Rodrigo Floriano V.
111954

1) $X \mid \mu, \Sigma \sim N_p(X \mid \mu, \Sigma)$

Sea $X_1 \in \mathbb{R}^{p_1}$, $X_2 \in \mathbb{R}^{p_2}$ donde $p_1 + p_2 = p$

Sabemos que:

$$X_1 \mid X_2 \sim N_{p_1}(X_1 \mid \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Para $p = 2$:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad \Sigma_{11} = \sigma_1^2,\ \Sigma_{12} = \Sigma_{21} = \rho\sigma_1\sigma_2,\ \Sigma_{22}^{-1} = \frac{1}{\sigma_2^2}$$

Por lo que:

$$\mathbb{E}[X_1 \mid X_2] = \mu_1 + \rho\sigma_1\sigma_2 \frac{1}{\sigma_2^2}(X_2 - \mu_2) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2)$$

$\therefore\ \mathbb{E}[X_1 \mid X_2] = L(X_2) = \alpha_0 + \alpha_1 X_2$ donde $\alpha_0 = \mu_1 - \rho\frac{\sigma_1}{\sigma_2}\mu_2$ y $\alpha_1 = \rho\frac{\sigma_1}{\sigma_2}$

Observando el problema de regresión lineal:

Sabemos $\mathbb{E}[X_1 \mid X_2] = \arg\min_f \mathbb{E}[\|X_1 - f(X_2)\|_2^2]$, suponiendo $X_1 = f(X_2) + \varepsilon$, $\mathbb{E}[\varepsilon] = 0$

Por otro lado, sea el problema de mínimos cuadrados (suponiendo que f es lineal, $f(X_2) = \beta_0 + \beta_1 X_2$)

$\min_{\beta_0, \beta_1} \mathbb{E}[(X_1 - \beta_0 - \beta_1 X_2)^2]$. Como $X_1, X_2 \sim N(X \mid \mu, \Sigma)$, se tiene por convergencia (dominada) que: $\frac{\partial}{\partial x_i}\mathbb{E}[g(X)] = \mathbb{E}\left[\frac{\partial}{\partial x_i} g(X)\right]$ (Además, $(\beta_0, \beta_1)^T$ son parámetros)

CPO:

β_0: $-2\mathbb{E}[X_1 - \beta_0 - \beta_1 X_2] = 0$ (1)

β_1: $-2\mathbb{E}[(X_1 - \beta_0 - \beta_1 X_2)X_2] = 0$ (2)

De (1) se tiene:

$$\mathbb{E}[X_1 - \beta_0 - \beta_1 X_2] = \mathbb{E}[X_1] - \beta_0 - \beta_1\mathbb{E}[X_2] = \mu_1 - \beta_0 - \beta_1\mu_2 = 0 \iff \beta_0 = \mu_1 - \beta_1\mu_2$$

De (2) se tiene:

$$\mathbb{E}[X_1X_2] - \beta_0\mathbb{E}[X_2] - \beta_1\mathbb{E}[X_2^2] = 0$$

Sabemos: $\rho\sigma_1\sigma_2 = \mathrm{Cov}(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$

$\iff \mathbb{E}[X_1X_2] = \rho\sigma_1\sigma_2 + \mu_1\mu_2$

$\sigma_2^2 = \mathrm{Var}(X_2) = \mathbb{E}[X_2^2] - (\mathbb{E}[X_2])^2$

$\iff \mathbb{E}[X_2] = \sigma_2^2 + \mu_2^2$

Por lo que $\rho\sigma_1\sigma_2 + \mu_1\mu_2 - \beta_0\mu_2 - \beta_1(\sigma_2^2 + \mu_2^2) = 0$

$\iff \rho\sigma_1\sigma_2 + \mu_1\mu_2 - (\mu_1 - \beta_1\mu_2)\mu_2 - \beta_1\sigma_2^2 - \beta\mu_2^2 = 0$

$\iff \rho\sigma_1\sigma_2 - \beta_1\sigma_2^2 = 0 \iff \beta_1 = \rho\frac{\sigma_1}{\sigma_2}$ $\therefore$ $\beta_0 = \mu_1 - \rho\frac{\sigma_1}{\sigma_2}\mu_2$

Esto implica que $\alpha_0 = \beta_0$ y $\alpha_1 = \beta_1$, por lo que son equivalentes.

