TEMA III

DISTRIBUCIONES CONDICIONADAS; CORRELACION Y REGRESION

ESQUEMA/RESUMEN

Nos proponemos en este capítulo caracterizar toda una serie de características estadísticas, tales como los coeficientes de correlación, correlación múltiple y parcial, así como las funciones de densidad condicionadas, exclusivamente en términos de los dos grupos de parámetros que determinan completamente la distribución normal n-dimensional, el vector de medias y la matriz de covarianzas.

Al final del capítulo se incluyen algunos resultados acerca de la función característica, que más adelante serán de utilidad.

3.1. DISTRIBUCION CONDICIONADA

En lo que sigue consideraremos un vector aleatorio n-dimensional ξ con distribución $N[\mu, \Sigma]$. Consideraremos también una partición de ξ en dos vectores, uno p-dimensional ξ_1 y otro n-p dimensional ξ_2. Sin pérdida alguna de generalidad, podemos suponer que ξ_1 está formado por las p primeras componentes de ξ, ya que en otro caso reordenando las variables, los elementos del vector de medias y las filas y columnas de la matriz de covarianzas, podemos transladar cualquier subconjunto de p componentes de ξ a los p primeros lugares.

Nuestro primer objetivo es calcular la distribución de ξ_1 condicionada a $\xi_2 = x_2$.

Una solución rápida y cómoda es alcanzable a partir de los resultados del capítulo anterior (ver ejercicios de autocomprobación). Hemos preferido aquí emplear un procedimiento más largo y engorroso, con el fin de utilizar algunos resultados sobre matrices, de cuya existencia es conveniente tener conocimiento.

Basándonos en la definición clásica de función de densidad condicionada

$$f(x/y) = \frac{f(x, y)}{f_2(y)}$$

tenemos que la función de densidad conjunta, por hipótesis será:

$$f(x_1, x_2) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x_1 - \mu_1)', (x_2 - \mu_2)') \Sigma^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right\} \quad [1]$$

por otra parte, como $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ la densidad marginal de ξ_2 será:

$$f_2(x) = (2\pi)^{-(n-p)/2} |\Sigma_{22}|^{-1/2} \exp\left\{(x_2 - \mu_2)' \Sigma_{22}^{-1} (x_2 - \mu)\right\} \quad [2]$$

la función de densidad condicionada vendrá dada por:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} \quad [3]$$

$$= \frac{(2\pi)^{-\frac{n}{2}} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}((x_1 - \mu_1)', (x_2 - \mu_2)') \Sigma^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right\}}{(2\pi)^{-(n-p)/2} |\Sigma_{22}|^{-1/2} \exp\left\{-\frac{1}{2}(x_1 - \mu_1)' \Sigma_{22}^{-1} (x_2 - \mu_2)\right\}}$$

Naturalmente no debemos conformarnos con esta primera expresión, sino que procuraremos simplificarla al máximo. Para ello vamos a emplear el siguiente resultado, sobre matrices, que el lector deberá probar (c.f. ejercicios de autocomprobación).

Propiedad 3.1.

Si $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ es una matriz cuadrada no singular tal que Σ_{22} es no singular, se verifica

$$\Sigma^{-1} = \begin{pmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & -(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1})\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1} \end{pmatrix} \quad [4]$$

y además,

$$|\Sigma| = |\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}| \, |\Sigma_{22}| \quad [5]$$

Naturalmente dado que Σ es una matriz definida positiva, Σ y Σ_{22} son no singulares, por tanto, estamos en condiciones de aplicar la propiedad 3.1.

De [3] tenemos

$$f(\mathbf{x}_1|\mathbf{x}_2) = (2\pi)^{-p/2} \frac{|\Sigma|^{-1/2}}{|\Sigma_{22}|^{-1/2}} \exp\left\{-\frac{1}{2}((\mathbf{x}_1 - \mu_1)', (\mathbf{x}_2 - \mu_2)')\, \Sigma^{-1} \begin{pmatrix} \mathbf{x}_1 - \mu_1 \\ \mathbf{x}_2 - \mu_2 \end{pmatrix}\right\} \cdot \qquad [6]$$

$$\cdot \exp\left\{\frac{1}{2}(\mathbf{x}_2 - \mu_2)'\, \Sigma_{22}^{-1} (\mathbf{x}_2 - \mu_2)\right\}$$

Teniendo en cuenta [4] obtenemos

$$((\mathbf{x}_1 - \mu_1)', (\mathbf{x}_2 - \mu_2)')\, \Sigma^{-1} \begin{pmatrix} \mathbf{x}_1 - \mu_1 \\ \mathbf{x}_2 - \mu_2 \end{pmatrix} =$$

$$= (\mathbf{x}_1 - \mu_1)' (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} (\mathbf{x}_1 - \mu_1) -$$

$$- (\mathbf{x}_1 - \mu_1)' (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \Sigma_{12}\Sigma_{22}^{-1} (\mathbf{x}_2 - \mu_2) -$$

$$- (\mathbf{x}_2 - \mu_2)' \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} (\mathbf{x}_1 - \mu_1) +$$

$$+ (\mathbf{x}_2 - \mu_2)' \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \Sigma_{12}\Sigma_{22}^{-1} (\mathbf{x}_2 - \mu_2) +$$

$$+ (\mathbf{x}_2 - \mu_2)' \Sigma_{22}^{-1} (\mathbf{x}_2 - \mu_2) \qquad [7]$$

Sustituyendo las expresiones [5] y [7] en [6] observamos que el último factor de [5] y el último sumando de [7] se cancelan, mientras que los restantes pueden reordenarse resultando:

$$f(\mathbf{x}_1|\mathbf{x}_2) =$$

$$= (2\pi)^{p/2} |\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}((\mathbf{x}_1 - \mu_1) - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2))'\right.$$

$$\left.\cdot (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \cdot ((\mathbf{x}_1 - \mu_1) - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2))\right\}$$

Si llamamos

$$\beta = \Sigma_{12}\Sigma_{22}^{-1} \quad \text{y} \quad \Sigma_{11.2} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Por tanto, la variable aleatoria $\xi_1|\xi_2 = \mathbf{x}_2$ tiene una distribución $N[\mu_1 + \beta(\mathbf{x}_2 - \mu_2), \Sigma_{11.2}]$.

