MATERIA
Bioestatística

TITULACIÓN
Grao en Medicina

unidade didáctica 4

Ciencias Experimentais

Variables aleatorias continuas

María Isabel Borrajo
Mercedes Conde Amboage

Área de Estatística e Investigación Operativa
Departamento de Estatística, Análise Matemática e Optimización
Facultade de Matemáticas

unidadesdidácticas
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
SERVIZO DE NORMALIZACIÓN LINGÜÍSTICA

Deseño e maquetación
J. M. Gairí

Edita
Edicións USC
usc.gal/publicacions

DOI
https://dx.doi.org/10.15304/9788419679383

MATERIA: Bioestatística
TITULACIÓN: Grao en Medicina
PROGRAMA XERAL DO CURSO

BLOQUE I: ESTATÍSTICA DESCRITIVA

Unidade 1. Estatística descritiva

A utilidade da estatística nas ciencias da saúde
Conceptos básicos: poboación, variable aleatoria e mostra
Tipos de variables aleatorias
Táboas de frecuencias
Medidas características
Representacións gráficas

BLOQUE II: TEORÍA DA PROBABILIDADE

Unidade 2. Fundamentos da teoría da probabilidade

Experimento aleatorio, sucesos e espazo mostral
Definición de probabilidade
Probabilidade condicionada. Independencia de sucesos
Resultados notables: regra do produto, probabilidades totais e teorema de Bayes
Prevalencia e incidencia
Probas diagnósticas: sensibilidade, especificidade e valores preditivos
Medidas de efecto: risco relativo e odds-ratio

Unidade 3. Variables aleatorias discretas

Definición de variable aleatoria discreta
Función de masa de probabilidade e función de distribución
Medidas características
Modelos de distribucións discretas

Unidade 4. Variables aleatorias continuas

Definición de variable aleatoria continua
Función de densidade e función de distribución
Medidas características
Modelos de distribucións continuas. A distribución Normal
Aproximación entre distribucións

BLOQUE III: INFERENCIA ESTATÍSTICA

Unidade 5. Introdución á inferencia estatística

Utilidade da inferencia estatística
Conceptos básicos: parámetro, estatístico e estimador
Distribución na mostraxe
Distribucións asociadas á Normal: T de Student e χ- cadrado

Unidade 6. Estimación de parámetros

Definición e propiedades dos estimadores puntuais
Estimación puntual da media, varianza e proporción
Estimación por intervalos de confianza

Unidade 7. Contrastes de hipóteses

Utilidade dos contrastes de hipóteses
Procedemento de contraste
Contrastes sobre unha poboación
Contrastes para comparación de poboacións

Unidade 8. Contrastes para datos categóricos

Introdución
Táboas de continxencia
Contrastes χ- cadrado

BLOQUE IV: MODELOS DE REGRESIÓN

Unidade 9. Modelo de regresión lineal simple

Introdución aos modelos de regresión
Conceptos previos: vector de medias, coeficiente de correlación e diagrama de dispersión
Modelo de regresión lineal simple
Método de mínimos cadrados
Coeficiente de determinación

Unidade 10. Análise da varianza

Modelo e hipóteses
Táboa de descomposición da variabilidade
Contraste de igualdade de medias
Comparacións múltiples

ÍNDICE

PRESENTACIÓN

A bioestatística é segundo o dicionario da RAE (Real Academia Española) a *ciencia que aplica el análisis estadístico a los problemas y objetos de estudio de la biología*. Partindo desta definición, a materia de Bioestatística (con pequenas variacións na súa nomenclatura) é unha materia que se imparte no primeiro curso de todos os graos da rama biosanitaria, tales como Medicina, Odontoloxía, Enfermaría, Bioloxía, Biotecnoloxía, Farmacia, Óptica, Psicoloxía ou Veterinaria.

Esta unidade didáctica caracterízase pola súa transversalidade no senso de que pode ser empregada polo alumnado e profesorado de moi diversas titulacións facendo mínimas adaptacións da mesma. Na Táboa 1 detállase a listaxe completa de titulacións e materias nas que se podería empregar esta unidade.

Táboa 1: Relación de titulacións e materias da rama biosanitaria, impartidas nalgún dos campus da USC, nas que se pode empregar esta unidade didáctica.

Titulación	Materia
Grao en Bioloxía	Bioestatística
Grao en Bioquímica	Bioestatística
Grao en Biotecnoloxía	Bioestatística
Dobre Grao en Farmacia e Óptica e Optometría	Matemáticas e estatística I
Grao en Enfermaría	Estatística
Grao en Farmacia	Matemáticas e estatística I
Grao en Nutrición Humana e Dietética	Bioestatística
Grao en Odontoloxía	Bioestatística e introdución á investigación
Grao en Óptica e Optometría	Bioestatística
Grao en Psicoloxía	Análise de datos en psicoloxía
Grao en Veterinaria	Bioestatística

Nestas materias realízase un percorrido polas ferramentas elementais da probabilidade e da estatística focalizándose nas aplicacións biosanitarias en xeral, e no ámbito máis específico de cada titulación en particular. No seu desenvolvemento hai que ter en conta que, por estar maioritariamente asentada no primeiro curso, é necesariamente unha materia de nivelación dos coñecementos do alumnado con procedencias diversas e, á súa vez, debe proporcionar unha competencia estatística sólida, pois en moitas das titulacións é o único contacto coa estatística que o alumnado vai ter ao longo deses anos de formación na correspondente facultade ou escola.

A programación docente proposta, da que esta unidade didáctica forma parte, está composta por catro bloques temáticos agrupando diversas unidades segundo se detalla a continuación:

— Bloque I: Estatística descritiva (1 UD);
— Bloque II: Teoría da probabilidade (3 UD);
— Bloque III: Inferencia estatística (4 UD);

— Bloque IV: Modelos de regresión (2 UD).

Esta proposta de programación docente cobre os contidos que se detallan na memoria de verificación de título do Grao en Medicina (pode consultarse na seguinte ligazón: https://pro-assets-usc.azureedge.net/cdn/ff/b4ycnzhOyuydvYl2rdg777pZ-5g_Z4rxapDN7SRNyAw/1632393404/public/plan/2021-09/Medicina_modif_2014.pdf), que son dos máis extensos de todas as titulacións biosanitarias da Universidade de Santiago de Compostela arriba mencionadas. Deste xeito, podemos cubrir os programas de todas as materias, se ben é certo que algunhas unidades non se empregarían en todas as titulacións.

