Seminario de Estadística 1
Tarea 10
REGRESION MULTPLE

Soriano Flores Antonio

Octubre 2019

Definición 0.1 (Distribuciones Elipticas) *Decimos que X es un vector aleatorio con distribución eliptica centrada en μ y con matriz de precisión $\boldsymbol{P}$ si su densidad está dada por:*

$$p\left(\boldsymbol{x}|\mu,\boldsymbol{P}\right) \propto |\boldsymbol{P}|^{\frac{1}{2}} g\left(\left(\boldsymbol{x}-\mu\right)^T \boldsymbol{P}\left(\boldsymbol{x}-\mu\right)\right)$$

Donde g es una función real positiva tal que:

$$\int_{\mathbb{R}^p} |\boldsymbol{P}|^{\frac{1}{2}} g\left(\left(\boldsymbol{x}-\mu\right)^T \boldsymbol{P}\left(\boldsymbol{x}-\mu\right)\right) < \infty$$

En cuyo caso usamos la siguiente notación:

$$X \sim EC\left(x|\mu,\boldsymbol{P},g\right)$$

Los vectores aleatorios que tienen distribuciones elípticas tienen el siguiente resultado:

Teorema 0.1 (Distribución Radial de una distribución eliptica) *Sea $X \sim EC\left(\mu,\boldsymbol{P},g\right)$ entonces:*

- $X \overset{d}{=} \mu + \boldsymbol{A}^T R U^{(p)}$*; donde* $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{P}^{-1}$*;* $U^{(p)}$ *es un vector aleatorio en* $\mathbb{R}^P$ *sobre la esfera unitaria y* $R^2 = \left(\boldsymbol{X}-\mu\right)^T \boldsymbol{P}\left(\boldsymbol{X}-\mu\right)$ *se le conoce como la distribución radial de* X
- R^2 *tiene por densidad:*
$$h_{R^2}\left(r\right) \propto r^{\frac{p}{2}-1} g\left(r\right)$$

Definición 0.2 (Normal Multivariada) *Decimos que $\boldsymbol{X}$ es un vector aleatorio con distribución Normal Multivariado con vector de medias μ y matriz de varianzas y covarianzas Σ si su densidad está dada por:*

$$p\left(\boldsymbol{x}|\mu,\Sigma\right) \propto |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}\left(\boldsymbol{x}-\mu\right)^T\Sigma^{-1}\left(\boldsymbol{x}-\mu\right)}; \quad \boldsymbol{x},\mu \in \mathbb{R}^p; \Sigma \in \mathbb{R}^{p\times p}$$

En cuyo caso usamos la siguiente notación:

$$X \sim N\left(x|\mu,\Sigma\right)$$

Como es costumbre en Bayesiana se trabaja con la matriz de precisión $\mathbf{P} = \Sigma^{-1}$ quedando entonces el modelo parametrizado como:

$$p\left(\mathbf{x}|\mu,\mathbf{P}\right) \propto |\mathbf{P}|^{\frac{1}{2}} e^{-\frac{1}{2}\left(\mathbf{x}-\mu\right)^T\mathbf{P}\left(\mathbf{x}-\mu\right)}; \quad \mathbf{x},\mu \in \mathbb{R}^p; \mathbf{P} \in \mathbb{R}^{p\times p}$$

Se puede probar en este caso que:

$$\mathbb{E}(X) = \mu \qquad \mathrm{Var}(X) = \mathbf{P}^{-1}$$

En este caso la notación para el enfoque bayesiano es:

$$X \sim N\left(x|\mu, \mathbf{P}\right)$$

Definición 0.3 (T - Multivariada) *Decimos que $\boldsymbol{X}$ es un vector aleatorio con distribución T Multivariada con parámtro de localización μ, matriz de escala $\boldsymbol{P}$ y grados de libertad d si su densidad está dada por:*

$$p\left(\boldsymbol{x}|d,\mu,\boldsymbol{P}\right) \propto |\boldsymbol{P}|^{\frac{1}{2}}\left(1+\frac{\left(\boldsymbol{x}-\mu\right)^T\boldsymbol{P}\left(\boldsymbol{x}-\mu\right)}{d}\right)^{-\left(\frac{d+p}{2}\right)} \quad ; \boldsymbol{x},\mu \in \mathbb{R}^p; \boldsymbol{P} \in \mathbb{R}^{p\times p}; d \in \mathbb{R}$$

