

Notas Curso Análisis de datos

Enric Pallas¹, Julio Sheinbaum^{1,*}

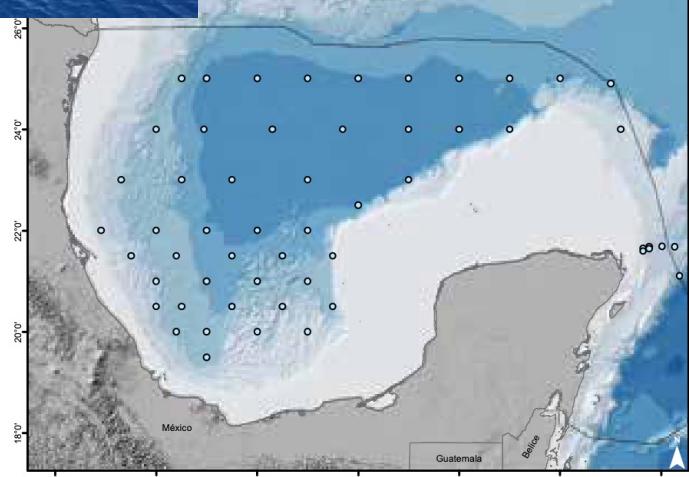
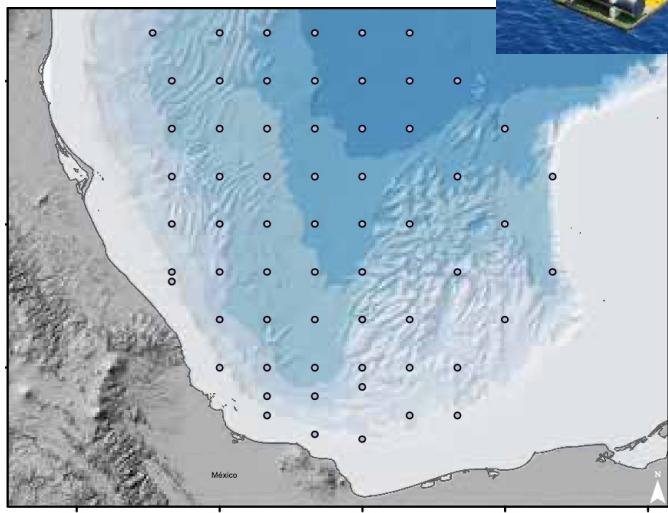
¹Physical Oceanography,, Centro de Investigación Científica y de Educación Superior de Ensenada,
CICESE, Ensenada, Baja California, MEXICO

*Translated template to Quarto.

5 Abstract

6 En el océano conviven una gran cantidad de corrientes de diferentes escalas espaciales
7 y temporales. Las escalas espaciales típicas de la circulación oceánica son la
8 larga escala, la mesoscala, submesoscala, y microescala. La larga escala es del
9 $O(1000\ km)$ y esta determinada por la circulación general en el océano como la ter-
10 mohalina y los grandes giros anticiclónicos de las grandes cuencas oceánicas; las
11 escalas temporales de la larga escala varía entre meses y años. La mesoscala esta
12 definida por corrientes del $O(100\ km)$ como remolinos, corrientes costeras, filamentos,
13 frentes, etc. Son corrientes mas regionales pero pueden tener gran influencia sobre la
14 circulación general o de larga escala. Sus escalas temporales son de semanas a meses.
15 La submesoscala corresponde a corrientes del $O(10\ km)$ de carácter local remolinos,
16 filamentos, frentes, corrientes en playas, puertos, y estuarios. La submesoscala varía
17 temporalmente con rapidez en tiempos que varían de horas a días. Finalmente pode-
18 mos hablar de la microescala que son remolinos del orden de centímetros a metros y
19 generalmente es la escala característica de la turbulencia que transfiere energía desde
20 la submesoscala hacia la disipación molecular. Aquí podemos hablar de fenómenos
21 del orden de segundos y minutos.

Estaciones Hidrográficas



ANCLAJES

1.1.- El proyecto CANEK (CNK)

Correntometría profunda: redes de anclajes

Monitoreo de corrientes y otras variables oceanográficas.

Microcat: Presión, conductividad, temperatura y oxígeno



Termistor: Temperatura



Sensor de presión



Correntómetro puntual

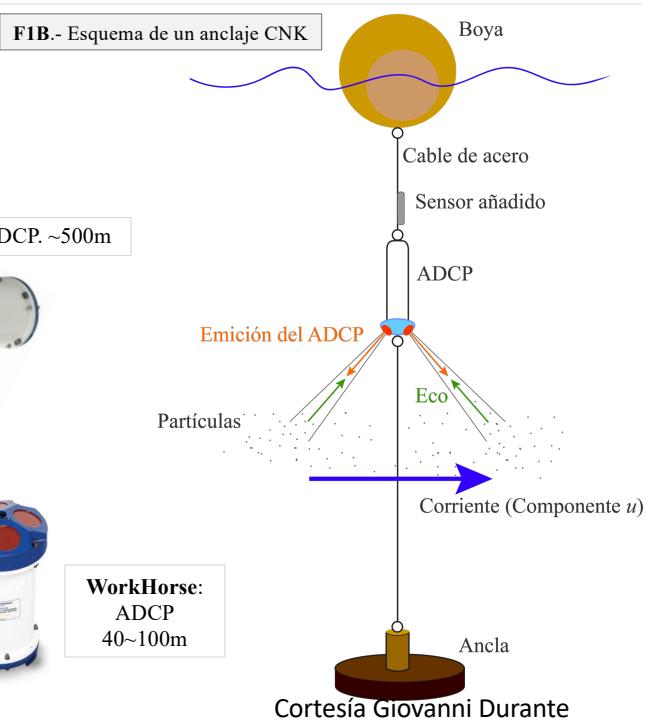


F1A.- Sensores comúnmente instalados en los anclajes CNK

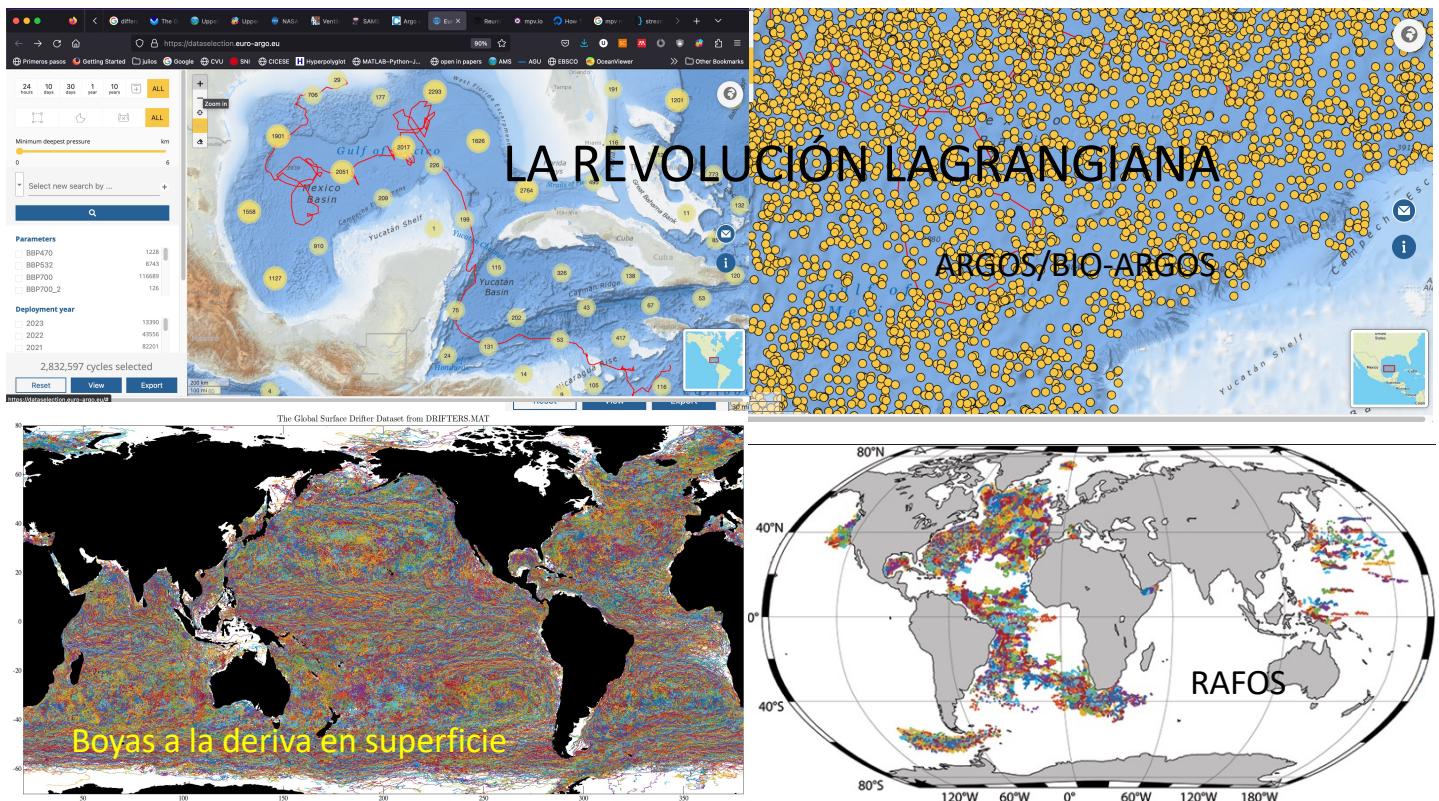
2

1.- Introducción

F1B.- Esquema de un anclaje CNK



Cortesía Giovanni Durante





26 **1 Estadística y conceptos de probabilidad**

27 **1.1 Porqué estudiar la estadística en oceanografía**

28 A pesar de nuestra formación determinista a la hora de resolver problemas matemáticos y aunque consideremos que las ecuaciones de Navier-Stokes que describen el
29 movimiento del océano son deterministas, la estadística es ampliamente utilizada en
30 oceanografía debido a diferentes razones:

- 32 • Para una descripción completa del océano es necesario especificar una gran
33 cantidad de variables, muchas de las cuales son desconocidas. Un ejemplo
34 de ello son las parametrizaciones que se hacen en oceanografía para describir
35 variables que no pueden medirse directamente. Una parametrización no es
36 nadamas que un modelo estadístico que explica la evolución de una variable
37 dependiente de otras variables independientes. Por ejemplo, parametrización
38 del esfuerzo del viento en función del corte vertical o parametrización del coefi-
39 ciente de arrastre en función de la velocidad del viento a 10m de la superficie
40 del océano.
- 41 • El océano es altamente no lineal. La evolución de una cierta variable no se
42 puede estudiar de forma aislada.

43 Ejemplo: Supongamos el término de aceleración horizontal en las ecuaciones
44 de Navier Stokes para fluidos incompresibles,

$$\frac{\partial \mathbf{u}_h}{\partial t} + \mathbf{u}_h \cdot \nabla \mathbf{u}_h$$

45 Como ya sabemos por el curso de Mecánica de Fluidos, la aceleración de un
46 fluido es una derivada material y consta de un término local (aceleración local)
47 y de un término advectivo o aceleración advectiva. En general, esta ecuación
48 no se aplica a partículas de agua individuales.

49 En oceanografía hablamos de continuo. No estamos interesados en las carac-
50 terísticas cinemáticas de las partículas individuales sino en la manifestación
51 promedia del movimiento molecular, es decir, del fluido como un conjunto
52 o continuo. Es decir, asumimos que el fluido es uniforme en el espacio que
53 ocupa sin considerar la estructura molecular.

54 Por ello debemos de promediar de alguna forma para explicar el compor-
55 tamiento conjunto del fluido y no de una partícula de agua específica? Y como
56 se realiza tal promedio? En general, el promediado se realiza de tal forma que
57 nos permite separar la larga escala que trataremos como determinística, de la
58 pequeña escala que consideramos un proceso aleatorio (turbulento). Supong-
59 amos entonces la separación de la velocidad horizontal en una velocidad
60 promedio y una velocidad fluctuante alrededor de la media

$$\mathbf{u}_h = \langle \mathbf{u}_h \rangle + \mathbf{u}'_h$$

61 donde $\langle \rangle$ denotan promedio. Si aplicamos esta descomposición a la compo-
62 nente x de la aceleración obtenemos:

$$\frac{\partial \langle \mathbf{u} \rangle}{\partial t} + \langle \mathbf{u} \rangle \cdot \nabla \langle \mathbf{u} \rangle + \langle \mathbf{u}' \cdot \nabla \mathbf{u}' \rangle$$

$$\frac{\partial \langle \mathbf{u} \rangle}{\partial t} + \nabla \cdot (\langle \mathbf{u} \rangle \langle \mathbf{u} \rangle) + \nabla \cdot \langle \mathbf{u}' \mathbf{u}' \rangle$$

63 <!-- comment

$$\frac{\partial \langle \mathbf{u} \rangle}{\partial t} + \langle \mathbf{u} \rangle \cdot \nabla \langle u \rangle + \langle \mathbf{u}' \cdot \nabla \mathbf{u}' \rangle$$

64 ->

65 Inevitablemente, las pequeñas escalas o fluctuaciones respecto a la larga es-
 66 cala aparecen en la expresión de la aceleración de larga escala. De forma que
 67 la separación que deseamos no es tan simple ya que debemos de conocer la
 68 estadística de la pequeña escala para poder describir la circulación media.

69 El término $\langle \mathbf{u}' \mathbf{u}' \rangle$ se denomina esfuerzo de Reynolds y nos informa de la
 70 correlación entre las componentes fluctuantes (alta frecuencia) de la velocidad.
 71 Por ejemplo, $\langle u' v' \rangle = 0$ significa que no existe i correlación y hablamos de
 72 isotropía. Si $\langle u' v' \rangle < 0$, significa que las fluctuaciones están inversamente
 73 correlacionadas, i.e., anisotropía.

74 Este es un gran problema no resuelto en la oceanografía física. El esfuerzo de
 75 Reynolds aparece porque la advección es no-lineal de tal forma que no pode-
 76 mos estudiar la larga escala sin conocer información de la pequeña escala que
 77 es un proceso aleatorio. Por similitud con el flujo laminar, los términos de
 78 esfuerzo de Reynolds se parametrizan estadísticamente como proporcionales
 79 a los gradientes de velocidad. El factor de proporcionalidad es el coeficiente
 80 de viscosidad, en este caso, turbulento. Es aqui donde utilizar herramientas
 81 estadísticas tiene sentido.

- 82 • No podemos controlar las variables oceanográficas; estan en constante cambio
 83 a medida que el sistema observado evoluciona.

84 Ejemplo: En el océano coexisten mareas, ondas internas, remolinos, turbulen-
 85 cia de pequeña escala,...las cuales enmascaran el fenómeno oceanográfico que
 86 estamos interesados en estudiar. Estos procesos incontrolables por el oceanó-
 87 grafo en ocasiones es útil considerarlos aleatorios y utilizar herramientas
 88 estadísticas para caracterizarlos.

89 Imaginemos que queremos conocer cual es la temperatura superficial promedio
 90 en la bahía de Todos Santos. Una forma de proceder sería promediar todos
 91 los datos de temperatura superficial que disponemos de los últimos 100 años
 92 y promediarnos? Pero, ¿es realmente lo que deseamos? ¿Deberíamos de con-
 93 siderar la estaciones del año y obtener un pormedio para cada estación? ¿Qué
 94 sucede en años Niño, el cual sabemos que afecta la temperatura del océano?
 95 En definitiva, debemos de definir sobre que conjunto de datos vamos a prome-
 96 diar, y dichos promedio va a reflejar efectivamente esa elección.

97 Este ejemplo precisa de la distinción entre lo que consideramos nuestra *señal*
 98 (temperatura media) de los procesos que son *ruido* (Estaciones del año, los
 99 años Niño, ondas internas, etc.). De esta forma, al definir el promedio esta-
 100 mos haciendo explícita la separación entre *señal* y *ruido*. Finalmente, una
 101 vez definido sobre que promediar, existen en literatura una gran cantidad de
 102 herramientas estadísticas que podemos utilizar. Definir *señal* y *ruido*, y deter-
 103 minar sobre que conjunto de datos vamos a calcular el promedio, es una tarea
 104 difícil. Conocer como debemos muestrear el océano también debe hacerse
 105 cuidadosamente.

106 hacerse cuidadosamente.

107 \end{itemize}

108 En oceanografía física se muestrea el océano de forma discontinua, es decir, se ob-
 109 tienen medidas puntuales en el espacio y en el tiempo. Como dijimos anteriormente,
 110 el océano contiene procesos de diferentes escalas espaciales y temporales, nolineales,

111 y aleatorios. Es por ello que es sumamente importante saber escoger el intervalo
 112 de muestreo Δt dependiendo del fenómeno que se quiere muestrear. Debemos de
 113 tener en mente que la frecuencia mas alta que podemos resolver es la frecuencia de
 114 Nyquist:

$$f_N = 1/(2\Delta t).$$

115 Por ejemplo, si medimos a intervalos de $\Delta t = 5$ h podremos como máximo resolver
 116 procesos que ocurren con frecuencia $f_N \leq 1/10$ cph. La frecuencia mas baja que
 117 podemos resolver va a depender de la longitud del registro. A esa frecuencia le lla-
 118 mamos frecuencia fundamental

$$f_0 = 1/(\Delta t N),$$

119 {}donde $T = \Delta t N$ es la duración del muestreo y N es el número de muestras o datos.
 120 En general, debemos de medir suficiente tiempo para registrar varios ciclos del
 121 fenómeno de estudio para tener significancia estadística. Por lo tanto, nuestra res-
 122 olución frecuencial va a depender del intervalo y duración del muestreo. El cociente
 123 $f_N/f_0 = (1/2\Delta t)/(1/N\Delta t) = N/2$ indica el número máximo de componentes de
 124 Fourier que podemos estimar. Una señal periódica se puede descomponer en la suma
 125 de un conjunto (infinito) de funciones oscilatorias de senos y cosenos o componentes
 126 de Fourier. Esto lo veremos en el capítulo~7. A cada muestreo de un fenómeno le
 127 denominamos realización, y a un conjunto de realizaciones se les denomina ensamble.

128 1.2 Estadística básica

129 La estadística trata de describir las características de una población continua a par-
 130 tir de muestras discretas de la misma. Hablamos de población y de muestra de una
 131 población. Si calculamos, por ejemplo, la media de una población, estamos calcu-
 132 lando un {}parámetro}. Cuando calculamos la media de una muestra le llamamos un
 133 estadístico de la población.

134 La estadística nos ayuda a organizar, analizar, presentar datos, y nos da información
 135 de cómo planear la recolección de los mismos, i.e. a muestrear.

136

137 (1) La media:

138 La media de una muestra de N valores $x_i = x_1, x_2, \dots, x_N$ es

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \langle x \rangle. \quad (1)$$

139 La media debe de diferenciarse de la mediana. La media es el momento de orden
 140 cero. La mediana de una población es aquel valor numérico que separa el 50% de val-
 141 ores mas altos del 50% de valores mas bajos. Se puede calcular ordenando de menor
 142 a mayor el conjunto de valores y escoger el valor central si el conjunto de datos es
 143 impar o el promedio de los dos centrales si es par.

144 (2) La varianza:

145 La varianza de un una muestra de N valores x_i es

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 = \langle x'^2 \rangle - \langle x' \rangle^2, \quad (2)$$

146 donde las primas indican fluctuaciones alrededor de la media. La varianza es una
 147 medida de cuán lejos están los diferentes puntos de la muestra de la media de la
 148 población. La varianza es el segundo momento alrededor de la media. Al dividir por
 149 N estamos subestimando la verdadera varianza de la población. Al dividir por $N - 1$
 150 obtenemos un estimador insesgado.\

151 NOTA: el sesgo de un estimador se refiere a la diferencia entre su esperanza matemática
 152 y el valor numérico (real) del parámetro que se estima. Un estimador que no tiene
 153 sesgo se dice insesgado. Por ejemplo, para la media:

$$E[x] - \mu \rightarrow 0$$

$$\bar{x} - \mu \rightarrow 0$$

154 EJERCICIO: Demostrar porqué hay que dividir por $N - 1$ en lugar de N para que la
 155 definición de varianza sea un estimador insesgado.

156 (3) La desviación típica:

157 Es la raíz cuadrada de la varianza. Se suele escribir como σ para referirse a la
 158 población o como s en estadística

$$s = \sqrt{s^2}. \quad (3)$$

159 {(4) Momentos de orden superior:}

160 Podemos definir un momento alrededor de la media como:

$$m_p = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^p = < x'^p >. \quad (4)$$

161 De esta forma m_2 es la varianza, m_3 es la asimetría, y m_4 la curtosis. El momento
 162 m_3 indica la asimetría de la muestra alrededor de la media ($m_3 > 0$ implica dis-
 163 tribución con cola larga en la parte positiva y viceversa). m_4 indica el grado de es-
 164 parcimiento de las muestras alrededor de la media. Una mayor curtosis indica mayor
 165 concentración de puntos alrededor de la media. Los momentos de orden superior
 166 (> 2) se suelen adimensionalizar dividiendo por la desviación estandar:

$$m_3 = \frac{1}{N} \sum_{i=1}^N \left[\frac{x_i - \bar{x}}{\sigma} \right]^3 = < (x/\sigma)'^3 > \quad (5)$$

$$m_4 = \frac{1}{N} \sum_{i=1}^N \left[\frac{x_i - \bar{x}}{\sigma} \right]^4 - 3 = < (x/\sigma)'^4 > - 3 \quad (6)$$

167 donde el factor -3 hace que la curtosis tome el valor cero para una distribución
 168 Normal.

169 Figura 1.1. Distribuciones con $m_3 < 0$ (izquierda) y $m_3 > 0$ (derecha).

170 Figura 1.2. Distribuciones con diferentes grados de curtosis; $m_4 > 0$ (Leptocúrtica),
 171 $m_4 = 0$ (Normal o Mesocúrtica), y $m_4 < 0$ (Platicúrtica).

174 { (5) Covarianza y correlación: }

175 La covarianza entre dos variables x e y puede definirse como un estadístico que
176 relaciona x e y de la siguiente forma

$$C_{xy} = \langle x'y' \rangle = \langle (x - \bar{x})(y - \bar{y}) \rangle = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}).$$

177 La correlación es una covarianza normalizada

$$\rho_{xy} = \frac{C_{xy}}{s_x s_y} = \frac{\langle x'y' \rangle}{\sqrt{\langle x'^2 \rangle \langle y'^2 \rangle}}.$$

178 Consideremos el modelo estadístico lineal de media cero (es una recta que pasa por
179 $(\bar{x}, \bar{y}) = (0, 0)$)

$$\hat{y} = \alpha x,$$

180 donde α es una constante. El error cometido por este estimador se define como el
181 error cuadrático medio

$$\epsilon = \langle (\hat{y} - y)^2 \rangle = \alpha^2 \langle x^2 \rangle + \langle y^2 \rangle - 2\alpha \langle xy \rangle$$

182 y si queremos minimizar dicho error entonces tenemos que encontrar que α es el que
183 provoca que la derivada $\partial\epsilon/\partial\alpha \rightarrow 0$. Es decir

$$\partial\epsilon/\partial\alpha = 2\alpha \langle x^2 \rangle - 2 \langle xy \rangle = 0,$$

184 y el α es

$$\alpha = \frac{\langle xy \rangle}{\langle x^2 \rangle}.$$

185 El error mínimo cuadrado se encuentra substituyendo el valor de α en la expresión
186 del error ϵ de arriba

$$\begin{aligned} \epsilon &= \frac{\langle xy \rangle^2}{\langle x^2 \rangle} + \langle y^2 \rangle - 2 \frac{\langle xy \rangle^2}{\langle x^2 \rangle} = \langle y^2 \rangle \left(\frac{\langle xy \rangle^2}{\langle x^2 \rangle \langle y^2 \rangle} + 1 - 2 \frac{\langle xy \rangle^2}{\langle x^2 \rangle \langle y^2 \rangle} \right) = \\ &= \langle y^2 \rangle (1 - \rho_{xy}^2). \end{aligned}$$

188 Si $\rho_{xy}^2 = 1$ entonces el error es cero, es decir, mínimo error. Opuestamente, si
189 $\rho_{xy}^2 = 0$ entonces el error es igual a la varianza, es decir, máximo error. Si ρ toma
190 valores intermedios, i.e., $\rho_{xy}^2 = 0.5$, entonces el error es $\epsilon = 0.5 \langle y^2 \rangle$, es decir,
191 el error del modelo lineal es un 50% de la varianza. Por lo tanto, la correlación al
192 cuadrado puede definirse también como la eficiencia relativa del estimador \hat{y}^2 o la
193 fracción de varianza explicada por el modelo lineal

$$\rho_{xy}^2 = \frac{\langle \hat{y}^2 \rangle}{\langle y^2 \rangle} = \frac{\text{varianza explicada}}{\text{varianza total}}.$$

194 A este parámetro se le puede encontrar en literatura inglesa como *{skill} del modelo*
195 *lineal*.

196 **2 Probabilidad**197 **2.1 Distribuciones de probabilidad:**

198 La *{función de distribución acumulativa}* $D_x(r)$ se define como la probabilidad que
 199 una variable aleatoria x sea menor o igual a r , es decir, $P(x \leq r)$. Matemáticamente

$$D(x) = \int_{-\infty}^r F(x)dx,$$

200 donde

$$F(x) = \frac{d}{dx} D(x)$$

201 es la función de densidad de probabilidad (PDF, por sus siglas en inglés). La PDF

202 nos informa de la probabilidad que x sea igual a un cierto valor r , $P(x = r)$.

203 Algunas propiedades de $D(x)$:

205 (1) $D(r) \leq D(s)$ if $r \leq s$

206 (2) $D(-\infty) = 0$

207 (3) $D(\infty) = 1$

208 Algunas propiedades de $F(x)$:

209 (1) $F(x) \geq 0$

210 (2) $\int_{-\infty}^{\infty} F(x)dx = 1$

211 La probabilidad que una variable aleatoria x esté contenida en el intervalo $[r, r + dr]$
 212 es la integral de la función de densidad de probabilidad

$$P(r \leq x \leq r + dr) = \int_r^{r+dr} F(x)dx.$$

213 *{}*Ambas definiciones son parecidas aunque no son lo mismo. Para ello veamos el
 214 ejemplo de la suma del lanzamiento de dos dados al aire.

215
 216 La densidad de probabilidad de que la suma de los dos dados sea 7 es máxima y
 217 que sea 2 o 12 es mínima. Este ejemplo describe dos propiedades fundamentales de
 218 funciones de probabilidad discretas: (i) $P(X = x) \geq 0$ y (ii) $\sum P(x) = 1$. La distribu-
 219 ción de probabilidad acumulativa y la función de densidad de probabilidad tienen las
 220 siguientes distribuciones

222 *{Momentos estadísticos de una función de densidad de probabilidad:}*
 223 Los momentos centrados (o alrededor de la media) de una distribución de probabili-
 224 dad se definen como

$$m_r = E[(x - E[x])^r] = \int_{-\infty}^{\infty} (x - \mu)^r F(x)dx.$$

225 Como caso particular, los momentos alrededor del origen (i.e., $\mu = 0$) son:

$$m_r^0 = E[x^r] = \int_{-\infty}^{\infty} x^r F(x)dx.$$

226 *Entonces, los primeros tres momentos centrados se definen como*

$$\begin{aligned} m_0 &= E[(x - E[x])^0] = E[1] = \int_{-\infty}^{\infty} F(x)dx = 1, \\ m_1 &= E[(x - E[x])^1] = E[x] - \mu = \int_{-\infty}^{\infty} (x - \mu)^1 F(x)dx = 0, \\ m_2 &= E[(x - E[x])^2] = E[x^2 + E[x]^2 - 2xE[x]] = E[x^2] + E[x]^2 - 2E[x]E[x] = \\ &E[x^2] - E[x]^2 = \underbrace{E[x^2]}_{\sigma^2} - \mu^2 = \int_{-\infty}^{\infty} (x - \mu)^2 F(x)dx = \sigma^2 - \mu^2. \end{aligned}$$

230 *Los momentos alrededor de cero ($\mu = 0$) tambien pueden ser estandarizados:*

$$m_r^* = m_r / \sigma^r = \frac{E[(x - E[x])^r]}{\underbrace{(E[(x - E[x])^2])^{r/2}}_{\sigma^2}}.$$

231 *Los cuatro primeros momentos estadísticos alrededor de cero estandarizados son:*

$$\begin{aligned} m_1^* &= m_1 / \sigma^1 = \frac{E[(x - \mu)^1]}{(E[(x - \mu)^2])^{1/2}} = \frac{\mu - \mu}{\sqrt{E[(x - \mu)^2]}} = 0, \\ m_2^* &= m_2 / \sigma^2 = \frac{E[(x - \mu)^2]}{(E[(x - \mu)^2])^{2/2}} = 1, \\ m_3^* &= m_3 / \sigma^3 = \frac{E[(x - \mu)^3]}{(E[(x - \mu)^2])^{3/2}}, \\ m_4^* &= m_4 / \sigma^4 = \frac{E[(x - \mu)^4]}{(E[(x - \mu)^2])^{4/2}}. \end{aligned}$$

235 *(1) Distribución uniforme:*

236 *La distribución de probabilidad uniforme viene dada por*

$$\begin{aligned} F(x) &= \frac{1}{b - a}, \quad a \leq x \leq b \\ &= 0, \quad \text{fuera del intervalo} \end{aligned}$$

238 *Se deduce de la expresión de área de un cuadrado:*

$$\text{Area} = \text{base} * \text{altura} = (b - a)F(x) = 1$$

239 *La función de distribución acumulativa es*

$$D(x) = 0, \quad x < a$$

$$\begin{aligned} D(x) &= \frac{x - a}{b - a}, \quad a \leq x \leq b \\ &= 1, \quad x \geq b \end{aligned}$$

242
243 *La media es $\mu = (b + a)/2$ y la varianza es $\sigma^2 = 1/3(a^2 + b^2 + ab)$. Demostración:
244 {}Los momentos estadísticos alrededor del origen de la distribución uniforme son*

$$m_r^0 = E[x^r] = \int_{-\infty}^{\infty} x^r F(x)dx = \int_a^b \frac{x^r}{b - a} dx =$$

245

$$\frac{1}{b-a} \int_a^b x^r dx = \frac{1}{b-a} \left[\frac{x^{r+1}}{r+1} \right]_a^b = \frac{1}{b-a} \left[\frac{b^{r+1}}{r+1} - \frac{a^{r+1}}{r+1} \right] = \frac{b^{r+1} - a^{r+1}}{(b-a)(r+1)}$$

246

y por lo tanto la media es

$$m_1^0 = E(x) = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2},$$

247

y la varianza

$$m_2^0 = E(x^2) = \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(a^2 + b^2 + ab)}{3(b-a)} = \frac{1}{3}(a^2 + b^2 + ab).$$

248

Ejemplo de distribución uniforme: La ruleta rusa. Supongamos que puede tomar 360 valores, es decir, $0 \leq x \leq 360$. Entonces

$$F(x) = \frac{1}{360}, \quad 0 \leq x \leq 360,$$

250

y, por ejemplo, la probabilidad de que la bola caiga entre el 50 y el 360 es

$$P(50 \leq x \leq 360) = \int_{50}^{360} \frac{1}{360} dx = \frac{1}{360} [x]_{50}^{360} = \frac{310}{360} = 0.8611 (\sim 86\%).$$

251

La función de distribución acumulativa es

$$D(x) = 0, \quad x < 0$$

252

$$D(x) = \frac{x}{360}, \quad 0 \leq x \leq 360$$

253

$$D(x) = 1, \quad x \geq 360$$

254

(2) Distribución normal o Gaussiana:

255

La distribución normal es una de las distribuciones mas recurrente en la naturaleza.

256

En general cualquier variable aleatoria medida, especialmente aquellas que son suma de otras variables aleatorias, tiene una distribución normal alrededor de la media

257

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(x-\bar{x})^2}{2\sigma^2} \right\}.$$

258

La distribución de función acumulativa normal se obtiene integrando la expresión de arriba. Para ello vamos a realizar el cambio de variable (que no es nada mas que estandarizar la variable aleatoria x)

$$z = \frac{x - \bar{x}}{\sigma\sqrt{2}}$$

262

y

$$dz = \frac{dx}{\sigma\sqrt{2}},$$

263

de lo que se deduce

$$D(z) = \frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^z \exp\{-z^2\} dz = \frac{1}{\sqrt{\pi}} \int_{-\infty}^z \exp\{-z^2\} dz,$$

264 donde $\frac{2}{\sqrt{\pi}} \int_0^z \exp\{-t^2\} dt = \text{erf}(z)$.

265

266 Los momentos estadísticos alrededor del origen de la función de distribución Normal
267 son

$$m_r^0 = E[x^r] = \int_{-\infty}^{\infty} x^r F(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^r \exp\left\{-\frac{(x-\bar{x})^2}{2\sigma^2}\right\} dx$$

268 Hagamos el cambio de variable

269

$$\begin{aligned} u &= \frac{x-\bar{x}}{\sigma\sqrt{2}} \\ du &= \frac{dx}{\sigma\sqrt{2}} \end{aligned}$$

270 Si substituimos en la expresión de m_r obtenemos

$$m_r^0 = \frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma\sqrt{2}u + \bar{x})^r e^{-u^2} du = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sigma\sqrt{2}u + \bar{x})^r e^{-u^2} du.$$

271 Ejercicio: Deducir los momentos estadísticos de orden 1 y 2 de la distribución
272 Normal, es decir, la media y la varianza. Integrales útiles:

273

$$\begin{aligned} \int e^{-ax^2} dx &= \frac{\sqrt{\pi}}{2\sqrt{a}} \text{erf}(x\sqrt{a}) \\ \int xe^{-ax^2} dx &= -\frac{1}{2a} e^{-ax^2}, \end{aligned}$$

274 donde la función de error se define cómo:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt.$$

275 La función de error cumple las siguientes identidades:

276

$$\text{erf}(0) = \frac{2}{\sqrt{\pi}} \int_0^0 e^{-t^2} dt = 0$$

$$\text{erf}(\infty) = \frac{2}{\sqrt{\pi}} \underbrace{\int_0^{\infty} e^{-t^2} dt}_{\frac{\sqrt{\pi}}{2}} = 1$$

277

$$\text{erf}(-\infty) = \frac{2}{\sqrt{\pi}} \underbrace{\int_0^{-\infty} e^{-t^2} dt}_{-\frac{\sqrt{\pi}}{2}} = -1$$

278

279 $\{\} \{\text{Respuesta}\}: \text{La media y la varianza son}$

280

$$\begin{aligned} m_1^0 &= E(x) = \mu \quad ; \quad m_1 = 0 \\ m_2^0 &= Var(x) = \mu^2 + \sigma^2 \quad ; \quad m_2 = \sigma^2 \end{aligned}$$

281

{ } { Dedución de la media}: El momento de orden 1 centrado es:

$$m_1 = \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} ue^{-u^2} du = -\frac{\sigma\sqrt{2}}{2\sqrt{\pi}} e^{-u^2} \Big|_{-\infty}^{\infty} = 0.$$

282

y alrededor de cero (momentos crudos):

$$\begin{aligned} m_1^0 &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sigma\sqrt{2}u + \mu)e^{-u^2} du = \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} ue^{-u^2} du + \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \mu e^{-u^2} du = \\ &= \underbrace{\frac{\sigma\sqrt{2}}{\sqrt{\pi}} \left(-\frac{1}{2} e^{-u^2} \right)}_0 \Big|_{-\infty}^{\infty} + \frac{\mu}{\sqrt{\pi}} \left(\frac{\sqrt{\pi}}{2} \operatorname{erf}(u) \right) \Big|_{-\infty}^{\infty} = \frac{\mu}{2} [1 - (-1)] = \mu, \end{aligned}$$

283

{ } { Dedución de la varianza}: El momento de orden 2 centrado es:

$$m_2 = \frac{(\sigma\sqrt{2})^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} u^2 e^{-u^2} du$$

284

$$\begin{aligned} x &= u \rightarrow dx = du \\ dy &= ue^{-u^2} \rightarrow y = \frac{1}{2} e^{-u^2} \end{aligned}$$

285

$$\begin{aligned} m_2 &= \frac{2\sigma^2}{\sqrt{\pi}} \left[-\frac{1}{2} e^{-u^2} \Big|_{-\infty}^{\infty} + \frac{\sqrt{\pi}}{4} \operatorname{erf}(u) \Big|_{-\infty}^{\infty} \right] = \frac{2\sigma^2}{\sqrt{\pi}} \left[0 + \frac{\pi}{4} (\operatorname{erf}(\infty) - \operatorname{erf}(-\infty)) \right] = \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \left[\frac{\pi}{4} + \frac{\pi}{4} \right] = \sigma^2 \end{aligned}$$

286

y alrededor de cero (momentos crudos):

$$\begin{aligned} m_2^0 &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sigma\sqrt{2}u + \mu)^2 e^{-u^2} du = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} 2\sigma^2 u^2 e^{-u^2} du + \\ &\quad + \frac{\mu^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2} du + \frac{2\sigma\sqrt{2}\mu}{\sqrt{\pi}} \int_{-\infty}^{\infty} ue^{-u^2} du = \\ &= \underbrace{\frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} u^2 e^{-u^2} du}_{m_2 = \sigma^2} + \underbrace{\mu \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \mu e^{-u^2} du}_{m_1^0 = \mu} + + 2\mu \underbrace{\frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} ue^{-u^2} du}_{m_1 = 0} = \sigma^2 + \mu^2, \end{aligned}$$

289

La probabilidad de que una variable normalmente distribuida caiga en una desviación estandar de su valor medio viene dada por

290

$$\begin{aligned} P(-1 \leq z \leq 1) &= \int_{-1}^{+1} F(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-1}^{+1} e^{-\frac{1}{2}z^2} dz = \\ &= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{2\sqrt{1/2}} [\operatorname{erf}(z\sqrt{1/2})]_{-1}^1 = \frac{1}{2} [\operatorname{erf}(1/\sqrt{2}) - \operatorname{erf}(-1/\sqrt{2})] = \\ &= \frac{1}{2} [0.6827 - (-0.6827)] = 0.6827 (68.27\%), \end{aligned}$$

291

293 *y similarmente para 2 y 3 desviaciones estandard*

$$P(-2 \leq z \leq 2) = \int_{-2}^{+2} F_x(z) dz = 95.45\%$$

$$P(-3 \leq z \leq 3) = \int_{-3}^{+3} F_x(z) dz = 99.73\%.$$

295 *Entonces solo hay un 4.55% de probabilidad de que una variable normalmente dis-
296 tribuida caiga fuera de dos desviaciones estandard respecto de la media. Puesto que
297 es una probabilidad con 2 colas, la probabilidad de que una variable normal exceda
298 su media por mas de 2σ es la mitad de esto, es decir 2.275%, ya que la distribución
299 normal es simétrica.*

300 *En la práctica una PDF se calcula como un histograma escalado. Es por ello que
301 necesitamos escojer el tamaño y localización de los bins en el histograma. La demo
302 muestra las consecuencias de esta elección (pdf_demo.m).*

303

304 *{(3) Distribución de Poisson:}*

305 *La distribución de Poisson expresa la probabilidad de que ocurra un determinado
306 número de eventos durante un cierto intervalo en el tiempo o distancia en el espacio.
307 Se usa generalmente para la ocurrencia de sucesos con muy poca probabilidad o muy
308 “raros”. La expresión para la función acumulativa es:*

$$D(x) = P(x \leq r) = e^{-\lambda} \sum_{k=0}^{|r|} \frac{\lambda^k}{k!},$$

309 *donde λ es el valor promedio*

310

311 *y la función de densidad de probabilidad*

$$F(x) = P(x = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

312

313 *Los momentos estadísticos alrededor del origen de la función de distribución de
314 Poisson se pueden calcular directamente con sumatorios y expansion de Taylor:*

$$\begin{aligned} m_r^0 &= E[x^r] = \sum_{k \geq 0} k^r P(X = k) = \sum_{k \geq 0} k^r F(x) = \sum_{k \geq 0} k^r \frac{\lambda^k e^{-\lambda}}{k!} = \\ &= \lambda e^{-\lambda} \sum_{k \geq 0} \frac{k^r}{k!} \lambda^{k-1} \end{aligned}$$

316 *{} Veamos el momento de orden 1 alrededor del origen:*

$$\begin{aligned} m_1^0 &= E[x] = \lambda e^{-\lambda} \sum_{k \geq 0} \frac{k}{k!} \lambda^{k-1} = \lambda e^{-\lambda} \sum_{k \geq 1} \frac{k}{(k-1)!k} \lambda^{k-1} = \lambda e^{-\lambda} \sum_{k \geq 1} \frac{1}{(k-1)!} \lambda^{k-1} \\ &= \lambda e^{-\lambda} \sum_{j \geq 0} \frac{\lambda^j}{j!}, \end{aligned}$$

317 *para $j = k - 1$. Finalmente expandiendo en series de Taylor la función exponencial*

$$e^\lambda = \sum_{j \geq 0} \frac{1}{j!} \lambda^j$$

³¹⁹, obtenemos:

$$m_1^0 = \lambda e^{-\lambda} e^\lambda = \lambda$$

³²⁰ *Ejercicio: Demostrar que el momento estadístico alrededor del origen de orden 2 de*
³²¹ *la distribución de Poisson es igual a $m_2^0 = \lambda + \lambda^2$.*

$$\begin{aligned} m_2^0 &= E[x^2] = \lambda e^{-\lambda} \sum_{k \geq 0} \frac{k^2}{k!} \lambda^{k-1} = \lambda e^{-\lambda} \left[\sum_{k \geq 1} (k-1) \frac{1}{(k-1)!} \lambda^{k-1} + \sum_{k \geq 1} \frac{1}{(k-1)!} \lambda^{k-1} \right] = \\ &= \lambda e^{-\lambda} \left[\lambda \sum_{k \geq 2} \frac{1}{(k-2)!} \lambda^{k-2} + \sum_{k \geq 1} \frac{1}{(k-1)!} \lambda^{k-1} \right] = \\ &= \lambda e^{-\lambda} \left[\lambda \sum_{j \geq 0} \frac{1}{j!} \lambda^j + \sum_{i \geq 0} \frac{1}{(i)!} \lambda^i \right] = \lambda e^{-\lambda} [\lambda e^\lambda + e^\lambda] = \\ &= \lambda(\lambda + 1) = \lambda^2 + \lambda. \end{aligned}$$

³²² *En el caso que fuera el momento de orden 2 centrado se escribiría:*

$$E[(x - E[x])^2] = E[x^2] - (E[x])^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

³²⁴ *{ Ejemplo: } En los últimos 160 años, han sucedido 680 tormentas intensas en el*
³²⁵ *Golfo de México, incluyendo depresiones, tormentas tropicales, y huracanes. Asumimos que la frecuencia de ocurrencia de una tormenta intensa en el Golfo de México*
³²⁶ *sigue una distribución de Poisson (eventos “raros”, poco frecuentes). Calcula la*
³²⁷ *probabilidad de que ocurran 2 huracanes en 1 año:*

- ³²⁸ (a) *El número promedio de tormentas intensas por año es: $\mu = 680/160 = 4.25$ huracanes/año.*
³²⁹ (b) *La probabilidad de que ocurran 2 huracanes en 1 año es:*

$$P(x = 2) = \frac{\lambda^2 e^{-\lambda}}{2!} = \frac{4.25^2 e^{-4.25}}{2!} = 0.1288 (\sim 12\%)$$

³³⁰ *La probabilidad es muy baja debido a que exigimos que sean exactamente 2 huracanes en un año y no, por ejemplo, > 2 . En el segundo caso, la probabilidad aumentaría considerablemente*

$$\begin{aligned} P(x > 2) &= P(x = 3) + P(x = 4) + \dots = 1 - P(x \leq 2) = 1 - [P(x = 0) + P(x = 1) + P(x = 2)] = \\ &= 1 - [0.0143 + 0.0606 + 0.1288] = 1 - 0.2037 = 0.7963 (\sim 80\%) \end{aligned}$$

³³⁶ *{(4) Distribución Binomial: }*
³³⁷ *Supongamos que tenemos un conjunto de n tiradas en los cuales pueden suceder únicamente dos cosas: ‘acertar’ o ‘fallar’. La probabilidad de acertar en una tirada es*

339 $p=P$. Si X es el número total de aciertos en n tiradas, entonces la probabilidad de
 340 que el número de aciertos sea k es:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, 3, \dots, n,$$

341 $\{\}$ donde la expresión

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

342 $\mid \{\}$ es el número de diferentes combinaciones de grupos de k objetos que pueden
 343 ser elegidos de un conjunto total de n objetos. Estos números se denominan coefi-
 344 cientes binomiales. La probabilidad de que el número de aciertos caiga en un rango
 345 de valores es

$$P(a \leq X \leq b) = \sum_a^b P(X)$$

346 $\{Ejemplo\ 1:\}$?‘Cual es la probabilidad de obtener exactamente 6 caras de 10 lanza-
 347 mientos de moneda? Respuesta:

$$P(x = 6) = C(10, 6)0.5^6(1 - 0.5)^{10-6} = \frac{10!}{(10-6)!6!}0.5^6(1 - 0.5)^{10-6} \simeq 0.205$$

348 $\{Ejemplo\ 2:\}$?‘Cual es la probabilidad de obtener mas de 15 caras de 20 lanzamien-
 349 tos de moneda? Respuesta:

$$\sum_{k=16}^{20} \binom{20}{k} 0.5^k (1 - 0.5)^{20-k} = 0.006.$$

350 Si realizas esta operación a mano se vuelve muy tediosa. Es por ello que se utiliza la
 351 aproximación Normal a la distribución Binomiala (DeMoivre-Laplace).