Dentro da programación, con esta unidade didáctica en particular, pretendemos fornecer o alumnado e o profesorado, das materias de primeiro curso das titulacións da rama biosanitaria, cun material tanto teórico como práctico que cubra os contidos sobre variables aleatorias continuas. Un material que sirva de referencia común, e que permita unificar tanto os conceptos básicos requiridos, como a notación e nomenclatura empregadas.

A presente unidade didáctica é a terceira das unidades que conforman o Bloque II. Teoría da Probabilidade. A través do seu desenvolvemento preténdese transmitir ao alumnado conceptos básicos como o de variable aleatoria continua, función de densidade, función de distribución, así como introducir as principais medidas características e o seu cálculo. Tamén se verán os principais modelos de distribucións continuas: uniforme, normal, exponencial e gamma, e as relacións de aproximación entre algún deles e os vistos na unidade didáctica previa.

COMPETENCIAS E OBXECTIVOS

Segundo a memoria de verificación do título de Grao en Medicina pola USC, a comprensión das ciencias básicas, amosar unha actitude ética, o establecemento dunha boa comunicación interpersoal e a adaptación a circunstancias cambiantes son obxectivos fundamentais a acadar polo alumnado durante esta etapa de formación.

As competencias xerais e específicas vencelladas coa materia de Bioestatística que se recollen na memoria de verificación de título veñen detalladas a continuación. Por coherencia coa memoria de verificación de título, decidíronse manter os seu códigos de nomenclatura. Debemos salientar que nesta listaxe tamén se inclúen as competencias CG32, CE31 e CE37, que a pesar de non vincularse coa materia de Bioestatística na memoria de verificación do título, si se traballan e se desenvolven ao longo da mesma. Polo tanto, as competencias que se traballarán ao longo da materia son:

CG28: obter e empregar datos epidemiolóxicos e valorar tendencias e riscos para a toma de decisións sobre saúde;

CG31: coñecer, valorar criticamente e saber empregar fontes de información clínica e biomédica para obter, organizar, interpretar e comunicar información científica e sanitaria,

CG32: saber empregar as tecnoloxías da información e a comunicación en actividades clínicas, terapéutica, preventiva e investigadora.
CG33: manter e utilizar rexistros con información do paciente para unha posterior análise, preservando a confidencialidade dos datos;
CG34: ter, na actividade profesional, un punto de vista crítico e creativo, con escepticismo construtivo e orientado á investigación;
CG35: comprender a importancia e as limitacións do pensamento científico no estudo, prevención e manexo de enfermidades;
CG36: ser capaz de formular hipóteses, recoller e valorar de xeito crítico información para resolver problemas, seguindo o método científico;
CG37: adquirir formación básica para a actividade investigadora;
CE31: coñecer, valorar criticamente e saber utilizar tecnoloxías e fontes de información clínica e biomédica, para obter, organizar, interpretar e comunicar información clínica, científica e de saúde;
CE32: coñecer os conceptos básicos da bioestatística e a súa aplicación ás ciencias médicas;
CE33: ser capaz de deseñar e realizar estudos estatísticos sinxelos mediante programas informáticos e interpretar os resultados;
CE34: comprender e interpretar datos estatísticos da literatura médica;
CE37: manexar con autonomía un ordenador persoal e as aplicacións informáticas máis comúns no campo da biomedicina.

Seguindo as competencias arriba indicadas, na guía docente da materia establécese que a finalidade desta é familiarizar ao alumnado cos conceptos e técnicas básicas da estatística descritiva, da teoría da probabilidade e da inferencia estatística. Amais, derívanse os obxectivos xerais da materia:

OX1: coñecer a linguaxe estatística básica: proporcionar ao estudantado os coñecementos teóricos básicos que lle permita comprender os distintos aspectos estatísticos e probabilistas implicados na investigación médico/sanitaria;
OX2: coñecer e aplicar algúns métodos estatísticos básicos para representar e analizar conxuntos de datos simples e poder obter conclusións destas análises;
OX3: coñecer, expresar e interpretar correctamente os niveis de precisión, confianza e de erro nas conclusións dun estudo estatístico.

Nesta unidade didáctica traballaremos o OX1, e del derívanse os seguintes obxectivos específicos:

OE1: coñecer o concepto de variable aleatoria continua;
OE2: distinguir entre variable aleatoria discreta e continua;
OE3: coñecer e saber representar as principais funcións que caracterizan as variables aleatorias continuas;
OE4: coñecer e saber obter as principais medidas características asociadas a unha variable aleatoria continua;
OE5: coñecer modelos notables de distribucións continuas;

OE6: saber identificar situacións reais nas que aparecen os modelos notables de distribucións continuas;
OE7: saber aplicar correctamente as posibles aproximacións entre distribucións;
OE8: desenvolver unha actitude crítica e responsable coa toma de decisións baseada en resultados estatísticos;
OE9: valorar a aprendizaxe autónoma;
OE10: amosar iniciativa na resolución de problemas;
OE11: amosar interese por adquirir novos coñecementos estatísticos.

A relación entre estes obxectivos específicos e as correspondentes competencias vén detallada na Táboa 2.

Táboa 2: Relación entre os obxectivos específicos desta unidade didáctica e as competencias vencelladas á materia segundo a memoria de verificación do Grao en Medicina.

Obxectivos	Competencias
OE1	CG31, CG37, CE32, CE34
OE2	CG31, CG35, CG37, CE32, CE34
OE3	CG37, CE32
OE4	CG37, CE32
OE5	CG37, CE32
OE6	CG31, CG35, CG37, CE32, CE34
OE7	CG35, CG37, CE32, CE37
OE8	CG34, CG35, CE31, CE32
OE9	CG37
OE10	CG37, CE31
OE11	CG37, CE31, CE34

Remarcar que os obxectivos OE1 a OE5 están ligados a competencias técnicas (aquelas que implican dominio como experta/o das tarefas e contidos do seu ámbito de traballo, así como os coñecementos e habilidades necesarios para o seu desempeño); o OE6 e OE7 a competencias metodolóxicas (implican a capacidade de reacción ao aplicar o procedemento adecuado ás tarefas encomendadas e ás irregularidades que xurdan atopando de xeito independente vías de solución, que ademais logo se transfiren a outros eidos da formación ou traballo); e os OE8 a OE11 a competencias actitudinais (vencelladas á colaboración interpersoal de forma comunicativa e construtiva, así como á participación na organización e á capacidade de decidir e asumir responsabilidades).

