típico de 80 litros.

- Cal é a porcentaxe de persoas que consomen menos de 80 litros por día?
- Que porcentaxe consome máis de 200 litros por día?
- Escollida unha persoa ao azar, cal é a probabilidade de que consoma entre 140 e 190?
- Por riba de que consumo estará o 10% que máis consome?
- Por baixo de que consumo estará o 7% que menos consome?

Aproximación da distribución binomial pola normal

Nas seguintes gráficas se representan os diagramas de barras para distintos valores de n (5, 15, 35, 75 e 150) da binomial de parámetro $p = 0,3$. Podemos observar como o perfil da gráfica, conforme aumenta n , vai adquirindo forma de campá, presentando simetría respecto á recta $x = np$, e imitando cada vez máis unha curva normal.



Este resultado é o que recolle o chamado **Teorema de De Moivre**, que a continuación enunciaremos:

Se X é unha variábel aleatoria de parámetros n e p , entón $Z = \frac{X - np}{\sqrt{npq}}$, cando n tende a infinito, é unha variábel aleatoria que segue unha $N(0,1)$.

En xeral, para valores de $n > 30$ adoita utilizarse a curva normal para aproximar a binomial.

Chamándolle X_b á variábel aleatoria binomial e X_n á correspondente normal, facemos

$$P(X_b = x) = P(x - 0,5 \leq X_n \leq x + 0,5) = F(x + 0,5) - F(x - 0,5),$$

onde $F(x)$ é a función de distribución normal de media np e varianza npq .

Do mesmo xeito, $P(a \leq X_b \leq b) = F(b + 0,5) - F(a - 0,5)$.

Para ver até que punto a aproximación é boa, imos calcular a probabilidade de obter entre 40 e 70 caras ao botar unha moeda 100 veces, utilizando as dúas distribucións

Binomial:

$$F(71) = \text{DISTR.BINOM}(71; 100; 0,5; \text{VERDADERO})$$

$$F(39) = \text{DISTR.BINOM}(39; 100; 0,5; \text{VERDADERO}) =$$

$$F(71) - F(39) = 40,99999371 - 0,0176001 = 0,98239361$$

Aproximación normal:

$$F(71,5) - F(39,5) = \text{DISTR.NORM}(70,5; 50; 5; \text{VERDADERO}) -$$

$$\text{DISTR.NORM}(39,5; 50; 5; \text{VERDADERO}) =$$

$$0,999979331 - 0,017864357 = 0,982114974.$$

Exemplo 13

Sábase que o 5% dos habitantes dun determinado país son analfabetos. Se escollemos ao azar unha mostra de 2000 persoas, calcúlese a probabilidade de que:

a) Polo menos 120 sexan analfabetos.

b) Entre 70 e 110 sexan analfabetos.

A variábel aleatoria binomial ten como parámetros $n = 2000$ e $p = 0,05$.

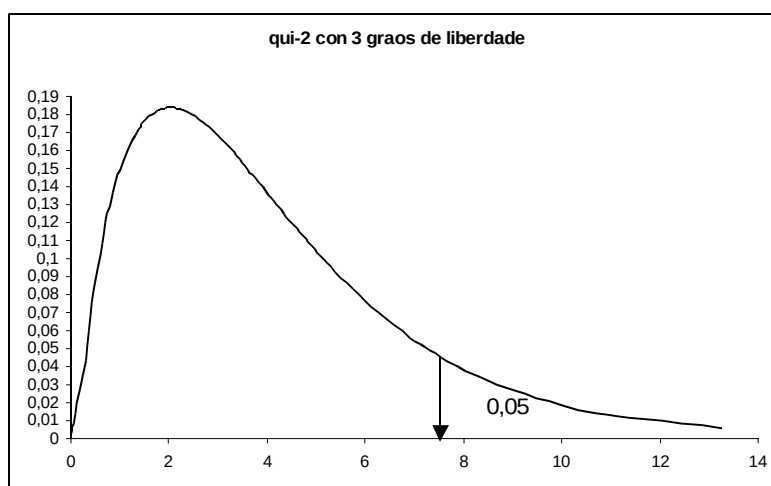
Aproximando pola normal, temos $\mu = 2000 \cdot 0,05 = 100$ e $\sigma = \sqrt{2000 \cdot 0,05 \cdot 0,95} = 9,7468$,

$$\text{daquela: } P(X \geq 120) = P\left(Z \geq \frac{119,5 - 100}{9,7468}\right) = 0,02271463$$

$$P(70 \leq X \leq 110) = P\left(\frac{69,5 - 100}{9,7468} \leq Z \leq \frac{110,5 - 100}{9,7468}\right) = 0,85844523$$

Distribución χ^2

Se $X_1, X_2, \dots, X_n$ son variábeis aleatorias $N(0,1)$ independentes, a distribución da variábel $\chi_n^2 = X_1^2 + X_2^2 + \dots + X_n^2$ recibe o nome de distribución **qui-cadrado de Pearson** con n graos de liberdade, e xoga un papel fundamental na resolución dalgúns problemas que formula a inferencia estatística.



Conforme n aumenta o máximo desprázase cada vez máis cara á dereita até acadar a forma da distribución normal para valores grandes de n .

Na práctica, o interesante é buscar as abscisas correspondentes a unha probabilidade dada. Así, para buscar un valor de x a partir do cal a área pechada entre a curva e o eixe de abscisas sexa 0,05, se utilizamos a folla de cálculo temos que inserir a función:

$$=PRUEBA.CHI.INV(0,05;3),$$

onde 0,05 é a probabilidade requirida, e 3 son os graos de liberdade. Neste caso, o valor obtido é: 7,8147247.

Esta distribución tamén está tabulada, daquela se pretendemos buscar o valor anterior utilizando as táboas, un a tal que $P(\chi_3^2 \leq a) = 0,95$, temos que buscar o valor na intersección da fila correspondente ao número 3 coa columna correspondente a $\chi_{0,95}^2$, que dá 7,81.

A función de densidade da χ_n^2 ven dada pola expresión :

$$f(x) = K_n \cdot x^{\frac{n-2}{2}} \cdot e^{-\frac{x}{2}}$$

onde n é o número de graos de liberdade e K_n é unha constante dependente de n que vale:

$$K_n = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}$$

A χ^2 é un caso particular dunha familia de distribucións chamadas Gamma, cuxas funcións de densidade, que dependen de dous parámetros α e β , veñen dadas por:

$$f(x, \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \cdot e^{-\frac{x}{\beta}}$$

Utilizando a folla de cálculo podemos representar a densidade da χ^2 . Pegamos DISTR.GAMMA e no cadro de diálogo poñemos $\alpha = n/2$ (sendo n o número de graos de liberdade escollido) e $\beta = 2$.

Temos a táboa seguinte:

	A	B	C	D
1	x	n = 3	n = 5	n = 10
2	0	0	0	0
3	1	0,24197072	0,08065691	0,00078975
4	2	0,20755375	0,13836917	0,00766416
5	3	0,15418033	0,15418033	0,02353326
6	4	0,10798193	0,14397591	0,04511176
7	5	0,07322491	0,12204152	0,06680094
8	6	0,04865217	0,09730435	0,08401568
9	7	0,0318734	0,07437127	0,09440614
0	8	0,02066699	0,05511196	0,09768341
11	9	0,01329555	0,03988664	0,09490381
12	10	0,00850037	0,02833456	0,08773368

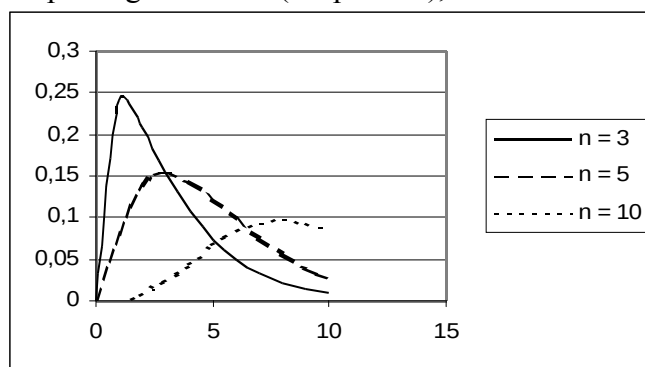
Na columna A escribimos as abscisas de 0 a 10.

En B2 pegamos a función DISTR.GAMMA. No cadro de diálogo poñemos : X = A2 ;

$\alpha = 1,5$, correspondente a $n = 3$ graos de liberdade. $\beta = 2$. Acumulado = Falso. Extendemos até B12.

Repetimos o proceso nas columnas C e D con $\alpha = 2,5$ e $\alpha = 5$, co cal temos os valores da función para 5 e 10 graos de liberdade.

Utilizando o comando GRÁFICO , sinalando como campo de gráfica a táboa anterior e escollendo como tipo de gráfico XY (Dispersion), resulta:



Observamos que conforme n aumenta, o máximo desprázase cada vez máis cara á dereita até acadar para valores grandes de n a forma da distribución normal.

Distribución t

Se Z é unha variábel $N(0,1)$ e Y unha variábel χ^2 con n graos de liberdade, independente de Z , entón $t = \frac{Z}{\sqrt{\frac{Y}{n}}}$ é unha variábel t de Student con n graos de liberdade.

A distribución t de Student, ten por función de densidade:

$$t_n(x) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

Onde n representa os graos de liberdade. Esta función existe para todos os valores de x reais, e é simétrica respecto ao eixe y .

A función gamma, que aparece na expresión da función de densidade, pode calcularse na folla de cálculo mediante GAMMA.LN(x), que devolve o logaritmo neperiano da función $\Gamma(x)$.

Para obter a función Gamma a partir de GAMMA.LN(x), teremos que calcular posteriormente =EXP(GAMMA.LN(x)).

O

Se, poñamos por caso, queremos calcular $\Gamma(6)$, partimos de =GAMMA.LN(6), que dá o logaritmo natural da función gamma en 6 (4,787491743). Daquela:

$$\Gamma(6) = \text{EXP}(\text{GAMMA.LN}(6)) = \text{EXP}(4,787491743) = 120.$$

Para calcular $\Gamma\left(\frac{1}{2}\right)$, partimos de =GAMMA.LN(1/2), e obtemos 0,572364942989577,

$$\text{co cal } \Gamma\left(\frac{1}{2}\right) = \text{EXP}(0,572364942989577) = 1,77245385102051$$

Podemos utilizar a folla de cálculo para computar os valores da función de densidade de calquera distribución t . Se queremos, por exemplo, representar t_3, t_{10} e t_{20} , teremos que comezar por calcular $\Gamma(2), \Gamma\left(\frac{3}{2}\right), \Gamma(5), \Gamma\left(\frac{11}{2}\right), \Gamma(10)$ e $\Gamma\left(\frac{21}{2}\right)$.

Procedendo da forma anterior, obtemos:

x	$\Gamma(x)$
2	1
1,5	0,88622693
5	24
5,5	52,3427778
10	362880
10,5	1133278,39

Colocamos os valores correspondentes na folla de cálculo da seguinte maneira

E	F	G
1	52,3427778	1133278,39
0,88622693	24	362880

E mediante o esquema:

Na columna A escribimos as abscisas de -5 a 5.

En B2 pegamos $= (1/(3*PI()))^{0,5} * (E\$1/E\$2) * (1+A2^2/3)^{-2}$, e extendemos até B12.

En C2 pegamos $= (1/(10*PI()))^{0,5} * (F\$1/F\$2) * (1+A2^2/10)^{-11/2}$ e extendemos até C12.

En D2 pegamos $= (1/(20*PI()))^{0,5} * (G\$1/G\$2) * (1+A2^2/20)^{-10}$ e extendemos até D12.

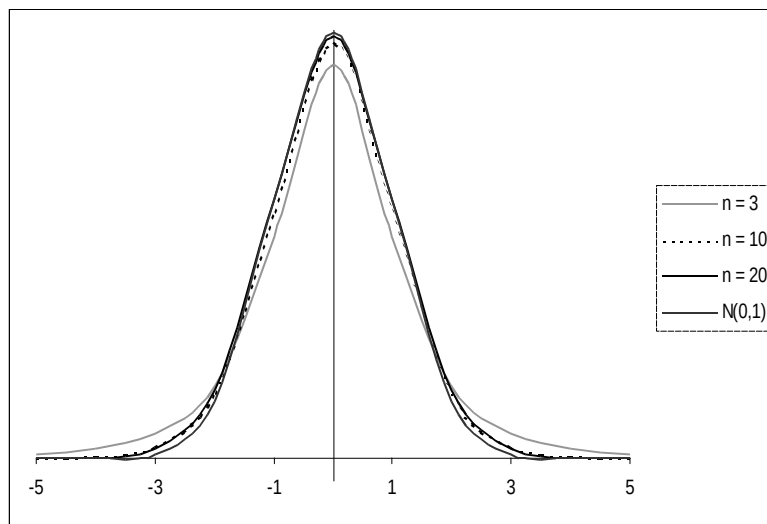
Obtemos a táboa seguinte:

	A	B	C	D	E
1	x	n = 3	n = 10	n = 20	N(0,1)
2	-5	0,004219354	0,000396001	0,000118484	1,48672E-06
3	-4	0,009163361	0,002031034	0,001103465	0,00013383
4	-3	0,022972037	0,011400549	0,009589669	0,004431848
5	-2	0,067509661	0,061145766	0,063631356	0,053990967
6	-1	0,206748336	0,230361989	0,241874814	0,241970725
7	0	0,367552597	0,389108384	0,393988586	0,39894228
8	1	0,206748336	0,230361989	0,241874814	0,241970725
9	2	0,067509661	0,061145766	0,063631356	0,053990967
10	3	0,022972037	0,011400549	0,009589669	0,004431848
11	4	0,009163361	0,002031034	0,001103465	0,00013383
12	5	0,004219354	0,000396001	0,000118484	1,48672E-06

Utilizando o comando GRÁFICO, sinalando como campo de gráfica a táboa anterior e escollendo como tipo de gráfico XY (Dispersion), obtemos o gráfico que figura enriba. Na columna E inclúense os valores obtidos para a función de densidade da

$N(0,1)$, que ao aumentar n pareceranse cada vez máis aos da t .

Como se pode ver no seguinte gráfico, a curva presenta simetría con respecto ao eixe y , ademais, cando n tende a infinito a distribución t tende á normal con media cero e desvío 1.

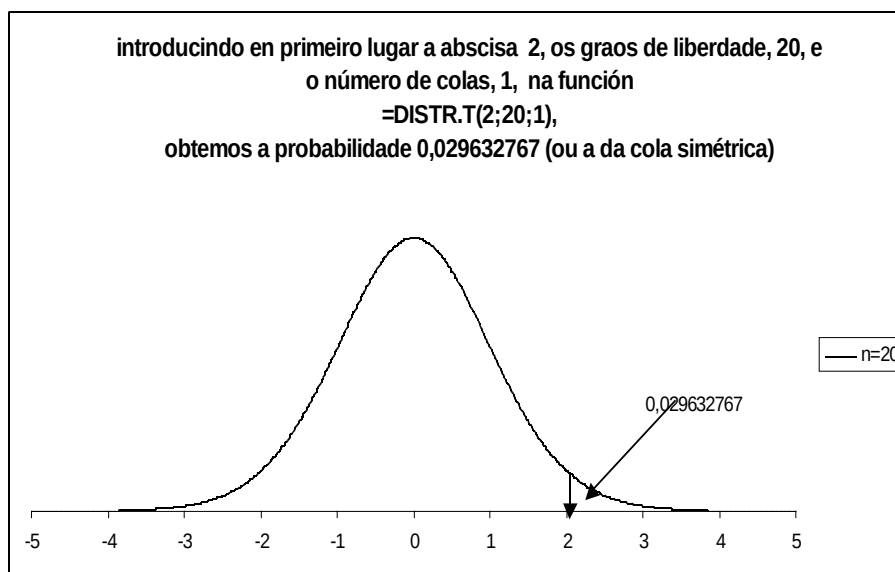


Obtención de probabilidades baixo a curva t

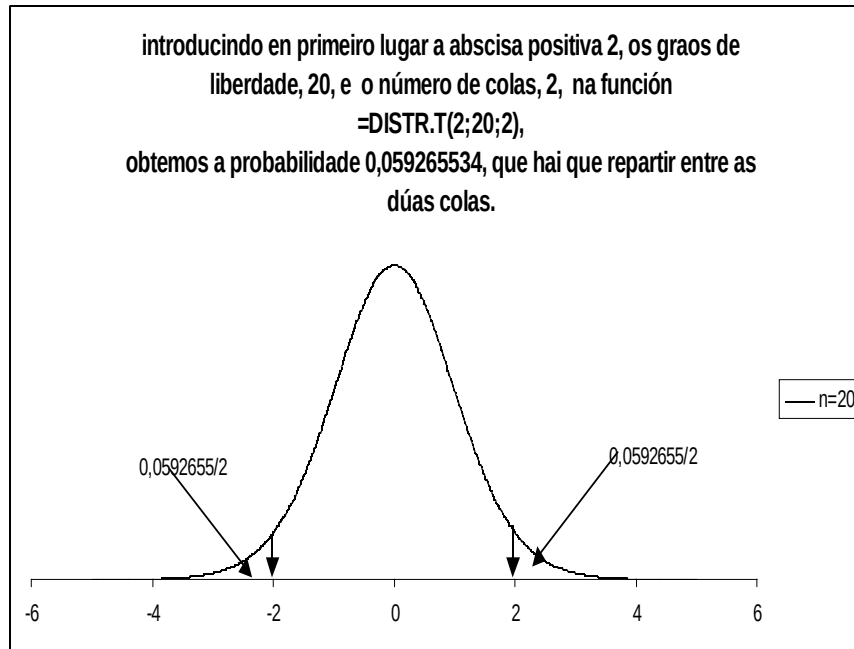
I) Coa folla de cálculo.

1. Dada unha abscisa achar a probabilidade.

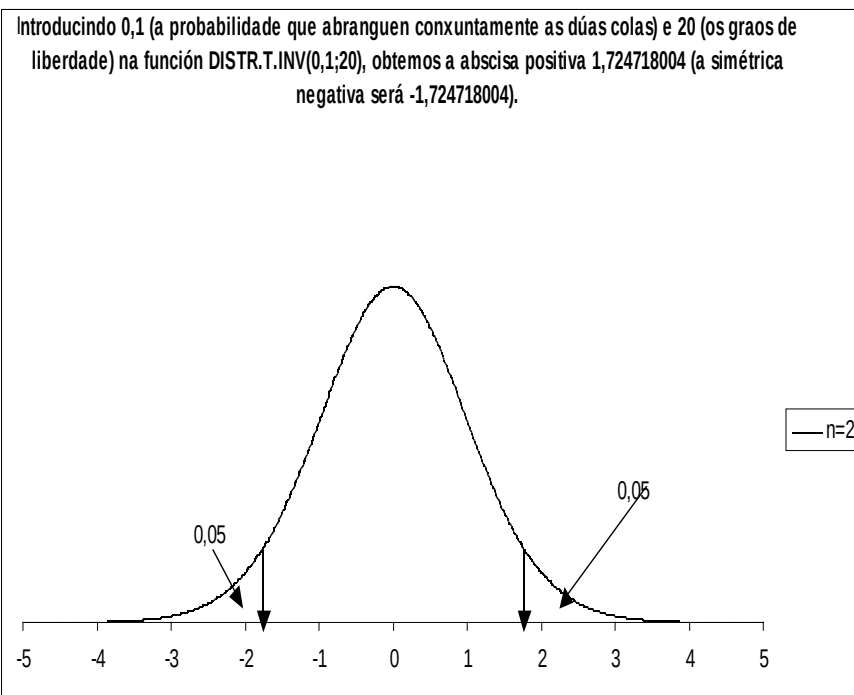
a) Unha cola.



b) Dúas colas.



2. Dada a probabilidade achar as abcisas



II) Utilizando as táboas.

Se queremos calcular, poñamos por caso, $a / P(t_{20} \leq a) = 0,80$, na intersección da fila correspondente a 20 graos de liberdade, e a columna $t_{0,80}$, obtemos $a = 0,860$

Exercicios

18. O 20% dos DVD que regalan cos xornais dan problemas ao reproducilos na computadora. Se temos 5 DVD desa procedencia, calcúlese a probabilidade de que:

- a) Dous se poidan reproducir sen problema.
- b) Como máximo 3 non se poidan reproducir.
- c) Polo menos 2 non se poidan reproducir.
- d) Ningún se poida reproducir.
- e) Todos se poidan reproducir.

19. Admítase que a probabilidade de que un cliente faga unha compra superior a 30€ nun supermercado é de 0,4. Calcula a probabilidade de que:

- a) De tres clientes ningún realice compra superior a 30€.
- b) De tres clientes, como mínimo dous gasten máis de 30€.
- c) De 15 clientes ningún realice compra superior a 30€.
- d) De 15 clientes tres gasten máis de 30€.
- e)

20. Tan só o 8% dos contratos de traballo realizados na comarca de Vigo no ano 2008 foron indefinidos. Calcúlese a probabilidade de que dun grupo de 12 contratados naquel ano

- a) Polo menos dous tivesen contrato indefinido.
- b) Seis tivesen contrato indefinido
- c) Ningún tivese contrato indefinido.

21. A máquina I produce ao día o dobre de pezas que a máquina II. O 6% das pezas producidas pola máquina I son defectuosas, mentres que das producidas pola II tan só o é o 2%. Extraído aleatoriamente un lote de 10 pezas da produción total, calcúlese a probabilidade de que:

- a) Dous sexan defectuosas.
- b) Entre 2 e 5 sexan defectuosas
- c) Cal é número esperado de defectuosas nun lote de 100?

22. Estímase que o 3% das píulas de certo medicamento perden a súa efectividade despois dun ano. Calcúlese, utilizando a aproximación pola normal, a probabilidade de que dun lote de 1000 píulas administradas despois dun ano de ser fabricadas, máis de 30 non sexan eficaces.

23. O número de chamadas que se reciben en cinco minutos na central dunha empresa é unha variábel de Poisson de parámetro 10. Calcúlese a probabilidade de que:

- a) En cinco minutos chegaran exactamente 8 chamadas.
- b) En cinco minutos chegaran menos de 5
- c) Nun minuto chegaran como mínimo dúas.
- d) En dez minutos chegaran 16 chamadas.
- e) En dez minutos chegaran menos de 10.

24. Admítase que o tempo de duración de cada anuncio nunha cadea de televisión segue unha distribución uniforme [5,12]. Calcúlese:

- a) A función de densidade.

- b) A probabilidade de que un anuncio dure menos de 8 segundos.
 - c) A probabilidade de que un anuncio dure entre 8 e 12 segundos.
 - d) Cal é a duración media por anuncio?
25. A duración en horas de certo tipo de pezas de precisión segue a función de densidade $f(x) = 100/x^2$ se $x > 100$, e 0 noutro caso. Calcúlese a probabilidade de que de tres pezas ningunha teña que ser substituída nas primeiras 150 horas de uso. Cal é a probabilidade de que haxa que substituír as tres?
26. Unha máquina fabrica parafusos de diámetros que se distribúen segundo a función de densidade $f(x) = k(x-0,24)^3 (x-0,26)^2$ se $x \in [0,24,0,26]$, e $f(x)=0$, noutro caso. Calcula k para que a función sexa de densidade. Se os parafusos se desbotan ao desviárense os seus diámetros de 0,25 en máis de 0,008, cal é a proporción de parafusos rexeitados?
27. Os suicidios en certa comarca francesa a principios do século XX, para persoas comprendidas entre 20 e 60 anos, distribuían-se segundo a función de densidade $f(x) = ke^{0,029x}$ se $x \in [20,60]$, e cero noutro caso. Calcúlese k para que a función sexa efectivamente de densidade. Cal é a proporción de suicidios para os menores de 25? E para os maiores de 55?
28. Calcúlese a función de distribución dunha variábel aleatoria de función de densidade $f(x) = \frac{k}{(1+x)^2}$ se $x \in [0,5]$, e cero noutro caso. Calcúlese tamén a esperanza e a varianza.
29. O tempo en horas de traballo sen fallo de certo dispositivo segue unha distribución exponencial de parámetro 0,03.
- a) Calcúlese a probabilidade de que o dispositivo traballe sen fallo nas primeiras cen horas.
 - b) Sabendo que non fallou nas primeiras 100 horas, cal é a probabilidade de que falle nas seguintes 200?
30. O tempo requirido para executar certa tarefa segue unha distribución normal de media 72 minutos e desvío 12 minutos. Calcúlese a probabilidade de que a tarefa:
- a) Leve máis de 93 minutos.
 - b) Non leve máis de 65.
 - c) Leve entre 63 e 78.
 - d) Determinar os percentís 25 e 80.
31. O tempo en minutos que un operario tarda en realizar certa tarefa segue unha distribución normal. Sábese que a probabilidade de que tarde máis de 13 minutos é 0,0668 e a de que tarde menos de 8 minutos é 0,1587.
- a) Calcúlese o tempo medio e o correspondente desvío típico.
 - b) Calcúlese a probabilidade de que tarde entre 9 e 12 minutos.
32. En determinada empresa a distribución semanal de materia prima é unha variábel aleatoria normal con media 600 kg e desvío padrón 40 kg. A empresa ten a comezos de semana 634 kg de materia prima e non lle é posíbel facer máis aprovisionamentos.
- a) Determínese a probabilidade de que a empresa quede sen materia prima.
 - b) Calcúlese cantos kg de materia prima debería ter para que a probabilidade de que non lle chegase fose de 0,01.
33. No ano 2007, segundo a enquisa de condicións de vida do IGE, o 19,22% de todos os galegos, tanto activos como inactivos, tiñan uns ingresos medios mensuais inferiores a 450€, mentres que o 7,48% superaba os 1500 €. Supoñendo que os ingresos medios seguen unha distribución normal,

- a) Cal era a porcentaxe de persoas con ingresos medios superiores a 1000€?
 - b) Cales eran os ingresos medios do 10% que máis ingresa?
 - c) Cal era o ingreso máis alto entre o 20% que menos ingresa?
 - d) Entre dez persoas elixidas ao azar, cal é a probabilidade de que 6 tivesen ingresos medios superiores a 900€.
34. Entre os algarismos 1,2,3,4,5,6,7,8,9 eliximos un ao azar. Se repetimos a operación 500 veces, cal é a probabilidade de que:
- a) Apareza o 5 máis de 120 veces?
 - b) Apareza o 3 entre 50 e 80 veces?

5. MOSTRAXE

—Tamén eu pensei nun modelo de cidade mediante o cal deduzo todas as outras —respondeu Marco. É unha cidade feita só de excepcións, exclusións, contradicións, incongruencias, contrasentidos. Se unha cidade así é absolutamente improbábel, diminuindo o número de elementos anormais aumentan as posibilidades de que a cidade verdadeira exista [...] Pero non podo levar a miña operación para alén de certos límites: obtería cidades demasiado verosímiles para seren verdadeiras.

Italo Calvino. *As cidades invisíbeis*.

Poboación e mostra

Algúns exemplos:

- 1) Co fin de deseñar unha campaña de prevención, preténdese coñecer o consumo de alcol entre a mocidade, comprendida entre 14 a 18 anos, dunha determinada cidade.
- 2) Perante a celebración dunhas eleccións quérese facer un prognóstico sobre os resultados que vai ter determinado partido.
- 3) Con vistas a tomar medidas contra a obesidade mórbida cómpre estimar a inxesta de graxas saturadas polos menores de 15 anos.
- 4) Para deseñar políticas de explotación pesqueira sustentábel quérese saber cal é a poboación aproximada de sardiña na ría de Arousa nunhas datas determinadas.

Todos os supostos anteriores teñen en común que resultaría practicamente imposible, ou moi custoso en todos os sentidos, ter acceso a cada un dos membros da correspondente poboación: mozos de 14 a 18 anos, membros do censo electoral, ou sardiñas da ría de Arousa. O que se fai na práctica é seleccionar determinados subconxuntos da poboación, *mostras*, e a partir delas tratar de obter resultados para a poboación ao completo.

As características dunha poboación identifícanse para o seu estudo cos *parámetros* que se consideran como os verdadeiros valores da característica estudada. Por exemplo, a proporción de votantes de determinado partido é un parámetro da poboación formada polos inscritos no censo electoral. O consumo medio de graxas saturadas é un parámetro da poboación de menores de 15 anos.

Polo xeral, sempre que non haxa ambigüidades, as poboacións obxecto de estudo na mostraxe consideraranse formadas polas medidas das magnitudes que tratamos de estudar. Falaremos, poñamos por caso, de poboación de alturas, de salarios, de calibres de parafusos, ou mesmo de poboacións formadas por ceros e uns, no caso de que nos

interese coñecer que elementos presentan ou non determinada característica: ser muller ou non selo, beber alcohol ou non bebelo, votar a un partido ou non votalo, etc.

A *inferencia estatística* é unha rama do cálculo de probabilidades que permite tirar conclusións para toda a poboación, con certas garantías, a partir das mostras. A escolla da mostra, para que sexa representativa, debe facerse segundo determinados procedementos dos que trata a teoría da *mostraxe*.

Mostraxe probabilística

O deseño da mostraxe é o plano que se levará a cabo para seleccionar os elementos da mostra de xeito que esta sexa o máis representativa posíbel. A mostraxe probabilística caracterízase por coñecermos previamente á escolla dunha mostra concreta a probabilidade que ten de ser seleccionada cada unha das posíbeis.

Se unha vez escollido un elemento para formar parte da mostra, se devolve á poboación de xeito que poida ser novamente seleccionado, diremos que se trata dunha *mostraxe con repetición*. Se o elemento escollido non pode volver ser seleccionado diremos que a mostraxe é *sen repetición*.

Mostraxe aleatoria simple

Dicimos que unha mostraxe é aleatoria simple se todas as mostras, dun determinado tamaño, que se poden extraer dunha poboación teñen a mesma probabilidade de seren seleccionadas. Así, se queremos extraer unha mostra de tamaño n dunha poboación de

tamaño N —sen repetición—, teremos $\binom{N}{n}$ posibilidades de escolla, co cal a probabilidade de elexir unha determinada mostra será $\frac{1}{\binom{N}{n}}$.

Exemplo 1

Nun grupo de 25 persoas o número de mostras de tamaño catro será $\binom{25}{4}=12650$, polo que a probabilidade de escoller a mostra formada precisamente polas persoas A, B, C e D será $1/12650$.

Selección da mostra

Na mostraxe aleatoria simple debe quedar garantida a aleatoriedade da mostra, e para iso hai que contar previamente co *listado* da poboación, que posibilite a potencial escolla de calquera elemento.

Para seleccionar os elementos dunha mostra botábase man das Táboas de números aleatorios. Nelas aparecen os algarismos 0, 1, 2,..., 7, 8, 9, colocados ao chou. Os díxitos poden lerse individualmente ou en grupos, e en calquera orde: en columnas cara abaixo, en columnas cara arriba, en filas, diagonalmente, etc., e están ordenados de

tal maneira que a probabilidade de que apareza cualquiera deles é igual á probabilidade de que apareza cualquiera outro. Ademais, os agrupamentos de certos díxitos teñen a mesma probabilidade de ocorrer que os outros agrupamentos dun número igual de díxitos. Na actualidade adoitan seleccionarse os números aleatorios utilizando a computadora. A diferenza das táboas, que eran resultados de sorteos de lotaría e procesos similares, os números aleatorios das computadoras están producidos por fórmulas, por iso se chaman pseudoaleatorios, mais cumpren os requisitos indispensábeis para que se lles poida considerar aleatorios.

No caso da folla de cálculo a función =ALEATORIO() dá un número aleatorio maior que cero e menor que 1, e a función =ALEATORIO.ENTRE(a;b) dá un número enteiro comprendido entre a e b.

Exemplo 2

Obter unha mostra de tamaño 20 dunha poboación de tamaño 1345, utilizando a folla de cálculo:

Pégase a función =ALEATORIO.ENTRE(1;1345) dende A1 até E4, obténdose nesta ocasión:

79	886	179	885	454
917	183	91	688	1205
98	900	631	829	636
509	1202	1255	485	614

No caso en que aparecese algún número repetido tirárase outro.

Outros tipos de mostraxe

A mostraxe aleatoria simple só é recomendábel no caso en que os elementos sexan doadamente identificábeis e a poboación sexa homoxenea respecto á característica estudada. Ademais, non sempre resulta posíbel, ou é moi custoso, numerar todos os elementos dunha poboación susceptible de estudo para os máis diversos fins: millonarios, indixentes, sardiñas, estreptococos,..., por iso na práctica se utilizan outros métodos como a *mostraxe sistemática*, a *mostraxe estratificada* e a *mostraxe por conglomerados*.

Mostraxe sistemática

Para tirar unha mostra de tamaño n aplicando este método, divídese a poboación en n grupos cada un deles de tamaño m ; daquela $m=N/n$ (se non dá exacto aproxímase ao enteiro seguinte). Tómase ao chou un elemento do primeiro grupo. Se o elemento seleccionado foi o k -ésimo, do segundo grupo escollerase o $k+m$, do terceiro o $k+2m$, e así até completar a mostra.

Exemplo 3

Seleccionar unha mostra de tamaño 7 dun grupo de 25 persoas.

Como $25/7 = 3,57$, facemos $m = 4$, daquela temos:

G1	G2	G3	G4	G5	G6	G7
----	----	----	----	----	----	----

1	5	9	13	17	21	25
<u>2</u>	<u>6</u>	<u>10</u>	<u>14</u>	<u>18</u>	<u>22</u>	...
<u>3</u>	7	11	15	19	23	...
4	8	12	16	20	24	...

Executamos =ALEATORIO.ENTRE(1;4), e neste caso obtivemos 2. Os seis primeiros elementos da mostra serían 2,6,10,14,18,22. Para obter o seguinte, executamos outra vez =ALEATORIO.ENTRE(1;4); nesta ocasión 3, que incluiremos na mostra. Se saíra o 1, incluíríamos na mostra o 25, e se saíra outra vez o 2, volveríamos tirar outro número.

Este método ten certas vantaxes se o criterio de ordenación da lista é tal que os elementos máis parecidos tenden a estar máis próximos, se é así, a mostraxe sistemática adoita ser máis precisa que a aleatoria simple, pero ten tamén o inconveniente de considerar numerada toda a poboación.

Mostraxe estratificada

Para levar a cabo este tipo de mostraxe consideramos a poboación dividida en subpoboacións –*estratos*- o máis homoxeneos posíbeis. Tírase en primeiro lugar unha mostra aleatoria simple de cada estrato. A continuación faise a *afixación*, ou reparto, do tamaño da mostra en cada estrato, que pode realizarse de distintos xeitos. Se a mostra se reparte en partes iguais diremos que a afixación é uniforme. Se o reparto é proporcional ao número de elementos diremos que a afixación é proporcional. Diremos, por último, que a afixación é óptima se o tamaño da mostra en cada estrato é inversamente proporcional á varianza; canto menos homoxeneo sexa un estrato maior será nel o tamaño da mostra. Para definir os estratos cómpre utilizar datos de experiencias anteriores ou tirar unha pequena *mostra piloto*.

Exemplo 4

Pretendemos estudar o gasto mensual en ocio dunha poboación de asalariados. Para iso consideramos catro estratos segundo os intervalos que aparecen na táboa, acompañados das frecuencias relativas correspondentes.

Se queremos tirar unha mostra de tamaño 1000, utilizando a afixación proporcional, repartirémola de acordo coas frecuencias relativas, redondeando para que a suma das submostras en cada estrato sexa 1000.

Salario en €	0-300	301-600	601-1000	1001-1500	>1500	totais
Frec. Relat.	0,06	0,16	0,38	0,31	0,09	1
Tamaño mostra por estrato	62	156	375	313	94	

Mostraxe por conglomerados.

Ás veces a poboación preséntase en grupos que conteñen toda a variabilidade respecto da característica a estudar, daquela poden seleccionarse só algúns destes grupos ou

conglomerados. En cada conglomerado están as unidades elementais, poñamos por caso, as persoas a entrevistar ou as pezas a inspeccionar. O instrumento de medición pode aplicarse ben a todas as unidades de cada grupo, ou só a algunhas deles seleccionadas ao azar. Neste segundo caso dicimos que se trata dunha mostraxe *bietápica*.

O número de etapas pode aumentarse segundo a tipoloxía do problema. Por exemplo, se queremos facer unha mostraxe que teña como unidades os habitantes dun país, pódense seleccionar cidades na primeira etapa, distritos na segunda, seccións na terceira, mázas na cuarta, edificios na quinta e vivendas na sexta.

Unha aplicación senlleira da mostraxe por conglomerados é o control de calidade. Supoñamos que determinado tipo de conserva se embala en caixas de a cen. Para inspeccionar as latas podemos seleccionar caixas (unidades de primeira etapa), e despois, dentro de cada caixa seleccionar as correspondentes latas (unidades de segunda etapa).

Mostraxes non probabilísticas.

Nalgunhas ocasións a escolla da mostra baséase na experiencia do propio investigador, quen decide segundo as súas tendencias e coñecementos se as unidades seleccionadas son representativa ou non. Aínda que esta estratexia deixe moito que desexar desde o punto de vista estatístico, utilízase con certo éxito para determinadas inspeccións, por exemplo as fiscais.

A cotío certas investigacións espurias preséntanse como científicas, aínda que na escolla da mostra se tivese moito cuidado en ocultar aqueles efectos ou opinións contrarias á hipótese que se trata de contrastar. Fronte a este tipo de mostraxe *opinática* debemos permanecer especialmente críticos.

Non é doado discernir á vista duns datos se proceden dunha mostraxe rigorosa, ora ben, sen ningún tipo de dúbida débense cumprir os seguintes criterios:

- 1) A poboación ten que estar definida sen ambigüidade.
- 2) Debemos ter acceso potencial a cada un dos elementos da poboación
- 3) Temos que coñecer a probabilidade de elección de cada elemento poboacional.

Se sometemos ao anterior filtro as mostraxes televisivas e as das páxinas web, podemos concluir que a case totalidade delas non son científicas, entendendo por isto que non seguen os cánones da teoría da mostraxe estatística.

No que segue imos tratar algúns aspectos da *mostraxe aleatoria*.

Exercicios

2. Queremos investigar sobre a utilización do transporte público en determinada cidade. Que plano de mostraxe se che ocorrería propor?

3. Quere se estimar sobre o desprazamento, tempo e medio utilizado, dos traballadores dunha determinada fábrica situada nun polígono industrial. Como se escollería a mostra? Que plano habería que seguir se se quere estender a investigación a todas as fábricas do polígono?

4. Unha fábrica de vidro produce tres tipos de garrafas de capacidade 1, $\frac{1}{2}$ e $\frac{3}{4}$ de litro. A produción diaria é de 50000 do primeiro tipo, 30000 do segundo e 20000 do terceiro. Que tipo de mostraxe sería conveniente neste caso? Como se seleccionaría a mostra?

5. A distribución, en miles, da poboación activa en Galicia no cuarto trimestre de 2009 (datos da EPA) era a seguinte:

Agricultura, gandería, caza e silvicultura	Pesca e acuicultura	Industria	Construción	Servizos	Total
67,4	24,5	199	136,1	809	1.307,30

Razóese como se obtería unha mostra representativa de todos os sectores produtivos.

Mostra aleatoria

Supoñamos unha poboación formada polos seguintes salarios expresados en euros: 500, 1000, 1500, 2000. Nunha mostraxe aleatoria simple con repetición, tomando mostras de tamaño 2, podemos obter 16 valores distintos. A mostra de tamaño 2 é unha variábel aleatoria bidimensional (X_1, X_2) que toma os valores indicados nas dúas primeiras filas da seguinte táboa, onde a terceira está formada pola media de cada mostra - *media da mostra*.

X_1	500	500	500	500	1000	1000	1000	1000	1500	1500	1500	1500	2000	2000	2000	2000
X_2	500	1000	1500	2000	500	1000	1500	2000	500	1000	1500	2000	500	1000	1500	2000
$\bar{X}$	500	750	1000	1250	750	1000	1250	1500	1000	1250	1500	1750	1250	1500	1750	2000

Como vemos, a *media da mostra* toma distintos valores, dependendo de cal fose a mostra escollida. Esta media é unha nova variábel aleatoria, da que podemos calcular a súa distribución de probabilidade sen máis que facer o recuento dos datos da terceira fila:

$\bar{X}_i$	f_i
500	1/16
750	2/16
1000	3/16
1250	4/16
1500	3/16
1750	2/16
2000	1/16

Se a mostraxe é con repetición, existen n^N maneiras distintas de tirar unha mostra de tamaño n dunha poboación de tamaño N . Por exemplo, nunha poboación de tamaño 1000 existiría 2^{1000} mostras de tamaño 2, polo que resultaría completamente imposible tratar de obter empiricamente a distribución da media. Para deseñar estratexias de mostraxe, na práctica, suponse que a poboación segue determinada distribución coñecida, o que permite establecer a *distribución teórica* da media da mostra, tendo en conta todos os resultados teoricamente posibles e as súas probabilidades.

Tamén é certo que utilizando o ordenador ás veces podemos simular a distribución teórica da media a partir da empírica. Por exemplo, se queremos calcular a distribución da media asociada ao experimento botar un dado dúas veces, aínda que os

$2^6 = 64$ casos posibles non serían demasiados para seguir o esquema anterior, podemos utilizar a folla de cálculo da seguinte maneira:

A	B	C
=ALEATORIO.ENTRE(1;6)	=ALEATORIO.ENTRE(1;6)	=(A1+B1)/2

onde en A e B se simula o lanzamento dos dados, e en C se calcula a media das puntuacións. Se extendemos, poñamos por caso até 8000, e facemos a táboa de frecuencias dos valores obtidos en C, temos:

$\overline{X}_i$	f_i
1	0,026625
1,5	0,05335
2	0,0825
2,5	0,113375
3	0,1405
3,5	0,164625
4	0,140375
4,5	0,112
5	0,086875
5,5	0,051375
6	0,02825

que, como podemos comprobar con algo de paciencia, é moi semellante á distribución teórica.