El subíndice de la matriz de covarianza es una notación clásica en estadística, el punto hace referencia al hecho de que las variables indicadas con números a la derecha de él, han sido fijadas.

La ecuación

$$(\mathbf{x}_1 - \mu_1) = \Sigma_{12}\Sigma_{22}^{-1} (\mathbf{x}_2 - \mu_2) \qquad [9]$$

representa la línea general de regresión de ξ_1 sobre ξ_2, a la matriz $\Sigma_{12}\Sigma_{22}^{-1}$ se la denomina matriz de regresión. El elemento que ocupa el lugar (i, j) en la matriz de regresión la denominaremos

$$\beta_{ij\cdot p+1,p+2,\ldots,n}.$$

El término «regresión» es debido a Galton. Al estudiar la correlación entre las estaturas de padres e hijos, observo que la altura media de los hijos de padres muy altos o muy bajos era más próxima a la media total que la de sus progenitores. Este fenómeno lo denomino «regresión a la mediocridad».

A los elementos de la matriz de covarianzas de la distribución condicionada, $\Sigma_{11.2}$, los denotaremos por $\sigma_{ij\cdot p+1,p+2,\ldots,n}$, dando a entender, como de costumbre, que las variables ξ_{p+1}, ξ_{p+2}, ..., ξ_n, componentes de ξ_2 han sido fijadas.

Por último el coeficiente de correlación entre ξ_i y ξ_j, calculado a partir de la matriz de covarianzas de la distribución condicionada, se denomina coeficiente de correlación parcial entre ξ_i y ξ_j cuando $\xi_{p+1}, \xi_{p+2}, \ldots, \xi_n$ han sido fijadas, lo representaremos por $\rho_{ij\cdot p+1,\ldots,n}$.

Es decir,

$$\rho_{ij\cdot p+1,p+2,\ldots,n} = \frac{\sigma_{ij\cdot p+1,p+2,\ldots,n}}{\sqrt{\sigma_{jj\cdot p+1,\ldots,n}\ \sigma_{ii\cdot p+1,\ldots,n}}} \qquad [10]$$

Resumiendo, hemos obtenido el siguiente resultado.

Proposición 3.1.

Sea ξ vector aleatorio n-dimensional, con distribución $N[\mu, \Sigma]$, si $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ donde ξ_1 es un vector p-dimensional y ξ_2 es (n − p)-dimensional. La distribución de ξ_1 condicionada a $\xi_2 = x_2$ es normal, de media $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ y matriz de covarianzas $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ donde Σ_{ij} es la matriz de covarianzas de ξ_i y ξ_j (i, j = 1, 2).

Antes de continuar adelante, vamos a profundizar un poco más en el sentido estadístico de la matriz de regresión.

Si el vector aleatorio ξ posee una distribución $N[\mu, \Sigma]$ y disponemos de la información adicional:

$$\xi_{p+1} = x_{p+1}, \xi_{p+2} = x_{p+2}, \ldots, \xi_n = x_n$$

o bien, vectorialmente

$$\xi_2 = x_2$$

la proposición 3.1 contiene la afirmación siguiente: El valor esperado de ξ_1 en estas condiciones es:

$$\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \qquad [11]$$

La variable aleatoria

$$\xi_{1.2} = \xi_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2) \qquad [12]$$

denominada residual, representa las discrepancias entre los verdaderos valores de ξ_1 y el valor esperado. Es fácil comprobar que la variable residual $\xi_{1.2}$ tiene una distribución $N[\mathbf{0}, \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}]$, además verifica

$$E[\xi_{1.2}(\xi_2 - \mu_2)'] = \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = \mathbf{0} \qquad [13]$$

Como la distribución conjunta de ξ_2 y $\xi_{1.2}$ es normal y según [13] son incorreladas, tenemos que (proposición 2.2 (bis)) ambas son independientes. Podemos considerar entonces, el siguiente modelo de previsión de ξ_1.

$$\xi_1 = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2) + \xi_{1.2} \qquad [14]$$

Es decir, conocido que $\xi_2 = \mathbf{x}_2$ nuestra previsión de ξ_1 es $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2)$ y la desviación del verdadero valor, es atribuida a un error aleatorio (variable residual), que independientemente del valor de ξ_2, tiene una distribución $N[\mathbf{0}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}]$.

Naturalmente, nuestro modelo se verá reforzado si, a las anteriores cualidades añadimos que entre todas las previsiones posibles que dependen linealmente de $\mathbf{x}_2$, nuestro modelo tiene una variable residual con dispersión la menor posible, o análogamente, que las correlaciones entre la variable observada y su previsión sean máxima en algún sentido.