En cuyo caso usamos la siguiente notación:

$$X \sim T\left(x|d,\mu,\Sigma\right)$$

Se puede probar en este caso que:

$$\mathbb{E}(x) = \mu; \qquad Var(x) = \frac{d}{d-2}\boldsymbol{P}^{-1} \quad d > 2$$

Definición 0.4 (Normal Multivariado - Gamma) *Diremos que el vector aleatorio $(\boldsymbol{x}, y)$ con $\boldsymbol{x} \in \mathbb{R}^p$ y $y \in \mathbb{R}$ sigue una distribución Normal Multivariada-Gamma de parametros $(\mu, \boldsymbol{P}, \alpha, \beta)$ si su densidad está dada por:*

$$\begin{aligned} p\left(\boldsymbol{x}, y|\mu, \boldsymbol{P}, \alpha, \beta\right) &\propto N\left(\boldsymbol{x}|\mu, y\boldsymbol{P}\right) Ga\left(y|\alpha,\beta\right) \\ &\propto |y\boldsymbol{P}|^{\frac{1}{2}} e^{-\frac{y}{2}\left(\boldsymbol{x}-\mu\right)^T\boldsymbol{P}\left(\boldsymbol{x}-\mu\right)} y^{\alpha-1}e^{-y\beta}\boldsymbol{1}_{(0,\infty)}(y) \\ &\propto y^{\frac{p}{2}+\alpha-1}e^{-\frac{y}{2}\left(\left(\boldsymbol{x}-\mu\right)^T\boldsymbol{P}\left(\boldsymbol{x}-\mu\right)+2\beta\right)}\boldsymbol{1}_{(0,\infty)}(y) \end{aligned}$$

Dadas las definiciones y teoremas anteriores, responda lo que se le pregunta:

1. Sea $X \in \mathbb{R}^p$ vector aleatorio tal que $X \sim N\left(x|\mu, \mathbf{P}\right)$. Demuestre que X pertenece a la familia de distribuciones elipticas y verifique que :

$$R^2 = \left(\mathbf{X}-\mu\right)^T\mathbf{P}\left(\mathbf{X}-\mu\right) \sim \chi^2_p$$

2. Sea $X \in \mathbb{R}^p$ vector aleatorio tal que $X \sim T\left(x|d,\mu,\Sigma\right)$. Demuestre que X pertenece a la familia de distribuciones elipticas y verifique que :

$$p^{-1}R^2 = p^{-1}\left(\mathbf{X}-\mu\right)^T\mathbf{P}\left(\mathbf{X}-\mu\right) \sim F_{(p,d)}$$

3. Sea (X, Y) con $X \in \mathbb{R}^p$ y $Y \in \mathbb{R}$ un vector aleatorio con distribución Normal Multivariada-Gamma de parametros $(\mu, \mathbf{P}, \alpha, \beta)$ demuestre explicando muy bien cada paso que:

   a) $X \sim T_p\left(\mathbf{x}\middle|2\alpha, \mu, \frac{\alpha}{\beta}\mathbf{P}\right)$

   b) $Y \sim Ga\left(\alpha, \beta\right)$

4. La tabla FootballLeague.csv contiene los datos sobre el desempeño de los equipos de la liga nacional de fútbol de E.U.A. durante 1976.

   Ajuste un modelo lineal bajo el enfoque bayesiano que relaciona el número de juegos ganados (y) con

   - Yardas por aire del equipo (x_2)

- El porcentaje de Yardas por Tierra (x_7)
- Las Yardas por tierra del contrario (x_8)

$$y = \beta_0 + \beta_1 x_2 + \beta_2 x_7 + \beta_3 x_8 + \varepsilon \quad \varepsilon \sim N(0, \tau)$$

Utilice como distribución inicial de los parámetros $(\beta_0, \beta_1, \beta_2, \beta_3, \tau)$ la Normal Multiviarada Gamma de parámetros:

- $\mu_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$
- $\mathbf{P}_0 = \begin{pmatrix} 0.001 & 0 & 0 & 0 \\ 0 & 0.001 & 0 & 0 \\ 0 & 0 & 0.001 & 0 \\ 0 & 0 & 0 & 0.001 \end{pmatrix}$
- $\alpha_0 = 0.001$
- $\beta_0 = 0.001$

  Con la distribución final encuentre los intervalos al 95 % de probabilidad para cada uno de los parámetros del modelo $(\beta_0, \beta_1, \beta_2, \beta_3,)$ y comparelos con los intervalos de confianza obtenidos por la estadística clásica.