

# ÍNDICE

<i>Presentación</i> .....	21
---------------------------	----

## I MÉTODOS NUMÉRICOS

<i>Capítulo 1. AJUSTE DE FUNCIONES CON POLINOMIOS: TÉCNICAS DE COLOCACIÓN Y DE MÍNIMOS CUADRADOS</i> .....	31
1.1. Introducción.....	32
A. <i>Polinomios de colocación</i> .....	34
1.2. Ajustes con polinomios de colocación.....	34
Opciones de ajuste polinómico.....	35
El criterio de colocación: casos simples.....	36
Observaciones de interés.....	38
1.3. La tabla de diferencias y los polinomios de Newton.....	39
El polinomio de avance de Newton.....	40
El polinomio de retroceso de Newton.....	42
Observaciones prácticas.....	43
1.4. El polinomio de Lagrange.....	44
1.5. Otras técnicas.....	46
B. <i>Mínimos cuadrados</i> .....	46
1.6. Concepto y aplicación al caso lineal.....	46
Estudio del caso lineal: determinación de los coeficientes.....	47
Unicidad de la solución.....	48
El carácter de mínimo.....	49
La «bondad» del ajuste.....	50
La utilidad extendida del caso lineal.....	51
Nota adicional sobre el error.....	56
1.7. Ajustes de mínimos cuadrados de orden superior.....	56
El caso cuadrático.....	57

El caso general.....	58
Observaciones prácticas.....	59
Bibliografía .....	61
Problemas teóricos y numéricos.....	62
<b>Capítulo 2. AJUSTE DE FUNCIONES CON POLINOMIOS ORTOGONALES.....</b>	<b>81</b>
2.1. Introducción.....	82
2.2. El caso discreto: Polinomios de Gram-Tschebyscheff.....	84
El sistema normal de ecuaciones .....	85
Forma de los polinomios de Gram-Tschebyscheff .....	86
2.3. El caso continuo: Producto escalar y distancia entre funciones.....	88
Producto escalar de funciones .....	89
Criterios de aproximación entre funciones.....	93
Desarrollos en serie de una base completa .....	93
El cálculo de los coeficientes del desarrollo.....	94
Observaciones de interés.....	96
2.4. Caso continuo: polinomios de Legendre.....	100
Ortogonalización constructiva de Gram-Schmidt .....	100
Forma de los polinomios normalizados de Legendre.....	102
Forma habitual de los polinomios de Legendre.....	104
Propiedades adicionales.....	105
2.5. Caso continuo: polinomios de Tschebyscheff.....	107
Definición.....	107
Propiedades adicionales.....	109
La economización de polinomios .....	111
Observaciones de interés.....	113
2.6. Caso continuo: polinomios de Hermite y de Laguerre.....	114
Bibliografía .....	117
Problemas teóricos y numéricos.....	118
<b>Capítulo 3. APLICACIONES NUMÉRICAS BÁSICAS.....</b>	<b>129</b>
3.1. Los errores en el cálculo numérico .....	130
Errores absoluto y relativo .....	131
El error de redondeo y conceptos asociados.....	132
Errores de entrada y cifras significativas «físico-químicas».....	135
Consideraciones adicionales.....	137

3.2. Interpolación y extrapolación.....	139
Observaciones prácticas en interpolación: elección de grado, selección de puntos de la tabla y tipo de polinomio, tabla desigualmente espaciada.....	141
Notas complementarias.....	145
El error de interpolación.....	146
3.3. Propagación de los errores en los datos de entrada.....	149
Alternancias de signo en una tabla de diferencias.....	150
Errores de entrada.....	151
3.4. Diferenciación numérica.....	152
Fórmulas de Newton.....	153
Fórmulas de Stirling.....	155
Extrapolación de Richardson.....	157
3.5. Integración numérica.....	159
Regla del trapecio.....	159
Regla de Simpson.....	162
Técnicas Gaussianas: Gauss-Legendre, Gauss-Hermite y Gauss-Laguerre.....	163
Tratamiento de integrales singulares.....	168
Tratamiento de integrales oscilantes.....	170
Complementos: Tablas para integración Gaussiana.....	173
Bibliografía.....	176
Problemas teóricos y numéricos.....	177
<b>Capítulo 4. RESOLUCIÓN NUMÉRICA DE ECUACIONES Y SISTEMAS.....</b>	<b>201</b>
4.1. Conceptos preliminares.....	202
Raíces (ceros) de ecuaciones no lineales.....	203
Sistemas de ecuaciones y diagonalización.....	205
A. <i>Ecuaciones no lineales</i> .....	206
4.2. Separación de raíces reales y estimación del error.....	206
4.3. Método de bisección.....	208
4.4. Método de la falsa posición ( <i>regula falsi</i> ).....	210
4.5. Método de Newton-Raphson.....	214
Definición del algoritmo.....	214
Condiciones suficientes de convergencia.....	215
Estimación del error.....	216