352 $\{Teorema\ de\ DeMoivre-Laplace\ (aproximación\ de\ Binomial\ a\ Normal)\}$
 353 La distribución binomial de una variable X definida por n tiradas independientes
 354 cada una de las cuales tienen una probabilidad p de acertar, es aproximadamente
 355 una distribución Normal de media np y desviación típica $\sqrt{np(1-p)}$, cuando n es
 356 suficientemente grande. Entonces se deduce que para cualquier número a y b ,

$$\lim_{n \rightarrow \infty} P \left(a < \frac{X - np}{\sqrt{np(1-p)}} < b \right) = \frac{1}{\sqrt{2\pi np(1-p)}} \int_a^b \exp \left[-\frac{(x - np)^2}{2np(1-p)} \right] dx.$$

357 Esto significa que el estadístico, $\frac{X-np}{\sqrt{np(1-p)}}$, tiene una distribución Normal. Este teo-
 358 rema es un caso particular del teorema del límite central y nos permite de simplificar
 359 la solución de un problema binomial.

360 $\{Ejemplo\ de\ la\ aproximación\ Normal\ a\ la\ distribución\ Binomial:\}$
 361 El 2% de los XBTs fabricados por una empresa presentan defectos. Si hemos
 362 adquirido 2000 XBTs, ?‘Cual es la probabilidad de que haya menos de 50 defectu-
 363 osos?

{ Respuesta}: Se trata de una distribución binomial ya que solo pueden ser defectuosos o no defectuosos. La probabilidad que sea defectuoso es $p = 0.02$ (2%) y $n = 2000$, lo que nos da una distribución Binomial $B(2000, 0.02)$. Puesto que la n es grande podemos hacer una aproximación a la distribución Normal. Calculamos la media y desviación estandar de la distribución Normal $\mu = np = 200 * 0.02 = 40$ y $\sigma = \sqrt{np(1-p)} = \sqrt{2000 * 0.02 * (1 - 0.02)} = 6.26$ x es $B(2000, 0.02)$ y x_N es $N(40, 6.26)$.

La probabilidad que $x < 50$ es

$$p(x < 50) = p(x_N \leq 49),$$

y si estandarizamos

$$p(x_N \leq 49) = p\left(z \leq \frac{49 - 40}{6.26}\right) = p(z \leq 1.44) = 0.9251.$$

EJERCICIOS de estadística y probabilidad:

Ejercicio 1: Calcule $E[x]$ si x tiene la función de densidad de probabilidad

$$f(x) = \begin{cases} \frac{1}{4}xe^{-\frac{x}{2}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}.$$

La esperanza $E[x]$ de la función $f(x)$ es entonces

$$E[x] = \int_0^\infty x \left(\frac{1}{4}xe^{-\frac{x}{2}}\right) dx = \frac{1}{4} \int_0^\infty x^2 e^{-\frac{x}{2}} dx.$$

Definamos $y = x/2$; entonces $x = 2y$ y $dx = 2dy$ y obtenemos

$$E[x] = \frac{1}{4} \int_0^\infty x^2 e^{-\frac{x}{2}} dx = \frac{1}{4} \int_0^\infty (2y)^2 e^{-y} 2dy = 2 \int_0^\infty y^2 e^{-y} dy.$$

Vamos ahora a resolver la integral por partes. Hacemos la siguiente sustitución: $u = y^2$, $dv = e^{-y}$ y por ende $du = 2ydy$ y $v = -e^{-y}$. La integral se puede reescribir usando la expresión general de integración por partes $h(x) = uv - \int vdu$:

$$E[x] = 2 \int_0^\infty y^2 e^{-y} dy = 2 \left[-y^2 e^{-y} - \int -e^{-y} (2y) dy \right] = 2 \left[-y^2 e^{-y} + 2 \int y e^{-y} dy \right].$$

Integramos de nuevo por partes. Usa $u = y$, $dv = e^{-y}$ y entonces $du = dy$ y $v = -e^{-y}$

$$\begin{aligned} E[x] &= 2 \left[-y^2 e^{-y} + 2 \left\{ -ye^{-y} - \int -e^{-y} dy \right\} \right] = \\ &= 2 \left[-y^2 e^{-y} + 2 \left\{ -ye^{-y} - e^{-y} \right\} \right] = \\ &= -2y^2 e^{-y} - 4ye^{-y} - 4e^{-y} = \\ &= [-2e^{-y}(y^2 + 2y + 2)]_0^\infty = \\ &= \lim_{n \rightarrow \infty} [-2e^{-y}(y^2 + 2y + 2)] - [-2e^{-y}(y^2 + 2y + 2)]_{y=0}. \end{aligned}$$

El límite es ahora del tipo ∞/∞ y entonces usamos la regla de l'Hopital

$$E[x] = -2 \lim_{y \rightarrow \infty} \frac{2y + 2}{e^y} + 2 [e^{-y}(y^2 + 2y + 2)]_{y=0}$$

387 *Usamos la regla de l'Hopital de nuevo*

$$E[x] = -2 \lim_{y \rightarrow \infty} \frac{2}{e^y} + 2 [e^{-y}(y^2 + 2y + 2)]_{y=0} = 0 + 2 [e^{-y}(y^2 + 2y + 2)]_{y=0} = 4$$

388 *{Ejercicio 2:} Calcule E[x] si x tiene la función de densidad de probabilidad*

$$f(x) = \begin{cases} c(1-x^2) & , -1 < x < 1 \\ 0 & , \text{otherwise} \end{cases} .$$

$$\begin{aligned} E[x] &= \int_{-1}^1 x[c(1-x^2)]dx = c \int_{-1}^1 x[(1-x^2)]dx = \\ 389 &= c \int_{-1}^1 x - x^3 dx = c \left[\frac{x^2}{2} + \frac{x^4}{4} \right]_{-1}^1 = 0 \end{aligned}$$

390 *{Ejercicio 3:} Calcule E[x] si x tiene la función de densidad de probabilidad*

$$f(x) = \begin{cases} \frac{5}{x^2} & , x > 5 \\ 0 & , x \leq 5 \end{cases} .$$

$$\begin{aligned} 391 E[x] &= \int_5^\infty x \frac{5}{x^2} dx = \int_5^\infty \frac{5}{x} dx = 5 \int_5^\infty \frac{1}{x} dx \\ &= 5[\ln x]_5^\infty = 5 \left[\left(\lim_{x \rightarrow \infty} \ln x \right) - \ln 5 \right] \rightarrow \infty \end{aligned}$$

392 *Ejercicio 4: La variable aleatoria x tiene la siguiente función de densidad de proba-*
393 *bilidad*

$$f(x) = \begin{cases} k(2x+3) & -1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases} .$$

394 • ¿Cuál es el valor de k?

$$\int_{-\infty}^{\infty} k(2x+3)dx = 1$$

$$\int_{-1}^2 k(2x+3)dx = (kx^2 + 3kx) \Big|_{-1}^2 = kx(3+x) \Big|_{-1}^2 = 10k + 2k = 1$$

$$12k = 1 \rightarrow k = \frac{1}{12}$$

395 • Calcula E[x]

$$\begin{aligned} E[x] &= \int_{-1}^2 xk(2x+3)dx = 2k \left. \frac{x^3}{3} \right|_{-1}^2 + 3k \left. \frac{x^2}{2} \right|_{-1}^2 = \\ &= k \left[\frac{18}{3} + \frac{9}{2} \right] = \frac{21}{24} . \end{aligned}$$

396 *Ejercicio 5: Sea la función $g(x)$ dada por*

$$g(x) = \begin{cases} x + 2\alpha & , x \leq -\alpha \\ x & , -\alpha \leq x \leq \alpha \\ x - 2\alpha & , x > \alpha \end{cases}$$

397 *donde asumimos que x esta normalmente distribuida. Calcula la media de $g(x)$.*

$$\begin{aligned} E[g(x)] &= \int_{-\infty}^{\infty} g(x)F(x)dx = \int_{-\infty}^{-\alpha} (x + 2\alpha)F(x)dx + \int_{-\alpha}^{\alpha} xF(x)dx + \int_{\alpha}^{\infty} (x - 2\alpha)F(x)dx = \\ &= \int_{-\infty}^{-\alpha} xF(x)dx + \int_{-\infty}^{-\alpha} 2\alpha F(x)dx + \int_{-\alpha}^{\alpha} xF(x)dx + \int_{\alpha}^{\infty} xF(x) - \int_{\alpha}^{\infty} 2\alpha F(x)dx = \\ &= \int_{-\infty}^{-\infty} xF(x)dx + 2\alpha \left[\int_{-\infty}^{-\alpha} F(x)dx - \int_{\alpha}^{\infty} F(x) \right] = \\ &= \int_{-\infty}^{-\infty} xF(x)dx + 2\alpha \left[D(x = -\alpha) - \left(1 - \int_{-\infty}^{-\alpha} F(x)dx \right) \right] = \\ &= \int_{-\infty}^{-\infty} xF(x)dx + 2\alpha \left[D(x = -\alpha) - \left(1 - D(x = \alpha) \right) \right] = \\ &= \mu + 2\alpha [D(-\alpha) - 1 + D(\alpha)] . \end{aligned}$$

403 *donde la media de la distribución Normal es*

$$E[x] = \int_{-\infty}^{-\infty} xF(x)dx = \int_{-\infty}^{-\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \mu .$$

404 *{Teorema central del límite:} \ | \ {Definición 1}: Sea $X_1, X_2, X_3, \dots, X_n$ un conjunto
405 de variables aleatorias, independientes e idénticamente distribuidas con media μ y
406 varianza σ^2 distinta de cero. Sea*

$$S_n = X_1 + X_2 + \dots + X_n ,$$

407 *entonces*

$$\lim_{n \rightarrow \infty} Pr(Z_n \leq z) = \Phi(z) ,$$

408 *donde $\Phi(z)$ es una distribución Normal estandar y $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ es una
409 estandarización del sumatorio S_n de tal forma que la media de la nueva variable Z_n
410 sea cero y su desviación estándar sea igual a 1. De esta forma, las variables Z_n
411 convergerán a una distribución normal estándar $N(0, 1)$, cuando n tienda a infinito.*
412 *Definición 2: Sea $X_1, X_2, X_3, \dots, X_n$ un conjunto de variables aleatorias, indepen-
413 dientes e idénticamente distribuidas con media μ y varianza σ^2 distinta de cero.
414 Entonces, si n es suficientemente grande, la variable aleatoria*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

415 *tiene aproximadamente una distribución normal con media $\mu(\bar{X}) = \mu$ y desviación
416 típica $\sigma(\bar{X}) = \sigma/\sqrt{n}$.*

417 *{NOTA}: Es importante remarcar que el teorema del límite central no dice nada
418 acerca de la distribución de X_i , solo de la distibución de su media muestral \bar{X} .*

419 % %

420 *Aplicación 1: Calculo de probabilidades sobre la media muestral.*421 *Ejemplo: La recolección de muestras de agua con una roseta es una variable aleato-*
422 *ria con media $\mu = 150 \text{ ml}$ y varianza de $\sigma^2 = 120 \text{ ml}^2$. Si tomamos $n = 40$ muestras*
423 *aleatorias de agua. (a) ¿Cuál es la media y la desviación estándar de la media*
424 *muestral?, (b) ¿Cuál es la probabilidad de que la media muestral contenga entre 145*
425 *y 153 ml de agua?*426 (a) $\mu(\bar{X}) = 150 \text{ ml}$ y $\sigma(\bar{X}) = \sigma/\sqrt{n} = \sqrt{120/40} = \sqrt{3} \text{ ml}$ 427 (b) Queremos calcular $Pr(145 \leq \bar{X} \leq 153)$. Si escribimos la probabilidad en forma
428 estandarizada, entonces:

$$Pr(145 \leq \bar{X} \leq 153) = Pr\left(\frac{145 - 150}{\sqrt{3}} \leq Z \leq \frac{153 - 150}{\sqrt{3}}\right) \simeq Pr(-2.89 \leq Z \leq 1.73) =$$

$$= Pr(Z \leq 1.73) - Pr(Z \leq -2.89) = 0.9582 - (1 - 0.9981) = 0.9582 - 0.0019 = 0.9563$$

429 *{Función de densidad de probabilidad conjunta}*430 *La probabilidad que dos variables aleatorias (x, y) caigan en la región R (como por*
431 *ejemplo un rectángulo) se obtiene integrando su función de probabilidad conjunta*

$$P((x, y) \in R) = \int \int_R F(x, y) dx dy .$$

432 En particular, si R es un rectángulo 2d $(x, y) : r \leq x \leq r + dr, s \leq y \leq s + ds$, en-
433 tonces

$$P((x, y) \in R) = P(r \leq x \leq r + dr, s \leq y \leq s + ds) = \int_r^{r+dr} \int_s^{s+ds} F(x, y) dx dy .$$

434 *Algunas propiedades:*

- 435 (1)
- $F(x, y) \geq 0$
- para todo
- x, y
- .
-
- 436 (2)
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x, y) dx dy = 1$

437 *Definición: La función de densidad de probabilidad marginal de variables aleatorias*
438 *x e y son:*

439 $F_x(x) = \int_{-\infty}^{\infty} F(x, y) dy$ y $F_y(y) = \int_{-\infty}^{\infty} F(x, y) dx .$

440 *{Ejemplo del uso de la función de densidad de probabilidad conjunta}*441 *Imaginemos que una empresa de instrumentación oceanográfica fabrica boyas La-*
442 *grangianas de grosor x y diámetro y , los cuales varían de una boya a la otra. Imag-*
443 *inemos que la función de densidad de probabilidad conjunta de la variable aleatoria*
444 *“dimensión del instrumento oceanográfico” es:*

$$F(x, y) = \frac{1}{6}(r + s) \quad \text{si } (x, y) \in R = \{1 \leq x \leq 2; 4 \leq y \leq 5\}$$

$$F(x, y) = 0 \quad \text{si } (x, y) \text{ fuera de } R$$

445 Ahora queremos saber que probabilidad hay de que una boyta tenga un grosor $1 \leq x \leq$
 446 $1.5m$ y un diámetro $4.5 \leq y \leq 5m$, es decir

$$P(1 \leq x \leq 1.5, 4.5 \leq y \leq 5) = \int_1^{1.5} \int_{4.5}^5 \frac{1}{6}(r+s)dsdr = 0.253 = 25\%$$

447 *{Significancia estadística utilizando la distribución Normal}*
 448 Como vimos en el teorema central del límite, para una población infinita ($N \rightarrow \infty$)
 449 la desviación estandar de la distribución de las medias muestrales es:

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{N}} = \text{error estandar del estimado de la media}.$$

450 Aquí, σ es la desviación estandar de la población y N es el número de datos (in-
 451 dependientes) utilizado para calcular la media muestral. Entonces, si promediamos
 452 observaciones de una población de desviación estandar σ , la desviación estandar de
 453 esos promedios disminuye como el inverso de la raíz cuadrada del tamaño muestral
 454 N .
 455 Si N es suficientemente grande podemos usar las estimaciones de σ y \bar{x} para calcular
 456 el denominado estadístico z que corresponde a una distribución normal estandarizada
 457 de media $\mu = 0$ y $\sigma = 1$

$$z = \frac{\bar{x} - \mu}{\sigma(\bar{x})} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

458 La fórmula de arriba puede modificarse convenientemente para darnos un test de
 459 significancia estadística para la diferencia entre medias muestrales con tamaños
 460 muestrales y desviación estandar diferentes:

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_{1,2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}},$$

461 donde $\Delta_{1,2}$ es la diferencia esperada entre las dos medias, lo que se suele asumir
 462 cero en la práctica.

463 Si el tamaño muestral N es menor de 30 entonces no podemos usar el estadístico
 464 z , pero podemos utilizar la distribución t-student; o cuando queremos comparar
 465 varianzas, podemos usar la distribución chi-cuadrada. La t-student converge a una
 466 distribución normal para largos tamaños muestrales y se define como

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N-1}}} = \frac{\bar{x} - \mu}{\frac{\hat{s}}{\sqrt{N}}}; \hat{s} = \sqrt{\frac{N}{N-1}}s.$$

467 Si consideramos una población normalmente distribuida de media μ la función de
 468 densidad de probabilidad de la t-student es

$$F_x(t) = \frac{f_0(\nu)}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}},$$

469 donde $\nu = N - 1$ es el número de grados de libertad y $f_0(\nu)$ es una constante que
 470 depende en ν y permite que el área bajo la curva $F_x(t)$ sea igual a la unidad. Los
 471 grados de libertad se definen como el número de muestras independientes N menos
 472 el número de parámetros del estadístico que queremos estimar.

473 A diferencia del estadístico z , la t-student depende del número de grados de libertad;
 474 la cola de la distribución es larga para números de grados de libertad bajos (o N
 475 pequeña). Para números altos de grados de libertad (o N grande), la distribución
 476 t-student se acerca al estadístico z o distribución Normal.

477 **{Intervalos de confianza}**

478 Para calcular valores de los estadísticos z y t-student debemos de fijar el nivel de
 479 confianza definido como $1 - \alpha$; porcentaje del nivel de confianza $100(1 - \alpha)\%$. Esto
 480 se puede escribir simbólicamente como

$$P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1 - \alpha$$

$$P(-t_{\alpha/2} < t < t_{\alpha/2}) = 1 - \alpha.$$

481 Una vez definido el nivel de confianza y los grados de libertad ν (para la t-student)
 482 podemos leer el valor de dichos estadísticos en tablas. En esas tablas $z_{\alpha/2}$ es el valor
 483 de z para el cual solo el $100 * \alpha/2\%$ de los valores de z es esperado ser mas grande
 484 (cola de la derecha de la distribución). Igualmente, $z_{-\alpha/2} = -z_{\alpha/2}$ es el valor de z
 485 para el cual solo el $100 * \alpha/2\%$ de los valores de z es esperado ser mas pequeño (cola
 486 de la izquierda de la distribución). O dicho de otra forma, $z_{\alpha/2}$ es el valor por encima
 487 del cual existe un área bajo la curva de $\alpha/2$. Los valores de z y t son las integrales
 488 bajo las correspondientes funciones de densidad de probabilidad.

489

490 (1) **{Intervalo de confianza para μ ($N > 30$, σ conocida)}**

491 Cuando $N > 30$ y σ es conocida, podemos usar el estadístico z para encontrar el
 492 intervalo de confianza para μ . Hay un $100 * (1 - \alpha)\%$ que cualquier estadístico z
 493 caiga en el intervalo

$$\begin{aligned} z_{-\alpha/2} &< \frac{\bar{x} - \mu}{\sigma} \sqrt{N} < z_{\alpha/2} \\ \frac{\sigma}{\sqrt{N}} z_{-\alpha/2} &< \bar{x} - \mu < \frac{\sigma}{\sqrt{N}} z_{\alpha/2} \\ -1 \frac{\sigma}{\sqrt{N}} z_{-\alpha/2} &> \mu - \bar{x} > -1 \frac{\sigma}{\sqrt{N}} z_{\alpha/2} \\ \bar{x} - 1 \frac{\sigma}{\sqrt{N}} z_{-\alpha/2} &> \mu > \bar{x} - 1 \frac{\sigma}{\sqrt{N}} z_{\alpha/2} \end{aligned}$$

497 y sabiendo que es simétrica $-z_{\alpha/2} = z_{-\alpha/2}$:

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{N}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}.$$

498 Supongamos que queremos encontrar el intervalo de confianza de μ al 95% de confi-
 499 anza, es decir, entonces $\alpha = 0.05$. Entonces $z_{\alpha/2} = 1.96$ (de tablas estadísticas). |

500 {Ejemplo}: $N = 40$, $\sigma = 0.5^\circ C$, y $\bar{x} = 12.7^\circ C$:

$$\bar{x} - z_{0.025} \frac{\sigma}{\sqrt{N}} < \mu < \bar{x} + z_{0.025} \frac{\sigma}{\sqrt{N}},$$

$$[12.7 - (1.96)0.5/\sqrt{40}]^\circ C < \mu < [12.7 + (1.96)0.5/\sqrt{40}]^\circ C$$

$$12.54^\circ C < \mu < 12.85^\circ C$$

503 | (2) {Intervalo de confianza para μ ($N < 30$, σ desconocida)} | Cuando $N < 30$
 504 y σ es desconocida, podemos usar el estadístico t -student para encontrar el intervalo
 505 de confianza para μ . Hay un $100 * (1 - \alpha)\%$ que cualquier estadístico t caiga en el
 506 intervalo

$$t_{-\alpha/2} < \frac{\bar{x} - \mu}{s} \sqrt{N-1} < t_{\alpha/2},$$

$$\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{N-1}} < \mu < \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{N-1}}.$$

507 Si $\alpha = 0.05$, hay un 95% de probabilidad que cualquier estadístico t caiga en el inter-
 508 valo

$$t_{-0.025} < \frac{\bar{x} - \mu}{s} \sqrt{N-1} < t_{0.025},$$

509 de lo cual podemos deducir que la verdadera media μ es de esperar con un 95% de
 510 confianza que caiga en el intervalo:

$$\bar{x} - t_{0.025} \frac{s}{\sqrt{N-1}} < \mu < \bar{x} + t_{0.025} \frac{s}{\sqrt{N-1}}.$$

511 De forma general, podemos definir el intervalo de confianza como:

$$\mu = \bar{x} \pm t_c \frac{\hat{s}}{\sqrt{N}},$$

512 donde t_c es el valor crítico del estadístico t (límites del intervalo), el cual depende
 513 del número de grados de libertad y del nivel de confiabilidad deseado. El intervalo
 514 de confianza con el estadístico z , el cual solo es apropiado para tamaños muestrales
 515 grandes ($N > 30$) donde la desviación estandar es conocida:

$$\mu = \bar{x} \pm z_c \frac{\sigma}{\sqrt{N}}.$$

516 Observamos que la teoría para tamaños muestrales pequeños reemplaza el estadístico
 517 z por el t y utiliza una desviación estandar muestral modificada

$$\hat{s} = \sqrt{\frac{N}{N-1}} s.$$

518 % {Diferencias entre medias}| %Supongamos dos muestras de tamaño N_1 y N_2

519 extraídas de una población %con distribución normal con desviaciones estandar
 520 iguales. Supongamos %que las medias muestrales son \bar{x}_1 y \bar{x}_2 y las %desviaciones
 521 estandar muestrales son s_1 y s_2 . Para comprobar %la hipótesis nula (H_0) que ambas
 522 muestras provienen de la misma población, %es decir $\mu_1 = \mu_2$ y $\sigma_1 = \sigma_2$ podemos
 523 usar la %siguiente expresión %

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}};$$

524 % donde $\nu = N_1 + N_2 - 2$. |

525 (3) Intervalo de confianza para la diferencia de medias $\mu_1 - \mu_2$ |

526 El teorema central del límite (TCL) para la diferencia de medias muestrales de dos
 527 poblaciones viene dado por

$$\bar{x}_1 - \bar{x}_2 \sim N(\mu_{\bar{x}_1 - \bar{x}_2}, \sigma_{\bar{x}_1 - \bar{x}_2}),$$

528 |

529 donde

530

$$\text{Media : } \mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

$$\text{Varianza : } \sigma_{\bar{x}_1 - \bar{x}_2}^2 = \sigma^2 \bar{x}_1 + \sigma^2 \bar{x}_2 = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

531 |

532 **3.1 Desviaciones estandar poblacionales (σ_1 y σ_2) conocidas; μ_1 y μ_2 desconocidas}**
533 {(estadístico z):}

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

534 | | {3.2} Desviaciones estandar poblacionales (σ_1 y σ_2) y medias poblacionales (μ_1
535 y μ_2) desconocidas} (estadístico t):

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

536 | | {3.3} Desviaciones estandar poblacionales (σ_1 y σ_2) desconocidas pero iguales}
537 {(estadístico t): | | Supongamos dos muestras de tamaño N_1 y N_2 extraídas de dos
538 poblaciones Normales con desviaciones estándar iguales ($\sigma_1 = \sigma_2$). Supongamos que
539 conocemos las medias y desviaciones estándar muestrales \bar{x}_1 y \bar{x}_2 y s_1 y s_2 . Para
540 comprobar la hipótesis nula H_0 que las muestran proceden de la misma población
541 ($\mu_1 = \mu_2$ y $\sigma_1 = \sigma_2$) usamos el estadístico t -score (o pooled t):}

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\hat{s}_d \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

542 |

$$\hat{s}_d = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}},$$

543 | donde $\nu = N_1 + N_2 - 2$ es el número de grados de libertad. |544 **Ejemplo:** Un ingeniero que diseña instrumentos oceanográficos está ineteresado en
545 aumentar el tiempo durante el cuál la pintura “antifouling” evita que los microor-
546 ganismos se pequen y crezcan sobre el instrumento oceanográfico. Se prueban dos
547 fórmulas de pintura: fórmula 1 estándar y fórmula 2 con un nuevo ingrediente que
548 aumenta el tiempo de acción.549 De la experiencia se sabe que la desviación estándar del tiempo de acción de la pin-
550 tura es de 8 días y ésta variabilidad no se vé afectada por el nuevo ingrediente. Se
551 pintan 35 instrumentos con la fórmula 1 y otros 35 con la fórmula 2. Los tiempos
552 promedios de acción del “antifouling” son de 116 días para la fórmula 1 y 112 días
553 para la fórmula 2. ¿A qué conclusión puede llegar el ingeniero diseñador del in-
554 strumento sobre la eficacia del nuevo ingrediente, con un nivel de significancia de
555 0.01? |556 $x_1 \equiv$ Tiempo de acción “antifouling” fórmula 1 | $x_2 \equiv$ Tiempo de acción “antifoul-
557 ing” fórmula 2 | $x_1 \sim$ Desconocida ($\mu_1, \sigma_1 = 8$ días) | $x_2 \sim$ Desconocida ($\mu_2, \sigma_2 = 8$
558 días) | $\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \sigma_1/\sqrt{N_1} + \sigma_2/\sqrt{N_2})$ (TCL) | $\bar{x}_1 = 116$ días | $\bar{x}_2 = 112$
559 días | $N_1 = N_2 = 35$ | $\alpha = 0.01$ | $H_0 : \mu_1 - \mu_2 = 0$ | $H_1 : \mu_1 - \mu_2 \neq 0$ |560 El intervalo de confianza de la diferencia de medias $\mu_1 - \mu_2$

$$\bar{x}_1 - \bar{x}_2 - z_{\alpha/2} \left(\frac{\sigma_1}{\sqrt{N_1}} + \frac{\sigma_2}{\sqrt{N_2}} \right) < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + z_{\alpha/2} \left(\frac{\sigma_1}{\sqrt{N_1}} + \frac{\sigma_2}{\sqrt{N_2}} \right)$$

$$\bar{x}_1 - \bar{x}_2 - 2.33(1.9124) < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + 2.33(1.9124)$$

$$\bar{x}_1 - \bar{x}_2 - 4.4559 < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + 4.4559$$

$$4 - 4.4559 < \mu_1 - \mu_2 < 4 + 4.4559,$$

561

|

562

y el intervalo de confianza es: |

$$-0.4559 < \mu_1 - \mu_2 < 8.4559 \text{ al } 99\% (\alpha = 0.01)$$

$$H_0 : \mu_1 - \mu_2 = 0 \text{ al } 99\%,$$

563

y aceptamos hipótesis nula H_0 . |

564

(4) | *Intervalo de confianza para la varianza:*

565

Distribución chi-cuadrada

566

En ocasiones queremos definir un intervalo de confianza para la varianza muestral.

567

Para ello podemos usar el estadístico chi-cuadrado. Definamos

$$\chi^2 = (N - 1) \frac{s^2}{\sigma^2}.$$

568

| **Propiedades:**

569

- No es simétrica.
- La forma de la distribución depende de los grados de libertad.
- A medida que los grados de libertad aumentan (N aumenta), la distribución se parece más a la Normal.
- $\chi^2 \geq 0$

574

Para definir el intervalo de confianza sabemos que hay un $100*(1-\alpha)\%$ que cualquier estadístico χ^2 caiga en el intervalo

576

$$\chi^2_{1-\alpha/2} < (N - 1) \frac{s^2}{\sigma^2} < \chi^2_{\alpha/2},$$

$$\frac{1}{\chi^2_{1-\alpha/2}} > \frac{\sigma^2}{(N - 1)s^2} > \frac{1}{\chi^2_{\alpha/2}},$$

577

y entonces:

578

$$\frac{(N - 1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(N - 1)s^2}{\chi^2_{1-\alpha/2}}.$$

579

Usamos $1 - \alpha/2$ porque la χ^2 es positiva. El valor $\chi^2_{\alpha/2}$ es mayor que el valor $\chi^2_{1-\alpha/2}$. Las tablas dan la probabilidad a la derecha del valor.

581 *Para una población normalmente distribuida con desviación estandar σ , la función
582 de densidad de probabilidad de la chi-cuadrada es:*

$$F_x(\chi) = f_0 \chi^{\nu-2} e^{-\frac{1}{2}\chi^2}; \nu = N - 1.$$

583 *Puesto que la distribución chi es asimétrica y positiva, si $\alpha = 0.05$ (95% confianza),
584 el intervalo de confianza para la varianza σ^2 como*

$$\frac{(N-1)s^2}{\chi_{0.025}^2} < \sigma^2 < \frac{(N-1)s^2}{\chi_{0.975}^2},$$

585 *y si leemos en las tablas para $\nu = 9$ grados de libertad:*

$$\frac{(N-1)s^2}{19.023} < \sigma^2 < \frac{(N-1)s^2}{2.700},$$

586 **Ejemplo:** Supongamos que tenemos $\nu = 9$ grados de libertad de nuestra estimación
587 espectral de la componente meridional de la velocidad de la corriente. Sabemos que
588 la varianza muestral de un pico espectral es $s^2 = 10 \text{ cm s}^{-2}$? Cuál es el intervalo de
589 confianza al 95% para la varianza?
590 De las tablas estadísticas vemos que para $\nu = N - 1 = 9$ grados de libertad, $\chi_{1-\alpha/2}^2 =$
591 $\chi_{0.095}^2 = 19.02$ y $\chi_{\alpha/2}^2 = \chi_{0.025}^2 = 2.70$. Entonces, el intervalo es:

$$\frac{(9)10}{19.023} < \sigma^2 < \frac{(9)10}{2.700}$$

592 $4.7 \text{ cm}^2 \text{ s}^{-2} < \sigma^2 < 33.3 \text{ cm}^2 \text{ s}^{-2}$

593 **{Grados de libertad}** El número de grados de libertad es el número de muestras in-
594 dependientes N menos el número de parámetros del estadístico que queremos estimar.
595 Por ejemplo en el estadístico t

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N-1}}} = \frac{\bar{x} - \mu}{\frac{\hat{s}}{\sqrt{N}}}; \hat{s} = \sqrt{\frac{N}{N-1}s},$$

596 calculamos la media muestral y la desviación estandar s a partir de los datos, pero
597 la verdadera media μ debe ser estimada, por lo que $\nu = N - 1$. Similarmente en el
598 estadístico chi-cuadrada

$$\chi^2 = (N-1) \frac{s^2}{\sigma^2},$$

599 conocemos la varianza muestral s^2 y el tamaño muestral N , pero debemos estimar la
600 verdadera varianza, y entonces $\nu = N - 1$.

601 **{Estadístico F}** Otro estadístico útil para tests espectrales es el estadístico F. Si
602 s_1^2 y s_2^2 son las varianzas de muestras aleatorias independientes de tamaño N_1 y N_2 ,
603 tomadas de dos poblaciones Normales con la misma varianza σ^2 , entonces

$$F = \frac{s_1^2}{s_2^2},$$

604 es el valor de una variable aleatoria cuya distribución es F con los parámetros
605 $\nu_1 = N_1 - 1$ y $\nu_2 = N_2 - 1$. Este estadístico es muy útil en tests de signifi-
606 cia para los picos de los espectros frecuenciales de potencia. Los dos parámetros son
607 los grados de libertad para las varianzas del cociente; ν_1 para s_1^2 y ν_2 para s_2^2 .

608 $\{\text{Tests para hipótesis}\}$ | Para usar los test de significancia estadística debemos de
 609 seguir 5 pasos: | (1) Definir el nivel de confianza | (2) Definir la hipótesis nulla
 610 H_0 y su alternativa H_1 | (3) Definir el estadístico que usaremos | (4) Definir la
 611 región crítica | (5) Evaluar el estadístico y concluir | | Es muy importante definir
 612 correctamente la hipótesis nula. Es decir, estar seguro que rechazar la hipótesis nula
 613 H_0 implica únicamente la existencia de su alternativa H_1 . Normalmente la hipótesis
 614 nula y su alternativa son mutualmente excluyentes. Ejemplos: |

615 H_0 : Las medias de dos muestras son iguales H_1 : Las medias de dos muestras no son
 616 iguales |

617 H_0 : El coeficiente de correlación es cero H_1 : El coeficiente de correlación no es cero
 618 |

619 **Ejemplo:** | En una muestra de 41 inviernos la temperatura media de Enero es 5.55°C
 620 y la desviación es de 0.65°C ? Cual es el intervalo de confianza al 95% de que la
 621 verdadera temperatura media sea esa? | (1) Nivel de confianza del 95% | (2) H_0
 622 es que la media verdadera se encuentra en el intervalo $5.55 \pm \Delta T$ y su alternativa
 623 H_1 es que se encuentra fuera de este intervalo. | (3) Usamos el estadístico t . | (4)
 624 La región crítica es $|t| < t_{0.025}$, lo cual para $\nu = N - 1 = 40$ es $|t| < 2.26$ (leido
 625 de tablas estadísticas). Escrito en términos de intervalo de confianza para la media
 626 poblacional:

$$\bar{x} - 2.0211 \frac{s}{\sqrt{N-1}} < \mu < \bar{x} + 2.0211 \frac{s}{\sqrt{N-1}}$$

627 | (5) Si ponemos en números el intervalo obtenemos $5.06 < \mu < 6.03$. Tenemos un
 628 95% de confianza que la verdadera temperatura media se encuentra en ese intervalo.

629 **{Teorema de Bayes:}** | Sea $E_i, i = 1, 2, 3, \dots, n$ un conjunto de n eventos que consti-
 630 tuyen una partición del espacio muestral S

$$\bigcup_{i=1}^n E_i \in S,$$

631 {}cada uno de los cuales tiene probabilidad positiva de ocurrir $P(E_i) > 0$ para
 632 $i = 1, 2, \dots, n$ y son exclusivos entre si

$$E_i \cap E_j = i \neq j.$$

633 Entonces dada la ocurrencia previa de un evento cualquiera B , la probabilidad de que
 634 suceda el evento E_j es

$$P(E_j|B) = \frac{P(B|E_j)P(E_j)}{\sum_{i=1}^n P(B|E_i)P(E_i)}, \quad (7)$$

635 {}donde

$$P(E_j|E_i) = \frac{P(E_i \cap E_j)}{P(E_i)},$$

636 es la probabilidad condicional, es decir, la probabilidad que ocurra el evento E_j si
 637 previamente ha ocurrido el evento E_i y $P(E_i \cap E_j)$ es la probabilidad que ambos
 638 eventos ocurran, i.e. la intersección de dos eventos. La intersección es puede escribir

$$P(E_i \cap E_j) = P(E_j|E_i) * P(E_i) = P(E_i|E_j) * P(E_j).$$

639 *Si ambos eventos son independientes (no interseccionan) tal que $P(E_i|E_j) = P(E_i)$*
 640 *obtenemos*

$$P(E_i \cap E_j) = P(E_i) * P(E_j),$$

641 *es decir, la definición de independencia estadística entre eventos.*
 642 *{Ejemplo teorema de Bayes:} Imaginemos que queremos saber si una muestra de*
 643 *agua contiene diatomeas o no. La probabilidad de que una muestra de agua tomada*
 644 *al azar en la bahía de Todos Santos tenga diatomeas es de 1/100 ($P(D) = 0.01$).*
 645 *La probabilidad de que si hay diatomeas el test de negativo es cero ($P(+|D) = 1$)*
 646 *y la probabilidad de que el test de un falso positivo es del 5% ($P(+|noD) = 0.05$).*
 647 *Si agarramos una muestra de agua y da positivo, ?Cual es la probabilidad de que*
 648 *hayan diatomeas? Intuitivamente, sabemos que existe un 5% de probabilidad de que*
 649 *el test me de un falso positivo y por lo tanto, un 95% de que si da positivo tenga di-*
 650 *atomeas en la muestra de agua. Veamos que nos dice el teorema de Bayes. Sabemos*
 651 *que $P(D) = 0.01$, $P(+|D) = 1$, y $P(+|noD) = 0.05$. Si usamos el teorema de Bayes:*

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|noD)P(noD)} = \frac{1 * 0.01}{1 * 0.01 + 0.05 * 0.99} \sim \frac{1}{6}.$$

652 *En verdad solamente existe 1 probabilidad de 6 (~ 16%) que si el test es pos-*
 653 *itivo existan diatomeas. De cada 100 muestras solamente 1 tiene diatomeas y*
 654 *5% * 100 = 0.05 * 100 = 5 muestras de las 100 dan positivo y no tienen diatomeas.*
 655 *Por ello, sabiendo que la muestra que tiene diatomeas dio positivo, de 6 positivos,*
 656 *solamente 1 es cierto.*

Table of the Chi-Squared Distribution.

v	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
75	47.206	49.475	52.942	56.054	59.795	91.061	96.217	100.839	106.393	110.285
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.170

Table of the F-statistic for a probability level of 0.01 (Denominator to left, Numerator across top)

d.f.	1	2	3	4	5	6	7	8	9	10	20	30	40	50	75	100
1	161	199	5404	5624	5764	5859	5928	5981	6022	6056	6209	6260	6286	6302	6324	6334
2	18.51	19.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.45	99.47	99.48	99.48	99.48	99.49
3	10.13	9.55	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	26.69	26.50	26.41	26.35	26.28	26.24
4	7.71	6.94	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.02	13.84	13.75	13.69	13.61	13.58
5	6.61	5.79	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.55	9.38	9.29	9.24	9.17	9.13
6	5.99	5.14	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.40	7.23	7.14	7.09	7.02	6.99
7	5.59	4.74	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.16	5.99	5.91	5.86	5.79	5.75
8	5.32	4.46	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.36	5.20	5.12	5.07	5.00	4.96
9	5.12	4.26	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.81	4.65	4.57	4.52	4.45	4.41
10	4.96	4.10	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.41	4.25	4.17	4.12	4.05	4.01
11	4.84	3.98	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.10	3.94	3.86	3.81	3.74	3.71
12	4.75	3.89	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	3.86	3.70	3.62	3.57	3.50	3.47
13	4.67	3.81	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.66	3.51	3.43	3.38	3.31	3.27
14	4.60	3.74	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.51	3.35	3.27	3.22	3.15	3.11
15	4.54	3.68	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.37	3.21	3.13	3.08	3.01	2.98
16	4.49	3.63	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.26	3.10	3.02	2.97	2.90	2.86
17	4.45	3.59	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.16	3.00	2.92	2.87	2.80	2.76
18	4.41	3.55	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.08	2.92	2.84	2.78	2.71	2.68
19	4.38	3.52	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.00	2.84	2.76	2.71	2.64	2.60
20	4.35	3.49	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	2.94	2.78	2.69	2.64	2.57	2.54
21	4.32	3.47	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	2.88	2.72	2.64	2.58	2.51	2.48
22	4.30	3.44	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	2.83	2.67	2.58	2.53	2.46	2.42
23	4.28	3.42	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	2.78	2.62	2.54	2.48	2.41	2.37
24	4.26	3.40	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	2.74	2.58	2.49	2.44	2.37	2.33
25	4.24	3.39	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.70	2.54	2.45	2.40	2.33	2.29
26	4.23	3.37	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.66	2.50	2.42	2.36	2.29	2.25
27	4.21	3.35	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.63	2.47	2.38	2.33	2.26	2.22
28	4.20	3.34	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.60	2.44	2.35	2.30	2.23	2.19
29	4.18	3.33	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.57	2.41	2.33	2.27	2.20	2.16
30	4.17	3.32	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.55	2.39	2.30	2.25	2.17	2.13
40	4.08	3.23	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.37	2.20	2.11	2.06	1.98	1.94
50	4.03	3.18	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.27	2.10	2.01	1.95	1.87	1.82
75	3.97	3.12	4.05	3.58	3.27	3.05	2.89	2.76	2.65	2.57	2.13	1.96	1.87	1.81	1.72	1.67
100	3.94	3.09	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.07	1.89	1.80	1.74	1.65	1.60

Table of the F-statistic for a significance level of 0.05 (Denominator to left, Numerator across top)

d.f.	1	2	3	4	5	6	7	8	9	10	20	30	40	50	75	100
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	248.02	250.10	251.14	251.77	252.62	253.04
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.45	19.46	19.47	19.48	19.48	19.49
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.66	8.62	8.59	8.58	8.56	8.55
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.80	5.75	5.72	5.70	5.68	5.66
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.56	4.50	4.46	4.44	4.42	4.41
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.87	3.81	3.77	3.75	3.73	3.71
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.44	3.38	3.34	3.32	3.29	3.27
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.15	3.08	3.04	3.02	2.99	2.97
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	2.94	2.86	2.83	2.80	2.77	2.76
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.77	2.70	2.66	2.64	2.60	2.59
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.65	2.57	2.53	2.51	2.47	2.46
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.54	2.47	2.43	2.40	2.37	2.35
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.46	2.38	2.34	2.31	2.28	2.26
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.39	2.31	2.27	2.24	2.21	2.19
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.33	2.25	2.20	2.18	2.14	2.12
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.28	2.19	2.15	2.12	2.09	2.07
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.23	2.15	2.10	2.08	2.04	2.02
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.19	2.11	2.06	2.04	2.00	1.98
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.16	2.07	2.03	2.00	1.96	1.94
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.12	2.04	1.99	1.97	1.93	1.91
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.10	2.01	1.96	1.94	1.90	1.88
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.07	1.98	1.94	1.91	1.87	1.85
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.05	1.96	1.91	1.88	1.84	1.82
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.03	1.94	1.89	1.86	1.82	1.80
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.01	1.92	1.87	1.84	1.80	1.78
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	1.99	1.90	1.85	1.82	1.78	1.76
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	1.97	1.88	1.84	1.81	1.76	1.74
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	1.96	1.87	1.82	1.79	1.75	1.73
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	1.94	1.85	1.81	1.77	1.73	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	1.93	1.84	1.79	1.76	1.72	1.70
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.84	1.74	1.69	1.66	1.61	1.59
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.78	1.69	1.63	1.60	1.55	1.52
75	3.97	3.12	2.73	2.49	2.34	2.22	2.13	2.06	2.01	1.96	1.71	1.61	1.55	1.52	1.47	1.44
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.68	1.57	1.52	1.48	1.42	1.39

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, k=1,2,3,\dots,n \quad (1.47)$$

Example: What is the probability of getting more than 15 heads in 20 tosses of a fair coin? Answer:

$$\sum_{k=16}^{20} \binom{20}{k} 0.5^k (1-0.5)^{20-k} = 0.006$$

The binomial distribution is helpful in assessing “field significance”, the significance of multiple tests as when an array of variables are tested against the same hypothesis. An example would be correlating the sunspot index with a map of pressure at many points over the earth. How many individual “significant” events do you expect to get by chance in such cases. As an example, consider the plot below, which shows for N tries of a test at the p=0.05 significance level, what the binomial distribution (1.47) says about how many you should get by chance.

