¿Qué ocurre cuando el rg$(\mathbf{X}) < p$

Si $\text{rg}(\mathbf{X}) = r < p$ tenemos que $\widehat{\beta}_1, \ldots, \widehat{\beta}_p$ no son únicos. Esta misma indeterminación afecta a los parámetros $\beta_1, \ldots, \beta_p$, en el sentido de que distintos conjuntos $b_1, \ldots, b_p$ darían origen al mismo η y por lo tanto al mismo modelo

$$\mathbf{Y} = \eta + \boldsymbol{\epsilon} = E(\mathbf{Y}) + \boldsymbol{\epsilon}\,.$$

Sin embargo, tal como vimos si $\mathbf{c}'\boldsymbol{\beta}$ es una función estimable tendrá el mismo valor independientente del $\boldsymbol{\beta}$ que usemos, en tanto

$$\mathbf{c}'\boldsymbol{\beta} = \mathbf{a}'\mathbf{X}\boldsymbol{\beta} = \mathbf{a}'\eta$$

expresión que sólo depende de η que es único.

¿Cómo podemos eliminar esta indeterminación?

a) Considerar un problema reducido con sólo r parámetros

Podríamos considerar r columnas l.i. de $\mathbf{X}$ que generen a $\mathcal{V}_r$ y mantener en el modelo sólo aquellos β_j asociados.

Así tendríamos una nueva matriz de diseño $\mathbf{X}_1 \in \Re^{n\times r}$ con rango máximo. En este caso tendríamos el modelo

$$\mathbf{Y} = \eta + \boldsymbol{\epsilon} \qquad \text{con } \eta \in \mathcal{V}_r$$

El estimador sería

$$\boldsymbol{\alpha} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}$$

y la matriz de proyección corresponiente $\mathbf{P} = \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$.

Si asumimos, s.p.g., que las columnas elegidas son las primeras r, tendríamos que

$$\mathbf{X} = [\mathbf{X}_1\mathbf{X}_2]$$

donde $\mathbf{X}_2 \in \Re^{n\times(p-r)}$ y además $\mathbf{X}_2 = \mathbf{X}_1\mathbf{B}$. Por lo tanto

$$\mathbf{X} = \mathbf{X}_1[\mathbf{I}_r \quad \mathbf{B}] = \mathbf{K}\mathbf{L}$$

con $\mathbf{K} \in \Re^{n\times r}, \quad \mathbf{L} \in \Re^{r\times p}$ y $\text{rg}(\mathbf{L}) = r$.

Por lo tanto el modelo original se obtiene como:

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{K}\mathbf{L}\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\alpha}$$

b) Considerar condiciones de contorno adecuadas para los β_j's y sus estimadores

Así podríamos pedir que $\boldsymbol{\beta}_{r+1} = \ldots = \boldsymbol{\beta}_p = 0$ y en este caso obtendríamos el mismo que en la situación a) (suponiendo que las r primeras son las columnas l.i.).

Sin embargo, en otras situaciones, como en la de ANOVA, es frecuente que se impongan otras restricciones lineales de manera de obtener la unicidad.

Consideremos el caso en que imponemos $t \geq p - r$ restricciones lineales a los $\boldsymbol{\beta}_j$, es decir

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{0} \quad \text{con } \mathbf{H} \in \Re^{\mathbf{t}\times\mathbf{p}}$$

Queremos encontrar dentro del conjunto de soluciones de $\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\eta}$ una sola que cumpla $\mathbf{H}\boldsymbol{\beta} = \mathbf{0}$, es decir buscamos $\tilde{\boldsymbol{\beta}}$ que sea única solución de

$$\begin{aligned} \mathbf{X}\tilde{\boldsymbol{\beta}} &= \mathbf{X}\boldsymbol{\beta} \quad (= \boldsymbol{\eta}) \\ \mathbf{H}\tilde{\boldsymbol{\beta}} &= \mathbf{0} \end{aligned}$$

De esta forma las primeras ecuaciones establecen que encontraremos una solución del sistema que nos interesa y las segundas que esta solución será única.

Lo que queremos es que

- toda función estimable del nuevo sistema lo sea en el viejo problema,
- un único conjunto de estimadores de mínimos cuadrados que satisfaga las condiciones de contorno.