2) Generalizar para $p_1+p_2>2$

Recordemos que $\mathbb{E}[X_1|X_2]=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(X_2-\mu_2)=\mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2+\Sigma_{12}\Sigma_{22}^{-1}X_2$

Sean $\alpha_0=\mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2$ y $\alpha=\Sigma_{12}\Sigma_{22}^{-1}$

Sabemos que un subconjunto $S=\{p_{s_1},\dots,p_{s_k}\}\subseteq\{1,\dots,p\}$, $\#S=k$ tiene una distribución normal k-variada

Sin pérdida de generalidad, asumimos que $P_1=\{1,\dots,p_1\}$ y $P_2=\{p_1+1,\dots,p\}$

$\#P_1=p_1$ y $\#P_2=p_2=p-p_1$

Sabemos que $S=\{i,p_1+1,\dots,p\}$ tiene una distribución p_2+1-variada

Retomando el problema de regresión lineal, suponemos $X_i=f_i(X_2)+\varepsilon_i$, $\mathbb{E}[\varepsilon]=0$ (recordemos $\mathbb{E}[X_i|X_2]=\operatorname{argmin}_f\mathbb{E}[\|X_i-f(X_2)\|_2^2]$) donde f_i es lineal $f_i(X_2)=\beta_{i0}+\sum_{j=p_1+1}^{p}\beta_{ij}X_j$ donde $i\in P_1$. Por ello se resuelve el siguiente problema:

$$\min_{\beta_{i0},\beta_i}\mathbb{E}[(X_i-\beta_{i0}-\beta_iX_2)^2]\quad,\ \beta_i=(\beta_{ip_1+1},\dots,\beta_{ip})$$

Recordando convergencia dominante y que β_{i0},β_i son parámetros:

CPO:

β_{i0}: $\mathbb{E}[X_i-\beta_{i0}-\beta_iX_2]=0$ $\quad\forall i\in P_1$

β_i: $\mathbb{E}[(X_i-\beta_{i0}-\beta_iX_2)X_2^T]=0$

Definiendo $\beta_0=(\beta_{10},\dots,\beta_{p_10})^T$ y $\beta=\begin{pmatrix}\beta_{1p_1+1}&\cdots&\beta_{1p}\\ \vdots&\ddots&\vdots\\ \beta_{p_1p_1+1}&\cdots&\beta_{p_1p}\end{pmatrix}=\begin{pmatrix}\beta_1\\ \vdots\\ \beta_{p_1}\end{pmatrix}\in\mathbb{R}^{p_1\times p_2}$

$\beta_0\in\mathbb{R}^{p_1\times 1}$

Se tiene que resolver todas las ecuaciones normales de P_1 es equivalente a resolver:

$$\mathbb{E}[X_1-\beta_0-\beta X_2]=0\iff\beta_0=\mu_1-\beta\mu_2$$
$$\mathbb{E}[(X_1-\beta_0-\beta X_2)X_2^T]=0\iff\mathbb{E}[X_1X_2^T]-\beta_0\mu_2^T-\beta\mathbb{E}[X_2X_2^T]=0$$

Sabemos $\Sigma_{12}=\operatorname{Cov}(X_1,X_2)=\mathbb{E}[X_1X_2^T]-\mu_1\mu_2^T\iff\mathbb{E}[X_1X_2^T]=\Sigma_{12}+\mu_1\mu_2^T$

$\Sigma_{22}=\operatorname{Var}(X_2)=\mathbb{E}[X_2X_2^T]-\mu_2\mu_2^T\iff\mathbb{E}[X_2X_2^T]=\Sigma_{22}+\mu_2\mu_2^T$

$$\therefore\ \mathbb{E}[X_1X_2^T]-(\mu_1-\beta\mu_2)\mu_2^T-\beta\mathbb{E}[X_2X_2^T]=0$$
$$\iff\Sigma_{12}+\mu_1\mu_2^T-\mu_1\mu_2^T+\beta\mu_2\mu_2^T-\beta\Sigma_{22}+\beta\mu_2\mu_2^T=0$$
$$\iff\Sigma_{12}=\beta\Sigma_{22}\iff\beta=\Sigma_{12}\Sigma_{22}^{-1}\quad\therefore\ \beta_0=\mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2$$

$\therefore\ \beta_0=\alpha_0$ y $\beta=\alpha$ por lo que son equivalentes.