Cando tomamos en consideración determinada poboación para o seu estudo, supomos que virá dada polos valores dunha variábel aleatoria que se comportará de acordo con determinado modelo probabilístico coñecido, como o normal ou o binomial. Se denotamos por X a variábel aleatoria correspondente á característica que se pretende estudar, as n observacións $(x_1, x_2, \dots, x_n)$, que se fan da poboación, poden considerarse como un valor da variábel $(X_1, X_2, \dots, X_n)$, onde as X_i son independentes e seguen a mesma distribución que X .

- O arranxo de variábeis aleatorias $(X_1, X_2, \dots, X_n)$ independentes, todas elas igualmente distribuídas, chámase *mostra aleatoria simple*.
- Calquera expresión que se obteña a partir de $X_1, X_2, \dots, X_n$ mediante operacións, chámase *estatístico*. (Formalmente será unha aplicación de $\Re^n \rightarrow \Re$)

Por exemplo a media da mostra $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$ é un estatístico, como tamén o

$$\text{son: } m_2 = \frac{\sum_{i=1}^n X_i^2}{n}, \quad s^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n}$$

Todas as expresións anteriores son variábeis aleatorias das que a distribución se pode obter a partir da distribución de X .

A poboación que se trata de estudar mediante a mostra terá a súa distribución teórica que quedará inequivocamente definida mediante determinadas cantidades (media, desvío,...) que se chaman *parámetros*. Por exemplo, se establecemos a hipótese de que unha poboación se distribúe normalmente, esta quedará completamente identificada unha vez se coñeza a súa media e o seu desvío.

Os parámetros estimaranse mediante os *estadísticos*. Supoñamos que temos unha poboación de persoas das que queremos coñecer a súa altura media. Se tomamos unha mostra de tamaño 50, e obtemos un valor $\bar{X} = 174$ cm, diremos que 174 é unha estimación da media. Para distintas mostras obteremos distintos valores de $\bar{X}$. É dicir, o valor do estatístico $\bar{X}$ muda de mostra en mostra e, como a continuación veremos, a media de todos os valores que toma é μ , a media poboacional.

A *inferencia estatística* estuda a problemática da estimación dos parámetros poboacionais a partir das mostras, dotándonos dos instrumentos necesarios para discernir entre as malas e as boas estimacións.

Esperanza e varianza da media da mostra

Consideremos a poboación, á que fixemos referencia anteriormente, formada polos salarios: 500, 1000, 1500, 2000. Mediante cálculos sinxelos podemos ver que $\mu = 1250$ e $\sigma^2 = 312500$.

A distribución da media da mostra figura nas dúas primeiras columnas da táboa. Engadimos outras dúas para o cálculo da súa media (media da media da mostra) e da súa varianza (varianza da media da mostra).

$\bar{X}_i$	f_i	$\bar{X}_i \cdot f_i$	$(\bar{X}_i - \mu)^2 \cdot f_i$
500	0,06	31,25	35156,25
750	0,13	93,75	31250
1000	0,19	187,5	11718,75
1250	0,25	312,5	0
1500	0,19	281,25	11718,75
1750	0,13	218,75	31250
2000	0,06	125	35156,25
Totais	1,00	1250	156250

Obtemos $\mu_{\bar{X}} = 1250$ e $\sigma_{\bar{X}}^2 = 156250$.

Se comparamos os valores anteriores cos respectivos valores poboacionais, observamos que $E(\bar{X}) = \mu_{\bar{X}} = \mu = 1250$, e $V(\bar{X}) = \sigma_{\bar{X}}^2 = 156250 = 312500/2 = \sigma^2/2$.

Comprobamos neste caso que a esperanza da media da mostra coincide coa media poboacional, mentres que a varianza da media da mostra é igual á varianza poboacional dividida polo tamaño da mostra.

Exercicio 6.

Facemos tres extraccións **sen repetición** da seguinte poboación formada por pesos en quilos: 50, 65, 70 e 80.

- Calcúlese a media poboacional.
- Calcúlese a varianza poboacional.
- Cantas mostrase de tamaño 3 non teñen elementos repetidos?
- Escríbanse todas as mostrase posibles.
- Calcúlese a media de cada mostra e fágase a correspondente distribución de frecuencias da variábel *media da mostra*.
- Calcúlese, como se fixo no anterior exemplo $\mu_{\bar{X}}$ e $\sigma_{\bar{X}}$, e compóbase que: a esperanza coincide coa media poboacional e a varianza é igual á poboacional partida por 3 (tamaño da mostra), e multiplicada por $(N-n)/(N-1)$. (Neste caso $N = 4$ e $n = 3$).

En xeral temos o seguintes resultados:

> Se tomamos mostrase aleatorias, con repetición, de tamaño n dunha poboación con media μ e desvío σ , temos $E(\bar{X}) = \mu_{\bar{X}} = \mu$ e $V(\bar{X}) = \sigma_{\bar{X}}^2 = \sigma^2 / n$.

$\sigma / \sqrt{n}$ chámase *erro medio* da mostra.

> Se a mostrase é sen repetición, temos: $E(\bar{X}) = \mu_{\bar{X}} = \mu$ e $V(\bar{X}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$.

O *erro medio* neste caso é: $\sigma(\bar{X}) = \sqrt{\frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}}$.

O cociente $\frac{N-n}{N-1}$ chámase factor de corrección da mostrase sen repetición. Adoita desprezarse se $n/N < 0,05$.

O erro medio será tanto menor canto maior sexa o tamaño da mostra, e daquela máis imitará a media da mostra á da poboación.

Distribución da media da mostra nunha poboación normal

Supoñamos unha poboación que se distribúe normal con media μ e desvío σ . Se escollemos mostrase sen repetición, demóstrase que a media da mostra $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$

tamén se distribúe normal con media μ e desvío $\frac{\sigma}{\sqrt{n}}$. Daquela, $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$

distribúese $N(0,1)$.

Admítese na práctica que moitas poboacións se distribúen normalmente polo que determinadas situacións problemáticas poden formularse e resolverse botando man desta hipótese, obteñéndose á par resultados de gran poder prognóstico que ao final do proceso son subsceptíbeis de ser contrastados experimentalmente.

Exemplo 5

Supoñamos que os paquetes de café envasados por determinada máquina seguen unha distribución normal de media 250g e desvío 5 g. Para facer un control de calidade tírase unha mostra de dez paquetes.

A distribución da media da mostra será $N(250, \frac{5}{\sqrt{10}}) = N(250, 1,58)$

Por exemplo, a probabilidade de que a media da mostra tome valores superiores a 252 gr será:

$$\begin{aligned} P(\bar{X} \geq 252) &= 1 - P(\bar{X} \leq 252) = 1 - P\left(\frac{\bar{X} - 250}{1,58} \leq \frac{252 - 250}{1,58}\right) = \\ &= 1 - 0,89721167 = 0,10278833 \end{aligned}$$

Teorema central do límite

Un dos resultados máis importantes da estatística é o que se coñece como *teorema central do límite* que di o seguinte:

Se $X_1, X_2, \dots, X_n$ son variábeis aleatorias independentes que teñen a mesma función de distribución –e polo tanto a mesma media μ e o mesmo desvío σ , entón, cando $n \rightarrow \infty$, $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ aproxímase a unha normal con media μ e desvío $\frac{\sigma}{\sqrt{n}}$.

No caso particular de que as variábeis $X_1, X_2, \dots, X_n$ sexan dicotómicas, cada unha delas con media P e varianza PQ , entón p , a proporción da mostra, tende a unha $N\left(P, \sqrt{\frac{PQ}{n}}\right)$, cando $n \rightarrow \infty$.

Exemplo 6

Consideremos a poboación formada polos 100 primeiros números naturais. A súa media é 50,5 e o seu desvío 28,9. Utilizando a folla de cálculo extraemos unha mostra de tamaño 9 utilizando a función:

=ALEATORIO.ENTRE(1;100),

que extendemos desde A1 até I1. En J1 inserimos a función =PROMEDIO(A1:I1), que extendemos até a fila 3000.

Obtemos:

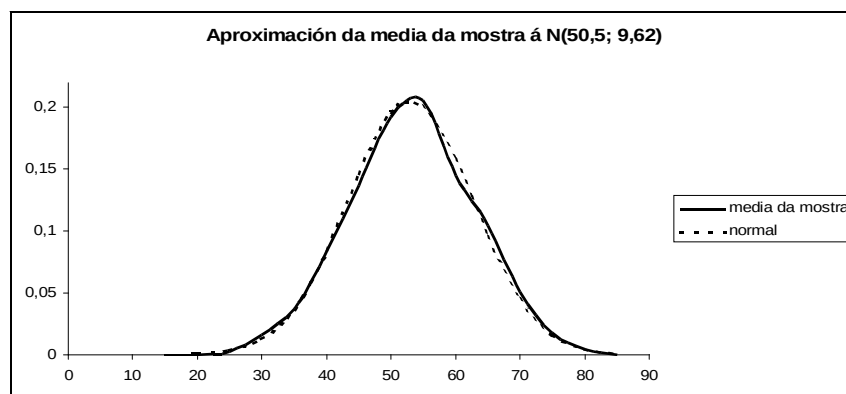
	A	B	C	D	E	F	G	H	I	J
1	72	17	2	9	92	50	71	60	75	49,78
2	82	10	17	85	94	31	46	17	29	45,67

2999	29	86	60	85	80	14	54	63	82	61,44
3000	48	89	14	19	70	72	82	89	6	54,33

Na columna J está a media de cada mostra, que podemos tabular mediante a función FRECUENCIA, obténdose:

Intervalo	F. Absolutas	F. Relativas
(0,15]	0	0
(15,20]	1	0,00033333
(20,25]	8	0,00266667
(25,30]	42	0,014
(30,35]	119	0,03966667
(35,40]	256	0,08533333
(40,45]	467	0,15566667
(45,50]	569	0,18966667
(50,55]	581	0,19366667
(55,60]	460	0,15333333
(60,65]	294	0,098
(65,70]	145	0,04833333
(70,75]	43	0,01433333
(75,80]	12	0,004
(85,100]	3	0,001

Escollendo no menú de gráficos da folla de cálculo, dispersión, observamos como a distribución empírica da mostra tende á normal $\mu, \frac{\sigma}{\sqrt{n}}$.



Distribución da diferenza de medias

En moitas ocasións nas ciencias sociais ou nas da natureza, trátase de comparar dúas poboacións distintas a través dun parámetro poboacional, xeralmente a media ou a proporción.

Se consideramos que ambas as dúas poboacións seguen unha distribución normal de xeito que os parámetros da primeira son μ_1, σ_1 , e os da segunda μ_2, σ_2 , demóstrase que :

$\bar{X}_1 - \bar{X}_2$ se distribúe segundo $N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$, sendo n_1 e n_2 os tamaños das respectivas mostras.

Exercicios

7. A vida media dunha especie de bolboreta é de 24 horas, cun desvío de 5 horas. Observadas 40 delas en cativeiro comprobouse que viviron 18 horas, cal é a probabilidade de que se obtivese un valor menor ou igual a este para a media da mostra?

8. Os seguintes datos representan os ingresos expresados en miles de euros de cinco familias: 1, 2, 3, 4, 5. Escríbanse todas as mostras con repetición de tamaño 2 e calcúlese:

- A media e o desvío poboacional.
- A distribución de frecuencias da media da mostra.
- A media do estatístico media da mostra.
- O erro estimado da media.

e) Compóbase que $E(\bar{X}) = \mu$ e $V(\bar{X}) = \frac{\sigma^2}{2}$

9. Fágase o mesmo que no exercicio anterior con mostras sen repetición de tamaño 3.

Cúmprese $V(\bar{X}) = \frac{\sigma^2}{3}$?

10. Calcúlese a probabilidade de que ao extraer unha mostra de tamaño 100 dunha poboación $N(10,9)$ a media da mostra estea comprendida entre 6 e 12.

11. Sabendo que o 75% dos alumnos de bacharelato saen de noite as fins de semana, se tomamos unha mostra de tamaño 200, cal é a probabilidade de que saian máis de 110?

12. O contido en nicotina de determinada marca de tabaco distribúese normal con media 12 e desvío 3. Tomada unha mostra de 40 cigarros, cal é a probabilidade de obter unha media maior que 14 mg.?

13. O director dunha oficina bancaria afirma que o tempo medio que un empregado necesita para xestionar a apertura dunha conta é de 20 minutos, cun desvío de 5 minutos. Observados datos relativos á apertura de 50 contas, obtívose unha media de 25 minutos e medio. Ten motivos o director para cambiar de opinión?

14. A media dun exame foi 5,4 cun desvío de 2,6. Cal debe ser o tamaño dunha mostra para que $P(\bar{X} \leq 4) = 0,2$?

15. Sabendo que a poboación de cochenos é do 20%, se tomamos unha mostra aleatoria simple de tamaño 100, cal é a probabilidade de obter máis de 50 cochenos?

16. Nas últimas eleccións municipais o partido A levou o 28% dos votos. Preguntados aleatoriamente 50 cidadáns con dereito a voto, 13 deles declararon que ían votar polo partido A. Hai motivos para pensar que o partido vai mellorar os resultados?

6. ESTIMACIÓN

Poderías preguntarlle que significa ter unha seguridade do noventa e cinco por cento. “Significa que estarei no certo polo menos en dezanove de vinte casos, ou, se non en dezanove de vinte, en dezanove mil de vinte mil”, responderá o investigador. E que caso é o presente?, preguntarlle: o decimo noveno ou o vixésimo, o décimo noveno milésimo ou o vixésimo milésimo?

J. M. Coetzee. *Diario dun mal ano*.

Na práctica, aínda que se saiba a distribución seguida por determinada poboación, descoñécese o valor dos seus parámetros (media, desvío,...), información indispensable para tomar decisións. O proceso de establecer valores “aproximados” dos parámetros poboacionais recibe o nome de *estimación estatística*, e pode ser de dous tipos: *puntual* ou *por intervalos*.

A *estimación puntual* consiste en dar un valor que se vai tomar como aproximación do correspondente parámetro poboacional, polo xeral: a proporción, a media ou o desvío. Así, por exemplo se as notas de determinado exame se distribúen normal e obtemos como media dunha mostra 5,7, diremos que esta é unha estimación da media poboacional.

Na *estimación por intervalos* o valor prognóstico para o parámetro poboacional acompáñase dun intervalo con centro nel, e da probabilidade de que o parámetro pertenza a ese intervalo. Seguindo co exemplo anterior, unha aseveración para este tipo de estimación sería: “a media poboacional pertence ao intervalo [5,5 , 5,9] cunha probabilidade do 95%”

Estimación puntual

Dada unha mostra aleatoria simple de tamaño n , lembremos que definiramos *estatístico* como calquera función de $\Re^n \rightarrow \Re$ cuxa expresión se obtivese a partir de $(X_1, X_2, \dots, X_n)$. Así, poñamos por caso, $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ e $s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$ son estatísticos

Un *estimador puntual* para un parámetro poboacional θ (media, proporción, desvío,...), será un estatístico que nos permita calcular valores *aproximados* de θ . Un valor do estimador obtido a partir dunha mostra concreta, diremos que é unha *estimación* de θ .

Escribimos ás veces $\hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = s^2$ ou $\hat{P} = p$, para indicar que o respectivo estatístico se utiliza como estimador do correspondente parámetro poboacional.

Tipos de estimadores

Ao tomarmos como valor do parámetro poboacional a estimación obtida a partir da mostra, prodúcense erros que serán tanto menores canto maior sexa o tamaño da mostra. Na práctica, a mostra non poderá ser tan grande como un quixer, debido a limitacións temporais e económicas, polo tanto haberá que ter coidado para escoller entre os posíbeis estatísticos aqueles que sexan bos estimadores.

Consideremos unha poboación que segue unha distribución dependente dun parámetro θ (media, proporción, desvío,...). Se tiramos unha mostra de tamaño n mediante o correspondente estimador $\hat{\theta}$ obteremos un valor aproximado de θ . Se por exemplo a poboación está formada polos 100 primeiros números naturais e para estimar a media tiramos unha mostra de tamaño 5, podemos obter

96	87	48	13	76
----	----	----	----	----

e, xa que logo, $\hat{\theta} = \bar{X} = 64$.

Tirando 20 mostras, obtemos distintas estimacións da media poboacional; neste caso obtivemos:

64	27,4	49,4	54,2	37,4	44	62	67,2	54	55,4	56,8	44,8	59,8	57	25	63,8	48,4	46,8	67	59,4
----	------	------	------	------	----	----	------	----	------	------	------	------	----	----	------	------	------	----	------

Como nesta ocasión coñecemos o valor real da media poboacional 50,5, podemos calcular a distancia, d , de cada estimación á media, que toma os valores:

13,5	2,7	-0,7	7,7	24,3	7,5	-7,9	3,5	18,1	1,3	7,7	18,7	15,3	12,1	19,7	20,7	-8,7	4,7	17,9	2,7
------	-----	------	-----	------	-----	------	-----	------	-----	-----	------	------	------	------	------	------	-----	------	-----

No caso xeral teremos: $d = \hat{\theta} - \theta$, que tomará valores distintos para as distintas mostras – é unha variábel aleatoria!-. Do que se trata é de escoller os estimadores $\hat{\theta}$ que minimicen esta distancia. Ora ben, para que non se compensen os valores positivos cos negativos, consideramos d^2 en vez de d , que tamén será unha variábel aleatoria. Se calculamos a súa esperanza, temos:

$$E(d^2) = E\left[(\hat{\theta} - \theta)^2\right] = V(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

$E\left[(\hat{\theta} - \theta)^2\right]$ chámase *erro cuadrático medio* do estimador e os bos estimadores serán aqueles que o minimicen.

$(E(\hat{\theta}) - \theta)^2$ chámase *nesgo* do estimador.

Un estimador, xa que logo, será tanto mellor canto menores sexan o seu *nesgo* e a súa varianza.

En xeral, diremos que $\hat{\theta}$ é un estimador *non esguellado* ou *innesgado* de θ se $E(\hat{\theta}) = \theta$.

Exemplo 1

Supoñamos unha poboación formada polos seguintes pesos expresados en quilos: 65, 70, 75, 90. A media poboacional é 75 e a varianza 87,5. Se extraemos mostras con repetición de tamaño 2, temos a seguinte distribución:

X_1	X_2	$\bar{X}$	S^2
65	65	65	0
65	70	67,5	6,25
65	75	70	25
65	90	77,5	156,25
70	65	67,5	6,25
70	70	70	0
70	75	72,5	6,25
70	90	80	100
75	65	70	25
75	70	72,5	6,25
75	75	75	0
75	90	82,5	56,25
90	65	77,5	156,25
90	70	80	100
90	75	82,5	56,25
90	90	90	0
medias		75	43,75

A esperanza (media) da *media da mostra*, 75, coincide coa media poboacional. Xa que logo, a media da mostra é un estimador innesgado da media poboacional. Porén, a esperanza da varianza da mostra é 43,75, distinta da varianza poboacional, 87,5.

Imos comprobar que $s^2 = \frac{n}{n-1} S^2$ é un estimador innesgado de σ^2 .

Como neste caso $n=2$, $s^2 = 2S^2$, temos:

																	media
S^2	0	6,25	25	156,25	6,25	0	6,25	100	25	6,25	0	56,25	156,25	100	56,25	0	43,75
s^2	0	12,5	50	312,5	12,5	0	12,5	200	50	12,5	0	112,5	312,5	200	112,5	0	87,5

co cal $E(s^2) = \sigma^2$.

s^2 é un estimador *non esguellado* da varianza poboacional que se chama *cuasi-varianza*. Dado que é o mellor estimador para σ^2 , polo xeral, sempre que non haxa ambigüidades, referirémonos a el como *varianza da mostra*.

Fixemonos en que:

$$s^2 = \frac{n}{n-1} S^2 = \frac{n}{n-1} \frac{\sum (X_i - \bar{X})^2}{n} = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

Exercicios

4. Utilizando os datos do exemplo anterior, e en mostras de tamaño 2, compróbase se a moda e a media xeométrica son estimadores innegados dos correspondentes parámetros poboacionais
5. Calcúlese a varianza da mostra (cuasi-varianza) de 23,56,45,32,43,28.

Xa dixemos antes que o erro cometido será menor canto máis pequena sexa a varianza do estimador. Daquela diremos que un estimador é máis **eficiente** que outro se a varianza do primeiro é menor que a do segundo. Demóstrase que entre todos os estimadores da media, $\bar{X}$ é o máis eficiente.

Exercicio 6

Dada a poboación de alturas en centímetros: 170, 180, 190, 200, calcúlese en mostras de tamaño 3 a distribución da media e da mediana. Compróbase que ambos estatísticos son estimadores innegados da media e que a media é máis eficiente que a mediana.

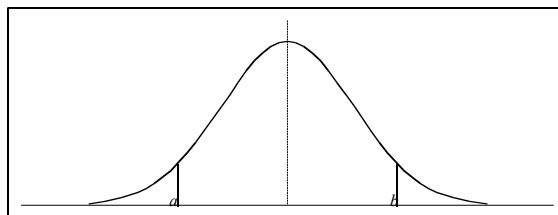
Diremos que un estimador é **consistente** se a medida que aumenta o tamaño da mostra faise maior a probabilidade de que o valor acadado se achegue ao parámetro. No caso en que o estimador sexa non esguellado demóstrase que para ser consistente abonda con que a súa varianza tenda a cero cando n tende a infinito.

Dado que $V(\bar{X}) = \sigma^2 / n$, entón a media da mostra é un estimador consistente da media poboacional. Tamén se demostra que s^2 é un estimador consistente de σ^2 .

Intervalos de Confianza

Calcular un *intervalo de confianza* para un estimador de determinado parámetro poboacional consiste en determinar un intervalo $[a, b]$ tal que o parámetro pertenza a el cunha probabilidade dada.

Consideremos unha poboación de prezos de pisos normalmente distribuídos con media μ e desvío σ . Tomada unha mostra da poboación de tamaño 36, obtivemos $\bar{X} = 182000\text{€}$. Dado que $\bar{X} \rightarrow N(\mu, \sigma / \sqrt{36})$, podemos calcular a e b tales que: $P(a \leq \bar{X} \leq b) = 0,90$, escollendo entre os infinitos intervalos que cumpren tal condición o simétrico respecto á media.



Tendo en conta $P(\sqrt{n} \frac{a-\mu}{\sigma} \leq \sqrt{n} \frac{\bar{X}-\mu}{\sigma} \leq \sqrt{n} \frac{b-\mu}{\sigma}) = 0,90$, buscando nas táboas da $N(0,1)$ e considerando un intervalo simétrico:

$$P(-1,65 \leq \sqrt{n} \frac{\bar{X}-\mu}{\sigma} \leq 1,65) = 0,90.$$

De onde:

$$\begin{aligned} P(-1,65 \leq \sqrt{n} \frac{\bar{X}-\mu}{\sigma} \leq 1,65) &= P(-1,65 \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq 1,65 \frac{\sigma}{\sqrt{n}}) = \\ P(-\bar{X} - 1,65 \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X} + 1,65 \frac{\sigma}{\sqrt{n}}) &= P(\bar{X} - 1,65 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1,65 \frac{\sigma}{\sqrt{n}}) = 0,9 \end{aligned}$$

Daquela, concluímos que o prezo medio poboacional pertencerá ao intervalo

$$(\bar{X} - 1,65 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1,65 \frac{\sigma}{\sqrt{n}}) = (182000 - 0,275\sigma \leq \mu \leq 182000 + 0,275\sigma)$$

cunha probabilidade do 90%.

Se o desvío poboacional fose coñecido, poñamos por caso $\sigma = 100000$, poderíamos concluir que (154500, 209500) é un intervalo de confianza do 90% para a media poboacional, **o que significa que de 100 mostras que collésemos, en 90 dos casos a media pertencerá ao intervalo así calculado.**

En xeral, para unha poboación normal de media μ e desvío σ o intervalo de confianza do 90%, para unha mostra de tamaño n ven dado por:

$$\left[\bar{X} - 1,65 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1,65 \frac{\sigma}{\sqrt{n}} \right]$$

Se tomamos o nivel de confianza do 95%, teremos o intervalo

$$\left[\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \right]$$

Aínda que a poboación non sexa normal, se a mostra é suficientemente grande, como consecuencia do teorema central do límite, aceptárase a hipótese de normalidade.

Exercicios

7. Supoñamos que o desvío dunha poboación de arandelas que se distribúe normal é 0,1mm. Obtéñanse intervalos para os niveis de confianza 0,90, 0,95 e 0,99 en mostras de tamaño n .

8. Nas mesmas hipóteses do exercicio anterior obtéñanse intervalos para mostras de tamaño 10, 20 e 30, tomando como nivel de confianza 0,95. Como varía a lonxitude do intervalo a respecto do tamaño da mostra?

Dos anteriores exercicios tiramos en conclusión dúas propiedades fundamentais dos intervalos de confianza:

- 1) Para un tamaño da mostra e un desvío dados, a lonxitude do intervalo aumenta a medida que o fai o nivel de confianza.
- 2) Para un nivel de confianza e un desvío dados, ao aumentar o tamaño da mostra diminúe a lonxitude do intervalo.

Intervalo de confianza para a media dunha poboación normal da que non se coñece a varianza.

Na maioría das ocasións nas que cómpre achar un intervalo de confianza, a varianza poboacional descoñécese. Ora ben, como $s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$ é un estimador innesgado de σ^2 , estimaremos o desvío poboacional mediante o desvío da mostra s .

William Sealy Gosset máis coñecido por *Student*, nome co que asinaba os seus traballos, propuxo o estatístico $t_{n-1} = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ para calcular intervalos de confianza para a media poboacional no suposto de varianza descoñecida, en poboacións normais.

Demostrou que t_{n-1} segue unha distribución *t de student con n-1 graos de liberdade*. Razoando como no caso da varianza poboacional coñecida, fixado un nivel de confianza $1 - \alpha$, para un tamaño de mostra n , obtense o intervalo:

$$\left[\bar{X} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$$

onde s é o cuasi-desvío da mostra e $t_{1-\frac{\alpha}{2}}$, neste caso igual a $t_{\frac{\alpha}{2}}$ por ser simétrica a función de densidade, a abscisa correspondente á distribución *t de Student* con $n-1 = 15$ graos de liberdade, para o nivel de confianza $1 - \alpha$.

Exemplo 2

Os seguintes datos corresponden ás alturas, expresadas en cm, dunha mostra 12 persoas, procedente dunha poboación normal.

151	184	168	177	183	178	152	177	181	161	186	167
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Determinemos un intervalo de confianza do 90% para a media.

Temos: $1 - \alpha = 0,90 \Rightarrow \alpha = 0,10$, graos de liberdade $= 12 - 1 = 11$

Utilizando a folla de cálculo:

$$t_{1-\alpha/2} = \text{DISTR.T.INV}(.0,10;11) = 1,7958837$$

(Inserimos DISTR.T.INV(α ;11) [introdúcese $\alpha = 0,10$ e non $1 - \alpha = 0,90$], obtendo o valor $t_{1-\frac{\alpha}{2}} = 1,7958837$).

Utilizando as táboas:

$1 - \frac{\alpha}{2} = 1 - \frac{0,10}{2} = 0,95$. Buscando na fila do 11 e na columna de $t_{0,95}$, obtemos: 1,80 (a folla de cálculo dá un resultado máis preciso).

Tendo en conta que $\bar{X} = 172,83$ e $s = 13,05$, temos:

$$\left[172,83 - 1,7958837 \cdot \frac{13,05}{\sqrt{11}}, 172,83 + 1,7958837 \cdot \frac{13,05}{\sqrt{11}} \right] = [165,77; 179,90]$$

Mostras grandes

Ao aumentar o tamaño da mostra, n , a distribución t de *Student* aproxímase a unha normal de media cero e desvío 1. Daquela, se as mostras son grandes (na práctica $n \geq 30$) utilízase a normal en troques da distribución t .

Exemplo 3

Consideremos a seguinte distribución de datos, agrupados por intervalos de clase, correspondente ao gasto semanal en combustíbel, en litros, de 40 vehículos.

Intervalo de clase	Marcas de clase	frecuencias
[0,20]	10	4
(20,40]	30	8
(40,60]	50	16
(60,80]	70	7
(80,100]	90	5

Temos: $\bar{X} = 50,5$ e $s = 10,13$. Utilizando a aproximación normal, por ser $n > 30$, para o nivel de confianza 0,95 ($1 - \frac{\alpha}{2} = 1 - \frac{0,05}{2} = 0,975$), obtemos $z_{0,95} = 1,96$, co cal o correspondente intervalo de confianza para a media será:

$$\left[50,5 - 1,96 \cdot \frac{10,13}{\sqrt{40}}, 50,5 + 1,96 \cdot \frac{10,13}{\sqrt{40}} \right] = [53,64; 57,36]$$

Exercicios

9. Unha multinacional de servizos modifica o sistema de xestión informática. Para comprobar se a adaptación ao novo sistema tivo éxito, escóllese unha mostra aleatoria de 15 traballadores e tomase nota do tempo que tardaron en adaptarse. Os resultados, en horas, foron os seguintes: 3,3 2,9 4,3 2,6 3,2 4,1 4,9 2,8 5,5 5,3 3,6 3,5 2,9 4,7. Determínese un intervalo de confianza ao 99% para o “verdadeiro” tempo medio de adaptación.

Intervalo de confianza para a proporción

Para establecer un intervalo de confianza para a proporción poboacional, P , debemos ter en conta que a proporción non é outra cousa que a media da mostra se a variábel é dicotómica, e dicir: $\bar{X} = p$, e ademais: $\frac{s}{\sqrt{n}} = \sqrt{\frac{pq}{n}}$. Daquela, supoñendo a mostra suficientemente grande para ter garantida a hipótese de normalidade, fixado o nivel de confianza $1-\alpha$, o intervalo buscado será:

$$\left[p - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}}, p + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}} \right]$$

Exemplo 4

Queremos estimar a proporción de votantes do partido A nunha próxima convocatoria electoral. Para iso tomamos unha mostra aleatoria simple de tamaño 100, da que se obtén un valor da proporción da mostra de 0,19.

Tomando o nivel de confianza 0,99, obtemos $z_{1-\frac{0,01}{2}} = z_{0,995} = 2,57$, co cal se obtén o intervalo $[0,18, 0,20]$.

Exercicios

10. Escollida unha mostra aleatoria da poboación activa de 16 a 24 anos, de tamaño 135, obtense como resultado que 48 están no paro. Constúase un intervalo de confianza do 95% para a proporción poboacional.

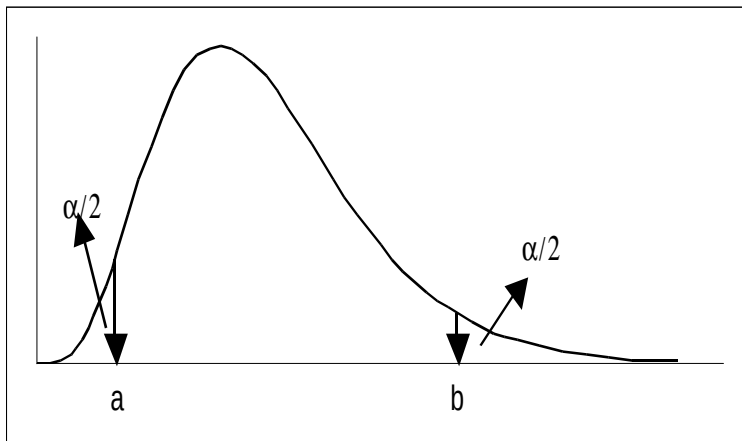
Intervalo de confianza para a varianza poboacional

En moitas ocasións, tanto en procesos de produción como nas investigacións máis diversas, a varianza poboacional non se coñece, polo que cómpre estimala, non só puntualmente, senón dando un intervalo ao que pertenza con determinado nivel de confianza. Demóstrase que se s^2 é a varianza dunha mostra aleatoria simple de tamaño

n , procedente dunha poboación normal, entón un intervalo de confianza $1-\alpha$ para σ^2 ven dado por:

$$\left[(n-1) \frac{s^2}{a}, (n-1) \frac{s^2}{b} \right]$$

onde a e b son tais que $1-\alpha/2 = F(a)$ e $\alpha/2 = F(b)$, sendo $F(x)$ a distribución χ^2 con $n-1$ graos de liberdade.



Exemplo 4

Consideremos a poboación formada polos gastos en ocio de fin de semana das familias de determinada cidade, expresados en euros, que suporemos se distribúen normalmente.

Tiramos unha mostra de tamaño catro:

60	70	65	62
----	----	----	----

Co cal $\bar{X} = 64,75$ e $s^2 = 48,25$.

Calculemos un intervalo de confianza para a varianza desta poboación a un nivel de confianza do 90%.

Temos :

$$\alpha = 0,10 \ ; \ \alpha/2 = 0,05$$

Mediante a folla de cálculo, utilizando a función PRUEBA.CHI.INV, obtemos:

$$a = \text{PRUEBA.CHI.INV}(0,05;3) = 7,8147247$$

Para calcular b , teñamos en conta que $1 - \alpha/2 = 1 - 0,10/2 = 0,95$, daquela:

$$b = \text{PRUEBA.CHI.INV}(0,95;3) = 0,35184596$$

O intervalo será:

$$\left[(4-1) \frac{48,25}{7,8147247}, (4-1) \frac{48,25}{0,35184596} \right] = [18,5; 411,4].$$

Como se pode observar, o pequeno tamaño da mostra dá lugar a un intervalo demasiado amplo, a partir do cal dificilmente se poden tirar conclusións respecto da varianza poboacional.

Os valores a e b tamén poderían obterse a partir das táboas, buscando na columna correspondente a $\chi^2_{0,05}$, e na fila correspondente a 3 graos, obtemos $b = 0,352$. Seguindo a mesma fila, na intersección coa columna $\chi^2_{0,95}$, obtemos $a = 0,81$.

Exercicio 11

Un fabricante de baterías de coche asegura que duran catro anos, cun desvío de 1 ano. Tírase unha mostra de cinco baterías, que duraron 3, 5, 5,8 6,4 e 8 anos. Determínese un intervalo de confianza para σ^2 e indíquese se é válida a afirmación do fabricante.

Intervalo para a diferenza de medias

En ocasións interesa definir un intervalo que permita establecer cales son os valores aceptables para a diferenza entre as medias de dúas poboacións, supoñendo as mostras independentes, e sendo as poboacións de orixe normais ou aproximadamente normais:

I) Varianzas coñecidas.

Tendo en conta a distribución da diferenza de medias de dúas mostras procedentes de poboacións normais, un intervalo de confianza para a diferenza de medias poboacionais virá dado por

$$\left[\bar{X}_1 - \bar{X}_2 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{X}_1 - \bar{X}_2 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

II) Varianzas descoñecidas.

1) Pode suporse que **as varianzas de ambas as poboacións son iguais**, o intervalo de confianza para a diferenza de medias poboacionais será:

$$\left[\bar{X}_1 - \bar{X}_2 - t_{1-\frac{\alpha}{2}} S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{X}_1 - \bar{X}_2 + t_{1-\frac{\alpha}{2}} S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

onde $t_{1-\frac{\alpha}{2}}$ é o valor crítico correspondente ao nivel de confianza $1-\alpha$ da distribución t de *Student* con $n_1 + n_2 - 2$ graos de liberdade e S é unha estimación do desvío común a ambas as poboacións obtida a partir das varianzas das dúas mostras mediante:

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Na práctica se n_1 e n_2 son grandes, a distribución t aproxímase, como xa se viu anteriormente, pola distribución normal.

2) Se non hai motivos para supor as varianzas poboacionais iguais, os límites do intervalo de confianza veñen dados por:

$$\left[\bar{X}_1 - \bar{X}_2 - t_{1-\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \bar{X}_1 - \bar{X}_2 + t_{1-\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]$$

Neste caso o valor de $t_{1-\frac{\alpha}{2}}$ corresponde a unha distribución t con k graos de liberdade. k calcúlase en base a ambos os dous tamaños das mostras e aos desvíos de cada grupo segundo a seguinte corrección proposta por Dixon e Massey:

$$k = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\left(\frac{S_1^2}{n_1} \right)^2 \left(\frac{1}{n_1 + 1} \right) + \left(\frac{S_2^2}{n_2} \right)^2 \left(\frac{1}{n_2 + 1} \right)} - 2$$

Intervalos de confianza para a diferenza de proporcións

Consideremos dúas poboacións de modo que en cada unha delas estudamos unha variábel aleatoria dicotómica de parámetros respectivos P_1 e P_2 . De cada poboación extraemos mostras de tamaño n_1 e n_2 suficientemente grandes para supoñer que é válida a aproximación mediante a normal. Daquela, un intervalo para a diferenza de proporcións virá dado por:

$$\left[p_1 - p_2 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}, p_1 - p_2 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \right]$$

Tamaño da mostra para a estimación da media

Ao abordarmos o problema da mostraxe desde unha perspectiva realista, é fundamental ter un método para decidir se o tamaño da mostra é o axeitado, xa que pretendemos que sexa suficientemente representativa para a partir dela tirar conclusións sobre toda a poboación. Tamén debemos ter en conta que a operación é demasiado custosa como para poder tomar unha mostra tan grande como quixeramos a prol dun aumento de calidade. O tempo e o diñeiro son elementos a ter sempre en consideración neste contexto.

No caso dunha poboación normal con varianza coñecida, un intervalo para a media viña dado por:

$$\left[\bar{X} - z \frac{\sigma}{\sqrt{n}}, \bar{X} + z \frac{\sigma}{\sqrt{n}} \right]$$

A semi-lonxitude do intervalo chámase erro máximo admisíbel ou *erro de estimación*: $\varepsilon = z \frac{\sigma}{\sqrt{n}}$, e despexando (*): $n = \frac{z^2 \sigma^2}{\varepsilon^2}$.

A anterior expresión só se pode calcular se coñecemos a varianza poboacional, despois de fixado un nivel de confianza e un erro máximo admisíbel. A fórmula é válida para a mostraxe con repetición, ou sen repetición se a poboación é suficientemente grande.

Se a mostraxe é sen repetición a fórmula modifícase do seguinte xeito:

$$n = \frac{\frac{z^2 s^2}{\varepsilon^2}}{1 + \frac{z^2 s^2}{N \varepsilon^2}},$$

onde N é o tamaño da poboación. Aínda así, tamén neste caso se utiliza a fórmula (*) sempre que $n/N < 0,05$.

Nas situacións ás que se ten que enfrontar polo xeral o investigador estatístico non se coñece a varianza poboacional e xa que logo haberá que estimala. Para este cometido, ou ben se utilizan estimacións baseadas en estudos anteriores, ou ben se tira unha pequena mostra preliminar, chamada *mostra piloto*, e estímase o valor de σ^2 a partir de

$$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$

Supoñamos por exemplo que queremos estimar o gasto mensual en alimentación, en euros, das familias de determinada zona residencial, cun erro de 30 €. Tírase para iso unha mostra piloto que deu os seguintes resultados: 500, 700, 650, 400, 550, 800, 900, 620, 500, 475.

Obtemos a partir dela: $\bar{X} = 609,5$ e $s^2 = 24458,06$.

a) Se supoñemos a poboación suficientemente grande, tomando o nivel de confianza 0,95, $n = \frac{z^2 s^2}{\varepsilon^2} = \frac{1,96^2 \cdot 24458,06}{30^2} = 104,3$. Redondeando, xa que n debe ser enteiro: $n = 105$.

A conclusión sería válida sempre que $n/N < 0,05$; ou sexa se o número de familias é maior que $105/0,05 = 2100$.

b) Se o número de familias o fixamos en $N = 500$, teríamos para o mesmo nivel de confianza:

$$n = \frac{\frac{z^2 s^2}{\varepsilon^2}}{1 + \frac{z^2 s^2}{N \varepsilon^2}} = \frac{\frac{1,96^2 \cdot 24458,06}{30^2}}{1 + \frac{1,96^2 \cdot 24458,06}{500 \cdot 30^2}} \approx 87.$$

Tamaño da mostra para a estimación da proporción

Cando se trata de estimar a proporción poboacional a partir da mostra, facendo as pertinentes adaptacións nas fórmulas anteriores, obtemos:

- a) Para o caso “infinito” (poboación suficientemente grande):

$$n = \frac{z^2 pq}{\varepsilon^2}.$$

- b) Para o caso de que o tamaño da poboación sexa N:

$$n = \frac{\frac{z^2 pq}{\varepsilon^2}}{1 + \frac{z^2 pq}{N\varepsilon^2}}$$

Se a varianza poboacional é descoñecida, sempre podemos pórnos no caso máis desfavorábel da máxima varianza, que se acada se $P=Q=0,5$; de onde $n = 0,25 \frac{z^2}{\varepsilon^2}$

Exemplo 5

Quérese investigar que porcentaxe da poboación maior de 65 anos consome alcol. De que tamaño debe tirarse unha mostra se o erro máximo que estamos dispostos a admitir é do 5%?

