

Durante el trabajo se menciona que la distribución posterior de  $\beta$  es conjugada, sin embargo, no sé da una definición formal aún pues se busca profundizar en los detalles y las implicaciones de la definición. Asimismo, este apéndice se concentra en demostrar formalmente el proceso de actualización de los hiperparámetros y su interpretación bayesiana.

**Definición 0.1.** Se dice que una distribución previa  $\pi(\beta) \in \mathcal{F}$  es *conjugada* bajo el muestreo  $\pi(\mathbf{X}|\beta)$ , si la distribución posterior  $\pi(\beta|\mathbf{X})$  pertenece a la misma familia de distribuciones  $\mathcal{F}$ . Es decir:

$$\pi(\beta) \in \mathcal{F} \Rightarrow \pi(\beta|\mathbf{X}) \in \mathcal{F},$$

con  $\mathbf{X}$  cualquier muestra aleatoria, **mendoza2011estadistica** y **barber2010bayesian**. Si se cumple esta definición, es usual referirse al modelo como conjugado o hablar de distribuciones conjugadas.

Como ejemplos clásicos de distribuciones conjugadas, se tiene el modelo *normal-normal* como el utilizado en este trabajo o el modelo *beta-bernoulli*. Para el segundo, la muestra consiste en observaciones binarias con parámetro  $\theta$ , el cual a su vez, se modela de forma previa con una distribución beta. Una vez llevado a cabo el procedimiento de actualización, la distribución posterior de  $\theta$  bajo el muestreo correspondiente vuelve a pertenecer a la familia de distribuciones beta. Asimismo, las distribuciones conjugadas tienen la ventaja que simplifican relativamente los cálculos permitiendo derivar distribuciones cerradas.

Esta definición es interesante pues retoma la idea central de la estadística bayesiana de actualización de la medida de incertidumbre ante la presencia de nueva información. Esto pues, bajo el modelado propuesto y dado que ya se piensa que  $\beta \sim \mathcal{F}$ , el introducir la nueva información bajo cierto muestreo, únicamente refina el entendimiento de  $\beta$  pues su forma funcional no cambia al seguir perteneciendo a la misma familia de distribuciones. Sin embargo, al modificar los hiperparámetros, si se está actualizando la asignación de incertidumbre asociada a  $\beta$ . Ante la eventualidad que se obtuviera más información (bajo el mismo muestreo), el proceso se podría repetir, iterando en la actualización para continuar mejorando el entendimiento de  $\beta$ . Para ejemplificar de forma más rigurosa, se completa la derivación de la distribución posterior usada para este trabajo.

**Teorema 0.2.** *Dado el modelo bpwpm (definición ??, página ??) donde se tienen las siguientes distribuciones:*

$$z_i \mid \mathbf{x}_i \sim \mathcal{N}(z_i \mid \eta(\mathbf{x}_i), 1) \quad (??)$$

$$\beta \sim \mathcal{N}_\lambda(\beta \mid \mu_\beta, \Sigma_\beta), \quad (??)$$

*la distribución posterior de  $\beta$  dado  $\mathbf{y}, \mathbf{z}$  y  $\mathbf{X}$  es conjugada. Es decir, esta también pertenece a la familia normal multivariada,*

$$\beta \mid \mathbf{y}, \mathbf{z}, \mathbf{X} \sim \mathcal{N}_\lambda(\beta \mid \mu_\beta^*, \Sigma_\beta^*), \quad (??)$$

con los hiper-parámetros actualizados:

$$\begin{aligned}\mu_{\boldsymbol{\beta}}^* &= \Sigma_{\boldsymbol{\beta}}^* \times (\Sigma_{\boldsymbol{\beta}}^{-1} \mu_{\boldsymbol{\beta}} + \tilde{\Psi}(\mathbf{X})^t \mathbf{z}) \\ \Sigma_{\boldsymbol{\beta}}^* &= \left[ \Sigma_{\boldsymbol{\beta}}^{-1} + \tilde{\Psi}(\mathbf{X})^t \tilde{\Psi}(\mathbf{X}) \right]^{-1}.\end{aligned}$$

*Demostración.* Se retoma la derivación comenzada en la página ??.