La variante Newton-secante .....	217
4.6 Método iterativo de punto fijo.....	218
4.7 El caso de las raíces múltiples.....	220
Métodos para determinar la multiplicidad .....	221
<i>B. Sistemas de ecuaciones</i> .....	223
4.8. Sistema lineal (no homogéneo).....	223
Método de Gauss con pivote .....	225
Estimación del error .....	227
4.9. Sistema no lineal.....	228
Método de Newton-Raphson.....	228
Método del gradiente .....	229
Bibliografía .....	231
Problemas teóricos y numéricos.....	232

## II

### INTRODUCCIÓN A LA TEORÍA Y APLICACIONES DE LA ESTADÍSTICA

<i>Capítulo 5. DISTRIBUCIONES DE PROBABILIDAD</i> .....	257
5.1. Probabilidad, Estadística y Química.....	258
El concepto de probabilidad .....	258
Breve presentación axiomática de la probabilidad .....	261
Otras observaciones y las aplicaciones en la Química.....	267
5.2. Variables aleatorias, población y muestra.....	270
5.3. Funciones de distribución de probabilidades .....	274
Variables monodimensionales (discretas y continuas) .....	274
Variables monodimensionales derivadas.....	281
5.4. Caracterización de una distribución de probabilidad.....	284
Valor medio y desviación típica (estándar).....	284
Momentos de una distribución.....	286
Medidas de asimetría y de exceso.....	288
Otros parámetros .....	289
5.5. Ejemplos de distribuciones discretas.....	291
La distribución binomial.....	291
La distribución de Poisson .....	294

La distribución multinomial.....	296
5.6. Ejemplos de distribuciones continuas.....	298
La distribución uniforme.....	298
La distribución Gaussiana (normal).....	300
La distribución logarítmico-normal (log-normal).....	305
5.7. Composición de variables aleatorias.....	307
Valores medios y varianzas de funciones aleatorias.....	307
Suma y producto de variables aleatorias.....	310
Distribuciones de probabilidad en $n$ dimensiones.....	315
Bibliografía.....	321
Problemas teóricos y numéricos.....	322
<i>Capítulo 6. MUESTREO, ESTIMACIÓN Y DECISIÓN ESTADÍSTICA</i> .....	341
6.1. Muestreo de poblaciones.....	342
Métodos generales de muestreo.....	343
Observaciones adicionales.....	345
6.2. Distribuciones muestrales.....	347
Media y varianza.....	347
Proporciones.....	351
Sumas y diferencias.....	351
Mediana.....	353
6.3. Inferencia estadística (I).....	354
Estimación por un punto.....	355
Estimación por intervalos de confianza.....	356
6.4. Inferencia estadística (II): formulación y verificación de hipótesis estadísticas.....	360
Cinco pasos a dar en hipótesis estadísticas.....	360
Observaciones adicionales.....	363
Principios de admisión y rechazo de hipótesis.....	365
6.5. Función de potencia y curva OC.....	366
6.6. Gráficos de control (Shewhart) y aleatoriedad.....	368
6.7. Comparación de muestras: medias y proporciones.....	371
6.8. Teoría de pequeñas muestras.....	374
Distribución $t$ de Student.....	375
Distribución chi-cuadrado.....	379
Distribución $F$ de Fisher.....	382