Se o nivel de confianza é 0,95, supoñendo a poboación suficientemente grande, temos:

$$n = \frac{z^2 pq}{\varepsilon^2} = \frac{1,96^2 \cdot 0,25}{0,05^2} \approx 385$$

Se a poboación fose, poñamos por caso, de 1000 persoas, teríamos:

$$n = \frac{\frac{z^2 pq}{\varepsilon^2}}{1 + \frac{z^2 pq}{N\varepsilon^2}} = \frac{384,16}{1 + \frac{384,16}{1000}} \approx 278.$$

No cálculo do tamaño da mostra para a estimación da proporción, e tamén da media, a maioría das veces adoita tomarse $z = 2$, que responde a un nivel de confianza 0,95449988.

Exercicios

12. Escóllase na aula unha mostra aleatoria de tamaño 10 e estímese a partir dela:
 - a) O peso medio, o desvío e o erro estándar (o desvío do peso medio)
 - b) Fágase o mesmo que no apartado anterior coa idade e coa altura.
13. Entre os estudantes de bacharelato e universitarios dunha cidade, escolléronse 111 ao chou e preguntóuselle se utilizaban o transporte público para acudir ás aulas. Se 47 contestaron afirmativamente, estímese a proporción, o desvío e o erro estándar.
14. Unha mostra aleatoria de 100 veciños de determinada cidade indícanos que camiñan diariamente 0,9 quilómetros cun desvío de 0,7. Calcúlese un intervalo de confianza do 99% para a media de quilómetros que andan diariamente os habitantes da cidade.
15. Supoñemos que a vida útil de determinadas zapatillas de deporte é normal. Se nunha mostra de tamaño 22 obtivemos unha media de 250 horas e un desvío de 25 horas, calcúlese un intervalo de confianza do 98% para a duración media das zapatillas.
16. Dunha mostra de 14 persoas con estudos secundarios obtívose unha media de 30 para número de páxinas que daban lido nunha hora, e un desvío de 10. Obtéñase un intervalo de confianza do 0,99 para a media de páxinas que pode ler nunha hora a poboación con estudos secundarios.
17. Unha enquisa realizada a 134 rapaces comprendidos entre 16 e 18 anos relativa á lectura do xornal deu os seguintes resultados: o 76% non le a prensa, e entre os que a len, o 2% interésase polos asuntos internacionais, mentres ao 85% só lle interesan os deportes. Calcúlese un intervalo do 92% nos seguintes casos:
 - a) Proporción de lectores de deportes.
 - b) Proporción de lectores de asuntos internacionais
 - c) Proporción dos que non len a prensa.
18. Sometidos dous grupos de persoas con alto índice de colesterol a dous tratamentos distintos, obtivéronse os seguintes resultados, expresados en mg/cm^3 :

GRUPO I	279	301	299	324	280	289
---------	-----	-----	-----	-----	-----	-----

GRUPO II	270	293	299	314	317	299	300	302
----------	-----	-----	-----	-----	-----	-----	-----	-----

- Obtéñase un intervalo de confianza do 95% para a diferenza das medias.
19. Supóñase que as alturas da poboación de determinado país se distribúen normal con desvío estándar de 5cm. De que tamaño se debe tirar unha mostra se se quere obter un intervalo do 99% cun erro máximo admisíbel de 0,5?
 20. Determinado laboratorio quere estimar o efecto dunha nova medicina contra as migrañas que por experiencias anteriores se sabe que fará efecto polo menos ao 43% dos doentes. De que tamaño se debe tomar unha mostra a un nivel de confianza 0,97 se se está disposto a admitir un erro do 2%?
 21. Cal debe ser o tamaño da mostra para estimar o gasto mensual en libros por unidade familiar, nun nivel de confianza do 90% e cun erro menor de 10€? [A información necesaria para responder, ou ben pode tirarse dunha mostra piloto –aínda que non sexa moi aleatoria-, ou buscando información ao respecto na rede].
 22. Un empresario, partidario do control total, quere determinar o tempo medio que os empregados necesitan para satisfacer as súas necesidades fisiolóxicas durante a xornada laboral. Que tamaño deberá ter unha mostra se se quere que a media da mostra difira en menos dun minuto da media real? O empresario non está disposto a admitir un desvío poboacional maior que un minuto. (Tómese como nivel de confianza 0,95).
 23. Un distribuidor ofrece ao xerente dun supermercado un produto para comercializar que só pode expoñer ao público en plano experimental durante dúas horas. Pasado ese

tempo ten que decidir se o quere ou non, e en caso afirmativo facer un considerábel desembolso económico. Cal debe ser o número de clientes que se definiu sobre o novo artigo se se quere tomar a decisión cun nivel de confianza do 95% e un erro máximo admisíbel do 3%?

24. Nun centro de investigación pesqueira experimentábase unha nova dieta para lograr o engorde rápido nos polbos. Para iso repártense polbos de similares características en dúas piscinas distintas. Ao cabo dun mes obtéñense os seguintes resultados, en quilos:

(alimentación convencional)	0,95	1,27	1,60	1,40	1,50	1,30	0,97	1,47	1,56	1,34	1,35	1
(alimentación experimental)	1,02	1,45	1,27	1,80	1,30	1,54	1,07	1,90	1,75	1,22	1,65	1,50

Tomando as anteriores mostras como piloto, de que tamaño se debe tirar unha mostra para construír un intervalo do 95% para a diferenza de medias?

25. A ficha técnica dunha enquisa de opinión sobre o voto é a seguinte:

- Poboación: persoas de 18 anos ou máis.
 - Ámbito: Galicia.
 - Mostra: Mil entrevistas cun erro do 3,2% para un nivel de confianza 95,5% e $p = q = 0,5$.
- a) Calcúlese o número de entrevistas necesarias.
- b) Se a unha das preguntas contestou afirmativamente o 23,34% dos entrevistados, cal será o intervalo de confianza para a correspondente proporción poboacional?

7. PROBAS DE HIPÓTESES

As ciencias experimentais teñen a grande vantaxe de que as súas posibilidades de experimentar son infinitas: unha hipótese confirmada pode ser perfeccionada por outras, de xeito que poida xerarse, ao final, unha teoría que explique e teña en conta moitos feitos e cuxa corrección siga a ser mellorábel a medida que aumenten os feitos coñecidos.

Luigi Luca Cavalli Sforza. *A evolución da cultura*

Contraste da proporción

Exemplo 1

Admítese polo xeral que a probabilidade de que un neonato sexa nena ou neno é a mesma. Porén, acudindo á páxina web do *IGE*, atopámonos con que dos 23062 nacementos que tiveron lugar en Galicia no ano 2008, foron nenas 11148, mentres que 11914 foron nenos. Falsarán estes datos a hipótese de que a probabilidade de nacer neno é igual á probabilidade de nacer nena?

Para someter a hipótese a proba consideraremos a variábel aleatoria dicotómica X , que toma os valores 0 ou 1, segundo que o naipele fose neno ou nena.

Consideramos:

$P(X=1) = p$ (proporción de nacemento de nenas)

$P(X=0) = q = 1-p$ (proporción de nacemento de nenos)

e queremos probar a hipótese $P = 0,5$ fronte a alternativa $P < 0,5$ (nacen menos nenas)

Se a hipótese é certa, a proporción da mostra distribuirase

$$N\left(P, \sqrt{\frac{PQ}{n}}\right) = N\left(0,5, \sqrt{\frac{0,5 \cdot 0,5}{23062}}\right)$$

ou o que é o mesmo

$$T = \frac{p - 0,5}{\sqrt{\frac{0,5 \cdot 0,5}{23062}}} = \frac{p - 0,5}{0,003292468} \rightarrow N(0,1)$$

A T chamáremoslle *estatístico de proba*. Para efectuar a proba escolleremos un *valor crítico* c , de maneira que se a proporción de nenas observadas é menor que c rexeitaremos a hipótese, e no caso contrario aceptáremola.

Para determinar c , supoñendo a hipótese certa, a probabilidade de obter na mostra unha proporción de nenas menor que a de nenos debe ser un acontecemento

cunha probabilidade moi pequena α , chamada *nivel de significación*, que adoita tomar os valores 0,1 0,05, 0,01.

Xa que logo, se o que queremos é un c tal que $P(T \leq c) = 0,05$, obtemos $c = z_{0,05} = -1,65$

Como substituíndo polos valores da mostra obtemos $T = -0,01660741 > -1,65$, non se rexeita a hipótese $p=0,5$ e daquela a nosa proba evidencia, para o nivel de significación escollido, que a proporción de nenos é igual á de nenas.

Contraste da media

1. Coñecida a varianza poboacional

Exemplo 2

Para realizar o control de calidade dun proceso de fabricación de parafusos tómanse mostras de tamaño 10 e mídese a lonxitude de cada unidade. Requírese que a súa lonxitude non se desvíe de 24mm en máis de 0,3mm, ficando este desvío garantido de antemán.

Para ver se o proceso cumpre as especificacións requiridas obtívose un valor da media da mostra de 24,5. Xa que logo, á vista deste dato trátase de contrastar a hipótese $\mu = 24$ contra a alternativa $\mu \neq 24$.

Daquela, tendo en conta que $T = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ se distribúe $N(0,1)$, este será o

estatístico de proba. Como o rexeitamento dos parafusos pode ser tanto por defecto como por exceso, se o nivel de significación é 0,05, aceptaremos os valores que pertencen a un intervalo $[-c, c]$ tal que $P(-c \leq T \leq c) = 0,95$, de onde $c = 1,96$.

Para os valores da mostra obtemos $T = \frac{24,5 - 24}{\frac{0,3}{\sqrt{10}}} = 5,20935532 > 1,96$, xa que

logo rexéitase a hipótese, o que quere dicir neste caso que o proceso está fóra de control.

2. Descoñecida a varianza poboacional

Exemplo 3

Unha mostra de nove carretes de tanza presentou unha resistencia media de 4,40 kg e un desvío de 1,15. Supoñendo que a resistencia á tensión é unha variábel aleatoria normal, á vista dos feitos, debemos fiarnos do fabricante cando afirma que a resistencia é de 4,5kg?

Se aceptamos a hipótese $\mu = 4,5$, o estatístico $T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = 3 \frac{\bar{X} - 4,5}{1,15}$

segue unha distribución t con $9-1=8$ graos de liberdade. Este será o estatístico de proba.

Para o nivel de significación 0,05, temos $P(t < c) = 0,05$, de onde $c = -1,86$. Substituíndo no estatístico polo valor obtido para a media da mostra:

$$T = 3 \frac{4,4 - 4,5}{1,15} = -0,26086957 > -1,86 ,$$

polo que non se rexeita a hipótese $\mu = 4,5$.

Exemplo 4

As cualificacións obtidas por 10 alumnos dun numeroso colectivo que realizou determinado exame, que se supoñen normais, foron as seguintes: 4,3 5,7 6,8 3 2 7 6 3 3 9. Trátase de contrastar a hipótese “a media do exame foi 6”, contra a alternativa “a media foi distinta de 6”.

$$\text{Neste caso } T = \sqrt{n} \frac{\bar{X} - \mu}{s} = \sqrt{10} \frac{\bar{X} - 6}{2,27} , \text{ será o estatístico de proba.}$$

Por outra banda, $c / P(-c \leq T \leq c) = 0,95 \Rightarrow c = 2,26$.

Para os valores da mostra obtemos $T = \sqrt{10} \frac{4,98 - 6}{2,27} = -1,42 > -2,26$, xa que logo non se rexeita a hipótese “a media do exame foi 6”.

Definicións

Nos tres exemplos anteriores fixéronse hipóteses estatísticas. Neles estudamos características que seguen determinada distribución, variábeis aleatorias sometidas á influencia dun grande número de factores cuxos efectos non se poden predicir con exactitude. Como vimos, a hipótese aceptábase se as desviacións respecto á media non eran significativas para determinado valor da probabilidade fixado de antemán.

Unha *hipótese estatística* é unha aseveración acerca do valor que poida tomar un parámetro poboacional ou sobre o tipo de distribución que segue determinada poboación.

Unha *proba de hipótese* é un procedemento para determinar se unha hipótese estatística pode ser aceptada ou rexeitada. A hipótese dirase *simple* se se especifica un único valor para o parámetro. En caso contrario dirase *composta*.

A hipótese que se somete a proba (no exemplo primeiro $p=0,5$) chámase *hipótese nula*; polo xeral denotarémola H_0 . A suposición respecto da cal se proba chamarémola *hipótese alternativa* (no mesmo exemplo $p<0,5$) e referirémonos a ela como H_1 . Utilizando esta notación, no primeiro exemplo contrastamos $H_0(p=0,5)$ contra $H_1(p<0,5)$.

Exercicio 1

Identifíquense as hipóteses nulas e as alternativas nos exemplos anteriores.

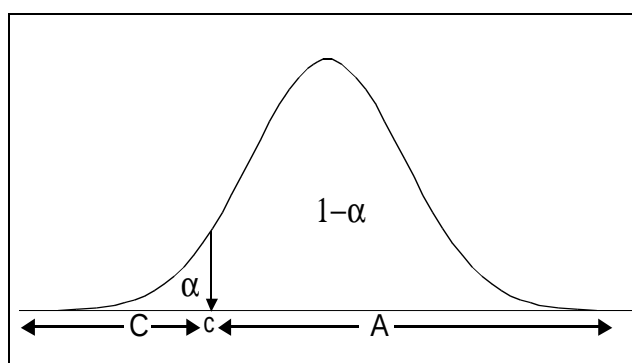
O número α , do que depende a probabilidade coa cal a hipótese se somete a proba, chámase *nivel de significación*. c e $-c$ chámanse *valores críticos*.

O intervalo de valores que toma o *estatístico de proba* T para os que se acepta H_0 , chámase *rexión de aceptación* (A), mentres que o intervalo, ou intervalos, para os que se rexeita, chámase *rexión crítica* (C).

Exercicio 2.

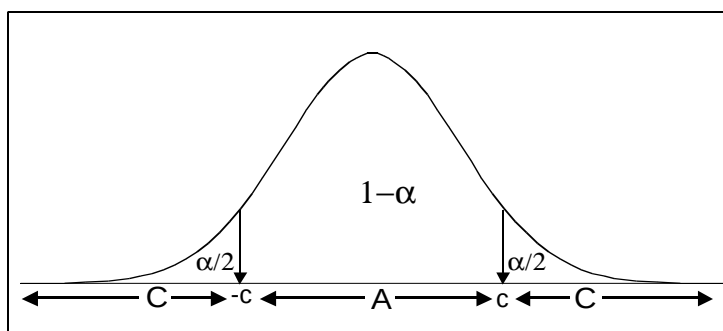
Especifíquense claramente os niveis de significación e as rexións críticas e de aceptación nos exemplos anteriores.

Segundo como sexan as rexións de aceptación e as críticas, as hipóteses poden ser *unilaterais* ou *bilaterais*. No exemplo 1 sometemos a proba $H_0(p=0,5)$ contra $H_1(p<0,5)$, trátase duna proba *unilateral (de lado esquerdo)*.



Se contrastásemos $H_0(p=0,5)$ contra $H_1(p>0,5)$, teríamos unha proba unilateral *de lado dereito*.

No caso de contrastar $H_0(p=0,5)$ contra $H_1(p \neq 0,5)$, teríamos unha proba *bilateral*.



Exercicio 3

Identifíquese a lateralidade dos exemplos anteriores.

p -valor

Cada vez é máis frecuente en moitos traballos científicos atopar referencias ao p -valor no canto de ao nivel de significación. O p valor é a probabilidade, calculada considerando H_0 certa, de que o estatístico de proba tome valores tan extremos ou máis que os calculados a partir da mostra tirada; ou sexa, o nivel de significación a partir do cal se rexeitaría a hipótese para a mostra actual. No primeiro exemplo tiñamos que o valor que o estatístico de proba tomaba para a mostra era $T=-0,01660741$. Tendo en conta que se distribúe $N(0;1)$,

$$P(T < -0,01660741) = 0,4934 \text{ será o } p\text{-valor}$$

Ora ben, xa que os estatísticos de proba se constrúen para cuantificar as desviación da hipótese nula aos valores da mostra, entón rexeitamos H_0 se o p -valor é pequeno, xa que daquela a mostra dá forte evidencia en contra de H_0 .

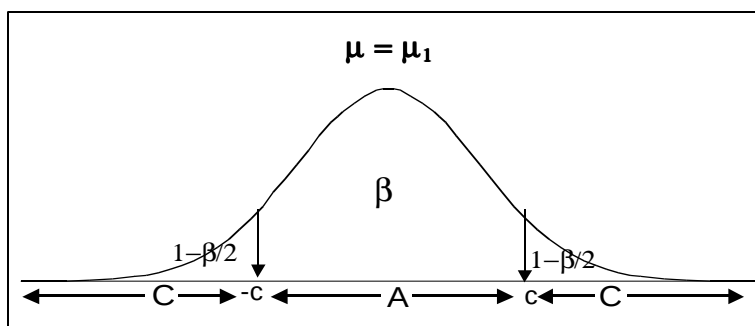
No segundo exemplo tiñamos $T = \frac{24,6 - 24}{\frac{0,3}{\sqrt{10}}} = 6,32455532$, onde T se distribuía

$N(0,1)$. Daquela o p -valor $p = P(T > 6,32455532) = 0,0000000001275$, sería o mínimo nivel de significación para o que se aceptaría a hipótese nula para a mostra tirada.

Erros de aceptación e rexeitamento

Ao aceptarmos unha hipótese estatística podemos errar en dous sentidos diferentes. Cometeremos un *erro de tipo I* ao rexeitarmos a hipótese nula sendo certa, por dar a mostra un valor do estatístico de proba que pertence á rexión crítica. Daquela, a probabilidade de cometer un erro de tipo I será α . Toda a estratexia de aceptación ou rexeitamento de hipóteses está baseada en asumir un erro de tipo I moi pequeno, de xeito que sexa moi pouco probábel rexeitar a hipótese nula no caso de ser certa.

Cometemos un *erro de tipo II* ao aceptarmos a hipótese nula sendo falsa, debido a que o estatístico de proba tomou un valor da rexión crítica. Supoñendo que contrastamos hipóteses simples, $H_0(\mu=\mu_0)$ contra $H_1(\mu=\mu_1)$, a probabilidade, de cometer un erro de tipo II, tal como se indica no seguinte gráfico, sería β



En resumo, as distintas situacións que se poden dar respecto a toma de decisións, verbo dun parámetro poboacional θ (media, desvío, proporción,...) onda as probabilidades correspondentes a cada unha delas, pódense ver no seguinte cadro.

		Estado(descoñecido) da realidade	
		$\theta = \theta_0$	$\theta = \theta_1$
Decisión a tomar	$\theta = \theta_0$	Aceptar H_0 sendo certa $p = 1-\alpha$	Aceptar H_0 sendo falsa $p = \beta$
	$\theta = \theta_1$	Rexeitar H_0 sendo certa $p = \alpha$	Rexeitar H_0 sendo falsa $p = 1-\beta$

Nas ciencias da natureza as probas de hipóteses son en certo sentido as que dan patente de científico ao método experimental. Por exemplo, se se pretende implantar un novo tratamento para determinada enfermidade, a experimentación nunha mostra haberá de dar suficiente evidencia (o correspondente estatístico de proba terá que tomar valores dentro da rexión crítica) para que se poida rexeitar a hipótese nula (seguir co mesmo tratamento), fixado o axeitado nivel de significación.

Para entender no seu xusto alcance o significado dos erros de primeiro e segundo tipo imos facer unha referencia, xa clásica, ao mundo xudicial.

Nun estado de dereito, cando se xulga a un cidadán A, a hipótese de partida (a nula) é: H_0 : “A é inocente”, e a alternativa H_1 : “A é culpábel”.

Se non hai probas suficientes non se rexeita H_0 , podendo cometerse daquela dous importantes tipos de erro:

- 1) Condenar a un inocente (rexeitar a hipótese nula sendo certa)- erro tipo I.
- 2) Absolver a un culpábel (aceptar a hipótese nula sendo falsa) – erro tipo II.

Como queira que o primeiro erro é moito máis grave que o segundo, toda a estratexia a seguir debe ter presente que a probabilidade de que se dea o erro tipo I, debe ser moi pequena, e é este criterio precisamente o que rexe ao fixar o nivel de significación.

Potencia dunha proba

E tamén interesante saber cal é a probabilidade de aceptar a *alternativa* cando é certa, e así para un mesmo nivel de significación diremos que unha proba de hipóteses é máis potente canto máis probábel sexa aceptar a alternativa en caso de ser certa. Ao concepto de *potencia dunha proba* achegarémonos mediante o seguinte exemplo

Exemplo 5

Nunha granxa quérese probar a hipótese de que o crecemento dos polos ao cabo dun mes de nacer é de 5 cm contra a alternativa de que é menor de 5 cm. Admítase un nivel de significación 0,05, e sábese por experiencias análogas que o desvío é de 1 cm.

Tomada unha mostra de 35 polos obtívose unha media de 4,8 cm. Para o nivel de significación requirido, temos $c = -1,65$.

O estatístico de proba toma o valor:

$$\frac{4,8 - 5}{1/\sqrt{35}} = -1,18 > -1,65,$$

polo tanto acéptase a hipótese nula.

Ora ben, cal foi a probabilidade de cometer erro de tipo II no caso de que fose certa a alternativa $\mu = 4,6$?

$$\text{Temos } \beta = P\left(\frac{\bar{X} - 4,6}{\frac{1}{\sqrt{35}}} \geq \frac{4,8 - 4,6}{\frac{1}{\sqrt{35}}}\right) = 1 - F(1,18) = 1 - 0,880999834 = 0,11900017$$

onde F é a función de distribución N(0,1). A probabilidade de rexeitar a hipótese nula sendo falsa sería, para esta alternativa, $\beta = 1 - 0,11900017 = 0,88099983$.

O que acontece é que se a hipótese alternativa é $H_1(\mu \neq 4,6)$, esta cúmprese sempre que a media tome valores μ_1 distintos a 4,6; é dicir, μ_1 é variábel, e xa que logo tanto β como a función de potencia dependerán dela.

Teñamos en conta que para o valor crítico $c = -1,65$, de $T = \frac{\bar{X} - 5}{1/\sqrt{35}} = -1,65$, desdexando, obteríamos un valor crítico para a media de 4,72. Daquela: $\beta(\mu_1) = 1 - F\left(\frac{4,72 - \mu_1}{1/\sqrt{35}}\right) = 1 - F(27,9 - 5,92\mu_1)$, tomará distintos valores segundo varíe μ_1 .

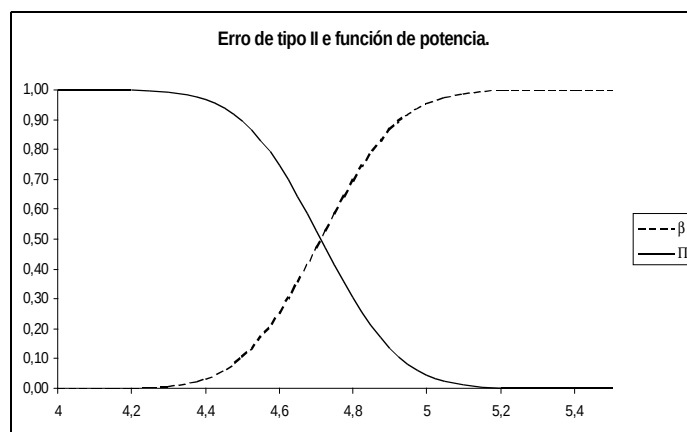
A función de potencia viría dada por $\Pi(\mu_1) = F(27,9 - 5,92\mu_1)$.

Na seguinte táboa podemos observar os valores de β e Π segundo os distintos valores de μ_1 .

μ_1	β	Π
4	1,2223E-05	0,99998778
4,1	0,00014285	0,99985715
4,2	0,00119877	0,99880123
4,3	0,00726272	0,99273728
4,4	0,03201285	0,96798715
4,5	0,10383475	0,89616525
4,6	0,25206673	0,74793327
4,7	0,46970948	0,53029052
4,8	0,69707281	0,30292719
4,9	0,86606905	0,13393095
5	0,95543457	0,04456543

5,1	0,98904721	0,01095279
5,2	0,99803664	0,00196336
5,3	0,99974548	0,00025452
5,4	0,99997628	2,3721E-05
5,5	0,99999842	1,5828E-06
5,6	0,99999992	7,5379E-08
5,7	1	2,5563E-09
5,8	1	6,1622E-11
5,9	1	1,0545E-12
6	1	1,2768E-14

Podemos representar as funcións β e Π no seguinte gráfico:



Proba de hipóteses para a varianza dunha poboación normal

En moitas ocasións cómpre someter as estimacións da varianza poboacional a unha proba de hipóteses para garantir a súa aceptación a un nivel de significación establecido.

Para iso faise uso do estatístico de proba $T = (n-1) \frac{s^2}{\sigma^2}$, que segue unha distribución χ^2_{n-1} , sendo o procedemento a utilizar semellante ao dos casos anteriores.

Consideremos o seguinte exemplo:

Exemplo 6.

Nunha planta de envasado de viño en botellas de $\frac{3}{4}$ de litro requírese que a varianza poboacional non exceda os 2,25 cl, para o cal vai someterse a proba a hipótese $H_0(\sigma^2=2,25)$ contra $H_1(\sigma^2>2,25)$. Tirada unha mostra de tamaño 25, da que se obtivo

$s^2=5,3$, escollido o nivel de significación $\alpha = 0,05$, se a hipótese é certa:

$$T = (n-1) \frac{s^2}{\sigma^2} = 24 \frac{s^2}{2,25} \text{ segue unha } \chi^2_{24}.$$

Para esta mostra, $T(5,3)=56,53$, e de $P(T>c) = 0,05$, é dicir: $P(T \leq c) = 0,95$, obtemos nas táboas un valor crítico $c=36,4$.
(utilizando a folla de cálculo teríamos $c = PRUEBA.CHI.INV(0,05;24)$, que dá o valor 36,4150265).

Como $56,53 > 36,4$, rexéitase a hipótese nula.

Exercicio 4.

A partir da seguinte mostra aleatoria, correspondente aos prezos, en euros, do litro de distintas marcas de leite, probar a hipótese $H_0(\sigma^2=0,4)$ contra $H_1(\sigma^2<0,4)$.

0,94	0,88	1,16	0,76	0,85	1,15	1,08	0,67	0,67	0,81	1,07	1,12	1,18	0,94	0,73
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Proba de hipóteses para a diferenza de medias.

1) No caso en que ambas mostras sexan independentes e proveñan de poboacións normais, en caso de igualdade de varianzas poboacionais, o estatístico:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ segue unha distribución } t_{n_1+n_2-2},$$

$$\text{sendo } S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}.$$

Exemplo 7

Mediuse o pH do cordón umbilical de 10 neonatos de mulleres normais e de 12 de mulleres preeclámicas, obténdose:

pH en neonatos, mulleres normais:

7,28 7,31 7,34 7,34 7,32 7,23 7,31 7,32 7,29 7,35

pH en neonatos, mulleres preeclámicas:

7,26 7,27 7,27 7,35 7,29 7,28 7,31 7,29 7,34 7,21 7,39 7,2

Quérese saber se existe unha diferenza significativa entre o pH dos nados de ambos tipos de mulleres.

Obtemos $\bar{X}_1 = 7,309$, $S_1^2 = 0,00125444$, $\bar{X}_2 = 7,296$, $S_2^2 = 0,00297879$.

$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2} = \frac{(10-1) \cdot 0,00125444 + (12-1) \cdot 0,00297879}{10 + 12 - 2} = 0,00220283$$

$$\text{daquela, } T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{0,00220283 \sqrt{\frac{1}{10} + \frac{1}{12}}} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{0,0009432}$$

Quérese probar a hipótese nula $\mu_1 = \mu_2$ contra a alternativa $\mu_1 \neq \mu_2$. Entón, baixo a hipótese nula $T = \frac{\bar{X}_1 - \bar{X}_2}{0,0009432}$, segue unha distribución t con 20 graos de liberdade, co cal fixado o nivel de significación 0,05, obtemos un valor crítico $c = 2,08596248$, polo que o intervalo de aceptación será: $[-2,08596248, 2,08596248]$. Como para as mostras tiradas $T = \frac{\bar{X}_1 - \bar{X}_2}{0,0009432} = \frac{7,309 - 7,296}{0,0009432} = 13,7828668$, non pertence á rexión de aceptación, rexéitase a hipótese de igualdade de medias.

2) No caso de seren as mostras suficientemente grandes, ou proveñentes de poboacións con varianzas coñecidas, o estatístico de proba sería:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{que segue unha distribución } N(0,1).$$

Se σ_1^2 e σ_2^2 son descoñecidas, estimaranse mediante as respectivas cuasi varianzas.

Proba de hipóteses para a diferenza de proporcións

Para contrastar a igualdade de proporcións en mostras suficientemente grandes, utilízase o estatístico:

$$T = \frac{p_1 - p_2 - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} = \frac{p_1 - p_2 - (P_1 - P_2)}{PQ \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

que segue unha distribución $N(0,1)$. P é a proporción poboacional que se supón común para as dúas mostras (o que se contrasta é a igualdade de proporcións). Ao ser descoñecida estímase mediante $P = \frac{k_1 + k_2}{n_1 + n_2}$, sendo k_1 e k_2 os éxitos obtidos en cadansúa das mostras, e $Q = 1 - P$.

Exemplo 8

Quérese saber se hai diferencias significativas para a proporción de votantes de determinado partido en dúas circunscripcións electorais. Para iso tírase unha mostra de

tamaño 100 da circunscrición I, obtendose $k_1 = 27$ votantes, e unha mostra de tamaño 80 na II, obténdose $k_2 = 14$ votantes.

Daquela temos $p_1 = 0,27$, $p_2 = 0,175$, $P = \frac{k_1 + k_2}{n_1 + n_2} = \frac{27 + 14}{100 + 80} = 0,22777778$

Contrastamos $P_1 = P_2$ contra $P_1 \neq P_2$. Fixado o nivel de significación 0,01, como o contraste é bilateral, obtemos $c = 2,57583451$, co cal o intervalo de aceptación será $[-2,57583451, 2,57583451]$.

Tendo en conta que:

$$\frac{p_1 - p_2}{PQ\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0,27 - 0,175}{0,22777778 \cdot 0,77222222\sqrt{\frac{1}{100} + \frac{1}{80}}} = 1,26138959$$

o valor obtido pertence á rexión de aceptación, daquela acéptase a hipótese de igualdade de proporción de votantes nas dúas circunscricións.

Probas para o axuste

Durante toda a exposición anterior viñemos supondo que as poboacións se distribuían *normalmente*. Aínda que tal hipótese é bastante xeral, en moitas ocasións cómpre deducir a partir dos datos de que se dispoña que tipo de distribución segue a poboación. Neste caso, a hipótese nula será que a poboación se distribúe de acordo cunha determinada función $F(x)$, para o cal a partir da mostra se obtén unha función de distribución empírica $F^*(x)$ que se compara con $F(x)$. Se a diferenza non é significativa, admitimos a igualdade de distribucións. Para iso temos que introducir unha nova variábel aleatoria que mida a diferenza entre F e F^* , da que coñecemos a distribución de probabilidade no caso de que se cumpra a hipótese nula. Isto permite calcular a rexión crítica e, daquela, decidir tal e como fixemos até aquí. Consideraremos dúas probas deste tipo, a χ^2 de Pearson e a de *Kolmogorov-Smirnov*.

Proba χ^2

Supoñamos unha variábel aleatoria definida nun espazo das mostras $\Omega = A_1 \cup A_2 \cup \dots \cup A_k$, onde os A_k son incompatíbeis dous a dous. Se tomamos unha mostra de tamaño n , chamaremos f_j á frecuencia observada para a j -ésima clase e denotaremos por P_j a probabilidade de que ocorra A_j , co cal o valor esperado de A_j será np_j .

Como medida das diferenzas entre as frecuencias observadas e as esperadas, utilízase o estatístico:

$$D = \sum_{j=1}^k \frac{(f_j - np_j)^2}{np_j}$$

que en caso de que o tamaño da mostra sexa grande segue unha distribución χ^2_{k-1} .

Na práctica prodúcese unha boa aproximación se para todo j , $np_j \geq 5$.

Exemplo 9

Compróbase no nivel de significación 0,05 se os seguintes pesos proveñen dunha distribución normal

Peso en kg	57-60	60-63	63-66	66-69	69-72	72-75	total
Frecuencias observadas	4	11	23	27	21	9	95

Da mostra obtemos unha media de 66,99 e un desvío de 20, que se tomarán como parámetros da distribución normal que se trata de axustar.

Peso en kg	57-60	60-63	63-66	66-69	69-72	72-75	total
f_j	0,04	0,12	0,24	0,28	0,22	0,09	
p_j normal	0,05	0,06	0,06	0,06	0,06	0,06	
$95p_j$	5,19	5,47	5,64	5,68	5,59	5,39	
$(f_j - 95p_j)$	1,42	30,58	301,51	454,63	237,37	13,05	Suma:
$(f_j - 95p_j)^2 / 95p_j$	0,27	5,59	53,5	80,07	42,44	2,42	184,30

Buscando nas táboas χ^2 con 6-1 graos de liberdade (ou na folla de cálculo), para o nivel de significación 0,05, obtemos $c = 11,07$. Como $184,3 > 11,07$, rexeitamos a hipótese de que os datos proveñan dunha poboación normal.

Exemplo 10

Nunha mostra procedente do cruzamento de chícharos verdes e chícharos amarelos obtivéronse 242 verdes. Se a mostra é de tamaño 956, dá evidencia suficiente para que se cumpra a hipótese mendeliana de que o 25% teñen que ser verdes e o 75% amarelos?

Neste caso a variábel aleatoria é dicotómica, téndose para as frecuencias esperadas $np_1 = 956 \cdot 0,75 = 717$ e $np_2 = 956 \cdot 0,25 = 239$, e para as observadas $f_1 = 710$ e $f_2 = 246$. Daquela:

$(f_j - 956p_j)^2$	49	49	
$956p_j$	717	239	Suma:
$(f_j - 956p_j)^2 / 956p_j$	0,0683	0,2050	0,2734

Tomando $\alpha = 0,05$ temos para a χ^2 con 2-1 graos de liberdade $c = 3,84$. Como $0,2734 < 3,84$, acéptase a hipótese de Mendel.

Exercicio 5

Unha revisión de sete procedementos en 50 auditorías permitiu observar determinados erros segundo como se recolle na seguinte distribución:

Nº de erros	0	1	2	3	4	5 ou máis
Nº informes	12	17	7	12	5	2

No nivel de significación 0,05, determínese se os datos proveñen dunha distribución binomial de parámetro $P = 0,2$.

Proba de Kolmogorov-Smirnov

Entre as probas que máis se utilizan para verificar se unha distribución empírica se axusta ou non a unha distribución esperada, en particular á distribución normal, destaca a de Kolmogorov-Smirnov. Esta proba é bastante potente con mostras grandes, e aplícase a variábeis continuas.

Exemplo 11

Un adestrador desexa coñecer se os tempos na carreira de 100 metros lisos están distribuídas normalmente. Para iso tira unha mostra aleatoria de 18 atletas de esa especialidade, obtendo os datos seguintes:

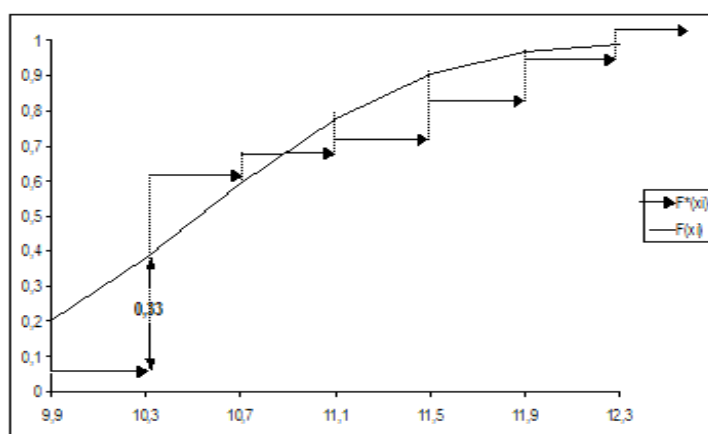
9,94	10,00	10,00	9,97	9,97	9,95	9,98	10,09	10,11	10,78	10,06	9,94	11,94	11,49	11,86	11,30	11,52	10,59
------	-------	-------	------	------	------	------	-------	-------	-------	-------	------	-------	-------	-------	-------	-------	-------

Do que se trata é de calcular a máxima discrepancia entre a función empírica e a teórica. En primeiro lugar calcúlase a media e o desvío da mostra, obténdose respectivamente: 10,52 e 0,75. A continuación dispomos os datos na seguinte táboa:

x_i	$F^*(x_i)$	$F^*(x_{i-1})$	$F(x_i)$	$d1_i$	$d2_i$	d_i
9,9	0,06	0	0,20	0,15	0,2	0,2
10,3	0,61	0,06	0,38	0,23	0,33	0,33
10,7	0,67	0,61	0,59	0,07	0,02	0,07
11,1	0,72	0,67	0,78	0,06	0,11	0,11
11,5	0,83	0,72	0,9	0,07	0,18	0,18
11,9	0,94	0,83	0,97	0,02	0,13	0,13
12,3	1	0,94	0,99	0,01	0,05	0,05
						0,33

Na 1ª columna colocamos en intervalos os valores da mostra. Neste caso escollemos a amplitude 0,4. Na segunda $F^*(9,9)=0,06$, indica que a frecuencia empírica acumulada no intervalo $[0, 9,9]$ é 0,06. Analogamente a acumulada até 11,9 é 0,94. Na terceira columna o primeiro valor corresponde á frecuencia acumulada no intervalo $[-\infty, 0]$, o segundo a acumulada até 9,9, e así sucesivamente. Na 4ª móstrase a probabilidade normal (de media 10,52 e desvío 0,75) acumulada para a abscisa da columna 1ª. Na 5ª o valor absoluto da 4ª menos a 2ª. Na 6ª o valor absoluto da 4ª menos a 3ª, e na 7ª o máximo dos correspondentes valores da 5ª e a 6ª. Por último calcúlase o máximo da columna 7ª, que é a máxima discrepancia entre as que na seguinte gráfica se sinalan mediante liñas de puntos.

Nas táboas obtemos para $n = 18$, e o nivel de significación 0,05, $c = 0,309$. Como $D = 0,33 > 0,309$ rexéitase a hipótese nula.



Exercicio 6

Na seguinte táboa clasifícanse as viaxes dun transportista segundo a súa duración en horas.

Duración en horas	[0,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8]
Nº de viaxes	2	7	12	13	8	6	4	3

Pódese afirmar que a duración das viaxes segue unha distribución normal?

A proba de Kolmogorov tamén se estende á comparación de funcións de distribucións empíricas para comprobar se dúas mostras seguen a mesma distribución. O procedemento é similar a este, buscando a diferenza máxima nas táboas existentes ao efecto. Outras probas de axuste son as de Stephens, Anderson-Darling e Cramer-von Mises.

Exercicios

- Unha fábrica de lámpadas garante que os seus produtos teñen unha vida media de 2000 horas e un desvío de 100. A un nivel de significación 0,10, próbese a hipótese $\mu = 2000$ contra a alternativa $\mu \neq 2000$, tendo en conta que nunha mostra de tamaño 40 se obtivo unha duración media de 1870 horas. Calcúlese o *p-valor*.
- Analizada unha mostra aleatoria de 10 libras de chocolate de determinada marca respecto ao seu contido porcentual de cacao, deu os seguintes resultados: 32, 25, 28, 31, 26,5, 28,3, 32,1, 30,4, 29,2. Debemos fiarnos do fabricante cando afirma que o contido é do 30%? Para que nivel de significación se pode aceptar a hipótese?
- Para someter a control do comprador unha remesa de latas de alcachofas en cuxo etiquetado figura un peso neto de 390 gr e un peso escorrido de 240, selecciónase unha mostra de 16 unidades, obéndose os seguintes pesos en escorrido: 200, 250, 245, 236, 214, 256, 230, 254, 230, 254, 230, 242, 233, 235, 236, 224, 253, 239, 241. Ten motivos fundados o comprador para rexeitar o lote a un nivel de significación 0,01?
- Unha mostra aleatoria de 100 veciños de determinada cidade indícanos que camiñan a diario 0,9 km cun desvío de 0,7. Pódese afirmar a un nivel de significación 0,05 que os veciños andan unha media de 1 km ao día?

11. Dunha mostra de 14 persoas con estudos de bacharelato obtívose que o número medio de páxinas que podían ler nunha hora era 37, cun desvío de 6. Pídesse: a) contrastar: a) $\mu = 40$ contra $\mu = 30$. b) $\mu = 40$ contra $\mu < 40$.c) a) $\mu = 40$ contra $\mu \neq 40$.

12. Unha enquisa realizada a 123 mozos de 16 a 18 anos dá como resultado que o 53% deles pasa máis de 4 horas de diante da computadora. Pódese afirmar a un nivel de significación 0,01 que a proporción de mozos que están máis de 4 horas na computadora non chega ao 40%? Calcúlese o *p-valor*.

13. Entre as persoas sometidas a un novo tratamento para controlar o colesterol, tírase unha mostra de tamaño 6 e obtéñense os seguintes datos: 279, 301, 299, 324, 280, 289. Tómase outra mostra, tamén de tamaño 6, entre pacientes non sometidos ao novo tratamento, obténdose os resultados: 270, 295, 299, 310, 296, 301. Nun nivel de confianza do 95%, suposta a hipótese de normalidade, podemos dicir que o tratamento deu resultados significativos? Calcúlese o *p-valor*.