El siguiente teorema nos dice como elegir $\mathbf{H}$ para cumplir con este objetivo:

Teorema: Sean $\mathbf{X} \in \Re^{n\times p}$ y $\mathbf{H} \in \Re^{t\times p}$ con $rg(\mathbf{X}) = r$, $p > r$ y $t \geq p - r$. Consideremos $\mathcal{V}_{\mathbf{X}}$ el espacio generado por las columnas de $\mathbf{X}$. El sistema

$$\begin{aligned} \mathbf{Xb} &= \mathbf{z} \\ \mathbf{Hb} &= \mathbf{0} \end{aligned}$$

tiene solución única $\mathbf{b}$ para todo $\mathbf{z} \in \mathcal{V}_{\mathbf{X}}$ si y sólo si se cumplen las siguientes dos condiciones:

i) si $\mathrm{rg}(\mathbf{G}) = \mathrm{rg}\begin{pmatrix} \mathbf{X} \\ \mathbf{H} \end{pmatrix} = p$

ii) ninguna combinación lineal de las filas de $\mathbf{H}$ es combinación lineal de las de $\mathbf{X}$, excepto el $\mathbf{0}$.

Observación:

- la condición ii) del Teorema nos dice que si $\mathbf{h}_i$ es la iésima fila de $\mathbf{H}$, entonces no existe $\mathbf{a}$ tal que $\mathbf{h}_i = \mathbf{a}'\mathbf{X}$, por lo tanto las $\mathbf{h}_i'\boldsymbol{\beta}$ no es una función estimable de los parámetros.

Se puede demostrar que:

- Si se cumplen las condiciones i) y ii) del Teorema, entonces los $\tilde{\beta}_j$ son funciones estimables.
- dada una función estimable ψ, para cualquier $\mathbf{H}$ que elijamos en las condiciones del Teorema anterior, $Var(\widehat{\psi})$ es la misma.

c) Computar una inversa generalizada de $\mathbf{X}'\mathbf{X}$: $(\mathbf{X}'\mathbf{X})^-$

En este caso tendríamos que $(\mathbf{X}'\mathbf{X})^-\mathbf{XY}$ es solución de las ecuaciones normales, por lo tanto otra forma de solucionar nuestro problema. En realidad puede verse que la opción b) y c) quedan ligadas a través del siguiente resultado:

Proposición: Sea $\mathbf{G} = \begin{pmatrix} \mathbf{X} \\ \mathbf{H} \end{pmatrix}$ una matriz que satisface las condiciones del Teorema anterior. Luego $(\mathbf{G}'\mathbf{G})^{-1}$ es una inversa generalizada de $\mathbf{X}'\mathbf{X}$, por lo tanto:

$$(\mathbf{X}'\mathbf{X})(\mathbf{G}'\mathbf{G})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{X}'\mathbf{X}$$

Mínimos Cuadrados Pesados y Mínimos Cuadrados Generalizados

¿ Qué ocurre cuando $\Sigma_{\mathbf{Y}} = \sigma^2 \mathbf{V}$ donde $\mathbf{V} \neq \mathbf{I}$?

Supongamos que $\mathbf{V} \in \Re^{n \times n}$ es una matriz definida positiva de constantes. Podemos entonces escribir: $\mathbf{V} = \mathbf{K}\mathbf{K}'$ con $\mathbf{K}$ una matriz invertible.

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \mathbf{K}^{-1}\mathbf{Y} &= \mathbf{K}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{K}^{-1}\boldsymbol{\epsilon} \end{aligned}$$

donde $E(\mathbf{K}^{-1}\boldsymbol{\epsilon}) = \mathbf{0}$ y $\Sigma_{\mathbf{K}^{-1}\boldsymbol{\epsilon}} = \sigma^2 \mathbf{I}$.

Por lo tanto, tenemos un nuevo problema:

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}}$$

que satisface las condiciones de Ω.

Hallar el estimador de mínimos cuadrados en el problema transformado equivale a:

$$
\begin{aligned}
\min_{\mathbf{b}} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{b}\|^2 &= \min_{\mathbf{b}} (\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{b})'(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\mathbf{b}) \\
&= \min_{\mathbf{b}} (\mathbf{Y} - \mathbf{X}\mathbf{b})'\mathbf{K}^{-1\prime}\mathbf{K}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{b}) \\
&= \min_{\mathbf{b}} (\mathbf{Y} - \mathbf{X}\mathbf{b})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{b})
\end{aligned}
$$

Si $\mathbf{V}$ es una matriz diagonal decimos que tenemos un problema de Mínimos Cuadrados Pesados, mientras que si $\mathbf{V}$ es una matriz definida positiva cualquiera, es de Mínimos Cuadrados Generalizados.

Las ecuaciones normales quedan:

$$
\begin{aligned}
\tilde{\mathbf{X}}\tilde{\mathbf{X}}'\mathbf{b} &= \tilde{\mathbf{X}}\tilde{\mathbf{Y}} \\
\mathbf{X}'\mathbf{K}^{-1\prime}\mathbf{K}^{-1}\mathbf{X}\mathbf{b} &= \mathbf{X}'\mathbf{K}^{-1\prime}\mathbf{K}^{-1}\mathbf{Y} \\
\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{b} &= \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}
\end{aligned}
$$

Observemos que si $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ tiene inversa, entonces

$$\tilde{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

y además

- $\tilde{\boldsymbol{\beta}}$ es un estimador insesgado de $\boldsymbol{\beta}$, es decir $E(\tilde{\boldsymbol{\beta}}) = \beta$.
- $\Sigma_{\tilde{\boldsymbol{\beta}}_e} = \sigma^2(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1} = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$

Veamos un ejemplo.