En un caso sencillo, cuando consideramos la previsión de una componente de ξ_1 a partir del conocimiento de ξ_2, las condiciones anteriores pueden fácilmente conducirnos a conclusiones aceptables. Este es el problema que abordaremos en la próxima sección.

3.2. COEFICIENTE DE CORRELACION MULTIPLE

Consideremos un vector aleatorio ξ con distribución $N[\mu, \Sigma]$ y una partición de sus componentes en dos vectores ξ_1 y ξ_2. Como ya vimos en el apartado anterior la distribución de ξ_1 condicionada a $\xi_2 = \mathbf{x}_2$ es

$$N[\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}].$$

Podemos suponer sin pérdida de generalidad que $\mu_1 = \mathbf{0}$ y $\mu_2 = \mathbf{0}$. En estas condiciones si ξ_i es una componente de ξ_1, la media de ξ_i condicionada a $\xi_2 = \mathbf{x}_2$ será:

$$E[\xi_i|\mathbf{x}_2] = \beta_i\mathbf{x}_2$$

donde β_i es el vector fila resultante de multiplicar la i-ésima fila de Σ_{12} por la matriz Σ_{22}^{-1}, llamaremos δ_i al vector fila i-ésima de Σ_{12}.

Como caso particular del resultado obtenido en el apartado anterior tenemos

$$E[(\xi_i - \beta_i \xi_2)\, \xi_2'] = \delta_i - \beta_i \Sigma_{22} = 0 \qquad [15]$$

Por tanto, ξ_2 y $(\xi_i - \beta_i \xi_2)$ son independientes.

Continuando con el razonamiento que seguíamos en el apartado anterior, interpretaremos $\xi_i - \beta_i \xi_2$ como la discrepancia entre el valor previsto en función de los datos conocidos, y el valor observado para ξ_i.

Nótese que en este caso la variable aleatoria que representa la discrepancia o el error cometido, es una variable escalar. Supongamos otra previsión lineal de ξ_i basada en el valor que toma ξ_2, cualquier previsión lineal de ξ_i basada en el valor que toma ξ_2, puede ser representada por: $\alpha \xi_2$, donde α es un vector fila, la varianza del error (error cuadrático medio) será:

$$\begin{aligned} E[\xi_i - \alpha \xi_2)^2] &= E[(\xi_i - \beta_i \xi_2)^2] + E[(\alpha - \beta_i)\, \xi_2 \xi_2'(\alpha - \beta_i)'] \\ &+ 2\, E[\xi_i - \beta_i \xi_2)\,(\alpha - \beta_2)\xi_2] = \\ &= (\sigma_{ii} - \delta_i \Sigma_{22}^{-1} \delta_i') + (\beta_i - \alpha)\, \Sigma_{22}(\beta_i - \alpha)' \end{aligned} \qquad [16]$$

donde σ_{ii} es la varianza de ξ_i.

Como Σ_{22} es definida positiva, se verifica que

$$(\beta_i - \alpha)\, \Sigma_{22}(\beta_i - \alpha) \geqslant 0$$

y la igualdad se cumple precisamente si $\beta_i - \alpha = \mathbf{0}$. Por tanto

$$E[(\xi_i - \beta_i \xi_2)^2] \leqslant E[(\xi_i - \alpha \xi_2)^2] \qquad [18]$$

verificándose que β_i es el vector de coeficientes de la función lineal de ξ_2 que mejor aproxima a ξ_i en el sentido de los mínimos cuadrados (la varianza residual es mínima).

Comprobemos que simultáneamente la correlación entre ξ_i y $\beta_i \xi_2$ es máxima.

Como ya se ha demostrado, cualquiera que sea c y α se cumple

$$E[(\xi_i - \beta_i \xi_2)^2] \leqslant E[(\xi_i - c\, \alpha \xi_2)^2] \qquad [19]$$

Desarrollando obtenemos

$$\sigma_{ii} + E[(\beta_i \xi_2)^2] - 2\, E[\xi_i(\beta_i \xi_2)] \leqslant$$

$$\leqslant \sigma_{ii} + c^2\, E[(\alpha \xi_2)^2] - 2c\, E[\xi_i]\, E[\alpha \xi_2]$$

De donde

$$\frac{2E[\xi_i(\beta_i\xi_2)]}{\sqrt{\sigma_{ii}}\sqrt{E[(\beta_i\xi_2)^2]}} - \frac{E[(\beta_i\xi_2)^2]}{\sqrt{\sigma_{ii}}\sqrt{E[(\beta_i\xi_2)^2]}} \geqslant \qquad [21]$$

$$\geqslant \frac{2c\ E[\xi_i(\alpha\,\xi_2)]}{\sqrt{\sigma_{ii}}\sqrt{E[(\beta_i\,\xi_2)^2]}} - \frac{c^2\ E[(\alpha\,\xi_2)^2]}{\sqrt{\sigma_{ii}}\sqrt{E[(\beta_i\,\xi_2)^2]}}$$

Si hacemos

$$c^2 = \frac{E[(\beta_i\,\xi_2)^2]}{E[(\alpha\,\xi_2)^2]} \qquad [22]$$

obtenemos

$$\frac{E[\xi_i(\beta_i\,\xi_2)]}{\sqrt{\sigma_{ii}}\sqrt{E[(\beta_i\,\xi_2)^2]}} \geqslant \frac{E[\xi_i(\alpha\,\xi_2)]}{\sqrt{\sigma_{ii}}\sqrt{E[(\alpha\,\xi_2)^2]}} \qquad [23]$$

de donde resulta que la correlación entre ξ_i y $\beta_i\,\xi_2$ es la mayor entre todas las funciones lineales de ξ_2.