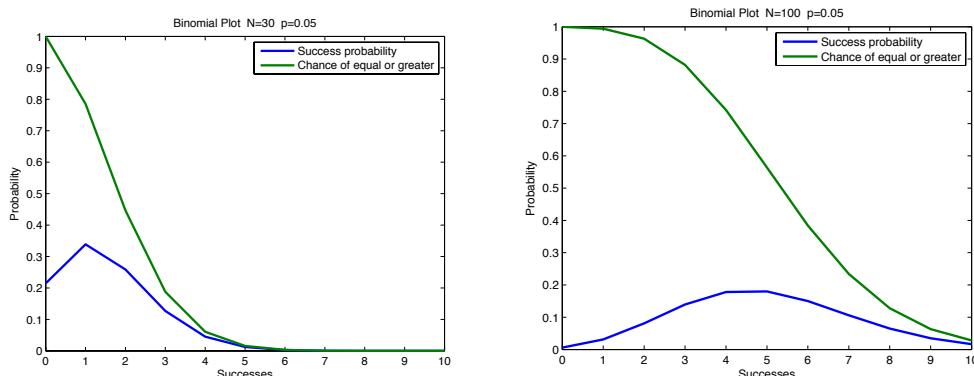


Figure Blue line shows probability of a given number of successes in N tries (N=30 on left and N=100 on right) where the probability of each event succeeding is p=0.05. Green line shows the probability of getting the number of successes or more.

Note that the probability of getting 5 successes or more in 30 tries is less than 0.05 and getting 10 successes or more in 100 tries is less than 0.05. That is 16.7% are successes for 30 tries and only 10% are successes for 100 tries at same probability level. For smaller samples, the fraction of total tries that can succeed by chance is greater. Even for 100 tries, 10% can succeed by chance, where the probability of each individual occurrence is p=5%. The most likely outcome is shown by the peak of the blue line and is what you expect, about 5% of the chances will succeed. But the chances of getting significantly more than that are quite good, and ten or fifteen percent of the field points could succeed by chance at the 5% level (See Wilks 2006, and Livezey and Chen 1983).

If you did the calculations above by hand you would find it tedious. This gets worse when the sample is larger. To assist in this it is useful to use the following Normal Approximation to the Binomial.

Normal Approximation to the Binomial (DeMoivre-Laplace)

From the central limit theorem, it can be shown that the distribution of sample means approaches the Normal Distribution, even if the population from which the means are derived is not normally distributed. This is also true for the Binomial distribution, for which values have a probability of being either zero or one, but nothing else. The distribution of sample means from a binomial population is nonetheless normally distributed about its mean value of 0.5.

Here is a Java Applet that allows you to play with the Normal Approximation http://www.ruf.rice.edu/~lane/stat_sim/normal_approx/. Try calculating the probability of getting more than 40 hits in a sample of 60 with P=0.5. That's 40 or more heads out of 60 coin tosses. You can compare the exact and approximate probabilities. Try this with small samples. For example, try 4 or more heads out of 6 tosses.

DeMoivre-Laplace Theorem.

X is a binomial variable defined on n independent trials each having success probability p . Then for any numbers a and b ,

$$\lim_{n \rightarrow \infty} P\left(a < \frac{X - np}{\sqrt{np(1-p)}} < b\right) = \frac{1}{\sqrt{2\pi np(1-p)}} \int_a^b e^{-x^2/2} dx \quad (1.48)$$

This means that the statistic, $\frac{X - np}{\sqrt{np(1-p)}}$, has the Normal distribution. We can use this fact to simplify the solution of binomial problems, as illustrated in the example below.

Example of the Normal approximation to the Binomial: An earthquake forecaster has forecast 200 earthquakes. How many times in 200 trials must she be successful so we can say with 95% certainty that she has nonzero skill?

The null hypothesis is that she has no skill and the confidence level is 0.05, or 95%. We then want,

$$P(s > s^* | H_0) = 0.025 = \sum_{s=s^*}^{200} \binom{200}{s} \left(\frac{1}{2}\right)^s \left(1 - \frac{1}{2}\right)^{200-s}$$

Solving this equation for $s > s^*$, the number of occurrences necessary to leave only a 0.025 probability to the right, is extremely tedious to do by hand, which nobody these days would do. However, we can use the Normal approximation to the Binomial to convert this to the problem,

$$P(s > s^* | H_0) = P\left(\frac{s - np}{\sqrt{np(1-p)}} > \frac{s^* - np}{\sqrt{np(1-p)}}\right) = P\left(Z > \frac{s^* - np}{\sqrt{np(1-p)}}\right)$$

Now $P(Z > 1.96) = 0.025$ (two-tailed 95%), so we want,

$$\frac{s - np}{\sqrt{np(1-p)}} > 1.96, \text{ or } s > 114$$

Where we have inserted $n=200$ and $p=0.5$ to get the numerical value shown. So to pass a no-skill test on a sample of this size, the forecaster must be right 57% of the time. Of course, this level of skill, while significantly different from zero, may not be practically useful.

1.8 Non-Parametric Statistical Tests

The statistical tests applied above mostly assume that the samples come from populations for which the statistical distributions are known, or assumed, a priori. We very often assume that the statistics we are testing are Normally distributed, so we can use the shape of the Normal distribution in our tests. Tests have also been developed that do not require the assumption of a theoretical distribution. These are called ‘non-parametric’ or ‘distribution-free’ statistical tests. This approach can be easily illustrated with the Signs Test.

663 3 Repaso de álgebra Lineal

664 Una matriz es un elemento matemático compuesto por filas y columnas. Una matriz
 665 rectangular de dimensiones $m \times n$ se define como

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

666 Un elemento de la matriz \mathbf{A} queda definida como A_{ij} , $i = 1, 2, \dots, m; j =$
 667 $1, 2, \dots, n$. El primer índice i denota filas y el segundo j columnas.

668 $\{\text{Definición}\}$ Un vector es una matriz con solo una columna ($m \times 1$)

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}.$$

669 $\{\text{Vectores ortogonales}\}$

670 Sean dos vectores $\mathbf{u} = [u_1, u_2, \dots, u_N]$ y $\mathbf{v} = [v_1, v_2, \dots, v_N]$ de longitud N , decimos que
 671 son ortogonales si

$$\mathbf{u} \cdot \mathbf{v} = (\mathbf{u}, \mathbf{v}) = \sum_{i=1}^N u_i v_i = \mathbf{u}^T \mathbf{v} = 0.$$

672 $\{\text{Norma, módulo, longitud de un vector}\}$

673 Sea el vector $\mathbf{u} = [u_1, u_2, \dots, u_N]$, entonces la norma de dicho vector se define como

$$\|\mathbf{u}\| = |\mathbf{u}| = \sqrt{u_1^2 + u_2^2 + \dots + u_N^2} = (\mathbf{u}, \mathbf{u})^{1/2} = (\mathbf{u}^T \mathbf{u})^{1/2}.$$

674 En general se puede obtener un $\{\text{vector unitario}\}$ (de módulo = 1) dividiendo el
 675 vector por su norma:

$$\mathbf{u}_I = \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

676 $\{\text{Vectores ortonormales}\}$

677 Dos vectores ortogonales \mathbf{u} y \mathbf{v} si tienen módulo la unidad, entonces se denominan
 678 vectores ortonormales.

679 $\{\text{Suma de vectores}\}$

680 La suma de 2 vectores $\mathbf{u} = [u_1, u_2, \dots, u_N]$ y $\mathbf{v} = [v_1, v_2, \dots, v_N]$ de longitud N se define como

$$\mathbf{u} + \mathbf{v} = [u_1 + v_1, u_2 + v_2, \dots, u_N + v_N] = \sum_{i=1}^N u_i + v_i.$$

682 *{Combinación lineal de vectores}*

683 *Un vector \mathbf{y} se dice que es combinación lineal de un conjunto de vectores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$*
 684 *si se puede expresar como la suma de los N vectores multiplicados por N coeficientes*
 685 *escalares a_1, a_2, \dots, a_N :*

$$\mathbf{y} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + \dots + a_N \mathbf{x}_N = \sum_{i=1}^N a_i \mathbf{x}_i.$$

686 *{Independencia lineal}*

687 *Un conjunto de vectores $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$ se dice que es linealmente independiente si*
 688 *existe una combinación lineal finita de los vectores del conjunto tal que:*

$$\sum_{i=1}^N a_i \mathbf{y}_i = a_1 \mathbf{y}_1 + a_2 \mathbf{y}_2 + \dots + a_N \mathbf{y}_N = 0,$$

689 *que se satisface cuando no todos los coeficientes son cero. En caso contrario, se dice*
 690 *que son linealmente dependientes.*

691 *{Ortonormalización Gram-Schmidt }*

692 *Es un método para convertir un conjunto de vectores \mathbf{v} en vectores ortonormales. De*
 693 *forma general el proceso definido por Gram-Schmidt para ortonormalizar el vector*
 694 *ortogonal \mathbf{w}_k a partir de un conjunto de vectores ortonormales $\mathbf{u}_i = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k-1}]$*
 695 *se define como:*

$$\mathbf{w}_k = \mathbf{v}_k - \sum_{i=1}^{k-1} \mathbf{u}_i \cdot \mathbf{v}_k \mathbf{u}_i.$$

696 $\mathbf{v}_k \equiv$ vectores originales

697 $\mathbf{u}_i \equiv$ vectores ortonormales

$\mathbf{w}_k \equiv$ vectores ortogonales a \mathbf{u}_j

698 *Ejemplo: | Convertir el conjunto de vectores de la base A en una base ortonormal:*

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 0 \\ 2 & 3 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

699 *Primero normalizamos el primer vector columna \mathbf{v}_1 :*

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \left[\frac{1}{\sqrt{6}}, 0, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right].$$

700 *| (1er vector ortonormal) | Ahora usamos la fórmula de arriba para encontrar el*
 701 *vector \mathbf{w}_2 ortogonal a \mathbf{u}_1*

$$\mathbf{w}_2 = \mathbf{v}_2 - \mathbf{u}_1 \cdot \mathbf{v}_2 \mathbf{u}_1 = [2, 2, 3, 1] - \left[\frac{1}{\sqrt{6}}, 0, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right] \cdot [2, 2, 3, 1] \left[\frac{1}{\sqrt{6}}, 0, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right] =$$

702 $= [2, 2, 3, 1] - \left(\frac{9}{\sqrt{6}} \right) \left[\frac{1}{\sqrt{6}}, 0, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right] = [2, 2, 3, 1] - \left[\frac{3}{2}, 0, 3, \frac{3}{2} \right] = \left[\frac{1}{2}, 2, 0, \frac{-1}{2} \right].$

703 (*vector ortogonal a u_1*)

704 *Normalizamos w_2 para obtener el primer vector ortonormal a u_1*

$$u_2 = \frac{w_2}{\|w_2\|} = \left[\frac{\sqrt{2}}{6}, \frac{2\sqrt{2}}{3}, 0, \frac{-\sqrt{2}}{6} \right]$$

705 | | (*2o vector ortonormal*) | *Ahora calculamos w_3 en términos de u_1 y u_2*

$$w_3 = v_3 - u_1 \cdot v_3 u_1 - u_2 \cdot v_3 u_2 = \left[\frac{4}{9}, \frac{-2}{9}, 0, \frac{-4}{9} \right],$$

706 (*vector ortogonal a u_1 y u_2*) | | *y si normalizamos* | |

$$u_3 = \left[\frac{2}{3}, \frac{-1}{3}, 0, \frac{-2}{3} \right].$$

707 | (*3er vector ortonormal*) | *La matriz o conjunto de vectores ortonormal es*

$$A = \begin{pmatrix} \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{6} & \frac{2}{3} \\ 0 & \frac{2\sqrt{2}}{3} & \frac{-1}{3} \\ \frac{\sqrt{6}}{3} & 0 & 0 \\ \frac{\sqrt{6}}{6} & \frac{-\sqrt{2}}{6} & \frac{-2}{3} \end{pmatrix}$$

708 (*base ortonormal*)

709 *{Aplicación lineal}* Sean dos espacios vectoriales V y W , decimos que una aplicación
710 $f : V \rightarrow W$ es lineal si la ‘imagen’ de la combinación lineal es la combinación lineal
711 de las imágenes. Es decir,

$$f(\alpha u + \beta v) = \alpha f(u) + \beta f(v).$$

712 *La imagen de una aplicación es el resultado de aplicar al vector una aplicación.*

713 *Ejemplos:*

714 (1) *La aplicación $f : R^3 \rightarrow R^2$ definida por $f(x, y, z) = (x + y, y + 2z)$ es lineal.*

715 *Demostración: Definimos $u = [u_1, u_2, u_3]$ y $v = [v_1, v_2, v_3]$. Entonces f es lineal si*

$$f(\alpha(u_1, u_2, u_3) + \beta(v_1, v_2, v_3)) = \alpha f(u_1, u_2, u_3) + \beta f(v_1, v_2, v_3) =$$

716 $f(\alpha u_1 + \beta v_1, \alpha u_2 + \beta v_2, \alpha u_3 + \beta v_3) = \alpha[u_1 + u_2, u_2 + 2u_3] + \beta[v_1 + v_2, v_2 + 2v_3]$

717 $[\alpha u_1 + \beta v_1 + \alpha u_2 + \beta v_2, \alpha u_2 + \beta v_2 + 2(\alpha u_3 + \beta v_3)] = [\alpha u_1 + \alpha u_2 + \beta v_1 + \beta v_2, \alpha u_2 + 2\alpha u_3 + \beta v_2 + 2\beta v_3]$

718 (2) *La aplicación $f : R^3 \rightarrow R^2$ definida por $f(x, y, z) = (x + y + 1, y + 2z)$ no es lineal.*

719 (3) *La aplicación $R : R^2 \rightarrow R^2$ definida por*

$$x' = Rx,$$

720 $R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix},$

721 rota el vector un ángulo θ en sentido contrario a las agujas del reloj. Si queremos 722 cambiar el sentido de la rotación solo debemos tomar el signo de θ negativo. Aquí 723 \mathbf{x}' es el vector rotado o la imagen de la aplicación rotación.

724 $\{\text{Matriz Identidad}\}$

725 Se define la matriz identidad \mathbf{I} como una matriz diagonal compuesta por unos

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

726 El producto matricial de cualquier matriz \mathbf{A} por la matriz identidad \mathbf{I} es igual a la 727 matriz original

$$\mathbf{A}\mathbf{I} = \mathbf{A}.$$

728 $\{\text{Matriz transpuesta}\}$

729 Sea \mathbf{A} una matriz $m \times n$ con elementos $[a_{ij}]$ para $i = 1, 2, \dots, m; j = 1, 2, \dots, n$. Se define 730 el elemento de su transpuesta como

$$\mathbf{A}^T = [a_{ji}].$$

731 La inversa de la matriz transpuesta es

$$(\mathbf{A}^T)^{-1} = \mathbf{A},$$

732 si \mathbf{A} es ortogonal. $\mid \{\text{Demostración}\}$:

$$(\mathbf{A}^T)^{-1} = \mathbf{A}$$

$$\mathbf{A}^T (\mathbf{A}^T)^{-1} = \mathbf{A}^T \mathbf{A} = \mathbf{I}$$

$$\mathbf{A} \mathbf{A}^T (\mathbf{A}^T)^{-1} = \mathbf{A} \mathbf{I}$$

735 $\{\text{Matriz Ortogonal}\}$

736 Una matriz \mathbf{A} es ortogonal si $\mathbf{A} \mathbf{A}^T = \mathbf{A}^T \mathbf{A} = \mathbf{I}$.

737 $\{\text{Matriz Diagonal}\}$

738 Una matriz es diagonal si únicamente contiene algunos elementos diferentes de cero 739 en la diagonal principal (ceros en la matriz triangular superior e inferior).

$$\mathbf{D} = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$$

740 $\{\text{Matriz Simétrica}\}$

741 Una matriz simétrica se define como aquella matriz \mathbf{A} tal que sea igual a su traspuesta

$$\mathbf{A} = \mathbf{A}^T \quad (a_{ij} = a_{ji})$$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

742 *{Matriz Antisimétrica}*

743 Una matriz antisimétrica se define como aquella matriz \mathbf{A} tal que sea igual a su
744 traspuesta

$$\mathbf{A} = -\mathbf{A}^T \quad (a_{ij} = -a_{ji})$$

$$\mathbf{A} = \begin{pmatrix} 0 & a_{12} & a_{13} \\ -a_{12} & 0 & a_{23} \\ -a_{13} & -a_{23} & 0 \end{pmatrix}$$

745 | NOTA: Cualquier matriz \mathbf{A} se puede descomponer en una parte simétrica y otra
746 antisimétrica:

$$A = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T) + \frac{1}{2}(\mathbf{A} - \mathbf{A}^T).$$

747 *{Matriz Singular}*

748 Una matriz singular es aquella matriz cuadrada cuyo determinante es igual a cero.
749 Las matrices singulares no tienen matriz inversa. | | Ejemplo:

$$\mathbf{S} = \begin{pmatrix} 3 & 2 \\ 6 & 4 \end{pmatrix}$$

750 Si nos fijamos las 2 filas de la matriz singular \mathbf{S} son linealmente dependientes, es
751 decir, podemos recuperar la segunda fila multiplicando la primera fila por 2. Si la
752 matriz tiene columnas o filas linealmente dependientes, el determinante es cero.

753 Si \mathbf{A} es ortogonal

$$\begin{aligned} \mathbf{A}^{-1} &= \mathbf{A}^T \\ \mathbf{A}^{-1}(\mathbf{A}^T)^{-1} &= \mathbf{I} \\ \mathbf{A}^{-1} \underbrace{(\mathbf{A}^T)^{-1}(\mathbf{A}^T)}_{\mathbf{I}} &= \mathbf{A}^T \end{aligned}$$

756 Una matriz cualquiera se puede convertir en matriz cuadrada si es multiplicada por
757 su transpuesta.

758 *{Determinante de una matriz}*

759 El determinante de una matriz \mathbf{A} 3×3 se puede calcular como

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - (a_{31}a_{22}a_{13} + a_{32}a_{23}a_{11} + a_{33}a_{21}a_{12}).$$

760 Para matrices de orden superior se puede utilizar la formula de adjuntos. El deter-
761 minante de una matriz $n \times n$ es el producto escalar entre cualquier fila o columna
762 con sus adjuntos

$$|\mathbf{A}| = a_{i1}C_{i1} + a_{i2}C_{i2} + \dots + a_{in}C_{in},$$

763 donde los adjuntos C_{ij} son subdeterminantes (de orden $n-1$, sin contar la columna j
 764 y fila i) con el signo adecuado

$$C_{ij} = (-1)^{i+j} |\mathbf{M}_{ij}|.$$

765 $\{\text{Rango de una matriz}\} \mid \mid \{\text{Definición 1}\}$: El rango de una matriz se define como
 766 el número de filas o columnas de la matriz que son linealmente independientes. $\mid \mid$
 767 $\{\text{Definición 2}\}$: El orden de la mayor submatriz cuadrada no nula

768 Así para calcular el rango debemos de hacer cero todos los elementos posibles de la
 769 matriz hasta obtener una submatriz no nula que nos indicará el rango de la matriz.

$$\left[\begin{array}{ccc} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{array} \right] \xrightarrow[2r_1+r_2]{\sim} \left[\begin{array}{ccc} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 3 & 5 & 0 \end{array} \right] \xrightarrow[-3r_1+r_3]{\sim} \left[\begin{array}{ccc} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & -1 & -3 \end{array} \right] \xrightarrow[r_2+r_3]{\sim} \left[\begin{array}{ccc} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{array} \right] \xrightarrow[-2r_2+r_1]{\sim} \left[\begin{array}{ccc} 1 & 0 & -5 \\ 0 & 1 & 3 \\ 0 & -1 & -3 \end{array} \right],$$

770 y entonces el rango de la matriz es 2. El rango lo podríamos haber encontrado simplemente viendo que sustrayendo la segunda fila de la matriz a la primera obtenemos la
 771 tercera fila; solamente hay 2 vectores linealmente independientes.

773 $\{\text{Matriz inversa}\}$

774 La matriz cuadrada \mathbf{A} es invertible si existe una matriz \mathbf{A}^{-1} tal que

$$\mathbf{A}^{-1} \mathbf{A} = \mathbf{I} \text{ and } \mathbf{A} \mathbf{A}^{-1} = \mathbf{I}.$$

775 $\{\text{Proposición}\}$ Si \mathbf{A} es invertible entonces la única solución de $\mathbf{A}\mathbf{x} = \mathbf{b}$ es

$$\mathbf{x} = \mathbf{A}^{-1} \mathbf{b}.$$

776 $\{\text{Proposición}\}$ Si \mathbf{A} es una matriz cuadrada de dimensión 2×2 y $|\mathbf{A}| = a_{11}a_{22} -$
 777 $a_{12}a_{21}$ no es cero, entonces

$$\mathbf{A}^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{|\mathbf{A}|} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

778 $\{\text{Cálculo de matriz inversa con método Gauss-Jordan}\}$

779 El método de Gauss-Jordan usa la siguiente identidad matemática $[\mathbf{A}|\mathbf{I}] \rightarrow [\mathbf{I}|\mathbf{A}^{-1}]$. \mid
 780 $\{\text{Ejemplo:}\}$ Calcula la matriz inversa de

$$\begin{pmatrix} 2 & 3 \\ 4 & 7 \end{pmatrix}$$

$$[\mathbf{A}\mathbf{I}] = \begin{pmatrix} 2 & 3 & 1 & 0 \\ 4 & 7 & 0 & 1 \end{pmatrix} \xrightarrow[-2r_1+r_2]{\sim} \begin{pmatrix} 2 & 3 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{pmatrix} \xrightarrow[-3r_2+r_1]{\sim} \begin{pmatrix} 2 & 0 & 7 & -3 \\ 0 & 1 & -2 & 1 \end{pmatrix} \rightarrow$$

$$\xrightarrow[r_1/2]{\sim} \begin{pmatrix} 1 & 0 & 7/2 & -3/2 \\ 0 & 1 & -2 & 1 \end{pmatrix}.$$

782 Entonces

$$\mathbf{A}^{-1} = \begin{pmatrix} 7/2 & -3/2 \\ -2 & 1 \end{pmatrix}.$$

783 **Ejemplo 2:**

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 0 & 3 \\ 0 & 1 & -1 \end{bmatrix} \rightarrow \mathbf{A}^{-1} = \begin{bmatrix} 3 & -1 & 3 \\ -2 & 1 & -1 \\ -2 & 1 & -2 \end{bmatrix}.$$

784 **{Factorización LU}**

785 *Es la descomposición de una matriz en una matriz triangular inferior \mathbf{L} y una ma-*
 786 *triz triangular superior \mathbf{U} , es decir*

$$\mathbf{A} = \mathbf{LU},$$

787 *donde*

$$\mathbf{L} = \begin{pmatrix} a_{12} & 0 \\ a_{21} & a_{22} \end{pmatrix}$$

788 *es una matriz triangular inferior (L; lower) y*

$$\mathbf{U} = \begin{pmatrix} a_{12} & a_{21} \\ 0 & a_{12} \end{pmatrix},$$

789 *es una matriz triangular superior (U; superior). | Esta factorización se suele usar*
 790 *para resolver un sistema de ecuaciones lineal $\mathbf{Ax} = \mathbf{b}$ y no es única. Es decir,*
 791 *pueden existir más de una factorización. Si sustituimos la definición de arriba*

$$\mathbf{Ax} = (\mathbf{LU})\mathbf{x} = \mathbf{b},$$

792 *lo que implica que*

$$\mathbf{L}(\mathbf{Ux}) = \mathbf{b}.$$

793 *| Si definimos $\mathbf{Ux} = \mathbf{z}$, entonces tenemos que*

$$\mathbf{Lz} = \mathbf{b}.$$

794 *Como \mathbf{L} es una matriz triangular inferior, podemos resolver para \mathbf{z} utilizando susti-*
 795 *tución hacia delante. Luego, como \mathbf{U} es una matriz triangular superior, resolvemos*
 796 *$\mathbf{Ux} = \mathbf{z}$ por sustitución en reversa.*

797 **{Ejemplo:}**

798 *Encuentre la descomposición LU de la matriz*

$$\begin{pmatrix} 2 & 5 \\ -3 & -4 \end{pmatrix}.$$

799 *Si multiplicamos la primera fila por $L_{21} = 3/2$ y le sumamos la segunda fila hacemos*
 800 *cero el elemento $a_{21} = -3$*

$$\mathbf{U} = \begin{pmatrix} 2 & 5 \\ 0 & 7/2 \end{pmatrix}.$$

801 *Esta matriz ya es una matriz triangular superior, es decir, la matriz \mathbf{U} . Para en-*
 802 *contrar la matriz triangular inferior solo debemos de conocer el valor del elemento*
 803 *L_{21} . Ese elemento es el multiplicador con signo opuesto usado en la eliminación de*
 804 *Gauss-Jordan. Es decir, $L_{21} = -3/2$. La matriz \mathbf{L} es*

$$\mathbf{L} = \begin{pmatrix} 1 & 0 \\ -3/2 & 1 \end{pmatrix}.$$

805 *Vamos a comprobar*

$$\begin{pmatrix} 1 & 0 \\ -3/2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 5 \\ 0 & 7/2 \end{pmatrix} = \begin{pmatrix} 2 & 5 \\ -3 & -4 \end{pmatrix}.$$

806 *{Ejemplo:}*

807 *Resuelva el siguiente sistema de ecuaciones con la descomposición LU*

$$2x_1 + 3x_2 + 4x_3 = 6$$

$$4x_1 + 5x_2 + 10x_3 = 16$$

$$4x_1 + 8x_2 + 2x_3 = 2$$

810 *La matriz de coeficientes es*

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 4 \\ 4 & 5 & 10 \\ 4 & 8 & 2 \end{pmatrix}.$$

811 *Si factorizamos*

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & -2 & 1 \end{pmatrix},$$

812 *y*

$$\mathbf{U} = \begin{pmatrix} 2 & 3 & 4 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

813 *Si utilizamos la identidad $\mathbf{L}\mathbf{z} = \mathbf{b}$ obtenemos*

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & -2 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 16 \\ 2 \end{pmatrix}.$$

814 *Si resolvemos para \mathbf{z}*

$$z_1 = 6$$

$$z_2 = 16 - 2z_1 = 4$$

$$z_3 = 2 + 2z_2 - 2z_1 = -2$$

817 *Así que*

$$\mathbf{z} = \begin{pmatrix} 6 \\ 4 \\ -2 \end{pmatrix}.$$

818 *Si utilizamos ahora la definición $\mathbf{U}\mathbf{x} = \mathbf{z}$,*

$$\begin{pmatrix} 2 & 3 & 4 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \\ -2 \end{pmatrix},$$

819 *y obtenemos*

$$x_3 = 1$$

$$x_2 = \frac{4 - 2x_3}{-1} = -2$$

821

$$x_1 = \frac{6 - 4x_3 - 3x_2}{2} = 4$$

822

Por lo tanto la solución del sistema de ecuaciones es

$$\mathbf{x} = \begin{pmatrix} 4 \\ -2 \\ 1 \end{pmatrix}.$$

823

{Valores propios y vectores propios}

824

Sea \mathbf{A} una matriz cuadrada, un número real λ se dice que es un valor propio de \mathbf{A} si existe un vector, diferente del vector cero, \mathbf{x} tal que

825

$$\mathbf{Ax} = \lambda\mathbf{x}.$$

826

Es decir, \mathbf{x} es un vector que al transformarlo mediante la multiplicación por \mathbf{A} el vector resultante mantiene la misma dirección; solamente se modifica su longitud (magnitud) y/o sentido. El valor propio λ nos informa si el vector propio \mathbf{x} se acorta o alarga o cambia de signo cuando es multiplicado por \mathbf{A} .

830

{Definición} *El número λ es un valor propio si y solo si*

$$|\mathbf{A} - \lambda\mathbf{I}| = 0.$$

831

{Ejemplo:}

832

Calcula los valores propios y vectores propios de la matriz

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

833

Sabemos

$$|\mathbf{A} - \lambda\mathbf{I}| = \begin{vmatrix} 1 - \lambda & 2 \\ 2 & 4 - \lambda \end{vmatrix} = \lambda^2 - 5\lambda = 0.$$

834

El polinomio de arriba se llama polinomio característico y es igual a cero cuando λ es un valor propio. Resolviendo obtenemos dos soluciones $\lambda = 0$ y $\lambda = 5$. Ahora para encontrar los vectores propios debemos resolver el sistema $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0$ separadamente para las dos λ :

$$(\mathbf{A} - 0\mathbf{I})\mathbf{x} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \mathbf{x} = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \text{ para } \lambda_1 = 0,$$

838

y

$$(\mathbf{A} - 5\mathbf{I})\mathbf{x} = \begin{pmatrix} -4 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow \mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ para } \lambda_2 = 5.$$

839

Ejercicio: Calcule los valores y vectores propios de la matriz de rotación

$$\mathbf{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

840 $\{\text{Multiplicación de matrices}\}$

841 *El producto matricial de dos matrices \mathbf{A} ($m \times n$) y \mathbf{B} ($n \times p$) se define como*

$$\mathbf{C} = \mathbf{AB},$$

842 *donde \mathbf{C} es una matriz $m \times p$, con el elemento (i, j) definido por*

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj},$$

843 *para todo $i = 1, 2, \dots, m; j = 1, 2, \dots, p$.*

844 $\{\text{Multiplicación de matrices}\}$

845 *Sea \mathbf{A} una matriz $m \times n$, y \mathbf{v} un vector $n \times 1$, entonces el elemento del producto*

$$\mathbf{z} = \mathbf{Av}$$

846 *viene dado por*

$$z_i = \sum_{k=1}^n a_{ik} v_k,$$

847 *para todo $i = 1, 2, \dots, m$. Similarmente, si \mathbf{u} es un vector $m \times 1$, entonces el elemento
848 del producto*

$$\mathbf{z}^T = \mathbf{u}^T \mathbf{A}$$

849 *viene dado por*

$$z_i = \sum_{k=1}^n a_{ki} u_k,$$

850 *para todo $i = 1, 2, \dots, n$. Finalmente, el escalar que resulta del producto $\alpha = \mathbf{u}^T \mathbf{Av}$,*
851 *viene dado por*

$$\alpha = \sum_{j=1}^m \sum_{k=1}^n a_{jk} u_j v_k.$$

852 $\{\text{Proposición:}\}$

853 *Sea $\mathbf{C} = \mathbf{AB}$ una matriz $m \times p$, entonces*

$$\mathbf{C}^T = \mathbf{B}^T \mathbf{A}^T.$$

854 $\{\text{Proposición}\}$ *Sea \mathbf{A} y \mathbf{B} matrices cuadradas $n \times n$ invertibles. Sea el producto
855 matricial $\mathbf{C} = \mathbf{AB}$, entonces*

$$\mathbf{C}^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}.$$

856 $\{\text{Derivada de matrices}\}$

857 $\{\text{Proposición:}\}$

858 *Sea*

$$\mathbf{y} = \mathbf{Ax},$$

859 *donde \mathbf{y} es $m \times 1$, \mathbf{x} es $n \times 1$, \mathbf{A} es $m \times n$, y \mathbf{A} no depende de \mathbf{x} , entonces la derivada
860 de \mathbf{y} es*

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}.$$

861 $\{\text{Proposición:}\}$

862 Sea

$$\mathbf{y} = \mathbf{A}\mathbf{x},$$

863 donde \mathbf{y} es $m \times 1$, \mathbf{x} es $n \times 1$, \mathbf{A} es $m \times n$, y \mathbf{A} no depende de \mathbf{x} . Supongamos que \mathbf{x}
864 es una función del vector \mathbf{z} , mientras que \mathbf{A} es independiente de \mathbf{z} . Entonces

$$\frac{\partial \mathbf{y}}{\partial \mathbf{z}} = \mathbf{A} \frac{\partial \mathbf{x}}{\partial \mathbf{z}}.$$

865 $\{\text{Proposición:}\}$

866 Sea el escalar α definido como

$$\alpha = \mathbf{x}^T \mathbf{A} \mathbf{x},$$

867 donde \mathbf{x} es $n \times 1$, \mathbf{A} es $n \times n$, y \mathbf{A} es independiente de \mathbf{x} , entonces

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}^T (\mathbf{A} + \mathbf{A}^T) = \mathbf{x}^T \mathbf{A}^T + \mathbf{x}^T \mathbf{A} = \mathbf{x}(\mathbf{A}^T + \mathbf{A})$$

868 $\{\text{Proposición:}\}$

869 Sea el escalar α definido como

$$\alpha = \mathbf{y}^T \mathbf{A} \mathbf{x},$$

870 donde \mathbf{y} es $m \times 1$, \mathbf{x} es $n \times 1$, y \mathbf{A} es $m \times n$, y \mathbf{A} es independiente de \mathbf{x} y \mathbf{y} , entonces

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y}^T \mathbf{A}$$

871 y

$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^T \mathbf{A}.$$

872 $\{\text{Definición}\}$ Sea \mathbf{A} una matriz $m \times n$ cuyos elementos son funciones de un escalar
873 α . Entonces la derivada de \mathbf{A} con respecto α es una matriz $m \times n$ compuesta por las
874 derivadas de elemento por elemento:

$$\frac{\partial \mathbf{A}}{\partial \alpha} = \begin{pmatrix} \frac{\partial a_{11}}{\partial \alpha} & \frac{\partial a_{12}}{\partial \alpha} & \dots & \frac{\partial a_{1n}}{\partial \alpha} \\ \frac{\partial a_{21}}{\partial \alpha} & \frac{\partial a_{22}}{\partial \alpha} & \dots & \frac{\partial a_{2n}}{\partial \alpha} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial a_{m1}}{\partial \alpha} & \frac{\partial a_{m2}}{\partial \alpha} & \dots & \frac{\partial a_{mn}}{\partial \alpha} \end{pmatrix}.$$

875 4 Cuadrados mínimos y regresión

876 En esta sección se van a introducir algunos modelos lineales estadísticos o modelos
877 de regresión. Aquí se incluyen ajustes lineales por mínimos cuadrados, coeficientes
878 de correlación, regresión múltiple, etc.

879 4.1 Métodos de cuadrados mínimos

880 Estos métodos son utilizados para ajustar un modelo dependiente de un conjunto
881 compuesto por k variables independientes $x_i; i = 1, 2, \dots, k$.

882 $\{(A) \text{ Mínimos cuadrados lineales}\} \mid$ Empezamos aplicando el método
883 en términos de estimación lineal. Nos referimos a lineal en cuanto a

884 los coeficientes b_0, b_1, \dots, b_k , es decir, $y = b_0 + b_1x_1 + \epsilon$ es lineal, pero
 885 $y = b_0 + \sin(b_1x_1)$ no lo es. |

886 (1) {Ajuste de una recta a un conjunto de datos} | Queremos usar
 887 los mejores coeficientes b_0 y b_1 en el sentido que se reduzca la
 888 desviación estandar de la recta ajustada versus los datos. Sea
 889 $i = 1, 2, \dots, N$ observaciones, entonces

$$y_i = \hat{y}_i + \epsilon,$$

890 donde

$$\hat{y}_i = b_0 + b_1x_i$$

891 es el estimador estadístico y ϵ es el residuo (medida de la diferen-
 892 cia de la recta ajustada versus conjunto de puntos). Para encon-
 893 trar b_0, b_1 debemos de minimizar la suma de los errores cuadrados
 894 (SEC), donde SEC es la varianza total que no es explicada por
 895 nuestro modelo de regresión lineal

$$SEC = \sum_{i=1}^N \epsilon_i^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N [y_i - (b_0 + b_1x_i)]^2.$$

896 |

897 La suma de los cuadrados totales se define como

$$SCT = \sum_{i=1}^N (y_i - \bar{y})^2,$$

898 y la suma de los cuadrados residuales

$$SCR = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2.$$

899 | La SCT es proporcional a la varianza de los datos y SCR es propor-
 900 cional a la cantidad de varianza explicada por nuestra regresión. La
 901 varianza total (o SCT) se puede descomponer en función de SCR y SEC
 902 como:

$$SCT = \sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N (\hat{y}_i + \epsilon_i - \bar{y})^2 = \sum_{i=1}^N ((\hat{y}_i - \bar{y}) + \epsilon_i)^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 + \\ + \sum_{i=1}^N \epsilon_i^2 + 2 \sum_{i=1}^N (\hat{y}_i - \bar{y})\epsilon_i,$$

903 904 y por lo tanto

$$SCT = SCR + SEC$$

905 | ya que por las ecuaciones normales $2 \sum_{i=1}^N (\hat{y}_i - \bar{y})\epsilon_i = 0$. |

906 El problema de mínimos cuadrados consiste en minimizar la SEC con
 907 respecto los coeficientes, cuyas condiciones son

$$\frac{\partial SEC}{\partial b_0} = 0; \frac{\partial SEC}{\partial b_1} = 0.$$

908

|

909 *Substituyendo obtenemos para b_0*

$$\begin{aligned} \frac{\partial SEC}{\partial b_0} &= \frac{\partial}{\partial b_0} \left\{ \sum_{i=1}^N [y_i - (b_0 + b_1 x_i)]^2 \right\} = -2 \sum_{i=1}^N [y_i - (b_0 + b_1 x_i)] = \\ 910 \quad &= -2 \left(\sum_{i=1}^N y_i - Nb_0 - b_1 \sum_{i=1}^N x_i \right) = 0, \end{aligned}$$

911

y para b_1

$$\begin{aligned} \frac{\partial SEC}{\partial b_1} &= \frac{\partial}{\partial b_1} \left\{ \sum_{i=1}^N [y_i - (b_0 + b_1 x_i)]^2 \right\} = -2 \sum_{i=1}^N x_i [y_i - (b_0 + b_1 x_i)] = \\ 912 \quad &= -2 \left(\sum_{i=1}^N x_i y_i - b_0 \sum_{i=1}^N x_i - b_1 \sum_{i=1}^N x_i^2 \right) = 0. \end{aligned}$$

913

|

914 *Si resolvemos para b_0*

$$b_0 = \bar{y} - b_1 \bar{x},$$

915

y sustituimos en la segunda ecuación, obtenemos:

$$\begin{aligned} b_1 &= \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2} = \frac{\sum_{i=1}^N x_i y_i - \frac{1}{N} \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2} = \\ &= \frac{(N-1)C_{xy}}{(N-1)s_x^2} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \frac{\langle x'y' \rangle}{\langle x'^2 \rangle}, \end{aligned}$$

916

donde $C_{xy} = \langle x'y' \rangle$ es la covarianza entre x e y , y $s_x^2 = \langle x'^2 \rangle$ es la varianza de x . Así, una vez hemos calculado las medias de las variables x e y , podemos encontrar el coeficiente b_1 y, a partir de este, calcular el segundo coeficiente b_0 . | | Si sustituimos $b_0 = \bar{y} - b_1 \bar{x}$ en la ecuación de la regresión lineal $\hat{y}_i = b_0 + b_1 x_i$ obtenemos

$$\hat{y}_i = \bar{y} + b_1(x_i - \bar{x}),$$

922

o

$$\hat{y}_i = \bar{y} + b_1 x'_i,$$

923

o finalmente

$$\hat{y}'_i = b_1 x'_i,$$

924

que nos informa que cuando $x'_i = 0$ ($x_i = \bar{x}$) entonces $\hat{y}'_i = 0$ ($\hat{y} = \bar{y}$), es decir, la recta pasa por el punto (\bar{x}, \bar{y}) , de tal forma, que puesto que $\partial SEC / \partial b_0 = 0$ minimiza la suma del error, $\sum \epsilon_i = 0$, los puntos están

927 dispersos respecto a la recta ajustada de tal forma que los residuos positivos ($\epsilon > 0$) siempre se cancelan con los residuos negativos ($\epsilon < 0$). El
 928 parámetro b_0 se interpreta como la intersección (corte con el eje y) y b_1
 929 es la pendiente de la recta ajustada. | | El cociente
 930

$$100(SCR/SCT),$$

931 es el porcentaje de varianza explicada por nuestra regresión lineal (varianza explicada/varianza total) y nos informa de la bondad del ajuste
 932 denominado coeficiente de correlación, r^2 . Si la regresión se ajusta per-
 933 fectamente a todos los datos, todos los residuos son cero y por lo tanto
 934 $SEC = 0$ y $SCR/SCT = r^2 = 1$. A medida que el ajuste empeora el
 935 coeficiente r^2 disminuye hasta un mínimo posible de $r^2 = 0$.
 936

937 {**Error estandar de la estimación}**

938 Una medida de la magnitud absoluta de la bondad del ajuste es el error
 939 estandar del estimado, s_ϵ , definido como

$$s_\epsilon = [SEC/(N - 2)]^{1/2} = \left[\frac{1}{N - 2} \sum_{i=1}^N (y - \hat{y})^2 \right]^{1/2}.$$

940 El número de grados de libertad, $N - 2$, se debe a que necesitamos es-
 941 timar dos parámetros para encontrar realizar la regresión lineal. Si ϵ es
 942 una variable aleatoria que sigue una distribución Normal de media cero
 943 y desviación estandar s_ϵ , entonces, el 68.3% de las observaciones caen
 944 dentro del intervalo $\pm 1s_\epsilon$ unidades de la recta ajustada, 95.4% caerá
 945 dentro del intervalo $\pm 2s_\epsilon$ unidades de la recta, y 99.7% caerán en el
 946 intervalo $\pm 3s_\epsilon$ unidades de la recta. s_ϵ es la desviación estandar de y
 947 alrededor de su media, i.e., la recta ajustada $b_0 + b_1x$.