PRINCIPIOS METODOLÓXICOS

Neste apartado especifícase a liña metodolóxica xeral e a dinámica de traballo que desexamos instaurar neste Bloque II. Teoría da Probabilidade. As actividades propostas clasifícanse en dous grandes grupos atendendo ao carácter presencial ou non das mesmas. Na Táboa 3 pódese ver en detalle a dedicación en horas presenciais e non presenciais que se espera por parte do alumnado para a superación desta unidade didáctica.

As actividades formativas con presenza do profesorado que se desenvolverán ao longo deste Bloque II son:

— **Sesións expositivas.** Para a exposición dos contidos teóricos o profesorado fará uso de presentacións en ordenador que o alumnado terá á súa disposición a través do Campus Virtual da USC. Estas presentacións son esquemas para que o alumnado poida seguir as sesións, pero é recomendable que cada alumna/o tome as súas propias notas e complete os contidos que se expoñen. As sesións expositivas desenvolveranse intercalando problemas prácticos e aplicacións entre as explicacións teóricas. Preténdese enriquecer a participación do alumnado na aula, por medio dun exercicio diario ao finalizar cada sesión que se poida resolver aplicando os coñecementos adquiridos durante a mesma.

  Xunto coa presentación de cada sesión, facilitaráselle ao alumnado ao concluír cada tema un material máis desenvolvido. Preténdese que o alumnado compare o material que recopilou durante as sesións con estes documentos, para repasar os conceptos estudados e completar os seus apuntamentos. Ademais, cada tema inclúe o seu correspondente boletín de exercicios nos que o alumnado poderá aplicar as ideas desenvolvidas nas sesións expositivas.

— **Sesións interactivas de seminario.** Estas sesións dedicaranse á resolución por parte do alumnado dos exercicios propostos nos boletíns temáticos. En ocasións nas que sexa necesario clarificar ou incidir en algún dos contidos, os problemas serán resoltos polo profesorado. Preténdese que a/o estudante adquira seguridade na exposición oral de coñecementos e valore o traballo en equipo.

— **Sesións interactivas de prácticas de ordenador.** Estas sesións serán realizadas nas aulas de informática do centro e introducirán o alumnado no manexo do *software* estadístico de libre distribución R. Na sesión, o alumnado disporá dun guión da práctica que servirá de fío condutor á vez que incluirá exercicios a realizar por parte da/o alumna/o de maneira autónoma.

Como actividades non presenciais asociadas a esta unidade considéranse a resolución de exercicios, as buscas bibliográficas ou a realización das lecturas recomendadas.

Táboa 3: Dedicación (en horas) do alumnado a esta unidade didáctica segundo o tipo de actividade.

Actividades presenciais	**Horas**	**Actividades non presenciais**	**Horas**
Sesións expositivas	3	Estudo individual	6
Seminarios	1	Resolución de exercicios	1.5
Prácticas de ordenador	1	Comprensión de código	1.5
Total horas presenciais	**5**	**Total horas non presenciais**	**9**

CONTIDOS

A definición e a estruturación dos contidos da presente unidade didáctica farase en primeiro lugar respectando as indicacións que se especifican na ficha da materia incluída na memoria de verificación do título. Ademais, pretende ser fiel a dous principios chave:

— non hipertrofiar os contidos disciplinarios, configurando un programa acorde co peso curricular da materia, proporcionando ao alumnado as ferramentas necesarias para o correcto desenvolvemento doutras materias da titulación e da súa futura actividade profesional;

— romper o que se coñece como «encefalograma plano», é dicir, dar a todos os contidos o mesmo nivel de importancia. Debemos ser capaces de amosar ao alumnado as chaves da nosa disciplina mediante contido esencial, necesario e de extensión.

Os contidos desta unidade didáctica atópanse distribuídos en cinco seccións, que deben de ser tratadas de xeito secuencial no tempo e que detallamos a continuación.

1. Definición de variable aleatoria continua

Comezamos esta sección recordando que, en xeral, unha variable aleatoria X é unha correspondencia que asocia un número a cada elemento do espazo mostral, vencellado este último a un experimento aleatorio.

Na unidade didáctica anterior, centrámonos en variables aleatorias discretas, é dicir, aquelas que toman unha cantidade finita ou infinita numerable de valores. Pola contra, cando unha variable aleatoria toma valores nun intervalo ou unión de intervalos da recta real, diremos que se trata dunha **variable aleatoria continua**.

2. Función de densidade e función de distribución

Dada unha variable aleatoria continua, X, o seu comportamento estará caracterizado pola función de distribución ou pola función de densidade. De maneira análoga ao caso de variables discretas, pódese definir a **función de distribución** da variable X, que denotaremos por F_X, como a función que asocia cada valor $x \in \mathbb{R}$ coa probabilidade de que a variable tome valores menores ou iguais a este, é dicir:

$$F_X(x) = \mathbb{P}\left(X \leq x\right).$$

A función de distribución é unha función non decrecente, con valores entre 0 e 1, e que verifica que $\text{lím}_{x\to-\infty} F_X(x) = 0$ e $\text{lím}_{x\to+\infty} F_X(x) = 1$.

Por outra banda, denominamos **función de densidade** dunha variable aleatoria continua X a unha aplicación $f_X : \mathbb{R} \to \mathbb{R}$ definida como:

$$f_X(x) = \underset{h\to 0}{\text{lím}} \frac{\mathbb{P}\left(x - h < X < x + h\right)}{2h}.$$

Unha caracterización da función de densidade que substitúe á definición formal, é que unha función $f : \mathbb{R} \to \mathbb{R}$ calquera é función de densidade se e só se verifican as seguintes condicións:

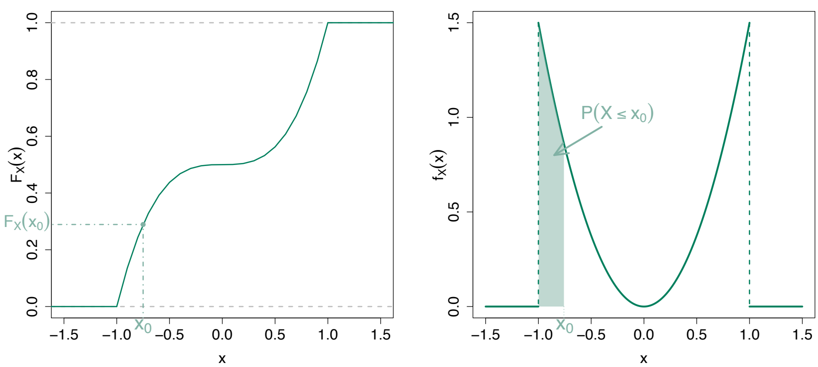
- $f(x) \geq 0, \forall x \in \mathbb{R}$ e
- $\int_{-\infty}^{+\infty} f(x)dx = 1.$

A partir destas definicións e caracterizacións é sinxelo ver que existe unha relación estreita entre a función de densidade e a función de distribución. É dicir:

$$f_X(x) = F_X'(x) \quad \text{ou ben} \quad F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Na Figura 1 mostramos a función de densidade e a función de distribución asociadas a unha certa variable continua. Ademais, ilustramos a utilidade da función de densidade para o cálculo de probabilidades a través da área que se encerra baixo esa curva e o eixe de abscisas.