14. Un laboratorio quere estimar o efecto dunha nova medicina para a migraña. Por experiencias anteriores sábese que polo menos fará efecto no 43% dos casos. Supoñendo que nunha mostra de tamaño 40, para un nivel de significación 0,05, a proporción de efecto positivo foi de 0,47, sométase a proba a hipótese $P = 0,5$ contra $P < 0,5$.

15. Unha mostra de tamaño dez para estimar o gasto en libros de texto, en euros, deu os seguintes resultados: 132,125,130,112,129,128,120,133,125,112. Que declaracións son máis fiábeis, as do responsábel de educación ao afirmar que o gasto medio é de 120€ ou ás da asociación de consumidores que afirma que é 130? (Supóñase normalidade e tómese o nivel de significación que se considere axeitado). Calcúlese o *p-valor*.

16. Proveñen os datos do exercicio anterior dunha distribución normal?

17. Próbese a hipótese de que ao botar unha moeda sobre unha superficie lisa ten a mesma probabilidade de saír cara que cruz, se nunha mostra aleatoria de tamaño 100 saíron 54 caras e 46 cruces.

18. Arthur Geissler publicou en 1889 a distribución segundo o sexo para familias de 8 fillos, que foi usada por R. A. Fisher en “Statistical Methods for Research Workers”, en 1925, como un exemplo de distribución binomial. Geissler compilou datos de Saxon, o estado libre alemán máis antigo, durante o período 1876-1885. Para as familias con 8 fillos dá a seguinte táboa:

Nº de nenos	0	1	2	3	4	5	6	7	8
Frec.	21485	5331	10649	14959	11969	6678	2092	342	

Pode considerarse que a anterior variábel “número de nenos en familias de 8 fillos” segue unha distribución binomial?

22. Sométense dous grupos de alumnos a dous métodos diferentes de aprendizaxe. Rematado o curso, realizadas as probas pertinentes, obtéñense os seguintes resultados:

Grupo A: $n = 25$, media = 5,7, desvío = 1,4

Grupo B: $n = 18$; media = 6,2, desvío = 0,5.

Existen diferencias significativas entre os dous métodos?

23. Nun complexo hospitalario empréganse dous procedementos distintos para certo tipo de intervención cirúrxica. Tiradas as correspondentes mostras de cada grupo, obtivéronse os seguintes tempos de recuperación, en días, para cada unha delas.

G1	7	8	9	7	7	9	9	7	7	9	7	9	9	9
G2	7	6	7	7	10	9	7	6	9	7	7	8	8	7

Supondo que o tempo de recuperación se distribúe normalmente, á vista dos datos, podemos dicir que existen diferencias significativas entre os dous métodos

8. SERIES DE TEMPO

“E agora –continuou Ulrich– fanse observacións con todas as aparencias de lei natural, pero sen o fundamento que a xustifique. A regularidade das series estatísticas é ás veces tan constante como a das leis.”

Robert Musil, *O home sen calidades*.

Para estudar certas variábeis económicas, meteorolóxicas ou demográficas, e outras moitas, séguese a súa evolución no tempo. Se representamos os datos observados en cada unidade temporal (neste caso meses e trimestres) nun sistema de coordenadas, uníndoos mediante un trazo continuo, obtemos gráficos do seguinte tipo:

Gráfico 1.

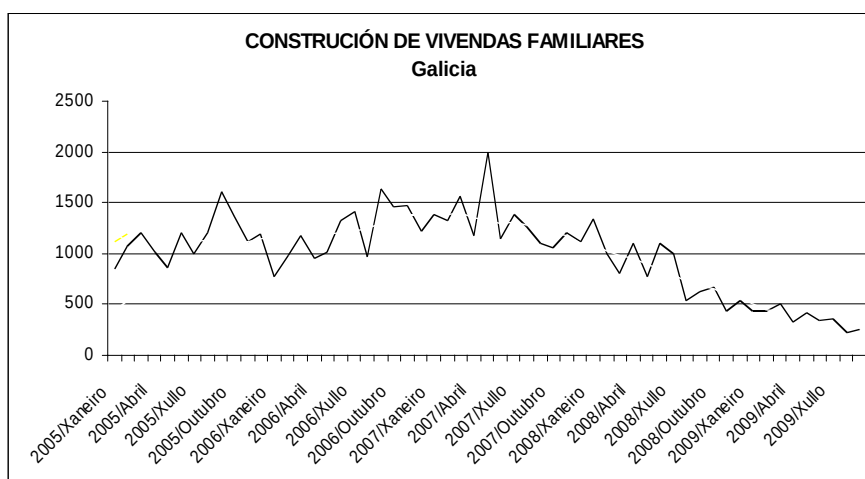


Gráfico 2

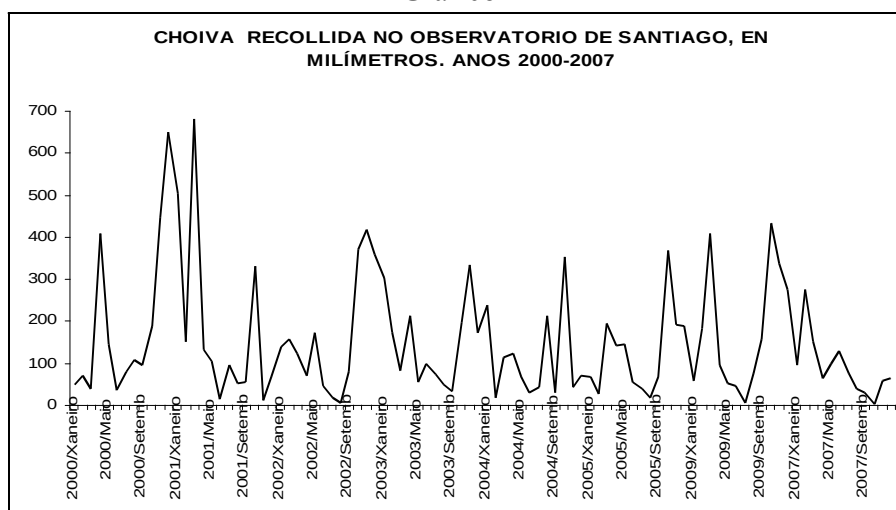
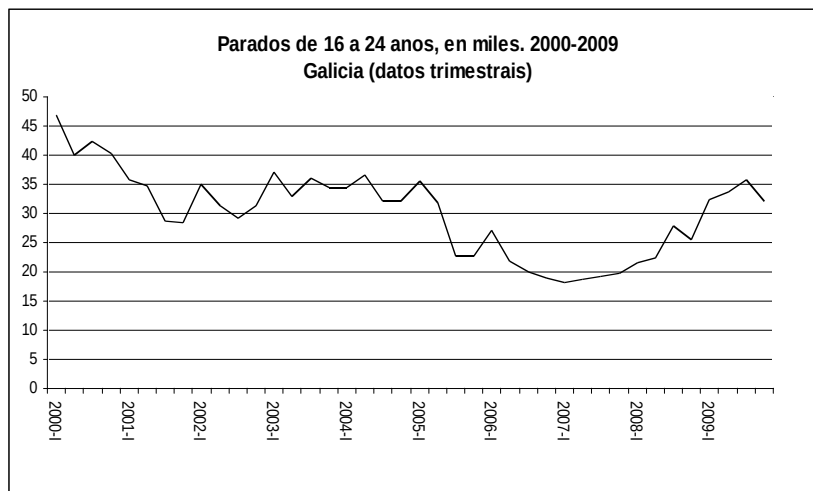


Gráfico 3.



Aos valores dunha magnitude ordenados respecto ao tempo chámase *serie de tempo* ou *serie cronolóxica*.

Existen diversos obxectivos que xustifican o estudo das series de tempo. Un deles é o eterno soño laplaciano de predicir o futuro baseándose no coñecemento do presente e do pasado –as sempre malfadadas previsións económicas!-, outro pode ser o control dun proceso dentro de certos valores admisíbeis das magnitudes observadas –poñamos por caso, as medidas sobre a natalidade. Noutras ocasións soamente se trata da utilización dun modelo axeitado para describir as características máis salientábeis de determinado fenómeno.

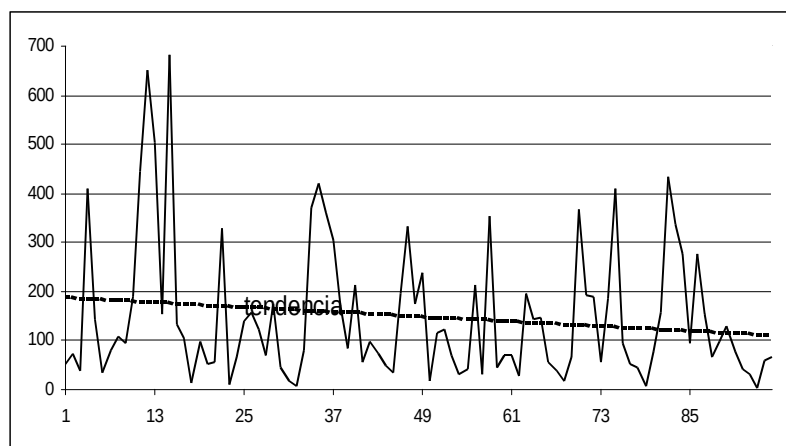
Achegarémonos ao estudo das series temporais dende o punto de vista determinístico, adoptando o modelo de análise clásica que parte do suposto de que a variábel estudada é combinación de catro compoñentes: *tendencia*, *variacións cíclicas*, *variacións estacionais* e *variacións residuais*.

Á marxe do anterior enfoque fica a perspectiva aleatoria, que interpreta as series temporais como observacións de **procesos estocásticos**, é dicir, de coleccións de variábeis aleatorias. Exemplos deste tipo poden ser os sinais de audio, o movemento browniano ou as flutuacións do prezo das accións no mercado de valores.

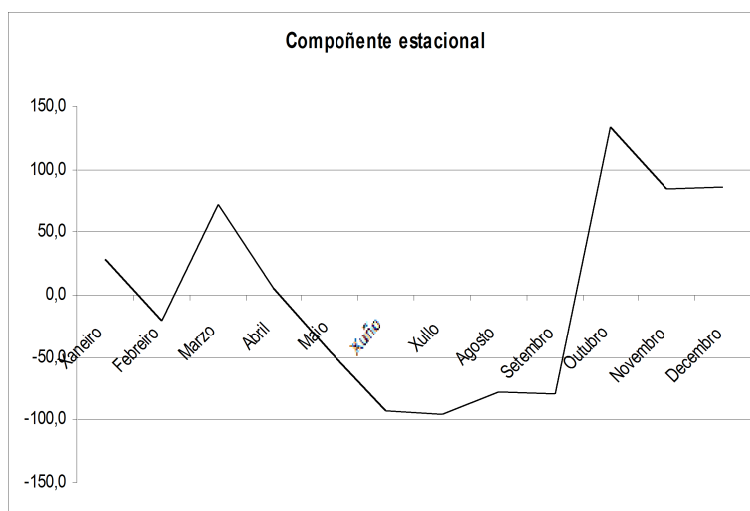
Análise clásica de series temporais

Para proceder ao seu estudo consideraremos a serie de tempo como resultante das seguintes compoñentes:

- A *tendencia* é o movemento suave e regular que amosa a serie ao estudala durante un período de tempo suficientemente longo. A tendencia mide a variación media da variábel por unidade de tempo.



- As *variacións estacionais* son movementos periódicos que se presentan con certa regularidade durante un período de tempo de duración igual ou inferior a un ano. Exemplos típicos deste comportamento son as vendas de determinados artigos, os gastos vacacionais ou a cantidade de choiva recollida.



- As *variacións cíclicas* son os movementos recorrentes, ascendentes e descendentes, que se dan en períodos superiores a un ano. Téñense proposto moitas teorías para explicar estas flutuacións, mais raramente gozan de aceptación xeralizada. No referente ás variábeis de natureza económica, é común aceptar que estas variacións se deben principalmente á alternancia de etapas de expansión e depresión (ciclos económicos), mais, como puidemos comprobar na crise actual [escribimos isto en xuño de 2010] non se pode predicir con exactitude o cambio de ciclo nin a súa duración e, xa que logo, o interese do estudo das variacións cíclicas estriba mais ben en modelizar o pasado con fins descritivos.

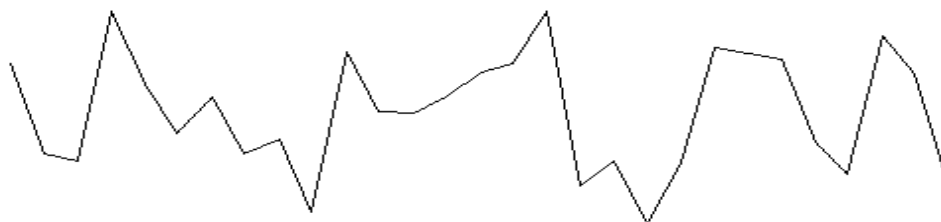
- Admíttese tamén a existencia dunhas *variacións residuais* que non amosan un carácter periódico recoñecíbel, considerándose que ou ben son aleatorias, ou ben se deben a certas perturbacións imprevisíbeis. Suponse tamén que estes movementos residuais son independentes e teñen media cero. (Un fenómeno deste tipo podería ser o efecto do volcán *Eyjafjalla* sobre o negocio das compañías aéreas, pola súa imprevisibilidade, mais non semella moi doado que se poida dar un fenómeno aleatorio que o contrarreste positivamente para conseguir a media cero).

Modelos para a análise

A maneira de integrar as catro compoñentes anteriores dá lugar a diferentes modelos de análise. O *esquema aditivo* parte da hipótese de que o valor da serie de tempo ven dado pola suma de catro compoñentes: $Y = T + E + C + R$, sendo Y o valor observado, T a tendencia, E as variacións estacionais, C as variacións cíclicas e R as residuais. O *esquema multiplicativo* supón que o valor da serie ven dado polo produto das catro compoñentes; é dicir: $Y = T \cdot E \cdot C \cdot R$.

A elección de esquema baséase en consideracións detectábeis na evolución gráfica da serie. O modelo aditivo debe empregarse se a composición da tendencia do movemento estacional conduce a unha variación de amplitude constante, mentres que o multiplicativo empregarase se as fluctuacións estacionais respecto á tendencia medran no tempo.

Modelo aditivo



Modelo multiplicativo



Nesta introdución limitaremonos ao esquema aditivo, mais se queremos aplicar o modelo multiplicativo, este pode ser reducido ao aditivo utilizando unha transformación logarítmica, é dicir, facendo $\ln Y = \ln T + \ln E + \ln C + \ln R$.

Identificación da tendencia

Axuste linear mediante mínimos cadrados.

Os seguintes datos responden ás precipitacións en mm (litros por metro cadrado) recollidas no observatorio meteorolóxico de Santiago (Gráfico 2), onde engadimos unha columna para calcular as medias de cada mes, da que máis adiante faremos uso.

	2000	2001	2002	2003	2004	2005	2006	2007	Media
Xaneiro	50,9	505,3	140,2	302,9	239,5	69,6	57,4	95,3	182,6
Febreiro	72,6	153,3	158,9	172,7	18,6	28,6	184	275,4	133,0
Marzo	39,1	682,9	122,9	83,5	114,6	194,7	408	152,7	224,8
Abril	409,4	133,8	70,2	214,3	123,9	143,5	96,2	66,3	157,2
Maio	144,4	104,5	174,2	56,8	68,4	146,7	52,2	99,6	105,9
Xuño	36,3	15,7	46	98,9	30,1	55,1	46,8	128,8	57,2
Xullo	81,1	96,4	17,9	73,7	42,2	39,4	6,4	76,7	54,2
Agosto	108,1	52,7	5,9	49,4	214,9	18,8	78,5	40,5	71,1
Setembro	94,8	54,5	80	35,4	32,3	68,2	157,6	31,7	69,3
Outubro	188	330,7	370,7	197,4	352,5	369,2	433	2,3	280,5
Novembro	444,8	11,4	419,2	333,4	43,9	193	337,2	58,3	230,2
Decembro	649,8	67,3	358,8	173,9	71,2	190,3	276,7	65,1	231,6

Trátase de axustar unha función que recolla a evolución da magnitude no tempo. Para iso, entre todas as rectas posíbeis, escolleremos a de regresión mínimo cuadrática, é dicir, aquela que fai mínima a suma das diferencias ao cadrado entre os valores observados e os obtidos na recta para a correspondente abscisa.

En primeiro lugar dispomos os datos da seguinte maneira:

	t	Y
2000/Xaneiro	1	50,9
2000/Febreiro	2	72,6
2000/Marzo	3	39,1
....	...	...
2007/Outubro	94	2,3
2007/Novembro	95	58,3
2007/Decembro	96	65,1

Na segunda columna numéranse consecutivamente os meses observados.

Se a ecuación da recta é $y = a + bt$, a e b poden calcularse mediante o sistema:

$$\begin{cases} \sum Y = aN + b \sum t \\ \sum tY = a \sum t + b \sum t^2 \end{cases}$$

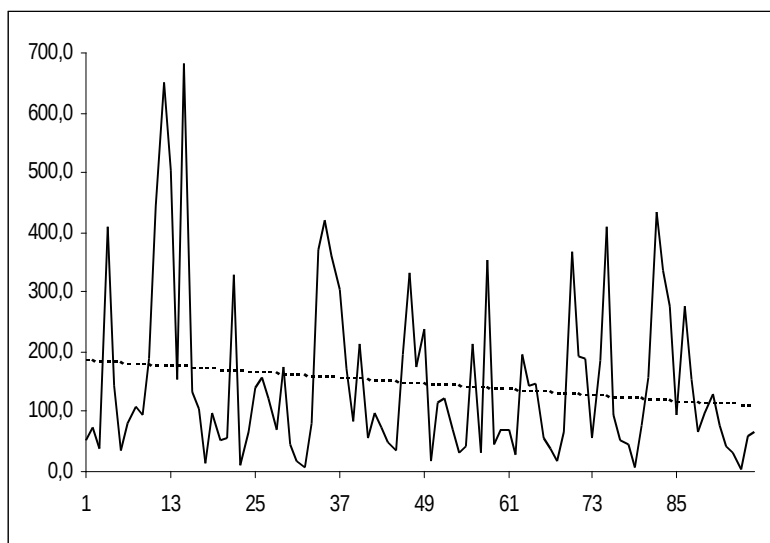
Máis adiante calcularemos a e b resolvendo o sistema, pero de momento botamos man da folla de cálculo, mediante o esquema:

	A	B	C
1	1	50,9	=TENDENCIA(\$B\$1:\$B\$96,\$A\$1:\$A\$96;A1;VERDADERO)
2	2	72,6	=TENDENCIA(\$B\$1:\$B\$96,\$A\$1:\$A\$96;A2;VERDADERO)
	...		
95	95	58,3	=TENDENCIA(\$B\$1:\$B\$96,\$A\$1:\$A\$96;A95;VERDADERO)
96	96	65,1	=TENDENCIA(\$B\$1:\$B\$96,\$A\$1:\$A\$96;A96;VERDADERO)

Na columna C obtemos os valores estimados de Y segundo a recta, que volvemos ordenar mensualmente. Tamén engadimos unha columna coa media mensual para os datos estimados.

Xaneiro	2000	2001	2002	2003	2004	2005	2006	2007	medias
Febreiro	188,1	178,4	168,7	159,1	149,4	139,7	130,1	120,4	154,2
Marzo	187,3	177,6	167,9	158,3	148,6	138,9	129,2	119,6	153,4
Abril	186,5	176,8	167,1	157,5	147,8	138,1	128,4	118,8	152,6
Maio	185,7	176,0	166,3	156,7	147,0	137,3	127,6	118,0	151,8
Xuño	184,9	175,2	165,5	155,8	146,2	136,5	126,8	117,2	151,0
Xullo	184,1	174,4	164,7	155,0	145,4	135,7	126,0	116,3	150,2
Agosto	183,3	173,6	163,9	154,2	144,6	134,9	125,2	115,5	149,4
Setembro	182,4	172,8	163,1	153,4	143,8	134,1	124,4	114,7	148,6
Outubro	181,6	172,0	162,3	152,6	142,9	133,3	123,6	113,9	147,8
Novembro	180,8	171,2	161,5	151,8	142,1	132,5	122,8	113,1	147,0
Decembro	180,0	170,4	160,7	151,0	141,3	131,7	122,0	112,3	146,2
Xaneiro	179,2	169,6	159,9	150,2	140,5	130,9	121,2	111,5	145,4

A continuación representamos os datos observados e a tendencia (líña de puntos)



Axuste parabólico

Se en vez dunha recta queremos axustar a tendencia a unha parábola $Y = a + bt + ct^2$, segundo o criterio dos mínimos cadrados, os coeficientes obtéñense da resolución do sistema:

$$\begin{cases} \sum Y = aN + b\sum t + c\sum t^2 \\ \sum tY = a\sum t + b\sum t^2 + c\sum t^3 \\ \sum t^2Y = a\sum t^2 + b\sum t^3 + c\sum t^4 \end{cases}$$

Exemplo

A seguinte táboa mostra a evolución dos postos de traballo na industria en Galicia de 2000 a 2008.

2000	2001	2002	2003	2004	2005	2006	2007	2008
155361	161344	163440	160709	161539	163420	166902	172044	169982

Para simplificar os cálculos, facemos un cambio de escala para os anos, obtendo:

t	y	t ²	t ³	t ⁴	ty	t ² y
1	155361	1	1	1	155361	155361
2	161344	4	8	16	322688	645376
3	163440	9	27	81	490320	1470960
4	160709	16	64	256	642836	2571344
5	161539	25	125	625	807695	4038475
6	163420	36	216	1296	980520	5883120
7	166902	49	343	2401	1168314	8178198
8	172044	64	512	4096	1376352	11010816
9	169982	81	729	6561	1529838	13768542
Σ	1474741	285	2025	15333	7473924	47722192

De onde:

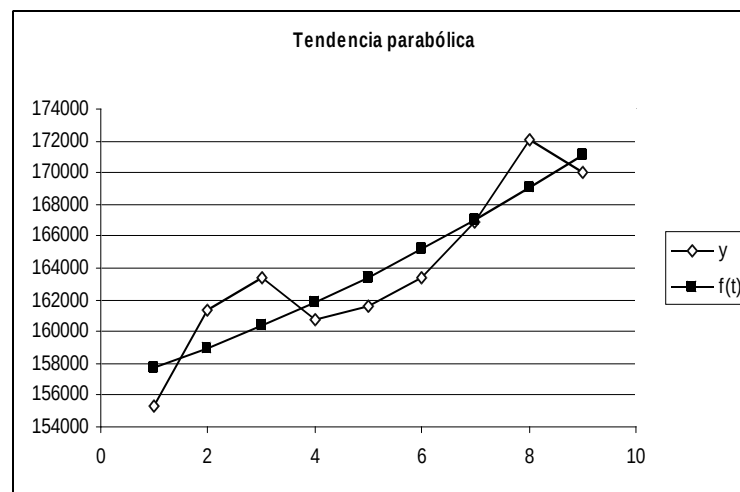
$$\begin{cases} 9a + 45b + 285c = 1474741 \\ 45a + 285b + 2025c = 7473924 \\ 285a + 2025b + 15333c = 47722192 \end{cases}$$

e resolvendo:

$$a = 156691; \quad b = 1025,2; \quad c = 64,5140693.$$

Como no caso anterior podemos representar graficamente os datos observados e a tendencia

t	y	$a + bt + ct^2$
1	155361	157780
2	161344	158999
3	163440	160347
4	160709	161824
5	161539	163430
6	163420	165164
7	166902	167028
8	172044	169021
9	169982	171143



Método das medias móbeis

A tendencia da serie pode tamén acharse mediante o cálculo reiterado dos valores medios. Do que se trata con este método e de restarlle peso a un valor determinado promediándoo cos que están ao seu carón. Daquela, para cada instante t teremos un valor Y_t que virá dado pola correspondente media.

A organización de datos para ao “suavizado” dunha serie con sete observacións utilizando medias móbeis de tamaño 3, sería a seguinte:

Tempo	observación	media
t_1	y_1	---
t_2	y_2	$(y_1 + y_2 + y_3)/3$
t_3	y_3	$(y_2 + y_3 + y_4)/3$
t_4	y_4	$(y_3 + y_4 + y_5)/3$
t_5	y_5	$(y_4 + y_5 + y_6)/3$
t_6	y_6	$(y_5 + y_6 + y_7)/3$
t_7	y_7	---

Fixémonos que nin ao primeiro valor nin ao derradeiro se lle adxudica ningunha media. O número de observacións sen media, ao principio e ao final da táboa, dependerá do tamaño elixido.

No caso de ser o tamaño das medias un número par, estas quedarán *desprazadas* con respecto ao tempo, polo que se procede a calcular as medias de tamaño unha unidade inferior, e unha unidade superior, para calcular posteriormente a media de ambas.

No seguinte exemplo, onde a información é mensual, calcularemos primeiro as medias de tamaño 11, e despois as de tamaño 13; o valor que lle adxudicamos á tendencia é a media das dúas.

Os datos da seguinte táboa corresponden ás temperaturas medias observadas, mensualmente, na estación meteorolóxica de Mabegondo

	2001	2002	2003	2004	2005	2006	2007	2008	2009
Xaneiro	10,5	10,8	7,9	10,2	9,2	7,1	9,4	9,9	9,4
Febreiro	9,3	10,2	8,6	8,5	6,1	6,3	10,4	10,9	8,8
Marzo	12,7	11,0	11,8	9,9	10,7	11,6	10,4	10,0	11,0
Abril	11,2	10,9	12,9	10,8	12,1	12,4	12,5	11,8	11,8
Maiο	13,6	13,2	13,8	14,2	14,2	14,7	14,1	13,9	14,0
Xuño	17,0	15,8	17,7	18,7	18,7	18,0	16,1	16,9	17,4
Xullo	18,0	17,4	18,1	18,5	19,4	20,4	17,6	18,1	18,4
Agosto	19,5	18,0	20,7	18,7	19,1	19,2	18,1	18,2	18,9
Setembro	16,3	17,4	17,7	17,4	16,7	18,5	16,3	16,0	17,0
Outubro	15,7	15,6	13,0	14,2	15,4	16,5	13,8	13,1	14,7
Novembro	9,3	11,9	12,0	9,9	10,6	13,4	9,0	9,3	10,7
Decembro	7,4	11,6	9,5	8,4	7,8	8,9	8,0	8,2	8,7

Para calcular a media móbil de período 12, na columna A da folla de cálculo numeramos correlativamente os meses dende 1 (xaneiro de 2001) até 98 (decembro de 2009). En B escribimos as observacións, en C calculamos a media móbil de período 11, en D a de período 13, en E facemos a media das dúas. Os seis primeiros meses e os seis últimos non teñen media móbil.

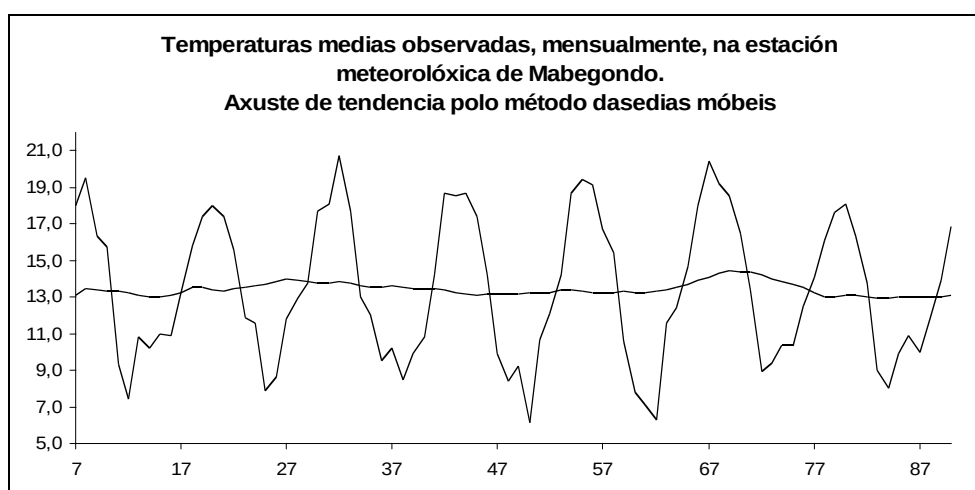
O esquema de folla de cálculo:

A	B	C	D	E
1	10,5			
2	9,3			
3	12,7			
4	11,2			
5	13,6			
6	17,0	=PROMEDIO(B1:B11)		
7	17,0	=PROMEDIO(B2:B12)	=PROMEDIO(B1:B13)	=PROMEDIO(C7:D7)
8	18,0	=PROMEDIO(B3:C13)	=PROMEDIO(B2:B14)	=PROMEDIO(D8:D8)
...				
88	14,0	=PROMEDIO(B83:B93)	=PROMEDIO(B82:B94)	=PROMEDIO(C88:D88)
89	17,4	=PROMEDIO(B84:B94)	=PROMEDIO(B83:B95)	=PROMEDIO(C89:D89)
90	18,4	=PROMEDIO(B85:C95)		

91	18,9			
92	17,0			
93	14,7			
94	10,7			
95	8,7			

proporciona os datos

A	B	C	D	E
1	10,5			
2	9,3			
3	12,7			
4	11,2			
5	13,6			
6	17,0	13,9		
7	18,0	13,0	13,2	13,1
8	19,5	13,8	13,2	13,5
9	16,3	13,5	13,3	13,4
...				
88	11,8	13,0	13,0	13,0
89	13,9	13,3	12,7	13,0
90	16,9	13,5	12,6	13,1
91	18,1	13,3		
92	18,2			
93	16,0			
94	13,1			
95	9,3			



Suavizado exponencial

Na actualidade este é un dos métodos de prognóstico formal que se utiliza con máis frecuencia, debido á súa sinxeleza e á reducida cantidade de datos que require. No suavizado exponencial as ponderacións asignadas ás observacións diminúen ao facerse

máis antigos os datos, isto é, dáselle moita máis importancia aos datos recentes que aos pasados. Para establecer o modelo necesítase un valor inicial (F_1) e unha constante de suavización (α , un valor arbitrario entre 0 e 1).

Para calcular a estimación no instante $t+1$, coñecida a do instante t , utilízase a seguinte fórmula, onde Y_t é a observación no instante t .

$$F_{t+1} = \alpha Y_t + (1 - \alpha) F_t$$

Para iniciar o proceso pode facerse $F_1 = Y_1$ (primeiro valor observado) ou $F_1 = \frac{\sum Y_i}{n}$.

Temos:

$$\begin{aligned} F_{t+1} &= \alpha Y_t + (1 - \alpha)(\alpha Y_{t-1} + (1 - \alpha) F_{t-1}) = \\ &= \alpha Y_t + (1 - \alpha) \alpha Y_{t-1} + (1 - \alpha)^2 F_{t-1} = \\ &= \alpha Y_t + (1 - \alpha)(\alpha Y_{t-1} + (1 - \alpha) F_{t-1}) = \\ &= \alpha Y_t + (1 - \alpha) \alpha Y_{t-1} + (1 - \alpha)^2 (\alpha Y_{t-2} + (1 - \alpha) F_{t-2}) = \\ &= \alpha Y_t + (1 - \alpha) \alpha Y_{t-1} + (1 - \alpha)^2 \alpha Y_{t-2} + (1 - \alpha)^3 F_{t-2} \end{aligned}$$

Como podemos ver na expresión recursiva que se obtén, os coeficientes das observacións tanto máis se aproximan a cero canto máis tempo transcurra.

Imos a ensaiar o método coa seguinte serie formada polo número de parados de 16 a 24 anos en Galicia, expresados en miles.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
TrimestreI	46,9	35,8	35,1	37,2	34,5	35,4	27,2	18,1	21,6	32,4
TrimestreII	40,1	34,7	31,3	32,9	36,6	31,8	21,8	18,7	22,4	33,6
TrimestreIII	42,4	28,7	29,1	36	32	22,7	20	19,3	28	35,7
TrimestreIV	40,3	28,3	31,4	34,4	32,2	22,7	19	19,7	25,6	32

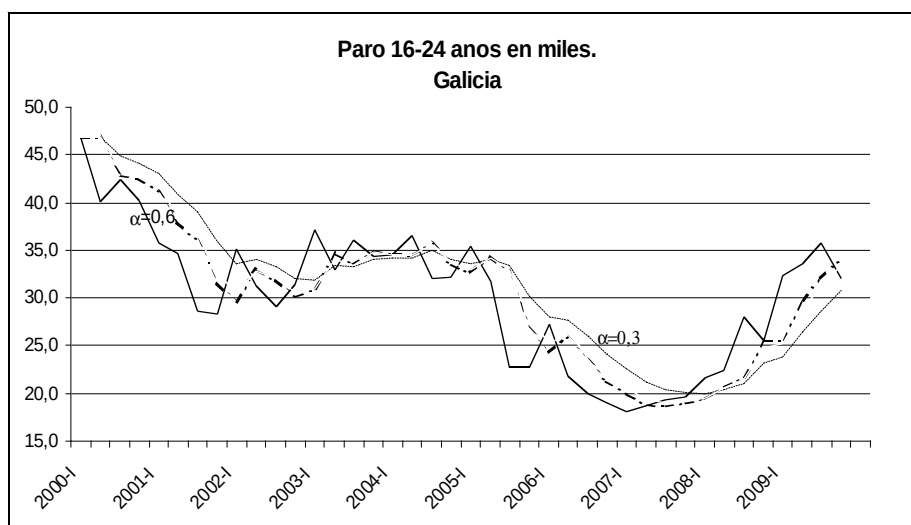
Na columna A colocamos consecutivamente as observacións trimestrais. Nas B, C e D ensaiamos con distintos valores do parámetro de suavización, pegando a fórmula correspondente que se extenderá até a fila 42 (son 40 trimestres). (En vez de escribir directamente a fórmula podemos seleccionar opción “suavizado exponencial” no menú da folla de cálculo, en “Herramientas para el análisis” - menú Herramientas->Análisis de datos->Suavizado exponencial.)

A	B	C	D
Y_t	$\alpha=0,6$	$\alpha=0,3$	$\alpha=0,7$
46,9	46,9	46,9	46,9
40,1	$=0,6*A2+0,4*B2$	$=0,3*A2+0,7*C2$	$=0,5*A2+0,5*D2$

Obtemos:

	A	B	C	D			A	B	C	D
1	Y_t	$\alpha=0,6$	$\alpha=0,3$	$\alpha=0,7$		22	35,4	32,7	33,5	33,0
2	46,9	46,9	46,9	46,9		23	31,8	34,3	34,1	34,2
3	40,1	46,9	46,9	46,9		24	22,7	32,8	33,4	33,0

4	42,4	42,8	44,9	43,5	25	22,7	26,7	30,2	27,8
5	40,3	42,6	44,1	43,0	26	27,2	24,3	27,9	25,3
6	35,8	41,2	43,0	41,6	27	21,8	26,0	27,7	26,2
7	34,7	38,0	40,8	38,7	28	20,0	23,5	25,9	24,0
8	28,7	36,0	39,0	36,7	29	19,0	21,4	24,2	22,0
9	28,3	31,6	35,9	32,7	30	18,1	20,0	22,6	20,5
10	35,1	29,6	33,6	30,5	31	18,7	18,8	21,3	19,3
11	31,3	32,9	34,1	32,8	32	19,3	18,8	20,5	19,0
12	29,1	31,9	33,2	32,1	33	19,7	19,1	20,1	19,2
13	31,4	30,2	32,0	30,6	34	21,6	19,5	20,0	19,4
14	37,2	30,9	31,8	31,0	35	22,4	20,7	20,5	20,5
15	32,9	34,7	33,4	34,1	36	28,0	21,7	21,1	21,5
16	36,0	33,6	33,3	33,5	37	25,6	25,5	23,1	24,7
17	34,4	35,0	34,1	34,7	38	32,4	25,6	23,9	25,2
18	34,5	34,7	34,2	34,6	39	33,6	29,7	26,4	28,8
19	36,6	34,6	34,3	34,5	40	35,7	32,0	28,6	31,2
20	32,0	35,8	35,0	35,6	41	32,0	34,2	30,7	33,4
21	32,2	33,5	34,1	33,8					



Na práctica próbase con distintos valores de α até acadar na gráfica unha boa aproximación. Neste caso observamos como con $\alpha = 0,6$ logamos un mellor axuste que con $\alpha = 0,3$.

Variacións estacionais. Compoñente estacionaria

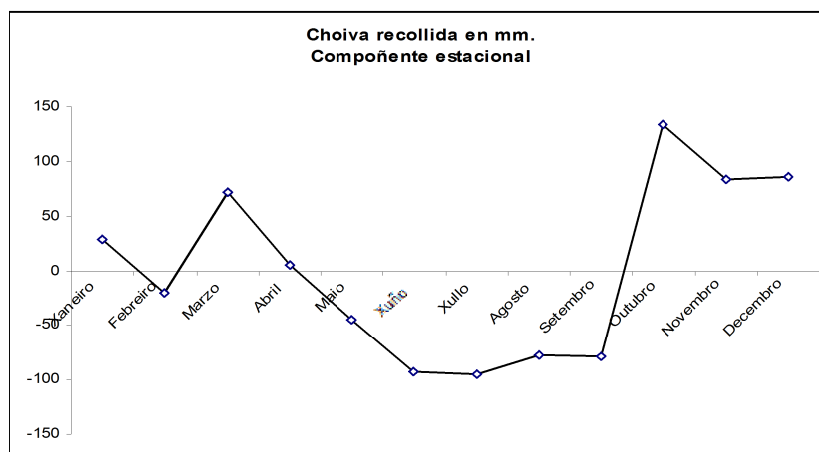
Definimos anteriormente as variacións estacionais como flutuacións de período inferior ou igual a un ano, debidas a factores climatolóxicos, hábitos de consumo, ou calquera outro fenómeno que xere unha periodicidade regular. En moitas series, as meteorolóxicas ou a maioría das series económicas, a compoñente estacional provoca

distorsións, polo que para seguir a evolución dos datos cómpre eliminala. Este proceso coñécese como desestacionalización da serie. Achegáremos a el mediante uns exemplos.

No caso da choiva recollida en Santiago, temos:

	Media observacións	Media estimacións	Compoñente estacional (Diferenza)
Xaneiro	182,6	154,2	28,4
Febreiro	133,0	153,4	-20,4
Marzo	224,8	152,6	72,2
Abril	157,2	151,8	5,4
Maio	105,9	151,0	-45,2
Xuño	57,2	150,2	-93
Xullo	54,2	149,4	-95,2
Agosto	71,1	148,6	-77,5
Setembro	69,3	147,8	-78,5
Outubro	280,5	147,0	133,5
Novembro	230,2	146,2	84
Decembro	231,6	145,4	86,3
		Suma	0

A diferenza entre as medias mensuais observadas e as estimadas é a compoñente estacional. Observemos que a súa suma é cero.



Para o caso dos parados, ordenando os datos por trimestre:

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Media estacional
TrimestreI	46,9	35,8	35,1	37,2	34,5	35,4	27,2	18,1	21,6	32,4	32,42
TrimestreII	40,1	34,7	31,3	32,9	36,6	31,8	21,8	18,7	22,4	33,6	30,39
TrimestreIII	42,4	28,7	29,1	36	32	22,7	20	19,3	28	35,7	29,39
TrimestreIV	40,3	28,3	31,4	34,4	32,2	22,7	19	19,7	25,6	32	28,56

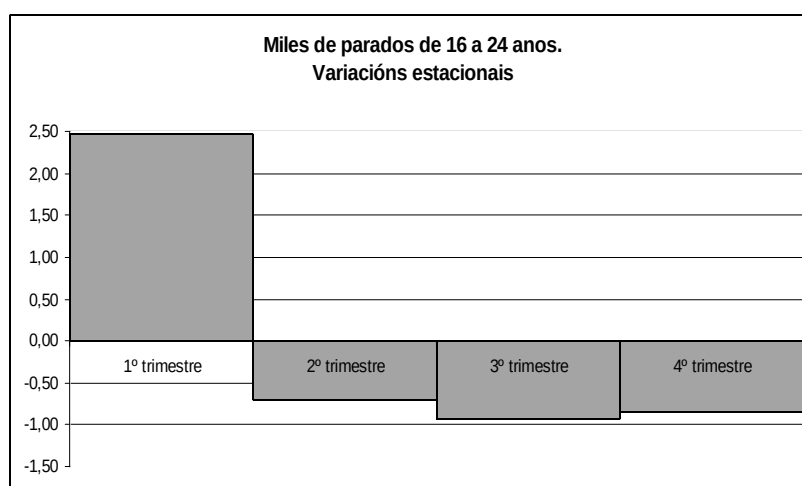
Tomando $\alpha = 0,6$ para o suavizado exponencial

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	media corrixida
1º trimestre	46,9	41,2	29,6	30,9	34,7	32,7	24,3	20,0	19,5	25,6	30,5
2º trimestre	46,9	38,0	32,9	34,7	34,6	34,3	26,0	18,8	20,7	29,7	31,7
3º trimestre	42,8	36,0	31,9	33,6	35,8	32,8	23,5	18,8	21,7	32,0	30,9
4º trimestre	42,6	31,6	30,2	35,0	33,5	26,7	21,4	19,1	25,5	34,2	30,0

	Media estacional	Media corrixida	Variacións estacionais
1º trimestre	32,4	30,5	1,9
2º trimestre	30,4	31,7	-1,3
3º trimestre	29,4	30,9	-1,5
4º trimestre	28,6	30,0	-1,4
		Suma	-2,3
		media	-0,58

Como a suma das variacións estacionais é $-2,3$ (non é cero!) calculamos a media do erro ($-0,58$) e restámoslla ás variacións (centrámoslas), obtendo así as variacións corrixidas que son as que temos que utilizar para a desestacionalización:

	Variacións estacionais corrixidas
1º trimestre	2,47
2º trimestre	-0,69
3º trimestre	-0,93
4º trimestre	-0,85
Suma	0



No caso en que a tendencia fose calculada polo método das medias móbeis procederíase do mesmo xeito, centrando as variacións estacionais para que teñan media cero.