948 {**Generalización de mínimos cuadrados en notación matricial**}

949 Supongamos el modelo dependiente de k variables independientes X_k

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + \epsilon,$$

950 y supongamos que hacemos N observaciones independientes y_1, y_2, \dots, y_N
 951 de Y . Por lo tanto podemos escribir el modelo como

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_kx_{ik} + \epsilon_i,$$

952 donde x_{ij} es la observación i del de la variable independiente j . Es de-
 953 cir,

954 N ; observaciones

955 k ; variables independientes

$k + 1$; coeficientes

956 Si escribimos en notación matricial

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ \dots \\ y_N \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 1 & x_{N1} & \dots & x_{Nk} \end{pmatrix}$$

957

$$\mathbf{B} = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ \dots \\ b_k \end{pmatrix} \quad \mathbf{E} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \dots \\ \epsilon_N \end{pmatrix}.$$

958 | | De esta forma, podemos escribir nuestro modelo en notación matri-
959 cial como

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}.$$

960 | | Si restringimos el modelo a una variable independiente ($k = 1$),
961 i.e. dos coeficientes (b_0, b_1), entonces

$$\mathbf{B} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix},$$

962 es la matriz de coeficientes, e

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \dots & \dots \\ \dots & \dots \\ 1 & x_{N1} \end{pmatrix},$$

963 es la matriz de variables independientes. Si utilizamos la definición de
964 los residuos como

$$\mathbf{E} = \mathbf{Y} - \mathbf{XB},$$

965 podemos encontrar la suma de los residuos al cuadrado como

$$SEC = \sum_{i=1}^N \epsilon_i^2 = \sum_{i=1}^N \epsilon_i \epsilon_i = \mathbf{E}^T \mathbf{E} = (\mathbf{Y} - \mathbf{XB})^T (\mathbf{Y} - \mathbf{XB}) =$$

966

$$\mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{XB} - \mathbf{B}^T \mathbf{X}^T \mathbf{Y} + \mathbf{B}^T \mathbf{X}^T \mathbf{XB}.$$

967 Si queremos minimizar esa suma de errores entonces

$$\frac{\partial SEC}{\partial \mathbf{B}} = 0.$$

968 Para calcular las derivadas de SEC respecto de los coeficientes \mathbf{b} vamos
969 a asumir las siguientes consideraciones | | (i) $\mathbf{Y}^T \mathbf{XB} = \mathbf{B}^T \mathbf{X}^T \mathbf{Y}$

970 ya que son matrices elemento (1×1) y siempre son simétricas. Por lo
 971 tanto $-\mathbf{Y}^T \mathbf{X} \mathbf{B} - \mathbf{B}^T \mathbf{X}^T \mathbf{Y} = -2\mathbf{B}^T \mathbf{X}^T \mathbf{Y}$, y su derivada

$$-2\mathbf{X}^T \mathbf{Y}.$$

972 | | (ii) $\partial \mathbf{B}^T \mathbf{X}^T \mathbf{X} \mathbf{B} / \partial \mathbf{B} = 2\mathbf{X}^T \mathbf{X} \mathbf{B}$. Para demostrar esto tomemos el
 973 caso para el ajuste lineal (2 parámetros). Definamos los elementos de
 974 $\mathbf{X}^T \mathbf{X}$ como c_{ij} $i, j = 1, 2$ y $c_{12} = c_{21}$ por ser simétrica. Entonces

$$\mathbf{B}^T \mathbf{X}^T \mathbf{X} \mathbf{B} = c_{11} b_0^2 + c_{22} b_1^2 + 2c_{12} b_0 b_1,$$

975 y su derivada respecto b_0 es

$$2c_{11} b_0 + 2c_{12} b_1$$

976 y respecto a b_1 es

$$2c_{12} b_0 + 2c_{22} b_1.$$

977 Si acomodamos las derivadas en un vector columna 2×1 obtenemos

$$\begin{pmatrix} 2c_{11} b_0 + 2c_{12} b_1 \\ 2c_{12} b_0 + 2c_{22} b_1 \end{pmatrix} = 2\mathbf{X}^T \mathbf{X} \mathbf{B}.$$

978 y por lo tanto

$$\frac{\partial S}{\partial \mathbf{B}} = -2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X} \mathbf{B},$$

979 y las ecuaciones normales quedan

$$-2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X} \mathbf{B} = 0.$$

980 Finalmente, el problema de mínimos cuadrados es

$$(\mathbf{X}^T \mathbf{X}) \mathbf{B} = \mathbf{X}^T \mathbf{Y}.$$

981 Y resolviendo para \mathbf{B} la forma general del método de regresión por mínimos cuadrados es

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y},$$

983 o equivalentemente

$$\mathbf{B}^T = \mathbf{Y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}.$$

984 **Propiedades estadísticas** | Algunas de las consideraciones de un modelo
 985 de regresión múltiple son:

- 986 1. *Distribuciones aleatorias con media cero.* El vector $N \times 1$ de er-
 987 rores, ϵ , sigue una distribución aleatoria de media cero tal que
 988 $E[\epsilon] = 0$, es decir, $E[\epsilon_i] = 0$ ($i = 1, 2, \dots, N$).
- 989 2. *Homoscedasticidad.* La matriz de covarianzas de los errores $E[\epsilon \epsilon']$
 990 existe y los elementos de la diagonal principal son igual a σ^2 , es
 991 decir, $E[\epsilon_i^2] = \sigma^2$ ($i = 1, 2, \dots, N$).
- 992 3. *No correlación.* Los elementos fuera de la diagonal principal de la
 993 matriz de errores $E[\epsilon \epsilon']$ son iguales a cero, es decir, $E[\epsilon_i \epsilon_j] = 0$
 994 for todo $i \neq j$.

995 4. *Parámetros constantes.* Los elementos del vector $k \times 1$, β , y el
996 escalar σ son números fijos desconocidos con $\sigma > 0$.

5. *Modelo lineal.* Los valores del vector dependiente y han sido generados a partir de el modelo de regresión lineal multivariado:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

6. *Normalidad.* Las distribuciones aleatorias de los errores siguen una distribución Normal. Las consideraciones 2 y 3 pueden expresarse en notación matricial

$$E[\epsilon\epsilon'] = \sigma^2 \mathbf{I},$$

donde \mathbf{I} es la matrix identidad $N \times N$. Si además la consideración 1 se cumple, entonces ϵ sigue una distribución Normal

$$\epsilon \sim N(0, \sigma^2 \mathbf{I}).$$

997 Las consideraciones 3 y 6 implican que las distribuciones aleatorias
998 ϵ_i , $i = 1, 2, \dots, N$ son linealmente independientes.

999 **Mínimos cuadrados no tienen sesgo** | La esperanza del vector de coe-
1000 ficientes $k \times 1$, b , se obtiene a partir de las consideraciones 1, 4, y 5.
1001 La consideración 5 implica que el estimador de los coeficientes $b =$
1002 $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ y se puede escribir como

$$b = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X}\beta + \epsilon) = \mathbf{I}\beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon.$$

1003 |

1004 De las consideraciones 1, y 4 obtenemos

$$E[b] = E[\beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon] = \beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \underbrace{E[\epsilon]}_0 = \beta,$$

1005 lo que implica que b es insesgado.

1006 **Matriz de varianza de b** | Usando el resultado de arriba, bajo las con-
1007 sideraciones 1-6, la matriz de varianza de los estimadores de los coefi-
1008 cientes b (o momento centrado de orden 2) viene dada por

$$\begin{aligned} \text{var}(b) &= E[(b - \beta)(b - \beta)^T] = E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon \epsilon^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}] = \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[\epsilon \epsilon^T] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\sigma^2 \mathbf{I}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}. \end{aligned}$$

1010 | | Los elementos de la diagonal principal de esta matriz son las vari-
1011 anzas asociadas a los estimadores de los coeficientes b , y los elementos
1012 fuera de la diagonal principal representan la covarianza entre esos es-
1013 timadores. | Si asumimos que las variables del problema han sido es-
1014 tandardizadas, es decir la media ha sido extraída de las variables x_{ij} e y_i

1015 *y hemos dividido por las desviaciones estandar*

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} \sum_{i=1}^N x_{i1}x_{i1} & \sum_{i=1}^N x_{i1}x_{i2} & \dots & \sum_{i=1}^N x_{i1}x_{ik} \\ \sum_{i=1}^N x_{i2}x_{i1} & \sum_{i=1}^N x_{i2}x_{i2} & \dots & \sum_{i=1}^N x_{i2}x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^N x_{ik}x_{i1} & \sum_{i=1}^N x_{ik}x_{i2} & \dots & \sum_{i=1}^N x_{ik}x_{ik} \end{pmatrix},$$

1016 *es una matriz $k \times k$ y en notación índice se puede escribir como*

$$[\mathbf{X}^T \mathbf{X}]_{nm} = (N - 1)\rho_{x_{in}x_{im}} ; (i = 1, 2, \dots, N),$$

1017 *que se puede interpretar como la matriz de correlaciones entre las vari-
1018 ables independientes. La matriz*

$$\mathbf{X}^T \mathbf{Y} = \begin{pmatrix} \sum_{i=1}^N y_i x_{i1} \\ \sum_{i=1}^N y_i x_{i2} \\ \vdots \\ \vdots \\ \sum_{i=1}^N y_i x_{ik} \end{pmatrix},$$

1019 *es una matriz $k \times 1$ y en notación índice es*

$$[\mathbf{Y}^T \mathbf{X}]_n = (N - 1)\rho_{y_i x_{in}} ; (i = 1, 2, \dots, N).$$

1020 *Esta matriz se puede interpretar como la matriz de correlación entre las
1021 variables independientes y dependientes. | El modelo multivariado en
1022 notación índice es*

$$(N - 1)\rho_{x_{in}x_{im}} b_m = (N - 1)\rho_{y_i x_{in}}$$

1023 *O*

$$\rho_{x_{in}x_{im}} b_m = \rho_{y_i x_{in}} ; m = n = 1, 2, \dots, k; i = 1, 2, \dots, N$$

1024 *y para una única observación i dada obtenemos*

$$\rho_{x_n x_m} b_m = \rho_{y x_n}$$

1025 *Ahora supongamos, por simplicidad, que solo tenemos dos variables in-
1026 dependientes. Entonces:*

$$\rho_{x_1 x_1} b_1 + \rho_{x_1 x_2} b_2 = \rho_{y x_1}$$

1027

$$\rho_{x_2 x_1} b_1 + \rho_{x_2 x_2} b_2 = \rho_{y x_2},$$

1028 *y puesto que* $\rho_{x_1x_1} = \rho_{x_2x_1} = 1$, *y* $\rho_{x_1x_2} = \rho_{x_2x_1}$, *podemos escribir el sistema como:*
 1029

$$\begin{pmatrix} 1 & \rho_{x_1x_2} \\ \rho_{x_1x_2} & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \rho_{yx_1} \\ \rho_{yx_2} \end{pmatrix}.$$

1030 *De forma que los coeficientes quedan como*

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 1 & \rho_{x_1x_2} \\ \rho_{x_1x_2} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_{yx_1} \\ \rho_{yx_2} \end{pmatrix} = \frac{1}{1 - \rho_{x_1x_2}^2} \begin{pmatrix} 1 & -\rho_{x_1x_2} \\ -\rho_{x_1x_2} & 1 \end{pmatrix} \begin{pmatrix} \rho_{yx_1} \\ \rho_{yx_2} \end{pmatrix}.$$

1031

$$b_1 = \frac{1}{1 - \rho_{x_1x_2}^2} (\rho_{yx_1} - \rho_{x_1x_2} \rho_{yx_2})$$

1032

$$b_2 = \frac{1}{1 - \rho_{x_1x_2}^2} (\rho_{yx_2} - \rho_{x_1x_2} \rho_{yx_1}).$$

1033 *Finalmente puntualizar que el problema de mínimos cuadrados se puede
 1034 resolver utilizando la descomposición LU. Para ello solo es necesario un
 1035 cambio de variable en la ecuación*

$$\mathbf{X}^T \mathbf{X} \mathbf{B} = \mathbf{X}^T \mathbf{Y}$$

1036 *para obtener un sistema de ecuaciones tipo* $\mathbf{A}\mathbf{x} = \mathbf{b}$, *donde ahora* $\mathbf{A} =$
 1037 $\mathbf{X}^T \mathbf{X}$, $\mathbf{x} = \mathbf{B}$, *y* $\mathbf{b} = \mathbf{X}^T \mathbf{Y}$.

1038 $\{(2) \text{ Mínimos cuadrados con restricciones}\}$

1039 *{NOTA: Multiplicadores de Lagrange} | Dada la función* $f(\mathbf{x}) = f(x_1, x_2, \dots, x_N)$
 1040 *que depende de N variables y p restricciones* $g_1(\mathbf{x}) = d_1, g_2(\mathbf{x}) =$
 1041 $d_2, \dots, g_p(\mathbf{x}) = d_p$ *entonces el teorema de Lagrange nos dice que para*
 1042 *minimizar la función* $f(\mathbf{x})$ *bajo esas p restricciones debemos resolver el*
 1043 *sistema de ecuaciones*

$$\frac{\partial}{\partial x_i} \left[f(\mathbf{x}) + \sum_{j=1}^p \lambda_j g_j(\mathbf{x}) \right] = 0 ; i = 1, 2, \dots, N$$

1044

$$g_j(\mathbf{x}) = d_j ; j = 1, 2, \dots, p$$

1045 *Vamos a usar los multiplicadores de Lagrange para resolver el problema
 1046 de mínimos cuadrados*

$$\mathbf{Y} = \mathbf{X} \mathbf{B},$$

1047 *pero incluyendo p restricciones de la forma*

$$\mathbf{G} \mathbf{B} = \mathbf{d}.$$

1048 *Queremos minimizar la función:*

$$L = (\mathbf{Y} - \mathbf{X} \mathbf{B})^T (\mathbf{Y} - \mathbf{X} \mathbf{B}) + \mathbf{d}^T (\mathbf{G} \mathbf{B} - \mathbf{d}).$$

1049 Aquí hemos introducido p incógnitas pero también tenemos p nuevas
 1050 ecuaciones $\mathbf{GB} = \mathbf{d}$. Derivando L e igualando a cero

$$\frac{\partial L}{\partial \mathbf{B}} = -2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{XB} + \mathbf{G}^T = 0,$$

1051 lo cual tiene la solución

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y} - \frac{1}{2} \mathbf{G}^T),$$

1052 que para $= 0$ se reduce a la expresión de los mínimos cuadrados sin re-
 1053 stricciones

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y}).$$

1054 Si sustituimos en la ecuación de las restricciones

$$\mathbf{G}(\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y} - \frac{1}{2} \mathbf{G}^T) = \mathbf{d},$$

1055 y resolvemos para $,$ obtenemos

$$\frac{1}{2} = [\mathbf{G}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{G}^T]^{-1} [\mathbf{G}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} - \mathbf{d}] .$$

1056 Finalmente, si sustituimos en la solución para la matriz de coeficientes
 1057 \mathbf{B} obtenemos

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \left(\mathbf{X}^T \mathbf{Y} - \mathbf{G}^T [\mathbf{G}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{G}^T]^{-1} [\mathbf{G}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} - \mathbf{d}] \right) .$$

1058 Un ejemplo sería ajustar una recta a un conjunto de observaciones pero
 1059 imponiendo que la recta pase por algún punto determinado del espacio
 1060 xy .

1061 $\{(3)\}$ Mínimos cuadrados pesados $\} \mid$ En ocasiones debido a las incer-
 1062 tidumbres asociadas a la medidas de las variables independientes es
 1063 conveniente asignarles un peso diferencial en el problema de mínimos
 1064 cuadrados. Supongamos que unas variables independientes son cono-
 1065 cidas con mayor precisión que otras. Entonces el modelo de mínimos
 1066 cuadrados pesados es:

$$SEC = (\mathbf{Y} - \mathbf{XB})^T \mathbf{W}_\epsilon (\mathbf{Y} - \mathbf{XB}),$$

1067 donde \mathbf{W}_ϵ es una matriz diagonal con los elementos σ_j^{-2} , el inverso de
 1068 la varianza de cada variable independiente. \mid Sin embargo, en general,
 1069 los errores en los datos están correlacionados, y \mathbf{W}_ϵ no es diagonal.
 1070 Una elección razonable para \mathbf{W}_ϵ es el inverso de la matriz de covarian-
 1071 zas. Si asumimos que \mathbf{Y} se compone de un valor medio más un error o
 1072 fluctuación respecto la media

$$\mathbf{Y} = \bar{\mathbf{Y}} + \mathbf{Y}',$$

1073 entonces la matriz de covarianzas es $\langle \mathbf{Y}' \mathbf{Y}'^T \rangle$ y

$$\mathbf{W}_\epsilon = \langle \mathbf{Y}' \mathbf{Y}'^T \rangle^{-1} .$$

1074 $\{Ejemplo\ (1):\}$ Emery and Thompson (sección 3.12.4). | En la tabla
 1075 adjunta se muestran 5 observaciones de la variable independiente (x_i) y
 1076 dependiente (y_i). Se pide ajustar una recta al conjunto de datos y cal-
 1077 cular las medidas de error (varianza s^2 y coeficiente de correlación al
 1078 cuadrado r^2) asociadas al ajuste lineal. % %insert table

x_i	y_i	$x_i y_i$	x_i^2
-2	0	0	4
-1	0	0	1
0	1	0	0
1	1	1	1
2	3	6	4

1080 Los coeficientes del modelo lineal se pueden calcular con las expresiones:

$$\hat{b}_1 = \frac{\left[N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i \right]}{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2} =$$

$$\frac{[(5)(7) - (0)(5)]}{[(5)(10) - 10^2]} = 0.7$$

$$b_0 = \bar{y} - b_1 \bar{x} = 5/5 - (0.7)(0) = 1$$

1082 En notación matricial obtenemos el mismo resultado:

$$\mathbf{Y} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$$

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}$$

$$\mathbf{X}^T \mathbf{Y} = \begin{pmatrix} 5 \\ 7 \end{pmatrix}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} 1/5 & 0 \\ 0 & 1/10 \end{pmatrix}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \begin{pmatrix} 1/5 & 0 \\ 0 & 1/10 \end{pmatrix} \begin{pmatrix} 5 \\ 7 \end{pmatrix} = \begin{pmatrix} 1 \\ 0.7 \end{pmatrix}.$$

1083 *Como ya vimos, para calcular la varianza de nuestro ajuste usamos*

$$s^2 = \frac{1}{N-2} \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \frac{1}{N-2} SEC,$$

1084 *donde SEC es la suma de los errores cuadrados y N – 2 resulta debido*

1085 *a que en la regresión lineal se requiere la estimación de dos parámetros.*

1086 *En notación matricial*

$$\begin{aligned} 1087 SEC &= \mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X} \mathbf{B} - \mathbf{B}^T \mathbf{X}^T \mathbf{Y} + \mathbf{B}^T \mathbf{X}^T \mathbf{X} \mathbf{B} = \mathbf{Y}^T \mathbf{Y} - 2\mathbf{B}^T \mathbf{X}^T \mathbf{Y} + \mathbf{B}^T \mathbf{X}^T \mathbf{X} \mathbf{B} = \\ &= \mathbf{Y}^T \mathbf{Y} - 2\mathbf{B}^T \mathbf{X}^T \mathbf{Y} + \mathbf{B}^T \mathbf{X}^T \mathbf{Y} = \mathbf{Y}^T \mathbf{Y} - \mathbf{B}^T \mathbf{X}^T \mathbf{Y}, \end{aligned}$$

1088 *donde hemos usado la identidad $\mathbf{X}^T \mathbf{X} \mathbf{B} = \mathbf{X}^T \mathbf{Y}$. Si sustituimos las ma-*

1089 *trices de nuestro ejemplo, obtenemos*

$$\begin{aligned} SEC &= (0 \ 0 \ 1 \ 1 \ 3) \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{pmatrix} - (1 \ 0.7) \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{pmatrix} = \\ &= 11 - (1 \ 0.7) \begin{pmatrix} 5 \\ 7 \end{pmatrix} = 11 - 9.9 = 1.1 \end{aligned}$$

1090 *SEC puede ser calculado directamente con la expresión:*

$$SEC = \sum_{i=1}^N (\hat{y}_i - y_i)^2 = (-0.4)^2 + (-0.3)^2 + (0)^2 + (0.7)^2 + (0.6)^2 = 1.1$$

1091 *Y entonces la desviación estandar de nuestro ajuste lineal es*

$$s = \sqrt{\frac{1}{N-2} SEC} = \sqrt{\frac{1}{5-2} 1.1} = \sqrt{1.1/3} \simeq 0.366.$$

1092 *El coeficiente de correlación se puede escribir como*

$$r^2 = \frac{SCR}{SCT} = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} = \frac{4.9}{6} \simeq 0.8167$$

1093 *{(4) Ajuste de curvas con mínimos cuadrados} | En general, podemos*

1094 *escribir nuestro modelo lineal como*

$$Y = b_0 + b_1 x + b_2 x^2 + \dots + b_k x^k + \epsilon.$$

1096 | | El procedimiento es el mismo que para el caso de la línea recta, pero
 1097 ahora la matriz \mathbf{X} tiene una columna mas. Es decir, para $k = 2$ y para
 1098 N observaciones independientes las ecuaciones independientes son

$$y_1 = b_0 + b_1 x_1 + b_2 x_1^2 + \epsilon_1$$

$$y_2 = b_0 + b_1 x_2 + b_2 x_2^2 + \epsilon_2$$

1100 ...

1101 ...

1102 ...

1103 ...

$$y_N = b_0 + b_1 x_N + b_2 x_N^2 + \epsilon_N$$

1104 y pueden resolverse matricialmente para \mathbf{B} como

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y},$$

1105 donde \mathbf{X} tiene una columna mas que el caso del ajuste de una recta,
 1106 i.e. $k + 1$ columnas.

1107 {Ejemplo (2):} Ajustes por mínimos cuadrados con restricciones. |
 1108 Supongamos que queremos ajustar dos polinomios $f(x)$ y $g(x)$ de or-
 1109 den $d-1$ a dos conjunto de datos continuos de M y N observaciones,
 1110 respectivamente, $x_1, x_2, \dots, x_M \leq a$ y $x_{M+1}, x_{M+2}, \dots, x_N > a$, tal que,
 1111 queremos minimizar:

$$\sum_{i=1}^M (f(x_i) - y_i)^2 + \sum_{i=M+1}^N (g(x_i) - y_i)^2,$$

1112 sujeto a las restricciones:

$$f(a) = g(a) \text{ y } f'(a) = g'(a).$$

1113 |

1114 Primero debemos de construir las matrices del problema lineal de cuadra-
 1115 dos mínimos:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & \dots & x_1^{d-1} & 0 & 0 & \dots & 0 \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ 1 & x_M & \dots & x_M^{d-1} & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & x_{M+1} & \dots & x_{M+1}^{d-1} \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & 0 & 1 & x_N & \dots & x_N^{d-1} \end{pmatrix}$$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_M \\ y_{M+1} \\ y_{M+2} \\ \dots \\ y_N \end{pmatrix},$$

$$\mathbf{G} = \begin{pmatrix} 1 & a & \dots & a^{d-1} & -1 & -a & \dots & -a^{d-1} \\ 0 & 1 & \dots & (d-1)a^{d-2} & 0 & -1 & \dots & -(d-1)a^{d-2} \end{pmatrix}, \mathbf{d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

1116 Segundo debemos de calcular los coeficientes del ajuste \mathbf{B} con la expresión para mínimos cuadrados con restricciones.
 1117

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \left(\mathbf{X}^T \mathbf{Y} - \mathbf{G}^T \left[\mathbf{G}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{G}^T \right]^{-1} [\mathbf{G}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} - \mathbf{d}] \right).$$

1118 | Apliquemos el caso particular del ajuste de dos rectas $f(x) = b_1 + b_2x$ y
 1119 $g(x) = b_3 + b_4x$: |

- 1120 • Minimizar: $\sum_{i=1}^M (f(x_i) - y_i)^2 + \sum_{i=M+1}^N (g(x_i) - y_i)^2$
 • Restricciones:

$$f(h) = g(h) \rightarrow b_1 + b_2h - b_3 - b_4h = 0$$

$$f'(h) = g'(h) \rightarrow b_1 - b_3 = 0,$$

1121 donde h es el valor donde ambas rectas interseccionan. Por ello exigimos que ambas rectas tomen el mismo valor y sus derivadas sean iguales
 1122 en $x = h$. Se resuelve el sistema de ecuaciones:
 1123

$$\mathbf{X} \mathbf{B} = \mathbf{Y}$$

1124
 1125 $(1)b_1 + (x_1)b_2 + (0)b_3 + (0)b_4 = y_1$
 1126 $(1)b_1 + (x_2)b_2 + (0)b_3 + (0)b_4 = y_2$
 1127 $(1)b_1 + (x_3)b_2 + (0)b_3 + (0)b_4 = y_3$
 1128 $(1)b_1 + (x_4)b_2 + (0)b_3 + (0)b_4 = y_4$
 1129 $(0)b_1 + (0)b_2 + (1)b_3 + (x_5)b_4 = y_5$
 1130 $(0)b_1 + (0)b_2 + (1)b_3 + (x_6)b_4 = y_6$
 1131 $(0)b_1 + (0)b_2 + (1)b_3 + (x_7)b_4 = y_7$
 $(0)b_1 + (0)b_2 + (1)b_3 + (x_8)b_4 = y_8,$

1132 | bajo el sistema de ecuaciones de restricciones:

$$\mathbf{G} \mathbf{B} = \mathbf{d}$$

1133

1134

$$(1)b_1 + (h)b_2 + (-1)b_3 + (-1)b_4 = 0$$

$$(1)b_1 + (0)b_2 + (-1)b_3 + (0)b_4 = 0$$

1135

En notación matricial:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ 1 & x_2 & 0 & 0 \\ 1 & x_3 & 0 & 0 \\ 1 & x_4 & 0 & 0 \\ 0 & 0 & 1 & x_5 \\ 0 & 0 & 1 & x_6 \\ 0 & 0 & 1 & x_7 \\ 0 & 0 & 1 & x_8 \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix}, \mathbf{G} = \begin{pmatrix} 1 & h & -1 & -h \\ 0 & 1 & 0 & -1 \end{pmatrix}, \mathbf{d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

1136

1137

1138

$\{(B)\}$ Mínimos cuadrados no-lineales} | Los mínimos cuadrados no-lineales se aplican cuando queremos ajustar un conjunto de observaciones a un modelo que no es lineal en cuanto a los coeficientes.

1139

1140

1141

1142

$\{(1)\}$ Linealización del problema no lineal} En ocasiones se puede linealizar el problema y resolverlo utilizando la solución de mínimos cuadrados lineales. Por ejemplo, supongamos que queremos ajustar un conjunto de N observaciones a una función no-lineal exponencial:

$$\hat{y} = ae^{bx}.$$

1143

Este ejemplo se puede linealizar:

$$\hat{y} = ae^{bx} \rightarrow \ln(\hat{y}) = \ln(a) + bx,$$

1144

y resolver el problema lineal:

$$Y = B_1 + B_2 X,$$

1145

donde hemos hecho el cambio de variables:

$$Y = \ln(\hat{y}), B_1 = \ln(a), B_2 = b \text{ y } X = x.$$

1146

Resolvemos para B_1 y B_2 :

$$B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = (X^T X)^{-1} X^T Y,$$

1147

y calculamos los coeficientes originales a y b con el cambio de variables.

1148

1149

1150

1151

$\{(2)\}$ Resolución numérica del problema no lineal} | Cuando no es posible linealizar la función no-lineal que queremos ajustar, debemos minimizar la suma de los errores cuadráticos medios y resolver numéricamente. La SEC del modelo exponencial

$$SEC = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N \underbrace{(y_i - ae^{bx_i})}_{\epsilon_i}^2$$

1152 es minimizada derivando con respecto los coeficientes:

$$\frac{\partial SEC}{\partial a} = 2 \sum_{i=1}^N (y_i - ae^{bx_i}) \frac{\partial \epsilon_i}{\partial a} = 0$$

1153

$$\frac{\partial SEC}{\partial b} = 2 \sum_{i=1}^N (y_i - ae^{bx_i}) \frac{\partial \epsilon_i}{\partial b} = 0.$$

1154

1155 Debemos resolver el sistema de ecuaciones de arriba, lo cual se puede
 1156 hacer con descomposiciones algebráicas tipo **LU**, minimizando *SEC*
 1157 con la función de Matlab fminsearch.m (o implementando un método
 1158 numérico manual), o directamente utilizando la función de mínimos
 1159 cuadrados no-lineales de Matlab nlmfit.m. | Otro ejemplo no-lineal es
 ajustar la siguiente función trigonométrica:

$$\hat{y} = \phi_1 e^{\phi_2 x} \cos(\phi_3 x + \phi_4),$$

1160

donde la suma de los errores cuadráticos es:

$$SEC = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N \underbrace{(y_i - \phi_1 e^{\phi_2 x_i} \cos(\phi_3 x_i + \phi_4))^2}_{\epsilon_i},$$

1161

y sus derivadas respecto los coeficientes ϕ_1 , ϕ_2 , ϕ_3 , y ϕ_4 son:

$$\frac{\partial SEC}{\partial \phi_1} = 2 \sum_{i=1}^N (y_i - \phi_1 e^{\phi_2 x_i} \cos(\phi_3 x_i + \phi_4)) \frac{\partial \epsilon_i}{\partial \phi_1} = 0$$

1162

$$\frac{\partial SEC}{\partial \phi_2} = 2 \sum_{i=1}^N (y_i - \phi_1 e^{\phi_2 x_i} \cos(\phi_3 x_i + \phi_4)) \frac{\partial \epsilon_i}{\partial \phi_2} = 0$$

1163

$$\frac{\partial SEC}{\partial \phi_3} = 2 \sum_{i=1}^N (y_i - \phi_1 e^{\phi_2 x_i} \cos(\phi_3 x_i + \phi_4)) \frac{\partial \epsilon_i}{\partial \phi_3} = 0$$

1164

$$\frac{\partial SEC}{\partial \phi_4} = 2 \sum_{i=1}^N (y_i - \phi_1 e^{\phi_2 x_i} \cos(\phi_3 x_i + \phi_4)) \frac{\partial \epsilon_i}{\partial \phi_4} = 0$$

1165

1166 {Relación entre regresión y correlación} | El coeficiente de correlación,
 1167 r, nos informa de que tanto dos (o mas) variables covarian en el espacio-
 tiempo. Para dos variables aleatorias x e y, el coeficiente de correlación
 1168 es

$$r = \rho_{xy} = \frac{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})} \sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})}} = \frac{C_{xy}}{s_x s_y},$$

1169

1170 donde C_{xy} es la covarianza de x e y, y s_x y s_y son las correspondientes
 1171 desviaciones estandar. | | {Propiedades del coeficiente de correlación}
 | (1) r es adimensional. | (2) la magnitud de r se encuantra acotada

entre -1 y 1 , ya que es una normalización de la covarianza por el producto de la desviación estandar de las dos variables aleatorias. | | Si $r = \pm 1$ entonces el ajuste es perfecto. Para $r = 0$ los puntos estan dispersos aleatoriamente y no existe relación alguna entre las variables. Normalmente encontramos el estadístico r^2 en lugar de r . r^2 se puede reescribir como

$$r^2 = SCR/SCT = \frac{SCT - SEC}{SCT} = 1 - \frac{SEC}{SCT} = \frac{C_{xy}^2}{(s_x s_y)^2},$$

lo que nos informa del porcentaje de la varianza explicada (r^2 = varianza explicada/varianza total) como vimos anteriormente. Un valor de $r = 0.75$ significa que la regresión lineal de y sobre x , es decir \hat{y} , explica $100 * r^2 = 56.25\%$ de la varianza total de la muestra. | Finalmente puntualizar que también podemos calcular el coeficiente de correlación utilizando los estimados de los coeficientes de regresión. Para el caso de una línea recta sabemos

$$\hat{b}_1 = \frac{C_{xy}}{s_x^2}.$$

Si sustituimos en la definición de r obtenemos

$$r = \hat{b}_1 \frac{s_x^2}{s_x s_y} = \hat{b}_1 \frac{s_x}{s_y}.$$

{Coeficiente de correlación ajustado} | Con el fin de considerar los grados de libertad del modelo de regresión lineal, el coeficiente de correlación debe ser ajustado en función del número de variables independientes k . Para ello hay que considerar la verdadera varianza de los errores

$$Var(SEC) = \frac{SEC}{N - k - 1},$$

y de la variable dependiente

$$Var(SCT) = \frac{SCT}{N - 1}.$$

Puesto que los grados de libertad no son iguales ($N - 1$ vs $N - k - 1$), el coeficiente de correlación ajustado al cuadrado (\tilde{r}^2) se define como

$$\begin{aligned} \tilde{r}^2 &= 1 - \frac{Var(SEC)}{Var(SCT)} = 1 - \frac{SEC/N - k - 1}{SCT/N - 1} = 1 - \frac{N - 1}{N - k - 1}(1 - r^2) = \\ &= 1 - (1 - r^2) \frac{N - 1}{N - k - 1}, \end{aligned}$$

donde para $k = 0$ obtenemos la definición clásica del coeficiente de correlación. | | {Intervalo de confianza para el coeficiente de correlación} | Podemos calcular intervalos de confianza para el coeficiente de correlación r por medio de la denominada transformación Z de Fisher. Básicamente transforma r en una variable normal estandar Z

$$Z = \frac{1}{2} \frac{\ln(1 + r)}{\ln(1 - r)},$$

1199 con desviación estandar

$$\sigma(Z) = \frac{1}{\sqrt{(N-3)}},$$

1200 y media

$$\mu(Z) = \frac{1}{2} \frac{\ln(1+\rho_0)}{\ln(1-\rho_0)},$$

1201 donde $\mu(Z)$ es la media esperada (media poblacional) del estadístico Z .

1202 %El intervalo de confianza se escribe entonces como %

$$Z - Z_{\alpha/2} < Z < Z + Z_{\alpha/2}$$

1203 **Ejemplo:** | Supongamos $N = 21$ y $r = 0.8$. Encuentra el intervalo de
1204 confianza al 95% para el coeficiente de correlación poblacional ρ_0 .

$$Z = \frac{1}{2} \frac{\ln(1+0.8)}{\ln(1-0.8)} = 1.0986$$

1205 Puesto que Z esta Normalmente distribuida, entonces todos los valores
1206 deben de caer dentro de 1.96 desviaciones estandar de Z . Entonces, al
1207 95%, la verdadera media μ_Z esta contenida en

$$Z - 1.96\sigma(Z) < \mu(Z) < Z + 1.96\sigma(Z)$$

$$Z - 1.96 \frac{1}{\sqrt{(21-3)}} < \mu(Z) < Z + 1.96 \frac{1}{\sqrt{(21-3)}}$$

$$0.6366 < \mu(Z) < 1.5606$$

1210 Los límites encontrados para $\mu(Z)$ los podemos transformar en términos
1211 de la verdadera correlación

$$\mu(Z) = 0.6366 = \frac{1}{2} \left\{ \frac{1+\rho}{1-\rho} \right\} \rightarrow \rho = 0.56,$$

1212 donde hemos usado

$$e^{2\mu} = \frac{1+\rho}{1-\rho}$$

$$e^{2\mu} - 1 = \rho(1 + e^{2\mu})$$

$$\rho = \frac{(e^{2\mu(Z)} - 1)}{(e^{2\mu(Z)} + 1)}.$$

1215 Podemos afirmar con un 95% de confianza que la verdadera correlación
1216 ρ esta en el intervalo $0.56 < \rho < 0.92$, dado un tamaño muestral $N = 21$
1217 y una correlación muestral $r = 0.8$.

1218 {Verdaderos grados de libertad} | Ya hemos definido anteriormente los
1219 grados de libertad se define como el número de muestras independientes

1220 *N* menos el número de parámetros que se quieren estimar. Esta definición
 1221 es un tanto incorrecta puesto que debemos de también asegurar
 1222 que las *N* muestras son efectivamente independientes, es decir, no es-
 1223 tan autocorreladas en el espacio-tiempo. Para considerar esto *N* debe
 1224 reescribirse como

$$N^* = \frac{N}{\left[\sum_{\tau=-\infty}^{\infty} C_{xx}(\tau)C_{yy}(\tau) + C_{xy}(\tau)C_{yx}(\tau) \right] / [C_{xx}(0)C_{yy}(0)]} = \\ \frac{N}{\left[\sum_{\tau=-\infty}^{\infty} \rho_{xx}(\tau)\rho_{yy}(\tau) + \rho_{xy}(\tau)\rho_{yx}(\tau) \right]}.$$

1225 | | En general, series de datos suelen estar correlacionados en el espacio-
 1226 tiempo y $N^* << N$. Cuanto mayores las escalas de correlación espaciales-
 1227 temporales, menores los N^* . Esto nos hace pensar que es muy impor-
 1228 tante la selección de las escalas espaciales-temporales sobre las que quer-
 1229 emos calcular un estadístico. Para extraer las escalas de interés podemos
 1230 usar métodos espectrales y filtros. El proceso de filtrado se encarga de
 1231 eliminar aquellas escalas que esperamos no contribuyen a la verdadera
 1232 correlación pero pueden adicionar correlación artificial debido a errores
 1233 instrumentales y de muestreo.

