Statistical Models 2021

Irving Gómez Méndez

January 22, 2021

CIMAT

Let be

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \right)$$

Then

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}\left(\mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \mu_x),\, \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \right)$$

Proof:

Consider the matrix

$$A = \begin{bmatrix} I & -\Sigma_{yx}\Sigma_{xx}^{-1} \\ \mathbf{0} & I \end{bmatrix}$$

and let be $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}$

Thus, $A\begin{pmatrix}\mathbf{y}\\ \mathbf{x}\end{pmatrix} \sim \mathcal{N}\left(A\mu,\ A\Sigma A^T\right)$. Now, let us compute this expressions

$$A\begin{pmatrix}\mathbf{y}\\ \mathbf{x}\end{pmatrix} = \begin{pmatrix}\mathbf{y} - \Sigma_{yx}\Sigma_{xx}^{-1}\mathbf{x}\\ \mathbf{x}\end{pmatrix} \equiv \begin{pmatrix}\mathbf{u}\\ \mathbf{x}\end{pmatrix}$$

$$A\mu = \begin{pmatrix}\mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}\mu_x\\ \mu_x\end{pmatrix}$$

$$\begin{aligned} A\Sigma A^T &= \begin{pmatrix}\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} & \Sigma_{yx} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xx}\\ \Sigma_{xy} & \Sigma_{xx}\end{pmatrix} A^T \\ &= \begin{pmatrix}\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} & \mathbf{0}\\ \Sigma_{xy} & \Sigma_{xx}\end{pmatrix}\begin{pmatrix}I & \mathbf{0}\\ -\Sigma_{xx}^{-1}\Sigma_{xy} & I\end{pmatrix} \\ &= \begin{pmatrix}\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} & \mathbf{0}\\ \mathbf{0} & \Sigma_{xx}\end{pmatrix} \end{aligned}$$

Because

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{x} \end{pmatrix}$$

is a (multivariate) normal variable and $\mathrm{Cov}(\mathbf{u}, \mathbf{x}) = 0$, it implies that $\mathbf{u}$ and $\mathbf{x}$ are independent, and hence $\mathbf{u}|\mathbf{x}$ has the same distribution than $\mathbf{u}$, that is

$$\mathbf{u}|\mathbf{x} \sim \mathcal{N}\left(\mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}\mu_x,\, \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\right)$$

That is,

$$\mathbf{y} - \Sigma_{yx}\Sigma_{xx}^{-1}\mathbf{x}|\mathbf{x} \sim \mathcal{N}\left(\mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}\mu_x,\, \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\right)$$

And from here, we conclude

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}\left(\mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \mu_x),\, \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\right)$$

$\square$

Define

$$\begin{aligned}\mu_{y|x} &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \mu_x) \\ &= \left[\mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}\mu_x\right] + \left[\Sigma_{yx}\Sigma_{xx}^{-1}\right]\mathbf{x} \\ &\equiv \beta_0 + \beta_1\mathbf{x}\end{aligned}$$

and

$$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$$

Remember that

$$\Sigma_{yy} = \mathbb{E}_{\mathbf{y}}\left[(\mathbf{y} - \mu_y)(\mathbf{y} - \mu_y)^T\right]^2 \quad \text{(resp. } \Sigma_{xx}\text{)}$$

and

$$\Sigma_{yx} = \mathbb{E}_{\mathbf{y},\mathbf{x}}\left[(\mathbf{y} - \mu_y)(\mathbf{x} - \mu_x)^T\right] \quad \text{(similarly } \Sigma_{xy} = \Sigma_{yx}^T\text{)}$$

Assume for now that $\mathbf{x} \equiv x \in \mathbb{R}$ and $\mathbf{y} \equiv y \in \mathbb{R}$, and that we count with a sample $\mathcal{D}_n = ((x_1, y_1), \ldots, (x_n, y_n))$ from independent and identically distributed random variables with the same distribution than the generic vector (x, y).

Thus (intuitively) good estimators of the previous quantities would be given by

- $\widehat{\Sigma}_{yy} \equiv S_{yy} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$ (resp. S_{xx})
- $\widehat{\Sigma}_{yx} \equiv S_{yx} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) = S_{xy}$
- $\hat{\mu}_{y|x} \equiv \hat{y}|x = \hat{\beta}_0 + \hat{\beta}_1 x$ where

$$\hat{\beta}_0 = \bar{y} - S_{yx} S_{xx}^{-1} \bar{x}$$

$$\hat{\beta}_1 = S_{yx} S_{xx}^{-1}$$

- $\widehat{\Sigma}_{y|x} \equiv \hat{\sigma}^2 = S_{yy} - S_{yx} S_{xx}^{-1} S_{xy}$

Let be $\hat{y}_i = \hat{y}_i|x_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $i = 1, \ldots, n$, we define the Sum of Squared Estimate Errors (SSE) (Suma de Cuadrados del Error (SCE), in Spanish) also known as Sum of Squared Residuals (SSR) or Residual Sum of Squares (RSS)

Definition 1 Sum of Squared Estimate Errors (SSE)

$$ESS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Remember that $\hat{y}_i = \bar{y} + S_{yx} S_{xx}^{-1} (x_i - \bar{x})$, thus applying simple algebra we get

$$
\begin{aligned}
SSE &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 + S_{yx}S_{xx}^{-1}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)S_{xx}^{-1}S_{xy} \\
&\quad - 2\sum_{i=1}^{n}(y_i - \bar{y})S_{yx}S_{xx}^{-1}(x_i - \bar{x}) \\
\frac{1}{n}SSE &= S_{yy} + S_{yx}\cancel{S_{xx}^{-1}}\cancel{S_{xx}}S_{xx}^{-1}S_{xy} - 2S_{yx}S_{xx}^{-1}S_{xy} \\
&= S_{yy} - S_{yx}S_{xx}^{-1}S_{xy}
\end{aligned}
$$