Predición

Ao utilizarmos unha serie para facer predicións asúmese que a tendencia que seguirá no futuro (canto máis inmediato mellor) é a determinada polas observacións pasadas. Así mesmo, no caso en que teñamos illada a compoñente estacional, esta deberá terse en conta para desestacionalizar o valor estimado.

Utilizando a recta mínimo cuadrática

Volvendo ao exemplo da choiva recollida na estación de Mabegondo, queremos predicir a cantidade recollida en xaneiro de 2009.

Dispomos os datos para calcular os coeficientes do sistema:

$$\begin{cases} \sum Y = aN + b\sum t \\ \sum tY = a\sum t + b\sum t^2 \end{cases}$$

t	Y	tY	t ²
1	50,9	50,9	1
2	72,6	145,2	4
3	39,1	117,3	9
...	...	...	...
94	2,3	216,2	8836
95	58,3	5538,5	9025
96	65,1	6249,6	9216
4656	14380,9	638045	299536

Substituíndo:

$$\begin{cases} 14380,9 = 96a + 4656b \\ 638045 = 4656a + 299536b \end{cases}$$

e resolvendo: $a = 194,123904$ e $b = -0,91406674$

Para xaneiro de 2009, $t = 109$, $f(109) = 194,123904 - 0,91406674 \cdot 109 = 122,9$ mm. E desestacionalizando $122,9 + 28,4 = 151,3$ mm.

Poderíamos chegar ao mesmo resultado inserindo na folla de cálculo:
`=TENDENCIA($B$1:$B$96;$A$1:$A$96;A1;VERDADERO)`,
 e sumándolle posteriormente a compoñente estacional.

Utilizando o suavizado exponencial

Partindo da serie de parados de 16 a 24 anos utilizada anteriormente, se queremos facer unha proxección para o primeiro trimestre de 2010, non hai máis que engadir unha fila na táboa utilizada anteriormente

41	32,0	=0,6*C40+0,4*D40
42		=0,6*C41+0,4*D41

obténdose:

42	32,9
----	------

e sumándolle o índice estacional: $32,9 + 1,9 = 34,8$.

Exercicios

- Déanse exemplos de series temporais que evolucionen: a) diariamente, b) semanalmente, c) mensualmente, d) anualmente.
- A venda de unidades de determinado produto, trimestralmente, ao longo dos anos 2008 e 2009 foi a seguinte:

	2008	2009
1º Trim	1900	3005
2º Trim	2210	3100
3º Trim	1950	3210
4º Trim	1870	2780

- Represéntese a serie. b) Determínese a tendencia axustando unha recta. c) Calcúlense as variacións estacionais para os catro trimestres. d) Estímense as unidades vendidas no cuarto trimestre de 2010 sen a compoñente estacional.
- Na seguinte táboa figuran os importes das hipotecas concedidas en Galicia por trimestre referidas aos anos 2006-2008.
 Áchese a tendencia mediante unha parábola e estímese o importe das hipotecas para o segundo trimestre de 2009. Son significativos os movementos estacionais?

2006-I	2049867
2006-II	2149826
2006-III	2071655
2006-IV	2109568
2007-I	2432061
2007-II	2589180
2007-III	2384031
2007-IV	2189651
2008-I	1968272
2008-II	1995367
2008-III	1663396
2008-IV	1389352

4. Desestacionalícese os seguintes datos e estímese a produción de enerxía eólica e solar para xuño de 2010.

Produción(=Consumo) interior: eólica e solar en España. En miles de tep			
	2007	2008	2009
xaneiro	189	237	324
febreiro	253	184	311
marzo	272	369	311
abril	122	307	314
maio	232	166	290
xuño	160	203	253
xullo	172	211	271
agosto	219	204	249
setembro	184	216	230
outubro	173	248	
novembro	245	333	
decembro	215	341	

tep (tonelada equivalente de petróleo) , enerxía liberada pola combustión dunha tonelada de cru de petróleo. $1 \text{ tep} = 41,84 \cdot 10^9 \text{ J}$.

4. Obtéñase da dirección:

<http://www.ine.es/jaxi/tabla.doper=03&type=db&divi=EPA&idtab=23&L=0>

a serie para a taxa de paro en Galicia a partir de 2005 até o último trimestre dispoñíbel. Representense os datos e calcúlese a tendencia mediante un suavizado exponencial probando con distintos valores de α . Desastionalícese a serie e fágase unha estimación para o próximo trimestre.

9. PROGRAMACIÓN LINEAR

Disolvíamos pois o estaño en ácido clorhídrico. Logo había que concentrar a solución até acadar un determinado peso específico, e deixar que se cristalizase por enfriamento. O cloruro de estaño apartábase en pequenos e grazosos prismas, incoloros e transparentes. Como a cristalización era lenta, necesitábanse moitos recipientes, e como o ácido clorhídrico ataca a todos os metais, estes recipientes tiñan que ser de metal ou barro. Nos períodos en que tiñamos moitos encargos, había que pór en circulación unha serie de recipientes de reserva [...]: unha sopeira, unha ola de ferro esmaltado, un portalámpadas estilo “liberty” e un penico.

Primo Levi. “Estaño”. *O sistema periódico*.

Introdución

Unha cooperativa pesqueira dispón de 725 tripulantes e tres tipos de barcos: A, B e C. Pretende planificar a próxima campaña en augas da NAFO de xeito que se obteña o máximo beneficio posíbel, mais ten que axustar as súas previsións a unha serie de restricións consecuencia dos acordos internacionais. O número máximo de barcos aos que se lles permite faenar é de 40 e as capturas non deben superar as 50 toneladas diarias. Mediante un estudo de rendabilidade, baseado en datos de campañas anteriores, pódese asegurar que un barco do tipo A, que precisa 10 tripulantes, producirá un beneficio de 42 000 euros. Un barco do tipo B con 15 tripulantes, 48 000 euros, e un do tipo C, con 25 tripulantes, 72 000 euros. Este último pesca diariamente, por termo medio, o dobre que cada un dos outros dous. Trátase de saber cantos barcos de cada un dos tipos deben ser enviados para obter un rendemento óptimo.

O problema consiste basicamente en **maximizar unha función linear** de tres variábeis

$$B(x,y,z) = 42x + 48y + 72z \quad \text{miles de euros}$$

onde x, y e z representan o número de barcos de cada tipo enviados a faenar. Os valores que maximicen esta función estarán limitados polas restricións explicadas anteriormente, que poden ser expresadas matematicamente mediante un sistema de inecuacións lineares. Lograr o punto óptimo que faga compatíbeis o beneficio máximo e as limitacións do acordo é o obxectivo buscado.

O enunciado exposto presenta unha situación típica dun problema de Programación Linear. En moitos campos da industria, transporte, comercio e mesmo loxística militar, preséntanse problemas análogos ao anterior, tan só diferentes na súa complexidade debido ao número de variábeis manexadas. Se o problema só contempla dúas variábeis pode ser abordado graficamente. Cando o número de variábeis aumenta, o método gráfico faise imposíbel e é preciso recorrer a un algoritmo de resolución. En 1947 George Danzig ideou para tal propósito o método simplex, que permite mediante

unha serie de aproximacións ir afinando a solución do problema até chegar á solución óptima. En 1984 Narendra Karmarkar introduciu un novo método moito máis rápido que o método simplex, e aplicábel a moitos máis supostos. Hoxe, coa axuda da computadora, podemos abordar de xeito satisfactorio a resolución de problemas de máis de dúas variábeis, mesmo cunha simple folla de cálculo.

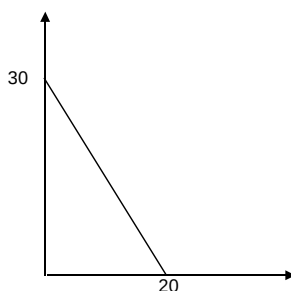
Inecuacións lineares con dúas incógnitas

Exemplo 1

Supoñamos que unha persoa dispón semanalmente de 60 horas para actividades de ocio e que quere repartir o seu tempo de lecer entre sesións deportivas de tres horas ou sesións de dúas horas de música. As posíbeis combinacións deberán satisfacer a relación:

$$3x + 2y = 60$$

onde x representa o número de sesións deportivas e y o número de sesións musicais. Ambas as dúas cantidades deben ser maiores ou iguais que cero. Graficamente a expresion anterior ven dada no plano polo segmento de liña recta seguinte:

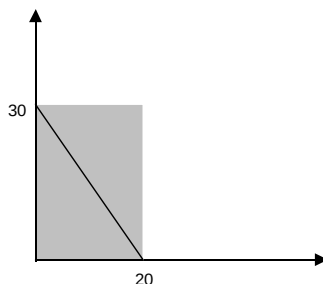


Todos os pares de puntos sobre o segmento verifican a ecuación anterior e, polo tanto, serven como solución de reparto de tempo entre as dúas actividades. Así, pares de valores como (6,21), (10,15) ou (12,12) representan posíbeis alternativas a esta reparto. Nos extremos, puntos como (20,0) e (0,30) significan dedicar todo o tempo de lecer a unha única actividade.

Supoñamos agora que a persoa non quere cubrir necesariamente todo o seu tempo libre nestas dúas actividades e contempla a posibilidade de deixar parte deste tempo para outras. As posíbeis combinacións virán agora dadas pola expresión:

$$3x + 2y \leq 60$$

Puntos como (5,0), (8,8) e mesmo (0,0) dánnos agora distribucións do tempo que son válidas co novo criterio. En realidade, toda a rexión do plano limitada pola recta e os eixos coordenados proporciona pares de valores que satisfán a relación de desigualdade formulada.



Unha desigualdade como a anterior na que aparece unha expresión linear nas variábeis x e y cos signos $>$, $<$, $\leq$, ou $\geq$ recibe o nome de **inecuación linear con dúas variábeis**. A forma xeral dunha inequación deste tipo será:

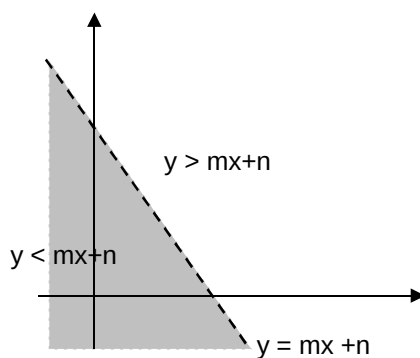
$$ax + by + c \leq 0 \quad (\text{ou } >, <, \geq)$$

Sendo a , b e c números reais, dos que a e b non se anulan simultaneamente.

Se na inequación despexamos y , adopta a forma :

$$y < mx + n$$

e mudando o signo “ $<$ ” polo “ $=$ ”, $y = mx + n$ representa unha liña recta que divide o plano en dous semiplanos.



Os puntos que están sobre a recta satisfán a igualdade $y = mx + n$. Os puntos do semiplano superior son as solucións da inequación $y > mx + n$ e os puntos do semiplano inferior as solucións de $y < mx + n$. Así, todas as solucións dunha inequación linear con dúas incógnitas están nun dos dous semiplanos dos que a recta é fronteira.

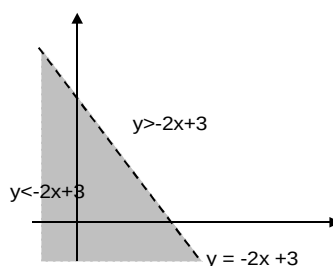
No caso en que a inequación inclúa os signos $\leq$ ou $\geq$, a mesma recta formará parte das solucións. Se, pola contra, na inequación aparecen os signos $>$ ou $<$, os puntos da recta non satisfarán a desigualdade e, polo tanto, deben excluírse. Neste caso adóitase representar a recta con trazos descontinuos.

É interesante observar que o semiplano que nos dá o conxunto de puntos solución dunha inequación linear con dúas incógnitas é un conxunto **convexo**, é dicir, tal que se contén dous puntos A e B , tamén contén todos os puntos do segmento AB .

Exempo 2

Representemos a rexión do plano dada pola inecuación: $2x + y < 3$. En primeiro lugar despexamos y e representamos a recta correspondente.

A zona sombreada corresponderá ao conxunto de solucións da inecuación dada.



Exercicios.

1. Debúxese a rexión do plano descrita polas inecuacións:

$$2x + 3y \geq 6$$

$$3x - 27 \geq 12$$

$$x + 2y \leq 7$$

$$y \geq 6 - 2x$$

$$-x \leq 2y - 4$$

$$2x + y \leq 10$$

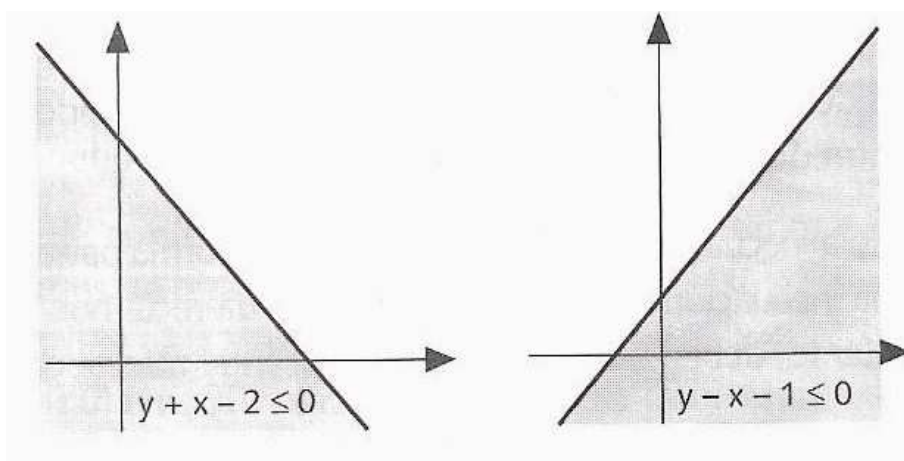
$$3x + y \geq 0$$

$$x + 25 \leq -5$$

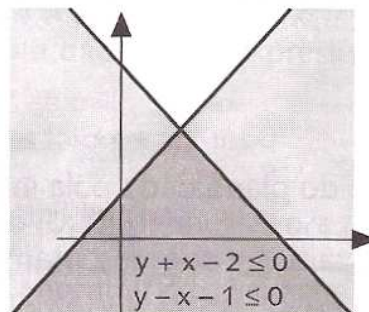
Sistemas de inecuacións lineares

Supoñamos que queremos atopar agora todos os puntos do plano que satisfagan simultaneamente as dúas inecuacións :

$$\begin{cases} y + x - 2 \leq 0 \\ y - x - 1 \leq 0 \end{cases}$$



A zona común ás dúas rexións daranos os puntos que satisfán ambas as dúas inecuacións.

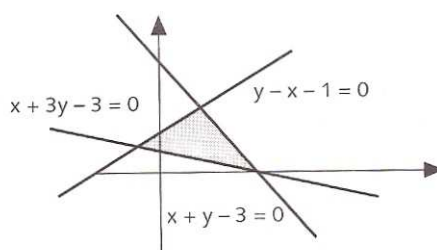


Observamos que a rexión solución é convexa posto que é intersección de dous semiplanos que son conxuntos convexos.

Resolvamos agora un sistema con tres inecuacións.

$$\begin{cases} x + y - 3 \leq 0 \\ x + 3y - 3 \leq 0 \\ y - x - 1 \leq 0 \end{cases}$$

representando as rectas correspondentes, resulta:



Os puntos da rexión sombreada serán as solucións do sistema.

Ás veces, as condicións impostas polo sistema de inecuacións son contraditorias e a rexión de solucións resulta polo tanto baleira. Sistemas deste tipo dinse incompatíbeis. Rexións como a do exemplo anterior, que poden ser rodeadas por un círculo, dinse acotadas. Moitos sistemas de inecuacións teñen conxuntos de solucións que non posúen esta propiedade. Nos exercicios seguintes aparecen exemplos de sistemas incompatíbeis e rexións de solucións non acotadas.

Exercicios

2. Resólvanse graficamente os sistemas de inecuacións seguintes.

$$\begin{cases} 2x + y \leq 1 \\ y \geq -x \\ 2x + 6 \leq 0 \end{cases} \quad \begin{cases} 4x + 3y \geq 12 \\ y \geq x \\ 3x - 2y + 6 \geq 0 \end{cases} \quad \begin{cases} 2x - 5y \geq -10 \\ 2x - 3y \leq 6 \\ y \geq 0 \end{cases}$$

Un problema de optimización

Unha comunidade veciñal propón levar a cabo a rexeneración dun monte queimado mediante repoboación forestal. Pretende plantar dous tipos de especies autóctonas: carballos e castiñeiros. Comeza a experimentar cunha pequena parcela de 180 metros cadrados. O tempo investido na reparación, transporte, plantación etc., de cada árbore é de 1 hora, precisando cada carballo 1kg de abono e cada castiñeiro 2. Ademais, a parcela adicada a cada carballo debe ter 2 metros cadrados de superficie e a adicada a cada castiñeiro 1 metro cadrado. Se dispón de 160 kg de fertilizante e man de obra por un total de 100 horas, e sabendo que o beneficio producido por cada exemplar de carballo é de 240 euros e por cada exemplar de castiñeiro 360 euros, cal será o número de árbores de cada especie que se debe plantar para tirar o máximo rendemento económico á parcela?

Resumimos os datos do problema na seguinte táboa:

	Metros	Fertilizante	Horas traballo	Beneficio
Carballos	2m	1kg	1h	240€
Castiñeiros	1m	2kg	1h	360€
Cant. Disponíbeis	180m	160kg	100h	

Se chamamos x ao número de carballos plantados e y ao número de castiñeiros, o problema consiste en buscar os valores de x e y que fagan máxima a función

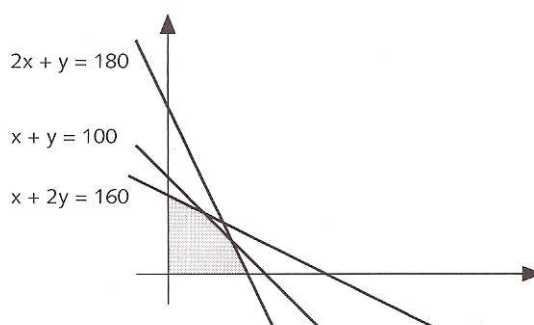
$$B = 240x + 360y,$$

que nos dará o beneficio producido. Esta función, linear de dúas variábeis recibe o nome de **función obxectivo**.

A optimización desta función está limitada por unha serie de restricións expresadas polos restantes datos da táboa. En primeiro lugar, os valores buscados de x e y deben ser enteiros e non negativos, debendo satisfacer ademais as limitacións en canto a metros cadrados, horas de traballo e cantidades de abono dispoñíbeis. A formulación destas condicións lévanos ao seguinte sistema de inecuacións :

$$\begin{cases} x \geq 0 \\ y \geq 0 \\ 2x + y \leq 180 \\ x + 2y \leq 160 \\ x + y \leq 100 \end{cases}$$

que recibe o nome de **sistema de restricións**, e ten como solucións a rexión sombreada do plano:

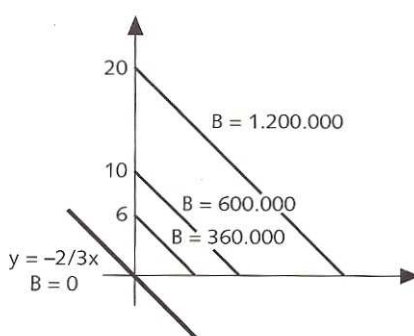


Cada un dos puntos con coordenadas enteiras dentro da rexión sombreada representa unha posíbel solución do problema. Por iso, chamáremola **rexión factíbel**, e cada punto dela será unha **solución factíbel**. O problema estará resolto se conseguimos seleccionar de entre todas as solucións factíbeis aquela que fai máxima a función obxectivo, que chamaremos **solución óptima**.

Se despexamos na función obxectivo: $B = 240x + 360y$, obtemos :

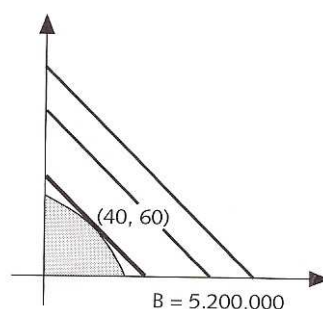
$$y = -\frac{240}{360}x + \frac{B}{360} = -\frac{2}{3}x + \frac{B}{360}$$

Esta relación representa unha familia de rectas paralelas, todas con pendente $-2/3$, e distintas ordenadas na orixe segundo os valores que asignemos a B . Na figura seguinte aparecen representadas algunhas.



Canto maior sexa a ordenada na orixe, maior será o valor de B e, polo tanto, maior o beneficio. A recta que teña maior ordenada na orixe e fique dentro da rexión de solucións factíbeis daranos a solución óptima do problema.

Se xuntamos nunha única gráfica as dúas anteriores, resulta:



Vemos que a recta da familia de maior ordenada corta á rexión factíbel no vértice intersección de dúas rectas fronteira. Resolvendo o sistema formado polas dúas rectas, obtemos o punto $(40,60)$ que corresponderá ao valor que optimiza a función.

Deberanse, pois, plantar 40 carballos e 60 castiñeiros para tirar o beneficio máximo de 30 120 euros.

Demóstrase que **unha función linear definida sobre unha rexión factíbel acoutada e non baleira ten polo menos un valor máximo ou mínimo que se atopa nun vértice**. Podemos entón atopar unha solución óptima calculando os valores da función obxectivo en cada un dos vértices da rexión factíbel e escollendo o máximo.

Os vértices do recinto teñen de coordenadas:

$$(0,0), (90,0); (80,20); (40,60); (0,80)$$

Se calculamos os valores de B para cada un destes puntos temos:

$$(0,0), B = 0; \quad (90,0), B = 18\ 100; \quad (80,20), B = 24\ 240; \quad (40,60), B = 30\ 120; \quad (0,80), B = 24\ 480.$$

Obsérvase que o valor máximo corresponde ao vértice (40,60) que é o calculado anteriormente.

Programación linear

O exemplo anterior representa un caso moi sinxelo dun problema de programación linear. Vemos que basicamente consiste en optimizar -facer máxima ou mínima- unha función linear que chamamos función obxectivo e está suxeita a un conxunto de restricións dadas por un sistema de inecuacións. Podemos enunciar:

Un problema de **Programación linear con dúas variábeis** consiste en:

1. **Optimizar** unha función linear de dúas variábeis :

$$B = ax + by$$

que se chama **función obxectivo**.

2. Sometida a un conxunto de **restricións** dado por un sistema de inecuacións :

$$\begin{cases} a_1x + b_1y \leq c_1 \\ a_2x + b_2y \leq c_2 \\ \dots\dots\dots \\ a_nx + b_ny \leq c_n \end{cases}$$

onde x e y deben ser non negativas.

Para resolver o problema graficamente seguimos os pasos:

1. Representamos o conxunto de solucións do sistema de inecuacións que resultará unha rexión convexa do plano chamada **rexión factíbel**.
2. Cada un dos puntos desta rexión satisfará todas as **restricións** do problema e, polo tanto, será unha **solución factíbel**.

3. Representamos a recta que nos defina a **función obxectivo** para o caso $B = 0$. Trazando rectas paralelas a esta poderemos obter a que ten ordenada na orixe maior e corta á rexión factíbel. O punto de corte daranos a **solución óptima**.
4. Podemos obviar este último paso tendo en conta que a **solución óptima se debe atopar nun dos vértices da rexión factíbel** e, polo tanto, chega con escoller o valor máximo entre os que a función obxectivo acada neles.

Exercicios

3. Optimícese as seguintes funcións suxeitas ás correspondentes restricións.

$B = 5x + 6y$	$B = 2x - 3y$	$B = 10 + 10y$	$B = x + 5y$
$\begin{cases} x, y \geq 0 \\ x + y \leq 6 \\ x - 2y \geq 0 \end{cases}$	$\begin{cases} x, y \geq 0 \\ x + y \geq 5 \\ 3x - y \leq 3 \end{cases}$	$\begin{cases} x, y \geq 0 \\ y - x \leq 1 \\ 2x + y \leq 8 \\ y - 2x \geq -4 \end{cases}$	$\begin{cases} 40 - 20y \geq -800 \\ 2x - y \leq 10 \\ 50x + 10y \leq 500 \end{cases}$

4. Unha fábrica produce dous modelos de artigos utilizando unha cadea de tres máquinas. O primeiro artigo precisa de 2 horas na primeira máquina e 1 hora en cada unha das outras dúas. O segundo, 1 hora nas máquinas primeira e segunda, e 3 horas na terceira. As horas semanais dispoñíbeis por máquina son 70, 40 e 90, respectivamente. Compróbase que o beneficio polo primeiro artigo é de 240 euros e polo segundo 360 euros. Cantas unidades de cada artigo deben fabricarse por semana para obter un beneficio máximo?

5. Unha cooperativa de produción de flores en invernadoiro ten unhas necesidades mínimas de 80 unidades semanais dun produto antifungos, 120 de desinfectante e 240 de desparasitante. No mercado existen dous produtos A e B, coas seguintes composicións: o produto A contén 2 unidades de antifungos, 6 de desinfectante e 4 de antiparasitante. A marca B contén 2,2 e 12 unidades respectivamente. Se o prezo do quilo de A é 24 euros e o de B 300 euros, cantos quilos de cada marca se deben comprar semanalmente para que o custo sexa mínimo?

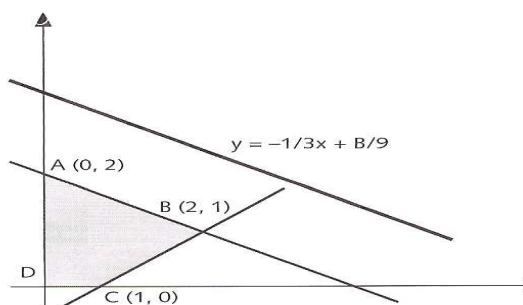
6. Os propietarios dunha granxa de coellos estiman que a dieta alimentaria destes debe conter un mínimo de 20 unidades de proteínas e 16 unidades de hidrocarbóns. Saben que o penso A contén 4 unidades de proteínas e 2 de hidrocarbóns e o penso B, 1 unidade de proteínas e 2 de hidrocarbóns. Se o penso A custa 0,72 euros por kg e o penso B 0,48 euros, cantos kg de cada penso deben comprarse para que o custo da alimentación sexa mínimo?

Casos especiais

Trátase de maximizar a función : $B = 3x + 9y$, coas condicións :

$$\begin{cases} x, y \geq 0 \\ x + 3y \leq 6 \\ x - y \leq 1 \end{cases}$$

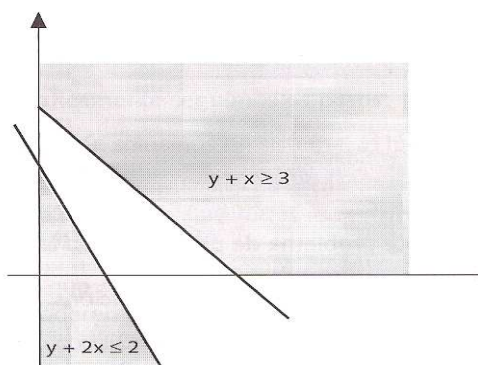
Se facemos unha representación gráfica do problema, resulta:



Observamos que a recta definida pola función obxectivo é paralela á recta que pasa por A e B e, en consecuencia, a intersección desta recta co conxunto de solucións factíbeis é o segmento de extremos A e B. A función obxectivo acada o seu valor máximo en todos os puntos do segmento e, polo tanto, temos infinitas solucións óptimas.

Consideremoss agora a maximización de $B = 5x + 3y$, suxeita ás restricións:

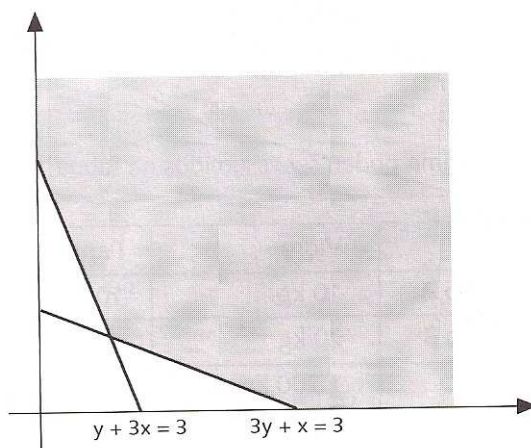
$$\begin{cases} x, y \geq 0 \\ y + 2x \leq 2 \\ y + x \geq 3 \end{cases}$$



A rexión de solucións factíbeis é baleira. Non existe ningún punto que verifique simultaneamente o sistema de restricións. Polo tanto, non hai solución óptima.

Consideremos, por último, a maximización de $B = 5x + 3y$, suxeita as restricións:

$$\begin{cases} x, y \geq 0 \\ y + 3x \geq 3 \\ 3y + x \geq 3 \end{cases}$$



A rexión de solucións factíbeis non está acoutada. Non existe unha solución óptima que maximice o problema. Porén, obsérvase que si é posíbel atopar un valor que minimice a función. Se unha rexión factíbel é non acoutada, se existe un valor que maximice ou minimice a función obxectivo, atoparase nun vértice da rexión.

Exercicio

7. Maximícense, se é posíbel, as seguintes funcións de acordo coas correspondentes restricións:

$$B = y + x$$

$$\begin{cases} x, y \geq 0 \\ 5x + 6y \geq 30 \\ 2x + 8y \geq 16 \end{cases}$$

$$B = 30 + 20y$$

$$\begin{cases} x, y \geq 0 \\ 5x - 2y \leq -10 \\ 2x - y \leq 2 \end{cases}$$

$$B = 5y - x$$

$$\begin{cases} x, y \geq 0 \\ x + y \leq 1 \\ y - x \geq -1 \end{cases}$$

$$B = 40y + 3x$$

$$\begin{cases} x, y \geq 0 \\ y \geq 3 \\ 4x - y \leq 12 \\ y - x \geq 0 \end{cases}$$

O problema dual

Un concello dispón de dous tipos de colectores de lixo. O tipo A consta de dous compartimentos que admiten diariamente 10 kg de vidro e 10 de papel, respectivamente; o colector B ten outros dous nos que caben 20 de vidro e 40 de papel. A planta de reciclado municipal está capacitada para tratar diariamente un máximo de 1200 Kg de vidro e 1800 de papel. Sábese que os colectores se enchen por completo todos os días. Se o beneficio do produto resultante do reciclado do lixo A é de 0,60

euros e o dun colector B 1,44 euros, cantos colectores de cada tipo se deben instalar par que este beneficio sexa máximo?

Os datos do problema pódense resumir na táboa:

	Vidro	Papel	
Colector tipo A	10 kg	10 kg	0,60 euros por colector
Colector tipo B	20 kg	40 kg	1,44 euros por colector
	1 200 kg máx	1 800 kg máx	

Se chamamos x ao número de colectores que se van instalar do tipo A e y ao número de colectores do tipo B, podemos formular o problema como maximizar a función beneficio : $B = 0,60x + 1,44y$, suxeita ás restricións:

$$\begin{cases} 10x + 1y \leq 1200 \\ 10x + 40y \leq 1800 \end{cases}$$

Mais tamén podemos tratalo desde outro punto de vista. Chamando x ao custo por kg do reciclado de vidro e y ao custo do reciclado dun kg de papel, a función que nos dá o custo máximo do reciclado diario virá dada por:

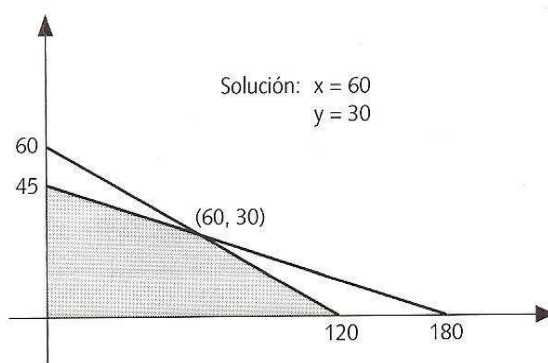
$$C = 1\,200x + 1\,800y,$$

f

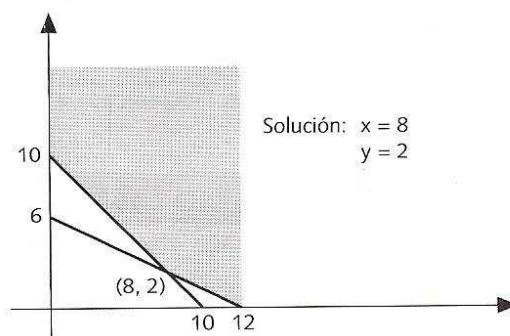
unción que corresponde á máxima capacidade da planta recicladora. A cuestión agora é minimizar esta función sometida á condición de que os prezos que nos dan a rendabilidade de cada colector, canto menos, se conserven:

$$\begin{cases} 10x + 10y \geq 0,60 \\ 20x + 40y \geq 1,44 \end{cases}$$

Se resolvemos o primeiro problema atopamos:



Neste caso, teremos:



Obsevamos que ambos puntos de vista conducen á mesma solución:

$$B = 79,2 \text{ euros}$$

O resultado pode ser aplicado a unha situación máis xeral. Dous problemas deste tipo dicense **duais**. Un problema e o seu dual caracterízanse por:

Problema orixinal	Problema dual
Maximizar: $B = px + qy$	Minimizar: $C = rx + sy$
Coas restricións:	Coas restricións:
$\begin{cases} x, y \geq 0 \\ ax + by \leq r \\ cx + dy \leq s \end{cases}$	$\begin{cases} x, y \geq 0 \\ ax + cy \geq p \\ bx + dy \geq q \end{cases}$

Observemos que:

Un problema de maximizar a función obxectivo tranfórmase noutro de minimizar a función obxectivo.

Os coeficientes da función obxectivo e os termos independentes das inecuacións intercámbianse entre si.

Os sistemas de inecuacións invirten os signos das desigualdade. O maior ou igual pasa a ser menor ou igual e viceversa.

Nas matrices dos sistemas de inecuacións cámbianse filas por columnas, isto é, son transpostas.

Pódese probar que **se un problema de programación linear ten solución óptima entón o problema dual tamén a ten, e ademais ambas solucións coinciden**. Isto permite optar pola resolución dun problema ou do seu dual segundo sexa máis conveniente.

Exercicios.

8. Nos seguintes casos fórmúlese o problema dual e compróbase que as solucións óptimas coinciden.

Maximizar

$$B = 2x + 3y$$

Minimizar

$$C = 8x + 12y$$

Maximizar

$$B = x - y$$

$$\begin{cases} x, y \geq 0 \\ x + y \leq 6 \\ y - x \leq 4 \end{cases} \quad \begin{cases} x, y \geq 0 \\ 2x + 27 \geq 1 \\ x + 3y \geq 2 \end{cases} \quad \begin{cases} x, y \geq 0 \\ 2x - y \leq 13 \\ y - x \geq 3 \\ x + y \geq 11 \end{cases}$$

O problema xeral de programación linear

O problema xeral de programación linear en **forma normal** vén dado por:

Maximizar a función obxectivo:

$$F(x_1, x_2, \dots, x_n) = c_1 x_1 + c_2 x_2 + \dots + c_n x_n$$

de xeito que verifique o sistema de restricións:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2 \\ \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m \end{cases}$$

sendo:

$$x_1, x_2, \dots, x_n, b_1, b_2, \dots, b_m \geq 0$$

O método xeométrico empregado para resolver un problema de programación linear con dúas incógnitas non resulta axeitado nos casos nos que aparecen máis. A representación dunha rexión factíbel para tres variábeis daríanos unha rexión convexa no espazo de tres dimensións limitada por certo número de planos. Resulta difícil interpretar en termos gráficos esta rexión e atopar as posíbeis solucións. Se o número de variábeis é maior, a solución por este método é xa imposíbel. A técnica de resolución para problemas de máis de dúas variábeis consiste en atopar unha primeira solución factíbel e íla mellorando con novas solucións que se aproximen cada vez máis á solución óptima. Esta idea levou a George Danzig a crear un método que dun xeito totalmente mecánico, mediante operacións aritméticas elementais, conduce á solución óptima dun problema de programación linear independentemente do número variábeis que conteña. Co tempo xurdiron novos métodos de resolución que o mesmo que o de Danzig son programábeis no computador. Aínda así, hai problemas de programación linear aparentemente inofensivos, que ao aumentar o número de variábeis, requirirían, mesmo optimizando todos os recursos informáticos mundiais, bastante máis tempo de execución que o transcorrido dende a *Big Bang*.

No que segue imos ilustrar como se utiliza o comando **SOLVER** ou **RESOLVEDOR**, que conteñen as follas de cálculo máis usuais, para resolver problemas de máis de dúas variábeis.

Resolvamos en primeiro lugar o seguinte problema en dúas variábeis xa proposto anteriormente utilizando a folia de cálculo. Tratábase de calcular o número de carballos e castiñeiros que maximizaban o beneficio dado pola función:

$$B = 240x + 360y$$

coas restricións:

$$\begin{cases} x \geq 0 \\ y \geq 0 \\ 2x + y \leq 180 \\ x + 2y \leq 160 \\ x + y \leq 100 \end{cases}$$

O procedemento de resolución é o seguinte:

As celas A1 e A2 representaran os valores de x e y que maximizan a función obxectivo.

Na cela B1 escribimos a función obxectivo: " $= 240*A1 + 360*A2$ "

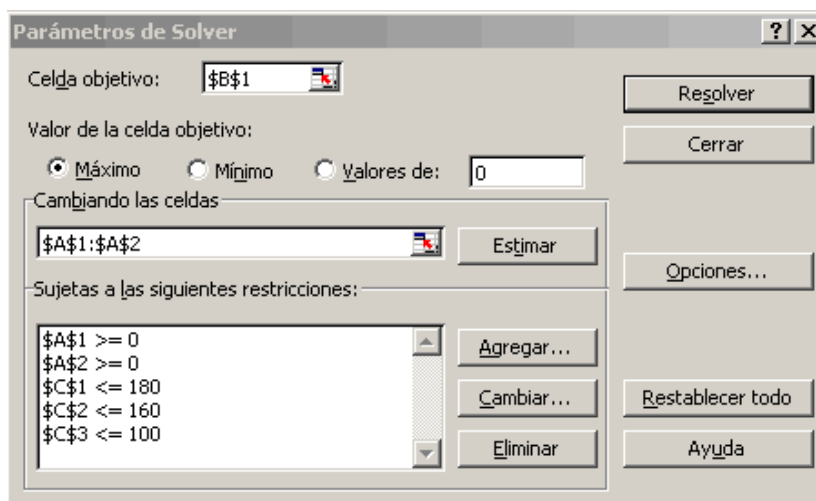
Nas celas C1:C3 escribimos as restricións:

En C1: " $= 2*A1+A2$ "

En C2: " $= A1 +2*A2$ "

En C3: " $=A1 *A2$ "

No menú Ferramentas escollemos o comando SOLVER. Aparecerá un cadro de diálogo como o que segue:



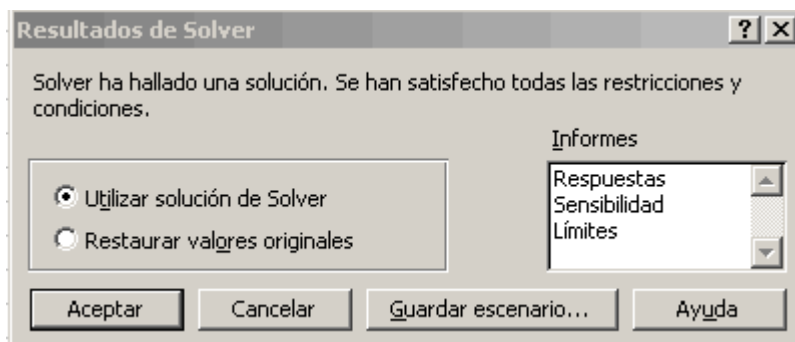
Na cela obxectivo escribimos a referencia de B1.

Señalamos máximo ou mínimo, segundo proceda.

No cadro "cambiando as celas" escribimos A1 e A2 na forma que aparece na ilustración.

No cadro das restricións utilizamos o botón AGREGAR para introducir cada unha delas. Na figura están xa introducidas.

Prememos RESOLVER e aparecerá o cadro:



aceptando, a folla de cálculo terá este aspecto:

	A	B	C	D
1	40	31200	140	
2	60		160	
3			100	
4				

En A1 e A2 aparecen os valores que maximizan a función. No noso caso 40 carballos e 60 castiñeiros.

En B1 o valor que acada a función beneficio: 31 200 euros

En C1, C2 e C3 os valores que acadan as restricións. Obsevemos que se utilizan completamente as horas de traballo e a cantidade de fertilizante. Unicamente plantamos en 140 m do total dos 180 que ten a parcela.