Resumiendo, hemos obtenido:

Proposición 3.2.

Sea $\xi = \begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$ vector aleatorio con distribución normal de media $\mu = \mathbf{0}$ y matriz de covarianzas $\mathbf{\Sigma} = \begin{pmatrix}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22}\end{pmatrix}$, sea ξ_i una componente de ξ_1. la combinación lineal de componentes de ξ_2 que minimiza la varianza de $\xi_i - \alpha\,\xi_2$ y maximiza la correlación entre ξ_i y $\alpha\,\xi_2$ es precisamente $\beta_i\,\xi_2$ donde β_i es igual a $\delta_i\,\mathbf{\Sigma}_{22}^{-1}$ siendo δ_i el i-ésimo vector fila de $\mathbf{\Sigma}_{12}$.

Al coeficiente de correlación entre ξ_i y $\beta_i\,\xi_2$ se le denomina coeficiente de correlación múltiple, lo representaremos por $\rho_{i\cdot p+1,p+2,\ldots,n}$.

De lo anterior se deduce

$$\rho_{i\cdot p+1,p+2,\ldots,n} = \frac{E[\xi_1(\beta_i\,\xi_2)]}{\sqrt{\sigma_{ii}}\ E[(\beta_i\,\xi_2)^2]}$$

$$= \frac{\delta_i\,\mathbf{\Sigma}_{22}^{-1}\,\delta_i'}{\sqrt{\sigma_{ii}}\sqrt{\delta_i\,\mathbf{\Sigma}_{22}^{-1}\,\delta_i'}}$$

$$= \frac{\sqrt{\delta_i\,\mathbf{\Sigma}_{22}^{-1}\,\delta_i'}}{\sqrt{\sigma_{ii}}} \qquad [24]$$

Una relación de interés es la siguiente:

Escribiendo la varianza de la varible $\xi_i|\xi_2 = \mathbf{x}_2$ tenemos

$$\begin{aligned}\sigma_{ii\cdot p+1,p+2,\ldots,n} &= \sigma_{ii} - \boldsymbol{\delta}_i \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\delta}'_i \\ &= \sigma_{ii}\left(1 - \frac{\boldsymbol{\delta}_i \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\delta}'_i}{\sigma_{ii}}\right) \\ &= \sigma_{ii}(1 - \rho^2_{i\cdot p+1,\ldots,n}) \qquad [25]\end{aligned}$$

Así pues, el coeficiente de correlación múltiple mide la proporción de varianza de ξ_i atribuible a la variación de $\xi_{p+1}, \xi_{p+2}, \ldots, \xi_n$. Simultáneamente esta expresión demuestra que la varianza condicional de una componente de ξ no puede superar a su varianza.

Para finalizar el tratamiento de estos temas y antes de entrar en el estudio de la función característica, vamos a considerar cierta relación entre los coeficientes de correlación parcial que resulta de gran utilidad de cara al cálculo efectivo de tales coeficientes si no se dispone de máquinas con gran capacidad de cálculo.

Supongamos un vector aleatorio tridimensional ξ con distribución $N[\boldsymbol{\mu}, \boldsymbol{\Sigma}]$.

La distribución de $\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$ condicionada a $\xi_3 = x_3$ tendrá como matriz de covarianzas

$$\boldsymbol{\Sigma}_{11} - \begin{pmatrix}\sigma_{13}\\ \sigma_{23}\end{pmatrix} \frac{1}{\sigma_{33}} (\sigma_{13}, \sigma_{23}).$$

Por tanto, tenemos

$$\sigma_{12.3} = \sigma_{12} - \frac{\sigma_{13}\sigma_{23}}{\sigma_{33}} \qquad [27]$$

$$\sigma_{11.3} = \sigma_{11} - \frac{\sigma_{13}^2}{\sigma_{33}}$$

$$\sigma_{22.3} = \sigma_{22} - \frac{\sigma_{23}^2}{\sigma_{33}}$$

de donde

$$\rho_{12.3} = \frac{\sigma_{12,3}}{\sqrt{\sigma_{11.3}}\sqrt{\sigma_{22.3}}} \qquad [28]$$

$$= \frac{\sigma_{33}(\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23})}{(\sigma_{11}\sigma_{33} - \sigma_{13}^2)(\sigma_{22}\sigma_{33} - \sigma_{23}^2)} \qquad [29]$$

$$= \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}}$$

Así hemos logrado expresar el coeficiente de correlación parcial de la distribución condicionada a $\xi_3 = x_3$ en función de los coeficientes de correlación de la distribución original, nuestra pretensión es obtener un resultado análogo que sea válido en general.

Consideremos

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}$$

con distribución $N[\mu, \Sigma]$ donde ξ_1, ξ_2, ξ_3 son vectores de dimensión p, q y $n - p - q$ respectivamente. Vamos a calcular la distribución de $(\xi'_1, \xi'_2)'|\xi_3 = x_3$.