$$\begin{aligned}\pi(\boldsymbol{\beta} | \mathbf{z}, \mathbf{y}, \mathbf{X}) &= \frac{\pi(\mathbf{z}, \boldsymbol{\beta} | \mathbf{y}, \mathbf{X})}{\pi(\mathbf{z})} \\ &= C \pi(\boldsymbol{\beta}) \prod_{i=1}^n \phi(z_i | \eta(\mathbf{x}_i), 1),\end{aligned}$$

utilizando un argumento de proporcionalidad sobre  $\boldsymbol{\beta}$  y expandiendo la función de densidad de una variable aleatoria normal  $\phi$  se tiene:

$$\propto \pi(\boldsymbol{\beta}) \times \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (z_i - \eta(\mathbf{x}_i))^2 \right\}.$$

Ahora, se sustituye la identidad (??),  $\eta(\mathbf{x}_i) = \boldsymbol{\beta}^t \tilde{\boldsymbol{\psi}}_i(\mathbf{x}_i)$ , y por las propiedades de la función exponencial, el producto se convierte en suma:

$$\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (z_i - \boldsymbol{\beta}^t \tilde{\boldsymbol{\psi}}_i(\mathbf{x}_i))^2 \right\} \times \pi(\boldsymbol{\beta}).$$

Posteriormente, se re-expresa la identidad en su forma matricial usando la siguiente propiedad

$$\|\mathbf{u}\|^2 = \mathbf{u}^t \mathbf{u} = \sum_{i=1}^n u_i^2$$

donde  $\mathbf{u}$  representa un vector de tamaño  $n$  arbitrario. En este caso  $u_i = (z_i - \beta^t \tilde{\psi}_i(\mathbf{x}_i))$ :

$$\propto \exp \left\{ -\frac{1}{2}(\mathbf{z} - \tilde{\Psi}(\mathbf{X})\beta)^t(\mathbf{z} - \tilde{\Psi}(\mathbf{X})\beta) \right\} \times \pi(\beta).$$

Se hace notar que se juntan todas las funciones  $\tilde{\psi}_i(\mathbf{x}_i)$  en su correspondiente matriz  $\tilde{\Psi}(\mathbf{X})$ . Lo anterior es análogo a que  $\mathbf{z}$  sea definida como una distribución normal multivariada de tamaño  $n$  con vector de medias  $\tilde{\Psi}(\mathbf{X})\beta$  y a la matriz identidad como matriz de varianza covarianza (preservando la independencia). Ahora, se desglosa  $\pi(\beta)$  en su correspondiente densidad normal multivariada:

$$\begin{aligned} &\propto \exp \left\{ -\frac{1}{2}(\mathbf{z} - \tilde{\Psi}(\mathbf{X})\beta)^t(\mathbf{z} - \tilde{\Psi}(\mathbf{X})\beta) \right\} \times \exp \left\{ -\frac{1}{2}(\beta - \mu_\beta)^t \Sigma_\beta^{-1}(\beta - \mu_\beta) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ (\mathbf{z} - \tilde{\Psi}(\mathbf{X})\beta)^t(\mathbf{z} - \tilde{\Psi}(\mathbf{X})\beta) + (\beta - \mu_\beta)^t \Sigma_\beta^{-1}(\beta - \mu_\beta) \right] \right\}. \end{aligned} \quad (1)$$

De lo anterior, basta concentrarse únicamente en la expresión dentro de los corchetes en (1) la cual se desarrolla usando propiedades estándar de álgebra lineal,

$$\begin{aligned} &(\mathbf{z} - \tilde{\Psi}(\mathbf{X})\beta)^t(\mathbf{z} - \tilde{\Psi}(\mathbf{X})\beta) + (\beta - \mu_\beta)^t \Sigma_\beta^{-1}(\beta - \mu_\beta) \\ &= \mathbf{z}^t \mathbf{z} - 2\beta^t \tilde{\Psi}(\mathbf{X})^t \mathbf{z} + \beta^t \tilde{\Psi}(\mathbf{X})^t \tilde{\Psi}(\mathbf{X})\beta + \beta^t \Sigma_\beta^{-1} \beta - 2\beta \Sigma_\beta^{-1} \mu_\beta + \mu_\beta^t \Sigma_\beta^{-1} \mu_\beta \\ &= \underbrace{\beta^t (\Sigma_\beta^{-1} + \tilde{\Psi}(\mathbf{X})^t \tilde{\Psi}(\mathbf{X})) \beta}_{\text{término cuadrado}} - \underbrace{2\beta (\Sigma_\beta^{-1} \mu_\beta + \tilde{\Psi}(\mathbf{X})^t \mathbf{z})}_{\text{término lineal}} + \underbrace{(\mathbf{z}^t \mathbf{z} + \mu_\beta^t \Sigma_\beta^{-1} \mu_\beta)}_{\text{término independiente}}. \end{aligned} \quad (2)$$

Estudiando a detalle la línea (2), se nota que está compuesta por un término cuadrático, un término lineal y un término independiente en su forma multi-dimensional.