O procedemento pódese aplicar sen ningunha dificultade a un número de variábeis maior que dous. Simplemente abonda utilizar máis celas da columna A para as novas variábeis e máis celas da columna C para as restricións.

Imos resolver un problema en tres varábeis. Maximizar a función:

$$B = x - 12y + 4z$$

suxeita ás restricións :

$$\begin{cases} x, y, z \geq 0 \\ 4x + 3y - z \leq 1 \\ -x - y + z \leq 2 \\ x - y - z \leq 1 \end{cases}$$

Utilizamos A1,A2 e A3 para x,y, e z, respectivamente.

En B1 Escribimos a función obxectivo: “ = A1 - 12*A2 + 4*A3 ”

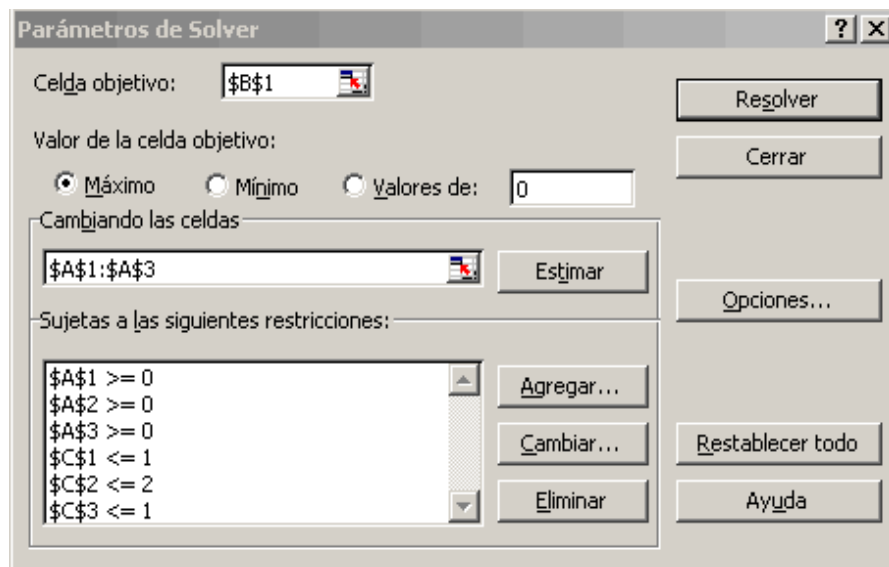
En C1,C2 e C3 escribimos as restricións :

En C1: “ = 4*A1 + 3*A2 - A3 “

En C2: “ = -A1 - A2 + A3 “

En C3: “ = A1 -A2 -A3

Prememos o comando SOLVER, enchemos o cadro de diálogo como no caso anterior:



obtendo os seguintes resultados:

	A	B	C	D
1	1	13	1	
2	0		2	
3	3		-4	
4				

dános as solucións: $x = 1$; $y = 0$; $z = 3$. Con $B = 13$ para a función obxectivo.

Exercicios

- Resólvase utilizando o comando Solver da folla de cálculo o problema da pesca formulado ao principio deste capítulo.
- Maximícense as seguintes funcións:

$$\begin{array}{lll}
 B = 10x + 10y & B = 15x + 60y & B = 9x + 5y \\
 \left\{ \begin{array}{l} x, y \geq 0 \\ 2x + y \leq 6 \\ y - x \leq 3 \end{array} \right. & \left\{ \begin{array}{l} x, y \geq 0 \\ 2x + y \leq 14 \\ 4x + y \leq 24 \\ y \leq 6 \end{array} \right. & \left\{ \begin{array}{l} x, y \geq 0 \\ x + y \leq 8 \\ x \leq 5 \\ y - x \leq 2 \end{array} \right.
 \end{array}$$

$$\begin{array}{lcl}
 B = y - x + 2z & B = 60y + 90v & B = 4x + y - 12z \\
 \left\{ \begin{array}{l} x, y, z \geq 0 \\ y + z \leq 1 \\ x + 2y - z \leq 2 \end{array} \right. & \left\{ \begin{array}{l} x, y, z, v \geq 0 \\ v - 2z \leq 7 \\ y - 2x \leq 2 \\ x + y \leq 5 \\ z + v \leq 4 \end{array} \right. & \left\{ \begin{array}{l} x, y, z \geq 0 \\ x - y - z \leq 2 \\ y - z - x \leq 1 \\ -x + 4y + 3z \leq 1 \end{array} \right.
 \end{array}$$

11. Unha empresa de transportes dispón de colectores de 24 metros cúbicos e 92 toneladas de carga máxima. A compañía dispón de dous clientes que fan envíos ao mesmo destino: o primeiro manda caixas de 30 kg de peso e un volume de 0,02 metros cúbicos; o segundo, paquetes de 50 kg de peso e 0,01 metros cúbicos de volume. Se o primeiro cliente paga 4,5 euros por paquete e o segundo 3 euros, cal debe ser o número de paquetes que debe cargar cada cliente no colector de xeito que o beneficio sexa máximo?
12. Un concello recibe tres propostas de tres empresas de recollida selectiva de lixo. A empresa A proporciona colectores que admiten 10 kg de vidro e 30 de plásticos, pagando 1,2 euros polo contido diario de cada un deles. A empresa B instalaría colectores con capacidade para 10 kg de papel e 20 de plásticos, pagando a 1,8 euros. Por outra banda, a capacidade diaria do servizo de recollida municipal límitase a 100 kg de vidro, 120 de papel e 300 de plásticos. Se supoñemos que os colectores instalados se enchen cada día, calcúlese o número de colectores que deben solicitar a cada empresa para que o beneficio pola recollido do lixo sexa máximo.
13. Unha compañía de alugueiro de avionetas ten dous tipos de aparellos: de pasaxe e de cabotaxe. As de pasaxe poden transportar 6 pasaxeiros e 100 quilos de carga; as de cabotaxe 3 pasaxeiros e 400 quilos de carga. O beneficio neto que se lles saca ás avionetas de cabotaxe é de 90 euros cada 100 km, e ás de pasaxe 60 euros cada 100 km. Cantos voos deben realizarse con cada tipo de avioneta para transportar a un grupo de 100 persoas e 4 000 quilos de equipaxe a un punto que dista 300 quilómetros? Cal será o beneficio obtido?
14. Un fabricante de cadaleitos fainos de dous tipos. Os de tipo especial requiren 8 horas de traballo/home e os de tipo normal 5 horas de traballo/home. Os materiais para os de tipo especial custan 180 euros e para os de tipo normal 120 euros. O beneficio que se obtén por cada venda do especial é de 120 euros, mentras que o normal reporta un beneficio de 60. Calcular o número de cadaleitos de cada tipo que debe fabricar á semana para obter o máximo beneficio, tendo en conta que debe fabricar como mínimo 15 do tipo especial e 10 do tipo normal, tan só dispón de 340 horas/home semanais e o custo dos materiais empregados semanalmente non deberá superar os 6 000 euros.
15. Un concesionario de motocicletas para abastecerse de nova mercadoría para á venda, pode escoller entre dous modelos. O modelo A cústalle 1 800 euros e o B 2 400 euros. Ten só espazo para almacenar 30 motos e ten un capital de 36 000 euros para gastar. Se pensa gañar 600 euros coa venda de cada moto do tipo A e 900 con cada moto do tipo B, cantas motos de cada tipo terá que mercar par obter o máximo beneficio?
16. Un fabricante fai dous tipos de produtos, A e B. Precísanse tres máquinas I, II e III para fabricar cada produto. Un artigo tipo A necesita de 2 horas da máquina I,

- 2 da máquina II e 5 da III, mentres que un do tipo B necesita, respectivamente, 3,4 e 2. Semanalmente dispón de 24 horas de máquina I, 45 da II e 50 da III e o beneficio por artigo é de 480 euros para o produto A e de 720 euros para o produto B. Cantas unidades de cada tipo se deben fabricar semanalmente para maximizar o beneficio?
17. Unha fábrica de plásticos produce dous tipos de cafeteiras para micro-ondas, unha para diario, D, e outra para cando hai invitados, I. Tárdase o duplo en fabricar unha do tipo I que unha do tipo D e a empresa podería realizar 1 000 por día se só producira o modelo D. O subministro de plástico é suficiente para producir 800 cafeteiras por día combinando a produción de I coa de D. A versión I require dun acabado especial que limita a súa produción a 500 diarias. Se a fábrica consegue un beneficio de 1,8 e 3 euros, respectivamente, por cada unidade de tipo D e de tipo I vendidas, cantas de cada tipo tería que producir diariamente para maximizar o beneficio?

10. ERROS. INTERPOLACIÓN. INTEGRACIÓN APROXIMADA.

*Se o nariz de Cleopatra fose máis curto,
tería mudado toda a face da terra.*

Blaise Pascal, *Penses*.

Segundo Plutarco, Cleopatra non era fisicamente moi agrazada, mais como descendente dos Ptolomeos, con antepasados gregos ou macedonios, era loura e iso facíaa moi atractiva. Aínda que sobre o talento político e a intelixencia de Cleopatra parece existir acordo, sobre o seu físico veñen arrastrándose até a actualidade moitas dúbidas. Parece ser que máis que pola súa beleza seducía pola súa conversa, pola súa voz melosa e polas súas extraordinarias habilidades cosméticas. Tratábase dunha princesa refinada e moi culta, que non necesitaba intérpretes para falar cos etíopes, árabes, hebreos, medos e partos.

O caso é que fose ou non debido ao seu nariz, a atracción de Marco Antonio cara a Cleopatra si está fóra de toda dúbida, e Pascal coa anterior reflexión só quere subliñar o feito de cómo ínfimas modificacións nas condicións iniciais – o rostro de Cleopatra– poden ter importantes consecuencias no devir temporal. Se a frota de Marco Antonio fuxiu en desbandada na batalla de Actium cando a vitoria estaba ao seu alcance, foi por seguir o navío de Cleopatra. A derrota de Marco Antonio propiciou a chegada de Augusto ao poder, durante cuxo mandado houbo un grande florecemento intelectual da man de Mecenas,....Sería distinta, de non ser por Cleopatra, o desenvolvemento da civilización occidental?

Nas matemáticas e nas ciencias que as utilizan acontece algo semellante, pequenas diferenzas nas observacións, nos redondeos ou nas estimacións de determinados valores descoñecidos, poden evolucionar no tempo de tal xeito que fan inservíbel calquera modelo de predición.

A teoría dos erros, e da súa propagación, é de grande importancia na física e na matemática aplicada. Cando se dá unha estimación do valor dunha magnitude descoñecida é fundamental saber cal é a máxima discrepancia que estamos dispostos a admitir, o que non é un problema menor cando descoñecemos o valor que se pretende estimar.

Aquí tratamos de métodos de aproximación de funcións, que nos permitirán facer interpolacións, e de cálculo de integrais mediante métodos aproximados. En ambos casos vai ser fundamental acoutar o erro, para o cal introduciremos brevemente a nomenclatura da teoría clásica e da probabilista, non sen antes facer unha incursión nalgúns aspectos da propagación dos erros desde o punto de vista do caos determinístico

Iteración da fórmula do interese composto que produce caos

Ao considerarmos o incremento dunha poboación durante n anos, se supoñemos que a inicial é p_0 , e ao cabo de n anos p_n , a taxa de crecemento anual sería:

$$r = \frac{p_{n+1} - p_n}{p_n}$$

Se o crecemento anual fose constante, despois teríamos:

$$p_{n+1} = (1 + r)p_n \quad (1),$$

Daquela, partindo da poboación inicial p_0 , ao cabo de n anos:

$$p_{n+1} = (1 + r)^n p_0$$

Este é un exemplo de crecemento exponencial que como sabemos tamén pode utilizarse para calcular o capital acumulado a partir do inicial p_0 , durante n anos a un xuro do $r \cdot 100\%$.

En 1845 o matemático belga Pierre François Verhulst formulou unha lei de crecemento para poboacións na que a taxa ía variando dende r a 0 até que se acadaba certa poboación obxectivo p . Esta situación pode modelizarse mudando en (1) r por $r - cp_n$, onde c é unha constante. Xa que pretendemos que a taxa de crecemento sexa

cero cando $p_n = p$, teremos $c = \frac{r}{p}$, co cal

$$p_{n+1} = (1 + r - cp_n)p_n = (1 + r)p_n - cp_n^2 \quad (2)$$

Se a poboación é menor que p medrará até p . Se é maior que p diminuírá até p . En ambos os dous casos o proceso permanecerá constante unha vez acade a poboación o tamaño p .

A función definida por (2) podemos representala utilizando a folla de cálculo, extendendo o seguinte esquema até a fila vixésima. Tomamos $r = 0,8$.

	B
1	0,8
2	1
3	$=(1+B\$1)*B2-(B\$1/3)*B2^2$
4	$(1+B\$1)*B3-(B\$1/3)*B3^2$
....	

de onde obtemos:

	B
1	0,8
2	1
3	1,53333333
4	2,13303704
5	2,62617413
6	2,88796929
....	
16	2,99999999
17	3
18	3
19	3
20	3

Podemos representar os datos anteriores seleccionando no menú gráfico: *dispersión* → *liñas suavizadas*:

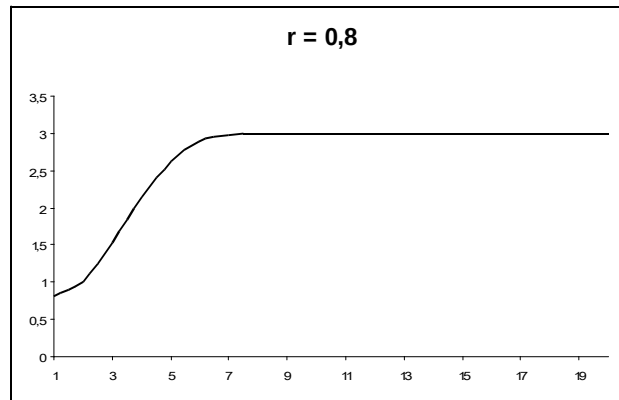
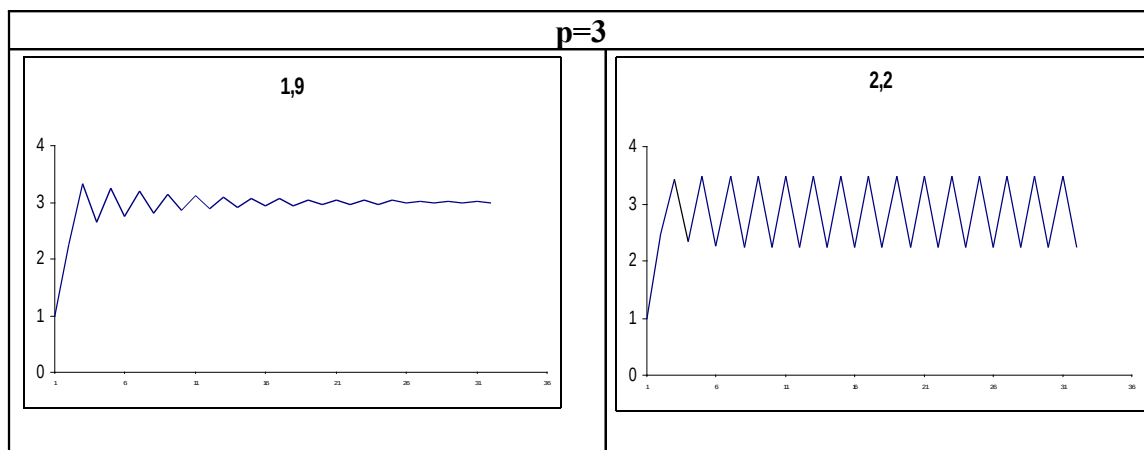


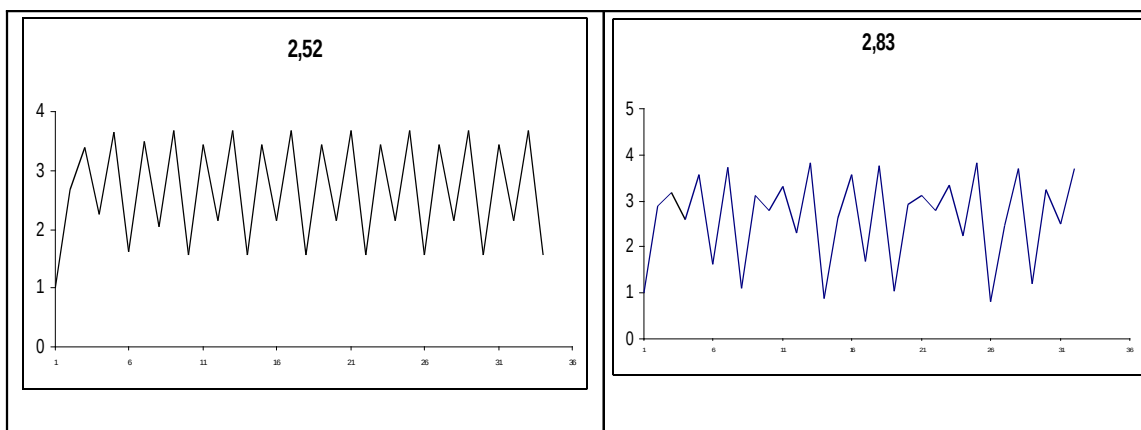
Gráfico 1.

Este era o comportamento esperado por Verlhust para calquera valor de r , mais, en realidade tal comportamento só se dá se $r \leq 2$. Para valores maiores que 2, un crecemento maior que o 200%, o proceso desestabilízase facendo que a sucesión de valores no canto de aproximarse a un valor límite, se concentre en torno a dous, catro,..., valores, dependendo do que valla r .

Exercicio

Fáganse representacións para distintos valores de r , repetindo o proceso anterior, e compárense cos seguintes gráficos.





O conxunto de Verhulst

O comportamento anterior foi estudado por Edward Lorenz en 1963 e ten especial interese en relación con determinados procesos como a hidrodinámica e a teoría cinética de reaccións. Os valores aos que converge o proceso para valores de r menores que 2,57 chámanse *atractores*. Se r é menor que 2 temos un único atractor. Para valores de r comprendidos entre 2 e 2,5 obtéñense dous atractores oscilantes. Para valores entre 2,5 e 2,55 hai un período de catro puntos. Para valores de r maiores de 2,57 o sistema presenta un comportamento caótico, mais, estudado a modioño revela novamente unha orde, a orde dun comportamento *fractal*.

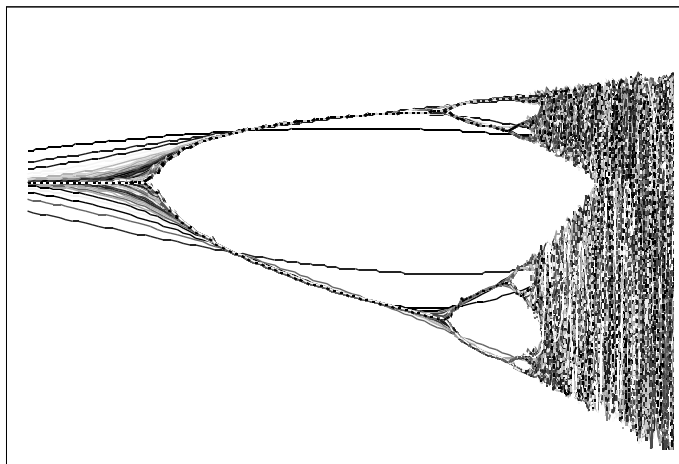
Fagamos agora na folla de cálculo o cómputo das funcións obtidas para os valores de r seleccionados –de centésima en centésima- segundo o seguinte esquema:

	A	B			DE	DF
1	1,8	=A1+0,01			=DD1+0,01	=DE1+0,01
2	1	1	...		1	1
3	=(1+A\$1)*A2-(A\$1/3)*A2^2	=(1+B\$1)*B2-(B\$1/3)*B2^2			=(1+DE\$1)*DE2-(DE\$1/3)*DE2^2	=(1+DF\$1)*DF2-(DF\$1/3)*DF2^2
...		...				
153	=(1+A\$1)*A152-(A\$1/3)*A152^2	=(1+B\$1)*B152-(B\$1/3)*B152^2	...	...	=(1+DE\$1)*DE152-(DE\$1/3)*DE152^2	=(1+DF\$1)*DF152-(DF\$1/3)*DF152^2

Obtemos:

	A	B			DE	DF
1	1,8	1,81			2,87	2,88
2	1	1	...		1	1
3	2,2	2,20666667			2,91333333	2,92
...		...				
152	3	3	...		0,456628608	3,85287991
153	3	3	...	...	1,5715497	3,85287991

Se seleccionamos o rango A1:DF153, e escollemos un gráfico de dispersión con liñas suavizadas, **por filas**, aparece o gráfico do chamado conxunto de Verlhust.



Integración gráfica de ecuaciones diferenciais

A expresión $\frac{dv}{dt} = 9,8 - 0,2v$ pode representar, poñamos por caso, a aceleración dun paracaidista na caída. Nela relaciónase a variación da velocidade coa propia velocidade, mediante unha igualdade que recibe o nome de *ecuación diferencial*.

Unha ecuación deste tipo pode resolverse graficamente por métodos aproximados sen máis que considerar taxas de variación media, con incrementos de t axeitados, en troques da taxa de variación instantánea $\frac{dv}{dt}$. Daquela, a ecuación anterior transformárase en:

$$\frac{V(t+h) - V(t)}{h} = 9,8 - 0,2V(t).$$

Tomando, por exemplo $h=0,1$, mediante o seguinte esquema de folla de cálculo

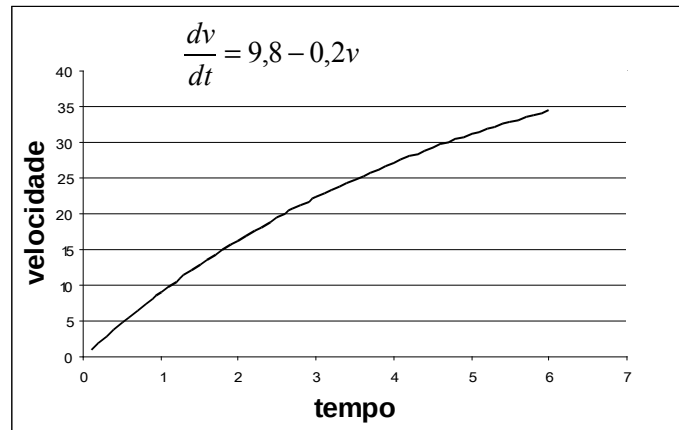
	A	B
1	t	v
2	0	0
3	0,1	=B2+0,1*(9,8-0,2*B2)
4	=A3+0,1	=B3+0,1*(9,8-0,2*B3)
...	...	

obtemos a velocidade en función do tempo

	A	B
1	t	v
2	0	0
3	0,1	0,98

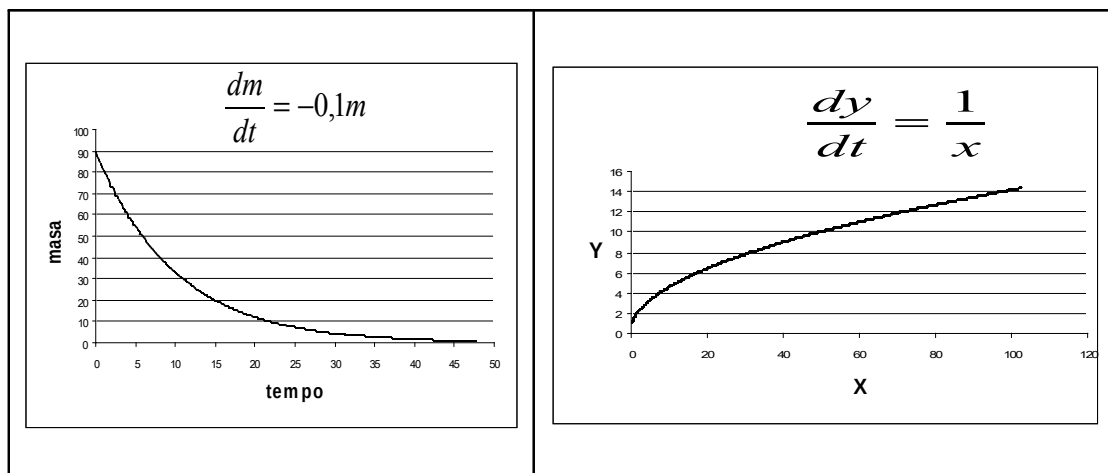
4	0,2	1,9404
...	...	
72	7	37,0869935

e representando graficamente:



Exercicio

Comproba, programando as correspondentes follas de cálculo, que se obteñen as seguintes gráficas para cadansúa das ecuacións diferenciais.



O efecto bolboreta

En 1972, no foro do “Congreso para o progreso da ciencia” preguntáronlle a Edward Lorenz, o pai da moderna meteoroloxía, se era posíbel que a perturbación producida polo bater das ás dunha bolboreta no Brasil producise un ciclón en Texas. Sen dar unha negativa rotunda, Lorenz respondeu que tamén outros animais, incluído o home, poderían producir perturbacións susceptíbeis de orixinar ciclóns, e que mesmo algunha outra bolboreta podería colaborar a que algún ciclón se desfixese.

O *efecto bolboreta*, como paradigma de pequenas perturbacións que poden producir co tempo tremendos efectos, ou incluso conducir a determinados sistemas regulados por ecuacións deterministas ao caos, pasou a ser reinterpretado nos máis variados contextos, a maioría das veces, por artistas e tertulianos, dun xeito completamente estrafalario.

Edward Lorenz foi quen de reducir as entrambilicadas relacións do fluxo atmosférico a tan só tres ecuacións diferencias, a partir das cales, transformadas lixeiramente, para determinados valores tamén podemos obter *bolboretas*.

$$\frac{dx}{dt} = -ax + ay$$

$$\frac{dy}{dt} = -xz + cx - y$$

$$\frac{dz}{dt} = xy - bz$$

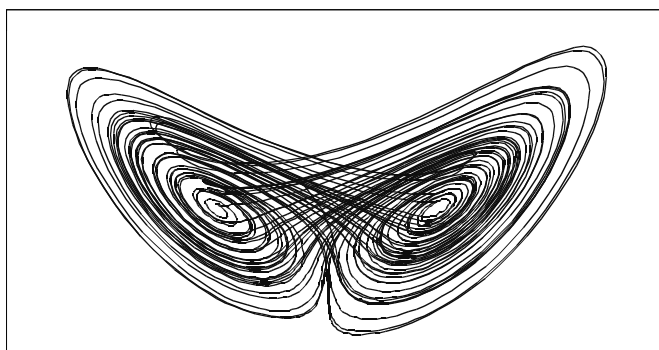
As tres ecuacións do sistema poden reinterpretarse como taxas de variación media, mediante o esquema seguinte, no que h é a amplitude do intervalo:

$X(t+h) = X(t) + h \cdot a \cdot (Y(t) - X(t))$
$Y(t+h) = Y(t) + h \cdot (X(t) \cdot (c - Z(t)) - Y(t))$
$Z(t+h) = Z(t) + h \cdot (X(t) \cdot Y(t) - b \cdot Z(t))$

que podemos programar na folla de cálculo:

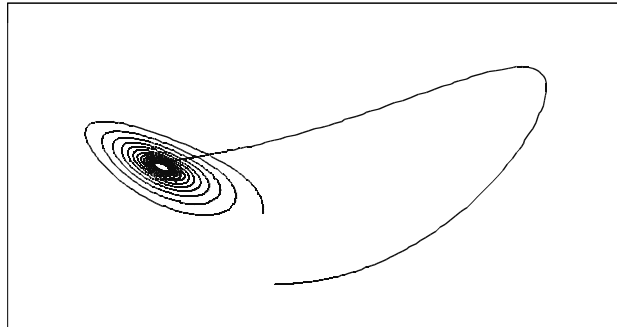
	A	B	C	D	E	F	G
1	x	y	z	a	b	c	h
2	-2	1	32	10	2,67	28	0,01930047
3	=A2+ \$G\$2*\$D\$2*(B2-A2)	=B2+ \$G\$2*(A2*(F\$2-C2)-B2)	=C2+\$G\$2*(A2*B2-\$E\$2*C2)				
4	=A3+ \$G\$2*\$D\$2*(B3-A3)	=B3+ \$G\$2*(A3*(F\$2-C3)-B3)	=C3+\$G\$2*(A3*B3-\$E\$2*C3)				
							
308 4	=A3083+ \$G\$2*\$D\$2*(B3083-A3083)	=B3083+ \$G\$2*(A3083*(F\$2-C3083)-B3083)	=C3083+ \$G\$2*(A3083*B3083-\$E\$2*C3083)				

e dá lugar á “bolboreta” :

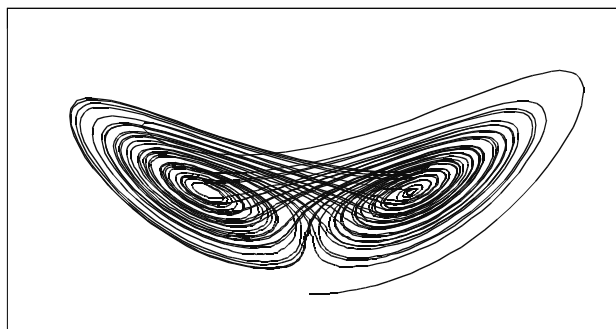


Cambiando os valores iniciais, obtemos distintas figuras:

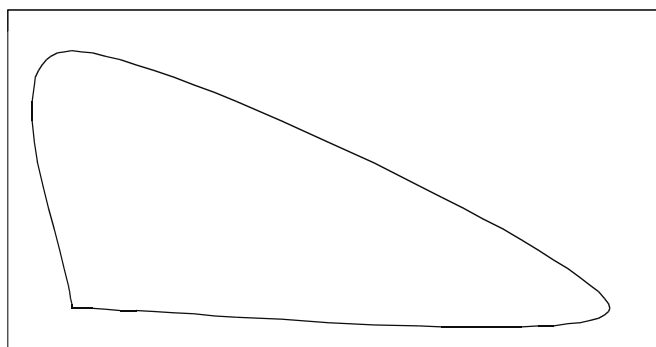
x	y	z	a	b	c	h
0	1	0	10	2,67	28	0,00325851



x	y	z	a	b	c	h
0	1	0	10	2,67	28	0,016099183



x	y	z	a	b	c	h
0	1	0	10	2,67	28	0,025180063



Erros

Dicimos que o número $\hat{a}$ é unha aproximación de a se $\hat{a}$ difire “lixeramente” de a , e pode usarse no seu lugar como resultado, ou pode substituílo en cálculos posteriores.

Poñamos por caso, 3,14 é unha aproximación de π , 2,4142 é unha aproximación de $\sqrt{2}$ e 2,7 é unha aproximación de e .

Se $\hat{a} > a$ diremos que a aproximación é por exceso e se $\hat{a} < a$ diremos que é por defecto. Así, por exemplo 2,414 é unha aproximación por defecto de $\sqrt{2}$, e 2,415 éo por exceso.

Chámase **erro** da aproximación á diferenza $a - \hat{a}$.

O **erro absoluto** defínese como $\Delta a = |a - \hat{a}|$.

Na maioría dos casos nos que se utiliza a aproximación, o número a non será coñecido, polo que nin o erro nin o erro absoluto poderán ser calculados polas expresións anteriores. Daquela, cómpre substituír o erro absoluto por unha estimación dunha cota superior súa.

Diremos que Δa é unha **cota de erro** da aproximación de a mediante $\hat{a}$ se $|a - \hat{a}| < \Delta a$.

Por exemplo, ao tomar 3,14 como estimación de π , temos:

$|\pi - 3,14| \leq 3,15 - 3,14 = 0,01$, polo que: $\Delta \pi = 0,01$. Analogamente, se tomamos 1,414 como aproximación de $\sqrt{2}$, a cota de erro será 0,001 xa que $1,414 \leq \sqrt{2} \leq 1,415$.

Defínese o **erro relativo** como $\delta = \frac{\Delta a}{|a|}$. Dado que tanto a cantidade que aparece no numerador como a do denominador terán que ser estimadas, trábállase polo xeral coa **cota de erro relativo**, definida por $\delta_a = \frac{\Delta a}{|\hat{a}|}$. Daquela, poñamos por caso, o erro relativo ao estimar π e $\sqrt{2}$ nos casos anteriores, estará acotado respectivamente por:

$$\delta_\pi = \frac{0,01}{3,14} \text{ e } \delta_{\sqrt{2}} = \frac{0,001}{1,41}.$$

Redondeo

-Ao dicirmos que a distancia (media) da Terra ao Sol é de 150 millóns de quilómetros, estamos tomando como cero todos os algarismos para alén da orde do millón de quilómetros.

-Se dicimos que temos 38,3 °C de temperatura, estamos tomando como cero todos os algarismos a partir da orde das centésimas.

En xeral, diremos que un número queda **redondeado** até a posición n , se se fan cero todos os algarismos situados á dereita da referida posición. Se temos en consideración que os números que redondeamos proveñen das medidas de certas magnitudes, semella razoábel tomar ao sumo unha cifra máis que a da orde inmediatamente inferior á medida utilizada. Así, ao buscar distintas posibilidades para ir en coche de Santiago a Lugo nunha páxina web, obtemos as distancias, 104, 107 e 106 quilómetros. Daquela, tomamos como estimación da distancia 106 e non

$$\frac{104 + 106 + 107}{3} = 105,6$$

Para redondear números existen certas regras que, aínda que de carácter aparentemente empírico, se fundamentan na teoría probabilística de erros. Son as seguintes:

1. Ao redondear un número a **n** algarismos, o algarismo na posición enésima ficará invariante se a parte truncada é menor que media unidade da posición enésima.
2. Se a parte truncada é maior que media unidade de orde **n**, o algarismo na posición **n** aumentarase nunha unidade.
3. Se a parte truncada é exactamente igual a media unidade de orde **n**, o algarismo que ocupa o lugar enésimo aumentarase nunha unidade se é impar e ficará invariante se é par.

Exemplos:

79 321 852 redondeado a 5 algarismos é 79 322 000.
3,141592 redondeado a 5 decimais é 3,14159.
9,52508 redondeado a dous decimais é 9,52
2,750 redondeado a un decimal é 2,8.

Algarismos significativos

As cifras significativas dun número son aquelas das que se está razoabelmente seguro da súa certeza e proporcionan, xa que logo, información real sobre a cantidade que representan. Un algarismo proporciona tanta máis información canto máis á esquerda estea situado. Se un número foi redondeado para que tan só conteña cifras significativas, estas comezarán polo primeiro algarismo non nulo. Se a parte decimal remata cun cero ou varios, estes serán significativos por definición. Se o número é enteiro e remata cun cero ou varios, ten que decidirse do contexto se estes son ou non significativos.

Se chegamos a un número como resultado dunha medida física ou de certos cálculos, deberán darse cifras significativas de xeito que a **cota de erro** non exceda nalgúns unidades a derradeira cifra significativa.

Por exemplo, 14 000 pode ter

- a) dous algarismos significativos se o expresamos como $1,4 \cdot 10^4$
- b) tres algarismos significativos se o expresamos como $1,40 \cdot 10^4$
- c) catro algarismos significativos se o expresamos como $1,400 \cdot 10^4$

7,090 kg ten catro cifras significativas

0,000987 ten tres cifras significativas.

Clasificación dos erros

Deixando á marxe as equivocacións ou erros de vulto, debido ao esquecemento ou substitución dalgún algarismo, á colocación dunha vírgula no lugar trabucado, etc, que adoitan ser detectados nun atento exame das cifras, as fontes esenciais de erro son:

1. Erros iniciais, que teñen a súa orixe nas deficiencias dos instrumentos de medida ou nas condicións en que esta se realiza.
2. Erros locais de truncadura, que se orixinan ao substituír procesos nalgún sentido infinitos por procesos finitos.
3. Erros locais de redondeo, que están motivados polo feito de que practicamente todos os números, ao calcular con eles, deben ser redondeados a un número limitado de algarismos.

Erros e probabilidade

Cada medida dunha magnitude ven afectada por un grande número de factores. Supoñendo eliminados os erros debidos a causas localizadas, quedarán moitas causas que teñen pequenos efectos que, conxuntamente, xerarán **erros aleatorios**. Estes erros, froito de causas parcialmente independentes, admítase que seguen unha distribución normal de media cero.

Consideramos que as medidas de calquera magnitude aparecen aumentadas ou diminuídas nunha parte ε , debida a flutuacións aleatorias. Se realizamos n medidas, a i -ésima virá dada por $x_i = \mu + \varepsilon_i$, sendo μ o verdadeiro valor. Ao realizarmos un número de medidas o suficientemente grande, os erros por exceso e por defecto tenderán a compensarse, obténdose daquela unha boa estimación de μ . Baixo estas hipóteses, medir determinada magnitude é estimar a media dunha poboación que se distribúe normal con media μ e desvío σ .

Como estimación da media μ tomaremos a media da mostra (de medidas efectuadas) que irá acompañada do correspondente intervalo de confianza $\left[\bar{X} - k \frac{\sigma}{\sqrt{n}}, \bar{X} + k \frac{\sigma}{\sqrt{n}} \right]$, onde k dependerá do coeficiente de confianza utilizado. A semilongitude do intervalo chámase **erro admisíbel** para o correspondente nivel de confianza. O desvío estándar, σ , ben se coñece de experiencias anteriores ou ben se estima a partir do cuasi-desvío da mostra.

Exemplo

Para unha arandela utilizada en radorreceptores obtivéronse as seguintes medidas para o seu espesor dadas en milímetros: 2,21, 2,30, 2,23, 2,19.

Obtense unha media de 2,2325 mm e un cuasi-desvío de 0,04787136, que para un coeficiente de confianza de 0,95, buscando na taboa da distribución t con 3 graos de liberdade, dános o intervalo [2,23087028, 2,23412972], o que supón un erro admisíbel de 0,00162972.

Exercicio.

Viator Volatus usuario da compañía *Amol-Ar*, farto de ter que desfacer a maleta por exceso de peso ao facturar, nesta ocasión chega ao aeroporto con moito tempo de antelación e dedícase a pesar a equipaxe en toda canta báscula ve libre, obtendo os seguintes pesos, en quilos:

15,180	15,075	14,700	15,120	15,000	14,800
15,160	15,100	14,500	14,600	14,990	14,770
14,560	14,875	14,675	14,900	14,700	14,900

14,500	14,470	15,200	14,900	15,200	14,800
14,600	14,600	14,850	14,800	14,625	14,900

Á hora de facturar, no mostrador pertinente, a maleta dá un peso de 15,100 kg, e como o máximo autorizado é de 15 quilos, o empregado invítao a alixeirar equipaxe. Viator, desta xeira non cede. Cal será a súa argumentación?

Interpolación

En moitas ocasións, a partir duns poucos datos correspondentes a observacións illadas dun determinado proceso, cómpre coñecer os seus valores para calquera outro instante dun intervalo de tempo ben delimitado. Situacións deste tipo son as seguintes:

1. O número de nacementos en Galicia no ano 2000 foi de 19 418 e no 2007 de 21 752, cal foi o número de nacementos en 2005?
2. Obsérvase que unha balsa de auga ten perdas desde os primeiros días do mes. O duodécimo día tiña 1 000 millóns de litros, o vixésimo primeiro 820 millóns e o trixésimo (o derradeiro do mes), 600. Canta auga habería o oitavo día? Canta haberá o décimo do mes seguinte?
3. Temos os seguintes datos, en miles, referentes a un cultivo de bacterias: no instante inicial había 7, ao pasar dous minutos 11, aos catro minutos 23 e aos cinco 32. Cantas bacterias haberá ao cabo de media hora?

Os problema presentados anteriormente resólvense asumindo a hipótese de partida de que os valores observados proveñen dunha certa función $f(x)$ que se trata de identificar.

No primeiro caso, tomando a orixe en 2 000 para facilitar os cálculos, podemos considerar que os datos proveñen dunha función $f(x)$ tal que $f(0)=19\,418$ e $f(7) = 21\,752$. Se pasamos revista ás funcións máis sinxelas que coñecemos, as polinómicas, atopámonos con que con tan só dúas observacións unicamente podemos identificar unha de primeiro grao, é dicir, unha recta.

Trátase entón de obter a ecuación da recta que pasa por $(0, 19418)$ e $(2, 21752)$.

Facendo os cálculos pertinentes: $f(x) = 19\,418 + 333,43\,x$. Daquela: $f(5) = 21\,085$.

No segundo suposto trátase de identificar unha función da que coñecemos os seguintes datos: $f(12) = 1000$, $f(21) = 820$ e $f(30) = 600$.

Botando man das funcións polinómicas, sabemos que tres puntos abundan para obter unha parábola: $f(x) = ax^2 + bx + c$.

Para identificar os coeficientes teñamos en conta que a anterior ecuación tera que ser verificada polos tres puntos cos que contamos. Daquela:

$$\begin{aligned} 1000 &= 144a + 12b + c \\ 820 &= 441a + 21b + c \\ 600 &= 900a + 30b + c \end{aligned}$$

Resolvendo: $a=-0,25$, $b= 63,75$ e $c=271$, co cal a función buscada é

$$f(x) = -0,25x^2 + 63,75x + 271$$

A cantidade de auga que queda o oitavo día será $f(8)=1\,066,98$ millóns de litros, e o décimo día do mes seguinte (pasaron 40 días) haberá $f(40) = 303,78$ millóns de litros.

Procedendo no terceiro suposto de modo análogo aos anteriores, teríamos: $f(0) = 7$, $f(2) = 11$, $f(4) = 23$ e $f(5) = 32$.