Si llamamos

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix}$$

tendremos que la distribución de $\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$ condicionada a $\xi_3 = \mathbf{x}_3$ será normal de media

$$E\left[\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} | \xi_3 = \mathbf{x}_3\right] = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Sigma_{13} \\ \Sigma_{23} \end{pmatrix} \Sigma_{33}^{-1} (\mathbf{x}_3 - \mu_3) \qquad [30]$$

$$= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \Sigma_{13} & \Sigma_{33}^{-1} \\ \Sigma_{23} & \Sigma_{33}^{-1} \end{pmatrix} (\mathbf{x}_3 - \mu_3)$$

$$= \begin{pmatrix} \mu_1 + \Sigma_{13}\Sigma_{33}^{-1}(\mathbf{x}_3 - \mu_3) \\ \mu_2 + \Sigma_{23}\Sigma_{33}^{-1}(\mathbf{x}_3 - \mu_3) \end{pmatrix}$$

y la matriz de covarianzas

$$E\left[\begin{pmatrix} \xi_1 - \mu_1 \\ \xi_2 - \mu_2 \end{pmatrix} ((\xi_1 - \mu_1)', (\xi_2 - \mu_2)')\right] = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} - \begin{pmatrix} \Sigma_{13} \\ \Sigma_{23} \end{pmatrix} \Sigma_{33}^{-1} (\Sigma'_{13}, \Sigma'_{23}) [31]$$

$$= \begin{pmatrix} \Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma'_{13} & \Sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma'_{23} \\ \Sigma_{21} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma'_{13} & \Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma'_{23} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32} \\ \Sigma_{21} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31} & \Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32} \end{pmatrix}$$

A partir de aquí calculamos la distribución de ξ_1 condicionada a $\xi_2 = \mathbf{x}_2$ y $\xi_3 = \mathbf{x}_3$, esta distribución será normal, de media

$$\mu_1 + \Sigma_{13}\Sigma_{33}^{-1}(\mathbf{x}_3 - \mu_3) + (\Sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32})(\Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}(\mathbf{x}_2 - \mu_2) \qquad [32]$$

y matriz de covarianzas

$$(\Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31}) - (\Sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32})(\Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32})^{-1}$$
$$(\Sigma_{21} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31}) \qquad [33]$$

Esta distribución debe coincidir con la distribución condicionada, calculada directamente por el método expuesto en este tema, así pues, la matriz de covarianzas debe ser igual a

$$\Sigma_{11} - (\Sigma_{12}, \Sigma_{13})\begin{pmatrix}\Sigma_{22} & \Sigma_{23}\\ \Sigma_{32} & \Sigma_{33}\end{pmatrix}^{-1}\begin{pmatrix}\Sigma_{21}\\ \Sigma_{31}\end{pmatrix} \qquad [34]$$

Consideremos el caso q = 1, es decir:

$$\xi_1 = \begin{pmatrix}\xi_1\\ \xi_2\\ \cdot\\ \cdot\\ \cdot\\ \xi_p\end{pmatrix} \qquad \xi_2 = \xi_{p+1} \qquad \xi_3 = \begin{pmatrix}\xi_{p+1}\\ \xi_{p+2}\\ \cdot\\ \cdot\\ \cdot\\ \xi_p\end{pmatrix}$$

El elemento (i, j) de la expresión [33] es precisamente $\sigma_{ij\cdot p+1,p+2,\dots,n}$, por otra parte, el elemento (i, j) de la expresión [31] es $\sigma_{ij\cdot p+2,p+3,\dots,n}$. Por tanto,

$$\sigma_{ij\cdot p+1,p+2,\dots,n} = \sigma_{ij\cdot p+2,\dots,n} - \frac{\sigma_{ip+1\cdot p+2,\dots,n}\,\sigma_{jp+1\cdot p+2,\dots,n}}{\sigma_{p+1p+1\cdot p+2,\dots,n}} \qquad [35]$$

Expresión que en particular implica

$$\sigma_{ii\cdot p+1,p+2,\dots,n} = \sigma_{ii\cdot p+1,\dots,n}(1 - \rho^2_{ip+1\cdot p+2,\dots,n}) \qquad [36]$$

Así podemos obtener

$$\rho_{ij\cdot p+1,p+2,\dots,n} = \frac{\sigma_{ij\cdot p+1,\dots,n}}{\sqrt{\sigma_{ii\cdot p+1,\dots,n}\,\sigma_{jj\cdot p+1,\dots,n}}} \qquad [37]$$

$$= \frac{\rho_{ij\cdot p+2,\dots,n} - \rho_{ip+1\cdot p+2,\dots,n}\,\rho_{jp+1\cdot p+2,\dots,n}}{\sqrt{1 - \rho^2_{ip+1\cdot p+2,\dots,n}}\sqrt{1 - \rho^2_{jp+1\cdot p+2,\dots,n}}}$$

3.3. FUNCION CARACTERISTICA DE LA DISTRIBUCION NORMAL n-DIMENSIONAL

Para finalizar vamos a incluir el cálculo de la función característica de la distribución normal n-dimensional.

El procedimiento de operar es el ordinario, efectuar un cambio lineal que reduzca el vector aleatorio ξ a otro η cuyas componentes sean independientes.

Dada la matriz de covarianzas $\mathbf{\Sigma}$ podemos encontrar una matriz $\mathbf{C}$ no singular tal que

$$\mathbf{C}' \mathbf{\Sigma}^{-1} \mathbf{C} = \mathbf{I} \qquad [38]$$

Si efectuamos el cambio

$$\xi - \mu = \mathbf{C}\eta$$

resulta que η se distribuye según una $N[\mathbf{0}, \mathbf{I}]$.