Bibliografía .....	387
Problemas teóricos y numéricos.....	388
<b>Capítulo 7. CORRELACIÓN, REGRESIÓN Y ESTADÍSTICA NO PARAMÉTRICA ...</b>	<b>407</b>
7.1. Experimentos con más de una variable aleatoria, correlación y regresión.....	408
7.2. Ecuaciones empíricas típicas en dos variables y su reducción a forma lineal.....	411
Tipos básicos con dos parámetros.....	412
Tipos con tres y cuatro parámetros.....	413
7.3. El coeficiente de correlación en dos variables.....	418
Correlación de poblaciones.....	418
Correlación lineal en muestras bivariantes.....	420
El coeficiente $r$ como estimador estadístico .....	425
7.4. Aspectos prácticos de la regresión lineal por mínimos cuadrados	430
7.5. Desestimación de puntos en el análisis de datos.....	434
Test de cuartiles con extensión «(box-and-whisker plot)».....	435
Test de distancias.....	436
7.6. Correlación lineal múltiple.....	437
7.7. Estadística no paramétrica .....	440
Test de signos .....	441
Correlación por rangos de Spearman .....	443
Bibliografía .....	447
Problemas teóricos y numéricos.....	448

### III. ANÁLISIS Y PROPAGACIÓN DE LOS ERRORES EXPERIMENTALES

<b>Capítulo 8. EL TRATAMIENTO DE ERRORES EN DATOS EXPERIMENTALES.....</b>	<b>475</b>
8.1. Introducción.....	476
8.2. Los errores en la medición experimental.....	478
8.3. Propagación del error de escala del aparato .....	480
8.4. Propagación de los errores sistemáticos .....	482

8.5. Propagación de los errores accidentales .....	485
Variables independientes .....	485
Variables dependientes.....	488
La inducción de errores sistemáticos.....	489
8.6. Un caso de estudio: cálculo del error total de un índice de refracción.....	491
Bibliografía .....	494
Problemas teóricos y numéricos.....	495

## IV

## SIMULACIÓN DE PROCESOS Y VALIDACION DE MÉTODOS

<i>Capítulo 9. MÉTODOS AVANZADOS DE CÁLCULO Y DE SIMULACIÓN</i>	
<b>NUMÉRICA</b> .....	511
<i>A. La aproximación trigonométrica</i> .....	513
9.1. Polinomios trigonométricos.....	513
Cambios de variable.....	514
Ortogonalidad en el caso de número impar de puntos .....	515
Ortogonalidad en el caso de un número par de puntos.....	516
Relaciones útiles.....	516
Cálculo de los coeficientes .....	517
Expresiones finales.....	519
<i>B. Simulación numérica de procesos deterministas</i> .....	519
9.2. Ecuaciones diferenciales: generalidades.....	519
9.3. Ecuaciones diferenciales ordinarias.....	522
Casos de estudio.....	522
Existencia y unicidad de la solución .....	523
9.4. Ecuación diferencial de primer orden y primer grado (valor inicial).....	524
Método de Euler .....	524
Estabilidad y error.....	525
Predictor-corrector de Euler.....	526
Métodos de Runge-Kutta.....	527
9.5. Ecuación diferencial de segundo orden (valores iniciales).....	531
Método de Runge-Kutta (IV).....	531

Método predictor-corrector de Adams.....	532
9.6. Problemas de valores de frontera.....	532
<i>C. Diagonalización numérica de matrices reales y simétricas</i> .....	533
9.7. Conceptos generales.....	533
Teorema básico para matrices reales y simétricas .....	534
Multiplicidad de raíces y degeneración.....	535
Observaciones prácticas.....	536
9.8. Método del polinomio característico: cálculo de autovectores.....	537
Caso no degenerado .....	537
Caso degenerado.....	537
9.9. Método de Jacobi.....	542
La transformación ortogonal .....	542
La construcción de la matriz ortogonal $O$ .....	543
Observaciones prácticas.....	547
9.10. Tests de diagonalización y técnicas complementarias .....	548
Bibliografía .....	551
Problemas teóricos y numéricos.....	552
 <i>Capítulo 10. MÉTODOS ESTADÍSTICOS DE SIMULACIÓN Y VALIDACIÓN</i> .....	589
<i>A. Integración numérica multidimensional</i> .....	590
10.1. Integración Monte Carlo .....	590
Aspectos numéricos: familias multiplicativas congruentes.....	591
Aspectos estadísticos: el error de integración.....	595
<i>B. Aplicaciones de los procesos de minimización</i> .....	596
10.2. Promedios con pesos muestrales.....	596
10.3. Ajuste lineal chi-cuadrado de datos.....	600
Aspectos numéricos .....	601
Aspectos estadísticos.....	603
Observaciones adicionales.....	604
10.4. Ajuste de datos a distribuciones de probabilidad.....	606
Caso continuo: ajuste Gaussiano .....	606
Caso discreto: ajuste binomial.....	609
10.5. Estadística robusta: ajuste de una línea recta.....	611
<i>C. Análisis de la varianza</i> .....	615

