

# PROBLEMAS DE PROBABILIDADES Y ESTADÍSTICA

## Vol. 2. Inferencia estadística

Carles M. Cuadras

Departamento de Estadística

# PROBLEMAS DE PROBABILIDADES Y ESTADÍSTICA

## Vol. 2. Inferencia estadística

---

Carles M. Cuadras

Departamento de Estadística

408

TEXTOS DOCENTS



UNIVERSITAT DE  
BARCELONA

Edicions

# Índice

<b>PRÓLOGO</b> .....	9
<b>1. MUESTREO</b> .....	11
1.1. Muestreo aleatorio simple .....	11
1.2. Concepto de estadístico. Características muestrales .....	11
1.3. Muestreo artificial .....	12
1.4. Imagen empírica de una distribución continua .....	13
1.5. Muestreo sobre poblaciones finitas .....	14
Problemas resueltos .....	14
Enunciados .....	22
<b>2. ESTIMACIÓN PUNTUAL</b> .....	25
2.1. Concepto de estimador .....	25
2.2. Propiedades de los estimadores .....	25
2.3. Cota de Cramér-Rao .....	27
2.4. Suficiencia .....	28
2.5. Estimación de varios parámetros .....	30
Problemas resueltos .....	30
Enunciados .....	49
<b>3. MÉTODOS DE ESTIMACIÓN</b> .....	53
3.1. Método de los momentos .....	53
3.2. Método de la máxima verosimilitud .....	53
3.3. Propiedades de los estimadores máximo verosímiles .....	54
3.4. Método de Bayes .....	55
3.5. Propiedades del estimador de Bayes .....	56
3.6. Familias conjugadas .....	56
Problemas resueltos .....	57
Enunciados .....	71
<b>4. CONTRASTE DE HIPÓTESIS</b> .....	75
4.1. Concepto de test. Hipótesis simple frente a alternativa simple .....	75
4.2. Regiones críticas .....	75
4.3. Lema de Neyman-Pearson .....	76
4.4. Hipótesis simple frente a alternativa compuesta .....	77
4.5. Hipótesis compuesta .....	78
4.6. Test de razón de verosimilitud .....	79
Problemas resueltos .....	81
Enunciados .....	98

<b>5. INTERVALOS DE CONFIANZA</b> .....	103
5.1. Definición de intervalo de confianza .....	103
5.2. Observaciones .....	103
5.3. Métodos de construcción .....	104
5.4. Contraste de hipótesis mediante intervalos de confianza .....	105
5.5. Método de Bayes .....	105
Problemas resueltos .....	106
Enunciados .....	115
<b>6. ANÁLISIS ESTADÍSTICO NORMAL</b> .....	119
6.1. Teorema de Fisher .....	119
6.2. Test $t$ sobre la media .....	119
6.3. Test $t$ de comparación de medias .....	120
6.4. Test $t$ para datos apareados .....	121
6.5. Test $F$ de comparación de varianzas .....	122
6.6. Intervalos de confianza .....	123
6.7. Test multivariantes sobre la media .....	124
Problemas resueltos .....	125
Enunciados .....	135
<b>7. ANÁLISIS ESTADÍSTICO EN REGRESIÓN</b> .....	139
7.1. Estimación de parámetros en regresión .....	139
7.2. Inferencia en el caso de dos variables .....	139
7.3. Comparación de rectas de regresión .....	143
7.4. Inferencia en regresión múltiple y correlación parcial .....	145
Problemas resueltos .....	146
Enunciados .....	155
<b>8. LA PRUEBA JI-CUADRADO</b> .....	159
8.1. Fundamento teórico .....	159
8.2. Caso de probabilidades conocidas .....	160
8.3. Caso de probabilidades dependientes de parámetros .....	161
8.4. Prueba de independencia en tablas de contingencia .....	161
8.5. Observaciones .....	163
8.6. Test de homogeneidad .....	163
8.7. Homogeneidad en distribuciones binomiales .....	164
8.8. Homogeneidad en distribuciones de Poisson .....	165
8.9. Grado de significación .....	166
Problemas resueltos .....	167
Enunciados .....	174
<b>9. PRUEBAS DE BONDAD DEL AJUSTE A UNA DISTRIBUCIÓN</b> .....	177
9.1. Bondad del ajuste mediante la prueba ji-cuadrado .....	177
9.2. Test de Kolmogorov .....	178
9.3. Test de Kolmogorov en muestras pequeñas .....	179
9.4. Test de Crámer-Von Mises .....	180
9.5. Pruebas de normalidad multivariante .....	180