Por tanto

$$\begin{aligned} \varnothing(\mathbf{u}) &= E[e^{iu'\eta}] \qquad [39] \\ &= \prod_{j=1}^{n} E[e^{iu_j y_j}] \\ &= \prod_{j=1}^{n} e^{-\frac{1}{2}u_j^2} \\ &= e^{-\frac{1}{2}u'u} \end{aligned}$$

Como $\xi - \mu = \mathbf{C}\eta$ se tiene que

$$\begin{aligned} \varphi(\mathbf{t}) &= E[e^{it'\xi}] \qquad [40] \\ &= e^{it'\mu} \cdot \varnothing(\mathbf{tC'}) \\ &= e^{it'\mu - \frac{1}{2}t'CC't} \end{aligned}$$

Como tenemos $\mathbf{C}'\mathbf{\Sigma}^{-1}\mathbf{C} = \mathbf{I}$ se verifica $\mathbf{\Sigma}^{-1} = (\mathbf{CC}')^{-1}$, por tanto,

$$\varphi(\mathbf{t}) = e^{it'\mu - \frac{1}{2}t'\Sigma't} \qquad [41]$$

Un bonito ejemplo de la importancia y utilidad de la función característica lo encontrará el lector en los ejercicios de autocomprobación.

EJERCICIOS DE AUTOCOMPROBACION

1. Proponer un ejemplo de una variable aleatoria bidimensional (ξ_1, ξ_2) que no sea normal bidimensional y, sin embargo, sus distribuciones marginales si lo sean.

2. Encontrar una distribución bidimensional tal que el coeficiente de correlación sea cero y las distribuciones marginales normales pero no independientes (véase proposición 2.2 y 2.2 (bis)).

3. Demuéstrese que $\rho_{i \cdot p+1, p+2, \dots, n}$ es invariante respecto de las transformaciones escalares de las componentes de ξ (llamaremos transformación escalar de las componentes a una aplicación φ tal que:

$$(\xi_1, \xi_2, \dots, \xi_n) \xrightarrow{\varphi} (c_1 \xi_1, c_2 \xi_2, \dots, c_n \xi_n) \qquad c_i \neq 0 \; 1 \leqslant i \leqslant n$$

4. Demostrar la forma de Σ^{-1} que aparece en la expresión [4].

5. Demostrar el siguiente resultado (que ha sido utilizado en el texto). Si Σ es una matriz definida positiva, entonces se cumple

$$|\Sigma| = |\Sigma_{22}| \cdot |\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}|$$

6. Utilizar la demostración del problema anterior, para dar una prueba muy sencilla de la forma de la distribución condicionada de una normal n-dimensional.

7. Demostrar que $|\Sigma| \leqslant \prod_{i=1}^{n} \sigma_{ii}$ siendo Σ una matriz definida positiva.

8. Sean $\xi_1, \xi_2, \ldots, \xi_n$ variables aleatorias independientes con distribución $N[\beta + \gamma z_i, \sigma^2]$ donde los z_i son tales que $\sum_{i=1}^{n} z_i = 0$. Se pide

   a) Calcular la distribución conjunta de $(\xi_1, \xi_2, \ldots, \xi_n)$.

   b) Calcular la distribución conjunta de

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^{n} \xi_i \quad \text{y} \quad \eta = \frac{\Sigma\, z_i \xi_i}{\Sigma\, z_i^2}$$

9. Sean $\xi_1, \xi_2, \ldots, \xi_n$ variables aleatorias independientes e igualmente distribuidas con función de densidad f(x) y tales que la función de densidad conjunta depende exclusivamente de $x_1^2 + x_2^2 + \ldots + x_n^2$. Probar que las variables aleatorias ξ_i son normales de media cero e igual varianza.

10. Sea ξ un vector aleatorio, probar que si la distribución de cualquier combinación lineal de componentes es normal entonces la distribución de ξ es normal n-dimensional.

SOLUCIONES A LOS EJERCICIOS DE AUTOCOMPROBACION

1. Ofrecemos en primer lugar un elegante ejemplo de E. Nelson.

Sea u(x) una función impar y continua tal que

(1) $u(x) = 0 \quad x \in [-1, 1]^c$

(2) $|u(x)| < (2\pi e)^{-1/2} \quad x \in [-1, 1]$

Entonces

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} + u(x) \cdot u(y)$$

es la función de densidad de una variable bi-dimensional (ξ_1, ξ_2). En efecto, como

$$\min_{(x,y)\in[-1,1]\times[-1,1]} \left\{\frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\right\} = \frac{1}{2\pi} e^{-1}$$

tenemos que $f(x, y) \geqslant 0 \quad \forall x \in \mathbb{R}, \quad \forall y \in \mathbb{R}$. Por otra parte,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)}\, dx\, dy + \int_{-1}^{1} \int_{-1}^{1} u(x)\, u(y)\, dx\, dy$$

$$= 1 + \left\{\int_{-1}^{1} u(t)\, dt\right\}^2$$

pero por ser u(t) impar

$$\int_{-1}^{1} u(t)\,dt = 0$$

de donde resulta

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$$

Las distribuciones marginales son normales.

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\,dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\,dy + \int_{-1}^{1} u(x)\,u(y)\,dy$$

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

Por simetría

$$f_2(y) = f_1(y)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

Un ejemplo más sencillo es el siguiente:

Consideremos la función de densidad de la distribución

$$N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right]$$

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \qquad \begin{array}{l} -\infty < x < \infty \\ -\infty < y < \infty \end{array}$$

Duplicamos la densidad de probabilidad en el 1.^{er} y 3.^{er} cuadrante a costa del 2.° y 4.° (figura 1 y 2).