1235 5 Propagación de errores

1236 **Regla 1** | Si x e y tienen errores aleatorios independientes δx y δy , entonces
 1237 el error en la suma $z = x + y$ es

$$\delta z = \sqrt{\delta x^2 + \delta y^2}.$$

1238 **Regla 2** | Si x e y tienen errores aleatorios independientes δx y δy ,
 1239 entonces el error en la multiplicación $z = xy$ es

$$\frac{\delta z}{z} = \sqrt{\left(\frac{\delta x^2}{x} \right) + \left(\frac{\delta y^2}{y} \right)}.$$

1240 **Regla 3** |

1241 Si $z = f(x)$, donde $f()$ es una función dada, entonces

$$\delta z = |f'(x)|\delta x.$$

1242 Formula general para propagación del error

1243 Sea x_1, x_2, \dots, x_N medidas con incertidumbres $\delta x_1, \delta x_2, \dots, \delta x_3$. Supong-
 1244 amos que queremos determinar q , el cual es una función de x_1, x_2, \dots, x_N :

$$q = f(x_1, x_2, \dots, x_N).$$

1245 *El error asociado a q es entonces*

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x_1} \delta x_1\right)^2 + \dots + \left(\frac{\partial q}{\partial x_N} \delta x_N\right)^2}$$

1246 | | *Si $q = x_1 + x_2$ entonces obtenemos la regla 1:*

$$\frac{\partial q}{\partial x_1} = 1,$$

$$\frac{\partial q}{\partial x_2} = 1,$$

$$\delta q = \sqrt{\delta x_1^2 + \delta x_2^2}.$$

1249 | | *Si $q = x_1 x_2$ entonces obtenemos la regla 2:*

$$\frac{\partial q}{\partial x_1} = x_2,$$

$$\frac{\partial q}{\partial x_2} = x_1,$$

$$\delta q = \sqrt{x_2^2 \delta x_1^2 + x_1^2 \delta x_2^2} = \sqrt{q^2 \left[\left(\frac{\partial x_1}{x_1}\right)^2 + \left(\frac{\partial x_2}{x_2}\right)^2 \right]}.$$

$$\frac{\delta q}{q} = \sqrt{\left(\frac{\delta x_1^2}{x_1}\right) + \left(\frac{\delta x_2^2}{x_2}\right)}.$$

1253 **Demstración:** | *Queremos calcular la desviación estándar de la función*

$$q = q(x_1, x_2, \dots, x_N),$$

1255 *que depende de N variables independientes x_1, x_2, \dots, x_N . El desarrollo
1256 de Taylor de la función q alrededor de la media \bar{q} se puede escribir:*

$$q - \bar{q} = (x_1 - \bar{x}_1) \frac{\partial q}{\partial x_1} + (x_2 - \bar{x}_2) \frac{\partial q}{\partial x_2} + \dots + (x_N - \bar{x}_N) \frac{\partial q}{\partial x_N}.$$

1257 *La varianza de la función q es:*

$$\begin{aligned} s_q^2 &= \frac{1}{N-1} \sum_{i=1}^N (q_i - \bar{q})^2 = \frac{1}{N-1} \left[(x_1 - \bar{x}_1) \frac{\partial q}{\partial x_1} + (x_2 - \bar{x}_2) \frac{\partial q}{\partial x_2} + \dots + (x_N - \bar{x}_N) \frac{\partial q}{\partial x_N} \right]^2 = \\ &= \frac{1}{N-1} \left[(x_1 - \bar{x}_1)^2 \left(\frac{\partial q}{\partial x_1}\right)^2 + (x_2 - \bar{x}_2)^2 \left(\frac{\partial q}{\partial x_2}\right)^2 + 2(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) \frac{\partial q}{\partial x_1} \frac{\partial q}{\partial x_2} + \dots \right] = \\ &= s_{x_1}^2 \left(\frac{\partial q}{\partial x_1}\right)^2 + s_{x_2}^2 \left(\frac{\partial q}{\partial x_2}\right)^2 + 2s_{x_1 x_2} \frac{\partial q}{\partial x_1} \frac{\partial q}{\partial x_2} + \dots \end{aligned}$$

1260 *y finalmente la expresión general para la propagación del error es:*

$$s_q = \sqrt{s_{x_1}^2 \left(\frac{\partial q}{\partial x_1} \right)^2 + s_{x_2}^2 \left(\frac{\partial q}{\partial x_2} \right)^2 + 2s_{x_1 x_2} \frac{\partial q}{\partial x_1} \frac{\partial q}{\partial x_2} + \dots}$$

1261 *{Ejemplo:} | La ecuación para el cálculo de la salinidad a partir de la*
 1262 *conductividad (C) y temperatura (T) es (Unesco EOS-80)*

$$S = a_0 + a_1 R_T^{1/2} + a_2 R_T + a_3 R_T^{3/2} + a_4 R_T^2 + a_5 R_T^{5/2} + \Delta S,$$

1263 *donde*

$$R_T = \frac{R}{R_p r_t}, \quad R = \frac{C(S, T, 0)}{C(35, 15, 0)},$$

1264 *$C(35, 15, 0)$ es la conductividad de un agua de salinidad práctica 35 a*
 1265 *los 15 °C,*

$$r_t = c_0 + c_1 T + c_2 T^2 + c_3 T^3 + c_4 T^4,$$

$$R_p = 1 + \frac{P(e_1 + e_2 P + e_3 P^2)}{(1 + d_1 T + d_2 T^2 + (d_3 + d_4 T)R)},$$

1266 *y*

$$\Delta S = \frac{T - 15}{1 + k(T - 15)} (b_0 + b_1 R_T^{1/2} + b_2 R_T + b_3 R_T^{3/2} + b_4 R_T^2 + b_5 R_T^{5/2}),$$

1268 *con los coeficientes a_i, b_i, c_i, d_i , y e_i | $a_0 = 0.0080$ $b_0 = 0.0005$ $c_0 =$*
 1269 *0.6766097 $d_1 = 3.426 e^{-2}$ $e_1 = 2.070 e^{-5}$ | $a_1 = -0.1692$ $b_1 =$*
 1270 *-0.0056 $c_1 = 2.00564 e^{-2}$ $d_2 = 4.464 e^{-4}$ $e_2 = -6.370 e^{-10}$*
 1271 *| $a_2 = 25.3851$ $b = -0.0066$ $c_2 = 1.104259 e^{-4}$ $d_3 =$*
 1272 *4.215 e^{-1}* $e_3 = 3.989 e^{-15}$ | $a_3 = 14.0941$ $b = -0.0375$ $c_3 =$
 1273 *-6.9698 e^{-7}* $d_4 = -3.107 e^{-3}$ | $a_4 = -7.0261$ $b_4 = 0.0636$ $c_4 =$
 1274 *1.0031 e^{-9}* | $a_5 = 2.7081$ $b_5 = -0.0144$ | $\sum a_i =$
 1275 *35.0000* $\sum b_i = 0.0000$ | $k = 0.0162$ | | *Si el error de precisión*
 1276 *del termistor y de la celda de conductividad del CTD es $\delta T = 0.001^\circ\text{C}$ y*
 1277 *$\delta C = 0.001 \text{ Sm}^{-1}$, respectivamente, calcule la incertidumbre asociada al*
 1278 *cálculo de la salinidad S a partir de la formula general de propagación*
 1279 *del error para $C = 5 \text{ Sm}^{-1}$, $T = 28^\circ\text{C}$, y $P = 50 \text{ dbar}$. Suponga*
 1280 *que $\delta P = 0$, es decir, el altímetro no tiene errores de precisión.* | |

1281 *{Resultado:} | | El error estándar asociado a la salinidad es*

$$\delta S = \sqrt{\left(\frac{\partial S}{\partial T} \delta T \right)^2 + \left(\frac{\partial S}{\partial C} \delta C \right)^2},$$

1282 *donde*

$$\frac{\delta S}{\delta T} = [(1 + k(T - 15))^{-1} - (k(T - 15))(1 + k(T - 15))^{-2}]$$

$$(b_0 + b_1 R_T^{1/2} + b_2 R_T + b_3 R_T^{3/2} + b_4 R_T^2 + b_5 R_T^{5/2}),$$

1284 y

$$\frac{\delta S}{\delta C} = \frac{1}{2} a_1 \frac{R_T^{-1/2}}{C(35, T, 0)} + a_2 \frac{1}{C(35, T, 0)} + \frac{3}{2} a_3 \frac{R_T^{1/2}}{C(35, T, 0)} +$$

1285 $+ 2a_4 \frac{R_T}{C(35, T, 0)} + \frac{5}{2} a_5 \frac{R_T^{3/2}}{C(35, T, 0)} + \frac{T - 15}{(1 + k(T - 15))}$

1286 $\left(b_1 \frac{R_T^{-1/2}}{C(35, T, 0)} + b_2 \frac{1}{C(35, T, 0)} + \frac{3}{2} b_3 \frac{R_T^{1/2}}{C(35, T, 0)} + 2b_4 \frac{R_T}{C(35, T, 0)} + \frac{5}{2} b_5 \frac{R_T^{3/2}}{C(35, T, 0)} \right).$

1287 | | Por lo tanto, dada una medida de conductividad C y temperatura
 1288 T podemos calcular $\delta S/\delta T$ y $\delta S/\delta C$ y obtener la desviación estándar
 1289 asociado al cálculo de la salinidad δS . | | {Ejercicio:} | | Calcule el
 1290 error estándar asociado a la densidad referenciada a la superficie que se
 1291 obtiene al usar el algoritmo del estado del agua de mar (Unesco EOS-
 1292 80)

$$\rho(S, T, 0) = \rho_w + (b_0 + b_1 T + b_2 T^2 + b_3 T^3 + b_4 T^4) S +$$

1293 $+ (c_0 + c_1 T + c_2 T^2) S^{3/2} + d_0 S^2,$

1294 con los coeficientes b_i , c_i , y d_0 | $b_0 = 8.24493 e^{-1}$ $c_0 = -5.72466 e^{-3}$
 1295 | $b_1 = -4.0899 e^{-3}$ $c_1 = 1.0227 e^{-4}$ | $b_2 = 7.6438 e^{-5}$ $c_2 =$
 1296 $-1.6546 e^{-6}$ | $b_3 = -8.2467 e^{-7}$ | $b_4 = 5.3875 e^{-9}$ $d_0 =$
 1297 $4.8314 e^{-4}$, | | y la densidad de referencia para agua pura (ρ_w) definida
 1298 como:

$$\rho_w = a_0 + a_1 T + a_2 T^2 + a_3 T^3 + a_4 T^4 + a_5 T^5,$$

1299 donde los coeficientes a_i son: | $a_0 = 999.842594$ | $a_1 = 6.793952 e^{-2}$ |
 1300 $a_2 = -9.095290 e^{-3}$ | $a_3 = 1.001685 e^{-4}$ | $a_4 = -1.120083 e^{-6}$ | $a_5 =$
 1301 $6.536332 e^{-9}$.

1302 6 Métodos de Interpolación}

1303 Interpolación es el procedimiento para el cual obtenemos valores de
 1304 propiedades en posiciones o tiempos que nunca fueron muestreados a
 1305 partir de observaciones existentes en otras localizaciones o tiempos. En
 1306 oceanografía necesitamos interpolar (i) para llenar huecos cuando el
 1307 instrumento dejó de medir, (ii) para obtener mapas espaciales de con-
 1308 tornos (2D), (iii) para calcular alguna propiedad derivada en un punto
 1309 concreto

1310 6.1 Interpolación Lineal

1311 La interpolación más sencilla, pero no por ello peor, es la interpolación
 1312 lineal. Esta interpolación se basa en ajustar una línea recta entre los
 1313 puntos conocidos, e interpolar cualquier punto intermedio como un
 1314 punto a lo largo de la recta. Este tipo de interpolación puede ser usado
 1315 para llenar huecos en nuestras series temporales. La interpolación
 1316 lineal de unas serie $y(t)$ se puede escribir como

$$y(t_i) = y(t_0) + \frac{y(t_1) - y(t_0)}{t_1 - t_0} (t_i - t_0).$$

1317 **6.2 Interpolación polinómica**

1318 *En el caso que queramos interpolar entre mas de dos puntos simultaneamente,*
 1319 *debemos de usar polinomios de orden superior a la recta (orden*
 1320 *1). Es decir,*

$$y(x) = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1} + a_nx_n.$$

1321 *Si tomamos N observaciones ($y_i(x)$; $i = 0, 2, \dots, N - 1$) obtenemos un*
 1322 *sistema de N ecuaciones*

$$\begin{pmatrix} 1 & x_0 & \dots & x_0^{N-1} & x_0^N \\ 1 & x_1 & \dots & x_1^{N-1} & x_1^N \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_N & \dots & x_N^{N-1} & x_N^N \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_N \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{pmatrix},$$

1323 *lo cual se puede resolver con eliminación Gauss-Jordan. Este método es*
 1324 *muy lento y por ello se usa el método de Lagrange.*

1325 **{Método de Lagrange}**

1326 *De nuevo asumimos*

$$p(x) = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1} + a_nx_n = \sum_{k=0}^N a_kx^k = \sum_{i=1}^{N+1} y_iL_i(x),$$

1327 *donde*

$$L_i(x) = \prod_{\substack{k=1 \\ k \neq i}}^{N+1} \frac{x - x_k}{x_i - x_k},$$

1328 *son los denominados polinomios de Lagrange, y \prod es el operador pro-*
 1329 *ducto. Puesto que este operador cuando $k \neq i$ no incluye el producto, a*
 1330 *pesar que varíe de 1 a $N + 1$, obtendremos un polinomio de orden N .*

1331 *Esta suma de polinomios de Lagrange es el polinomio de menor grado*
 1332 *que interpola un conjunto de datos, es decir,*

$$p(x_j) = \sum_{i=1}^{N+1} y_iL_i(x_j) = y_j.$$

1333 **{Demostración:}**

1334 *Los polinomios de Lagrange para $i \neq j$ son iguales a cero, y para $i = j$*
 1335 *son iguales a 1. Veamos esto:*

$$(1) \text{ Para } i \neq j : L_i(x_j) = \prod_{\substack{k=1 \\ k \neq i}}^{N+1} \frac{x_j - x_k}{x_i - x_k} = \\ = \frac{x_j - x_1}{x_i - x_1} \frac{x_j - x_2}{x_i - x_2} \dots \frac{x_j - x_j}{x_i - x_j} \dots \frac{x_j - x_{N+1}}{x_i - x_{N+1}}$$

$$(2) \text{ Para } i = j : L_i(x_j) = \prod_{\substack{k=1 \\ k \neq i}}^{N+1} \frac{x_j - x_k}{x_j - x_k} = 1$$

1337 *De modo que*

$$L_i(x_j) = \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}.$$

1338 *Finalmente podemos concluir entonces que*

$$p(x_j) = \sum_{i=1}^{N+1} y_i L_i(x_j) = \sum_{i=1}^{N+1} y_i \delta_{ij} = y_j.$$

1339 *es un polinomio de grado no mayor a N y que $p(x_j) = y_j$.*

1340 *El polinomio se puede reescribir en términos de la función Q_i como*

$$p(x) = \sum_{i=1}^{N+1} y_i [Q_i(x)/Q_i(x_i)],$$

1341 *donde*

$$Q_i(x) = (x - x_1)(x - x_2) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_{N-1}),$$

1342 *es el producto de todas las diferencias excepto la posición i (i.e., $x - x_i$).*

1343 *Si expandemos $p(x)$*

$$\begin{aligned} p(x) = y_1 \frac{(x - x_2)(x - x_3) \dots (x - x_{N+1})}{(x_1 - x_2)(x_1 - x_3) \dots (x_1 - x_{N+1})} + y_2 \frac{(x - x_1)(x - x_3) \dots (x - x_{N+1})}{(x_2 - x_1)(x_2 - x_3) \dots (x_2 - x_{N+1})} + \\ + \dots + y_{N+1} \frac{(x - x_1)(x - x_2) \dots (x - x_N)}{(x_{N+1} - x_1)(x_{N+1} - x_2) \dots (x_{N+1} - x_N)}. \end{aligned}$$

1345 *{Ejemplo:}*

1346 *Considere los puntos (0,2), (1,2), (2,0) y (3,0) para los cuales queremos*
 1347 *ajustar un polinomio de orden 3*

$$\begin{aligned} y(x) = 2 \frac{(x - 1)(x - 2)(x - 3)}{(0 - 1)(0 - 2)(0 - 3)} + 2 \frac{(x - 0)(x - 2)(x - 3)}{(1 - 0)(1 - 2)(1 - 3)} + 0 + 0 = \\ = \frac{2}{3}x^3 - 3x^2 + \frac{7}{3}x + 2. \end{aligned}$$

1349 6.3 Spline cúbico

1350 *La interpolación por spline cúbicos es un método de ajuste de polí-*
 1351 *nomios de orden 3 por segmentos. Veamos este método con un ejemplo.*

1352 *Supongamos que queremos interpolar los siguientes datos (x_i, y_i)* =
 1353 *$[(2, -1), (3, 2), (5, -7)]$ con una spline cúbica. Primero definimos un*
 1354 *polinomio cúbico para cada intervalo:*

$$s(x) = a_1 x^3 + b_1 x^2 + c_1 x + d_1 \quad \text{si } x \in [2, 3]$$

1355

$$s(x) = a_2x^3 + b_2x^2 + c_2x + d_2 \quad \text{si } x \in [3, 5]$$

1356

A continuación debemos de asegurar que los polinomios pasan por los puntos del problema:

1358

$$s(2) = 8a_1 + 4b_1 + 2c_1 + d_1 = -1$$

1359

$$s(3) = 27a_1 + 9b_1 + 3c_1 + d_1 = 2$$

1360

$$s(3) = 27a_2 + 9b_2 + 3c_2 + d_2 = 2$$

1361

$$s(5) = 125a_2 + 25b_2 + 5c_2 + d_2 = -7$$

1362

Ahora calculamos la primera y segundas derivadas para cada intervalo

1363

$$s'(x) = 3a_1x^2 + 2b_1x + c_1 \quad \text{si } x \in [2, 3]$$

1364

$$s'(x) = 3a_2x^2 + 2b_2x + c_2 \quad \text{si } x \in [3, 5]$$

1365

$$s''(x) = 6a_1x + 2b_1 \quad \text{si } x \in [2, 3]$$

1366

$$s''(x) = 6a_2x + 2b_2 \quad \text{si } x \in [3, 5]$$

y aseguramos que sean continuas. Para ello debemos de igualar las derivadas entre intervalos de manera que no hayan discontinuidades

$$3a_1(3)^2 + 2b_1(3) + c_1 = 3a_2(3)^2 + 2b_2(3) + c_2 \rightarrow 27a_1 + 6b_1 + c_1 = 27a_2 + 6b_2 + c_2$$

1367

$$6a_1(3) + 2b_1 = 6a_2(3) + 2b_2 \rightarrow 18a_1 + 2b_1 = 18a_2 + 2b_2$$

1368

En este momento tenemos 6 ecuaciones y 8 incógnitas. Debemos por lo tanto encontrar dos ecuaciones mas. Para ello vamos a forzar que en los extremos la segunda derivada sea nula, es decir, no haya curvatura

1369

$$s''(x_0) = s''(2) = 0 \rightarrow 6a_1(2) + 2b_1 = 0 \rightarrow 12a_1 + 2b_1 = 0$$

1370

$$s''(x_N) = s''(5) = 0 \rightarrow 6a_2(5) + 2b_2 = 0 \rightarrow 30a_2 + 2b_2 = 0$$

1371

Ahora ya tenemos un sistema determinado, es decir, 8 ecuaciones y 8 incógnitas

1372

$$8a_1 + 4b_1 + 2c_1 + d_1 = -1$$

1373

$$27a_1 + 9b_1 + 3c_1 + d_1 = 2$$

1374

$$27a_2 + 9b_2 + 3c_2 + d_2 = 2$$

1375

$$125a_2 + 25b_2 + 5c_2 + d_2 = -7$$

1376

$$27a_1 + 6b_1 + c_1 = 27a_2 + 6b_2 + c_2$$

1377

$$18a_1 + 2b_1 = 18a_2 + 2b_2$$

1378

$$12a_1 + 2b_1 = 0$$

1379

$$30a_2 + 2b_2 = 0$$

1380

Lo cual en notación matricial se puede escribir

$$\begin{pmatrix} 8 & 4 & 2 & 1 & 0 & 0 & 0 & 0 \\ 27 & 9 & 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 27 & 9 & 3 & 1 \\ 0 & 0 & 0 & 0 & 125 & 25 & 5 & 1 \\ 27 & 6 & 1 & 0 & -27 & -6 & -1 & 0 \\ 18 & 2 & 0 & 0 & -18 & -2 & 0 & 0 \\ 12 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 30 & 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ b_1 \\ c_1 \\ d_1 \\ a_2 \\ b_2 \\ c_2 \\ d_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 2 \\ -7 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

1382 *Este sistema se puede resolver facilmente si la matriz de datos es invertible. Los coeficientes resultantes son*
 1383

$$a_1 = -1.25 \quad b_1 = 7.5 \quad c_1 = -10.75 \quad d_1 = 0.5$$

$$1384 \quad a_1 = 0.625 \quad b_1 = -9.375 \quad c_1 = 39.875 \quad d_1 = -50.125,$$

1385 *y los polinomios de nuestra spline cúbica es*

$$s(x) = -1.25x^3 + 7.5x^2 - 10.75x + 0.5 \quad \text{si } x \in [2, 3]$$

$$1386 \quad s(x) = 0.625x^3 + -9.375x^2 + 39.875x + 50.125 \quad \text{si } x \in [3, 5]$$

Figura. Comparación entre ajuste de un polinomio de orden 6 por método de cuadrados mínimos, interpolación de Lagrange, e interpolación spline cúbica.

7 Interpolación Objetiva

1388 *Un mapa objetivo se obtiene como una regresión múltiple (donde el error cuadrático medio es mínimo) de un conjunto de observaciones discretas. Se utiliza en oceanografía para obtener mapas continuos (mallas regulares) a partir de datos discretos distribuidos irregularmente en el espacio. Las variables representadas en el mapa objetivo pueden modificarse de una realización a otra, con lo que deben de considerarse las anomalías de las variables en lugar de las variables en si mismas. Se debe de definir un ensemble (promedio), climatología, o candidato y extraerlo a cada variable para obtener anomalías. Este proceso de elección de la media es una parte delicada de la interpolación objetiva. En general, para el océano, la media es desconocida ya que solamente tenemos pocas realizaciones de nuestro muestreo. Una forma de operar es estimar la media ajustando un polinomio de bajo orden a nuestros datos discretos, extraer esta a los datos y proceder con la interpolación objetiva. La media se añade de nuevo después de calcular el mapa objetivo de las fluctuaciones.* |

1404 *Típicamente se debe asumir dos condiciones:* |

- 1405 (1) *El error (o ruido) asociado a la interpolación no esta correlacionado con nuestra señal (variable a interpolar)*

$$<\phi_i \epsilon_i> = 0$$

1408 (2) *El error no esta correlacionado de una estación a la otra*

$$\langle \epsilon_i \epsilon_j \rangle = \begin{cases} < \epsilon^2 > & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

1409 *Supongamos entonces fluctuaciones respecto un estado climatológico o*
 1410 *media*

$$\phi'_i = \phi_i - \bar{\phi}; i = 1, 2, \dots, N$$

1411 *Ahora vamos a intentar aproximar el valor de ϕ' en un punto de una*
 1412 *malla, ϕ_g , en términos de una combinación lineal de los valores en esta-*
 1413 *ciones vecinas ϕ_i (señal). Entonces el problema de mínimos cuadrados*
 1414 *es*

$$\phi'_g = \sum_{i=1}^N b_i \phi'_i$$

1415 *donde ϕ'_g son anomalías en la malla regular y ϕ'_i son anomalías en las*
 1416 *estaciones. Los mejores coeficientes son aquellos que minimizan el error*
 1417 *cuadrático medio, es decir,*

$$\begin{aligned} SEC &= \sum_{i=1}^N \left(\phi'_g - \sum_{i=1}^N b_i \phi'_i \right)^2 = \sum_{i=1}^N \left(\phi'_g - \sum_{i=1}^N b_i \phi'_i \right) \left(\phi'_g - \sum_{i=1}^N b_i \phi'_i \right) = \\ &\quad \sum_{i=1}^N \left[\phi'_g \phi'_g - 2 \sum_{i=1}^N b_i \phi'_g \phi'_i + \sum_{i=1}^N \sum_{j=1}^N b_i b_j \phi'_i \phi'_j \right] = \\ &= \sum_{i=1}^N \phi'_g \phi'_g - 2 \sum_{i=1}^N b_i \sum_{i=1}^N \phi'_g \phi'_i + \sum_{i=1}^N \sum_{j=1}^N b_i b_j \sum_{i=1}^N \phi'_i \phi'_j. \end{aligned}$$

1420 *El error normalizado se puede escribir como*

$$\begin{aligned} \epsilon &= \frac{SEC}{\sum_{i=1}^N \phi'_g \phi'_g} = 1 - 2 \sum_{i=1}^N b_i \frac{\sum_{i=1}^N \phi'_g \phi'_i}{\sum_{i=1}^N \phi'_g \phi'_g} + \sum_{i=1}^N \sum_{j=1}^N b_i b_j \frac{\sum_{i=1}^N \phi'_i \phi'_j}{\sum_{i=1}^N \phi'_g \phi'_g} = \\ &= 1 - 2 \sum_{i=1}^N b_i r_{gi} + \sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{ij}, \end{aligned}$$

1421 *donde*

$$r_{gi} = \frac{\sum_{i=1}^N \phi'_g \phi'_i}{\sum_{i=1}^N \phi'_g \phi'_g} \quad \text{y} \quad r_{ij} = \frac{\sum_{i=1}^N \phi'_i \phi'_j}{\sum_{i=1}^N \phi'_g \phi'_g},$$

1423 son matrices de correlación entre el punto de malla y las estaciones, y
 1424 entre las estaciones, respectivamente.

1425 Si derivamos ϵ respecto los coeficientes, obtenemos la condición de mini-
 1426 b_i

$$\frac{\partial \epsilon}{\partial b_i} = -2r_{gi} + 2 \sum_{j=1}^N b_j r_{ij} = 0 \quad i = 1, 2, \dots, N.$$

$$r_{gi} = \sum_{j=1}^N b_j r_{ij},$$

1428 y los coeficientes en notación índice son

$$b_j = (r_{ij})^{-1} r_{gi}$$

1429 Finalmente, el valor de la medida en el punto de malla es

$$\phi'_g = r_{gi}(r_{ij})^{-1}\phi'_i$$

1430 o en notación matricial para un único punto de malla

$$\phi'_g = r_{gs}(r_{ss})^{-1}\phi',$$

1431 donde ϕ'_g es el valor de la anomalía en un punto de malla (matriz ele-
 1432 mento, 1×1), r_{gs} es un vector fila compuesto por las correlaciones entre
 1433 el punto de malla y las estaciones de medida, r_{ss} es una matriz de cor-
 1434 relaciones entre todas las estaciones, y ϕ' es un vector columna con las
 1435 anomalías en las estaciones.

1436 En el caso de que tengamos N puntos de malla y k estaciones de me-
 1437 dida, las ecuaciones básicas de la interpolación objetiva se pueden es-
 1438 cribir en notación matricial como

$$\begin{pmatrix} \phi'_{g_1} \\ \phi'_{g_2} \\ \vdots \\ \phi'_{g_N} \end{pmatrix} = \begin{pmatrix} r_{g_11} & r_{g_12} & \dots & r_{g_1k} \\ r_{g_21} & r_{g_22} & \dots & r_{g_2k} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ r_{g_N1} & r_{g_N2} & \dots & r_{g_Nk} \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1k} \\ r_{21} & r_{22} & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & r_{kk} \end{pmatrix}^{-1} \begin{pmatrix} \phi'_1 \\ \phi'_2 \\ \vdots \\ \phi'_k \end{pmatrix}.$$

$$(N \times 1) \qquad (N \times k) \qquad (k \times k) \qquad (k \times 1)$$

1439

|

1440 El error normalizado ϵ asociado al mapa interpolado es

$$\epsilon = 1 - 2 \sum_{j=1}^N b_j r_{gi} + \sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{ij} = 1 - 2 \sum_{j=1}^N b_j \sum_{i=1}^N b_i r_{ij} + \sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{ij} =$$

1442

$$= 1 - 2 \sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{ij} + \sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{ij} = 1 - \sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{ij} =$$

1443

$$= 1 - \sum_{j=1}^N b_j \sum_{j=1}^N b_j r_{ij} = 1 - \sum_{i=1}^N r_{gi} b_i,$$

1444

o en notación matricial para un único punto de malla

$$\epsilon = 1 - r_{gs}(r_{ss})^{-1} r_{gs}^T.$$

1445

En el caso de N puntos de malla y k estaciones

1446

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} = \text{Diag} \left[\mathbf{I} - \begin{pmatrix} r_{g_11} & r_{g_12} & \dots & r_{g_1k} \\ r_{g_21} & r_{g_22} & \dots & r_{g_2k} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ r_{g_N1} & r_{g_N2} & \dots & r_{g_Nk} \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1k} \\ r_{21} & r_{22} & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & r_{kk} \end{pmatrix}^{-1} \begin{pmatrix} r_{g_11} & r_{g_21} & \dots & r_{g_N1} \\ r_{g_12} & r_{g_22} & \dots & r_{g_N2} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ r_{g_1k} & r_{g_2k} & \dots & r_{g_Nk} \end{pmatrix} \right] (N \times 1) \quad (N \times k) \quad (k \times k) \quad (k \times N)$$

1447

*Este error es tal vez la característica mas importante de la interpolación objetiva. En general solo se muestra el mapa interpolado en las regiones donde ϵ es menor que un cierto valor. El error solo depende en las localizaciones de las estaciones y no en un valor particular de la medida. Es por ello que esta técnica puede ser usada para el diseño de experimentos. Es decir, nos sirve para saber cual debe ser la distribución óptima de las estaciones de medida para obtener el error mínimo en el mapa. *

1454

Nota: Si un punto de observación, $i = k$, coincide con un punto de malla, entonces $r_{gk} = r_{kk} = 1$, y esperamos que el método de regresión nos de $b_k = 1$ y todos los demás pesos sean igual a cero. En este caso el valor interpolado en el punto de malla es igual al valor medido en la estación $\$ g = r\{gk\} (r_{\{kk\}})^{-1} _k = 1(1)^{-1} _k = _k \$$. En este caso el error es cero, $\epsilon = 1 - r_{gk}(r_{kk})^{-1} r_{gk}^T = 1 - 1(1)^{-1} 1 = 0$, ya que hemos asumido que los datos son perfectos. Si los puntos de las estaciones están decorrelacionados con el punto de malla en cuestión (i.e., muy lejos de las estaciones), entonces $b_i = 0$ y $\epsilon = 1$, y recuperamos la media o climatología

1455

1456

1457

1458

1459

1460

1461

1462

1463

$$\phi'_g = \sum_{i=1}^N b_i \phi'_i = 0 \rightarrow \phi'_g = \phi_g - \bar{\phi} = 0 \quad \text{and} \quad \phi_g = \bar{\phi}$$

1464

{Error observacional:}

1465

1466

Vamos asumir ahora que las mediciones en las estaciones no son perfectas, es decir,

$$\phi'_i = E[\phi'_i] + \delta_i.$$

¹⁴⁶⁷ *Igual que anteriormente asumimos que el error de instrumentación δ_i*
¹⁴⁶⁸ *no esta correlacionado con la señal verdadera*

$$\langle E[\phi_i]\delta_i \rangle = 0,$$

¹⁴⁶⁹ *y que el error instrumental entre estaciones tampoco esta correlacionado*

$$\langle \epsilon_i \epsilon_j \rangle = \begin{cases} \langle \delta^2 \rangle & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

¹⁴⁷⁰ *En este caso obtenemos*

$$\epsilon = 1 - 2 \sum_{i=1}^N b_i r_{gi} + \sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{ij} + \eta \sum_{i=1}^N b_i^2,$$

¹⁴⁷¹ *donde η es el cociente entre la varianza del error (ruido) y la varianza*
¹⁴⁷² *de las medidas, es decir el cociente ruido-señal*

$$\eta = \frac{\langle \delta^2 \rangle}{\langle \phi_g' \phi_g' \rangle}.$$

¹⁴⁷³ *La minimización del error nos da la condición*

$$r_{gi} = \sum_{j=1}^N b_j r_{ij} + \eta b_i,$$

¹⁴⁷⁴ *y los coeficientes en notación índice son*

$$b_j = (r_{ij} + \eta I_{ij})^{-1} r_{gi},$$

¹⁴⁷⁵ *donde I_{ij} es la matriz identidad. De esta expresión se deduce que cuando*
¹⁴⁷⁶ *η es grande (mucho ruido en la medida) entonces los coeficientes son*
¹⁴⁷⁷ *mas pequeños respecto al caso de observaciones perfectas. Consecuentemente*
¹⁴⁷⁸ *nuestro mapa interpolado se acerca mas a la climatología ya que*
¹⁴⁷⁹ *la anomalía es menor para coeficientes mas pequeños*

$$\downarrow \sum_{i=1}^N b_i \phi_i' \rightarrow \downarrow \phi_g' \quad y \quad \phi_g \rightarrow \bar{\phi}.$$

¹⁴⁸⁰ *Incluyendo errores observacionales, la interpolación objetiva tenderá a*
¹⁴⁸¹ *la climatología o media y las nuevas observaciones serán incluidas pero*
¹⁴⁸² *con menos pesos. También podemos añadir diferentes errores ruido-*
¹⁴⁸³ *señal en la diagonal principal para darle menos peso a las estaciones de*
¹⁴⁸⁴ *medida que tienen mas incertidumbre asociada. Es conveniente entonces*
¹⁴⁸⁵ *añadir errores observacionales al esquema de interpolación objetiva.*
¹⁴⁸⁶ *Otra razón para ello es el caso en que existan dos estaciones que coin-*
¹⁴⁸⁷ *ciden exactamente con un punto de malla. En este caso, si no hemos*
¹⁴⁸⁸ *añadido error observacional la matriz de correlaciones entre las esta-*
¹⁴⁸⁹ *ciones se convierte singular y el esquema de interpolación objetiva no se*
¹⁴⁹⁰ *puede resolver. Por ejemplo para la interpolación en un único punto de*

1491 *malla a partir de dos estaciones de medida, la matriz de correlaciones*
 1492 *sería singular*

$$r_{ss} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

1493 *La anomalía interpolada en notación matricial en un punto de malla es*
 1494 *ahora*

$$\phi'_g = r_{gs}(r_{ss} + \eta I)^{-1}\phi',$$

1495 *y el error asociado al mapa interpolado en notación matricial en un*
 1496 *único punto de malla es*

$$\epsilon = 1 - r_{gs}(r_{ss} + \eta I)^{-1}r_{gs}^T.$$

1497 *Para el caso de N puntos de malla y k estaciones obtenemos el mismo*
 1498 *sistema que para el caso de medidas perfectas pero añadiendo el error*
 1499 *ruido-señal en la diagonal principal de la matriz de correlaciones entre*
 1500 *estaciones*

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} = \text{Diag} \left[\mathbf{I} - \begin{pmatrix} r_{g_11} & r_{g_12} & \dots & r_{g_1k} \\ r_{g_21} & r_{g_22} & \dots & r_{g_2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{g_N1} & r_{g_N2} & \dots & r_{g_Nk} \end{pmatrix} \begin{pmatrix} r_{11} + \eta & r_{12} & \dots & r_{1k} \\ r_{21} & r_{22} + \eta & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & r_{kk} + \eta \end{pmatrix}^{-1} \begin{pmatrix} r_{g_11} & r_{g_21} & \dots & r_{g_N1} \\ r_{g_21} & r_{g_22} & \dots & r_{g_N2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{g_N1} & r_{g_N2} & \dots & r_{g_Nk} \end{pmatrix} \right].$$

(1501) $(N \times 1)$ $(N \times k)$ $(k \times k)$ $(k \times N)$

1502 *8 Funciones Empíricas Ortogonales (FEOs)*

1503 *{Interpretación de los sistemas ‘propios’ }*

1504 *Antes de entrar en la teoría para las FEOs, vamos a ver que los vec-*
 1505 *tores propios son equivalentes a modos de oscilación de sistemas físicos.*
 1506 *Imaginemos oscilaciones verticales de bolas en una cuerda. Las bolas*
 1507 *tienen masa m y están separadas por cuerdas elásticas de longitud d en*
 1508 *el equilibrio. Supongamos que los desplazamientos y_n son tan pequeños*
 1509 *que la tensión de la cuerda T se puede considerar constante. El ángulo*
 1510 *de cada cuerda es θ como se ilustra en la figura. Entonces, la ecuación*
 1511 *del movimiento para la bola n es*

$$m \frac{d^2 y_n}{dt^2} = -T \sin \theta_{n-1} - T \sin \theta_n.$$

1512

1513 *Bajo la asunción que los desplazamientos son pequeños, el $\sin \theta_n =$*
 1514 *$\tan \theta_n$, es decir, $\sin \theta_{n-1} = \tan \theta_{n-1} = y_n - y_{n-1}/d$ y $\sin \theta_n = \tan \theta_n =$*
 1515 *$(y_n - y_{n+1})/d$. Entonces la ecuación queda*

$$m \frac{d^2 y_n}{dt^2} = -T \frac{y_n - y_{n-1}}{d} - T \frac{y_n - y_{n+1}}{d},$$

1516 *y si reagrupamos*

$$m \frac{d^2 y_n}{dt^2} = \frac{T}{d} (y_{n-1} - 2y_n + y_{n+1}).$$

1517 *Vamos ahora a substituir una solución osculatoria del tipo*

$$y_n = Y_n e^{i\omega t}$$

1518 *en la ecuación del movimiento*

$$-m\omega^2 Y_n e^{i\omega t} = \frac{T}{d} (Y_{n-1} e^{i\omega t} - 2Y_n e^{i\omega t} + Y_{n+1} e^{i\omega t})$$

$$\frac{-m\omega^2 d}{T} Y_n e^{i\omega t} = \frac{T}{d} (Y_{n-1} e^{i\omega t} - 2Y_n e^{i\omega t} + Y_{n+1} e^{i\omega t}).$$

1520 *Si definimos*

$$\lambda = \frac{m\omega^2 d}{T},$$

1521 *el sistema de ecuaciones a resolver es*

$$-\lambda Y_n = (Y_{n-1} - 2Y_n + Y_{n+1})$$

$$Y_n(2 - \lambda) - Y_{n-1} - Y_{n+1} = 0,$$

1523 *con las condiciones de frontera $Y_0 = Y_{n+1} = 0$ en las paredes.*

1524 *Vamos a suponer ahora el caso de dos bolas. Para la primera bola $n = 1$*

$$Y_1(2 - \lambda) - Y_0 - Y_2 = 0,$$

1525 *y para la bola $n = 2$*

$$Y_2(2 - \lambda) - Y_1 - Y_3 = 0.$$

1526 *Si aplicamos las condiciones de frontera $Y_0 = Y_3 = 0$, nos queda un sistema “propio”, donde λ son los autovalores.*

$$\begin{vmatrix} 2 - \lambda & -1 \\ -1 & 2 - \lambda \end{vmatrix} = 0.$$

1528 *El polinomio característico es*

$$\lambda^2 - 4\lambda + 3 = 0,$$

1529 *y tienes las raíces $\lambda_1 = 1$ y $\lambda_2 = 3$.*

1530 *Resolvamos para los vectores propios:* |

1531 *(1) $\lambda_1 = 1$ entonces*

$$Y_1 - Y_2 = 0$$

1532 *y el vector propio es*

$$\mathbf{y} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

1533

|

1534

(2) $\lambda_2 = 3$ entonces

$$-Y_1 - Y_2 = 0,$$

1535

y el vector propio es

$$\mathbf{y} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

1536

Estas soluciones representan los modos oscilatorios de un sistema físico de dos bolas oscilando verticalmente. Los modos de oscilación se muestran en la figura. Los modos oscilan independientemente uno de otro, y la evolución del sistema es una combinación lineal de los dos modos. De esta forma, lo que estamos haciendo al resolver el problema de las bolas en una cuerda elástica es precisamente la solución de un sistema “propio”.

1543

{Definición de FEOs}

1544

Un análisis en FEOs busca estructuras en los datos que explican la mayor cantidad de varianza de un conjunto de datos bidimensional. La primera dimensión es la dimensión en la que deseamos encontrar una estructura, y la otra dimensión es la dimensión en la que se muestran las diferentes realizaciones. Por ejemplo, un conjunto de series temporales de datos distribuidos espacialmente (i.e. arreglo de anclajes). La primera dimensión es espacio y la segunda es tiempo. Las estructuras en la dimensión espacial son las FEOs, mientras que las estructuras en la dimensión de muestreo se denominan Componentes Principales (CPs). | Tanto las FEOs como las PCs son ortogonales en sus dimensiones. Las FEOs/PCs pueden entenderse de diferentes formas:

- | (i) Transforma variables correlacionadas en un conjunto de variables no correlacionadas que expresan mejor la relación dinámica entre los datos originales. | (ii) Identifica y ordena los vectores ortogonales (o dimensiones) a lo largo de los cuales nuestro conjunto de datos presenta la mayor varianza. | (iii) Una vez definidas las FEOs, es posible encontrar la mejor aproximación de los datos originales con el mínimo número de vectores ortogonales.*

1563

{En general, aplicaremos el análisis en FEOs para describir de manera mas sencilla conjuntos de datos organizados en matrices $M \times N$:} |

- (1) Una matriz espacio-tiempo: Medidas de una variable en M localizaciones y N tiempos. | (2) Una matriz parámetro-tiempo: Medidas de M variables en una localización y N tiempos. | (3) Una matriz parámetro-espacio: Medidas de M variables tomadas en N localizaciones al mismo tiempo. | **Nota:** Un error común es considerar que las FEOs se corresponden con modos de físicos oceánicos. Eso no es cierto!.*

1571 *Los modos físicos en el océano son modos de oscilación que se obtienen
 1572 considerando las ecuaciones que rigen el movimiento y condiciones de
 1573 frontera; las FEOs son simplemente funciones ortogonales que explican
 1574 la mayor cantidad de varianza de un conjunto de datos.*

1575 *Aunque los procesos físicos dominantes son representados por los primeros
 1576 modos de oscilación, no existe una correspondencia uno a uno entre mo-
 1577 dos físicos y FEOs.*

1578 **Teoría |**

1579 *Supongamos M localizaciones de medición con series temporales de tem-
 1580 peratura de N elementos. Queremos descomponer la serie temporal de
 1581 temperatura en una localización dada k como una combinación lineal
 1582 de M funciones ortogonales \mathbf{b}_i cuyas amplitudes son pesadas con M
 1583 coeficientes dependientes del tiempo, es decir,*

$$\mathbf{T}(t) = \sum_{i=1}^M [\alpha_i(t) \mathbf{b}_i],$$

1584 *donde $\alpha_i(t)$ es la amplitud del modo ortogonal i al tiempo $t = t_n$ ($1 \leq$
 1585 $n \leq N$). Los coeficientes $\alpha_i(t)$ nos informan de como varian los modos
 1586 \mathbf{b}_i con el tiempo. Necesitamos tantas funciones ortogonales como esta-
 1587 ciones con series temporales tenemos para poder describir la varianza
 1588 total de los datos originales de temperatura a cada tiempo. Sin embargo,
 1589 en términos prácticos podemos explicar una gran cantidad de varianza
 1590 de los datos originales con las primeras FEOs. Podemos ver el problema
 1591 al revés, es decir, tenemos N funciones temporales cuyas amplitudes
 1592 son pesadas por M coeficientes que varian en el espacio. En este caso
 1593 hablamos de PCs. Ya sea la reducción de los datos en funciones ortog-
 1594 onales espaciales (FEOs) o temporales (PCs), obtenemos los mismos
 1595 resultados.*

1596 *Puesto que queremos \mathbf{b}_i ser ortogonal, requerimos*

$$\mathbf{b}_i^T \mathbf{b}_j = \delta_{ij},$$

1597 *y los coeficientes temporales $\alpha_i = \mathbf{b}_i^T \mathbf{T}(t)$. Son precisamente estos
 1598 coeficientes temporales las CPs, es decir, la proyección de los datos origi-
 1599 nales sobre las FEOs o la expresión de los datos originales en la nueva
 1600 base de vectores ortogonales (el nuevo sistema de coordenadas).*

1601 *El objetivo del análisis es encontrar una base de vectores ortogonales tal
 1602 que las funciones $\alpha_i(t)$ no estén correlacionadas*

$$\langle \alpha_i \alpha_j \rangle = \langle \mathbf{b}_i^T \mathbf{T} \mathbf{b}_j^T \mathbf{T} \rangle = \langle \mathbf{b}_i^T \mathbf{T} \mathbf{T}^T \mathbf{b}_j \rangle = \mathbf{b}_i^T \langle \mathbf{T} \mathbf{T}^T \rangle \mathbf{b}_j = \delta_{ij} \langle \alpha_i^2 \rangle,$$

1603 *donde*

$$\langle \alpha_i^2 \rangle = \frac{1}{N} \sum_{n=1}^N = \alpha_i^2(t_n).$$

1604 *Es decir, la matriz de covarianza de $\alpha_i(t)$ será una matriz diagonal \mathbf{D} .*
 1605 *Para M posiciones se puede reescribir como*

$$\mathbf{B}^T \langle \mathbf{T} \mathbf{T}^T \rangle \mathbf{B} = \mathbf{D} \quad \text{o} \quad \mathbf{B}^T \mathbf{T} \mathbf{T}^T \mathbf{B} = N \mathbf{D}$$

1606 *donde \mathbf{B} es una matriz ortogonal cuyas columnas son los vectores or-
 1607 *togonales \mathbf{b}_i y \mathbf{D} es una matriz diagonal compuesta por las varianzas de
 1608 *las funciones temporales $\alpha_i(t)$. Si multiplicamos por \mathbf{B} llegamos a un
 1609 *sistema propio****

$$\langle \mathbf{T} \mathbf{T}^T \rangle \mathbf{B} = \mathbf{B} \mathbf{D}.$$

1610 *Este tipo de sistema propio es conocido. La diagonal de \mathbf{D} esta com-
 1611 *puesta por valores propios y las columnas de \mathbf{B} son los vectores propios.
 1612 Los vectores propios son denominados FEOs, y los valores propios son
 1613 las varianzas de las amplitudes α_i . Básicamente hemos realizado una
 1614 transformación de coordenadas de tal forma que los vectores propios \mathbf{b}_i
 1615 indican combinaciones lineales de datos que no están correlacionados
 1616 (*i.e.*, $\langle \alpha_i \alpha_j \rangle = \delta_{ij} \langle \alpha_i^2 \rangle$). Esta descomposición de los datos en
 1617 FEOs es óptima en el sentido de mínimos cuadrados. Imaginemos que
 1618 queremos un conjunto de K vectores que mejor aproxima \mathbf{T}**

$$\langle (\hat{\mathbf{T}} - \mathbf{T})^T (\hat{\mathbf{T}} - \mathbf{T}) \rangle = \langle \mathbf{T}^T \mathbf{T} \rangle - \sum_{i=1}^K \langle \alpha_i^2 \rangle.$$