De resultados estándar de álgebra básica se piensa que si se logra completar el trino-

mio cuadrado perfecto se puede simplificar mucho la expresión. En el caso multivariado, completar el un trinomio cuadrado perfecto es análogo al utiliza la propiedad de *rectificación elipsoidal*:

$$\mathbf{u}^t A \mathbf{u} - 2 \mathbf{u}^t \boldsymbol{\alpha} = (\mathbf{u} - A^{-1} \boldsymbol{\alpha})^t A (\mathbf{u} - A^{-1} \boldsymbol{\alpha}) - \boldsymbol{\alpha}^t A^{-1} \boldsymbol{\alpha}$$

donde  $\mathbf{u}$  y  $\boldsymbol{\alpha}$  son vectores de igual tamaño y  $A$  una matriz invertible. Por lo tanto, basta con reemplazar las variables  $\mathbf{u} = \boldsymbol{\beta}$ ,  $A = \Sigma_{\boldsymbol{\beta}}^{-1} + \tilde{\Psi}(\mathbf{X})^t \tilde{\Psi}(\mathbf{X})$  y  $\boldsymbol{\alpha} = \Sigma_{\boldsymbol{\beta}}^{-1} \mu_{\boldsymbol{\beta}} + \tilde{\Psi}(\mathbf{X})^t \mathbf{z}$ , para derivar que (2) es igual a:

$$= (\boldsymbol{\beta} - \mu_{\boldsymbol{\beta}}^*)^t \Sigma_{\boldsymbol{\beta}}^* (\boldsymbol{\beta} - \mu_{\boldsymbol{\beta}}^*) + \gamma \quad (3)$$

con:

$$\begin{aligned} \mu_{\boldsymbol{\beta}}^* &= \Sigma_{\boldsymbol{\beta}}^* \times (\Sigma_{\boldsymbol{\beta}}^{-1} \mu_{\boldsymbol{\beta}} + \tilde{\Psi}(\mathbf{X})^t \mathbf{z}) \\ \Sigma_{\boldsymbol{\beta}}^* &= \left[ \Sigma_{\boldsymbol{\beta}}^{-1} + \tilde{\Psi}(\mathbf{X})^t \tilde{\Psi}(\mathbf{X}) \right]^{-1} \\ \gamma &= \mathbf{z}^t \mathbf{z} + \mu_{\boldsymbol{\beta}}^t \Sigma_{\boldsymbol{\beta}}^{-1} \mu_{\boldsymbol{\beta}} + (\Sigma_{\boldsymbol{\beta}}^{-1} \mu_{\boldsymbol{\beta}} + \tilde{\Psi}(\mathbf{X})^t \mathbf{z})^t \mu_{\boldsymbol{\beta}}^*. \end{aligned}$$

donde  $\gamma$  es un término independiente que no depende de  $\boldsymbol{\beta}$ . Retomando la notación de (1) y sustituyendo (3), se tiene:

$$\begin{aligned} &\propto \exp \left\{ -\frac{1}{2} \left[ (\boldsymbol{\beta} - \mu_{\boldsymbol{\beta}}^*)^t \Sigma_{\boldsymbol{\beta}}^{*-1} (\boldsymbol{\beta} - \mu_{\boldsymbol{\beta}}^*) + \gamma \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mu_{\boldsymbol{\beta}}^*)^t \Sigma_{\boldsymbol{\beta}}^{*-1} (\boldsymbol{\beta} - \mu_{\boldsymbol{\beta}}^*) \right\}, \end{aligned}$$

1. La matriz  $A$  es de rango completo y por ende invertible pues tanto  $\Sigma_{\boldsymbol{\beta}}^{-1}$  como  $\tilde{\Psi}(\mathbf{X})^t \tilde{\Psi}(\mathbf{X})$  son simétricas positivas definidas, por lo tanto, su suma también lo es.

lo cual es el kernel de una distribución normal multivariada con vector de medias  $\mu_{\beta}^*$  y matriz de varianza y covarianza  $\Sigma_{\beta}^*$ . Dado que la distribución previa y posterior pertenecen a la misma familia de distribuciones, estas son conjugadas y se termina así la prueba. Q.E.D.