10.6. Homogeneidad de un conjunto de varianzas muestrales .....	616
10.7. Homogeneidad de un conjunto de medias (ANOVA-1) .....	617
Estimación entre muestras .....	618
Estimación dentro de la muestra.....	619
Observaciones adicionales.....	619
10.8. Análisis de la varianza con dos factores de variación indepen- dientes (ANOVA-2) .....	621
Caso de dos efectos fijos.....	623
Caso de dos efectos aleatorios .....	624
10.9. Análisis de la varianza en ajustes de regresión.....	625
Bibliografía .....	626
Problemas teóricos y numéricos.....	627
<i>Apéndice I: Tratamiento de datos experimentales mediante computa- ción (Modelos de Prácticas en Centros Asociados).....</i>	<i>647</i>
<i>Apéndice II: La base ortogonal de Fourier.....</i>	<i>651</i>
<i>Apéndice III: Tablas estadísticas .....</i>	<i>665</i>
<i>Bibliografía general .....</i>	<i>669</i>
<i>Glosario de términos .....</i>	<i>679</i>
<i>Índice alfabético de materias .....</i>	<i>703</i>

### C. ANÁLISIS DE LA VARIANZA

En esta última sección se consideran cuestiones complementarias importantes en el tratamiento de datos, como son las relacionadas con la *homogeneidad* y con ellas se generaliza lo ya visto en el análisis de muestras anteriormente. Ya se mencionó al discutir la distribución de Fisher (Cap.6) que ésta constituía la base de una técnica muy poderosa conocida como el *análisis de la varianza* (ANOVA). Esta es una técnica perfectamente adecuada para realizar comparaciones entre varias series de datos que pudieran haberse obtenido bajo diferentes condiciones, equipamientos, calidades o purezas de los productos empleados, etc. La potencia de ANOVA radica en que permite estudiar simultáneamente estos diversos *factores* de variación que influyen en los datos obtenidos, separando sus efectos particulares, tanto si son puramente *aditivos* (efectos independientes) como si *interaccionan* entre sí. Estos factores de variación se clasifican en *controlables* (o de efecto fijo) y *no controlables* (aleatorios). Ejemplos típicos de factores de variación, en principio, controlables son los equipos o máquinas de medición o producción, los operarios que realizan las mediciones, y las determinadas condiciones experimentales seguidas para la obtención de la muestra (los posibles agentes reductores, oxidantes, quelantes, etc.). Por otra parte, ejemplos típicos de factores de variación no controlables son las inhomogeneidades en el sistema del que se toman las muestras y el día en el que se realizan las mediciones (por los cambios aleatorios en la presión, temperatura, etc.).

Es interesante aquí fijarse en que además de los factores anteriores siempre va a estar presente el *error aleatorio del proceso de medición* (factor aleatorio o residual) como un factor o fuente de inevitables variaciones en los datos tomados. ANOVA también permite evaluar el efecto de este último factor. Incluso ANOVA puede utilizarse ventajosamente en la decisión del mejor ajuste de regresión entre varios de ellos. La complejidad que puede alcanzarse con ANOVA es muy grande y de ahí que goce hoy de una popularidad creciente en las ciencias experimentales (¡fue introducida por Fisher entre 1937 y 1950!). Aquí se van a considerar sólo algunas aplicaciones comunes y sencillas de ANOVA, remitiendo al lector, como de costumbre, a la bibliografía especializada para más detalles.