Estes catro puntos permiten obter unha función polinómica de terceiro grao: $f(x) = ax^3 + bx^2 + cx + d$ cuxos coeficientes se obteñen resolvendo o sistema:

$$\begin{aligned} 7 &= d \\ 11 &= 8a + 4b + 2c + d \\ 23 &= 46a + 16b + 4c + d \\ 32 &= 125a + 25b + 5c + d \end{aligned}$$

e daquela:

$$f(x) = -0,18x^3 + 2,43x^2 - 2,15x + 7, \text{ e } f(30) = 6989.$$

Conforme aumenta o grao do polinomio, os cálculos fanse máis entrambilicados por termos que resolver sistemas con cada vez maior número de ecuacións e incógnitas. Aínda que na actualidade unha simple folla de cálculo nos permite a súa resolución inmediata, os esforzos feitos no pasado tendían por unha banda a facilitar os cálculos e pola outra a aproveitar os xa feitos para calcular o polinomio dun determinado grao no caso de termos máis información para calcular o de grao superior. Neste contexto é no que temos que considerar as achegas de Newton e Lagrange para obter os polinomios de interpolación.

Polinomio de interpolación de Newton

Partindo dos datos $f(3) = 7$ e $f(5) = 11$ podemos escribir o polinomio de interpolación de primeiro grao, que pretendemos calcular, da forma:

$$y = a + b(x-3)$$

Substituíndo os valores iniciais, temos: $a = 7$ e $11 = 7 + b(5-3)$, ou sexa $b = 2$. O polinomio de interpolación será: $y = 7 + 2(x-3)$.

Consideremos agora que temos unha nova observación $f(9) = 20$ para engadir ás anteriores.

Os datos con que contamos son: $f(3) = 7$, $f(5) = 11$ e $f(9) = 20$.

Podemos escribir o polinomio de interpolación de segundo grao da forma:

$$y = a + b(x-3) + c(x-3)(x-5)$$

que substituíndo dá lugar ao sistema: $a = 7$, $a + 2b = 11$, $a + 6b + 24c = 20$.

Para cuxa solución podemos aproveitar os valores de a e b anteriormente calculados. Daquela, obtense $c = 0,04$, e o polinomio de interpolación será:

$$y = 7 + 2(x-3) + 0,04(x-3)(x-5)$$

Se engadimos agora o dato $f(12) = 42$, podemos calcular o polinomio de terceiro grao sen máis que obter o coeficiente d a partir da expresión:

$$y = a + b(x-3) + c(x-3)(x-5) + d(x-3)(x-5)(x-9)$$

facendo $x = 12$, obtemos:

$$42 = a + b(12-3) + c(12-3)(12-5) + d(12-3)(12-5)(12-9)$$

$$42 = 7 + 2(12-3) + 0,04(12-3)(12-5) + d(12-3)(12-5)(12-9), \text{ de onde } d = 0,08.$$

O polinomio de terceiro grao será:

$$y = 7 + 2(x-3) + 0,04(x-3)(x-5) + 0,08(x-3)(x-5)(x-9)$$

En xeral, se queremos calcular o polinomio de interpolación de grao n a partir de $n+1$ puntos de abscisas $x_0, x_1, \dots, x_{n-1}, x_n$, escribiremolo da forma:

$$y = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)\dots(x - x_{n-1})$$

Deste xeito os cálculos fanse moito máis doados e, ademais, en caso de contar con novos datos serven todos os coeficientes, até o grao $n-1$, calculados anteriormente.

Polinomio de interpolación de Lagrange.

Outra forma de obter o polinomio de interpolación é mediante o método de Lagrange. No caso de que teñamos os valores $f(3) = 7$, $f(5) = 11$ e $f(9) = 20$, se queremos interpolar un polinomio de segundo grao que tome os mesmos valores que $f(x)$ nos puntos de abscisas 3, 5 e 9, podemos facer

$$F(x) = \frac{(x-5)(x-9)}{(3-5)(3-9)} 7 + \frac{(x-3)(x-9)}{(5-3)(5-9)} 11 + \frac{(x-3)(x-5)}{(9-3)(9-5)} 20$$

Evidentemente, $F(x)$ é un polinomio de segundo grao que cumpre: $F(3) = 7$, $F(5) = 11$ e $F(9) = 20$. Simplificando queda:

$$F(x) = 0,58(x-5)(x-9) - 1,38(x-3)(x-9) + 0,83(x-3)(x-5), \text{ de onde}$$

$$F(x) = 0,03x^2 + 1,8x + 1,22 \text{ é o polinomio de Lagrange.}$$

En xeral, se queremos calcular o polinomio de interpolación de Lagrange de grao n a partir dos $n+1$ puntos $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$, definimos:

$$F_n(x) = \frac{(x-x_1)(x-x_2)\dots(x-x_n)}{(x_0-x_1)(x_0-x_2)\dots(x_0-x_n)} y_0 + \frac{(x-x_0)(x-x_2)\dots(x-x_n)}{(x_1-x_0)(x_1-x_2)\dots(x_1-x_n)} y_1 + \dots$$

$$+ \frac{(x-x_0)(x-x_1)\dots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\dots(x_n-x_{n-1})} y_n$$

Ao facermos unha interpolación $F_n(x)$ cométese un erro que se pode acoutar. O problema é que para iso hai que coñecer a función da que proveñen os datos. A expresión para o erro ven dada por:

$$|\varepsilon_n| = |f(x) - F_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |(x-x_0)(x-x_1)\dots(x-x_n)|, \text{ sendo } M_{n+1} = \max_{[a,b]} |f^{n+1}(x)|.$$

Se queremos, poñamos por caso, obter unha cota para o erro cometido ao interpolar $\sqrt{20}$ a partir do polinomio de Lagrange de grao dous, tomando como puntos de interpolación 16, 25 e 36, teremos que construír o polinomio que pasa por $(16, \sqrt{16}), (25, \sqrt{25}), (36, \sqrt{36})$, que será:

$$F(x) = \frac{(x-25)(x-36)}{(16-25)(16-36)} 4 + \frac{(x-16)(x-36)}{(25-16)(25-36)} 5 + \frac{(x-16)(x-25)}{(36-16)(36-25)} 20,$$

daquela $F(20) = 4,46464646$

Para calcular a cota do erro: $f(x) = \sqrt{x}, f'(x) = \frac{1}{2}x^{-1/2}, f''(x) = -\frac{1}{4}x^{-3/2}, f'''(x) = \frac{3}{8}x^{-5/2},$

$$M_3 = \max_{[16,36]} \left| \frac{3}{8}x^{-5/2} \right| = \left| \frac{3}{8}16^{-5/2} \right| = 0,00036621 \text{ e, polo tanto:}$$

$$|\varepsilon_2| = |\sqrt{20} - F_2(20)| \leq \frac{3,3321 \times 10^{-4}}{3!} |(20-16)(20-25)\dots(20-36)| = 0,0177712, \text{ sería}$$

unha cota para o erro de interpolación. Neste caso, sen máis que botar man da computadora, ou da calculadora, sabemos que o erro cometido anda por 0,00748949, que é bastante inferior á cota achada.

Exemplos e exercicios

- 1) Calculemos $\ln 5$ utilizando o polinomio de Lagrange de grao 2, a partir dos puntos: $(1, \ln 1), (e, \ln e), (e^2, \ln e^2)$, e calculemos o erro cometido.

Os puntos dados son: $(1, 0), (e, 1), (e^2, 2)$, daquela:

$$F_2(x) = \frac{(x-e)(x-e^2)}{(1-e)(1-e^2)} 0 + \frac{(x-1)(x-e^2)}{(e-1)(e-e^2)} 1 + \frac{(x-1)(x-e)}{(e^2-1)(e^2-e)} 2$$

$$F_2(5) = \frac{(5-1)(5-e^2)}{(e-1)(e-e^2)} 1 + \frac{(5-1)(5-e)}{(e^2-1)(e^2-2)} 2 = 1,72$$

O erro ven dado por $\varepsilon_2(5) \leq \frac{M_3}{3!} |(5-1)(5-e)(5-e^2)|$, sendo $M_3 = \max_{[1,e^2]} |f^3(x)|$

Tendo en conta que $f^3(x) = \frac{2}{x^3}$ e estritamente decrecente no intervalo $[1, e^2]$, o máximo absoluto acadará en 1, daquela

$$\varepsilon_2(5) \leq \frac{2}{1^3} \frac{1}{3!} |(5-1)(5-e)(5-e^2)| = 7,27$$

Podemos observar como a acoutación do erro é demasiado “conservadora”, xa que, utilizando a calculadora ou a computadora, $\ln(5)=1,61$, co cal o erro cometido en realidade é $1,72-1,61=0,11$, moito máis pequeno que o vaticinado pola fórmula; o 7% fronte ao 350%!

2) Calculemos $\cos 0,8$ a partir de $\cos 0, \cos \frac{\pi}{6}$ e $\cos \frac{\pi}{3}$.

Sabemos que os puntos dados para construír o polinomio de segundo grao son: $(0,1)$, $(\frac{\pi}{6}, \frac{\sqrt{3}}{2})$, $(\frac{\pi}{3}, \frac{1}{2})$, entón:

$$F_n(0,8) = \frac{\left(0,8 - \frac{\pi}{6}\right)\left(0,8 - \frac{\pi}{3}\right)}{\left(0 - \frac{\pi}{6}\right)\left(0 - \frac{\pi}{3}\right)} 1 + \frac{(0,8-0)\left(0,8 - \frac{\pi}{3}\right)}{\left(\frac{\pi}{6} - 0\right)\left(\frac{\pi}{6} - \frac{\pi}{3}\right)} \frac{\sqrt{3}}{2} + \frac{(0,8-0)\left(0,8 - \frac{\pi}{6}\right)}{\left(\frac{\pi}{3} - 0\right)\left(\frac{\pi}{3} - \frac{\pi}{6}\right)} \frac{1}{2} = 0,7017213.$$

3) Segundo o Instituto Galego de Estatística, no relativo a papel e cartón, a cantidade de residuos urbanos recollidos (selectivamente e mesturados) nos anos 2002, 2004 e 2006, foron respectivamente 27138, 37807 e 40860 toneladas. A vista dos datos, cal sería a cantidade recollida en 2005?

Tomando o 2002 como ano orixe, trátase de axustar unha función polinómica aos puntos $(0, 27138)$, $(2, 37807)$, $(4, 40860)$. Se a supoñemos da forma

$P_2(x) = a + b(x-0) + c(x-0)(x-2)$, substituíndo temos o seguinte sistema:

$$a = 27138$$

$$a+2b = 37807$$

$$a+4b+8c=40860, \text{ de onde } b = 5334,5 \text{ e } c = -952$$

Para interpolar o ano 2005, calculamos

$$P_3(3) = 27138 + 5334 \cdot 3 - 952 \cdot 3 \cdot (3-2) = 40284.$$

4. Áchese unha función polinómica de cuarto grao que pase polos puntos (2,-5),(3,9), (7,8), (8,10).

Solución:

De $P_3(x) = a + bx + cx^2 + dx^3$ tal que $P_3(2) = -5$, $P_3(3) = 9$, $P_3(7) = 8$, $P_3(8) = 10$, obtemos o sistema

$$\begin{pmatrix} 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 7 & 49 & 343 \\ 1 & 8 & 64 & 512 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} -5 \\ 9 \\ 7 \\ 10 \end{pmatrix}, \text{ co cal :}$$

$$P_3(x) = -73,2 + 50,8x - 9,45x^2 + 0,55x^3$$

5. Dada a sucesión $\{2, 5, 7, 9, \dots\}$ atópese unha fórmula para o termo xeral utilizando o polinomio de Lagrange).

$$\text{Solución: } a_n = 2,17n^3 - 22,33n^2 + 6,33n - 3$$

6. A función de distribución normal con media 0 e desvío 1, ven dada por

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

Sabendo que $F(0) = 0,5$, $F(1,65) = 0,95$ e $F(1,96) = 0,975$, calcúlese $F(0,6)$.

$$\text{Solución: } F_2(0,6) = 0,72537704$$

7. Encóntrense as fórmulas de Lagrange e de Newton do polinomio de interpolación para os seguintes datos: $(-2,1)$, $(0,2)$, $(1,-2)$. Exprésense da forma $P_2(x) = a + bx + cx^2$. Coinciden ambos os dous?

$$\text{Solución: } -\frac{3}{2}x^2 - \frac{5}{2}x + 2. \text{ Si, coinciden.}$$

8. Dada a sucesión $\left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right\}$ utilícese o polinomio de Lagrange para calcular o termo xeral. Converxe a sucesión que responde a tal termo xeral?

Polinomio de Taylor

Dado que as funcións polinómicas son as máis “tratábeis”, definidas en toda a recta real e con derivada de calquera orde, utilízanse para aproximar outras funcións. O teorema de Taylor demostra que se $f(x)$ ten derivada de orde $n+1$, se pode aproximar mediante un polinomio de grao n nunha veciñanza dun punto a :

$$P_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^n(a)}{n!}(x-a)^n + R_n(x)$$

sendo $R_n(x)$ o resto, cuxa fórmula para o cálculo é:

$$R_n(x) = \frac{f^{(n+1)}(z)}{(n+1)!} (x-a)^{n+1}, \text{ onde } z \in (a, x)$$

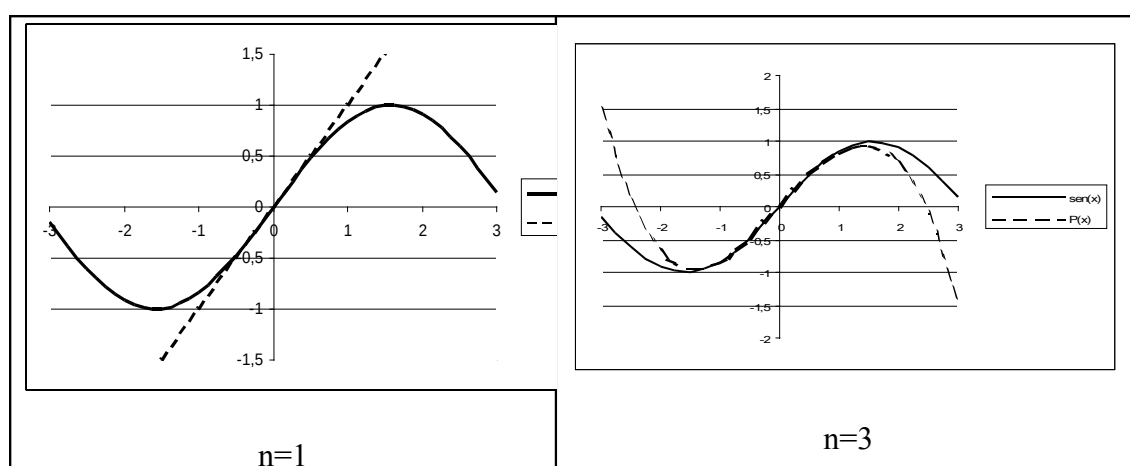
Por exemplo, a función $y = \text{sen}x$, pode ser aproximada mediante o polinomio de Taylor no punto $x=0$. Sen máis que facer as derivadas sucesivas e substituíndo, obtemos a expresión:

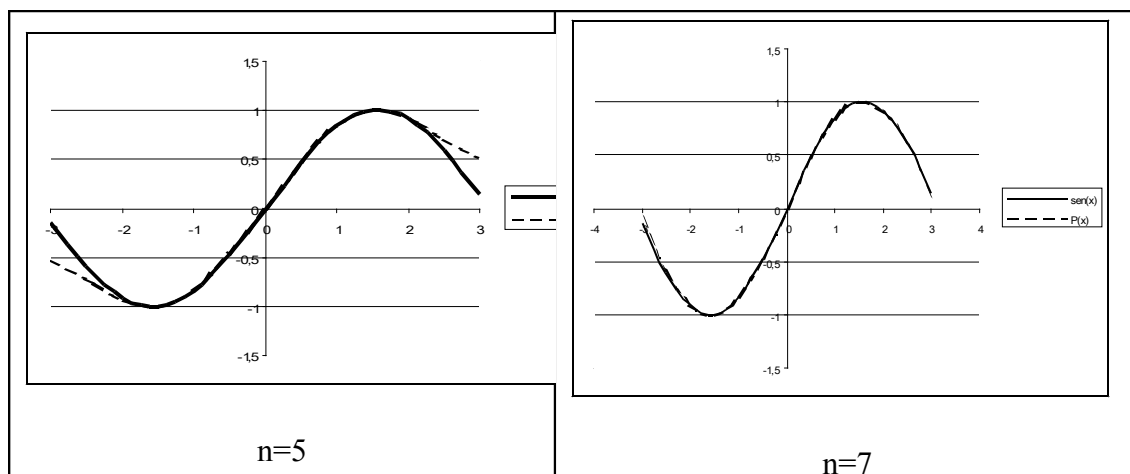
$$P_n(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \dots + (-1)^{n+1} \frac{x^{2n-1}}{(2n-1)!} + R_{2n-1}(x).$$

Podemos ver a súa aproximación a $\text{sen}x$ na veciñanza de $x = 0$, facendo as seguintes representacións gráficas a partir do esquema de folla de cálculo:

	A	B	C	D	E	F
1	x	sen(x)	P(x) n=1	P(x) n=3	P(x) n=5	P(x) n=7
2	-3	=SENO(A2)	A2	=A2-A2^3/FACT(3)	=A2- A2^3/FACT(3)+A2^5/FACT(5)	=A2- A2^3/FACT(3)+A2^5/FAC T(5)-A2^7/FACT(7)
3	=A2+0,5	=SENO(A3)	A3	=A3-A3^3/FACT(3)	=A3- A3^3/FACT(3)+A3^5/FACT(5)	=A3- A3^3/FACT(3)+A3^5/FAC T(5)-A3^7/FACT(7)
...			...		...	
14	=A13+0,5	=SENO(A14)	A14	=A14- A14^3/FACT(3)	=A14- A14^3/FACT(3)+A14^5/FAC T(5)	=A14- A14^3/FACT(3)+A14^5/F ACT(5)-A14^7/FACT(7)

Selecciónanse as columnas correspondentes e no menú gráfico escóllese: *dispersión* → *líneas suavizadas*:





Exercicio

Obtéñanse os polinomios de Taylor de grao 6 no punto $x = 0$ para as funcións:

- a) $f(x) = e^x$, b) $f(x) = \cos x$, c) $f(x) = \log(x+1)$

Represéntese, utilizando a folla de cálculo, cada función onda o polinomio calculado.

Integración numérica

En matemática aplicada, física, enxeñería ou economía é moi frecuente ter que calcular a área pechada por unha función $y = f(x)$ entre as abscisas $x = a$ e $x = b$.

No caso de que a función $f(x)$ teña unha primitiva $F(x)$ doada de calcular, temos:

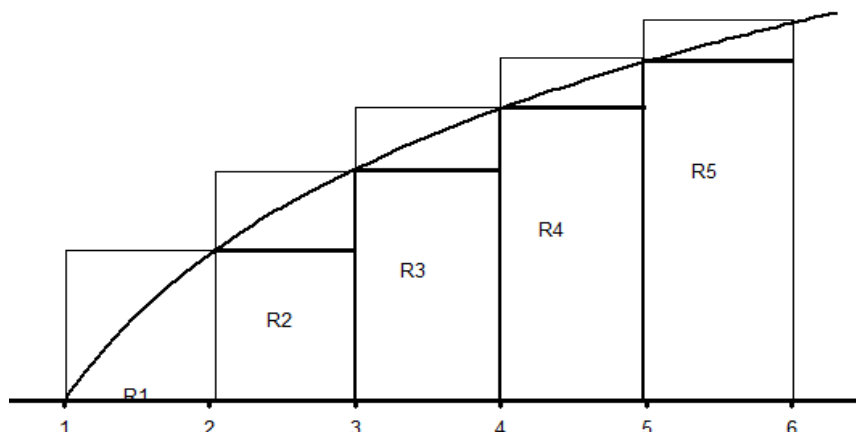
$$\int_a^b f(x)dx = F(b) - F(a)$$

Se $f(x)$ é definida positiva a interior integral é a área pechada por $f(x)$ entre a e b . E se é definida negativa o oposto da área. O problema preséntase no caso en que $F(x)$ non exista ou, aínda que exista, sexa de cálculo difícil. Neste caso non fica outro remedio que recorrer a métodos aproximados.

Aproximación mediante rectángulos

Exemplo 1

Cálculo da área pechada por $f(x) = \ln x$ no intervalo $[1,6]$



Unha primeira aproximación podemos facela mediante a suma dos rectángulos:
 $S_I = R_1 + R_2 + R_3 + R_4 + R_5 = 1 \cdot f(1) + 1 \cdot f(2) + 1 \cdot f(3) + 1 \cdot f(4) + 1 \cdot f(5) = 4,78749$
 que se chama **suma inferior**.

De modo análogo podemos considerar a **suma superior**:
 $S_S = R_2 + R_3 + R_4 + R_5 + R_6 = f(2) + f(3) + f(4) + f(5) + f(6) = 6,57925$,
 onde R_6 é o rectángulo de base $[5,6]$ e altura $f(6)$.

Temos daquela: $S_I \leq \int_1^6 \ln x dx \leq S_S$, co que $A = \frac{S_I + S_S}{2}$ pode tomarse como un valor aproximado da integral. O erro cometido estará acoutado por:

$$\left| \int_1^6 \ln x dx - A \right| < \frac{S_S - S_I}{2}$$

Neste caso $A = 5,68337$, e o erro estaría acoutado por 0,895.

Para conseguir unha mellor aproximación poderíamos dividir o intervalo en moitos máis subintervalos que servirán de base aos correspondentes rectángulos. Por exemplo, se queremos partir o intervalo $[5,6]$ en 100 subintervalos, cada un terá de lonxitude $h = \frac{6-1}{100} = 0,05$

Obteríamos así: $S_I = 0,25[f(1) + f(1,05) + f(1,10) + \dots + f(5,95)]$

e $S_S = 0,25[f(1,05) + f(1,10) + \dots + f(6)]$

Os cálculos poden disporse nunha folla de cálculo segundo o seguinte esquema:

	A	B	C	D
1	x	0,05*lnx		
2	1	=0,05*LN(A2)		
3	=A2+0,05	=0,05*LN(A3)		
....				
101	=A100+0,05	=LN(A101)*0,05	=SUMA(C2:C101)	S. Inferior
102	=A101+0,05	=LN(A102)*0,05	=SUMA(C3:C102)	S. superior
103			=(D101+D102)/2	Aproximación
			=(D102-D101)/2	erro

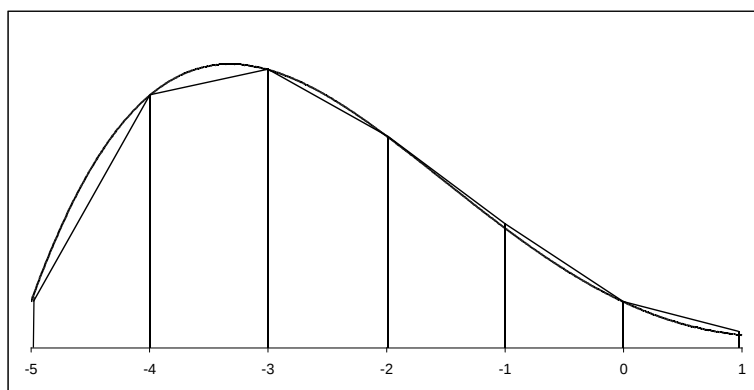
obténdose:

	A	B	C	D
1	x	0,05*lnx		
2	1	0		
3	1,05	0,00243951		
....				
101	5,95	0,08916956	5,70558923	S. Inferior
102	6	0,08958797	5,79517721	S. superior
103			5,75038322	Aproximación
104			0,04479399	erro

Método dos trapezos

Á aproximación establecida no apartado anterior pódesele dar outra interpretación.

Podemos aproximar $\int_{-5}^1 f(x)dx$, onde $f(x) = 50 - x(x-4)(x-3)(x+5)$, mediante a suma dos seis trapezos que aparecen na figura,



entón:

$$\int_{-5}^1 f(x)dx \approx \frac{f(-5)+f(-4)}{2} + \frac{f(-4)+f(-3)}{2} + \frac{f(-3)+f(-2)}{2} +$$

$$+ \frac{f(-2)+f(-1)}{2} + \frac{f(-1)+f(0)}{2} + \frac{f(0)+f(1)}{2}$$

xa que as alturas de todos os trapezos valen 1.

Mais, facendo contas decatámonos de que a expresión non é outra que a do apartado anterior: $\frac{S_I + S_S}{2}$. Daquela, mediante o esquema de folla de cálculo exposto enriba aproximamos a integral sumando áreas de trapezos.

Se queremos aproximar $\int_{-5}^1 (50 - x(x-4)(x-3)(x+5))dx$ intercalando 30 trapezos, a altura de cada un –amplitudes dos subintervalos- será $h = \frac{1 - (-5)}{30} = 0,2$, obténdose:

	A	B	C	D
--	---	---	---	---

1	x	0,2*f(x)		
2	-5	50		
...				
31	0,8	17,3344	5255,8016	
32	1	14	5219,8016	
33			5237,8016	Aprox.

En xeral, se queremos aproximar $\int_a^b f(x)dx$ mediante n trapezos, a expresión para a súa suma será: $S_n = \frac{h}{2} \left[f(a) + f(b) + \sum_{k=1}^{n-1} 2f(a+kh) \right]$, onde $h = \frac{b-a}{n}$.

Demóstrase que unha cota para o erro da aproximación ven dada por:

$|\epsilon_n| \leq \frac{m(b-a)^3}{12n^2}$, sendo m o máximo do valor absoluto da derivada segunda sobre o intervalo $[a,b]$.

Exemplos

2. Imos calcular $\int_1^3 \left(\frac{\text{sen}x}{x} + 4 \right) dx$ cun erro menor que 0,001.

$$f'(x) = \frac{x \cos x - \text{sen}x}{x^2}, \quad |f''(x)| = \left| \frac{-x^2 \text{sen}x - 2x \cos x + 2 \text{sen}x}{x^3} \right| \leq \left| \frac{1}{x} \right| + \left| \frac{2}{x^2} \right| + \left| \frac{2}{x^3} \right| \leq 5$$

Tendo en conta $|\epsilon_n| \leq \frac{5(3-1)^3}{12n^2} < 0,001$, obtense $n=58$, e xa que logo:

$$h = \frac{3-1}{58} = 0,03448276.$$

Facendo os cálculos como nos casos anteriores, obtemos:

$$\int_1^3 \left(\frac{\text{sen}x}{x} + 4 \right) dx \approx 8,90256455$$

3. Calcúlese $\int_1^4 e^{-x^2} dx$ cun erro menor que 0,05.

A cota do erro ven dada por $|\epsilon_n| \leq \frac{m(4-1)^3}{12n^2}$, onde m é o máximo do valor absoluto da segunda derivada da función no intervalo $[1,3]$. Daquela, para os valores de x pertencentes ao intervalo, temos:

$$|f''(x)| = \left| -2e^{-x^2}(1-2x) \right| \leq \left| -2e^{-x^2} \right| |1-2x| \leq \frac{2}{e} \times 5 = \frac{10}{e} = 3,68, \text{ co cal}$$

$$n^2 \geq \frac{3,68 \times 27}{12 \times 0,05} = 165,6, \text{ de onde } n=13 \text{ e } h = \frac{3}{13}.$$

Procedendo como nos casos anteriores, obtemos:

$$\int_1^4 e^{-x^2} dx \approx 0,61825295.$$

4. Calcúlese $\int_2^7 \ln(x^2 + 1)dx$ cun erro menor que 0,02.

$$|f''(x)| = \left| -\frac{2x}{(x^2 + 1)^2} \right| \leq |2x| \left| \frac{1}{(x^2 + 1)^2} \right| \leq \frac{14}{25}$$

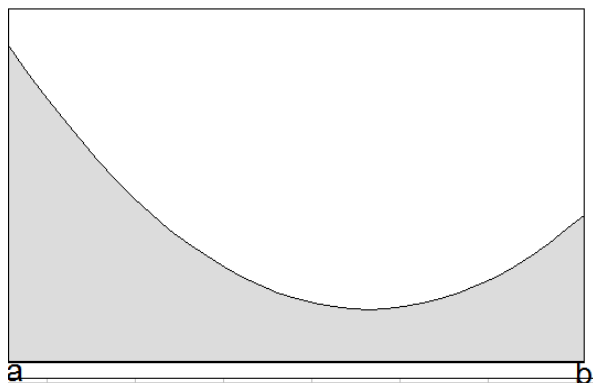
Procedendo como no caso anterior, obtemos $n = 18$, $h = \frac{5}{18}$ e

$$\int_2^7 \ln(x^2 + 1)dx \approx 11,62.$$

Aproximación mediante o método de Montecarlo

Á resolución de problemas da física ou da matemática aplicada mediante a repetición sistemática de experimentos aleatorios, simulados con computadora, chámase *método de Montecarlo*. Aquí faremos unha pequena introdución a el utilizando as prestacións da folia de cálculo para a aproximación de integrais.

Para tratar de avaliar integrais de funcións dunha variábel real partamos do seguinte esquema:



O rectángulo da figura ten de base $[a, b]$ e de altura unha cota superior, c , da función no intervalo. Se obtemos moitas parellas de números aleatorios (x, y) coa condición de que x pertenza ao intervalo $[a, b]$, e y pertenza ao intervalo $[0, c]$, baseándonos na lei de estabilidade das frecuencias, podemos facer a consideración:

$$\frac{\int_a^b f(x)dx}{\text{Area}([a, b] \times [0, c])} \approx \frac{D}{N} \quad (*),$$

sendo D o número de parellas que quedaron dentro do recinto pechado pola curva, o eixo x e as abscisas $x = a$ e $x = b$, e N o número total de parellas obtidas.

A converxencia da integral sería unha consecuencia da lei dos grandes números: “as frecuencias relativas, D/N , tendén á probabilidade teórica –primeiro membro de(*)”.

Daquela:

$$\int_a^b f(x)dx \approx \text{Área}([a,b] \times [0,c]) \times \frac{D}{N}$$

Exemplo

Imos aproximar $\int_6^{14} 200 + (x-5)(x-10)(x-17)dx$ polo método de Montecarlo.

Trátase de “meter” a gráfica da función nun rectángulo de base $[6,14]$ e altura unha cota superior da función no intervalo (neste caso poderíamos ter calculado o máximo da función no intervalo pechado).

Imos calcular unha cota superior:

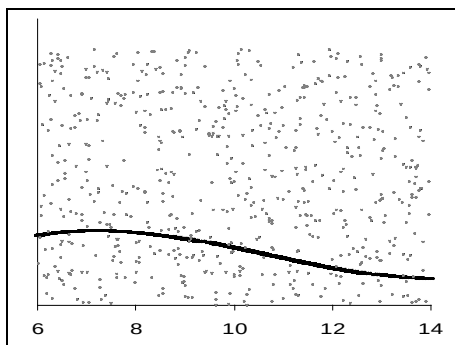
$$200 + (x-5)(x-10)(x-17) \leq 200 + |x-5||x-7||x-17| \leq 200 + 9 \times 7 \times 11 = 893.$$

Agora tiramos pares de números aleatorios no rectángulo $[6,14] \times [0,893]$, e facemos a seguinte consideración:

$$\frac{\text{puntos dentro}(T)}{\text{total de tiradas}} \approx \frac{\int_6^{14} f(x)dx}{\text{área do rectángulo}},$$

facendo 3000 tiradas:

$$\int_6^{14} 200 + (x-5)(x-10)(x-17)dx = \frac{8 \times 893 \times T}{3000}$$



Seguimos o esquema:

A	B	C	D	E
1	=6+8*ALEATORIO()	=200+(B1-5)*(B1-10)*(B1-17)	=893*ALEATORIO()	=SI(D1<C1;1;0)
...	...	...	...	...
3000	=6+8*ALEATORIO()	=200+(B1000-5)*(B1000-10)*(B1000-17)	=893*ALEATORIO()	=SI(D1000<C1000;1;0)
				SUMA(E1:E1000)
				=8*893*E3001/3000

Na columna E compárase a ordenada aleatoria (D) co valor da función no punto de abscisa aleatoria (B), e no caso de que a segunda sexa menor, selecciónase (aparece un 1 na columna E).

1	12,43004088	117,488003	745,174717	0
2	11,69036053	139,952643	414,606525	0
3	8,927318737	234,008279	852,392342	0
...	...	...	...	...
3000	7,628032885	258,421165	617,515513	0
				645
				1535,96

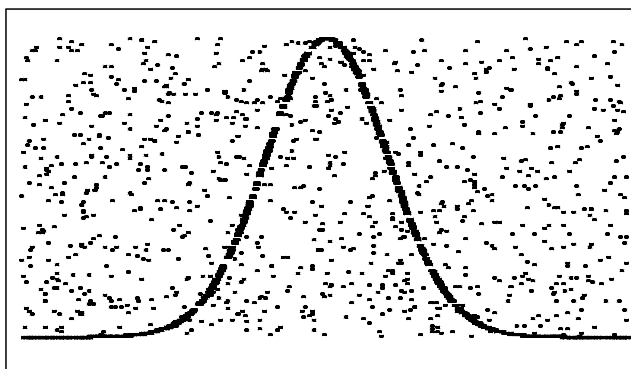
Neste caso poderíamos ter calculado o valor exacto mediante a primitiva, que dá 1514,7; xa que logo o erro cometido foi do 3,6%.

Aproximación da área pechada pola curva N(0,1)

Imos calcular polo método de Montecarlo $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx$.

E doado comprobar que a función ten un máximo en $x=0$ onde toma o valor $\frac{1}{\sqrt{2\pi}}$. Ademais, considerando o comportamento asíntótico da función non teremos en conta os valores de x fóra do intervalo $[-5,5]$, co cal o esquema de cálculo quedaría:

A	B	C	D	E
1	=5+10*ALEATORIO()	=(1/(2*PI())^0,5)*EXP(-(B1^2)/2)	=(1/(2*PI())^0,5)*ALEATORIO()	=SI(D1<C1;1;0)
...	...	...	...	...
4255	=5+10*ALEATORIO()	=(1/(2*PI())^0,5)*EXP(-(B4255^2)/2)	=(1/(2*PI())^0,5)*ALEATORIO()	=SI(D4255<C4255;1;0)
4256				SUMA(E1:E4255)
				=10*(1/(2*PI())^0,5)*E4256/4255



4254	2,74936063	0,009109564	0,29848638	0	
4255	3,58403666	0,000648014	0,09117445	0	
4256				1068	1,00134044

Sabemos que a área pechada pola curva normal entre $-\infty$ e $+\infty$ é 1, polo tanto cometeuse un erro de 0,001.

Exercicios

1. Calcúlese polo método de Montecarlo $\int_0^4 \sqrt{x^3 - 12x + 100} dx$.

Solución: A derivada da función anúlase en -2 e en 2 . Daquela, os posíbeis máximos e mínimos, no intervalo $[0,4]$, estarán en 0 , 2 ou 4 . Como $f(0)=10$, $f(2)=9,17$ e $f(4)=10,77$, o máximo acádase en 4 . Seguindo o esquema xa esposto, ulilizando o rectángulo $[0,4] \times [0,11]$, e tirando 5000 pares aleatorios, obtemos:

4998	1,45766659	9,25230993	10,1959764	0	
4999	0,08272005	9,95027263	1,61937714	1	
5000	0,6464629	9,61834764	3,407133	1	
				4355	38,324

2. Aproxímese $\int_1^3 \frac{1}{x} dx$ tomando $n=10$.

Solución. $h=2/10=0,2$, co cal $\int_1^3 \frac{1}{x} dx \approx 1,10156233$.

3. Calcúlese $\int_1^4 \frac{1}{x^2+1} dx$ para $n=10$ e acóutese o erro.

Solución. $h=0,3$; $\int_1^4 \frac{1}{x^2+1} dx \approx 0,54396172$. $|\epsilon_{10}| \leq \frac{m(4-1)^3}{12 \times 10^2}$, onde m é o máximo da

segunda derivada no intervalo $[1,4]$. Tendo en conta que:

$$|f''(x)| = \left| \frac{6x^2 - 2}{(x^2 + 1)^3} \right| \leq \left| \frac{1}{(x^2 + 1)^3} \right| |6x^2 - 2| \leq \frac{1}{8} \times 62 = 7,75, \text{ obtemos}$$

$$|\epsilon_{10}| \leq \frac{7,75 \times 27}{12 \times 10^2} = 0,174375.$$

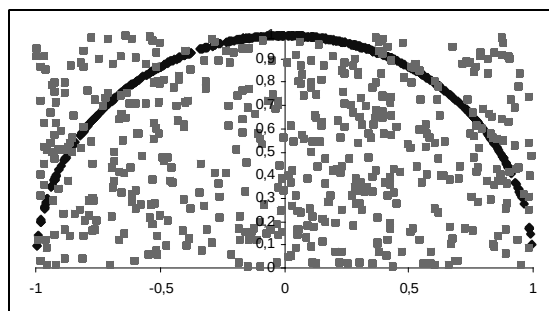
Como podemos comprobar a acoutación do erro e moi conservadora, xa que neste caso calculando a integral exacta: $\arctan(4) - \arctan(1) = 0,5404195$, co cal o erro cometido en realidade foi 0,00354222.

4. Cal debe ser o número de subintervalos, n , para lograr unha aproximación menor que unha dezmilésima de $\int_1^2 \arctan x dx$?

Solució.: Tendo en conta que $|f''(x)| = \left| \frac{-2x}{1+x^2} \right| \leq 2, x \in [1,2]$, de $\frac{2 \times 1}{12 \times n^2} \leq 0,0001$, obtemos $n = 41$, e o valor aproximado da integral é 0,97073903.

5. Tendo en conta que $\int_{-1}^1 \sqrt{1-x^2} dx = \frac{\pi}{2}$ (o integrando é un semicírculo de raio 1 con centro na orixe), estímese π polo método de Montecarlo.

Solución.



A	B	C	D	E
....	...	...	...	
65530	0,87256117	0,48850487	0,892549125	0
65531	-0,30653746	0,9518586	0,422956086	1
65532	0,88130239	0,47255274	0,063049057	1
65533	-0,19769284	0,98026402	0,416115238	1
				51488
			4*51488/65533 = 3,14272199	

6. Aproxímese $\int_{-1}^1 e^{3x^3} dx$ escollendo n para que o erro cometido sexa menor que 0,005.

Solución. A segunda derivada $f''(x) = e^{3x^3} (81x^4 + 18x)$, no intervalo $[-1,1]$, acada o máximo en $x=1$, daquela $m=f''(1) = 1988,47$ co cal $0,005 \geq \frac{1988,47(1-(-1))^3}{12 \times n^2}$, de aí $n = 515$, para o cal o valor da integral obtido é 3,92588092.

7. Aproxímese $\int_2^{2,5} \frac{1}{1-x^2} dx$ escollendo n para que o erro cometido sexa menor que 0,0001.

Solución. Teñamos en conta que no intervalo $[2, 2,5]$

$$|f''(x)| = \left| \frac{2+4x^2}{(1-x^2)^3} \right| \leq \left| \frac{1}{(1-x^2)^3} \right| |2+4x^2| \leq \frac{1}{27} \times 27 = 1. \text{ Co que } 0,0001 \geq \frac{1 \times 0,5^3}{12 \times n^2}, \text{ de}$$

$$\text{onde } n=11 \text{ e } \int_2^{2,5} \frac{1}{1-x^2} dx = -0,09388536.$$

11. RESOLUCIÓN DE ECUACIONES

Hai quen progresa respecto ao que se sabía antes que el, hai quen comenta o que se conserva dos seus predecesores e é difícil, expón claramente os seus textos, clarexa o seu método e fai accesíbeis as fontes, hai quen atopa erros nalgún libro, xunta o disperso e mellora os descubrimentos dos seus colegas sen fachenda cara a eles, nin orgullecéndose do que fixo por el mesmo.

Mohammed ibn-Musa al-Khwarizmi. *O libro da Álgebra.*

A táboa cuneiforme conservada na actualidade no Museo Británico, clasificada como BM13901, contén 24 problemas que dan orixe, na notación alxébrica actual, a ecuacións de segundo grao ou a sistemas de ecuacións, nos que unha delas é de segundo grao; unha clara mostra dos coñecimentos que sobre o particular tiñan os matemáticos babilonios arredor do ano 1800 a. C.



O matemático indio Brahmagupta, 628 d. C., tamén propuxo problemas que se resolvían mediante ecuacións de segundo grao. Xa no século XI o matemático árabe Mohamed ibn Musa al-Khwarizmi en *O libro da Álgebra – al-Mukhtasar fi hisab al-khabr wa-lmuqabala-*, resolve infinidade de problemas, moitos deles xeométricos e outros relacionados con herdanzas, achando as solucións das ecuacións de segundo grao a que dan lugar.