Figura 1: Exemplo de función de distribución (a) e función de densidade (b) para unha certa variable aleatoria continua X.



(a) Función de distribución.

(b) Función de densidade.

3. Medidas características

As principais medidas características asociadas a unha variable aleatoria en xeral, e a unha variable aleatoria continua en particular son a media, a varianza e a desviación típica.

Media
Sexa X unha variable aleatoria continua con función de densidade f_X, definimos a **media** ou **esperanza poboacional** de X como

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

A media poboacional interprétase como o valor central da distribución de probabilidade da variable aleatoria.

Varianza
Sexa X unha variable aleatoria continua con función de densidade f_X, definimos a **varianza poboacional** de X como

$$\sigma_X^2 = \mathbb{V}ar(X) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx.$$

A varianza poboacional é un valor non negativo que expresa a dispersión da distribución en torno á media poboacional.

Desviación típica
A **desviación típica poboacional** defínese a partir da varianza poboacional como a súa raíz cadrada, isto é,

$$\sigma_X = \sqrt{\mathbb{V}ar(X)} = \sqrt{\int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx}.$$

A vantaxe principal da desviación típica fronte a varianza é que a primeira mídese nas mesmas unidades que a variable aleatoria de interese, polo que os seus valores son máis intuitivos e máis facilmente interpretables.

Ademais destas tres medidas características fundamentais, existen outras medidas características de interese como a moda, a mediana ou, en xeral, os cuantís.

- **Moda:** é o(s) valor(res) no(s) que se alcanza o(s) máximo(s) da función de densidade, entendendo que cando só existe un máximo absoluto existirá unha única moda (e falaremos de distribución unimodal), mentres que se a función de densidade presenta varios máximos relativos estaremos ante varias modas (e falaremos de distribución multimodal).
- **Mediana:** é o valor que deixa unha probabilidade de 0.5 tanto á súa esquerda como á súa dereita. En termos da función de distribución a mediana pode

expresarse como $F_X^{-1}(0.5)$, onde F_X^{-1} denota a inversa da función de distribución, tamén coñecida como función cuantil.

- **Cuantil:** dado un valor $\tau \in (0,1)$ defínese o cuantil de orde τ como o valor que deixa á súa esquerda unha probabilidade a lo menos τ. De novo, tendo en conta a función de distribución podemos expresar o cuantil de orde τ como $F_X^{-1}(\tau)$. A mediana é un caso particular de cuantil, sendo nese caso $\tau = 0.5$.

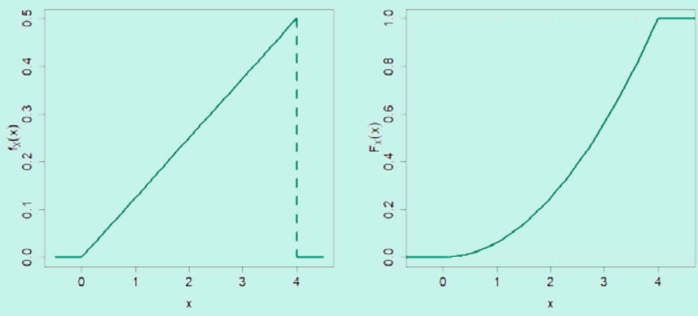
Exemplo 1 *Consideremos unha variable aleatoria X con función de densidade $f_X(x) = \frac{x}{8}$ no intervalo $(0,4)$, e cero noutro caso. Calcular a súa función de distribución, a súa media, a súa varianza e a súa desviación típica.*

Para calcular a función de distribución non teriamos máis que integrar a función de densidade tendo en conta os límites:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0, & \text{se } x < 0, \\ \int_0^x \frac{t}{8}dt = \frac{x^2}{16}, & \text{se } 0 \leq x < 4, \\ 1, & \text{se } x \geq 4. \end{cases}$$

Na Figura 2 representamos a función de densidade e a función de distribución asociadas á variable aleatoria continua X que estamos estudando.

Figura 2: Función de densidade (a) e función de distribución (b) asociadas á variable aleatoria continua X.



(a) Función de densidade. (b) Función de distribución.

Para o cálculo da media aplicaremos a definición correspondente:

$$\mu_X = \int_0^4 x f_X(x)dx = \int_0^4 x\frac{x}{8}dx = \int_1^2 \frac{x^2}{8}dx = \left[\frac{x^3}{24}\right]_0^4 = \frac{8}{3} \text{ unidades.}$$

De xeito análogo obteremos a varianza e a desviación típica, tendo en conta *que neste cálculo precisamos incluír o valor da media previamente obtido:*

$$\begin{aligned}
\sigma_X^2 &= \int_0^4 (x-\mu_X)^2 f_X(x)dx = \int_0^4 \left(x - \frac{8}{3}\right)^2 \frac{x}{8}dx \\
&= \int_0^4 \left(\frac{x^3}{8} + \frac{8x}{9} - \frac{2x^2}{3}\right) dx = \left[\frac{x^4}{32} + \frac{8x^2}{18} - \frac{2x^3}{9}\right]_0^4 \\
&= \frac{8}{9} \text{ unidades}^2, \\
\sigma_X &= \sqrt{\int_0^4 (x-\mu_X)^2 f_X(x)dx} = \sqrt{\frac{8}{9}} = \frac{\sqrt{8}}{3} \text{ unidades}.
\end{aligned}$$

4. Modelos de distribucións continuas. A distribución Normal

Nesta sección faremos un percorrido por algunhas das principais distribucións continuas: uniforme continua, normal, exponencial e gamma. De entre todas elas focalizarémonos con máis detalle na distribución Normal, pois é a que xoga un papel máis importante no campo da estatística.