1619 *El problema en mínimos cuadrados, bajo las restricciones que las fun-
 1620 *ciones \mathbf{b}_i sean ortogonales, se puede escribir a partir de los multipli-
 1621 *cadores de Lagrange***

$$L = \sum_{i=1}^K \left[\mathbf{b}_i^T \langle \mathbf{T} \mathbf{T}^T \rangle \mathbf{b}_i - \lambda_i (\mathbf{b}_i^T \mathbf{b}_i - 1) \right].$$

1622 *Si derivamos L respecto de \mathbf{b}_i obtenemos el sistema propio*

$$\langle \mathbf{T} \mathbf{T}^T \rangle \mathbf{b}_i = \lambda_i \mathbf{b}_i.$$

1623 *De este análisis deducimos que las primeras K funciones ortogonales o
 1624 *FEOs son las mejores funciones que explican la máxima varianza de los
 1625 *datos originales, donde los valores propios están ordenados de mayor
 1626 a menor. Es decir, no existe un subset de datos más pequeño que K
 1627 *funciones ortogonales que produce el error cuadrático medio menor. En
 1628 *este sentido las FEOs son los mejores ‘descriptores’ de la varianza de
 1629 *los datos.******

1630 *Una matriz de datos de M localizaciones y N tiempos se puede descom-
 1631 *poner**

$$\mathbf{C}\mathbf{B} = \mathbf{B} ,$$

1632 donde

$$\mathbf{C} = \mathbf{T}\mathbf{T}^T = \begin{pmatrix} \sum_{i=1}^N T_1(t_i)T_1(t_i) & \sum_{i=1}^N T_1(t_i)T_2(t_i) & \dots & \sum_{i=1}^N T_1(t_i)T_M(t_i) \\ \sum_{i=1}^N T_2(t_i)T_1(t_i) & \sum_{i=1}^N T_2(t_i)T_2(t_i) & \dots & \sum_{i=1}^N T_2(t_i)T_M(t_i) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^N T_M(t_i)T_1(t_i) & \sum_{i=1}^N T_M(t_i)T_2(t_i) & \dots & \sum_{i=1}^N T_M(t_i)T_M(t_i) \end{pmatrix},$$

1633 es la matriz de covarianza entre las series temporales de temperatura en
1634 localizaciones espaciales, unas con otras;

$$\mathbf{T} = \begin{pmatrix} T_1(t_1) & T_1(t_2) & \dots & T_1(t_N) \\ T_2(t_1) & T_2(t_2) & \dots & T_2(t_N) \\ \vdots & \vdots & \ddots & \vdots \\ T_M(t_1) & T_M(t_2) & \dots & T_M(t_N) \end{pmatrix}$$

1635 es la

1636 matriz de datos de temperatura;

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1M} \\ b_{21} & b_{22} & \dots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{M1} & b_{M2} & \dots & b_{MM} \end{pmatrix},$$

1637 es la matriz con los vectores propios \mathbf{b}_i como columnas, y

$$= \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_M \end{pmatrix},$$

1638 es la matriz diagonal compuesta por los valores propios $\lambda_i = <\alpha_i^2>$
1639 | | Otra forma de entender las FEOs (PCs) consiste en un análisis en
1640 valores propios de las matrices de dispersión de nuestras matrices de

1641 *datos. La matriz de dispersión es el producto matricial de la matriz con
1642 su transpuesta, o a la inversa. La primera matriz de dispersión es*

$$\begin{aligned} \mathbf{C} = \mathbf{T}\mathbf{T}^T &= \begin{pmatrix} T_1(t_1) & T_1(t_2) & \dots & T_1(t_N) \\ T_2(t_1) & T_2(t_2) & \dots & T_2(t_N) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ T_M(t_1) & T_M(t_2) & \dots & T_M(t_N) \end{pmatrix} \begin{pmatrix} T_1(t_1) & T_2(t_1) & \dots & T_M(t_1) \\ T_1(t_2) & T_2(t_2) & \dots & T_M(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ T_1(t_N) & T_2(t_N) & \dots & T_M(t_N) \end{pmatrix} = \\ &= \begin{pmatrix} \sum_{i=1}^N T_1(t_i)T_1(t_i) & \sum_{i=1}^N T_1(t_i)T_2(t_i) & \dots & \sum_{i=1}^N T_1(t_i)T_M(t_i) \\ \sum_{i=1}^N T_2(t_i)T_1(t_i) & \sum_{i=1}^N T_2(t_i)T_2(t_i) & \dots & \sum_{i=1}^N T_2(t_i)T_M(t_i) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^N T_M(t_i)T_1(t_i) & \sum_{i=1}^N T_M(t_i)T_2(t_i) & \dots & \sum_{i=1}^N T_M(t_i)T_M(t_i) \end{pmatrix}, \\ &\quad (M \times M) \end{aligned}$$

1645 *y la segunda es*

$$\begin{aligned} \mathbf{C} = \mathbf{T}^T \mathbf{T} &= \begin{pmatrix} T_1(t_1) & T_2(t_1) & \dots & T_M(t_1) \\ T_1(t_2) & T_2(t_2) & \dots & T_M(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ T_1(t_N) & T_2(t_N) & \dots & T_M(t_N) \end{pmatrix} \begin{pmatrix} T_1(t_1) & T_1(t_2) & \dots & T_1(t_N) \\ T_2(t_1) & T_2(t_2) & \dots & T_2(t_N) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ T_M(t_1) & T_M(t_2) & \dots & T_M(t_N) \end{pmatrix} = \\ &= \begin{pmatrix} \sum_{i=1}^M T_i(t_1)T_i(t_1) & \sum_{i=1}^M T_i(t_1)T_i(t_2) & \dots & \sum_{i=1}^M T_i(t_1)T_i(t_N) \\ \sum_{i=1}^M T_i(t_2)T_i(t_1) & \sum_{i=1}^M T_i(t_2)T_i(t_2) & \dots & \sum_{i=1}^M T_i(t_2)T_i(t_N) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^M T_i(t_N)T_i(t_1) & \sum_{i=1}^M T_i(t_N)T_i(t_2) & \dots & \sum_{i=1}^M T_i(t_N)T_i(t_N) \end{pmatrix}. \\ &\quad (N \times N) \end{aligned}$$

1648 *Ambas matrices de dispersión obtenidas del producto de la matriz de
1649 datos consigo misma son matrices de covarianza simétricas. La primera
1650 matriz de dispersión $\mathbf{T}\mathbf{T}^T$ es una matriz $M \times M$ con lo que hemos
1651 eliminado la dimensión temporal o dimensión de muestreo. En este
1652 caso, la matriz de dispersión es una matriz de covarianzas temporales
1653 (determinadas por sus variaciones temporales) de las estaciones unas
1654 con otras. En la segunda matriz $\mathbf{T}^T \mathbf{T}$ se invierten las dimensiones, la
1655 matriz resultante es $N \times N$, y es una matriz de covarianzas espaciales*

(determinadas por sus variaciones espaciales) entre los diferenetes tiempos. \ \ {\bf Ejemplo:} Un ejmplo sencillo viene dado por el diagrama de dispersión de la figura. La primera EOF o vector propio que explica la mayor varianza sería la recta que se ajusta al conjunto de puntos y la segunda EOF sería la línea perpendicular a la ajustada.

1661

{Relación entre FEOs y CPs}

1663

Como ya hemos visto nuestro sistema propio es

$$\mathbf{B}^T < \mathbf{T} \mathbf{T}^T > \mathbf{B} = \mathbf{D} \quad \text{o} \quad \mathbf{B}^T \mathbf{T} \mathbf{T}^T \mathbf{B} = N \mathbf{D},$$

1664

donde \mathbf{B} es una matriz cuyas columnas son los vectores propios y \mathbf{D} es una matriz cuadrada con los M valores propios en la diagonal principal. Si queremos expresar los datos originales en términos de los vectores propios, entonces debemos usar la definición

$$\alpha_i(t) = \mathbf{b}_i^T \mathbf{T}(t),$$

1668

que en notación matricial se puede expresar como

$$\mathbf{Z} = \mathbf{B}^T \mathbf{T},$$

1669

y finalmente para recuperar los datos originales a partir de la base de FEOs usamos

$$\mathbf{T} = \mathbf{BZ},$$

1671

ya que $\mathbf{BB}^T = \mathbf{I}$. La matriz \mathbf{Z} contiene los vectores de las CPs, que no son mas que las amplitudes por las cuales multiplicamos las FEOs para obtener los datos originales de vuelta. De esta forma podemos ir de un espacio (vectores propios) al otro (datos originales) con la matriz de FEOs. Supongamos que tenemos un conjunto de vectores propios ortogonales y normalizados. El primero de ellos por ejemplo sería

$$\mathbf{e} = \begin{pmatrix} e_{11} \\ e_{21} \\ \vdots \\ \vdots \\ e_{M1} \end{pmatrix}.$$

1677

Si ponemos todos los vectores propios en columna obtenemos la matriz cuadrada \mathbf{B} , la cual es ortonormal $\mathbf{B}^T \mathbf{B} = \mathbf{I}$. Si queremos proyectar un vector propio sobre los datos originales y obtener la amplitud de este vector propio en cada tiempo, debemos de hacer

$$\mathbf{e}^T \mathbf{T} = (e_{11} \ e_{21} \ \dots \ e_{M1}) \begin{pmatrix} T_1(t_1) & T_1(t_2) & \dots & T_1(t_N) \\ T_2(t_1) & T_2(t_2) & \dots & T_2(t_N) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ T_M(t_1) & T_M(t_2) & \dots & T_M(t_N) \end{pmatrix} = (z_{11} \ z_{12} \ \dots \ z_{1N}),$$

1681 donde, por ejemplo, $z_{11} = e_{11}T_1(t_1) + e_{21}T_2(t_1) + \dots + e_{M1}T_M(t_1)$.
 1682 Si hacemos lo mismo para todos los otros vectores propios, obtenemos
 1683 series temporales de longitud N para cada FEO, lo cual se denominan
 1684 componentes principales de cada EOF

$$\mathbf{Z} = \mathbf{B}^T \mathbf{T}.$$

1685

$$M \times N$$

1686 Las CPs también son ortogonales. Si substituimos $\mathbf{T} = \mathbf{BZ}$, en $\mathbf{B}^T \mathbf{TT}^T \mathbf{B} =$
 1687 \mathbf{ND} obtenemos

$$\mathbf{B}^T \mathbf{BZ}(\mathbf{BZ})^T \mathbf{B} = \mathbf{ND}$$

1688

$$\mathbf{IZZ}^T \mathbf{B}^T \mathbf{B} = \mathbf{ND}$$

1689

$$\mathbf{IZZ}^T \mathbf{I} = \mathbf{ND}$$

1690

$$\mathbf{ZZ}^T = \mathbf{ND} \quad \text{o} \quad <\mathbf{ZZ}^T> = \mathbf{D}.$$

1691 Por lo tanto no solo las FEOs sino también las CPs son ortogonales.

1692 {Equivalecia con descomposición en valores singulares (SVD)}

Definición de SVD:

La descomposición en valores singulares es básicamente un método de reducción dimensional de un conjunto de datos inicialmente definidos sobre un espacio multidimensional. Esta puede ser usada para encontrar simultáneamente las FEOs y PCs de una matriz de datos.

Una matriz \mathbf{A} puede ser descompuesta en un producto de matrices:
 (i) una matriz ortogonal \mathbf{U} (i.e., $\mathbf{U}^T \mathbf{U} = \mathbf{I}$), una matriz diagonal \mathbf{S} ,
 y la transpuesta de una matriz ortogonal \mathbf{V} (i.e., $\mathbf{V}^T \mathbf{V} = \mathbf{I}$)

$$\mathbf{A} = \mathbf{USV}^T.$$

Las columnas de \mathbf{U} son vectores propios ortogonales de \mathbf{AA}^T o los vectores singulares izquierdos de \mathbf{A} , las columnas de \mathbf{V} son vectores propios ortogonales de $\mathbf{A}^T \mathbf{A}$ o los vectores singulares derechos de \mathbf{A} , y \mathbf{S} es una matriz diagonal que contiene las raíces cuadradas de los valores propios positivos de las matrices \mathbf{U} o \mathbf{V} en orden descendente (o los valores singulares no-nulos de \mathbf{A}). con orden descendente nos referimos a que la primera columna esta asociada al valor propio mas grande, la segunda columna al segundo valor propio mas grande, y así sucesivamente.

Interpretación geométrica: La SVD descompone la matriz \mathbf{A} en tres transformaciones algebraicas sucesivas: (i) una rotación inicial \mathbf{U} , un escalamiento \mathbf{S} a lo largo de los ejes coordenados, una rotación final \mathbf{V} . Las longitudes σ_1 y σ_2 de la elipse son los valores singulares de \mathbf{A} (ver script *interpreta_svd.m*).

1693 Supongamos la matriz de datos compuesta por M series temporales de
 1694 temperatura de longitud N

$$\mathbf{T} = \begin{pmatrix} T_1(t_1) & T_1(t_2) & \dots & T_1(t_N) \\ T_2(t_1) & T_2(t_2) & \dots & T_2(t_N) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ T_M(t_1) & T_M(t_2) & \dots & T_M(t_N) \end{pmatrix}.$$

1695 La matriz de dispersión es

$$\langle \mathbf{T}\mathbf{T}^T \rangle = \frac{1}{N} \mathbf{T}\mathbf{T}^T.$$

1696 Sabemos que la matriz \mathbf{T} puede descomponerse en una matriz ortogonal
 1697 \mathbf{U} , una matriz diagonal \mathbf{S} , y la transpuesta de una matriz ortogonal \mathbf{V} .
 1698 Esto es

$$\mathbf{T} = \mathbf{U}\mathbf{S}\mathbf{V}^T,$$

1699 donde el número de valores singulares no nulos indican el rango de \mathbf{T} .
 1700 Si $K < N$ y el número de filas (i.e., los datos) son linealmente inde-
 1701 pendientes entonces el rango será K . Ahora la matriz de covarianza
 1702 es

$$\frac{1}{N} \mathbf{T}\mathbf{T}^T = \frac{1}{N} (\mathbf{U}\mathbf{S}\mathbf{V}^T)(\mathbf{U}\mathbf{S}\mathbf{V}^T)^T = \frac{1}{N} \mathbf{U}\mathbf{S}\mathbf{V}^T \mathbf{V}(\mathbf{U}\mathbf{S})^T = \frac{1}{N} \mathbf{U}\mathbf{S}\mathbf{V}^T \mathbf{V}\mathbf{S}^T \mathbf{U}^T = \frac{1}{N} \mathbf{U}\mathbf{S}\mathbf{S}^T \mathbf{U}^T.$$

1703 La derecha del igual es la descomposición en valores propios de la ma-
 1704 triz de covarianza, donde $\mathbf{S}\mathbf{S}^T$ es cuadrada y diagonal con los elementos
 1705 igual a $N\lambda_i$; y las columnas de \mathbf{U} son las FEOs. Las amplitudes de las
 1706 FEOs vienen dadas por las filas de la matriz

$$\mathbf{U}^T \mathbf{T} = \mathbf{S}\mathbf{V},$$

1707 asociado con valores singulares no nulos.

1708 {Interpretación de las FEOs}

1709 Como comentario final decir que las FEOs no son muy fáciles de in-
 1710 terpretar. Matemáticamente son estructuras que representan la mayor
 1711 cantidad de varianza de los datos originales y que son ortogonales entre
 1712 ellas. En ocasiones estas estructuras nos dan estructuras con sentido
 1713 físico en un conjunto de datos, u otras no. Las estructuras particu-
 1714 lares encontradas dependerán de como hemos acomodado nuestra matriz
 1715 bidimensional de datos. Algunas sugerencias para detectar si las FEOs
 1716 tienen sentido físico son las siguientes: | (1) ?La varianza de tu FEO
 1717 es mas grande que lo que esperabas si los datos originales
 1718 no tenían estructura alguna? \| (2) ?Existe una explicación
 1719 apriori para las estructuras que has encontrado? ?Se pueden explicar
 1720 las estructuras en términos de alguna teoría? ?Las estruc-
 1721 turas se comportan consistentemente con la teoría apriori? | (3) ?Cuán
 1722 robustas son las estructuras a la elección del dominio de

1723 *la estructura? Si cambias el dominio del análisis, esas estructuras
 1724 cambian significativamente? Si las estructuras están definidas
 1725 en un espacio geográfico, y cambias el tamaño de la región,
 1726 las estructuras cambian significativamente? Si las estruc-
 1727 turas están definidas en el espacio de parámetros y añades o eliminas un
 1728 parámetros, los resultados cambian de forma suave o aleatoriamente?
 1729 | (4) Cuán robustas son las estructuras a los datos usados?
 1730 ?Si divides los datos originales en fracciones menores y haces el análisis
 1731 de cada fracción, obtienes las mismas estructuras?*

1732 *Ejemplo de descomposición en valores singulares | Supongamos la ma-
 1733 triz*

$$\mathbf{A} = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}.$$

1734 *Para encontrar \mathbf{U} debemos resolver el problema en vectores y valores
 1735 propios de la matriz $\mathbf{A}\mathbf{A}^T$*

$$\mathbf{A}\mathbf{A}^T = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 11 & 1 \\ 1 & 11 \end{pmatrix}$$

1736 *El sistema de ecuaciones ‘propio’ es*

$$\begin{pmatrix} 11 & 1 \\ 1 & 11 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

1737 *Si resolvemos para λ*

$$\begin{vmatrix} 11 - \lambda & 1 \\ 1 & 11 - \lambda \end{vmatrix} = 0,$$

1738 *lo que deja el polinomio característico*

$$(\lambda - 10)(\lambda - 12) = 0,$$

1739 *con raíces (valores propios) $\lambda_1 = 12$ y $\lambda_2 = 10$.*

1740 *Si sustituimos en el sistema de ecuaciones ‘propio’ el primer valor pro-
 1741 pio $\lambda_1 = 12$*

$$(11 - 12)x_1 + x_2 = 0$$

$$x_1 = x_2$$

1743 *Para $x_1 = 1$ obtenemos que $x_2 = 1$. Entonces obtenemos el vector pro-
 1744 pio $\mathbf{v}_1 = [1, 1]$.*

1745 *Si sustituimos en el sistema de ecuaciones ‘propio’ el primer valor pro-
 1746 pio $\lambda_1 = 10$*

$$(11 - 10)x_1 + x_2 = 0$$

$$x_1 = -x_2$$

¹⁷⁴⁸ Para $x_1 = 1$ obtenemos que $x_2 = -1$. Entonces obtenemos el vector propio $\mathbf{v}_2 = [1, -1]$.

¹⁷⁵⁰ Si organizamos la matriz con columnas correspondientes a los vectores propios asociados a los valores propios de mayor a menor, obtenemos

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

¹⁷⁵² Finalmente, sabemos que \mathbf{U} tiene que ser ortonormal. Vamos a usar el proceso de Gram-Schmidt para ortonormalizar las columnas de \mathbf{U} . ¹⁷⁵³ Empezamos normalizando la primera columna

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|} = \frac{[1, 1]}{\sqrt{2}} = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right],$$

¹⁷⁵⁵ y calculamos el vector ortonormal como

$$\mathbf{w}_2 = \mathbf{v}_2 - \mathbf{u}_1 \cdot \mathbf{v}_2 \mathbf{u}_1 = [1, -1] - \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right] \cdot [1, -1] \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right] = [1, -1] - [0, 0] = [1, -1].$$

¹⁷⁵⁶ Si lo normalizamos

$$\mathbf{u}_2 = \frac{\mathbf{w}_2}{|\mathbf{w}_2|} = \left[\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right],$$

¹⁷⁵⁷ dejando la matriz

$$\mathbf{U} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

¹⁷⁵⁸ Similarmente, para calcular la matriz \mathbf{V} debemos resolver el problema en vectores y valores propios de la matriz $\mathbf{A}^T \mathbf{A}$

$$\mathbf{A}^T \mathbf{A} = \begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{pmatrix}$$

¹⁷⁶⁰ El sistema de ecuaciones ‘propio’ es

$$\begin{pmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$(10 - \lambda)x_1 + 2x_3 = 0$$

$$(10 - \lambda)x_2 + 4x_3 = 0$$

$$2x_1 + 4x_2 + (2 - \lambda)x_3 = 0$$

¹⁷⁶³ Si resolvemos para λ

$$\begin{vmatrix} 10 - \lambda & 0 & 2 \\ 0 & 10 - \lambda & 4 \\ 2 & 4 & 2 - \lambda \end{vmatrix} = 0$$

1764 *lo que deja el polinomio característico*

$$\lambda(\lambda - 10)(\lambda - 12) = 0,$$

1765 *con raíces (valores propios) $\lambda_1 = 12$, $\lambda_2 = 10$, y $\lambda_3 = 0$. Substituyendo*
 1766 *en el sistema ‘propio’ para $\lambda_1 = 12$*

$$(10 - 12)x_1 + 2x_3 = -2x_1 + 2x_3 = 0$$

$$1767 \quad x_1 = 1; \quad x_3 = 1$$

$$1768 \quad (10 - 12)x_2 + 4x_3 = -2x_2 + 4x_3 = 0$$

$$1769 \quad x_2 = 2x_3$$

$$1770 \quad x_2 = 2$$

1771 *Entonces, para $\lambda_1 = 12$ obtenemos el vector propio $\mathbf{v}_1 = [1, 2, 1]$. | Para*
 1772 *$\lambda_2 = 10$*

$$(10 - 10)x_1 + 2x_3 = 2x_3 = 0$$

$$1773 \quad x_3 = 0$$

$$1774 \quad 2x_1 + 4x_2 = 0$$

$$1775 \quad x_1 = -2x_2$$

$$1776 \quad x_1 = 2; \quad x_2 = -1$$

1777 *y obtenemos $\mathbf{v}_2 = [2, -1, 0]$ para $\lambda_2 = 10$. | Finalmente, para $\lambda_3 = 0$*

$$1778 \quad 10x_1 + 2x_3 = 0$$

$$1779 \quad x_3 = -5$$

$$1780 \quad 10x_1 - 20 = 0$$

$$1781 \quad x_2 = 2$$

$$1782 \quad 2x_1 + 8 - 10 = 0$$

$$1783 \quad x_1 = 1$$

1784 *lo que implica que para $\lambda_3 = 0$ $\mathbf{v}_3 = [1, 2, -5]$. | | Si organizamos*
 1785 *los vectores de acuerdo con el valor de los valores propios (de mayor a*
menor) obtenemos la matriz

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & 0 & -5 \end{pmatrix}$$

1786 *Ahora vamos a ortonormalizarla con el proceso de Gram-schmidt*

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|} = \frac{[1, 2, 1]}{\sqrt{6}},$$

1787 *y calculamos el vector ortonormal como*

$$\mathbf{w}_2 = \mathbf{v}_2 - \mathbf{u}_1 \cdot \mathbf{v}_2 \mathbf{u}_1 = [2, -1, 0]$$

1788 *Si lo normalizamos*

$$\mathbf{u}_2 = \frac{\mathbf{w}_2}{|\mathbf{w}_2|} = \left[\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}, 0 \right]$$

1789 *El último vector ortonormal a calcular es*

$$\mathbf{w}_3 = \mathbf{v}_3 - \mathbf{u}_1 \cdot \mathbf{v}_3 \mathbf{u}_1 - \mathbf{u}_2 \cdot \mathbf{v}_3 \mathbf{u}_2 = \left[\frac{-2}{3}, \frac{-4}{3}, \frac{10}{3} \right]$$

1790 *Si lo normalizamos*

$$\mathbf{u}_3 = \frac{\mathbf{w}_3}{|\mathbf{w}_3|} = \left[\frac{1}{\sqrt{30}}, \frac{2}{\sqrt{30}}, \frac{-5}{\sqrt{30}} \right]$$

1791 *dejando la matriz*

$$\mathbf{V} = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{pmatrix}.$$

1792 *Finalmente, lo que realmente queremos es*

$$\mathbf{V}^T = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{pmatrix}.$$

1793 *Para calcular \mathbf{S} debemos tomar las raíces cuadradas de los valores pro-*
 1794 *pios diferentes de cero ($\lambda_i \neq 0$) y colocarlos en la diagonal principal en*
 1795 *orden descendente. Es decir, el valor propio mayor en la posición s_{11} ,*
 1796 *el siguiente mas grande en s_{22} , y así sucesivamente. Los valores propios*
 1797 *diferentes de cero son iguales para \mathbf{U} y \mathbf{V} con lo que no importa de cual*
 1798 *los tomemos. Puesto que solo hay dos valores propios diferentes de cero*
 1799 *y el orden de las matrices \mathbf{U} y \mathbf{V} es 3×3 , debemos añadir una columna*
 1800 *de ceros a \mathbf{S}*

$$\mathbf{S} = \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix}.$$

1801 *Ahora ya tenemos todas las matrices de la descomposición en valores*
 1802 *singulares:*

$$\begin{aligned} \mathbf{A} &= \mathbf{USV}^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{pmatrix} = \\ &= \begin{pmatrix} \frac{12}{\sqrt{2}} & \frac{\sqrt{10}}{\sqrt{2}} & 0 \\ \frac{12}{\sqrt{2}} & -\frac{\sqrt{10}}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{pmatrix} = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}. \end{aligned}$$

1804 **9 Análisis espectral o análisis de Fourier**

1805 **9.1 Introducción**

1806 Una función periódica es aquella cuyos valores se repiten a intervalos
 1807 regulares. El tiempo entre las sucesivas repeticiones se denomina periodo
 1808 τ . Normalmente lo definimos entre sucesivas crestas. Matemáticamente,
 1809 una función es periódica si $f(t) = f(t + T)$ para todo valor de T .

1810
 1811
$$\text{Función periódica } f(t) = f(t + \tau).$$

1812 La frecuencia de una función periódica se define como el inverso del pe-
 1813 ríodo, $f = 1/\tau$, es decir el número de ciclos por unidad de tiempo (si es
 1814 por segundo hablamos de Hercios, Hz). Si un ciclo equivale a 2π radi-
 1815 anes, entonces el número de radianes por segundo es lo que se conoce
 1816 por la frecuencia angular fundamental:

$$\omega = \frac{2\pi}{T}.$$

1817 Las funciones periódicas también se pueden definir en el espacio. En-
 1818 tonces el periodo se define como

$$\tau = \lambda/v_p,$$

1819 donde λ es la longitud de onda y $v_p = \lambda/\tau = \omega/k$ es la velocidad de
 1820 fase. La longitud de onda es una distancia entre estados de la onda que
 1821 se repiten, e.j. entre dos crestas. El número de onda k es el número de
 1822 ondas contenidas en una unidad de distancia

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{v_p}$$

1823 El valor promedio de una función periódica es:

$$f_m = \frac{1}{\tau} \int_0^\tau f(t) dt,$$

1824 y su valor cuadrático medio (o RMS, en inglés) es:

$$f_{rms} = \sqrt{\frac{1}{\tau} \int_0^\tau f^2(t) dt},$$

1825 donde las integrales se han definido en el intervalo $0, \tau$, aunque se
 1826 pueden definir en cualquier intervalo que abarque un periodo, e.j. de
 1827 $-\tau/2$ a $\tau/2$.

1828 Una de las ondas periódicas mas utilizadas es la sinusoidal o cosenosoidal.

$$f(t) = A \sin(\omega t + \theta),$$

1829 siendo A la amplitud y θ su fase inicial. En este caso el valor medio
 1830 es cero y su rms es $A/\sqrt{2}$. % Recordar que existen dos frecuen-
 1831 cias básicas: (i) Frecuencia de Nyquist $f_N = 1/(2\Delta t)$ (la frecuen-
 1832 cia mas alta que podemos resolver) y (ii) Frecuencia fundamental
 1833 $f_0 = 1/(\Delta t N) = 1/T$ (la frecuencia mas baja que podemos resolver).

1834 9.2 Serie de Fourier

1835 El principio básico del análisis de Fourier es que cualquier función peri-
 1836ódica $f(t)$ definida en el intervalo $[0, T]$ se puede descomponer en suma
 1837 de funciones simples, sinusoidales y cosinusoidales, o series de Fourier
 1838 de la forma

$$f(t) = \bar{f} + \sum_p [A_p \cos(\omega_p t) + B_p \sin(\omega_p t)],$$

1839 donde \bar{f} es el valor medio de la serie temporal, A_p y B_p son con-
 1840 stantes denominados coeficientes de Fourier, y $\omega_p = 2\pi p f_0 = 2\pi p/\tau$ es
 1841 múltiplo de la frecuencia angular fundamental.

1842 Si tenemos suficientes componentes de Fourier cada valor de la serie
 1843 original se puede reconstruir. La contribución que cada componente
 1844 tiene sobre la varianza de la serie temporal es una medida de la impor-
 1845 tancia de una frecuencia particular en la serie original. El punto clave
 1846 aquí es que el conjunto de coeficientes de Fourier con amplitudes A_p
 1847 y B_p forman un espectro el cual define la contribución de cada com-
 1848 ponente oscilatoria ω_p sobre la ‘energía’ total de la señal original. En
 1849 concreto, el espectro de potencia (Power spectrum) define la energía por
 1850 unidad de banda de frecuencia de una serie temporal. Puesto que debe-
 1851 mos definir dos amplitudes A_p y B_p , hay dos grados de libertad por
 1852 estimación espectral.

1853 Como ya hemos dicho el primer armónico ($p = 1$) oscila con frecuencia
 1854 fundamental $\omega_1 = 2\pi f_1$. El armónico $N/2$, el cual nos da la compo-
 1855 nente con la frecuencia más alta que puede ser resuelta tiene frecuencia
 1856 $f_N = N/2/N\Delta t = 1/(2\Delta T)$ ciclos por unidad de tiempo y un periodo
 1857 de $2\Delta t$. Esta es la frecuencia de Nyquist.

Integrales de funciones trigonométricas útiles:

$$\int \cos ax dx = \frac{\sin ax}{a} + C$$

$$\int \sin ax dx = -\frac{\cos ax}{a} + C$$

$$\int \sin^2 ax dx = \frac{x}{2} - \frac{\sin 2ax}{4a} + C$$

$$\int \cos^2 ax dx = \frac{x}{2} + \frac{\sin 2ax}{4a} + C$$

1858 *Las series de Fourier se definen como*

$$f(t) = \frac{1}{2}A_0 + \sum_{p=1}^{\infty} [A_p \cos(\omega_p t) + B_p \sin(\omega_p t)],$$

1859 *en la cual*

$$\omega_p = 2\pi f_p = 2\pi p f_1 = 2\pi p/T; \quad p = 1, 2, \dots,$$

1860 *es la frecuencia de la componente pésima en radianes por unidad de*
 1861 *tiempo (f_p es en ciclos por unidad de tiempo) y $A_0/2$ es la media de la*
 1862 *serie temporal.*

1863 *Para obtener los coeficientes A_p , debemos multiplicar la expresión de la*
 1864 *descomposición de Fourier por $\cos(\omega_r t)$ e integrar sobre la serie com-*
 1865 *pleta.*

$$\begin{aligned} \int_0^T f(t) \cos(\omega_r t) dt &= \frac{1}{2} A_0 \int_0^T \cos(\omega_r t) dt + \\ &+ \sum_{p=1}^{\infty} A_p \int_0^T \cos(\omega_p t) \cos(\omega_r t) dt + \\ &+ \sum_{p=1}^{\infty} B_p \int_0^T \sin(\omega_p t) \cos(\omega_r t) dt. \end{aligned}$$

1866 *Si usamos las siguientes condiciones de ortogonalidad:*

$$\int_0^T \sin(\omega_p t) \cos(\omega_r t) dt = 0$$

$$\int_0^T \cos(\omega_p t) \cos(\omega_r t) dt = \begin{cases} T, & p = r = 0 \\ T/2, & p = r > 0 \\ 0, & p \neq r \end{cases}$$

$$\int_0^T \sin(\omega_p t) \sin(\omega_r t) dt = \begin{cases} 0, & p = r = 0 \\ T/2, & p = r > 0 \\ 0, & p \neq r \end{cases}$$

1871 *entonces encontramos que para $r = 0; p \neq r$ la ecuación de arriba se re-*
 1872 *duce a*

$$\int_0^T f(t) dt = \frac{A_0}{2} T,$$

1873 *es decir,*

$$A_0 = \frac{2}{T} \int_0^T f(t) dt = 2\overline{f(t)},$$

¹⁸⁷⁴ dos veces el valor medio de la serie $f(t)$. Es por ello que se añade el
¹⁸⁷⁵ factor de $1/2$ en la serie de Fourier. Es decir, para que el primer tér-
¹⁸⁷⁶ mino de la serie de Fourier sea igual a la media de la serie temporal
¹⁸⁷⁷ $\bar{f(t)} = 1/2A_0$.

¹⁸⁷⁸ Cuando $p \neq 0$ el único término no despreciable de la derecha de la
¹⁸⁷⁹ expresión de arriba sucede cuando $r = p$

$$\int_0^T f(t) \cos(\omega_p t) dt = \frac{A_p}{2} T,$$

¹⁸⁸⁰ y entonces

$$A_p = \frac{2}{T} \int_0^T f(t) \cos(\omega_p t) dt, \quad p = 1, 2, \dots$$

¹⁸⁸¹ Los otros coeficientes B_p son obtenidos igualmente multiplicando por
¹⁸⁸² $\sin(\omega_r t)$ en lugar de $\cos(\omega_r t)$

$$B_p = \frac{2}{T} \int_0^T f(t) \sin(\omega_p t) dt, \quad p = 1, 2, \dots$$

Relaciones trigonométricas útiles:

$$\cos(\alpha - \beta) = \cos\alpha \cos\beta + \sin\alpha \sin\beta$$

$$\sin(\alpha - \beta) = \sin\alpha \cos\beta - \cos\alpha \sin\beta$$

$$\operatorname{tg}\gamma = \frac{\sin\gamma}{\cos\gamma}$$

De estas relaciones de arriba se puede llegar a:

$$A \cos\theta + B \sin\theta = C \cos(\theta - \theta_0),$$

donde

$$C = \sqrt{A^2 + B^2},$$

y

$$\theta_0 = \operatorname{arctg}\left(\frac{B}{A}\right)$$

A partir de esta relación

$$e^{i\theta} = \cos\theta + i \sin\theta,$$

llegamos a

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}$$

1883 También podemos representar la serie de Fourier en notación compacta
 1884 como:

$$f(t) = \frac{1}{2}C_0 + \sum_{p=1}^{\infty} C_p \cos(\omega_p t - \theta_p),$$

1885 en la cual la amplitud de la pésima componente es

$$C_p = \sqrt{A_p^2 + B_p^2}, \quad p = 1, 2, \dots$$

1886 donde $C_0 = A_0$ ($B_0 = 0$) es dos veces el valor promedio de la serie y

$$\theta_p = \text{arctg}[B_p/A_p], \quad p = 1, 2, \dots$$

1887 es el ángulo de fase de la componente al tiempo $t = 0$. El ángulo de fase
 1888 nos informa del “desfase” lag relativo de las componentes en radianes
 1889 (o grados) medido en el sentido contrario a las agujas del reloj desde
 1890 el eje real definido por $B_p = 0, A_p > 0$. El correspondiente tiempo de
 1891 desfase para la componente pésima es $t_p = \theta_p/2\pi f_p$ en el cual θ_p esta
 1892 medida en radianes. La energía espectral se define como las amplitudes
 1893 de los coeficientes de Fourier al cuadrado, lo cual representa la varianza
 1894 y entonces la energía

$$C_p^2 = A_p^2 + B_p^2.$$

1895 | | De igual forma con las relaciones trigonométricas de arriba se puede
 1896 expresar las series de Fourier en notación compleja. Usando

$$\text{sen}\omega_p t = \frac{e^{i\omega_p t} - e^{-i\omega_p t}}{2i} \quad y \quad \text{cos}\omega_p t = \frac{e^{i\omega_p t} + e^{-i\omega_p t}}{2},$$

1897 obtenemos

$$\begin{aligned} f(x) &= \frac{1}{2}A_0 + \sum_{p=1}^{\infty} [A_p \cos(\omega_p t) + B_p \text{sen}(\omega_p t)] = \\ &= \frac{1}{2}A_0 + \sum_{p=1}^{\infty} \left[A_p \frac{e^{i\omega_p t} + e^{-i\omega_p t}}{2} + B_p \frac{e^{i\omega_p t} - e^{-i\omega_p t}}{2i} \right] = \\ &= \frac{1}{2}A_0 + \sum_{p=1}^{\infty} \left[\frac{A_p e^{i\omega_p t}}{2} + \frac{A_p e^{-i\omega_p t}}{2} - \frac{iB_p e^{i\omega_p t}}{2} + \frac{iB_p e^{-i\omega_p t}}{2} \right] = \\ &= \frac{1}{2}A_0 + \sum_{p=1}^{\infty} e^{i\omega_p t} \frac{A_p - iB_p}{2} + \sum_{p=1}^{\infty} e^{-i\omega_p t} \frac{A_p + iB_p}{2} = \\ &= C_0^* + \sum_{p=1}^{\infty} C_p^* e^{i\omega_p t} + \sum_{p=1}^{\infty} C_{-p}^* e^{-i\omega_p t} = \sum_{p=-\infty}^{\infty} C_p^* e^{i\omega_p t}, \end{aligned}$$

1902 donde hemos definido las siguientes relaciones entre los coeficientes de
 1903 Fourier complejos y reales:

$$C_0^* = \frac{1}{2}A_0,$$

$$C_p^* = \frac{1}{2}(A_p - iB_p),$$

1905

$$C_{-p}^* = \frac{1}{2}(A_p + iB_p).$$

1906

En resumen, podemos reconstruir la serie periódica $f(t)$ con la transformada de Fourier

1907

$$f(x) = \sum_{p=-\infty}^{\infty} C_p^* e^{i\omega_p t},$$

1908

e inversamente podemos calcular los coeficientes de Fourier C_p^ a partir de la $f(t)$*

1909

$$\begin{aligned} C_p^* &= \frac{1}{2}(A_p - iB_p) = \frac{1}{2} \left(\frac{2}{T} \int_0^T f(t) \cos(\omega_p t) dt - i \frac{2}{T} \int_0^T f(t) \sin(\omega_p t) dt \right) = \\ &= \frac{1}{T} \int_0^T f(t) [\cos(\omega_p t) - i \sin(\omega_p t)] dt = \frac{1}{T} \int_0^T f(t) e^{-i\omega_p t} dt, \end{aligned}$$

1910

es decir, podemos pasar del espacio temporal $f(t)$ al espacio espectral o de Fourier C_p^ e inversamente regresar al espacio temporal de nuevo.*

1911

El teorema de Parseval es precisamente el que demuestra que el valor cuadrático medio de la serie de Fourier es igual al error cuadrático medio de los coeficientes de Fourier. La varianza de la serie de Fourier es

1912

$$\begin{aligned} \frac{1}{T} \int_0^T f^2(t) dt &= \frac{1}{T} \int_0^T \left(\frac{1}{2} A_0 + \sum_{p=1}^{\infty} [A_p \cos(\omega_p t) + B_p \sin(\omega_p t)] \right) dt = \\ &= \frac{1}{T} \frac{1}{4} A_0^2 \int_0^T dt + \frac{1}{T} A_0 \sum_{p=1}^{\infty} A_p \int_0^T \cos(\omega_p t) dt + \frac{1}{T} A_0 \sum_{p=1}^{\infty} B_p \int_0^T \sin(\omega_p t) dt + \\ &\quad + \frac{1}{T} \sum_{p=1}^{\infty} A_p^2 \int_0^T \cos^2(\omega_p t) dt + \frac{1}{T} \sum_{p=1}^{\infty} B_p^2 \int_0^T \sin^2(\omega_p t) dt + \\ &\quad + \frac{1}{T} 2 \sum_{p=1}^{\infty} A_p \sum_{p=1}^{\infty} B_p \int_0^T \sin(\omega_p t) \cos(\omega_p t) dt = \frac{1}{T} \frac{1}{4} A_0^2 T + \frac{1}{T} \sum_{p=1}^{\infty} A_p^2 \left[\frac{t}{2} + \frac{\sin(2\omega_p t)}{4\omega_p} \right]_0^T + \\ &\quad + \frac{1}{T} \sum_{p=1}^{\infty} B_p^2 \left[\frac{t}{2} - \frac{\sin(2\omega_p t)}{4\omega_p} \right]_0^T = \frac{1}{4} A_0^2 + \frac{1}{2} \sum_{p=1}^{\infty} A_p^2 + \frac{1}{2} \sum_{p=1}^{\infty} B_p^2 = \\ &= \frac{1}{4} A_0^2 + \frac{1}{2} \sum_{p=1}^{\infty} A_p^2 + B_p^2. \end{aligned}$$

1913

Utilizando las siguientes identidades

$$|C_p^*|^2 = |C_{-p}^*|^2 = \frac{1}{4}(A_p^2 + B_p^2),$$

1923

$$C_0^* = \frac{1}{2}A_0,$$

1924

el teorema de Parseval en términos de los coeficientes de Fourier complejos es

1925

$$\begin{aligned} \frac{1}{T} \int_0^T f^2(t) dt &= (C_0^*)^2 + \frac{1}{2} \sum_{p=1}^{\infty} 4|C_p^*|^2 = (C_0^*)^2 + \sum_{p=1}^{\infty} |C_p^*|^2 + \sum_{p=1}^{\infty} |C_p^*|^2 = \\ &= (C_0^*)^2 + \sum_{p=1}^{\infty} |C_p^*|^2 + \sum_{p=1}^{\infty} |C_{-p}^*|^2 = \sum_{p=-\infty}^{\infty} |C_p^*|^2 \end{aligned}$$

1926

El teorema de Parseval es una ley de conservación que demuestra que la suma de los módulos cuadrados de los coeficientes de Fourier complejos es igual al valor promedio de $f^2(x)$

$$\frac{1}{T} \int_0^T f^2(t) dt = \frac{1}{4} A_0^2 + \frac{1}{2} \sum_{p=1}^{\infty} A_p^2 + B_p^2 = \sum_{p=-\infty}^{\infty} |C_p^*|^2.$$

1927 *Esto da lugar a la relación entre la amplitud de las componentes de*
 1928 *Fourier en el dominio espectral (de frecuencia) y la varianza de la serie*
 1929 *en el dominio temporal. \ \{NOTA:\} En el análisis de Fourier es im-*
 1930 *portante recalcar que debemos de eliminar la tendencia de la serie antes*
 1931 *de calcular los coeficientes. Sino lo hacemos, el análisis de Fourier pon-*
 1932 *drá erróneamente la varianza de la tendencia en las componentes de*
 1933 *baja frecuencia de la expansión de Fourier. En Matlab eso lo pode-*
 1934 *mos hacer con el comando detrend.m o bien simplemente extrayendo*
 1935 *el promedio temporal.*

1936 *{Ejemplo de cálculo de coeficientes de Fourier}*

1937 *Imaginemos la siguiente onda cuadrada representada por la función*

$$f(t) = \begin{cases} -1, & \text{para } -\frac{1}{2}T \leq t < 0 \\ +1, & \text{para } 0 \leq t < \frac{1}{2}T \end{cases}$$

1938

1939 *Onda cuadrada.*

1940 *Puesto que función de arriba es impar, es decir cumple la condición de*
 1941 *simetría $-f(t) = f(-t)$, entonces la serie de Fourier resultante sola-*
 1942 *mente contendrá componentes sinusoidales. Entonces*

$$\begin{aligned} B_p &= \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin(\omega_p t) dt = \\ &= \frac{2}{T} \int_{-T/2}^0 f(t) \sin(\omega_p t) dt + \frac{2}{T} \int_0^{T/2} f(t) \sin(\omega_p t) dt = \\ &= \frac{2}{T} \left[(-1) \int_{-T/2}^0 \sin(\omega_p t) dt + (1) \int_0^{T/2} \sin(\omega_p t) dt \right] = \\ &= \frac{2}{T} \left[\left| \frac{\cos(\omega_p t)}{\omega_p} \right|_{-T/2}^0 - \left| \frac{\cos(\omega_p t)}{\omega_p} \right|_0^{T/2} \right] = \end{aligned}$$

1943

1945

$$= \frac{2}{\omega_p T} [1 - \cos(\omega_p T/2) - \cos(\omega_p T/2) + 1] = \frac{2}{\omega_p T} [2 - 2\cos(\omega_p T/2)] =$$

1946

$$= \frac{4}{\omega_p T} [1 - \cos(\omega_p T/2)] = \frac{2}{\pi p} [1 - \cos(\pi p)]$$

1947 | | Los coeficientes son cero ($B_p = 0$) si p es par y $B_p = 4/\pi p$ si p es
1948 impar. Finalmente nuestra serie de Fourier es

$$f(t) = \sum_{p=1}^{\infty} B_p \sin(\omega_p t) =$$

1949

$$= \sum_{p=1}^{\infty} B_p \sin\left(\frac{2\pi p}{T}t\right) = B_1 \sin\left(\frac{2\pi 1}{T}t\right) + B_3 \sin\left(\frac{2\pi 3}{T}t\right) +$$

1950

$$+ B_5 \sin\left(\frac{2\pi 5}{T}t\right) + \dots = \frac{4}{\pi} \sin\left(\frac{2\pi 1}{T}t\right) +$$

1951

$$+ \frac{4}{3\pi} \sin\left(\frac{2\pi 3}{T}t\right) + \frac{4}{5\pi} \sin\left(\frac{2\pi 5}{T}t\right) =$$

1952

$$= \frac{4}{\pi} \left(\frac{\sin \omega_1 t}{1} + \frac{\sin 3\omega_1 t}{3} + \frac{\sin 5\omega_1 t}{5} \right) + \dots$$

**Código Matlab para $N=100$ componentes*

```
N=100;
t=-30:0.1:30;
f=0;
for p=1:2:N;
f=f+(4/(p*pi))*sin(((2*pi*p)/(10))*t);
end
figure; plot(t,f);
```

1953

1954 Serie de Fourier para Onda cuadrada. Arriba para 5 componentes, medio
1955 para 50 componentes, y abajo para 100 componentes.