Probar: $X_1 | X_2 \sim N_{p_2}(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2-\mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ si

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_p\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right), \quad X_1 \in \mathbb{R}^{p_1 \times 1};\ X_2 \in \mathbb{R}^{p_2 \times 1},\ p_1 + p_2 = p$$

Buscamos $Z = Z(X_1, X_2) \in \mathbb{R}^{p_1 \times 1}$ tal que Z y X_2 son independientes

$Z = X_1 + AX_2$; $A \in \mathbb{R}^{p_1 \times p_2}$. Se debe cumplir $Cov(Z, X_2) = 0$

$$Cov(Z, X_2) = Cov(X_1 + AX_2, X_2) = Cov(X_1, X_2) + A\,Var(X_2)$$
$$= \Sigma_{12} + A\Sigma_{22} = 0$$

$\therefore A = -\Sigma_{12}\Sigma_{22}^{-1}$

Sabemos que $\mathbb{E}[Z|X_2] = \mathbb{E}[Z]$ y $Var(Z|X_2) = Var(Z)$

Probando que Z es normal:

$$M_Z(t) = \mathbb{E}\left[e^{t^T Z}\right] = \mathbb{E}\left[e^{t^T(X_1 + AX_2)}\right] = \mathbb{E}\left[e^{t^T X + t^T A X_2}\right] = M_X\left(\begin{bmatrix} t \\ A^T t \end{bmatrix}\right)$$

Sabemos $M_X(t) = e^{\mu^T t + \frac{1}{2} t^T \Sigma t}$

Por lo que:

$$M_Z(t) = e^{\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}^T \begin{bmatrix} t \\ A^T t \end{bmatrix} + \frac{1}{2}\begin{bmatrix} t \\ A^T t \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\begin{bmatrix} t \\ A^T t \end{bmatrix}}$$

$$= e^{t^T\mu_1 + t^T A\mu_2 + \frac{1}{2}\left[t^T\Sigma_{11} + t^T A\Sigma_{21},\ t^T\Sigma_{12} + t^T A\Sigma_{22}\right]\begin{bmatrix} t \\ A^T t \end{bmatrix}}$$

$$= e^{t^T\mu_1 \underbrace{- t^T A\mu_2}_{-t^T\Sigma_{12}\Sigma_{22}^{-1}\mu_2} + \frac{1}{2}\left[t^T\Sigma_{11}t + \underbrace{t^T A\Sigma_{21}t}_{-t^T\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}t} + t^T\Sigma_{12}A^T t + \underbrace{t^T A\Sigma_{22}A^T t}_{\substack{t^T\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21}t \\ t^T\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}t}}\right]}$$

$$= e^{t^T\mu_1 + t^T A\mu_2 + \frac{1}{2}\left[t^T\Sigma_{11}t - t^T\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}t\right]}$$

$$= e^{t^T(\mu_1 - \Sigma_{12}\Sigma_{22}^{-}\mu_2) + \frac{1}{2}t^T(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})t}$$

De lo anterior se observa: $\mathbb{E}[Z] = \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2$

$Var(Z) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

Sabemos que $X_1 | X_2$ es una normal al ser comb. lineal de Z, X_2:

$$\mathbb{E}[X_1 | X_2] = \mathbb{E}[Z - AX_2 | X_2] = \mathbb{E}[Z|X_2] - \mathbb{E}[AX_2|X_2] = \mathbb{E}[Z] - AX_2$$
$$= \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 + \Sigma_{12}\Sigma_{22}^{-1}X_2$$
$$= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$$

$$Var(X_1|X_2) = Var(Z - AX_2|X_2) \underset{\text{independ.}}{=} Var(Z|X_2) + Var(AX_2|X_2)$$
$$= Var(Z|X_2) + 0 = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \quad \blacksquare$$