Consideremos el caso sencillo de una regresión simple por el origen:

$$\mathbf{Y} = \mathbf{x}\beta + \boldsymbol{\epsilon}$$

donde $\mathbf{Y} = (y_1, \ldots, y_n)'$, $\mathbf{x} = (x_1, \ldots, x_n)'$ y $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ con $E(\boldsymbol{\epsilon}) = \mathbf{0}$ y $\Sigma_{\boldsymbol{\epsilon}} = \sigma^2\mathbf{V} = \sigma^2 diag(w_1, \ldots, w_n)$ con $w_i > 0$.

Probaremos que

$$\boxed{\tilde{\boldsymbol{\beta}} = \frac{\sum_{i=1}^n y_i x_i / w_i}{\sum_{i=1}^n x_i^2 / w_i}}$$

y además

$$\boxed{\Sigma_{\tilde{\boldsymbol{\beta}}} = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2/w_i}}$$

Si $\mathrm{rg}(\mathbf{X}) = p$ se puede probar fácilmente que el estimador $\tilde{\boldsymbol{\beta}}$ conserva las propiedades del estimador de mínimos cuadrados: dada una función lineal estimable $\mathbf{c}'\boldsymbol{\beta}$ tenemos que

- $\mathbf{c}'\tilde{\boldsymbol{\beta}}$ es el estimador lineal insesgado de $\mathbf{c}'\boldsymbol{\beta}$ de menor varianza.

Una pregunta muy natural es:

¿ Hay situaciones en las que $\tilde{\beta}$ y $\widehat{\beta}$ coinciden?

Los siguientes resultados nos dan la respuesta

Teorema: Una condición necesaria y suficiente para que $\tilde{\boldsymbol{\beta}}$ y $\widehat{\boldsymbol{\beta}}$ coincidan es que $\mathcal{V}_{\mathbf{V}^{-1}\mathbf{X}} = \mathcal{V}_{\mathbf{X}}$.

Corolario: $\tilde{\boldsymbol{\beta}}$ y $\widehat{\boldsymbol{\beta}}$ coinciden $\Longleftrightarrow \mathcal{V}_{\mathbf{V}\mathbf{X}} = \mathcal{V}_{\mathbf{X}}$.

Corolario: Si tenemos un modelo de regresión simple por el origen, $\mathbf{Y} = \mathbf{x}\beta + \boldsymbol{\epsilon}$, entonces

$$\tilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} \quad \forall \mathbf{x} \Longleftrightarrow \mathbf{V} = c\mathbf{I}_n$$

Forma Canónica del Modelo Ω

Dada una base ortonormal de $\mathcal{V}_r = \mathcal{V}_{\mathbf{X}}$, digamos $\{\boldsymbol{\alpha}_1, \ldots \boldsymbol{\alpha}_r\}$, sabemos que podemos extenderla a una base ortonormal de $\Re^n$: $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+1}, \ldots, \boldsymbol{\alpha}_n\}$.

Por lo tanto,

$$\mathbf{y} \in \Re^n : \mathbf{y} = \sum_{j=1}^{n} z_j \boldsymbol{\alpha}_j$$

y tenemos que

$$\boldsymbol{\alpha}_i' \mathbf{y} = \sum_{j=1}^{n} z_j \boldsymbol{\alpha}_i' \boldsymbol{\alpha}_j = z_i \boldsymbol{\alpha}_i' \boldsymbol{\alpha}_i = z_i \quad \forall i = 1, \ldots, n$$

Luego, si definimos a $\mathbf{T}$ como la matriz que tiene filas $\boldsymbol{\alpha}_i'$, entonces

$$\mathbf{z} = \mathbf{T}\mathbf{y}$$

Observemos que

$$E(z_i) = \begin{cases} \boldsymbol{\alpha}_i' \boldsymbol{\eta} = \xi_i & \text{si } 1 \leq i \leq r \\ 0 & \text{si } r+1 \leq i \leq n \end{cases}$$

$$\Sigma_{\mathbf{z}} = \mathbf{T} \Sigma_{\mathbf{y}} \mathbf{T}' = \sigma^2 \mathbf{I}$$

Por lo tanto, ahora podemos reescribir a Ω como

$$\boxed{\begin{aligned} &\Omega : \\ &\qquad E(z_i) = \begin{cases} \xi_i & \text{si } 1 \leq i \leq r \\ 0 & \text{si } r+1 \leq i \leq n \end{cases} \\ &\qquad \Sigma_{\mathbf{z}} = \sigma^2 \mathbf{I} \end{aligned}}$$

donde $\boldsymbol{\xi}$ y σ^2 son parámetros desconocidos.