1956 { Series de Fourier Discretas}

1957 En general, vamos a muestrear de forma discreta el océano y conse-
1958 cuentemente las series temporales que obtenemos son discretas en el
1959 tiempo. Segun el teorema de Parseval, la varianza de estas series dis-
1960 cretas

$$\sigma^2 = \frac{1}{N-1} \sum_{t=1}^N (f(t) - \bar{f(t)})^2$$

1961 se puede obtener sumando las contribuciones individuales de los armóni-
1962 cos de Fourier. La descomposición de series temporales discretas en
1963 armónicos específicos da lugar al concepto de espectro de Fourier. Para

1964 encontrar el espectro de Fourier debemos calcular los coeficientes A_p, B_p
 1965 o, equivalentemente, las amplitudes C_p y el ángulo de fase θ_p .

1966 Supongamos la serie de Fourier para un registro finito de longitud par N
 1967 definido en los tiempos t_1, t_2, \dots, t_N

$$f(t_n) = \frac{1}{2}A_0 + \sum_{p=1}^{N/2} [A_p \cos(\omega_p t_n) + B_p \sin(\omega_p t_n)],$$

1968 donde ya sabemos $\omega_p = 2\pi f_p = 2\pi p/T$. Sabiendo que $t_n = n\Delta t$, esta
 1969 serie se puede reescribir como

$$\begin{aligned} f(t_n) &= \frac{1}{2}A_0 + \sum_{p=1}^{N/2} [A_p \cos(2\pi pn\Delta t/\Delta t N) + B_p \sin(2\pi pn\Delta t/\Delta t N)] = \\ &= \frac{1}{2}A_0 + \sum_{p=1}^{N/2} [A_p \cos(2\pi pn/N) + B_p \sin(2\pi pn/N)] = \frac{1}{2}C_0 + \sum_{p=1}^{N/2} [C_p \cos((2\pi pn/N) - \theta_p)], \end{aligned}$$

1970 donde los términos $A_0/2$ y $C_0/2$ son los valores medios de toda la serie
 1971 $f(t)$. Los coeficientes se calculan de igual forma usando las condiciones
 1972 de ortogonalidad. La única diferencia es que en lugar de tratar con inte-
 1973 grales (serie continua) tratamos con sumatorios (serie discreta)

$$A_p = \frac{2}{N} \sum_{n=1}^N f(t_n) \cos(2\pi pn/N), \quad p = 1, 2, \dots, N/2$$

$$A_0 = \frac{2}{N} \sum_{n=1}^N f(t_n), \quad B_0 = 0$$

$$A_{N/2} = \frac{1}{N} \sum_{n=1}^N f(t_n) \cos(n\pi), \quad B_{N/2} = 0$$

$$B_p = \frac{2}{N} \sum_{n=1}^N f(t_n) \sin(2\pi pn/N), \quad p = 1, 2, \dots, N/2$$

1978 El $N/2$ se debe a que es el armónico con mayor frecuencia que pode-
 1979 mos resolver, es decir, aquel que oscila con frecuencia de Nyquist. Para
 1980 $p > N/2$ las funciones trigonométricas simplemente darán coeficientes
 1981 de Fourier repetidos ya obtenidos en el intervalo $1 \leq p \leq N/2$. Para
 1982 calcular la serie discreta de Fourier, primero debemos calcular los ar-
 1983 gumentos de las funciones trigonométricas $2\pi np/N$ para cada entero p
 1984 y n . Segundo, evaluamos las funciones $\cos(2\pi np/N)$ y $\sin(2\pi np/N)$,
 1985 y sumamos para los términos $f(t_n) \cos(2\pi np/N)$ y $f(t_n) \sin(2\pi np/N)$.
 1986 Por último, incrementamos p y repetimos los dos pasos anteriores.

1987 $\{Ejemplo\ de\ Series\ Temporales\ Discretas\ (modificado\ de\ Emery\ and\ Thompson,\ p387)\}$

1989 Considera la serie temporal de temperatura promedia mensual por un
 1990 periodo de tres años (ver tabla y Figura).

1991	Temperatura ($^{\circ}$ C)	9.6	9.3	9.8	10.3	10.7	11.3	11.5	12	12	9.8	8.1	7.3
	Año 2	7.6	8.7	10.5	12	14	14.3	15.8	16.2	15.4	12.9	11.9	8.5
	Año 3	9.1	8.9	10	9.9	10.1	11.3	11.5	12	12	14	14.3	14.8

1992
 1993
 1994 Utilizando las expresiones de arriba podemos calcular las frecuencias
 1995 f_p , amplitudes A_p , B_p , C_p , las fases θ_p y finalmente la serie de Fourier
 1996 $f(t)$. Los valores para las primeras 8 componentes estan reflejados en la
 1997 tabla

p	f (cpm)	T (mes)	A_p ($^{\circ}$ C)	B_p ($^{\circ}$ C)	C_p ($^{\circ}$ C)	θ_p ($^{\circ}$)
0	0	-	22.68	0	22.68	0
1	0.0278	36.0	-0.263	-0.61	0.67	-113
2	0.0556	18.0	1.95	0.55	2.03	15.76
3	0.1111	12.0	-1.16	-1.69	2.05	-124.48
4	0.1389	9.0	0.91	-0.25	0.95	-15.33
5	0.1667	7.2	0.59	-0.31	0.67	-27.53
6	0.1944	6	-0.11	-0.28	0.30	-112.69
7	0.2222	5.14	0.39	-0.3	0.49	-36.83
8	0.2500	4.5	0.34	-0.31	0.46	-42.78

1998
 1999 Serie temporal de temperatura promedia mensual por tres años de muestreo
 2000 (línea azul) y la reconstrucción con la serie de Fourier de 5 componentes
 2001 (línea roja).
 2002

2003 { Serie de Fourier para variables vectoriales (complejas) }

2004 En este caso la transformada de Fourier se aplica a una cantidad vec-
 2005 torial en lugar de una cantidad escalar como temperatura, salinidad,
 2006 densidad, etc. Supongamos que tenemos las dos componentes de la ve-
 2007 locidad u y v las cuales expandemos en series de Fourier

$$2008 u(t) = \frac{1}{2}A_0 + \sum_{p=1}^{N/2} [A_p \cos(\omega_p t_n) + B_p \sin(\omega_p t_n)]$$

$$v(t) = \frac{1}{2}C_0 + \sum_{p=1}^{N/2} [C_p \cos(\omega_p t_n) + D_p \sin(\omega_p t_n)],$$

2009 lo cual se puede escribir en versión compleja como

$$2010 R(t) = u(t) + iv(t) = \frac{1}{2}A_0 + \sum_{p=1}^{N/2} [A_p \cos(\omega_p t_n) + B_p \sin(\omega_p t_n)] + \\ + i \left(\frac{1}{2}C_0 + \sum_{p=1}^{N/2} [C_p \cos(\omega_p t_n) + D_p \sin(\omega_p t_n)] \right) =$$

2011

$$= \left[\frac{1}{2}A_0 + i\frac{1}{2}C_0 \right] + \sum_{p=1}^{N/2} [(A_p + iC_p)\cos(\omega_p t_n) + (B_p + iD_p)\sin(\omega_p t_n)] ,$$

2012

donde $\frac{1}{2}A_0 + i\frac{1}{2}C_0 = \overline{u(t)} + i\overline{v(t)}$ es la velocidad media, $\omega_p = 2\pi f_p = 2\pi p/N\Delta t$ la frecuencia angular, $t_n = n\Delta t$ es el eje de tiempo, y (A_p, B_p, C_p, D_p) y son las amplitudes y fases de cada componente de Fourier, tanto las reales como las imaginarias. A diferencia de la serie de Fourier real en este caso las componentes van de $p = 1$ hasta $p = N$ y, por lo tanto, estamos cubriendo ambas frecuencias {positivas} y {negativas}. Si extraemos la velocidad media

2013

2014

2015

2016

2017

2018

$$R'(t) = R(t) - [\overline{u(t)} + i\overline{v(t)}] = \sum_{p=1}^{N/2} [(A_p + iC_p)\cos(\omega_p t_n) + (B_p + iD_p)\sin(\omega_p t_n)] .$$

2019

2020

2021

2022

Ahora vamos a escribir la anomalía de la serie compleja $R'(t)$ en términos de dos componentes rotatorias ortogonales, es decir, una componente que gira en el sentido de las agujas del reloj con amplitud R^- y otra que gira en el sentido opuesto a las agujas del reloj y amplitud R^+

2023

$$\begin{aligned} R'(t) &= \sum_{p=1}^{N/2} [e^{i\omega_p t} + e^{-i\omega_p t}] = \\ &= \sum_{p=1}^{N/2} R_p^+ [\cos(\omega_p t_n) + i\sin(\omega_p t_n)] + \sum_{p=1}^{N/2} R_p^- [\cos(\omega_p t_n) - i\sin(\omega_p t_n)] = \\ &= \sum_{p=1}^{N/2} [(R_p^+ + R_p^-)\cos(\omega_p t_n) + (R_p^+ - R_p^-)i\sin(\omega_p t_n)] \end{aligned}$$

2024

2025

2026

2027

2028

Note que $e^{i\omega_p t} = \cos(\omega_p t_n) + i\sin(\omega_p t_n)$ rota en el sentido contrario a las agujas del reloj y $e^{-i\omega_p t} = \cos(\omega_p t_n) - i\sin(\omega_p t_n)$ rota en el sentido de las agujas del reloj. Si comparamos las dos expresiones obtenemos las siguientes identidades

2029

$$\begin{aligned} A_p + iC_p &= R_p^+ + R_p^- \\ B_p + iD_p &= (R_p^+ - R_p^-)i \end{aligned}$$

2030

y de ahí obtenemos que

2031

$$\begin{aligned} R_p^+ &= \frac{1}{2} [A_p + D_p + i(C_p - B_p)] \\ R_p^- &= \frac{1}{2} [A_p - D_p + i(C_p + B_p)] , \end{aligned}$$

2032

y las magnitudes de las componentes rotatorias es

2033

$$\begin{aligned} |R_p^+| &= \frac{1}{2} [(A_p + D_p)^2 + (C_p - B_p)^2]^{1/2} \\ |R_p^-| &= \frac{1}{2} [(A_p - D_p)^2 + (C_p + B_p)^2]^{1/2} , \end{aligned}$$

2034 *y las fases de las componentes rotatorias son*

$$\epsilon_p^+ = \operatorname{actan} \left(\frac{C_p - B_p}{A_p + D_p} \right),$$

$$\epsilon_p^- = \operatorname{actan} \left(\frac{C_p + B_p}{A_p - D_p} \right).$$

2035
2036 *La rotación de las componentes {clockwise} y {anticlockwise} dibujan*
2037 *una elipse en el plano u vs v. Puesto que ambas componentes rotan en*
2038 *sentido contrario pero con la misma frecuencia, habrán momentos que*
2039 *ambás apuntaran en la misma dirección (aditivas), otras ocasiones en*
2040 *dirección opuesta (cancelativas). Esos tiempos de adición y cancelación*
2041 *definen el eje mayor de la elipse $L_E = R_p^+ + R_p^-$ y el eje menor de la*
2042 *elipse $L_e = R_p^+ - R_p^-$. La orientación (inclinación) y la fase de estas*
2043 *elipses a $t = 0$ es*

$$\theta_e = \frac{1}{2}(\epsilon_p^+ + \epsilon_p^-),$$

$$\phi_e = \frac{1}{2}(\epsilon_p^+ - \epsilon_p^-).$$

2044
2045 *Una propiedad interesante es el coeficiente rotatorio*

$$r(\omega) = \frac{R_p^+ - R_p^-}{R_p^+ + R_p^-},$$

2046 *que toma valores entre 0 y 1. Para $r = -1$ tenemos movimiento en*
2047 *el sentido de las agujas del reloj, para $r = 0$ tenemos un flujo unidirec-*
2048 *cional, y para $r = +1$ tenemos movimiento en el sentido contrario de las*
2049 *agujas del reloj.*

2050 *{ Transformada Rápida de Fourier y espectros de potencia}*

2051 *La FFT (por sus siglas en inglés) es un algoritmo para calcular la se-*
2052 *rie de Fourier discreta de forma mas eficiente computacionalmente*
2053 *hablando. En este caso la FFT se debería aplicar a series temporales*
2054 *con longitudes múltiples de 2. En caso contrario, es útil llenar de*
2055 *ceros nuestra serie para obtener longitudes múltiplos de 2. A eso se le*
2056 *llama ‘padding’. Básicamente, el algoritmo obtiene los coeficientes de la*
2057 *serie discreta de Fourier*

$$f(t) = \sum_{p=-\infty}^{\infty} C_p^* e^{i\omega_p t_n} ; \quad \omega_p = 2\pi p f_0 = 2\pi p/T,$$

$$F(p) = C_p^* = \frac{1}{2}(A_p - iB_p) = \frac{1}{N} \sum_{n=0}^{N-1} f(t_n) e^{-i\omega_p t_n} = \frac{1}{N} \sum_{n=0}^{N-1} f(t_n) [\cos(\omega_p t_n) - i \sin(\omega_p t_n)].$$

2058
2059 *La parte real de la FFT me da las amplitudes A_p y la parte imaginaria*
2060 *me da las amplitudes $-B_p$*

$$Re[F(p)] = A_p = \frac{2}{N} \sum_{n=0}^{N-1} f(t_n) \cos(2\pi p n / N), \quad p = 0, 1, \dots, N/2$$

$$Im[F(p)] = -B_p = -\frac{2}{N} \sum_{n=0}^{N-1} f(t_n) \sin(2\pi pn/N), \quad p = 0, 1, \dots, N/2$$

2061 *En general, debemos de normalizar las amplitudes A_p y B_p por la longi-
2062 tud del registro N . Así que en Matlab la amplitud de la FFT es*

$$\text{abs}(\text{fft}(f(t)))/N,$$

2063 *y la potencia de la FFT es*

$$\text{abs}(\text{fft}(f(t)))^2/N^2.$$

2064 *La FFT ha descompuesto una señal de N elementos, $f(t)$, en un con-
2065 junto de $N/2 + 1$ ondas cosinusoidales y $N/2 + 1$ ondas sinusoidales,
2066 con las frecuencias definidas por el índice $p = 0, 1, \dots, N/2$, i.e. $\$_p =$
2067 $2f_p = 2p/T = 2p/N$ t \$. Las amplitudes de los cosenos estan
2068 contenidas en $\text{Re}[F(p)]$ y las amplitudes de los senos en $\text{Im}[F(p)]$.
2069 Note que las frecuencias son siempre {positivas}, es decir, los índices
2070 k siempre van de cero a $N/2$. Las frecuencias entre $N/2$ y $N - 1$ son
2071 {negativas}. Recuerda que el espectro frecuencial de una señal discreta
2072 es periódico, y entonces las frecuencias son negativas entre $N/2$ y $N - 1$
2073 al igual que en el intervalo $-N/2$ y -1 . Los puntos 0 y $N/2$ separan
2074 las frecuencias negativas de las positivas. Es por ello que, generalmente,
2075 solamente centramos nuestra atención en la parte positiva del espectro.
2076 La magnitud (o norma) de la transformada de Fourier discreta es*

$$\text{Magnitud} = |F(p)| = \sqrt{\text{Re}[F(p)]^2 + \text{Im}[F(p)]^2},$$

2077 *y la fase es*

$$\text{Phase} = \tan^{-1} \left(\frac{\text{Im}[F(p)]}{\text{Re}[F(p)]} \right).$$

2078 *La FFT organiza los coeficientes de Fourier (imaginarios y reales) en
2079 frecuencias {negativas} y {positivas} y reparte la varianza de la señal
2080 equitativamente entre ellas. En $p = 0$ tenemos la media de la serie
2081 temporal, aunque debido a que hemos eliminado la media y la tenden-
2082 cia de la serie temporal no debemos de preocuparnos por ella. Entre
2083 $p = 1, \dots, N/2$ tenemos los valores de los coeficientes de Fourier reales
2084 y entre $p = N/2 + 1, \dots, N - 1$ tenemos los complejos conjugados de los
2085 primeros $N/2$ coeficientes. Si calculamos el valor absoluto de la trans-
2086 formada de Fourier (en Matlab $\text{abs}(\text{fft}(f(t)))/N$) estamos calculando
2087 $A_p^2 + B_p^2$ y si solo nos quedamos con los primeros $N/2$ elementos de
2088 la FFT, debemos de multiplicar por un factor de 2 para conservar la
2089 energía espectral.*

2090 *{ Estimaciones espectrales o autoespectros }*

2091 *{(1) Espectro de amplitud}*

2092 *La gráfica de la magnitud de los coeficientes complejos $|C_p^*|$ de la serie
2093 de Fourier*

$$f(t) = \sum_{p=-\infty}^{\infty} C_p^* e^{i\omega_p t_n}$$

2094 *frente a (versus) la frecuencia ω_p se denomina espectro de amplitud de
2095 la función periódica $f(t)$. En Matlab (para las primeras $N/2$ compo-
2096 nentes),*

$$\text{Amplitud} = 2 * |C_p^*| = \text{abs}(\text{fft}(f(t)))/N.$$

2097 $\{(2) \text{ Espectro de densidad de potencia (Power Spectral Density, PSD)}\}$

2098 *El espectro de densidad de potencia (PSD, por sus siglas en inglés) es la
2099 potencia de la FFT por unidad de frecuencia*

$$\text{PSD}(p) = 2 * |C_p^*|^2 / \Delta f$$

2100 *donde $\Delta f = 1/N\Delta t$ es la frecuencia fundamental.*

2101 *La gráfica de PSD(p) de la frente a (versus) la frecuencia ω_p se denom-
2102 ina espectro de densidad de potencia de la función periódica $f(t)$. Si
2103 solo nos quedamos con las $N/2$ primeras componentes en Matlab se
2104 escribe*

$$\text{PSD}(p) = 2 * \text{abs}(\text{fft}(f(t)))^2 / N^2 / \Delta f.$$

2105 *Esta normalización tiene su fundamento en el cumplimiento del teorema
2106 de Parseval, de tal forma que la energía total de la señal en el dominio
2107 temporal $f(t)$ (por unidad de tiempo) sea igual a la energía total de la
2108 señal en el dominio frecuencial definido por C_p^* :*

$$\frac{1}{T} \sum_{n=1}^N |f(t_n)|^2 \Delta t = \frac{1}{N} \sum_{n=1}^N |f(t_n)|^2 = \text{var}(f(t)) = \sum_{p=-\infty}^{\infty} |C_p^*|^2 = \sum_{p=0}^{N/2} \text{PSD}(p) * \Delta f.$$

2109 $\mid \mid$ *Este teorema de conservación de energía nos informa de que la inte-
2110 gral bajo la curva espectral PSD(p) debe ser igual a la varianza total de
2111 la serie temporal.*

2112 $\{(Efectos de los extremos en estimaciones espectrales)\}$

2113 *En general, para calcular un espectro promedio debemos fragmentar
2114 nuestra serie temporal en bloques de igual tamaño que contengan las
2115 frecuencias de interés, realizar espectros individuales de dichos fragmen-
2116 tos, y promediar todos ellos. Este método se le conoce como Welch. Los
2117 fragmentos pueden ser únicos, es decir, sin superposición o bien pueden
2118 ser recursivos, es decir, cuando utilizamos superposición de fragmentos.
2119 Por ejemplo, una superposición del 50% significa que cada fragmento
2120 empieza en la mitad del fragmento anterior. Además de eliminar el
2121 ruido, el método de Welch también reduce la transferencia de energía
2122 de las frecuencias pico hacia frecuencias colindantes ('leakage' en*

2123 *inglés). El método de Welch reduce el ruido causado por el*
 2124 *uso de datos imperfectos y por el efectoleakage'. Aquí les mue-*
 2125 *stro un ejemplo de como promediar un espectro de densidad espectral*
 2126 *con fregmentos de tamaño M:*

Matlab

Ejemplo del método de pWelch para calcular espectros de frecuencia promediados o suavizados de series temporales o espaciales, u(t) o u(x), de tamaño N.

% Suavizado con pwelch (a mano y con subrutina de librería MATLAB)

M=500;

M2=floor(M/2);

nintervals=N/M;

PWuu=zeros(M2+1,1);

s=0;

for k=0:M:N-M

s=s+1;

uu=[];

%seleccionamos la serie sobre la ventana M

uu=u(k+1:k+M);

%Frecuencia de muestreo

*dfM=1./(M*dt);*

%PSD

spctuu0=[];

spctuu=abs(fft(uu)).^2/M.^2/dfM;

*PSD=[spctuu(1),2*spctuu(2:M2),spctuu(M2+1)];*

PWuu = PWuu + PSD(:);

end

PWuu=PWuu/nintervals;

*fPWuu=[0:M2]*dfM;*

%

% Librería MATLAB

%

*%Fs=M*dfM;*

%[PWuuMatlab,FMatlab]=pwelch(u(:,M,0,[]),Fs,'psd','oneside');

2127 *Este problema de transferencia de energía o 'leakage' es intrínseco al*
 2128 *problema de que las series temporales oceanográficas son finitas y, por lo*
 2129 *tanto, no son necesariamente periódicas, condición necesaria en el análi-*
 2130 *sis de Fourier. Veamos esto con un ejemplo de una onda cosinusoidal*
 2131 *periódica y no-periódica.*

2132

2133 *Podemos observar como en el caso de la serie no-periódica existe una*
 2134 *transferencia de energía del pico espectral hacia las frecuencias colin-*
 2135 *dantes de forma que se reduce la amplitud del pico de interés.*

2136 $\{Correlación\}$ **RECORDATORIO:**

La función de covarianza cruzada se debe definir en función de un desfase τ como

$$C_{xy}(\tau) = \frac{1}{N-m} \sum_{n=0}^{N-m} x(n\Delta t)y(n\Delta t + \tau),$$

donde $m = 0, 1, \dots, M$ es el número de desfases $\tau_m = m\Delta t$ y $M < N$. Su correspondiente correlación cruzada es

$$\rho_{xy}(\tau) = \frac{C_{xy}}{C_{xx}(0)C_{yy}(0)} = \frac{C_{xy}}{s_x s_y},$$

donde s_x y s_y son las varianzas de las variables x e y , respectivamente.

2137 *Kundu (1976b) define la función de correlación cruzada desfasada entre
2138 dos series de velocidad en las profundidades 1 y 2 como*

$$\rho_\tau = \frac{\overline{u'_1(t)u'_2(t-\tau)}}{\left[\overline{u'_1(t)^2} \overline{u'_2(t)^2}\right]^{1/2}},$$

2139 *donde las primas' indican anomalías. Esta función fue utilizada para
2140 estudiar la propagación vertical de ondas inercio-gravitatorias y la ve-
2141 locidad de fase de estas ondas $c = \Delta_{12}/\tau$. Kundu (1976a) introduce el
2142 coeficiente de correlación complejo*

$$\rho = \frac{\overline{w_1^*(t)w_2(t)}}{\left[\overline{w_1^*(t)w_1^*(t)} \overline{w_2^*(t)w_2(t)}\right]^{1/2}},$$

2143 *donde $w = u + iv$, los asteriscos * indican complejo conjugados, y los
2144 subíndices 1 y 2 se refieren a dos estaciones de medida. La cantidad
2145 ρ es un número complejo cuya magnitud (≤ 1) nos da una medida de
2146 correlación promedio y cuyo ángulo de fase da el angulo promedio, me-
2147 dido en el sentido contrario a las agujas del reloj, del segundo vector
2148 con respecto del primero. Por ejemplo, un ángulo de fase negativo en-
2149 tre las profundidades 50 y 100 m implica que la señal llega primero a
2150 $z = -100$ m y luego a $z = -50$ m, es decir, podría tratarse de una onda
2151 interna cuyas fases se propagan hacia arriba.*2152 $\{(3)\ Espectro\ cruzado\}$ 2153 *Con análisis de espectros cruzados pretendemos comprender la relación
2154 entre dos series temporales en función de la frecuencia. Por ejemplo,
2155 observamos en dos localizaciones espectros con picos en las mismas fre-
2156 cuencias y queremos saber si dichos armónicos están relacionados.*

2157 Supongamos dos series de Fourier $x(t)$ e $y(t)$

$$x(t) = \bar{x} + \sum_{p=1}^{N/2} A_{xp} \cos(\omega_p t_n) + B_{xp} \sin(\omega_p t_n),$$

2158

$$y(t) = \bar{y} + \sum_{p=1}^{N/2} A_{yp} \cos(\omega_p t_n) + B_{yp} \sin(\omega_p t_n).$$

2159

2160 Utilizando las condiciones ortogonalidad entre las funciones sinusoidales y cosinusoidal, la covarianza entre las variables x e y es

$$\overline{x'y'} = \sum_{p=1}^{N/2} \frac{1}{2} (A_{xp} A_{yp} + B_{xp} B_{yp}) = \sum_{p=1}^{N/2} Co(p),$$

2161

donde $Co(p)$ es el co-espectro de x e y .

2162

2163 Supongamos dos series de Fourier $x(t)$ e $y(t)$ definidas en la forma compleja (el asterisco ha sido eliminado en esta notación)

$$x(t) = \bar{x} + \sum_{p=1}^{N/2} C_{xp} e^{i\omega_p t_n} = \bar{x} + \sum_{p=1}^{N/2} \frac{1}{2} (A_{xp} - iB_{xp}) e^{i\omega_p t_n} = \bar{x} + \sum_{p=1}^{N/2} F_x(p)$$

2164

$$y(t) = \bar{y} + \sum_{p=1}^{N/2} C_{yp} e^{i\omega_p t_n} = \bar{y} + \sum_{p=1}^{N/2} \frac{1}{2} (A_{yp} - iB_{yp}) e^{i\omega_p t_n} = \bar{y} + \sum_{p=1}^{N/2} F_y(p).$$

2165

Si ahora calculamos las varianzas

$$\overline{x'^2} = \sum_{p=-N/2}^{N/2} F_{xx}(p),$$

2166

donde

$$F_{xx}(p) = 2 \frac{1}{2} (A_{xp} - iB_{xp}) e^{i\omega_p t_n} \frac{1}{2} (A_{xp} + iB_{xp}) e^{-i\omega_p t_n} = 2F_x(p)F_x^*(p) = |C_{xp}|^2,$$

2167

2168 y el asterisco indica complejo conjugado. Para la variable y de igual forma obtenemos:

$$\overline{y'^2} = \sum_{p=-N/2}^{N/2} F_{yy}(p); \quad F_{yy}(p) = 2F_y(p)F_y^*(p) = |C_{yp}|^2$$

2169

2170 De las expresiones anteriores se deduce que covarianza se puede calcular en el espacio espectral como

$$\overline{x'y'} = Re \left[\sum_{p=-N/2}^{N/2} F_{xy}(p) \right],$$

2171

donde

$$F_{xy}(p) = 2 \frac{1}{2} (A_{xp} - iB_{xp}) e^{i\omega_p t_n} \frac{1}{2} (A_{yp} + iB_{yp}) e^{-i\omega_p t_n} = 2F_x(p)F_y^*(p) = |C_{xp}||C_{yp}|e^{i(\theta_{xp}-\theta_{yp})}.$$

²¹⁷² *El factor $e^{i(\theta_{xp}-\theta_{yp})}$ aparece para considerar que ambas series periódicas*
²¹⁷³ *no están en fase.*

²¹⁷⁴ *{(4) Espectro cruzado complejo}*

²¹⁷⁵ *Si escribimos $F_{xy}(p)$ en términos de los coeficientes de Fourier reales*

$$\begin{aligned} F_x(p)F_y^*(p) &= \frac{1}{2} (A_{xp} - iB_{xp}) e^{i\omega_p t_n} \frac{1}{2} (A_{yp} + iB_{yp}) e^{-i\omega_p t_n} = \\ &= \frac{1}{4} [A_{xp}A_{yp} + B_{xp}B_{yp} + i(A_{xp}B_{yp} - A_{yp}B_{xp})] . \end{aligned}$$

²¹⁷⁷ *Para el caso de series $x(t)$ e $y(t)$ reales sabemos que las frecuencias*
²¹⁷⁸ *negativas son los complejos conjugados de las frecuencias positivas y*
²¹⁷⁹ *entonces*

$$A_k = A_{-k} \quad y \quad B_k = B_{-k},$$

²¹⁸⁰ *y*

$$F_x(p)F_y^*(p) = F_x(-p)F_y^*(-p),$$

²¹⁸¹ *y como conclusión*

$$F_{xy}(p) + F_{xy}(-p) = \frac{1}{2} [A_{xp}A_{yp} + B_{xp}B_{yp} + i(A_{xp}B_{yp} - A_{yp}B_{xp})],$$

²¹⁸² *que es espectro cruzado de x e y para el armónico p . De esta expresión*
²¹⁸³ *encontramos que*

$$F_{xy}(p) + F_{xy}(-p) = 2F_{xy}(p) = Co(p) + iQ(p),$$

²¹⁸⁴ *donde $Co(p) = \frac{1}{2}(A_{xp}A_{yp} + B_{xp}B_{yp})$ es el co-espectro del armónico p y*
²¹⁸⁵ *$Q(p) = \frac{1}{2}(A_{xp}B_{yp} - A_{yp}B_{xp})$ es el espectro de cuadratura del armónico*
²¹⁸⁶ *k .*

²¹⁸⁷ *En notación compleja el espectro cruzado*

$$F_{xy}(p) = C_{xp}C_{yp}e^{i(\theta_{xp}-\theta_{yp})} = C_{xp}C_{yp} (\cos(\theta_{xp} - \theta_{yp}) + i\sin(\theta_{xp} - \theta_{yp})).$$

²¹⁸⁸ $\theta_{xp} = \theta_{yp}$ *entonces $F_{xy}(p)$ es real*

$\theta_{xp} \neq \theta_{yp} = \pm \frac{\pi}{2}$ *entonces $F_{xy}(p)$ es complejo*

²¹⁸⁹ *Entonces el co-espectro (la parte real del espectro cruzado) está en fase*
²¹⁹⁰ *con la señal y el espectro de cuadratura está totalmente desfasado.*

²¹⁹¹ *{(5) Espectro de coherencia}*

²¹⁹² *Para una única componente p , el espectro de coherencia al cuadrado*
²¹⁹³ *entre dos series x e y se define*

$$Coh^2(p) = \frac{|F_{xy}(p)|^2}{F_{xx}F_{yy}} = \frac{|C_{xp}C_{yp}|^2}{C_{xp}^2C_{yp}^2},$$

2194 donde $|Coh^2(p)|^{1/2}$ es su magnitud y $\phi_{xy}(p)$ es el ángulo de desfase entre las dos componentes p de x e y .
 2195

2196 El espectro de coherencia al cuadrado nos indica el grado de correlación
 2197 existente entre dos señales. Dos señales están altamente correlacionadas
 2198 si la magnitud del espectro de coherencia al cuadrado es $\simeq 1$ y su fase es
 2199 $\phi_{xy}(p) \simeq 0$.

2200 **10 Métodos de filtrado y suavizado**

2201 **10.1 Convolución y funciones respuesta (ventanas {espectrales})**

2202 La convolución de dos funciones $f(t)$ y $g(t)$ sobre un registro finito
 2203 $[0, T]$ se define como

$$[f * g](t) = \frac{1}{T} \int_0^T f(\tau)g(t - \tau)d\tau.$$

2204 O también se puede expresar sobre un registro infinito como

$$[f * g](t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau = \int_{-\infty}^{\infty} g(\tau)f(t - \tau)d\tau.$$

2205 La convolución satisface las siguientes propiedades

$$f * g = g * f$$

$$f * (g * h) = (f * g) * h$$

$$f * (g + h) = (f * g) + (f * h)$$

2208 Ahora retomemos las definiciones de serie de Fourier y la transformada
 2209 de Fourier:

$$f(t) = F^{-1}[F(\omega)](t) = \sum_{p=-\infty}^{\infty} F(\omega_p)e^{i\omega_p t}; \quad \text{serie de Fourier}$$

y

$$F(\omega_p) = C_p^* = \frac{1}{T} \int_0^T f(t)e^{-i\omega_p t}dt; \quad \text{Transformada de Fourier}$$

También se pueden escribir alternativamente en forma integral (continua) como:

$$f(t) = \int_{-\infty}^{\infty} F(\omega)e^{i\omega t}d\omega,$$

y

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt$$

2210 Vamos ahora a deducir el teorema de la convolución. Para ello vamos a
 2211 partir de la definición de convolución:

$$\begin{aligned} f * g &= \frac{1}{T} \int_0^T f(\tau)g(t-\tau)d\tau = \frac{1}{T} \int_0^T f(\tau) \sum_{p=-\infty}^{\infty} G(\omega_p)e^{i\omega_p(t-\tau)}d\tau = \\ 2212 &= \sum_{p=-\infty}^{\infty} G(\omega_p) \left[\frac{1}{T} \int_0^T f(\tau)e^{-i\omega_p\tau}d\tau \right] e^{i\omega_p t} = \sum_{p=-\infty}^{\infty} G(\omega_p)F(\omega_p)e^{i\omega_p t} = F^{-1}[G(\omega)F(\omega)](t). \end{aligned}$$

2213 Si aplicamos transformada de Fourier a ambos lados del igual obten-
 2214 emos:

$$F(f * g) = G(\omega_p)F(\omega_p) = F[g(t)]F[f(t)],$$

2215 es decir, la transformada de Fourier de la convolución de f y g es equiv-
 2216 alente a multiplicar en el espacio espectral las transformadas de Fourier
 2217 de las funciones individuales. La correlación cruzada desfasada de $f(t)$
 2218 y $g(t)$ en forma integral se puede definir como

$$C_{fg}(\tau) = \frac{1}{T} \int_0^T f(\tau)g(t+\tau)d\tau = F^{-1}[G(\omega)F(-\omega)](t) = F^{-1}[G(\omega)F^*(\omega)](t),$$

2219 es decir, si multiplicamos la transformada de Fourier de una función
 2220 por el complejo conjugado de la transformada de Fourier de otra función
 2221 es equivalente a la transformada de Fourier de la correlación cruzada
 2222 desfasada entre ellas. Este se le conoce por el teorema de correlación.
 2223 Para el caso particular que sea la misma función $g(t)$ la que se corre-
 2224 ciona, entonces:

$$F[C_{gg}(\tau)] = G(\omega)G^*(\omega) = |G(\omega)|^2,$$

2225 es decir, la transformada de Fourier de la autocorrelación es igual al
 2226 espectro de potencia de la función $g(t)$. Este se denomina el teorema de
 2227 Weiner-Khinchin.

2228 El concepto de convolución es útil en el filtrado de señales periódicas.
 2229 En general vamos a convolucionar nuestra señal $f(t)$ con la denominada
 2230 función respuesta $r(t)$. La función $r(t)$ es típicamente una función pico
 2231 que cae a cero en ambas direcciones desde el máximo (o pico).

2232

2233 Puesto que la función respuesta es mas ancha que algunas estructuras
 2234 de pequeña escala de nuestra señal original, estas serán suavizadas tras
 2235 realizar la convolución. \ \ {\NOTA:} Por el teorema de convolución fil-
 2236 trar en el dominio temporal convolucionando es equivalente a multiplicar
 2237 la transformada de Fourier de la señal con la transformada de Fourier
 2238 de la función respuesta.