Uniforme continua

Sexa X unha variable aleatoria que toma valores nun intervalo (a,b), diremos que X segue unha distribución uniforme continua en (a,b), e denotarase por $X \in U(a,b)$, se a súa función de densidade é:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{se } x \in (a,b), \\ 0, & \text{noutro caso.} \end{cases}$$

A función de distribución vén polo tanto dada por:

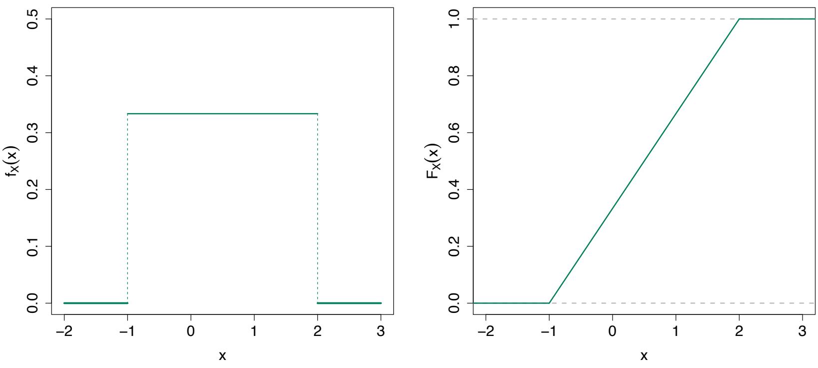
$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0, & \text{se } x < a, \\ \frac{x-a}{b-a}, & \text{se } a \leq x < b, \\ 1, & \text{se } x \geq b. \end{cases}$$

Delas dedúcese que a esperanza e a varianza desta distribución son, respectivamente:

$$\mu_X = \mathbb{E}(X) = \frac{a+b}{2} \quad \text{e} \quad \sigma_X^2 = \mathbb{V}ar(X) = \frac{(b-a)^2}{12}.$$

A modo de exemplo, na Figura 3 represéntanse a función de densidade e a función de distribución asociadas a unha variable aleatoria continua que segue unha distribución uniforme no intervalo $(-1,2)$.

Figura 3: Representación gráfica da función de densidade (a) e da función de distribución (b) para unha variable aleatoria continua que segue unha distribución uniforme no intervalo $(-1, 2)$.



(a) Función de densidade.

(b) Función de distribución.

Normal

A distribución normal é a máis importante e de máis uso de todas as distribucións continuas, pois permite modelizar unha gran cantidade de medicións en distintos ámbitos como a Física, a Química ou a Bioloxía.

Sexa X unha variable aleatoria continua, dise que segue unha distribución normal de media μ_X e varianza σ_X^2, que denotaremos por $X \in N(\mu_X, \sigma_X^2)$, se ten por función de densidade

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}}, \text{ con } -\infty < x < +\infty.$$

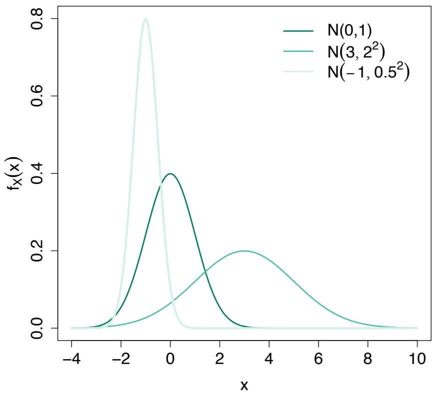
Modificar os valores da media (parámetro de localización) ou da varianza (parámetro de escala) dá lugar a variacións na forma da función de densidade, tal e como se pode ver na Figura 4. Tamén podemos observar como a distribución normal é sempre simétrica respecto da súa media, μ_X, e presenta dous puntos de inflexión en $\mu_X \pm \sigma_X$. No caso particular no que $\mu_X = 0$ e $\sigma_X^2 = 1$, isto é, $X \in N(0, 1)$, diremos que X segue unha distribución normal estándar, e habitualmente denótase pola letra Z.

Unha propiedade importante que verifica a distribución normal, e que é esencial na resolución de numerosos problemas, é a coñecida como estandarización:

$$X \in N(\mu_X, \sigma_X^2) \Rightarrow Z = \frac{X - \mu_X}{\sigma_X} \in N(0, 1).$$

Esta propiedade, entre outras cuestións, permítenos calcular probabilidades dunha distribución normal con calquera parámetros, dispondo unicamente das probabilidades asociadas á distribución normal estándar, por exemplo resumidas nunha táboa como a que pode atoparse no Anexo.

Figura 4: Representación gráfica das funcións de densidade de distintos exemplos de distribucións normais, incluída a normal estándar.



Exemplo 2 *A concentración de calcio en sangue en individuos sans segue unha distribución normal de media* 9.35 *mg/dl e de desviación típica* 0.5 *mg/dl. A afirmación "máis do* $95\,\%$ *dos individuos ten unha concentración de calcio en sangue comprendida entre* 8.4 *e* 10.3 *mg/dl", é correcta?*

Neste problema comezaremos por definir a variable de interese e a súa distribución, que é X="Nivel de calcio en sangue en individuos sans" (mg/dl) verificándose que $X \in N(9.35, 0.5^2)$.

Para poder cotexar a afirmación formulada, temos que calcular a probabilidade de que a variable de interese, X*, se atope entre os dous valores proporcionados. Para isto procederemos como segue:*

$$
\begin{aligned}
\mathbb{P}\left(8.4 \leq X \leq 10.3\right) &= \mathbb{P}\left(\frac{8.4-9.35}{0.5} \leq \frac{X-9.35}{0.5} \leq \frac{10.3-9.35}{0.5}\right) \\
&= \mathbb{P}\left(-1.9 \leq Z \leq 1.9\right) = \mathbb{P}\left(Z \leq 1.9\right) - \mathbb{P}\left(Z < -1.9\right) \\
&= \mathbb{P}\left(Z \leq 1.9\right) - \mathbb{P}\left(Z > 1.9\right) \\
&= \mathbb{P}\left(Z \leq 1.9\right) - \left(1 - \mathbb{P}\left(Z \leq 1.9\right)\right. \\
&= 2 \cdot \mathbb{P}\left(Z \leq 1.9\right) - 1 = 2 \cdot 0.97128 - 1 = 0.94256
\end{aligned}
$$

onde a probabilidade $\mathbb{P}\left(Z \leq 1.9\right)$ *a podemos coñecer grazas á táboa do Anexo. Polo tanto, a afirmación formulada non é correcta.*

Exponencial

Na unidade didáctica anterior estudabamos a distribución de Poisson, $Pois(\lambda)$, que contaba o número de sucesos nun intervalo de tempo. Se ademais de contar o número, queremos medir coa variable X o tempo que transcorre entre dous sucesos consecutivos, esa variable seguirá unha distribución exponencial de parámetro λ, o mesmo que o da Poisson, e denótase por $X \in Exp(\lambda)$.