La aplicación significativa de ANOVA a un conjunto de datos  $\{x_{ij}\}$  procedentes de varias muestras ( $i =$  muestra,  $j =$  dato) requiere de ciertas condiciones generales:

- El carácter Gaussiano de la distribución de probabilidad de la que se hayan extraído las muestras, para poder aplicar con fiabilidad el test  $F$  de Fisher.
- La homogeneidad de las varianzas de todas las medidas analizadas, pues en el análisis se mezclan todas estas medidas.
- Las variaciones debidas a factores no controlados deben ser verdaderamente aleatorias.
- Los errores aleatorios tienen que distribuirse de forma aproximadamente Gaussiana.

Como una parte importante es la homogeneidad de las varianzas se va a comenzar presentando un criterio que la garantiza.

### 10.6. Homogeneidad de un conjunto de varianzas muestrales

Dado un conjunto de  $k$  varianzas muestrales  $\{s_1^2, s_2^2, \dots, s_k^2\}$  cada una fijada con un número diferente de mediciones  $\{N_1, N_2, \dots, N_k\}$  extraídas de una población normal, su homogeneidad se puede analizar con el estadístico  $B_s$  de Bartlett

$$B_s = \frac{1}{c} \left[ N \ln S^2 - \sum_{i=1}^k (N_i - 1) \ln s_i^2 \right] \quad (10.6.1)$$

en donde suele exigirse que  $N_i \geq 5$  y cada símbolo es

$$N = \sum_{i=1}^k (N_i - 1) \quad (10.6.2)$$

$$S^2 = \frac{1}{N} \sum_{i=1}^k (N_i - 1) s_i^2 \quad (10.6.3)$$

$$c = 1 + \frac{-\frac{1}{N} + \sum_{i=1}^k \frac{1}{N_i - 1}}{3(k - 1)} \quad (10.6.4)$$

El estadístico  $B_s$  sigue aproximadamente una distribución  $\chi^2$  con  $\nu = k - 1$  grados de libertad. Una vez seleccionado el nivel de significación unilateral  $\alpha$  ( $= 0,05, 0,01, \text{etc.}$ ) se ensaya la hipótesis nula como es habitual. Así, para

$H_0$  : el conjunto  $\{s_1^2, s_2^2, \dots, s_k^2\}$  es homogéneo y todas las varianzas proceden de una misma población  $(\mu, \sigma^2)$

$$\langle s_1^2 \rangle = \langle s_2^2 \rangle = \dots = \langle s_k^2 \rangle = \sigma^2 \quad (10.6.5)$$

se tiene la regla de decisión siguiente

- Si  $B_s < \chi_{1-\alpha}^2(k - 1)$ , se acepta  $H_0$  al nivel  $\alpha$ .
- Si  $B_s > \chi_{1-\alpha}^2(k - 1)$  se rechaza  $H_0$  al nivel  $\alpha$ .

Este criterio puede aplicarse al caso de diferentes poblaciones con la misma  $\sigma^2$ .

## 10.7. Homogeneidad de un conjunto de medias (ANOVA-1)

En un experimento con un único *factor* de variación, controlable o no, cuando se dispone de un conjunto de medias  $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k\}$  cada una de ellas fijada con una serie de  $N$  mediciones extraídas de una población normal,  $x_{ij}$  ( $i = 1, 2, \dots, k; j = 1, 2, \dots, N$ ), su homogeneidad puede evaluarse con un estadístico  $F$  de Fisher definido como

$$F = \frac{s_n^2}{s_d^2} = \frac{\frac{N}{k-1} \sum_{i=1}^k (\bar{x}_i - \bar{x})^2}{\frac{1}{k(N-1)} \sum_{i=1}^k \sum_{j=1}^N (x_{ij} - \bar{x}_i)^2}; \quad s_n^2 > s_d^2 \quad (10.7.1)$$

en donde  $\bar{x}$  es la media global de todas las  $\bar{x}_i$

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k \bar{x}_i \quad (10.7.2)$$

**Tabla 1. ANOVA-1**

Muestra	Mediciones				Medias
1	$x_{11}$	$x_{12}$	.....	$x_{1N}$	$\bar{x}_1$
2	$x_{21}$	$x_{22}$	.....	$x_{2N}$	$\bar{x}_2$
.....	.....	.....	.....	.....	.....
$k$	$x_{k1}$	$x_{k2}$	.....	$x_{kN}$	$\bar{x}_k$
Media global					$\bar{x}$