2239 Las ventanas mas comunes para suavizar señales son las de \ \ (1)
 2240 'Boxcar'

$$r(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq T \\ 0 & \text{if } t > T \end{cases}$$

2241

2242 *La transformada de Fourier es la función {sinc}*

$$R(\omega) = \text{sinc} = \frac{\sin\left(\frac{\omega T}{2}\right)}{\frac{\omega T}{2}}.$$

2243

2244 *Esta función respuesta tiende a cero cuando $\omega T/2$ se acerca a cero, es*
 2245 *dicho, para $\omega T = 2n\pi$; para $n = 1, 2, 3, \dots$. Esta no es una ventana*
 2246 *o función respuesta recomendable debido a los lóbulos de menor ampli-*
 2247 *tud alrededor del pico. En general respuestas tipo ondas sinusoidales o*
 2248 *cosinusoidales amortiguadas a ambos lados del pico no son deseables.*

2249 *{}(2) Hanning*

$$r(t) = \begin{cases} \frac{1}{2} \left(1 - \cos\frac{2\pi t}{T}\right) & \text{if } -T/2 \leq t \leq T/2 \\ 0 & \text{if contrario} \end{cases}$$

2250 *(3) Hamming*

$$r(t) = \begin{cases} (0.54 + 0.46\cos(\frac{\pi t}{T})) & \text{if } -T/2 \leq t \leq T/2 \\ 0 & \text{if contrario} \end{cases}$$

2251

2252 *Es evidente que este tipo de filtro es un suavizado o filtro pasa bajo. Sin*
 2253 *embargo, siempre podemos recuperar facilmente la señal de alta frecuen-*
 2254 *cia (filtro pasa altas) restando a la señal original la serie suavizada con*
 2255 *convolución.*

2256 *Imaginemos que tenemos una serie temporal $u(t)$ con un paso temporal*
 2257 *de $dt = 1$ h. Entonces, para suavizar $u(t)$ de tal forma que se eliminen*
 2258 *las señales con periodos menores de $T = 48$ h, es decir, un filtro pasa*
 2259 *baja con frecuencia de corte $1/48$ h $^{-1}$ debemos convolucionar $u(t)$ con*
 2260 *una función de respuesta $r(t)$ o ventana. En este ejemplo {Matlab} se*
 2261 *muestra como programar un suavizado*

Matlab:

```
%Crea función respuesta r(t) o ventana "espectral"
T=48; p1=hanning(48);
r1=p1./sum(p1); %Normaliza (0,1)
p2=hamming(48);
r2=p2./sum(p2); %Normaliza (0,1)
%Convolución u(t) y r(t)
%'same' mantiene las dimensiones de la serie filtrada
us1=conv(u,r1,'same');
us2=conv(u,r2,'same');
```

2262 *Ya hemos visto que podemos suavizar una señal simplemente con la*
 2263 *convolución en el dominio temporal de la señal con una ventana o fun-*
 2264 *ción respuesta (Boxcar, Hanning, Hamming, etc.). Suavizar una señal*

2265 es comparable a un filtro de pasa baja, es decir, un filtro que solamente
 2266 deja pasar las frecuencias bajas y elimina (pone a cero) las altas fre-
 2267 cuencias. De forma ideal los filtros en el dominio frecuencial los repre-
 2268 sentamos como: |

2269 (1) Pasa baja (filtrar las altas frecuencias),

$$|R(\omega)| = \begin{cases} 1 & \text{si } |\omega| \leq \omega_c \\ 0 & \text{si } \omega_c \leq \omega \end{cases}$$

2270 |
 2271 (2) Pasa banda (filtrar las frecuencias fuera de la banda)

$$|R(\omega)| = \begin{cases} 1 & \text{si } \omega_{c1} \leq |\omega| \leq \omega_{c2} \\ 0 & \text{(lo contrario)} \end{cases}$$

2272 |
 2273 (3) Pasa alta (filtrar las bajas frecuencias)

$$|R(\omega)| = \begin{cases} 0 & \text{si } |\omega| \leq \omega_c \\ 1 & \text{si } \omega_c \leq \omega \end{cases}$$

2275 10.2 Promedio corrido

2276 Veamos primero un ejemplo sencillo de filtrado paso bajo con un prome-
 2277 dio corrido de dos puntos. Este sería el caso de filtrar utilizando la
 2278 función `{smooth.m}` de Matlab. Para dos puntos sería `>> f_suavizada =`
 2279 `smooth(f, 2);`. | Promedio corrido con dos puntos es simplemente el
 2280 valor promedio

$$y(n) = \frac{s(n) + s(n-1)}{2},$$

2281 donde $s(n)$ es una señal periódica, $s(-1) = s(N)$, $N \geq 2$. La función
 2282 $y(n)$ es una versión suavizada con altas frecuencias eliminadas y bajas
 2283 frecuencias mantenidas. Para ver esto definimos $s(n) = \sin(2\pi fn/N)$,
 2284 y entonces

$$\begin{aligned} y(n) &= \frac{1}{2}s(n) + \frac{1}{2}s(n-1) = \frac{1}{2}\sin(2\pi fn/N) + \frac{1}{2}\sin(2\pi f(n-1)/N) = \\ &= \frac{1}{2}\sin(2\pi fn/N) + \frac{1}{2}[\sin(2\pi fn/N)\cos(2\pi f/N) - \cos(2\pi fn/N)\sin(2\pi f/N)] = \\ &= \frac{1}{2}[1 + \cos(2\pi f/N)]\sin(2\pi fn/N) - \frac{1}{2}\cos(2\pi fn/N)\sin(2\pi f/N) = \\ &= A_1\sin(2\pi fn/N) - A_2\cos(2\pi fn/N), \end{aligned}$$

2288 donde

$$A_1 = \frac{1}{2}[1 + \cos(2\pi f/N)]; \quad A_2 = \sin(2\pi f/N).$$

2289 Para bajas frecuencias, es decir, $f \sim 0$ se cumple que $A_1 \sim 1$ y $A_2 \sim 0$ y
 2290 entonces

$$y(n) \sim \sin(2\pi fn/N) = s(n)$$

2291 *y las bajas frecuencias son prácticamente mantenidas. Por el contrario*
 2292 *para altas frecuencias, es decir, $f \sim N/2$, $A_1 \sim 0$, $A_2 \sim 0$, y entonces*
 2293 *y(n) ~ 0 y consecuentemente las altas frecuencias son prácticamente*
 2294 *eliminadas.*

2295 *La fórmula general para el promedio corrido es*

$$y(n) = \frac{1}{M} \sum_{p=-(M-1)/2}^{(M-1)/2} s(n+p),$$

2296 *donde $y()$ es el valor de la serie filtrada y $s()$ es la serie original sin fil-
 2297 trar, M es el número de puntos usados en el promedio. Por ejemplo, en
 2298 un promedio corrido de 5 puntos, el valor en el punto 30 será*

$$y(30) = \frac{s(28) + s(29) + s(30) + s(31) + s(32)}{5}.$$

2299 *Es evidente que este tipo de filtro es un suavizado o filtro pasa bajo.
 2300 Sin embargo, podemos recuperar facilmente la señal de alta frecuencia
 2301 restando a la señal original la serie filtrada.*

Matlab:

```
%Suavizado/filtrado con ventana corrida (moving average)
us3=[];
Ws=48;
for n=Ws:length(u)
  us3_Hamming(n)=sum(u(n-Ws+1:n).*hamming(Ws))/Ws;
  us3_Hanning(n)=sum(u(n-Ws+1:n).*hanning(Ws))/Ws;
end
```

2302 10.3 Filtros generales coseno

2303 *Supongamos un simple filtro simétrico obtenido como la convolución
 2304 entre una función de pesos $r(t)$ y la señal $x(t)$*

$$y_n = \sum_{p=-\infty}^{\infty} r_p x_{n-p} \text{ donde } r_p = r_{-p},$$

2305 *son pesos elegidos adecuadamente. El efecto de filtrado se observa
 2306 mejor en el dominio frecuencial. Queremos calcular la transformada de
 2307 Fourier de una serie temporal $f(t)$, la cual ha sido desfasada un tiempo
 2308 $\Delta t = a$:*

$$f(t \pm a) = \int_{-\infty}^{\infty} F(\omega) e^{i\omega(t \pm a)} d\omega = \int_{-\infty}^{\infty} [F(\omega) e^{i\omega t}] e^{\pm i\omega a} d\omega$$

2309 *De esta expresión deducimos que la transformada de Fourier de una se-
 2310 rie desfasada por un intervalo de tiempo Δt es igual a la transformada
 2311 de Fourier de la serie no desfasada multiplicada por un factor*

$$e^{\pm i\omega \Delta t}.$$

²³¹² *Usando este resultado, la transformada de Fourier de y_n se puede es-*
²³¹³ *cribir como*

$$Y(\omega) = F[y_n] = \sum_{p=-\infty}^{\infty} r_p e^{-i\omega_p \Delta t} X(\omega),$$

²³¹⁴ *donde $X(\omega)$ y $Y(\omega)$ son la transformada de Fourier de $y(t)$ y $x(t)$, y la*
²³¹⁵ *función respuesta en el dominio frecuencial es*

$$R(\omega) = \frac{Y(\omega)}{X(\omega)} = \sum_{p=-\infty}^{\infty} r_p e^{-i\omega_p \Delta t}.$$

²³¹⁶ *{}Puesto que $r_p = r_{-p}$ y*

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2},$$

²³¹⁷ *podemos escribir la función respuesta del filtrado deseado como*

$$\begin{aligned} R(\omega) &= \sum_{p=-\infty}^{\infty} r_p e^{-i\omega_p \Delta t} = r_0 + \sum_{p=1}^{\infty} r_p e^{i\omega_p \Delta t} + \sum_{p=1}^{\infty} r_{-p} e^{-i\omega_p \Delta t} = r_0 + \sum_{p=1}^{\infty} r_p [e^{i\omega_p \Delta t} + e^{-i\omega_p \Delta t}] = \\ &= r_0 + 2 \sum_{p=1}^{\infty} r_p \left[\frac{e^{i\omega_p \Delta t} + e^{-i\omega_p \Delta t}}{2} \right] = r_0 + 2 \sum_{p=1}^{\infty} r_p \cos(\omega_p \Delta t) \end{aligned}$$

²³¹⁹ *En general los pesos r_p se van a calcular utilizando la siguiente expre-*
²³²⁰ *sión:*

$$r_p = \frac{1}{\omega_N} \int_0^{\omega_N} R(\omega_p) \cos(\omega_p \Delta t) d\omega$$

²³²¹ *Por ejemplo para un filtro pasa bajo $R(\omega) = 1$ para $0 < |\omega_p| \leq \omega_c$ y la*
²³²² *integral para calcular los pesos queda*

$$\begin{aligned} r_p &= \frac{1}{\omega_N} \int_0^{\omega_c} R(\omega_p) \cos(\omega_p \Delta t) d\omega = \frac{\omega_c}{\omega_N} \frac{\sin(\omega_c p \Delta t)}{\omega_c p \Delta t} = \frac{1}{\omega_N} \frac{\sin(\pi p \omega_c / \omega_N)}{\pi p / \omega_N} = \\ &= \frac{\sin(\pi p \omega_c / \omega_N)}{\pi p} \quad p = 1, 2, \dots, N \end{aligned}$$

²³²⁴ *Para $p = 0$ entonces*

$$r_0 = \frac{1}{\omega_N} \int_0^{\omega_c} R(\omega_p) \cos(\omega_p \Delta t) d\omega = \frac{1}{\omega_N} \int_0^{\omega_c} d\omega = \frac{\omega_c}{\omega_N}.$$

²³²⁵ *Y la función respuesta es*

$$R(\omega) = \frac{\omega_c}{\omega_N} + 2 \sum_{p=1}^{\infty} \frac{\sin(\pi p \omega_c / \omega_N)}{\pi p} \cos(\pi p \omega / \omega_N)$$

2326 *Veamos que forma tiene este filtro en el dominio frecuencial, asumiendo*
 2327 *un número finito de N coeficientes de Fourier, i.e., $p = 1, 2, \dots, N$,*
 2328 *frecuencia de Nyquist $f_N = 1$ y frecuencia de corte $f_c = 1$:*

2329 *En la figura observamos oscilaciones con longitud de onda*

$$\lambda = \frac{4f_N}{2N+1},$$

2331 *Esta longitud de onda coincide con el ancho de banda de transición del*
 2332 *filtro, es decir, del pico hasta la base indicado en la figura por las líneas*
 2333 *rojas. Para filtrar únicamente debemos:* \ (i) *multiplicar la respuesta*
 2334 *espectral $R(f)$ por la transformada de Fourier de la señal y regresar con*
 2335 *la transformada inversa*

$$x(t)[\text{filtrado}] = F^{-1}[R(f)X(f)],$$

2336 \ (ii) *convolucionar la respuesta en el dominio temporal $r(t)$ por la serie*
 2337 *temporal.*

$$x(t)[\text{filtrado}] = [r(t) * x](t).$$

2338 \ \{ } *Si queremos un filtro pasa alta, usamos r_p (pasa alto) = $1 - r_p$. Y la*
 2339 *función respuesta sería*

$$R(\omega)[\text{pasa alto}] = 1 - R(\omega) = 1 - \frac{\omega_c}{\omega_N} - 2 \sum_{p=1}^{\infty} \frac{\sin(\pi p \omega_c / \omega_N)}{\pi p} \cos(\pi p \omega / \omega_N)$$

2340 \ Veamos ahora de nuevo el promedio corrido pero esta vez usando el
 2341 método de Fourier. De nuevo decir que el promedio corrido reemplaza
 2342 el valor central de la ventana por el promedio de los valores que rodean
 2343 a ese punto. Para este ejemplo los pesos son siempre iguales $r_p = 1/T$
 2344 para el intervalo $-N < p < N$, donde $T = 1/(2N+1)$ es el tamaño de
 2345 la ventana ‘boxcar’. De esta forma \

$$T = 2N + 1 = 3 \quad R(\omega) = \frac{1}{3} + \frac{2}{3} \cos(\omega \Delta t) \quad 0 < \omega < \pi / \Delta t$$

$$T = 2N + 1 = 5 \quad R(\omega) = \frac{1}{5} + \frac{2}{5} \cos(\omega \Delta t) + \frac{2}{5} \cos(2\omega \Delta t) \quad 0 < \omega < \pi / \Delta t$$

$$T = 2N + 1 = 7 \quad R(\omega) = \frac{1}{7} + \frac{2}{7} \cos(\omega \Delta t) + \frac{2}{7} \cos(2\omega \Delta t) + \frac{2}{7} \cos(3\omega \Delta t) \quad 0 < \omega < \pi / \Delta t$$

2346
 2347 *Como volvemos a observar, las transformadas de Fourier de funciones*
 2348 *de peso cuadradas (‘boxcar’) no son adecuadas debido a las oscilaciones*
 2349 *o lóbulos menores.*

2350 10.4 Filtro Lanczos pasabajo

2351 Sea $R(f)$ la función respuesta de un filtro pasa baja, en donde f corre-
 2352 sponde a la frecuencia en ciclos por unidad de tiempo, f_N es la frecuen-
 2353 cia de Nyquist, y f_c es la frecuencia de corte.

2354
 2355 Lanzcos se dio cuenta que las oscilaciones de los filtros cosinusoidales
 2356 con longitud de onda $\lambda(f) = 4f_N/2N + 1$ podían ser reducidas si se
 2357 realiza un suavizado de la función respuesta $H(f)$. Para ello realizó un
 2358 promedio corrido de tamaño igual a la longitud de onda de las oscila-
 2359 ciones, es decir, λ . Esto se puede escribir como

$$\widetilde{R}(f) = \frac{1}{\lambda(f)} \int_{f-\lambda/2}^{f+\lambda/2} R(f) df,$$

2360 donde ya hemos visto que

$$R(f) = \frac{f_c}{f_N} + 2 \sum_{p=1}^N r_p \cos(\pi p f / f_N).$$

2361 Un filtro de media corrida no tiene efecto en el promedio, entonces

$$\begin{aligned} \widetilde{R}(f) &= \frac{f_c}{f_N} + \frac{1}{\lambda} \int_{f-\lambda/2}^{f+\lambda/2} 2 \sum_{p=1}^N r_p \cos(\pi p f / f_N) df = \\ &= \frac{f_c}{f_N} + \frac{2}{\lambda} \sum_{p=1}^N r_p \left[\frac{1}{\pi p / f_N} \sin\left(\frac{\pi p f}{f_N}\right) \right]_{f-\lambda/2}^{f+\lambda/2} = \\ &= \frac{f_c}{f_N} + \frac{2}{\lambda} \sum_{p=1}^N r_p \frac{f_N}{\pi p} \left[\sin\left(\frac{\pi p(f + \lambda/2)}{f_N}\right) - \sin\left(\frac{\pi p(f - \lambda/2)}{f_N}\right) \right] = \\ &= \frac{f_c}{f_N} + \frac{2}{\lambda} \sum_{p=1}^N r_p \frac{f_N}{\pi p} \left[2 \cos\left(\frac{\pi p f}{f_N}\right) \sin\left(\frac{\lambda \pi p}{2f_N}\right) \right] = \frac{f_c}{f_N} + \frac{2}{\lambda} \sum_{p=1}^N r_p \frac{f_N}{\pi p} \left[2 \cos\left(\frac{\pi p f}{f_N}\right) \sin\left(\frac{2\pi p}{2N+1}\right) \right] = \\ &= \frac{f_c}{f_N} + \frac{2(2N+1)}{4f_N} \sum_{p=1}^N r_p \frac{f_N}{\pi p} \left[2 \cos\left(\frac{\pi p f}{f_N}\right) \sin\left(\frac{2\pi p}{2N+1}\right) \right] = \\ &= \frac{f_c}{f_N} + 2 \sum_{p=1}^N r_p \frac{2N+1}{2\pi p} \left[\cos\left(\frac{\pi p f}{f_N}\right) \sin\left(\frac{2\pi p}{2N+1}\right) \right] = \\ &= \frac{f_c}{f_N} + 2 \sum_{p=1}^N r_p \frac{1}{\frac{2\pi p}{2N+1}} \left[\cos\left(\frac{\pi p f}{f_N}\right) \sin\left(\frac{2\pi p}{2N+1}\right) \right] = \frac{f_c}{f_N} + 2 \sum_{p=1}^N r_p \sigma_p \cos\left(\frac{\pi p f}{f_N}\right), \end{aligned}$$

2368 donde

$$\sigma_p = \frac{\sin\left(\frac{2\pi p}{2N+1}\right)}{\frac{2\pi p}{2N+1}} = \text{sinc}\left(\frac{2\pi p}{2N+1}\right),$$

2369 es una función $\{sinc\}$ como la respuesta espectral de un filtro rectangular
 2370 o ‘boxcar’. A este factor de suavizado se le suele llamar peso sigma.

2371

2372 En la figura se observa que la frecuencia de corte es la la frecuencia que
 2373 corta el 50% de la magnitud de la respuesta. La frecuencia efectiva es
 2374 $f = f_c + \lambda/2$.

MATLAB:

Programa para crear el filtro de Lanczos (Autor: Dr. Modesto Ortiz)

```
N=10; %tamaño filtro
p=1:N; %índices de tiempo
fc=0.5; %frecuencia de corte a 0.5
fn=1.0; %frecuencia de Nyquist
f=-fn:0.01:fn; %generamos vector de frecuencias (positivas y negativas)
k=length(f); %dimension vector frecuencias
%funcion pesos o respuesta en el dominio temporal
rp=sin(pi*p*fc/fn)./(pi*p);
%calculamos el factor sigma de suavizado (factores lanczos)
sigma=sin(2*pi*p/(2*N+1))./(2*pi*p/(2*N+1));
%Calculamos la función respuesta de un filtro general en el %dominio espectral
R=[];
for m=1:k; %recorremos todas las frecuencias
R(m) = fc/fn +
2*sum(rp.*sigma.*cos(p*pi*f(m)/fn));
end
lambda=4*fn/(2*N+1);
y0=[-0.2:0.1:1.2];x0=0.4*ones(1,length([-0.2:0.1:1.2]));
y1=[-0.2:0.1:1.2];x1=0.4+lambda*ones(1,length([-0.2:0.1:1.2]));
%Graficamos
f1=figure;
plot(f,R,'b');
hold on;
plot(x0,y0,'r-');
hold on;
plot(x1,y1,'r-');
grid on;
xlabel('f(ciclos/tiempo)', 'fontsize', 18);
ylabel('R(f)', 'fontsize', 18);
set(gca, 'fontsize', 18);
axis([-1 1 -0.2 1.2]);
```

2375

10.5 Filtro Lanczos pasabanda

2376 En el dominio de las frecuencias, el filtro de pasa-banda se obtiene con
 2377 volucionando el filtro pasa-bajas con la transformada de Fourier de la

2378 función coseno:

$$\widetilde{R}_b(f) = \widetilde{R}(f) * [\delta(f - f_o) + \delta(f + f_0)] ,$$

2379 donde * significa convolución,

$$\begin{aligned} F[\cos(2\pi f_0 x)](f) &= \int_{-\infty}^{\infty} e^{-2\pi i f x} \cos(2\pi f_0 x) dx = \int_{-\infty}^{\infty} e^{-2\pi i f x} \left(\frac{e^{2\pi i f_0 x} + e^{-2\pi i f_0 x}}{2} \right) dx = \\ &= \frac{1}{2} \int_{-\infty}^{\infty} [e^{-2\pi i (f-f_0)x} + e^{-2\pi i (f+f_0)x}] dx \\ &= \frac{1}{2} [\delta(f - f_0) + \delta(f + f_0)] , \end{aligned}$$

2380 y la delta de dirac se define como

$$\delta(x) = F[1](f \pm f_0) = \int_{-\infty}^{\infty} e^{-2\pi i (f \pm f_0)x} dx$$

2383 Vemos que la transformada de Fourier del coseno se ha multiplicado
2384 por un factor de 2 para que la respuesta del filtro sea unitaria (normal-
2385 ización). El resultado de la convolución es

$$\begin{aligned} \widetilde{R}_b(f) &= \widetilde{R}(f) * [\delta(f - f_o) + \delta(f + f_0)] = \widetilde{R}(f - f_0) + \widetilde{R}(f + f_0) = \\ &= \frac{f_c}{f_N} + 2 \sum_{p=1}^N r_p \sigma_p \cos \left(\frac{\pi p (f - f_0)}{f_N} \right) + \frac{f_c}{f_N} + 2 \sum_{p=1}^N r_p \sigma_p \cos \left(\frac{\pi p (f + f_0)}{f_N} \right) = \\ &= 2 \frac{f_c}{f_N} + 2 \sum_{p=1}^N r_p \sigma_p \left[\cos \left(\frac{\pi p f}{f_N} - \frac{\pi p f_0}{f_N} \right) + \cos \left(\frac{\pi p f}{f_N} + \frac{\pi p f_0}{f_N} \right) \right] = \\ &= 2 \frac{f_c}{f_N} + 2 \sum_{p=1}^N r_p \sigma_p \left[2 \cos \left(\frac{\pi p f}{f_N} \right) \cos \left(\frac{\pi p f_0}{f_N} \right) \right] = \\ &= 2 \frac{f_c}{f_N} + 4 \sum_{p=1}^N r_p \sigma_p \left[\cos \left(\frac{\pi p f}{f_N} \right) \cos \left(\frac{\pi p f_0}{f_N} \right) \right] \end{aligned}$$

2389

MATLAB:

Programa para crear el filtro de Lanczos pasa-banda (Autor: Dr. Modesto Ortiz)

```

N=10; %tamaño filtro
p=1:N; %índices de tiempo
fc=0.2; %frecuencia de corte a 0.5
f0=0.2; %frecuencia central del filtro
fn=1.0; %frecuencia de Nyquist
f=-fn:0.01:fn; %generamos vector de frecuencias (positivas y negativas)
k=length(f); %dimension vector frecuencias
%función pesos o respuesta en el dominio temporal
rp=sin(pi*p*fc/fn)./(pi*p);
%calculamos el factor sigma de suavizado (factores lanczos)
sigma=sin(2*pi*p/(2*N+1))./(2*pi*p/(2*N+1));
%dominio espectral
R=[];
for m=1:k; R(m) = 2*fc/fn + 4*sum(rp.*sigma.*cos(p*pi*f0/fn).*cos(p*pi*f(m)/fn)); end
lambda=4*fn/(2*N+1);
y0=[-0.2:0.1:1.2];x0=0.4*ones(1,length([-0.2:0.1:1.2]));
y1=[-0.2:0.1:1.2];x1=0.4+lambda*ones(1,length([-0.2:0.1:1.2]));
%Graficamos
f1=figure;
plot(f,R,'b');
hold on;
plot(x0,y0,'r-');
hold on;
plot(x1,y1,'r-');
grid on;
xlabel('f(ciclos/tiempo)', 'fontsize', 18);
ylabel('R(f)', 'fontsize', 18);
set(gca, 'fontsize', 18);
axis([-1 1 -0.2 1.2]);

```

2391 Para el filtro pasa bajas:

$$x(t)[\text{filtrado}] = [r_p \sigma_p * x](t),$$

2392 o

$$x(t)[\text{filtrado}] = F^{-1}[\widetilde{R}(f)X(f)].$$

2393 Para el filtro pasa-banda:

$$x(t)[\text{filtrado}] = [r_p \sigma_p \cos\left(\frac{\pi p f_0}{f_N}\right) * x](t),$$

2394 o

$$x(t)[\text{filtrado}] = F^{-1}[\widetilde{R}_b(f)X(f)].$$

2395 *Todo esto se puede programar facilmente en {Matlab} como se muestra
2396 en el siguiente:*

MATLAB:

Programa para calcular los coeficientes de Lanczos pasa-baja y pasa-banda (Autor: Dr. Modesto Ortiz)

function [h,Fh,Rh]=lnz_co(fo,fc,fn,N,df);

% INPUT:

% fo = frecuencia central del filtro

% fc = frecuencia de corte (al 50% de la señal)= medio ancho de banda del filtro.

% fn = frecuencia de Nyquist de los datos que se van a filtrar.

*% df = 1/(N*dt) de los datos que vamos a filtrar*

% N = longitud del filtro

% El número total de coeficientes es (N+1+N). Con esta definicion

% aseguramos que el número total de coeficientes sea impar. Esto significa

% que el filtro no cambia la fase de la señal filtrada.

2397

% OUTPUT

% h = coeficientes del filtro Lanczos en el dominio del tiempo

```

% Fh = vector de frecuencias
% Rh = respuesta en el dominio espectral
%*****
Fh=[]; Rh=[];
lambda = 4*fn/(2*N+1); % ancho de banda de transición
% este valor debe ser menor que 1
lambda/(2*fc)
p= -N : N;
rp=sin(pi*p*fc/fn)./(pi*p); %coeficientes de Fourier del pulso
cuadrado.
%esta operación da una división por cero pero
%no te asustes, se evalua con la función sinc.
%rp=(fc/fn)*sinc(p*fc/fn); % esta es la misma pero no da division
por cero
Sigma=sin(2*pi*p/(2*N+1))./(2*pi*p/(2*N+1)); %factores de
Lanczos, division por cero
%Sigma=sinc(2*p/(2*N+1)); % lo mismo pero sin division por cero
Co=fc/fn;
ind=find(p==0); % se encuentra el índice de la división por cero.
h=rp.*Sigma;
h(ind)=Co; % se suple el valor del coeficiente cero y ya tenemos
%*****
if abs(fo) > 0
banda=cos(2*pi*p*fo/(2*fn));
h=2*h.*banda; % coeficientes del filtro lanczos pasa banda centrado
en fo
end
%*****

```

```

% Por comodidad generamos la respuesta en frecuencia del filtro.
Fh=-fn:df:fn; Fh=Fh';
% for k=1:length(Fh);
% Rh(k)=sum( h.*cos( 2*pi*p*Fh(k)/(2*fn) ) );
% end;
%
xdum=length(Fh); ydum=length(p);
A = zeros(xdum,ydum);
dum=2*pi/(2*fn);
A = cos(dum*Fh*p);
M=length(h);
h=reshape(h,M,1);
Rh = A*h;%multiplicamos la matriz de cosenos por la señal (= a
suma comentada de arriba)

```

MATLAB:

Programa para filtrar con Lanczos en el dominio temporal (Autor: Dr. Modesto Ortiz)

```
% Función lnz_fi.m
%INPUT:
% z = serie de tiempo sin filtrar
% h = coeficientes del filtro
%OUTPUT
% zf = serie de tiempo filtrada
%*****
function [zf] = lnz_fi(z,h);
m = length(z);
zf = conv (z , h);
% se obtiene la serie filtrada convolucionando la serie original con los pesos o coeficientes del filtro.
zf = wkeep(zf,m); % esta rutina trunca los extremos de zf para que el número de datos de la serie filtrada sea igual al número de datos de la serie original. El objetivo de esto es poder graficar en un mismo eje de tiempo ambas series para poder observar la bondad del filtro.
```

2400 Aquí se muestra un ejemplo sintético del uso de las subrutinas presentadas.
 2401

MATLAB:

Ejemplo de filtrado con Lanczos (Autor: Dr. Modesto Ortiz)

```
dt=1; k=200; T=k*dt; df=1/(k*dt); t=(0:1:k-1)*dt; %eje de tiempo.
f1=0.005; f2=0.1; f3=0.25; % tres frecuencias
z=cos(2*pi*f1*t)+cos(2*pi*f2*t)+cos(2*pi*f3*t); %serie de tiempo sintética.
%filtramos pasa-banda para quedarnos con la frecuencia f2.
fo=f2; %frecuencia central del filtro
%Medio ancho de banda del filtro
fc=0.03;
fn=0.5; % frecuencia de Nyquist
N = 30; % recordemos que el número total de coeficientes será N+1+N.
[h, Fh, Rh]=lnz_co(fo,fc,fn,N,df);
zf = lnz_fi(z,h); %serie filtrada y lista.
```

2402 **11 Temas selectos**

11.1 Análisis Armónico

2403 Se trata de un ajuste por mínimos cuadrados de una serie temporal
 2404 dominada por armónicos específicos. Por ejemplo, en el caso del océano,

2406 *es muy común encontrar en series temporales de temperatura, salinidad,
 2407 velocidad, etc.. señales de las mareas que no son nada mas que corri-
 2408 entes periódicas generadas por fuerzas astronómicas con una frecuencia
 2409 de oscilación determinada (24h, 12h, etc...).*

2410 *El método consiste en elejir las frecuencias de los armónicos y usar
 2411 cuadrados mínimos para ajustarlos a la serie temporal. Supongamos M
 2412 armónicos a ajustar*

$$y(t_n) = \overline{y(t)} + \sum_{q=1}^M C_q \cos(\omega_q t_n - \theta_q) + y_r(t_n),$$

2413 *donde $\overline{y(t)}$ es el promedio de la serie, y $y_r(t_n)$ es el residuo de la se-
 2414 rie temporal (donde hay el resto de armónicos presentes en la serie),
 2415 $\omega_q = 2\pi q/N\Delta t$. En términos de las amplitudes A_q y B_q*

$$y(t_n) = \overline{f(t)} + \sum_{q=1}^M [A_q \cos(\omega_q t_n) + B_q \sin(\omega_q t_n)] + y_r(t_n),$$

2416 *donde*

$$C_q = \sqrt{A_q^2 + B_q^2}, \quad q = 1, 2, \dots$$

$$\theta_q = \arctg[B_q/A_q], \quad q = 1, 2, \dots$$

2418 *Antes de empezar el análisis debemos de extraer la media, \bar{y} , a la serie
 2419 temporal. El método de mínimos cuadrados consiste en minimizar la
 2420 suma de los errores cuadrados SEC, es decir,*

$$\begin{aligned} SEC &= \sum_{n=1}^N y_r^2(t_n) = \sum_{n=1}^N \left(y(t_n) - \left[\overline{y(t)} + \sum_{q=1}^M A_q \cos(\omega_q t_n) + B_q \sin(\omega_q t_n) \right] \right)^2 = \\ &= \sum_{n=1}^N \left(y(t_n) - \left[\overline{y(t)} + \sum_{q=1}^M A_q \cos(2\pi q n / N) + B_q \sin(2\pi q n / N) \right] \right)^2 \end{aligned}$$

2422 *Como siempre derivamos respecto los coeficientes e igualamos a cero
 2423 para obtener un sistema de $2M + 1$ ecuaciones*

$$\frac{\partial SEC}{\partial A_q} = 0 = 2 \sum_{n=1}^N \left(y(t_n) - \left[\overline{y(t)} + \sum_{q=1}^M A_q \cos(2\pi q n / N) + B_q \sin(2\pi q n / N) \right] - \cos(2\pi q n / N) \right)$$

$$\frac{\partial SEC}{\partial B_q} = 0 = 2 \sum_{n=1}^N \left(y(t_n) - \left[\overline{y(t)} + \sum_{q=1}^M A_q \cos(2\pi q n / N) + B_q \sin(2\pi q n / N) \right] - \sin(2\pi q n / N) \right)$$

2425 Soluciones del sistema requiere una ecuación matricial de la forma
 2426 $\mathbf{Dz} = \mathbf{y}$, donde

$$\mathbf{D} = \begin{pmatrix} N & c_1 & c_2 & \dots & c_M & s_1 & s_2 & \dots & s_M \\ c_1 & cc_{11} & cc_{12} & \dots & cc_{1M} & cs_{11} & cs_{12} & \dots & cs_{1M} \\ c_2 & cc_{21} & cc_{22} & \dots & cc_{2M} & cs_{21} & cs_{22} & \dots & cs_{2M} \\ \dots & \dots \\ \dots & \dots \\ c_M & cc_{M1} & cc_{M2} & \dots & cc_{MM} & cs_{M1} & cs_{M2} & \dots & cs_{MM} \\ s_1 & sc_{11} & sc_{12} & \dots & sc_{1M} & ss_{11} & ss_{12} & \dots & ss_{1M} \\ s_2 & sc_{21} & sc_{22} & \dots & sc_{2M} & ss_{21} & ss_{22} & \dots & ss_{2M} \\ \dots & \dots \\ \dots & \dots \\ s_M & sc_{M1} & sc_{M2} & \dots & sc_{MM} & ss_{M1} & ss_{M2} & \dots & ss_{MM} \end{pmatrix}$$

2427 y

$$\mathbf{y} = \begin{pmatrix} yc_0 \\ yc_1 \\ yc_2 \\ \dots \\ \dots \\ yc_M \\ ys_1 \\ ys_2 \\ \dots \\ \dots \\ ys_M \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ \dots \\ \dots \\ A_M \\ B_1 \\ B_2 \\ \dots \\ \dots \\ B_M \end{pmatrix}$$

2428 Los coeficientes de las matrices son:

$$yc_i = \sum_{n=1}^N y(t_n) \cos(\omega_i t_n) , \quad ys_i = \sum_{n=1}^N y(t_n) \sin(\omega_i t_n)$$

$$2429 \quad c_i = \sum_{n=1}^N \cos(\omega_i t_n) , \quad s_i = \sum_{n=1}^N \sin(\omega_i t_n)$$

$$2430 \quad cc_{ij} = cc_{ji} = \sum_{n=1}^N [\cos(\omega_i t_n) \cos(\omega_j t_n)]$$

$$2431 \quad ss_{ij} = ss_{ji} = \sum_{n=1}^N [\sin(\omega_i t_n) \sin(\omega_j t_n)]$$

$$2432 \quad cs_{ij} = sc_{ji} = \sum_{n=1}^N [\cos(\omega_i t_n) \sin(\omega_j t_n)],$$

2433 donde $t_n = n\Delta$, $\omega_i = 2\pi f_i$ es la frecuencia angular de las componentes
 2434 de interés i , y $\phi_i(n) = \omega_i t_n$ es el argumento de las funciones de Fourier.

2435 {Ejemplo de Ajuste de armónicos (Emery and Thompson, p395)}

2436 Asumamos la siguiente serie temporal de temperatura promedia mensual.

2437 |

T ($^{\circ}$ C)	7.6	7.4	8.2	9.2	10.2	11.5	12.4	13.4	13.7	11.8	10.1	9
Año 2	8.9	9.5	10.6	11.4	12.9	12.7	13.9	14.2	13.5	11.4	10.9	8.1

2439 | |

2440 Deseamos encontrar las componentes mareales dominantes en la serie
 2441 temporal de temperatura. A simple vista podemos ver que existe una fre-
 2442 cuencia dominante semianual. Por tanto, vamos buscar las amplitudes
 2443 y frecuencias de interés, es decir, de las componentes anual y semian-
 2444 ual que tienen unas frecuencias de $f_1 = 1/12$ meses ($= 0.0833$ cpm) y
 2445 $f_2 = 1/24$ meses ($= 0.1667$ cpm). Los argumentos de las funciones de
 2446 Fourier son $\phi_1(n) = \omega_1 t_n = 2\pi(1/12) * n * \Delta t = (\pi/6) * n * 1 = n\pi/6$
 2447 y $\phi_2(n) = \omega_2 t_n = 2\pi(1/6) * n * \Delta t = (\pi/3) * n * 1 = n\pi/3$ Para este
 2448 problema las matrices son |

$$\begin{aligned}
 \mathbf{D} &= \begin{pmatrix} N & c_1 & c_2 & s_1 & s_2 \\ c_1 & cc_{11} & cc_{12} & cs_{11} & cs_{12} \\ c_2 & cc_{21} & cc_{22} & cs_{21} & cs_{22} \\ s_1 & sc_{11} & sc_{12} & ss_{11} & ss_{12} \\ s_2 & sc_{21} & sc_{22} & ss_{21} & ss_{22} \end{pmatrix} = \\
 &= \begin{pmatrix} N & c_1 & c_2 & s_1 & s_2 \\ c_1 & \sum_{n=1}^N [\cos(\phi_1(n))\cos(\phi_1(n))] & cc_{12} & cs_{11} & cs_{12} \\ c_2 & cc_{21} & \sum_{n=1}^N [\cos(\phi_1(n))\cos(\phi_2(n))] & cs_{21} & cs_{22} \\ s_1 & sc_{11} & sc_{12} & \sum_{n=1}^N [\sin(\phi_1(n))\sin(\phi_1(n))] & ss_{12} \\ s_2 & sc_{21} & sc_{22} & ss_{21} & \sum_{n=1}^N [\sin(\phi_2(n))\sin(\phi_2(n))] \end{pmatrix} = \\
 &= \begin{pmatrix} 24 & 0 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 \\ 0 & 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 0 & 12 \end{pmatrix}
 \end{aligned}$$

2451 La matriz

$$\mathbf{y} = \begin{pmatrix} \sum_{n=1}^N y(t_n)\cos(\omega_0 t_n) \\ \sum_{n=1}^N y(t_n)\cos(\omega_1 t_n) \\ \sum_{n=1}^N y(t_n)\cos(\omega_2 t_n) \\ \sum_{n=1}^N y(t_n)\sin(\omega_1 t_n) \\ \sum_{n=1}^N y(t_n)\sin(\omega_2 t_n) \end{pmatrix} = \begin{pmatrix} 262.5 \\ -21.45 \\ -5.4 \\ -23.76 \\ -0.51 \end{pmatrix} {}^{\circ}\text{C}$$

2452 *Finalmente encontramos las amplitudes de los armónicos resolviendo el*
 2453 *sistema*

$$\mathbf{z} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{y} = \mathbf{D}^{-1} \mathbf{y} = \begin{pmatrix} 10.93 \\ -1.78 \\ -0.45 \\ -1.98 \\ -0.04 \end{pmatrix}$$

2454 *El coeficiente de correlación entre la señal original y la serie de Fourier*
 2455 *con 2 armónicos es $r^2 = 0.92$, es decir, solamente con 2 armónicos*
 2456 *podemos explicar el 92% de la varianza total.*

2457 11.2 Demodulación compleja

2458 *Este método es utilizado para conocer el comportamiento de una compo-*
 2459 *nente o armónico con frecuencia particular ω , tal como la marea diurna,*
 2460 *o semidiurna, o las ondas inerciales. Aquí vamos a mostrar la forma*
 2461 *clásica de demodular que consiste en ajustar por fragmentos de la serie*
 2462 *un armónico teórico utilizando mínimos cuadrados. Cada fragmento de*
 2463 *la serie debe, como mínimo, contener un ciclo del armónico a demodu-*
 2464 *lar. Para cada segmento, la anomalía de la componente de velocidad a*
 2465 *la frecuencia de interés ω es*

$$\mathbf{u} - \bar{\mathbf{u}} = [u(t) - \overline{u(t)} + iv(t) - \overline{v(t)}] = \\ R^+ e^{i(\omega t + \epsilon^+)} + R^- e^{-i(\omega t + \epsilon^-)},$$

2466 *donde $\overline{u(t)}$, $\overline{v(t)}$ son las componentes de la velocidad promedio, R^+ , R^-
 2467 y ϵ^+ , ϵ^- son las amplitudes y fases de las componentes rotatorias que
 2468 giran en el sentido de las agujas del reloj (+) y en el sentido contrario
 2469 (-). La serie temporal está definida para cada t_k ($k = 1, 2, \dots, N$) y las
 2470 soluciones son encontradas resolviendo el sistema de ecuaciones*

$$\mathbf{z} = \mathbf{D}^{-1} \mathbf{y},$$

2472 *donde*

$$\mathbf{y} = \begin{pmatrix} u(t_1) \\ u(t_2) \\ \dots \\ u(t_n) \\ v(t_1) \\ v(t_2) \\ \dots \\ v(t_n) \end{pmatrix}; \quad \mathbf{z} = \begin{pmatrix} R^+ \cos(\epsilon^+) \\ R^+ \sin(\epsilon^+) \\ R^- \cos(\epsilon^-) \\ R^- \sin(\epsilon^-) \end{pmatrix} = \begin{pmatrix} ACP \\ ASP \\ ACM \\ ASM \end{pmatrix},$$

2473 *y la matriz \mathbf{D} es*

$$\mathbf{D} = \begin{pmatrix} \cos(\omega t_1) & -\sin(\omega t_1) & \cos(\omega t_1) & \sin(\omega t_1) \\ \cos(\omega t_2) & -\sin(\omega t_2) & \cos(\omega t_2) & \sin(\omega t_2) \\ \dots & & & \\ \cos(\omega t_n) & -\sin(\omega t_n) & \cos(\omega t_n) & \sin(\omega t_n) \\ \sin(\omega t_1) & \cos(\omega t_1) & -\sin(\omega t_1) & \cos(\omega t_1) \\ \sin(\omega t_2) & \cos(\omega t_2) & -\sin(\omega t_2) & \cos(\omega t_2) \\ \dots & & & \\ \sin(\omega t_n) & \cos(\omega t_n) & -\sin(\omega t_n) & \cos(\omega t_n) \end{pmatrix}.$$

2474 Una vez los valores de z son encontrados a partir de la solución de mínimos cuadrados de arriba, podemos encontrar los parámetros de la elipse
 2475 como:
 2476

$$R^+ = \sqrt{(ASP^2 + ACP^2)}; \quad R^- = \sqrt{(ASM^2 + ACM^2)}$$

$$\epsilon^+ = \tan^{-1} \left(\frac{ASP}{ACP} \right); \quad \epsilon^- = \tan^{-1} \left(\frac{ASM}{ACM} \right)$$

2478 Por ejemplo, si queremos demodular la amplitud y fase de las ondas in-
 2479 erciales observadas en un anclaje situado en latitudes medias, debemos
 2480 de usar una frecuencia $\omega = 2\Omega \sin \phi$ y ajustar por mínimos cuadrados
 2481 segmentos de 24 h sin superposición. La serie temporal medida por el
 2482 anclaje debería de ser horaria para que existan mas datos por segmento
 2483 que parámetros a ajustar.

2484 Otra forma, tal vez mas sencilla, es la siguiente. Imaginemos que la se-
 2485 rie original es $X(t)$ y se asume como una señal periódica con frecuencia
 2486 igual a la de interés mas otras cosas que llamamos $Z(t)$

$$X(t) = A(t) \cos(\omega t + \varphi(t)) + Z(t) = \frac{1}{2} A(t) [e^{i(\omega t + \varphi(t))} + e^{-i(\omega t + \varphi(t))}] + Z(t),$$

2487 donde la amplitud $A(t)$ y la fase $\varphi(t)$ de la señal periódica se asumen
 2488 que dependen del tiempo pero que varían “lentamente” en comparación a
 2489 la frecuencia ω .

2490 Para demodular tenemos que: |

2491 (1) Multiplicar $X(t)$ por $e^{-i\omega t}$:

$$\begin{aligned} Y(t) &= X(t)e^{-i\omega t} = \frac{1}{2} A(t) [e^{i(\omega t + \varphi(t))} + e^{-i(\omega t + \varphi(t))}] e^{-i\omega t} + Z(t)e^{-i\omega t} = \\ &= \frac{1}{2} A(t) e^{i(\omega t + \varphi(t))} e^{-i\omega t} + \frac{1}{2} A(t) e^{-i(\omega t + \varphi(t))} e^{-i\omega t} + Z(t)e^{-i\omega t} = \\ &= \underbrace{\frac{1}{2} A(t) e^{i\varphi(t)}}_{(a)} + \underbrace{\frac{1}{2} A(t) e^{-i(2\omega t + \varphi(t))}}_{(b)} + \underbrace{Z(t)e^{-i\omega t}}_{(c)}. \end{aligned}$$

2492 El término (a) varía lentamente ya que $\varphi(t)$ también lo hace y no tiene
 2493 energía (potencia espectral) a la frecuencia de demodulación ω o arriba
 de ella. El término (b) oscila a dos veces la frecuencia de demodulación,
 i.e., 2ω . El término (c) varía a la frecuencia ω . Debido a que $Z(t)$ no
 2494 tiene energía a la frecuencia ω , entonces el término (c) no tendrá tam-
 2495 poco energía en la frecuencia cero, i.e., $\omega = 0$. |

2500 (2) Filtro pasa-bajas de la serie $Y(t)$ para eliminar las ondas con fre-
 2501 cuencia ω o por encima de ω . Esto eliminará prácticamente los
 2502 términos (b) y (c), y suavizará (a). El resultado es

$$Y_s(t) = \frac{1}{2} A_s(t) e^{i\varphi_s(t)},$$

2503 donde el subíndice s significa suavizado o filtro pasa-bajas. |

2504 (3) Extraer $A_s(t)$ y $\varphi_s(t)$:

$$\frac{1}{2}A_s(t) = |Y_s(t)| = 2(Re[Y_s]^2 + Im[Y_s]^2)^{1/2}$$

$$e^{i\varphi(t)} = 2 \frac{Y_s(t)}{A_s(t)}; \quad \varphi_s(t) = \tan^{-1} \left(\frac{Im[Y_s]}{Re[Y_s]} \right)$$

2506 *Al suavizar hacemos dos cosas. Primero, eliminamos los términos no
 2507 deseados (a) y (b). El tipo de filtrado o suavizado determina la anchura
 2508 de la banda de frecuencias de las oscilaciones retenidas. Por ejemplo,
 2509 si usamos un triángulo (ventana triangular) de longitud $2T - 1$ donde
 2510 $T = 2\pi/\omega$ es el periodo de demodulación, entonces para la banda para la
 2511 potencia-media (3 dB desde el pico) será $\omega \in [T/(1 + 0.44295), T/(1 -$
 2512 $0.44295)]$. Potencia media se refiere a la frecuencia a la cual la poten-
 2513 cia se ha reducido a la mitad de su valor medio de la banda. Segundo,
 2514 el filtrado suaviza las series de amplitud y la fase.* \

2515 (4) La elección de la frecuencia de demodulación ω se puede validar
 2516 ajustando localmente una línea a la fase, $\varphi \simeq a + bt$. Típicamente
 2517 esto lo haremos en fragmentos de longiutd T . De esta forma si
 2518 seleccionamos el origen en el tiempo central de cada fragmento (tal
 2519 que $a \simeq 0$) obtenemos que $\cos(\omega t + \varphi) \simeq \cos(\omega t + bt) = \cos(\hat{\omega}t)$.

2520 *La frecuencia ajustada $\hat{\omega} = \omega + a$ es una validación de la elección inicial
 2521 de nuestra frecuencia de demodulación ω .*

2522 **12 Acknowledgments**

2523 *Ayuda de Muchos*

2525 **13 Open research**

2526 *Disponible para todos*

2527 **References**