En términos de esta forma caónica es sencillo demostrar que

$$\boxed{s^2 = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2}{n-r} == \frac{\|\mathbf{Y} - \widehat{\boldsymbol{\eta}}\|^2}{n-r}}$$

es un estimador insesgado de σ^2. Sólo habíamos demostrado hasta ahora el caso de rango completo.

Distribución Normal Multivariada

Definición 1: Se dice que un vector $\mathbf{X}$, $k-$dimensional tiene distribución normal multivariada $N_k(\boldsymbol{\mu}, \mathbf{Q})$ donde $\boldsymbol{\mu}$ es un vector $k-$dimensional, $\mathbf{Q}$ una matriz de $k \times k$ definida positiva, si su densidad es de la forma

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^k |\mathbf{Q}|^{1/2}} \, e^{-\frac{(\mathbf{x}-\boldsymbol{\mu})'\mathbf{Q}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}}$$

donde $|\mathbf{Q}|$ indica determinante de $\mathbf{Q}$.

Por ejemplo, si X_i son k v.a. independientes tales que $X_i \sim N(\mu_i, \sigma_i^2)$, entonces el vector $\mathbf{X}' = (X_1, \ldots, X_k)$ tiene densidad

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^k \, \Pi_{j=1}^k (\sigma_i^2)^{1/2}} \, e^{-1/2\{\Sigma_{i=1}^k (x_i-\mu_i)^2/\sigma_i^2\}}$$

Luego, resulta que $\mathbf{X}$ es $N_k(\boldsymbol{\mu}, \mathbf{Q})$ donde $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_k)$ y

$$\mathbf{Q} = diag(\sigma_1^2, \ldots, \sigma_k^2) = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_k^2 \end{pmatrix}$$

Más aún, en el caso en que las k v.a. X_i son todas $N(0,1)$, $\mathbf{X}$ es $N(\mathbf{0_k}, \mathbf{I_k})$ donde $\mathbf{0_k'} = (\mathbf{0}, \ldots, \mathbf{0}) \in \Re^{\mathbf{k}}$ y $\mathbf{I}_k$ es la matriz identidad de $k \times k$.

Recordemos el enunciado del Teorema de Cambio de Variable:

Sean $\mathbf{x}$ es un vector aleatorio con densidad f y $\mathbf{y} = g(\mathbf{x})$, tal que $g^{-1}(\mathbf{y}) = \mathbf{x}$.

Supongamos que en un abierto $\mathcal{G}$ existen las derivadas parciales $\dfrac{\partial x_i}{\partial y_j}$ y sea

$J = det\left\{\dfrac{\partial x_i}{\partial y_j}\right\}$, entonces

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y}))|J|$$

Teorema N1: Si $\mathbf{X}$ es un vector aleatorio $k-$dimensional con distribución $N_k(\boldsymbol{\mu}, \mathbf{Q})$, $\mathbf{A}$ es una matriz no singular de $k \times k$ y $\mathbf{b}$ un vector $k-$dimensional, entonces

$$\mathbf{Y} = \mathbf{AX} + \mathbf{b} \quad \text{es} \quad N_k(\mathbf{A}\boldsymbol{\mu} + \mathbf{b},\ \mathbf{AQA}')$$

Teorema N2:

i) Un vector aleatorio $k-$dimensional $\mathbf{X}$ es $N_k(\boldsymbol{\mu}, \mathbf{Q})$ si y sólo si $\mathbf{X} = \mathbf{BY} + \boldsymbol{\mu}$, donde $\mathbf{Y}$ es $N_k(\mathbf{0_k}, \mathbf{I_k})$ y $\mathbf{B}$ es una matriz de $k \times k$ no singular tal que $\mathbf{BB}' = \mathbf{Q}$.

ii) Si $\mathbf{X}$ es $N_k(\boldsymbol{\mu}, \mathbf{Q})$ entonces

$$E(\mathbf{X}) = \boldsymbol{\mu} \quad \text{y} \quad \Sigma_{\mathbf{X}} = \mathbf{Q}$$

Teorema N3: Sea $\mathbf{X}$ un vector aleatorio $k-$dimensional $N_k(\boldsymbol{\mu}, \mathbf{Q})$ y $\mathbf{A}$ una matriz de $h \times k$ con rango h, luego si $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ entonces

$$\mathbf{Y} \sim N_h(A\boldsymbol{\mu} + \mathbf{b}, \mathbf{AQA}')$$