Dado que esta variable mide o tempo transcorrido, non pode tomar valores negativos, polo que o soporte da mesma é $(0, +\infty)$. A función de densidade asociada a esta variable aleatoria vén dada por:

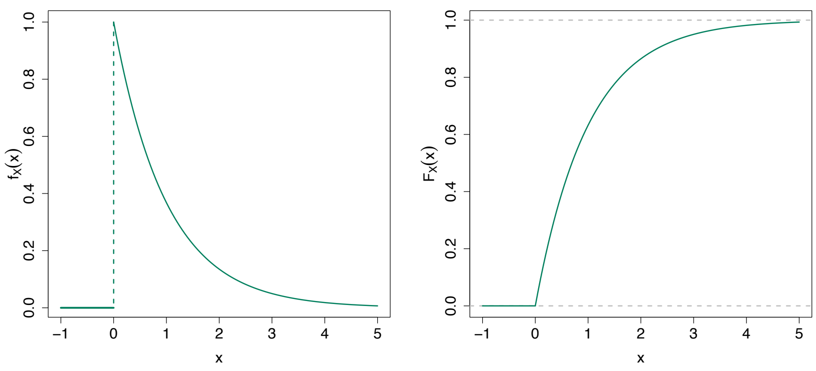
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x > 0, \\ 0, & \text{noutro caso.} \end{cases}$$

A función de distribución obtense integrando por partes a función de densidade anterior, e teriamos que:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{se } x > 0, \\ 0, & \text{noutro caso.} \end{cases}$$

A modo de exemplo, na Figura 5 represéntanse a función de densidade e a función de distribución asociadas a unha variable aleatoria continua que segue unha distribución exponencial de parámetro $\lambda = 1$.

Figura 5: Representación gráfica da función de densidade (a) e da función de distribución (b) para unha variable aleatoria continua que segue unha distribución exponencial de parámetro $\lambda = 1$.



(a) Función de densidade.

(b) Función de distribución.

A media e a varianza, que tamén se poden obter integrando por partes a función de densidade, e son, respectivamente:

$$\mu_X = \mathbb{E}(X) = \frac{1}{\lambda} \quad \text{e} \quad \sigma_X^2 = \mathbb{V}ar(X) = \frac{1}{\lambda^2}.$$

Gamma

Dado un experimento de Poisson, se en lugar de interesarnos medir o tempo entre dous sucesos consecutivos, queremos medir o tempo ata o suceso m-ésimo, obtense a xeneralización da distribución exponencial, que se coñece co nome de distribución gamma con parámetros m e λ, que denotaremos por $X \in \Gamma(m, \lambda)$. Neste caso o soporte segue sendo $(0, +\infty)$ e a función de densidade vén dada por:

$$f_X(x) = \begin{cases} \frac{\lambda^m}{\Gamma(m)} e^{-\lambda x} x^{m-1}, & \text{se } x \in (a, b), \\ 0, & \text{noutro caso,} \end{cases}$$

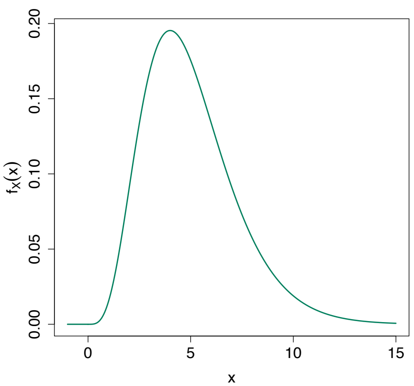
onde $\Gamma(m) = (m-1)!$ con $m \in \mathbb{N}$.

Se nos fixamos nesta expresión da densidade veremos como $Exp(\lambda) \equiv \Gamma(1, \lambda)$. Neste contexto non existe unha expresión explícita manexable da función de distribución, polo que simplemente indicaremos que a súa media e a súa varianza veñen dadas, respectivamente, por:

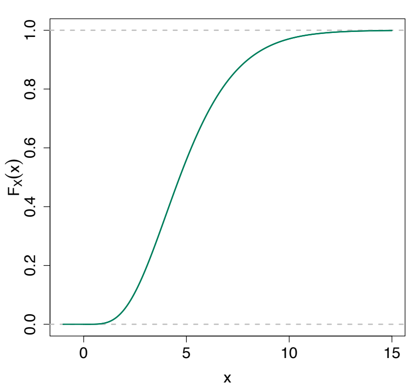
$$\mu_X = \mathbb{E}(X) = \frac{m}{\lambda} \quad \text{e} \quad \sigma_X^2 = \mathbb{V}ar(X) = \frac{m}{\lambda^2}.$$

A modo de exemplo, na Figura 5 represéntanse a función de densidade e a función de distribución asociadas a unha variable aleatoria continua que segue unha distribución gamma de parámetros $m = 5$ e $\lambda = 1$.

Figura 6: Representación gráfica da función de densidade (a) e da función de distribución (b) para unha variable aleatoria continua que segue unha distribución gamma de parámetros $m = 5$ e $\lambda = 1$.



(a) Función de densidade.

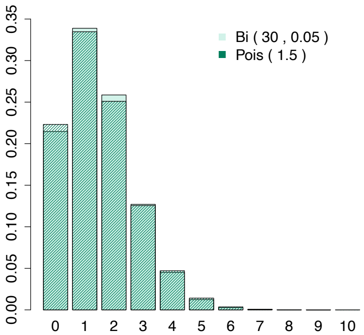


(b) Función de distribución.

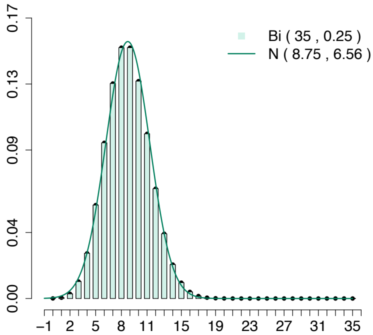
5. Aproximación entre distribucións

Se retomamos os experimentos de Bernoulli vistos na Unidade Didáctica 3, para un número, m, de probas dese experimento, suficientemente grande, podemos aproximar a distribución binomial pola Poisson ou pola normal, facilitando así o cálculo de probabilidades. Tamén na distribución de Poisson, cando o parámetro λ é suficientemente grande, poden aproximarse as súas probabilidades polas da distribución normal.