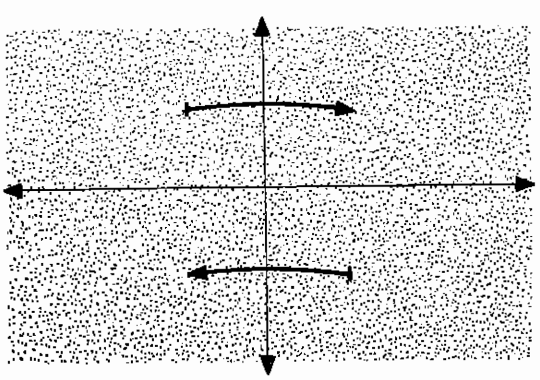

Figura 1

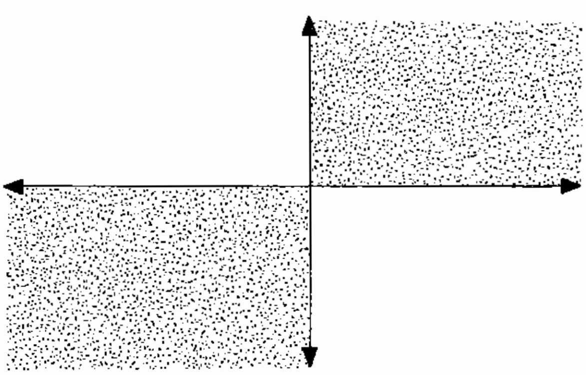

Figura 2

Es decir, definimos

$g(0, y) = f(0, y) \qquad \forall y \in \mathbb{R}$

$g(x, y) = f(x, y) + f(x, -y) \quad$ si $\quad (x > 0$ y $> 0)$ ó $(x < 0$ y $< 0)$

$g(x, y) = 0 \quad$ si $\quad (x > 0$ y $< 0)$ ó $(x < 0$ y $> 0)$

$g(x, 0) = f(x, 0) \qquad \forall x \in \mathbb{R}$

g(x, y) es función de densidad de una variable aleatoria que no es normal bidimensional, sin embargo, verifica:

$$\text{Si } x > 0 \;\; g_1(x) = \int_{-\infty}^{\infty} g(x, y)\, dy$$

$$= \int_{0}^{\infty} (f(x, y) + f(x, -y))\, dy$$

$$= \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$$

Si $x < 0$ es análogo.

2. Propongo el siguiente ejemplo que aprendí de I. Yáñez. Consideremos la función de densidad normal bi-dimensional

$$f(x, y) = f_1(x)\, f_1(y)$$

siendo

$$f_1(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad -\infty < x < \infty$$

Construimos una nueva función de densidad de cierta variable (ξ^*, η^*) de la siguiente manera

$g(x, y) = f(x, y) + f(x, -y) \quad$ si $\quad x \in [-\lambda_0, \lambda_0] \;\;$ y < 0

$g(x, y) = f(x, y) + f(x, -y) \quad$ si $\quad x \notin [-\lambda_0, \lambda_0] \;\;$ y > 0

$g(x, 0) = f(x, 0) \quad$ si $\quad x \in \mathbb{R}$

Donde λ_0 es el único número real que verifica:

$$\int_{0}^{\lambda_0} f_1(x)\, dx = \frac{1}{4}$$

La comprobación de que g(x, y) es función de densidad es inmediata, comprobemos que las marginales son normales

$$\text{Si } y > 0 \quad g_2(y) = \int_{-\infty}^{-\lambda_0} (f(x, y) + f(x, -y))\, dx + \int_{\lambda_0}^{\infty} (f(x, y) + f(x, -y)\, dx$$

como $f(x, y) = f(y, x)$

$$= f(x, -y)$$

$$= f(-y, x)$$

$$g_2(y) = 2\left\{\int_{-\infty}^{-\lambda_0} f(x, y)\, dx + \int_{\lambda_0}^{\infty} f(x, y)\, dx\right\}$$

$$= 2f_1(y)\left\{\int_{-\infty}^{-\lambda_0} f_1(x)\, dx + \int_{\lambda_0}^{\infty} f_1(x)\, dx\right\}$$

$$= f_1(y)$$

Un razonamiento idéntico se aplica si $y < 0$. De igual manera obtenemos
$$g_1(x) = f_1(x)$$

Ahora, teniendo en cuenta que

$$E[\xi^*|\eta^* = y] = \int_{-\infty}^{-\lambda_0} 2x\, f_1(x)\, dx + \int_{\lambda_0}^{\infty} 2x\, f_1(x)\, dx$$

$$= 0 \qquad\qquad \text{si} \quad y < 0$$

y

$$E[\xi^*|\eta^* = y] = \frac{1}{f_1(0)} \int_{-\infty}^{\infty} x\, f_1(x)\, dx = 0 \qquad \text{si} \qquad y = 0$$

y

$$E[\xi^*|\eta^* = y] = \int_{-\lambda_0}^{\lambda_0} 2x\, f_1(x)\, dx = 0$$

Obtenemos que la línea general de regresión de ξ^* sobre η^* es $x = 0$, por tanto, coincide con la línea de regresión

$$(x - \mu_x) = \rho \frac{\sigma_y}{\sigma_x} (y - \mu_y)$$

por consiguiente $\rho = 0$, sin embargo, es claro que ξ^* y η^* no son independientes.

Nótese que este ejemplo no contradice en absoluto las afirmaciones de la proposición 2.2. (bis).