Os resultados aos que xa chegaran tan antigas civilizacións podemos expresalos hoxe, utilizando a notación introducida polo matemático francés Viète –1540-1603-, decindo que toda ecuación de segundo grao $ax^2 + bx + c = 0$, pode resolverse mediante a fórmula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

No ano 1545 Gerolamo Cardano publica en Nuremberg *Ars magna sive de regulis algebraicis, liber unus*, no que se dan a coñecer as solucións das ecuacións alxébricas de terceiro e cuarto grao. Recoñece que a primeira delas aprendéaa de Niccoló Fontana (1499-1557), máis coñecido por Tartaglia, pero este non fora o primeiro en dar coa fórmula, xa que Scipione de Ferro, Bolonia(1465-1526), sen tela publicado déralla a coñecer no leito da morte a un dos seus discípulos. Aínda así,

Cardano é quen de reducir calquera ecuación cúbica a unha sen coeficiente de segundo grao, tendo idea ademais do número de solucións que cabe esperar dela, aínda que ficaba desconcertado fronte as solucións negativas e complexas. Indica tamén Cardano algunhas relacións entre as solucións dunha ecuación e os seus coeficientes, e senta as bases da teoría das funcións simétricas, xermolo da teoría de Galois

A exposición que fai Cardano no su libro para resolver a ecuación de terceiro grao é demasiado longa e pesada. Se alguén ten interese en entrar no detalle dela, adaptada á linguaxe alxébrica actual, non ten máis que consultar un dos múltiples libros ao efecto ou buscar na rede.

A ecuación $z^3 + az^2 + bz + c = 0$, facendo o cambio $z = x - a/3$, convértese en $x^3 + px + q = 0$, o que significa que se resolvemos as deste tipo, podemos resolver calquera ecuación cúbica. A célebre fórmula de Tartaglia para este tipo de ecuacións ven dada por:

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

O discípulo e secretario de Cardano, Ludovico Ferrari (1522-1565) foi quen deu a solución para a ecuación de cuarto grao, descompoñéndoa mediante un enxeño procedemento en dúas de segundo.

Respecto á ecuación de quinto grao, o matemático noruegués Abel (1802, 1829) demostrou a imposibilidade de resolvela utilizando operacións aritméticas e radicais. Uns anos despois, en 1832, Galois (1811, 1832) deu a condición necesaria e suficiente para que a ecuación $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0$ puidese ser resolta mediante radicais, da cal se desprende que non existe fórmula xeral para resolver as ecuación de grao superior a catro.

Se as únicas ecuacións que se puidesen resolver fosen aquelas para as que existen fórmulas, non se lle podería atopar solución a unha grande parte dos problemas formulados pola ciencia e a técnica. Dado que na práctica aparecen ecuacións de grao superior a catro, e mesmo con expresións trigonométricas, logarítmicas ou exponenciais, cómpre abordar métodos que permitan aproximar a solución de ecuacións para as que non exista fórmula. Reparemos, ademais, que aínda no caso en que utilizemos a fórmula para resolver a ecuación, poñamos por caso, de segundo grao, exceptuando os poucos caso nos que a raíz dea exacta, o procedemento non é máis que unha achega á solución cun certo grao de aproximación (décimas, centésimas,...).

Método de iteración

Exemplo 1

Queremos resolver a ecuación $9x - 1 - \cos x = 0$.

Podemos despegar x como función da propia x da seguinte maneira:

$$x = \frac{1 + \cos x}{9}, \text{ ou sexa } x = g(x)$$

Partindo de $x=0$, obtemos a seguinte sucesión de valores

$$x_1 = g(0) = 0,2222222$$

$$x_2 = g(0,2222222) = 0,21949001$$

$$x_3 = g(0,21949001) = 0,21955651$$

$$x_4 = g(0,21955651) = 0,2195549$$

$$x_5 = g(0,2195549) = 0,21955494$$

Vemos así que $g(x_4) - x_4 = 0,00000004$, co que :

$$x_4 \approx \frac{1 + \cos x_4}{9}$$

Podemos considerar entón que $x_4 = 0,2195549$ é unha solución aproximada da ecuación.

O procedemento utilizado para aproximar a solución chámase **método de iteración**.

1. Resólvase: $x^5 - x^3 + x^2 - x + 3 = 0$

Solución. Facemos $x = \sqrt[5]{x^3 - x^2 + x - 3}$, e deseñamos a seguinte folla de cálculo onde na cela A1 introducimos un valor inicial arbitrario.

1	
2	$= (A1^3 - A1^2 + A1 - 3)^{(1/5)}$
...	...

Partindo do valor inicial 5, obtemos

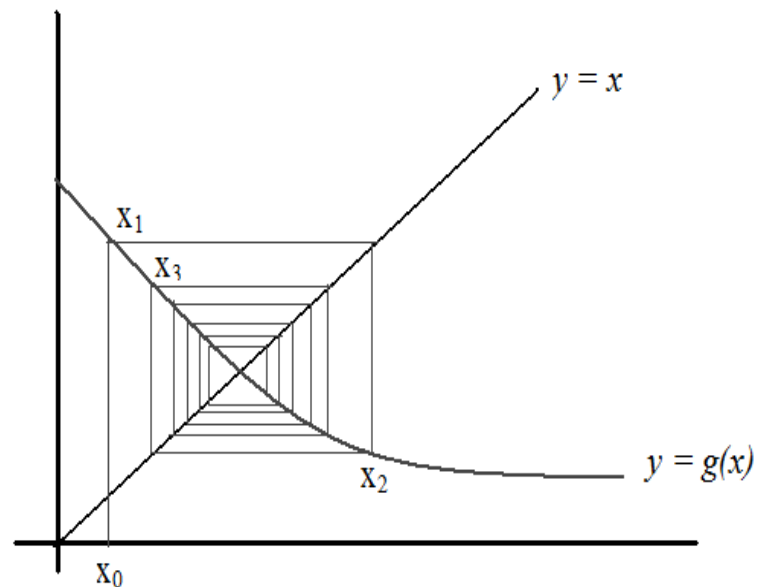
	A
1	5
2	2,52185455
3	1,55869807
4	-0,60922256
	
22	-1,63478757
23	-1,63478757

Daquela, -1,63478757 é unha solución aproximada. Substituíndo, obtense:

$$(-1,63478757)^5 - (-1,63478757)^3 + (-1,63478757)^2 + 1,63478757 + 3 = -0,00162437$$

En xeral, se queremos resolver a ecuación $f(x) = 0$, en primeiro lugar teremos que escribirla da forma $x = g(x)$, e a partir dun valor inicial x_0 obter a sucesión: $x_1 = g(x_0)$, $x_2 = g(x_1)$, ..., $x_n = g(x_{n-1})$. Se despois de realizado certo número de aproximacións chegamos a $|g(x_n) - g(x_{n-1})| < \varepsilon$, para un certo ε prefixado: 0,1, 0,01, 0,001, tomaremos x_n como solución aproximada de $x = g(x)$.

Como pode verse no seguinte gráfico, o método de iteración trata de buscar a intersección de $y = g(x)$ coa bisectriz do primeiro cadrante, o que non sempre está garantido.



Exercicios

2. Resólvase utilizando o método iterativo:

$$\ln x - 0,001x^2 = 0$$

Solución. Facendo $x = g(x) = \sqrt{\frac{\ln x}{0,001}}$, obtense:

	$g(x)$			$g(x)$
x_0	2	x^7		64,5562945
x_1	26,3276885	x^8		64,5564685
x_2	57,189345	x^9		64,5564894
x^3	63,6110651	x^{10}		64,5564919
x^4	64,4421247	x^{11}		64,5564922
x^5	64,5427574	x^{12}		64,5564922
x^6	64,5548442			

3. Resólvase iterativamente $x + \text{sen}x - e^x = -5$.

Solución. Facendo $x = -\text{sen}x + e^x - 5$

x_0	2
x_1	1,47975867
x_2	-1,60397327
x^3	-3,79945438
....	
x_7	-5,61551629
x_7	-5,61551629

4. Pódese resolver polo método de iteración $x - \ln x = 0$?

Solución. Non, nin por ningún outro, porque as gráficas de $y = x$ e $y = \ln x$ non se cortan

Criterio de converxencia

Para garantir unha solución polo método de iteración, cómpre que a sucesión $x_1, x_2, \dots, x_n, \dots$, teña límite, para o que $g(x_n) - g(x_{n-1})$ debe tender a cero. Isto cúmprese con tal de que a función sexa derivábel no intervalo (a, b) ao que pertencen os termos da sucesión, e o valor absoluto da derivada sexa menor que 1 en todo o intervalo. Aplicando reiteradamente o teorema do valor medio, temos:

$$|g(x_2) - g(x_1)| = |g'(c)(x_2 - x_1)| < \alpha |x_2 - x_1|, \text{ con } \alpha < 1$$

$$|g(x_3) - g(x_2)| < \alpha |x_3 - x_2| = \alpha |g(x_2) - g(x_1)| < \alpha^2 |x_2 - x_1|$$

e $|g(x_{n+1}) - g(x_n)| < \alpha^n |x_2 - x_1|$, expresión que tende a cero por ser $\alpha < 1$, co que o límite da sucesión sería unha raíz de $x = g(x)$ ou, o que é o mesmo, de $f(x) = 0$.

Por exemplo, na ecuación $x = \frac{1 + \cos x}{9}$, resolta anteriormente, temos:

$$|g'(x)| = \left| \frac{-\sin x}{9} \right| \leq \frac{1}{9}, \forall x \in \mathbb{R}.$$

Separación de raíces

Se tratamos de resolver a ecuación $f(x) = 0$, onde supoñemos que $f(x)$ é continua polo menos nun intervalo no que pretendemos localizar unha raíz, se atopamos a e b tales que $f(a)f(b) < 0$ (f toma valores con distinto signo en a e en b), entón o teorema de

Bolzano garante que existe $c \in (a,b) / f(c) = 0$. Daquela dicimos que temos unha raíz separada no intervalo $[a,b]$.

Por exemplo, se queremos separar unha raíz da ecuación $x^5+2x^4-x^3+5 = 0$, temos $f(0) = 5$.

Se damos valores negativos, obtemos $f(-2)=13$ e $f(-3) = -49$, entón o intervalo $[-3,-2]$ contén unha raíz.

Método da bisección

Queremos resolver $f(x) = 0$ sabendo que xa temos separada unha raíz en $[a,b]$ de xeito que $f(a)>0$ e $f(b)<0$. Calculamos $f\left(\frac{a+b}{2}\right)$. Se $f\left(\frac{a+b}{2}\right)>0$, entón no intervalo $\left[\frac{a+b}{2},b\right]$ haberá unha solución. Se fose $f\left(\frac{a+b}{2}\right)<0$, a solución estaría en $\left[a,\frac{a+b}{2}\right]$.

Chamando $[a_1,b_1]$ ao intervalo así obtido, procedendo do mesmo xeito que co intervalo $[a,b]$, obteríamos un intervalo $[a_2,b_2]$ no que f tomaría signos distintos nos extremos. Procedendo reiteradamente obteríamos unha sucesión de intervalos $[a_n,b_n]$, tales que a súa lonxitude $b_n - a_n = \frac{b-a}{2^n}$ tería límite cero. Daquela, o punto medio do

intervalo enésimo será unha boa aproximación da solución cun erro menor que $\frac{b-a}{2^{n+1}}$.

O razoamento anterior dótanos dun método para calcular unha solución aproximada, tanto máis efectivo canto mellores ferramentas de cálculo teñamos. Na actualidade coa estandarización da folla de cálculo, para a resolución de ecuacións seguindo este procedemento, mestúranse na práctica separación e aproximación, e en troques de *avanzar* mediante potencias de 2, podemos facelo, poñamos por caso, utilizando as de 10. Vexamos a continuación un caso práctico:

Para resolver a ecuación $x^7-32x^6+121x^5-x+2 = 0$, na columna A tratamos de separar unha raíz nun intervalo de lonxitude 1. Neste caso, por ser o polinomio de grao impar, sabemos que o cambio de signo está garantido. O valor -10 do que partimos é arbitrario.

A	B	C	D	E	F	G	H	I	J
X	F(x)	x	F(x)	x	F(x)	x	F(x)	x	F(x)
-10	-74099988	-1	-153	-0,5	-1,8046875	-0,45	-0,0597209	-0,448	-0,00519067
-9	-38499937	-0,9	-86,9902927	-0,49	-1,3912184	-0,449	-0,0323256	-0,4479	-0,00249142
-8	-18644982	-0,8	-45,8670336	-0,48	-1,01211896	-0,448	-0,00519067	-0,4478	0,00020523
-7	-8269035	-0,7	-21,6483009	-0,47	-0,66520817	-0,447	0,02168572	-0,4477	0,00289931
-6	-3273688	-0,6	-8,3859328	-0,46	-0,34840404	-0,446	0,04830539	-0,4476	0,00559081
-5	-1112493	-0,5	-1,8046875	-0,45	-0,0597209	-0,445	0,07467016	-0,4475	0,00827973
-4	-304122	-0,4	1,0249728	-0,44	0,20273331	-0,444	0,10078182	-0,4474	0,01096607
-3	-59287	-0,3	1,9819859	-0,43	0,44075966	-0,443	0,12664219	-0,4473	0,01364984
-2	-6300	-0,2	2,1591936	-0,42	0,6560709	-0,442	0,15225304	-0,4472	0,01633104
-1	-153	-0,1	2,0987577	-0,41	0,850294	-0,441	0,17761616	-0,4471	0,01900966
0	2	0	2	-0,4	1,0249728	-0,44	0,20273331	-0,447	0,02168572
1	93	0,1	1,9011783	-0,39	1,18157046	-0,439	0,22760626	-0,4469	0,02435921

2	2208	0,2	1,8367104	-0,38	1,32147204	-0,438	0,25223675	-0,4468	0,02703013
3	12635	0,3	1,9713581	-0,37	1,44598692	-0,437	0,27662653	-0,4467	0,02969849
4	41982	0,4	2,7128832	-0,36	1,55635131	-0,436	0,30077733	-0,4466	0,03236429

Localízase unha raíz no intervalo $[-1,0]$. Na columna C queremos achar un intervalo de lonxitude 0,1 que conteña unha raíz. Daquela partimos do valor -1 en C1, facemos $C2=C1+0,1$ e arrastramos, en D avaliamos a función e obtemos que en $[-0,5,-0,4]$ cambia de signo. Nas columnas adxacentes continuamos coa mesma estratexia obtendo intervalos de separación de lonxitude 0,01, 0,001, 0,0001, etc

Neste caso paramos en 0,0001, atopando unha raíz en $[-0,4479, -0,4478]$ co cal

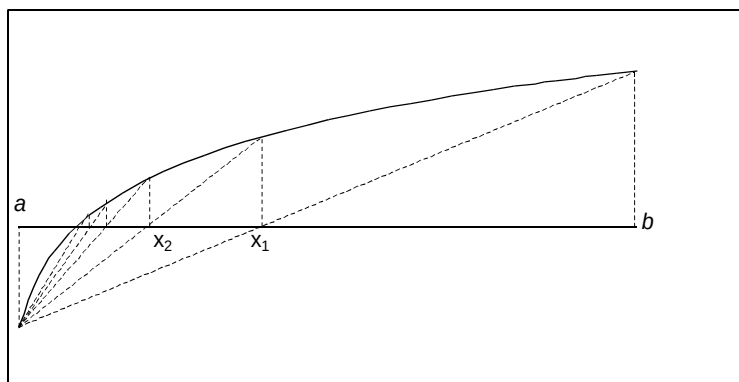
$$x = \frac{-0,4479 + (-0,4478)}{2} = -0,44785 \text{ é unha solución aproximada da ecuación cun}$$

erro menor que 0,00005, a semiampitude do intervalo.

Método da corda

Con este método de resolución trátase de conseguir boas aproximacións en menos pasos que coa bisección – ou mellor dito, *decasección*–.

Supoñamos que xa temos separada unha raíz da ecuación $f(x) = 0$ no intervalo $[a,b]$



Calculamos en primeiro lugar a ecuación da corda que une os puntos $(a, f(a))$ con $(b, f(b))$:

$$y - f(a) = \frac{f(b) - f(a)}{b - a}(x - a)$$

A continuación calculamos a intersección da corda co eixe $y = 0$, obtendo:

$$x_1 = a - \frac{f(a)}{f(b) - f(a)}(b - a)$$

$$\text{Se } f(x_1) > 0, \text{ facemos } x_2 = x_1 - \frac{f(x_1)}{f(b) - f(x_1)}(b - x_1) = g(x_1) \quad (1)$$

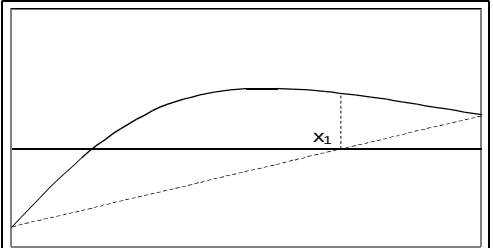
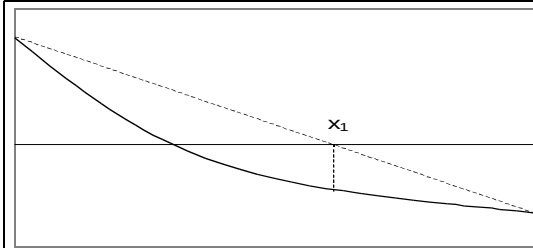
$$\text{En caso de ser } f(x_1) < 0, \text{ sería } x_2 = x_1 - \frac{f(x_1)}{f(x_1) - f(a)}(x_1 - a) = g(x_1) \quad (2)$$

Reiterando o proceso, tal como se indica no gráfico, obtemos unha sucesión:

$$x_3 = g(x_2), x_4 = g(x_3), \dots, x_n = g(x_{n-1}), \dots$$

que se aproxima á solución requirida.

A substitución de x_i nas fórmulas (1) ou (2) dependerá da convexidade de $f(x)$ no intervalo considerado. Se $f''(x)$ ten signo constante en todo o intervalo, pódense dar as seguintes situacións:

 Substitúese x_1 polo extremo en que $f(x)$ toma valor positivo. Cónca $f''(x) < 0$ (1)	 Substitúese x_1 polo extremo en que $f(x)$ toma valor negativo. Convexa: $f''(x) > 0$ (2)
---	---

Exemplo

Resólvese $x^4 - 4x + 2 = 0$

A ecuación ten unha raíz en $[1,2]$ xa que $f(2) = 10$ e $f(1) = -1$.

Transcribindo a fórmula a unha folla de cálculo, temos

	A
1	1
2	=A1-(2-A1)*(A1^4-4*A1+2)/(10-A1^4-4*A1+2)

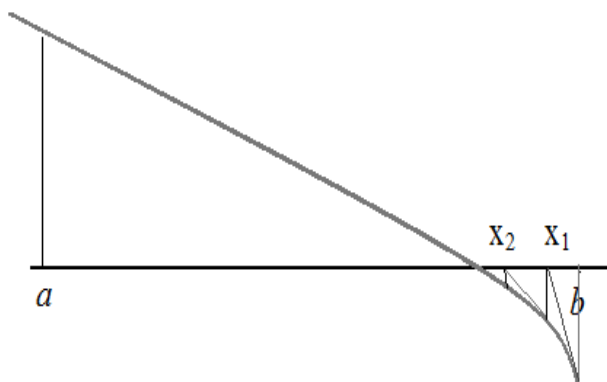
e extendéndoa até que se establece o resultado:

1	1,00000000	10	1,36311046
2	1,14285714	11	1,36311401
3	1,27248908	12	1,36311308
4	1,35189770	13	1,36311332
5	1,36532311	14	1,36311326
6	1,36250517	15	1,36311328
7	1,36327012	16	1,36311327
8	1,36307209	17	1,36311327
9	1,36312404		

Co que $x_{17} = 1,36311327$ é unha solución aproximada cun erro da orde das 100 millonésimas.

Método de Newton

Suposta separada unha raíz de $f(x) = 0$ en $[a, b]$ utilizaremos agora a recta tanxente nun dos extremos, en vez da corda, para acadar a aproximación.



Trazamos a tanxente no punto x_0 , a ou b , tal que $f(x_0)f'(x_0) > 0$, e calculamos o seu punto de corte con $y = 0$.

A ecuación da recta tanxente en $(x_0, f(x_0))$ será:

$y - f(x_0) = f'(x_0)(x - x_0)$, e a súa abscisa de corte con $y = 0$: $x_1 = x_0 - f(x_0)/f'(x_0)$.

Calculando agora a tanxente en x_1 e reiterando o proceso, obtemos a sucesión:

$$x_2 = x_1 - f(x_1)/f'(x_1)$$

....

$$x_n = x_{n-1} - f(x_{n-1})/f'(x_{n-1}),$$

que converxe á solución da ecuación.

Se $f(a)f(b) < 0$ e $f'(x)$ máis $f''(x)$ non se anulan, e teñen signo constante, en $[a, b]$, comezando con x_0 tal que $f(x_0)f'(x_0) > 0$, o algoritmo de Newton tende á raíz da ecuación (sempre que non sexa múltiple). Esta condición non sempre se cumpre, polo que o método se aplica sen ter a seguridade de que converxa. En caso de diverxencia, o que se adoita facer é buscar un punto cercano á solución, utilizando outro método (o da bisección), e logo continúaase co de Newton.

Exemplo

Resólvase $x^4 - 4x + 2 = 0$.

Separamos unha raíz no intervalo $[1, 2]$ e dispoñemos na folla de cálculo:

2
$=A1-(A1^4-4*A1+2)/(4*A1^3-4)$

Obtemos:

1	2
2	1,64285714
3	1,4453446
4	1,37321385
5	1,36329394
6	1,36311333
7	1,36311327
8	1,36311327

Observamos que $x_8 = 1,36311327$ é unha solución aproximado co mesmo orde de erro que a obtida polo método da corda con 17 iteracións.

En xeral, demóstrase que o erro cometido ao tomar x_n como solución exacta é :

$$\varepsilon \leq \frac{|f(x_n)|}{m}, \text{ onde } m = \min_{[a,b]} |f'(x)|.$$

1. Resólvase $x^5 + 2x^4 - x^3 + 5 = 0$ polo método de Newton e calcúlese o erro.

Solución.

Pódese comprobar que o intervalo $[-3,-2]$ contén unha raíz. Para $n = 5$ obtense a solución aproximada $x = -2,52059819$. A derivada no intervalo considerado é decrecente, xa que logo acada o mínimo en $x = -2$ con $f'(-2) = 4$. Daquela:

$$\varepsilon \leq \frac{f(-2,52059819)}{4} = \frac{1,939 \times 10^{-7}}{4} = 4,848 \times 10^{-8}.$$

2. Émpreguese o método de Newton para aproximar as raíces de $x^2 - x - 2 = 0$ e compárense os resultados cos obtidos ao empregar a fórmula de resolución da ecuación de segundo grao.

Solución:

x_0	1	0
x_1	3	-2
x_2	2,2	-1,2
x_3	2,01176471	-1,01176471
x_4	2,00004578	-1,00004578
x_5	2	-1
x_6	2	-1

3. Émpreguese o método de Newton para aproximar as raíces de $x^3 - x - 2 = 0$ e compárense os resultados cos obtidos ao empregar a fórmula de resolución da ecuación de terceiro grao dada na introdución. Calcúlese o erro.

Solución:

x_0	1
x_1	3
x_2	1,70588235
x_3	1,44482919
x_4	1,57985106

....	
x ₄₂	1,52137971
x ₄₃	1,52137971

O resultado que dá a fórmula é: $x = 1,52137971$. Considerado o intervalo $[1,3]$,
 $\varepsilon \leq \frac{f(1,52137971)}{1} = 6,206 \times 10^{-9}$.

4. Resólvanse as seguintes ecuacións utilizando indistintamente o método da corda, da bisección ou de Newton. Acóutese o erro.

a) $x^4 - x - 2 = 0$, b) $x^3 - 3x - 7 = 0$, c) $x^4 - 4x^3 + 2 = 0$, d) $x^4 - x^3 + 2x^2 - 4x + 2 = 0$

Sol: a) 1,35320996, -1, b) 2,42598876, c) 3,96798754, 0,86043959, d) 1, 0,770917.

5. Os babilónicos no ano 1 700 a. C. calculaban $\sqrt{a}$ (iterando) mediante a fórmula

$$x_{n+1} = \frac{1}{2} \left(x_n + \frac{a}{x_n} \right).$$

Xustifiquese a fórmula polo método de Newton e calcúlese $\sqrt{2}$ con catro decimais.

Sol: Trátase de resolver a ecuación $x^2 - a = 0$, daquela, segundo o método de Newton

$$x_{n+1} = x_n - \frac{x_n^2 - a}{2x_n} = \frac{x_n^2 + a}{2x_n} = \frac{1}{2} \left(x_n + \frac{a}{x_n} \right).$$

x ₀	1
x ₁	1,5000
x ₂	1,4167
x ₃	1,4142

6. Resólvese $\arctan x - 2e^{-2x} = 0$.

Sol: Facendo $x = -\ln\left(\frac{\arctan x}{2}\right)$, iterando obtemos:

x ₀	0,5
x ₁	1,46177766
x ₂	0,72275894
x ₃	1,16181142
x ₄	0,8438436
...	...
x ₅₈	0,96061669
x ₅₉	0,96061669

7. Resólvese $\ln x + 3 = e^x$.

Sol: Despexando $x = e^{e^x - 3}$ e iterando:

x_0	0
x_1	0,13533528
x_2	0,15644052
x_3	0,16030792
x_4	0,16103594
...	...
x_{13}	0,16120584
x_{14}	0,16120584

8. Resólvese $\sqrt{x^2 + 1} - x = \tan x$.

Sol:

x_0	0
x_1	1
x_2	0,44259228
x_3	0,72193671
x_4	0,64069236
...	...
x_{13}	0,65960623
x_{14}	0,65960623

9. Un investigador ten dous cultivos de bacterias de distinto tipo que pretende mesturar en partes iguais para probar a hipótese de aniquilación mutua. O número de bacterias, en millóns, do primeiro cultivo vai diminuindo en función do tempo, expresado en días, segundo a función $f(x) = 3e^{-0,2t}$, mentres que o segundo vai aumentando segundo a función $f(x) = 0,05e^{0,5t}$. Canto tempo terá que pasar para comprobar a hipótese?

(Indicación: en $0,05e^{0,5t} = 3e^{-0,2t}$, despexar t en función de t)

t_0	2
t_1	7,38868912
t_2	5,23321347
t_3	6,09540373
t_4	5,75052763
...	...
t_{27}	5,84906366
x_{28}	5,84906366

10. Calcúlese o prezo de equilibrio (corte da curva de oferta coa de demanda) e a cantidade correspondente de unidades ofrecidas e demandadas se estas se rexen respectivamente por $S(p) = p^4 - 4p - 100$ e $D(p) = 1200 - p^2$.

Solución: $p = 5,99€$.

11. Un promotor de vivendas despois de facer moitas contas pretende vender $10 + \frac{p^3}{1000}$ pisos, onde o prezo, p , está expresado en miles de euros. A demanda é de $90037 - p^2$. A que prezo a oferta igualará a demanda? Cantos pisos se venderán a ese prezo?
 Solución: $p = 300000\text{€}$. Venderá 37 pisos.

Resolución de sistemas

Moitos problemas da matemática aplicada e xa que logo da técnica e das ciencias, incluídas as sociais, requiren da resolución de sistemas de ecuacións. Estes sistemas en ocasións son lineares e teñen un número razoábel de ecuacións e de incógnitas, podendo ser resoltos utilizando métodos estándar como o de Newton-Gauss ou a regra de Cramer, ou mesmo calculando directamente a matriz inversa dos coeficientes e multiplicala pola matriz columna de termos independentes, se a solución é única. Mais, ás veces os sistemas que cómpre resolver non son lineares, ou mesmo aparecen neles expresións trigonométricas, logarítmicas ou exponenciais, polo que hai que acudir a outros métodos. Un deles é o de iteración, unha xeralización do de resolución de ecuacións visto anteriormente, do cal no que segue faremos unha pequena introdución.

Método de iteración

Exemplo

Dado o sistema $\begin{cases} 3x + 2y = 7 \\ -x + y = 1 \end{cases}$, é inmediato comprobar que $x=1, y=2$ é unha solución.

Agora ben, o sistema anterior ten as mesmas solucións que $\begin{cases} x = \frac{7-2y}{3} \\ y = 1+x \end{cases}$

ou sexa, pode escribirse da forma: $\begin{cases} x = g(y) \\ y = f(x) \end{cases}$

Partindo de valores iniciais arbitrarios de x e y , poñamos por caso $x=0, y=0$, obtemos: $\begin{cases} x_1 = g(0) = 2,33333333 \\ y = f(0) = 1 \end{cases}$

$$\begin{cases} x_2 = g(2,33333333) = 1,66666667 \\ y_2 = f(1) = 3,33333333 \end{cases}$$

Para continuar coas iteracións utilizamos a folla de cálculo segundo o esquema:

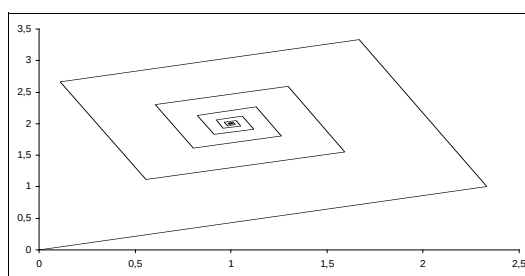
	A	B
1	0	0
2	$=(7-2*B1)/3$	$=1+A1$

Sinalando o rango A2:B2 e arrastrando obtemos:

	A	B
1	0	0
2	2,33333333	1
3	1,66666667	3,33333333
...		
96	1,02312204	1,98265847
97	1	1,99999999
98	1	2
99	1	2

Neste caso sabíamos a solución exacta, mais no caso xeral consideraremos que a aproximación é boa cando a diferenza entre os termos consecutivos acade o erro fixado de antemán.

No gráfico seguinte, seleccionando as columnas A e B, e a modalidade dispersión, podemos observar a converxencia cara á solución a partir das condicións iniciais (0,0).



Exemplo

Consideremos o sistema
$$\begin{cases} \text{sen } x + \text{sen } y = 0,89 \\ \ln x + 5y = 2 \end{cases}$$

Na folla de cálculo, temos:

	A	B
1	0,7	0,8
2	$=\text{ASENO}(0,89-\text{SENO}(B1))$	$=(2-\text{LN}(A1))/5$

e arrastrando até a fila 76:

	A	B
1	0,7	0,8
2	0,173513256	0,47133499
3	0,451064579	0,75030026
...		...
74	0,29193841	0,64624248

75	0,291938411	0,64624248
76	0,291938411	0,64624248

De onde concluímos que $x = 0,291938411$ e $y = 0,64624248$ son solucións aproximadas cun erro da orde das dez mil millonésimas.

Exemplo

Consideremos agora o sistema $\begin{cases} 2x + 3y = 7 \\ x - y = 1 \end{cases}$ do que sabemos que a solución é $x=2$ e $y=1$, porén, se intentamos resolvelo como os anteriores, mesmo partindo de condicións

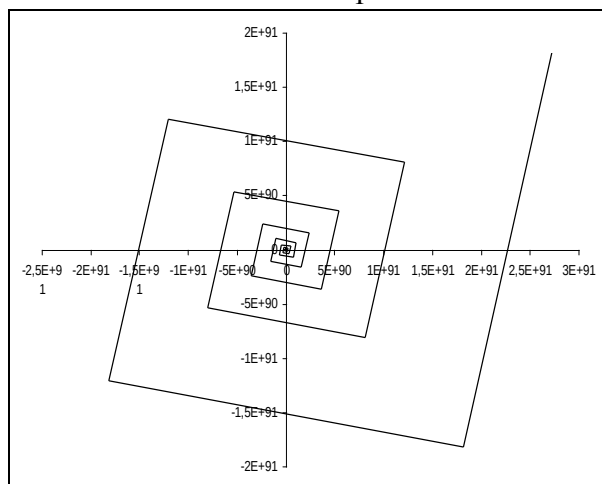
iniciais moi próximas á solución, vexamos o que acontece:

	A	B
1	2,00000001	0,99999999
2	$=(7-3*B1)/2$	$=A1-1$

	A	B
1	2,00000001	0,99999999
2	2,000000015	1,00000001
3	1,999999985	1,000000015
4	1,999999978	0,999999985
5	2,000000023	0,999999978
6	2,000000034	1,000000023
....	...	...
368	-2,5165E+25	-1,6776E+25
369	2,5165E+25	-2,5165E+25
370	3,7747E+25	2,5165E+25
371	-3,7747E+25	3,7747E+25

A sucesión que obtemos non converge á solución.

Graficamente, iterando até a fila 3500, observamos a diverxencia malia partir dunhas condicións iniciais moi próximas á solución.



Os criterios para determinar que tipo de sistemas se poden resolver por este método non son doados de xustificar. Aínda así imos ver unha condición suficiente que garanta a solución para certo tipo de sistemas lineares aplicando este método.

Consideremos o sistema de n ecuacións e n incógnitas:

$$\begin{cases} x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + x_2 + \dots + a_{2n}x_n = b_2 \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + x_n = b_n \end{cases}$$

onde os coeficientes da diagonal son todos 1. Se ademais $\max_{i \dots n} \sum_{i \neq j} |a_{ij}| < 1$, é dicir, a suma dos valores absolutos dos coeficientes de cada fila, excluindo o diagonal, é menor que 1, entón o sistema pode resolverse polo método de iteración.

Imos resolver:

$$\begin{cases} 100x_1 + 30x_2 + 25x_3 - 20x_4 = 300 \\ -x_1 + 10x_2 - 2x_3 + 5x_4 = -20 \\ 10x_1 - 2x_2 + 20x_3 + 6x_4 = 140 \\ -8x_1 - 20x_2 + 10x_3 + 40x_4 = 200 \end{cases}$$

Dividindo a primeira ecuación por 100, a segunda por 10, a terceira por 20 e a cuarta por 40, obtemos o sistema equivalente:

$$\begin{cases} x_1 + 0,3x_2 + 0,25x_3 - 0,2x_4 = 3 \\ -0,1x_1 + x_2 - 0,2x_3 + 0,5x_4 = -2 \\ 0,5x_1 - 0,1x_2 + x_3 + 0,3x_4 = 7 \\ -0,2x_1 - 0,5x_2 + 0,25x_3 + x_4 = 5 \end{cases}$$

Fixémonos en que os coeficientes da diagonal son iguais a un, e ademais as sumas

$$0,3 + 0,25 + 0,2 = 0,75$$

$$0,1 + 0,2 + 0,5 = 0,8$$

$$0,5 + 0,1 + 0,3 = 0,9$$

$$0,2 + 0,5 + 0,25 = 0,95$$

son todas menores que 1.

Mediante o esquema de folla de cálculo:

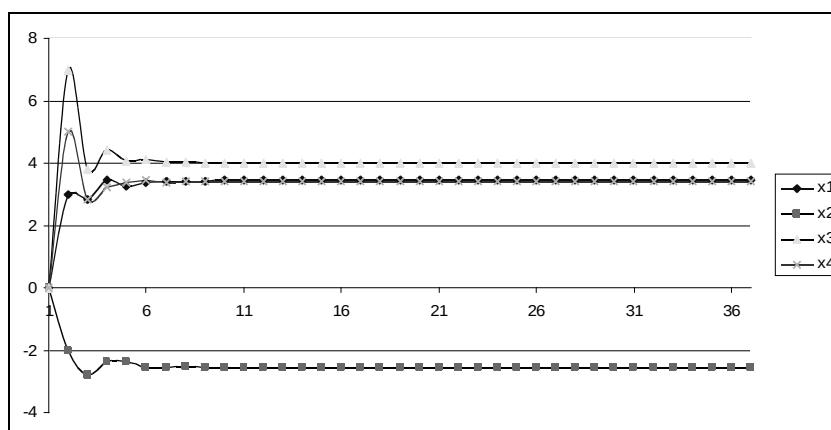
	A	B	C	D
1	0	0	0	0

2	$=3-0,3*B1-0,25*C1+0,2*D1$	$=-2+0,1*A1+0,2*C1-0,5*D1$	$=7-0,5*A1+0,1*B1-0,3*D1$	$=5+0,2*A1+0,5*B1-0,25*C1$
---	----------------------------	----------------------------	---------------------------	----------------------------

extendendo até a fila 36, obtemos:

	A	B	C	D
1	0	0	0	0
2	3	-2	7	5
3	2,85	-2,8	3,8	2,85
...				
34	3,45200182	-2,56141947	3,9945405	3,4110555
35	3,45200182	-2,56141947	3,99454049	3,4110555
36	3,45200182	-2,56141947	3,99454049	3,4110555

cuxa rápida converxencia podemos observar no seguinte gráfico:



Co cal $x_1 = 3,45200182$, $x_2 = -2,56141947$, $x_3 = 3,99454049$, $x_4 = 3,4110555$ é unha solución aproximada, que podemos comparar coa obtida utilizando directamente as funcións MMULT e MINVERSA da folia de cálculo:

$x_1 =$	$x_2 =$	$x_3 =$	$x_4 =$
3,45200182	-2,56141947	3,99454049	3,41105551

Exercicios

16. Resólvase se é posíbel polo método iterativo:

$$\begin{cases} 24x - 12y + 3z - 33 = 0 \\ 14x + 32y - 4z + 23 = 0 \\ 4x + 2y + 9z - 12 = 0 \end{cases}$$

Sol: $x = 0,77221439$, $y = -0,90761636$, $z = 1,19181946$

17. Resólvese se é posíbel polo método iterativo:

18.

$$\begin{cases} 4x + 12y + z = 7 \\ 7x + y - 14z = 8 \\ 4x + 2y + 3z = 4 \end{cases}$$

Solución. O sistema non se pode resolver iterativamente. Utilizando cálculo matricial, na folla de cálculo obtemos $x = 0,92531646$, $y = 0,28227848$, $z = -0,08860759$.

Se tomamos estes valores como condicións iniciais no procedemento iterativo o sistema desestabilízase a partir da undécima iteración (o efecto bolboreta aparece en calquera sitio).

Exercicios

18. Resólvanse: a) $x = \frac{1}{(x+2)^2}$, b) $x = \frac{1}{\text{sen} x}$

Solución: a) 0,20556943 b) 1,11415714

19. Resólvanse: a) $x = 3 + \sqrt[3]{\frac{x-1}{x+1}}$, b) $x = \sqrt{\ln(x+1)}$

Solución: a) 3,83707157, b) 0,74688174

20. Resólvanse: a) $x = 3 - \sqrt{x}$, b) $3x^4 = e^x + 1$

Solución: a) 1,69722436, b) 1,068553384.

21. Resólvanse: a) $(4-x)\cos x = \text{sen} x$, b) $x = \frac{e^{-x}}{8}$, c) $\cos x = x^2$

a) A ecuación é equivalente a $\tan x = 4 - x$ ou $x = \arctan(4 - x)$.

Sol: $x = -1,387262203$. b) 0,111780109. c) 0,82413231.

22. Resólvanse utilizando indistintamente o método de Newton ou o da corda.

a) $x^3 - 5x + 1 = 0$ b) $x^3 - 9x^2 + 20x - 11 = 0$

Solución: a) 0,201639676, b) 0,9274.

23. Resólvanse utilizando indistintamente o método de Newton ou o da corda.

a) $x^3 - 3x^2 + 11 = 0$ b) $x^5 + 3x + 1 = 0$

c) $x^2 - 3\ln x - 2 = 0$ d) $\cos x + x = 1$

Solución: a) -1,55414922, b) -0,33198903, c) 2,03124131, d) $x=0$, evidentemente.

24. Resólvese se é posíbel polo método iterativo:

$$\begin{cases} 10x - 2y + z = 8,98 \\ 3x - 10y + 1,5z = -13,8 \\ 25x - 40y - 100z = 367,7 \end{cases}$$

Solución: $x=1,53204376$, $y=1,26935356$, $z = -3,80173048$.

25. Resólvese polo método iterativo:

$$\text{a) } \begin{cases} 0,25 \cos(x + y) + 0,336 = x \\ y = -0,25 \cos(x - y) + 0,362 \end{cases}$$

$$\text{b) } \begin{cases} x = \sqrt{2x + y - 0,710} \\ y = \sqrt{x - y} + 1 \end{cases}$$

Solución: a) $x=0,53278832$, $y=0,13182832$.

b) $x = 0,53278832$, $y = 0,13182832$.

26. Áchese o punto de equilibrio e a cantidade correspondente de unidades ofrecidas e demandadas dun artigo cuxas curvas de oferta e demanda son respectivamente:

$$s(p) = p^4 + 3p^3 - 460 \text{ e } d(p) = 3200 - p^3$$

Solución: $p = 6,99335885$.

27. Resólvase se é posíbel polo método iterativo:

$$\begin{cases} 12x - 4y + z - 3t = 23 \\ 2x + 8y + 2z + 2t = 12 \\ 3x + 4y + 15z - 47 = -10 \\ -3x + 5y + 6z + 30t = 40 \end{cases}$$

Sol: $x=2,62969563$, $y=0,6551142$, $z=-0,92158702$, $t=1,6714346$.

$$28. \text{ Resólvase: } \begin{cases} x - \ln y = -3 \\ \cos x + y = 1 \end{cases}$$

Solución: $x = -2,43453069$, $y = 1,76027371$.

$$29. \text{ Resólvanse: a) } \sqrt{\sin x} + \cos x = \frac{\sqrt{2}}{2} \quad \text{b) } \begin{cases} 0,5 + 16^{\sin x} = y \\ 0,7y - x = 2 \end{cases}$$

Solución: a) $1,84795693$. b) $x = -1,60617403$, $y = 0,56260852$.

$$30. \text{ Resólvase } \begin{cases} e^{\frac{-x^2}{2}} - 2y = 0 \\ y^2 + x = 1 \end{cases}$$

Solución: $x = 0,88596192$, $y = 0,33769524$.

$$31. \text{ Resólvase } \begin{cases} x + 0,8 = \sqrt{x + 2y} \\ y - 1 = \sqrt{y - x} \end{cases}$$

Solución: $x = 1,24939015$, $y = 1,47530492$.