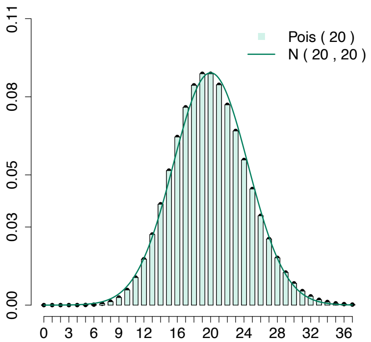
Figura 7: Ilustración, mediante tres exemplos concretos, das posibles aproximacións entre as distribucións binomial, Poisson e normal.



(a) Aproximación binomial por Poisson.



(b) Aproximación binomial por normal.



(c) Aproximación Poisson por normal.

As condicións prácticas para poder facer as devanditas aproximacións recóllense na Táboa 4. Ademais, na Figura 7 ilústranse a través duns exemplos concretos estas tres posibles aproximacións.

Táboa 4: Condicións de aproximación entre varias distribucións notables.

Distribución	Condicións	Aproximación
$Bi(m,p)$	$m > 30$ e $p < 0.1$ ou $p > 0.9$	$Pois(mp)$
$Bi(m,p)$	$m > 30,\ 0.1 < p < 0.9$	$N\left(mp, mp(1-p)\right)$
$Bi(m,p)$	$p < 0.1$ e $np > 5$	$N\left(mp, mp(1-p)\right)$
$Bi(m,p)$	$p > 0.9$ e $n(1-p) > 5$	$N\left(mp, mp(1-p)\right)$
$Pois(\lambda)$	$\lambda \geq 10$	$N(\lambda, \lambda)$

AVALIACIÓN DA UNIDADE DIDÁCTICA

Na memoria de verificación do Grao en Medicina especifícase que en todas as materias do Grao a cualificación de cada alumna/o farase mediante avaliación continua e a realización dun exame final. En concreto, no caso da materia de Bioestatística, a avaliación continua terá un peso do $30\,\%$ da nota final, mentres que o exame final terá un peso do $70\,\%$.

A avaliación continua desta unidade didáctica realizarase conxuntamente coas outras unidades didácticas que forman parte do Bloque II. Teoría da Probabilidade supoñendo un peso do $6\,\%$ na nota final da materia. Esta avaliación engloba os once obxectivos específicos desta unidade e consistirá na entrega dun exercicio realizado de xeito individual por cada alumna/o como traballo non presencial.

Así mesmo, no exame final da materia poderanse incluír preguntas de tipo teórico e/ou práctico sobre os contidos desta unidade didáctica.

BIBLIOGRAFÍA

Bibliografía básica

BORRAJO, M. I., CONDE-AMBOAGE, M. E CRUJEIRAS-CASAIS, R. M. (2021). *Fundamentos da teoría da probabilidade*. Esenciais USC. Universidade de Santiago de Compostela. https://www.usc.gal/libros/es/categorias/1025-fundamentos-da-teoria-da-probabilidade-334620-fundamentos-da-teoria-da-probabilidade.html#/29-formato-pdf/34-transaccion-gratis.

CAO, R. E OUTROS (1998). *Introducción a las estadística y sus aplicaciones.* Pirámide.

CRUJEIRAS-CASAIS, R. M. E FARALDO-ROCA, P. (2010). *Manual de estadística básica para ciencias de la salud.* Universidade de Santiago de Compostela, Dpto. de Estadística e Investigación Operativa.

MILTON, J.S. (1994) *Estadística para biología y ciencias de la salud*. (2ª ed). Interamericana, McGraw-Hill.

ROSNER, B. (2006) *Fundamentals of biostatistics* (6ª edición). Wadsworth Publishing Company. Duxbury Press.

ROSS, S. (2014) *Introduction to Probability and Statistics for Engineers and Scientists* (5ª edición). Elsevier Academic Press.

Bibliografía complementaria

ALTMAN, D. G. (1990). *Practical statistics for medical research*. Chapman & Hall.

ÁLVAREZ-CÁCERES, R. (2007). *Estadística Aplicada a las Ciencias de la Salud.* Editorial Diaz de Santos.

COBO, E., MUÑOZ, P., GONZÁLEZ, J. A., BIGORRA, J., CORCHERO, C., MIRAS, F., SELVA, A. E VIDELA, S. (2007). *Bioestadística para no estadísticos.* Elsevier.

DANIEL, W. (1991). *Bioestadística. Base para el análisis de las ciencias de la salud.* Limusa Wiley.

DURRETT, R. (2009). *Elementary probability for applications.* Cambridge University Press.

RECURSOS NA REDE

Hoxe en día, tanto alumnado como profesorado teñen acceso a multitude de recursos a través da Rede. Aínda que é certo que na liberdade que reina no mundo dixital hai que saber cales merece a pena consultar. É por iso que se esta sección pretende proporcionar recursos de calidade para diferentes fins dentro do proceso de aprendizaxe.

— ***Academic Skills Kit*** **(ASK)**
https://internal.ncl.ac.uk/ask/numeracy-maths-statistics/statistics/
A Universidade de Newcastle pon a disposición de calquera usuario da Rede cursos de diversas temáticas. En particular, a páxina de estatística ten un curso de iniciación bastante completo, no que para cada tema comeza coas definicións teóricas básicas, seguidas dalgunha actividade extra, que poden ser vídeos, exemplos ilustrativos, exercicios, outros recursos electrónicos ... Ademais, todos os exercicios propostos son acompañado dunha solución detallada.

— ***GeoGebra***
https://www.geogebra.org/
GeoGebra é un conxunto de ferramentas dixitais gratuítas para a docencia das matemáticas en xeral, e da estatística en particular, á que lle foron concedidos numerosos premios e mencións. En relación á temática desta unidade, destacamos ferramentas para visualizar as distribucións notables: https://www.geogebra.org/m/rZ7NHCsS para a Normal, https://www.geogebra.org/m/GhVNUhHf para a Exponencial, ou https://www.geogebra.org/m/yCZgR3dh para a Gamma. Estes son só algúns exemplos, pero existe unha gran cantidade e variedade de ferramentas sinxelas e moi útiles de cara á visualización e interpretación dos conceptos vistos ao longo desta unidade.

— **SOCR**
http://www.distributome.org/tools.html
SOCR é un recurso interactivo gratuíto e accesible a través da Rede para exploración, modelado, análise e interpretación de datos. Foi deseñado pola Universidade de California no 2006. Inclúe ferramentas e recursos baseados en *applets* interactivos como aplicacións de cálculo e gráficos para a impartición de cursos estatísticos. A súa filosofía é a dunha ferramenta aberta e ampliable para a comunidade educativa. É unha das maiores coleccións de *applets* Java, moi útil para a aprendizaxe autónoma, motivación, modernización e mellora do modelo de ensino estatístico. Promove a aprendizaxe activa e práctica, co/coa estudante como axente centrado/a no proceso de aprendizaxe. Para a unidade de variables aleatorias continuas dispón de varias aplicacións nas que se poden simular diversos experimentos, ver representacións das funcións características e mesmo visualizar as aproximacións límite.