Problemas resueltos .....	182
Enunciados .....	190
<b>10. ANÁLISIS DE LA VARIANZA .....</b>	<b>193</b>
10.1. Introducción .....	193
10.2. Diseño de un factor .....	193
10.3. Diseño de dos factores. Bloques aleatorizados .....	195
10.4. Diseño de cuadrados latinos .....	197
10.5. Diseño de dos factores con interacción .....	199
10.6. Diseños anidados .....	201
10.7. Diseños no balanceados y observaciones faltantes .....	201
Problemas resueltos .....	202
Enunciados .....	210
<b>11. ANÁLISIS DE LA VARIANZA (AMPLIACIÓN) .....</b>	<b>213</b>
11.1. Descomposición ortogonal de la variabilidad en diseños multifactoriales .....	213
11.2. Diseño de parcelas divididas. Otros diseños .....	216
11.3. Estimación de parámetros y cálculo del residuo .....	219
11.4. Análisis de componentes de la varianza .....	220
11.5. Estimación de las componentes de la varianza .....	223
11.6. Correlación intraclásica .....	223
Problemas resueltos .....	224
Enunciados .....	233
<b>12. COMPARACIONES MÚLTIPLES Y ANÁLISIS DE LA COVARIANZA .....</b>	<b>237</b>
12.1. Intervalos de confianza simultáneos .....	237
12.2. Método de Bonferroni .....	237
12.3. Comparaciones múltiples en diseños experimentales .....	238
12.4. Método de Scheffé .....	239
12.5. Análisis de la covarianza .....	241
12.6. Un factor y una variable concomitante .....	241
12.7. Dos factores y una variable concomitante .....	242
12.8. Otros diseños .....	244
Problemas resueltos .....	246
Enunciados .....	259
<b>13. ESTADÍSTICA NO PARAMÉTRICA .....</b>	<b>263</b>
13.1. Introducción .....	263
13.2. Test de los signos .....	263
13.3. Test de McNemar .....	264
13.4. Test del signo-rango de Wilcoxon .....	264
13.5. Test U de Mann-Whitney .....	265
13.6. Test de Smirnov .....	267
13.7. Test de Kruskal-Wallis .....	268
Problemas resueltos .....	268
Enunciados .....	275

<b>14. ESTADÍSTICA NO PARAMÉTRICA (AMPLIACIÓN)</b> .....	279
14.1. Test de Friedman .....	279
14.2. Correlación por rangos de Kendall .....	280
14.3. Correlación por rangos de Spearman .....	281
14.4. Estimación no paramétrica .....	283
Problemas resueltos .....	286
Enunciados .....	296
<b>SOLUCIONES</b> .....	299
<b>PROBLEMAS DE RECAPITULACIÓN</b> .....	309
<b>RESUMEN DE ESTIMACIÓN PUNTUAL DE PARÁMETROS</b> .....	325
<b>BIBLIOGRAFÍA</b> .....	327
<b>TABLAS ESTADÍSTICAS</b> .....	329
I. Tabla de valores de $e^{-x}$ .....	331
II. Función de distribución normal reducida .....	332
III. Tabla de la distribución normal (dos colas) .....	333
IV. Distribución ji-cuadrado .....	334
V. Distribución t de Student .....	335
VI. Valores críticos de la distribución F $P(F_n^m > \text{valor tabulado}) = 0.1$ .....	336
VII. Valores críticos de la distribución F $P(F_n^m > \text{valor tabulado}) = 0.05$ .....	337
VIII. Valores críticos de la distribución F $P(F_n^m > \text{valor tabulado}) = 0.025$ .....	338
IX. Valores críticos de la distribución F $P(F_n^m > \text{valor tabulado}) = 0.01$ .....	339
X. Valores críticos de la distribución F $P(F_n^m > \text{valor tabulado}) = 0.005$ .....	340
XI. Distribución de Kolmogorov-Smirnov. Tablas de Massey y Lilliefors .....	341
XII. Distribución de Mann-Whitney-Wilcoxon .....	342
XIII. Distribución de Mann-Whitney-Wilcoxon (cont.) .....	343
XIV. Valores críticos del test del signo-rango de Wilcoxon.	
Tabla de Massey para el test de Smirnov .....	344
<b>ÍNDICE ANALÍTICO</b> .....	345