Este estadístico sigue la distribución de Fisher con  $\nu_n = k - 1$  y  $\nu_d = k(N - 1)$  grados de libertad. Fijado el nivel de significación unilateral  $\alpha$  la verificación de la hipótesis nula se realiza de la forma habitual. Esquemáticamente

$H_0$  : el conjunto  $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k\}$  es homogéneo y todas las medias proceden de una misma población  $(\mu, \sigma^2)$

$$\langle \bar{x}_1 \rangle = \langle \bar{x}_2 \rangle = \dots = \langle \bar{x}_k \rangle = \mu \tag{10.7.3a}$$

$$\langle s_1^2 \rangle = \langle s_2^2 \rangle = \dots = \langle s_k^2 \rangle = \sigma^2 \tag{10.7.3b}$$

se tiene la regla de decisión siguiente

- Si  $F < F_{\nu_1, \nu_2}^C(\alpha)$ , se acepta  $H_0$  al nivel  $\alpha$ .
- Si  $F > F_{\nu_1, \nu_2}^C(\alpha)$ , se rechaza  $H_0$  al nivel  $\alpha$ .

El procedimiento anterior contiene de manera compacta lo que se conoce como la aplicación ANOVA-1 y es muy instructivo analizar la naturaleza de  $F$  que debe ser un cociente entre dos varianzas como se indica en (10.7.1).

Como la población es común se puede estimar  $\sigma^2$  de dos maneras significativas: *entre muestras* y *dentro de la muestra global*. Estos dos cálculos van a dar origen al numerador y al denominador de (10.7.1) respectivamente y se detallan a continuación.

i) *Estimación entre muestras*

Ya se vio en el Cap. 6 que la varianza de una media muestral se relaciona con la varianza poblacional como

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{N} \tag{10.7.4}$$

Por otra parte, en el supuesto de que  $H_0$  fuera cierta es claro que la varianza de las medias muestrales  $s_{\bar{x}}^2$  es también un estimador de  $\sigma_{\bar{x}}^2$  con  $k - 1$  grados de libertad

$$s_{\bar{x}}^2 = \frac{1}{k-1} \sum_{i=1}^k (\bar{x}_i - \bar{x})^2 \approx \frac{\sigma^2}{N} \quad (10.7.5)$$

De este modo la estimación de  $\sigma^2$  por esta vía es el numerador de (10.8.1)

$$s_n^2 = \frac{N}{k-1} \sum_{i=1}^k (\bar{x}_i - \bar{x})^2; \quad v = k-1 \text{ grados de libertad} \quad (10.7.6)$$

## ii) Estimación dentro de la muestra

En cada muestra una estimación de la varianza poblacional  $\sigma^2$  está dada por

$$s_i^2 = \frac{1}{N-1} \sum_{j=1}^N (x_{ij} - \bar{x}_i)^2; \quad i = 1, 2, \dots, k \quad (10.7.7)$$

y promediando los  $k$  valores anteriores se tiene otro estimador para  $\sigma^2$  que es justamente el denominador de (10.8.1)

$$s_d^2 = \frac{1}{k(N-1)} \sum_{i=1}^k \sum_{j=1}^N (x_{ij} - \bar{x}_i)^2; \quad v = k(N-1) \text{ grados de libertad} \quad (10.7.8)$$

en donde hay que notar que hay  $k$  muestras cada una con  $N - 1$  grados de libertad.

## Observaciones adicionales

Las dos explicaciones anteriores detallan la construcción de  $F$  (10.7.1) y la regla de decisión asociada. Conviene además notar en este contexto las observaciones que se indican a continuación.

- Al estar fijada con las diferencias entre valores medios la estimación  $s_n^2$  no depende explícitamente de las variaciones dentro de cada muestra.
- Igualmente, la estimación  $s_d^2$  al depender de las diferencias  $x_{ij} - \bar{x}_i$  no depende de las medias muestrales  $\bar{x}_i$  (ni de  $\bar{x}$ ).