3. Sin pérdida de generalidad, podemos suponer a ξ centrada en la media. Por definición de coeficiente de correlación múltiple tenemos

$$\rho_{i \cdot p+1, p+2, \dots, n} = \frac{E[\xi_i(\alpha\, \xi_2)]}{\sqrt{E[\xi_i^2]\, E[\alpha\, \xi_2\, \xi_2'\, \alpha']}}$$

Donde $\xi_2' = (\xi_{p+1}, \xi_{p+2}, \dots, \xi_n)$ y α es un vector fila (n-p)-dimensional que minimiza la expresión

$$E[(\xi_i - \alpha\, \xi_2)^2]$$

Definamos:

$$\eta_1 = c_1\, \xi_1, \qquad \eta_2 = c_2\, \xi_2, \ \dots, \ \eta_n = c_n\, \xi_n$$

supongamos β vector fila (n-p)-dimensional tal que $E[(\eta_i - \beta\eta_2)^2]$ es mínima, por consiguiente

$$E[(c_i\, \xi_i - \eta_2)^2] = c_i^2\, E\left[\left(\xi_i - \frac{1}{c_i}\, \beta\eta_2\right)^2\right]$$

es mínima, de donde resulta que

$$E\left[\left(\xi_i - \frac{1}{c_i}\, \beta\eta_2\right)^2\right]$$

es también mínima y, por tanto, se cumplen

$$E[(\eta_i - \beta\eta_2)^2] = c_i^2\, E[(\xi_i - \alpha\, \xi_2)^2]$$

$$\left(\beta_{p+1} \frac{C_{p+1}}{C_1}, \beta_{p+2} \frac{C_{p+2}}{C_1}, \dots, \beta_n \frac{C_n}{C_1}\right) \equiv (\alpha_{p+1}, \alpha_{p+2}, \dots, \alpha_n)$$

de aquí obtenemos inmediatamente

$$\begin{aligned} E[\eta_i(\beta\eta_2)] &= c_i^2\, E[\xi_i(\alpha\, \xi_2)] \\ E[\eta_i\, \eta_i] &= c_i^2\, E[\xi_i\, \xi_i] \\ E[\beta\eta_2\eta_2'\beta'] &= c_i^2\, E[\alpha\, \xi_2\, \xi_2'\, \alpha'] \end{aligned}$$

por lo cual

$$\rho_{i \cdot p+1, p+2, \dots, n} = \frac{E[\xi_i(\alpha\, \xi_2)]}{\sqrt{E[\xi_i\, \xi_i]\, E[\alpha\, \xi_2\, \xi_2'\, \alpha']}}$$

$$= \frac{E[\eta_i(\beta\eta_2)]}{\sqrt{E[\eta_i\eta_i]\, E[\beta\eta_2\eta_2'\, \beta']}}$$

4. Consideremos $\mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$ donde $\mathbf{A}_{ij}$ i, j = 1, 2 son submatrices a determinar.

Como $\mathbf{\Sigma}\mathbf{\Sigma}^{-1} = \mathbf{I}$ tenemos el conjunto de igualdades

(1) $\mathbf{\Sigma}_{11}\mathbf{A}_{11} + \mathbf{\Sigma}_{12}\mathbf{A}_{21} = \mathbf{I}$

(2) $\mathbf{\Sigma}_{11}\mathbf{A}_{12} + \mathbf{\Sigma}_{12}\mathbf{A}_{22} = \mathbf{0}$

(3) $\mathbf{\Sigma}_{21}\mathbf{A}_{11} + \mathbf{\Sigma}_{22}\mathbf{A}_{21} = \mathbf{0}$

(4) $\mathbf{\Sigma}_{21}\mathbf{A}_{12} + \mathbf{\Sigma}_{22}\mathbf{A}_{22} = \mathbf{I}$

De (3) obtenemos $\mathbf{A}_{21} = \mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{A}_{11}$

sustituyendo en (1) $\mathbf{A}_{11} = (\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})^{-1}$

por tanto $\mathbf{A}_{21} = \mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})$

Si hicieramos uso de las igualdades (2) y (4) aparecerían términos en $\mathbf{\Sigma}_{11}^{-1}$ que no son cσnvenientes para el propósito actual de la matriz inversa, a saber, el cálculo de la distribución condicionada. Por ello aplicando $\mathbf{\Sigma}^{-1}\mathbf{\Sigma} = \mathbf{I}$ obtenemos

(1)' $\mathbf{A}_{11}\mathbf{\Sigma}_{11} + \mathbf{A}_{12}\mathbf{\Sigma}_{21} = \mathbf{I}$

(2)' $\mathbf{A}_{11}\mathbf{\Sigma}_{12} + \mathbf{A}_{12}\mathbf{\Sigma}_{22} = \mathbf{0}$

(3)' $\mathbf{A}_{21}\mathbf{\Sigma}_{11} + \mathbf{A}_{22}\mathbf{\Sigma}_{21} = \mathbf{0}$

(4)' $\mathbf{A}_{21}\mathbf{\Sigma}_{12} + \mathbf{A}_{22}\mathbf{\Sigma}_{22} = \mathbf{I}$

De donde se deduce inmediatamente

$$\mathbf{A}_{12} = (\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}$$

$$\mathbf{A}_{22} = \mathbf{\Sigma}_{22}^{-1} + \mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}$$

5. En primer lugar, nótese que por ser $\mathbf{\Sigma}$ definida positiva, son $\mathbf{\Sigma}_{11}$ y $\mathbf{\Sigma}_{22}$ no singulares.

Consideremos la matriz

$$\mathbf{C} = \begin{pmatrix} \mathbf{I} & -\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

Evidentemente se tiene que $|\mathbf{C}| = 1$.

Como

$$\mathbf{C}\,\mathbf{\Sigma}\,\mathbf{C}' = \begin{pmatrix} \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22} \end{pmatrix}$$