# Prólogo

La primera versión de esta obra fue publicada por la Universidad de Barcelona en 1974, formando parte de la colección Publicaciones del Laboratorio de Cálculo. Pronto se convirtió en libro y después de un periplo de cinco lustros por las editoriales EUNIBAR, PPU y EUB S. L., y tras una docena de ediciones, la última (revisada y ampliada), apareció en 1999-2000. Agotadas estas ediciones, y ante la insistencia de varios colegas, la obra regresa a la Universidad de Barcelona, que se ha ofrecido a editarla y publicarla con nuevo formato, dentro de la colección de Textos Docentes.

Es este un libro de ejercicios y, por lo tanto, destinado a proponer y resolver problemas de probabilidades y estadística. Los problemas de naturaleza teórica se alternan con los aplicados. Se ha procurado presentarlos en orden de dificultad creciente. La extensión de la obra ha obligado a dividirla en dos volúmenes: el primero dedicado a Problemas de probabilidades y el segundo, a Problemas de inferencia estadística. La actual edición EUB de la Universidad de Barcelona mejora la edición de 1999-2000 en la presentación y revisión de algunos apartados.

A fin de facilitar la consulta del libro, se incluye en cada capítulo un extenso resumen de la teoría. Para ampliar la teoría, el lector deberá consultar libros adecuados y puede orientarse en la bibliografía que aparece al final de cada volumen. Los apartados de mayor dificultad teórica y los problemas de resolución más difícil se indican con un asterisco.

La Estadística se ha ido imponiendo como materia fundamental en la mayoría de los estudios universitarios. Esta obra, con un planteamiento teórico-práctico, sintetiza los principales temas de la Probabilidad y la Estadística. A pesar de los grandes avances tecnológicos, la base no ha cambiado, por lo que el contenido de este libro sigue teniendo plena actualidad.

Este segundo libro, dedicado a la inferencia estadística, contiene problemas teóricos y aplicados sobre la base de datos de procedencia muy variada.

Los dos volúmenes constituyen una obra que resulta adecuada para estudiantes y profesores de primer ciclo de carreras y grados universitarios (Estadística, Biología, Matemáticas, Física, Química, Geología, Ingeniería, Arquitectura, Informática, Economía, Veterinaria, Psicología, Pedagogía y Medicina).

Mi más sincero agradecimiento a todos aquellos que me han hecho comentarios, sugerido cambios y correcciones, lo que ha hecho posible la versión actual. En especial debo mencionar a J. Fortiana, P. Sánchez Algarra, G. Alonso, J. M. Oller, F. Carmona, J. Ocaña, C. Ruiz-Rivas, M. Ríos, R. Vélez, J. M. Font, S. Salvo y D. Cuadras.

Respecto a la presente edición, deseo agradecer a M. Aicart, J. M. Oller, M. C. Pallejá y Edicions de la Universitat de Barcelona su valiosa colaboración en la preparación del original.

Carles M. Cuadras  
Barcelona, diciembre de 2016

# 1. MUESTREO

## 1.1. Muestreo aleatorio simple

Sea  $X$  una variable aleatoria (v.a.) definida sobre una población  $\Omega$ . Dados  $n$  valores de  $X$

$$x_1, x_2, \dots, x_n$$

observados en condiciones independientes, diremos que constituyen una *muestra aleatoria simple* de  $X$ . Los valores de una muestra son aleatorios (a muestras distintas tendremos valores distintos), y desde el punto de vista probabilístico, deben considerarse como  $n$  v.a.  $X_1, X_2, \dots, X_n$  tales que:

1. Son v.a. independientes.
2. Cada  $X_i$  tiene la misma distribución que  $X$ .

Por lo tanto, si  $f_X(x)$  es la función de densidad de  $X$ , la función de densidad de una muestra aleatoria simple  $x_1, x_2, \dots, x_n$  es

$$f(x_1, x_2, \dots, x_n) = f_X(x_1) \cdot f_X(x_2) \cdots f_X(x_n).$$

El conjunto de posibles valores que podemos obtener mediante un muestreo aleatorio simple de tamaño  $n$  es un subconjunto del espacio euclídeo  $\mathbb{R}^n$ , llamado *espacio muestral*.

## 1.2. Concepto de estadístico. Características muestrales

Dada una muestra  $(X_1, \dots, X_n)$  se llama *estadístico* a toda variable que sea función de la muestra

$$U_n = g(X_1, \dots, X_n).$$

Los estadísticos se presentan de forma natural en problemas de estimación de parámetros (y entonces reciben el nombre de estimadores), y en contraste de hipótesis. Son aspectos relacionados con la obtención de un estadístico:

1. Encontrar el estadístico adecuado al problema de inferencia estadística que pretendemos resolver.
2. Encontrar la distribución exacta de  $U_n$  para cualquier  $n$ .
3. Encontrar la distribución asintótica de  $U_n$  para  $n \rightarrow \infty$ .
4. Determinar el tamaño de la muestra  $n$  para obtener una precisión suficiente en estimación de parámetros, o construir un test con potencia alta.

Los parámetros poblacionales  $m = E(X)$ ,  $\sigma^2 = \text{var}(X)$ ,  $m_k = E(X^k)$  están relacionados con ciertos estadísticos especialmente importantes llamados características muestrales:

$$\text{Media muestral} \quad \bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$

$$\text{Varianza muestral} \quad S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

$$\text{Momento muestral} \quad A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

Se verifica:

$$E(\bar{X}_n) = m \quad \text{var}(\bar{X}_n) = \frac{\sigma^2}{n} \quad E(S_n^2) = \frac{n-1}{n} \sigma^2 \quad E(A_k) = m_k$$

Las características muestrales son v.a. (que varían con el muestreo), algunos casos con distribución exacta conocida y otros con distribución asintótica, generalmente normal.  $\bar{X}_n$ ,  $S_n^2$  y  $A_k$  no deben confundirse con  $m$ ,  $\sigma^2$  y  $m_k$ , que son las características poblacionales de  $X$  y tienen valor constante. Las características muestrales calculadas sobre una muestra dada se indican por  $\bar{x}_n$ ,  $s_n^2$  y  $a_k$ .

Análogamente, se define un muestreo aleatorio simple de tamaño  $n$  del par  $(X, Y)$  con distribución bivalente, como  $n$  pares ordenados

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

obtenidos en condiciones independientes. Son características muestrales importantes:

$$\text{Covarianza muestral:} \quad S_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$\text{Correlación muestral:} \quad r = \frac{S_{XY}}{S_X S_Y}$$

donde  $S_X = \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}$ ,  $S_Y = \left( \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \right)^{1/2}$  son las desviaciones típicas muestrales. Tampoco debe confundirse  $r$  (que es una v.a.) con la correlación poblacional  $\rho$ .

### 1.3. Muestreo artificial

En ciertos problemas es útil simular la distribución de una v.a.  $X$  con función de distribución  $F_X(x)$ , obteniendo una muestra artificial de  $X$ . El proceso es el siguiente:

1. Se genera una muestra aleatoria simple

$$z_1, z_2, \dots, z_n$$

de una v.a. uniforme en  $(0, 1)$ .

2. Si  $F_X$  es continua y creciente, entonces

$$x_1 = F_X^{-1}(z_1), x_2 = F_X^{-1}(z_2), \dots, x_n = F_X^{-1}(z_n)$$

es una muestra aleatoria simple de  $X$ .

3. Si  $X$  tiene distribución discreta, se genera la muestra haciendo corresponder a  $z_i$  el valor  $x_i$  tal que  $F_X(x_{i-1}) < z_i \leq F(x_i)$ .

Un procedimiento para generar una muestra sobre  $(0, 1)$  es iniciarla con un valor  $z_1$  cualquiera, y utilizar la fórmula de recurrencia

$$z_i = \text{parte fraccionaria de } [(\pi + z_{i-1})^8].$$

El muestreo artificial se aplica para comprobar la calidad de un estimador, la potencia de un test, simular la distribución de un estadístico  $U_n$  cuya distribución exacta resulta complicada, simulación en modelos genéticos, etc.

### 1.4. Imagen empírica de una distribución continua

La función de distribución  $F(x)$  describe totalmente la distribución de una v.a.  $X$ . Sin embargo, si nuestra información sobre  $X$  es una muestra aleatoria simple  $x_1, \dots, x_n$ , podemos obtener una imagen empírica de la distribución ordenando la muestra

$$x_{(1)} < x_{(2)} < \dots < x_{(n)}$$

y definiendo la *función de distribución empírica*

$$F_n(x) = \begin{cases} 0 & \text{si } x < x_{(1)}, \\ \frac{k}{n} & \text{si } x_{(k)} \leq x < x_{(k+1)}, \quad k = 1, \dots, n-1, \\ 1 & \text{si } x \geq x_{(n)}. \end{cases}$$

Como  $F_n(x)$  es la frecuencia relativa de la presencia del suceso  $[X \leq x]$ , se verifica, para cada  $x \in \mathbb{R}$ , como consecuencia del teorema de Bernoulli, la convergencia en probabilidad

$$F_n(x) \xrightarrow{P} F(x). \tag{1}$$

Por lo tanto, para  $n$  grande,  $F_n(x)$  proporciona una imagen de la distribución de la variable  $X$ .

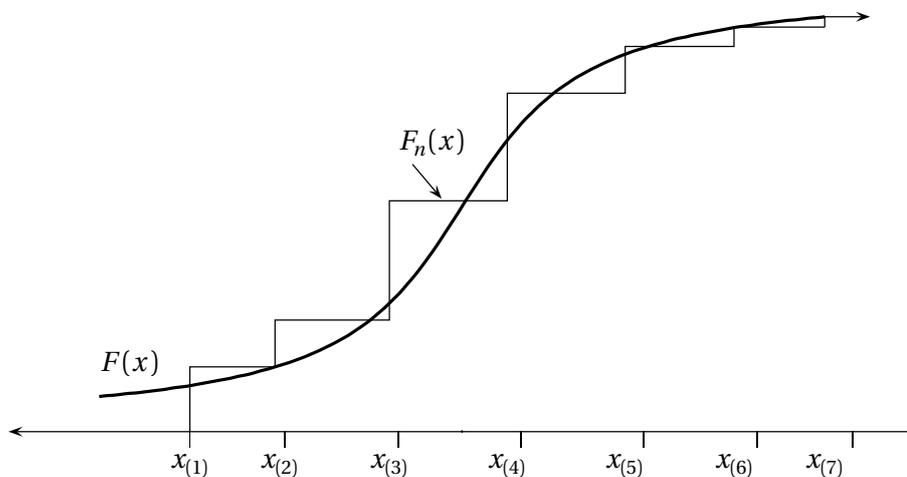


Figura 1.1.

Una manera de estudiar la diferencia entre  $F_n(x)$  y  $F(x)$  es introducir el estadístico

$$D_n = \sup_x |F_n(x) - F(x)|.$$

Se verifica, entonces, una propiedad todavía más fuerte que (1), conocida como teorema de Glivenko-Cantelli, también llamado teorema fundamental de la estadística.

**Teorema.**  $F_n(x)$  converge casi seguramente a  $F(x)$  uniformemente en  $x$ , es decir,

$$P\left(\lim_{n \rightarrow \infty} D_n = 0\right) = 1.$$

## 1.5. Muestreo sobre poblaciones finitas

Se entiende por población finita un conjunto finito de  $N$  elementos de naturaleza homogénea ( $N$  familias,  $N$  piezas fabricadas,  $N$  individuos de una especie). Si obtenemos una muestra de tamaño  $n$ , *con reemplazamiento*, cada elemento tiene una probabilidad  $1/N$ , las muestras son independientes y el muestreo es aleatorio simple. Si la muestra obtenida es *sin reemplazamiento*, las diferentes muestras no son independientes.

Si el muestreo se realiza sobre una variable cuantitativa  $X$ , son características poblacionales importantes la media y la varianza

$$m = \frac{1}{N} \sum_{i=1}^N x_i \quad \sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - m)^2.$$

En el muestreo sin reemplazamiento, la media muestral  $\bar{x}'$  verifica

$$E(\bar{x}') = m \quad \text{var}(\bar{x}') = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}.$$

Si lo que se desea estudiar es la presencia de una característica cualitativa de probabilidad  $p$ , de modo que  $N_p$  es el número de elementos que poseen esta característica, si  $\bar{x}$  es la frecuencia relativa de esta característica en el muestreo con reemplazamiento tendremos:

$$E(\bar{x}) = p \quad \text{var}(\bar{x}) = \frac{pq}{n},$$

siendo  $q = 1-p$ . Si  $\bar{x}'$  es la frecuencia relativa en el muestreo sin reemplazamiento, entonces:

$$E(\bar{x}') = p \quad \text{var}(\bar{x}') = \frac{N-n}{N-1} \cdot \frac{pq}{n}.$$

## Problemas resueltos

1.1. Sea  $X$  una v.a. con distribución de Poisson de parámetro  $\lambda$ . Dada una muestra aleatoria simple de tamaño  $n$

$$X_1, \dots, X_n$$

resolver las siguientes cuestiones:

- a) Hallar la función de densidad conjunta de la muestra.  
 b) Decir cuáles de las siguientes funciones son estadísticos:

$$U = X_1 \cdot X_2 \cdots X_n - n\lambda$$

$$V = X_1 + X_2$$

$$W = X_1 + \cdots + X_n$$

$$Z = \max\{X_1, \dots, X_n\}$$

y hallar la distribución de  $W$  y  $Z$ .

- c) ¿Cuál es la característica muestral correspondiente a  $\lambda$ ?

### Solución

- a)  $f_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}$ ,  $x = 0, 1, 2, \dots$  luego la densidad de una muestra  $x_1, \dots, x_n$  es

$$f(x_1, \dots, x_n) = e^{-n\lambda} \frac{\lambda^{\sum x_i}}{x_1! \cdots x_n!}.$$

- b)  $U$  no es un estadístico porque depende de  $\lambda$  y su valor dada una muestra  $x_1, \dots, x_n$ , no puede ser calculado si  $\lambda$  es desconocido.

$V$  es estadístico. Sin embargo, es un estadístico que no utiliza toda la información de la muestra.

$W$  y  $Z$  son también estadísticos. Como la suma de v.a. Poisson independientes es también una Poisson, la función de densidad de  $W$  es

$$f_W(k) = e^{-n\lambda} \frac{(n\lambda)^k}{k!} \quad k = 0, 1, 2, \dots$$

Por otra parte, como  $X_1, \dots, X_n$  son independientes,

$$\begin{aligned} P(Z \leq k) &= P(\max\{X_1, \dots, X_n\} \leq k) \\ &= P([X_1 \leq k] \cap \cdots \cap [X_n \leq k]) \\ &= P(X_1 \leq k) \cdots P(X_n \leq k) = [F_X(k)]^n, \end{aligned}$$

siendo  $F_X(k) = \sum_{i \leq k} e^{-\lambda} \frac{\lambda^i}{i!}$ , luego la densidad de  $Z$  es

$$f_Z(k) = [F_X(k)]^n - [F_X(k-1)]^n.$$

- c) Como

$$\lambda = E(X) = \text{var}(X)$$

tanto  $\bar{X}_n$  como  $S_n^2$  son características muestrales relacionadas con  $\lambda$ . Veremos más adelante que  $\bar{X}_n$  es más eficiente en problemas de inferencia sobre  $\lambda$ .

- 1.2. Una v.a. tiene una distribución  $N(m, \sigma^2)$ . Hallar la función de densidad conjunta de una muestra aleatoria simple de tamaño  $n$  y calcular  $\bar{x}_n$ ,  $s_n^2$  y  $a_3$  para la muestra

1,74 0,45 2,52 1,19 1,24 2,68 3,51 1,83 1,00 0,87.

**Solución**

$$f_X(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}$$

$$f(x_1, \dots, x_n) = \frac{1}{(\sigma \sqrt{2\pi})^n} e^{-\sum_{i=1}^n \frac{(x_i-m)^2}{2\sigma^2}}$$

$$n = 10 \quad \sum_{i=1}^n x_i = 17,03 \quad \sum_{i=1}^n x_i^2 = 37,14 \quad \sum_{i=1}^n x_i^3 = 95,23$$

$$\bar{x}_n = 1,703 \quad s_n^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_n^2 = 0,814 \quad a_3 = 9,523.$$

- 1.3. Hallar la distribución exacta del estadístico media muestral  $\bar{X}_n$  en los casos de que la distribución de la variable  $X$  sea:

- Bernoulli de parámetro  $p$ .
- Exponencial de parámetro  $\alpha$ .
- Cauchy.
- Normal  $N(m, \sigma^2)$ .

¿Cuál es la distribución asintótica de  $\bar{X}_n$  si la distribución  $X$  es desconocida?

**Solución**

- a)  $X$  toma el valor 1 con probabilidad  $p$  y el valor 0 con probabilidad  $1-p$ . Una muestra de  $X$  estará formada por una sucesión de ceros y unos. Luego,  $\bar{X}_n$  será la frecuencia relativa de unos en una muestra de tamaño  $n$  y su función de densidad será:

$$f_{\bar{X}_n}\left(\frac{k}{n}\right) = \binom{n}{k} p^k (1-p)^{n-k} \quad \frac{k}{n} = 0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1.$$

- b) Por las propiedades de la distribución exponencial  $S = X_1 + \dots + X_n$  sigue la distribución gamma  $G(\alpha, n)$ , con función de densidad

$$f_S(x) = \frac{\alpha^n}{\Gamma(n)} e^{-\alpha x} x^{n-1} \quad \text{si } x > 0,$$

luego, la función de densidad de  $\bar{X}_n = S/n$  es

$$f_{\bar{X}_n}(x) = n f_S(nx) = n \frac{\alpha^n}{\Gamma(n)} e^{-\alpha nx} (nx)^{n-1} \quad \text{si } x > 0.$$

- c) Si  $X$  sigue la distribución de Cauchy, entonces  $\bar{X}_n$  sigue también la distribución de Cauchy (véase problema 10.6., volumen 1).

d)  $\bar{X}_n$  es  $N\left(m, \frac{\sigma^2}{n}\right)$ .

Finalmente, si  $X$  posee esperanza  $m$  y varianza  $\sigma^2$ , por el teorema central del límite, la distribución asintótica de  $\bar{X}_n$  (distribución para  $n$  grande) es  $N(m, \sigma^2/n)$ .

- 1.4. Los siguientes números: 0,8663, 0,9948, 0,0367, 0,1918, 0,2859, 0,1493, 0,2187, 0,6653, 0,8778 y 0,7912 han sido extraídos al azar e independientemente del intervalo (0, 1). Generar una muestra aleatoria simple del mismo tamaño de una v.a.  $X$  con distribución exponencial de parámetro  $\alpha = 2$ .

**Solución.** Las funciones de densidad y de distribución son:

$$f(x) = 2e^{-2x} \quad \text{si } x > 0,$$

$$F(x) = 1 - e^{-2x} \quad \text{si } x > 0.$$

Si  $Z$  es uniforme en (0, 1), se verifica que  $X = F^{-1}(Z)$  es exponencial de parámetro  $\alpha = 2$ . Entonces

$$Z = 1 - e^{-2X} \Rightarrow e^{-2X} = 1 - Z \Rightarrow X = -\ln(1 - Z)/2.$$

Aplicando esta transformación inversa obtenemos la muestra:

$$1,0062, \quad 2,6295, \quad 0,0187, \quad 0,1064, \quad 0,1683$$

$$0,0808, \quad 0,1234, \quad 0,5472, \quad 1,051, \quad 0,7832.$$

- 1.5. La distribución de una v.a.  $X$  es uniforme en el intervalo (0, 1). En muestras aleatorias simples de tamaño  $n$  hallar la distribución asintótica del estadístico «media geométrica»

$$G_n = (X_1 \cdot X_2 \cdot \dots \cdot X_n)^{1/n}.$$

**Solución**

$$Y_n = \ln G_n = \frac{1}{n} (\ln X_1 + \ln X_2 + \dots + \ln X_n)$$

$$E(\ln X_i) = \int_0^1 \ln x \, dx = -1$$

$$E(\ln^2 X_i) = \int_0^1 \ln^2 x \, dx = 2$$

$$\text{var}(\ln X_i) = 2 - (-1)^2 = 1.$$

Luego cada v.a.  $\ln X_i$  tiene un valor medio  $m = -1$  y varianza  $\sigma^2 = 1$ ; por lo tanto, el valor medio de  $Y_n$  es también  $-1$  y la varianza es  $1/n$ . Por el teorema central del límite, la distribución de  $Y_n$  es asintóticamente normal  $N(-1, 1/n)$ , es decir, para  $n$  grande

$$F_{Y_n}(\tilde{y}) \simeq \int_{-\infty}^{\tilde{y}} \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-n(y+1)^2/2} \, dy.$$

Además,

$$F_{G_n}(x) = P(G_n \leq x) = P(Y_n \leq \ln x) = F_{Y_n}(\ln x)$$

$$f_{G_n}(x) = f_{Y_n}(\ln x) \cdot \frac{1}{x} \quad 0 < x < 1.$$

Luego, para  $n$  grande, la función de densidad asintótica de  $G_n$  es

$$f(x) = \frac{\sqrt{n}}{\sqrt{2\pi} x} e^{-n(\ln x + 1)^2/2} \quad 0 < x < 1.$$

**1.6.** El porcentaje de individuos daltónicos varones de una población es  $P$ . Se desea estimar el porcentaje de la población a partir del porcentaje observado sobre una muestra de tamaño  $n$ . Calcular el tamaño de la muestra a fin de que el error cometido sea inferior al 1% con probabilidad 0,90 en los casos:

- a) Se sabe que  $P$  es inferior al 16%.
- b) No se sabe nada acerca de  $P$ .

**Solución.** Podemos suponer  $n \geq 30$ . Entonces, la frecuencia  $f$  del número de daltónicos sobre una muestra de  $n$  individuos es aproximadamente normal  $N(np, npq)$ , siendo  $p = P/100$ . Por lo tanto, el porcentaje de daltónicos ( $100f/n$ ) es aproximadamente  $N(P, 100^2 pq/n)$ . Luego, con probabilidad 0,90 tendremos:

$$P - 1,64 \times 100 \sqrt{pq/n} < \hat{P} < P + 1,64 \times 100 \sqrt{pq/n},$$

siendo  $\hat{P} = 100f/n$ .

- a) Si se sabe que  $p < 0,16$ , el máximo de  $pq$  se alcanza para  $p = 0,16$ ,  $q = 1 - 0,16$ . Tomemos pues  $p = 0,16$ ,  $q = 0,84$ . Como el error debe ser inferior al 1%, hacemos:

$$1,64 \times 100 \sqrt{0,16 \times 0,84/n} = 1 \Rightarrow n = (1,64 \times 100)^2 \times 0,16 \times 0,84$$

$$= 3.614,82$$

**Primera respuesta:**  $n = 3.615$  individuos.

- b) Si no se tiene información sobre  $P$ , estudiemos en qué casos  $pq = p(1-p)$  es máximo. Derivando e igualando a cero:

$$\frac{d}{dp} [p(1-p)] = 1 - 2p = 0 \Rightarrow p = 1/2$$

$$1,64 \times 100 \sqrt{0,5 \times 0,5/n} = 1 \Rightarrow n = (1,64 \times 100)^2 \times 0,5 \times 0,5$$

$$= 6.724$$

**Segunda respuesta:**  $n = 6.724$  individuos.