— ***Shiny apps***
http://www.showmeshiny.com/
Son aplicacións sinxelas feitas con código R, pero cunha interface gráfica moi amigable que permite á/ao usuaria/o «xogar» cos números. Os temas que tratan son moi variados, pero hai moitos dedicados a conceptos relacionados con variables aleatorias continuas.

ANEXO: TÁBOA DA DISTRIBUCIÓN NORMAL ESTÁNDAR

Dada $Z \in N(0,1)$ preséntase $\mathbb{P}(Z \leq z)$ para todo $z \in \mathbb{R}$.

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	$0.9^{2}0097$	$0.9^{2}0358$	$0.9^{2}0613$	$0.9^{2}0863$	$0.9^{2}1106$	$0.9^{2}1344$	$0.9^{2}1576$
2.4	$0.9^{2}1802$	$0.9^{2}2024$	$0.9^{2}2240$	$0.9^{2}2451$	$0.9^{2}2656$	$0.9^{2}2857$	$0.9^{2}3053$	$0.9^{2}3244$	$0.9^{2}3431$	$0.9^{2}3613$
2.5	$0.9^{2}3790$	$0.9^{2}3963$	$0.9^{2}4132$	$0.9^{2}4297$	$0.9^{2}4458$	$0.9^{2}4614$	$0.9^{2}4766$	$0.9^{2}4915$	$0.9^{2}5060$	$0.9^{2}5201$
2.6	$0.9^{2}5339$	$0.9^{2}5473$	$0.9^{2}5604$	$0.9^{2}5731$	$0.9^{2}5854$	$0.9^{2}5975$	$0.9^{2}6093$	$0.9^{2}6207$	$0.9^{2}6319$	$0.9^{2}6427$
2.7	$0.9^{2}6533$	$0.9^{2}6636$	$0.9^{2}6736$	$0.9^{2}6833$	$0.9^{2}6928$	$0.9^{2}7020$	$0.9^{2}7110$	$0.9^{2}7197$	$0.9^{2}7282$	$0.9^{2}7365$
2.8	$0.9^{2}7445$	$0.9^{2}7523$	$0.9^{2}7599$	$0.9^{2}7673$	$0.9^{2}7744$	$0.9^{2}7814$	$0.9^{2}7882$	$0.9^{2}7948$	$0.9^{2}8012$	$0.9^{2}8074$
2.9	$0.9^{2}8134$	$0.9^{2}8193$	$0.9^{2}8250$	$0.9^{2}8305$	$0.9^{2}8359$	$0.9^{2}8411$	$0.9^{2}8462$	$0.9^{2}8511$	$0.9^{2}8559$	$0.9^{2}8605$
3	$0.9^{2}8650$	$0.9^{2}8694$	$0.9^{2}8736$	$0.9^{2}8777$	$0.9^{2}8817$	$0.9^{2}8856$	$0.9^{2}8893$	$0.9^{2}8930$	$0.9^{2}8965$	$0.9^{2}8999$
3.1	$0.9^{3}0324$	$0.9^{3}0646$	$0.9^{3}0957$	$0.9^{3}1260$	$0.9^{3}1553$	$0.9^{3}1837$	$0.9^{3}2112$	$0.9^{3}2378$	$0.9^{3}2636$	$0.9^{3}2886$
3.2	$0.9^{3}3129$	$0.9^{3}3363$	$0.9^{3}3591$	$0.9^{3}3811$	$0.9^{3}4024$	$0.9^{3}4230$	$0.9^{3}4429$	$0.9^{3}4623$	$0.9^{3}4810$	$0.9^{3}4991$
3.3	$0.9^{3}5166$	$0.9^{3}5335$	$0.9^{3}5499$	$0.9^{3}5658$	$0.9^{3}5811$	$0.9^{3}5959$	$0.9^{3}6103$	$0.9^{3}6242$	$0.9^{3}6376$	$0.9^{3}6505$
3.4	$0.9^{3}6631$	$0.9^{3}6752$	$0.9^{3}6869$	$0.9^{3}6982$	$0.9^{3}7091$	$0.9^{3}7197$	$0.9^{3}7299$	$0.9^{3}7398$	$0.9^{3}7493$	$0.9^{3}7585$
3.5	$0.9^{3}7674$	$0.9^{3}7760$	$0.9^{3}7842$	$0.9^{3}7922$	$0.9^{3}7999$	$0.9^{3}8074$	$0.9^{3}8146$	$0.9^{3}8215$	$0.9^{3}8282$	$0.9^{3}8347$
3.6	$0.9^{3}8409$	$0.9^{3}8469$	$0.9^{3}8527$	$0.9^{3}8583$	$0.9^{3}8637$	$0.9^{3}8689$	$0.9^{3}8739$	$0.9^{3}8787$	$0.9^{3}8834$	$0.9^{3}8879$
3.7	$0.9^{3}8922$	$0.9^{3}8964$	$0.9^{4}0039$	$0.9^{4}0426$	$0.9^{4}0799$	$0.9^{4}1158$	$0.9^{4}1504$	$0.9^{4}1838$	$0.9^{4}2159$	$0.9^{4}2468$
3.8	$0.9^{4}2765$	$0.9^{4}3052$	$0.9^{4}3327$	$0.9^{4}3593$	$0.9^{4}3848$	$0.9^{4}4094$	$0.9^{4}4331$	$0.9^{4}4558$	$0.9^{4}4777$	$0.9^{4}4988$
3.9	$0.9^{4}5190$	$0.9^{4}5385$	$0.9^{4}5573$	$0.9^{4}5753$	$0.9^{4}5926$	$0.9^{4}6092$	$0.9^{4}6253$	$0.9^{4}6406$	$0.9^{4}6554$	$0.9^{4}6696$
4	$0.9^{4}6833$	$0.9^{4}6964$	$0.9^{4}7090$	$0.9^{4}7211$	$0.9^{4}7327$	$0.9^{4}7439$	$0.9^{4}7546$	$0.9^{4}7649$	$0.9^{4}7748$	$0.9^{4}7843$

unidadesdidácticas
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Unha colección orientada a editar materiais docentes de calidade e pensada para apoiar o traballo do profesorado e do alumnado de todas as materias e titulacións da universidade

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA