Regresión Lineal: Enfoque Bayesiano

Departamento de Probabilidad y Estadística

Antonio Soriano Flores

[1] UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

asoriano@sigma.iimas.unam.mx

ÍNDICE

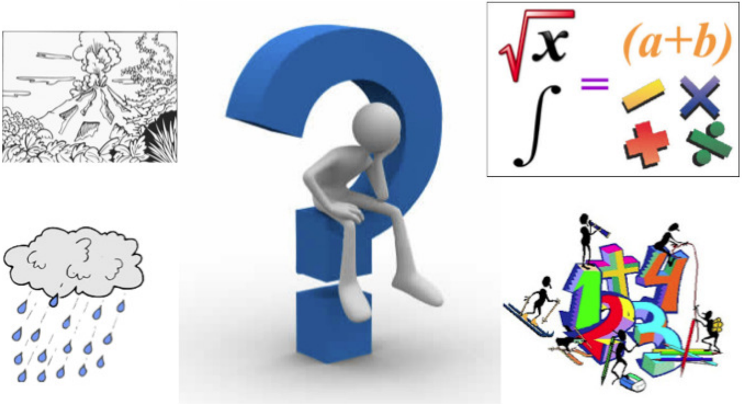


Figura : Modelación Matemática

En general un modelo busca encontrar una relación funcional entre ciertas variables independientes $\underline{z} = (z_1, z_2, ..., z_k)$ y una variable respuesta Y la cual queremos explicar o modelar.

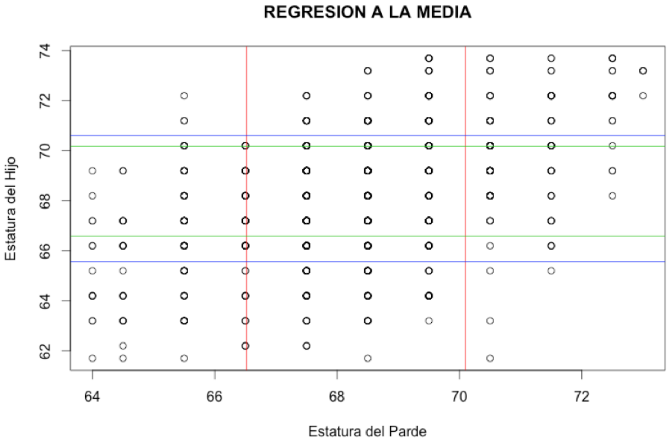
$$Y = f\left(z_1, z_2, ..., z_k\right) \quad (Determinista)$$

$$Y = f\left(z_1, z_2, ..., z_k\right) + \varepsilon \quad (Probabilistico)$$

- ¿Qué función f tomar?
- ¿El modelo es adecuado? ¿Existe el mejor modelo?
- ¿Qué supuestos debemos tomar en cuenta?
- ¿Las variables z_i si explican a Y? ¿Está respaldado con buenos datos?

Un poco de Historia....

El término **regresión** fue acuñando por Francis Galton en el siglo XIX en su artículo **Regression towards mediocrity in hereditary stature**, en donde observó que las alturas de los descendientes de ancestros altos tienden a regresar hacia abajo, hacia un promedio normal (un fenómeno conocido como regresión hacia la media).



Dada una variable repuesta Y y un conjunto de covariables $\underline{z}$, surge de manera natural preguntarnos cuál deberá ser la relación funcional para modelar dicha relación.
Una forma de modelar podría ser:

$$\mathbb{E}(Y \mid \underline{z}) = \mu(\underline{z})$$

donde, en general, $\mu(.)$ es una función desconocida. En la práctica es común aproximar a $\mu(.)$ a traves de una función más simple:

$$\mu(\underline{z}) = \psi(\underline{z}; \underline{\beta})$$

donde $\underline{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^T$ denota a un vector de parámetros desconocidos.

La forma mas simple para modelar la relación es suponer una función lineal de $\underline{\beta}$ es decir:

$$\psi\left(\underline{z};\underline{\beta}\right)=\beta_0+\beta_1 s_1\left(\underline{z}\right)+\ldots+\beta_k s_k\left(\underline{z}\right)$$

Donde s_i son funciones conocidas.
Finalmente esta función $\psi\left(\underline{z};\underline{\beta}\right)$ es tratada como si fuera la verdadera función de regresión $\mu\left(.\right)$, por lo que el problema se reduce a hacer inferencias sobre el valor del vector de parámetros $\underline{\beta}$.

Adicional al supuesto de la relación funcional entre Y y $\underline{z}$, el modelo lineal general supone normalidad en la variable respuesta Y lo que se traduce en:

$$Y|\underline{z} \sim N\left(\mu\left(\underline{z}\right), \sigma^2\right)\left(\sigma^2 > 0, desconocida\right)$$

Otro supuesto importante es suponer ademas independencia entre cada una de las observaciones, es decir, $Y_i \perp Y_j$. Finalmente imaginando que recibiremos n observaciones de este modelo para distinto nivel de las covariables $\underline{z}_i$ podemos escribir:

$$Y_i = \beta_0 + \beta_1 s_1\left(\underline{z}_i\right) + \ldots + \beta_k s_k\left(\underline{z}_i\right) + \varepsilon_i \quad \varepsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right) \tag{1}$$

Definamos:

$$x_{ij} = s_j\left(\underline{z}_i\right) \quad i \in \{1,\ldots,n\}; \quad j \in \{1,\ldots,k\}$$

Entonces el modelo (1) lo podemos escribir como:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right) \tag{2}$$

Ejemplo:
Suponiendo que sólo tenemos una covariable $z \in \mathbb{R}$, y haciendo $s_j(z) = z^j$ entonces (2) toma la forma:

$$Y_i = \beta_0 + \beta_1 z + \ldots + \beta_k z^k + \varepsilon_i \quad \varepsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right) \tag{3}$$

El modelo anterior pretende modelar el valor de Y a través de una función polinomial de la covariable z.

Dada la relación lineal que hemos impuesto resulta conveniente utilizar una notación matricial y escribir (2) como: (haciendo $p = k + 1$)

$$\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}; \qquad \underline{\varepsilon} \sim N_n\left(\underline{0}, \sigma^2\mathbf{I}_n\right) \Rightarrow \underline{Y} \sim N_n\left(\mathbf{X}\underline{\beta}, \sigma^2\mathbf{I}_n\right)$$

Donde:

$$\underline{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}_{n\times p} \qquad \underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n\times 1}$$

El vector de parámetros:

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}_{p\times 1}$$

Solución clásica del problema de estimación:
Dado que ahora conocemos la forma de distribución de $\underline{Y}$, podemos encontrar la función de verosimilitud

$$f\left(\underline{Y};\underline{\beta},\sigma^2\right)=\frac{1}{(2\pi)^{\frac{n}{2}}\left|\sigma^2\mathbf{I}_n\right|^{\frac{1}{2}}}e^{-\frac{1}{2}\left(\underline{Y}-\mathbf{X}\underline{\beta}\right)^t\left(\sigma^2\mathbf{I}_n\right)^{-1}\left(\underline{Y}-\mathbf{X}\underline{\beta}\right)}$$

Como $\left|\sigma^2\mathbf{I}_n\right|=\sigma^{2n}$ y $\left(\sigma^2\mathbf{I}_n\right)^{-1}=\frac{1}{\sigma^2}\mathbf{I}_n$ entonces la verosimilitud es:

$$\mathscr{L}\left(\underline{\beta},\sigma^2;\underline{Y}\right)=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\left(\underline{Y}-\mathbf{X}\underline{\beta}\right)^t\left(\underline{Y}-\mathbf{X}\underline{\beta}\right)}$$

Sacamos logaritmo de la verosimilitud:

$$\log\mathscr{L}\left(\underline{\beta},\sigma^2;\underline{Y}\right)=-\frac{n}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2\sigma^2}\left(\underline{Y}-\mathbf{X}\underline{\beta}\right)^t\left(\underline{Y}-\mathbf{X}\underline{\beta}\right)$$

Maximizamos con respecto a $\left(\underline{\beta},\sigma^2\right)$ para ello derivamos e igualamos a cero

$$\frac{\partial}{\partial\underline{\beta}}\log\mathscr{L}\left(\underline{\beta},\sigma^2;\underline{Y}\right)=-\frac{1}{2\sigma^2}\left(-2\mathbf{X}^t\underline{Y}+2\mathbf{X}^t\mathbf{X}\underline{\beta}\right)\tag{4}$$

$$\frac{\partial}{\partial\sigma^2}\log\mathscr{L}\left(\underline{\beta},\sigma^2;\underline{Y}\right)=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\left(\underline{Y}-\mathbf{X}\underline{\beta}\right)^t\left(\underline{Y}-\mathbf{X}\underline{\beta}\right)\tag{5}$$

De la ecuación (4) obtenemos las ecuaciones normales, es decir:

$$\mathbf{X}^t\mathbf{X}\underline{\beta} = \mathbf{X}^t\underline{Y} \tag{6}$$

Notamos que (6) tiene solución única si y solo si la matrix $\mathbf{X}^t\mathbf{X}$ es invertible ($\mathbf{X}$ sea de rango completo) en cuyo caso el estimador máximo verosímil para $\underline{\beta}$ es:

$$\underline{\hat{\beta}}_{MV} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\underline{Y} \tag{7}$$

Mientras que para σ^2, de la ecuación (5) obtenemos:

$$-n + \frac{1}{\sigma^2}\left(\underline{Y} - \mathbf{X}\underline{\beta}\right)^t\left(\underline{Y} - \mathbf{X}\underline{\beta}\right) = 0 \Rightarrow \sigma^2 = \frac{\left(\underline{Y} - \mathbf{X}\underline{\beta}\right)^t\left(\underline{Y} - \mathbf{X}\underline{\beta}\right)}{n}$$

Por lo tanto al sustituir en la última igualdad lo que obtuvimos en la ecuación (7) obtenemos que estimador máximo verosímil para σ^2 es:

$$\hat{\sigma}^2_{MV} = \frac{\left(\underline{Y} - \mathbf{X}\underline{\hat{\beta}}_{MV}\right)^t\left(\underline{Y} - \mathbf{X}\underline{\hat{\beta}}_{MV}\right)}{n}$$

Definiendo $\underline{\hat{Y}} := \mathbf{X}\underline{\hat{\beta}}_{MV}$ entonces:

$$\hat{\sigma}^2_{MV} = \frac{\left(\underline{Y} - \underline{\hat{Y}}\right)^t\left(\underline{Y} - \underline{\hat{Y}}\right)}{n} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

Los estimadores máximo verosimiles gozan de las siguientes propiedades:

- $\mathbb{E}\left(\underline{\hat{\beta}}_{MV}\right) = \underline{\beta}$ (insesgado)
- $\text{Var}\left(\underline{\hat{\beta}}_{MV}\right) = \sigma^2 \left(\mathbf{X}^t\mathbf{X}\right)^{-1}$
- $\underline{\hat{\beta}}_{MV} \sim N_p \left(\underline{\beta}, \sigma^2 \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\right)$
- $\mathbb{E}\left(\hat{\sigma}^2_{MV}\right) = \sigma^2 \frac{n-p}{n}$ (sesgado)
- $\text{Var}\left(\hat{\sigma}^2_{MV}\right) = 2\frac{n-p}{n^2}\sigma^4$
- $\hat{\sigma}^2_{MV} \sim Gamma\left(\frac{n-p}{2}, \frac{n}{2\sigma^2}\right)$
- $\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$ (Insesgado)
- Como $\underline{\hat{\beta}}_{MV} \sim N_p \left(\underline{\beta}, \sigma^2 \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\right)$, entonces haciendo $\mathbf{C} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}$ y definiendo C_{ij} al elemento (i,j) de la matriz $\mathbf{C}$ se tiene que:

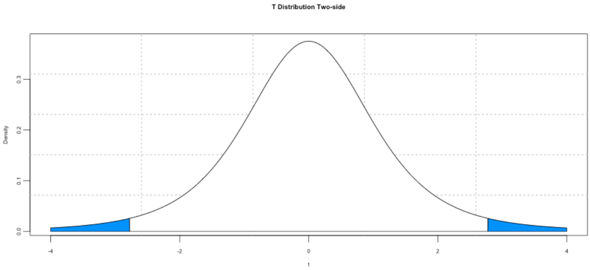
$$\hat{\beta}_i \sim N\left(\beta_i, \sigma^2 C_{(i+1)(i+1)}\right) \Rightarrow \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}} \sim t_{(n-p)} \tag{8}$$

De la última expresión de (8) la inferencia clásica desprende las pruebas de hipótesis e intervalos de confianza correspondientes para el parámetro β_i con $i \in \{0, \ldots, k\}$

Análisis inferencial para $\underline{\beta}$::Pruebas para un coeficiente

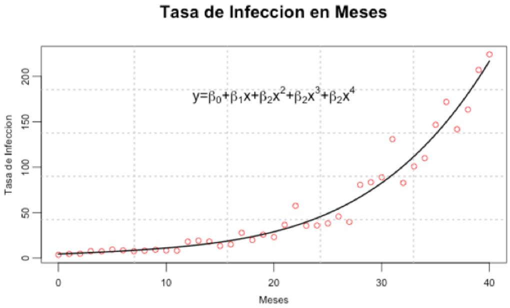
$$H_0 : \beta_j = b_j \quad vs \quad H_1 : \beta_j \neq b_j \quad j \in \{0, \ldots, k\}$$

La regla es rechazar H_0 cuando $|t| \geq \tau_{n-p)}^{1-\alpha/2}$, donde $t = \frac{\hat{\beta}_j - b_j}{\sqrt{\hat{\sigma}^2 C_{(j+1)(j+1)}}}$;



Importante: Esta es la prueba de hipótesis mas importante dentro del análisis de regresión lineal pues cuando hacemos $b_i = 0$ nos ayuda a determinar si la variable asociada a ese coeficiente es *estadísticamente* diferente de cero lo que se traduce en verificar si hay un efecto de la variable X_i en la variable Y.

Ejemplo de pruebas T




```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.571e+00  8.366e+00   0.546    0.588
X[, 2]       4.106e-01  2.974e+00   0.138    0.891
X[, 3]       2.498e-02  3.078e-01   0.081    0.936
X[, 4]      -8.855e-04  1.163e-02  -0.076    0.940
X[, 5]       8.301e-05  1.442e-04   0.576    0.568

Residual standard error: 12.32 on 36 degrees of freedom
Multiple R-squared:  0.963,     Adjusted R-squared:  0.9589
F-statistic: 234.1 on 4 and 36 DF,  p-value: < 2.2e-16
```

Análisis inferencial para $\underline{\beta}$::Intervalo para un coeficiente

La estadística $t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}}$ es una cantidad pivotal y por tanto puede ser utilizada para construir intervalos de confianza:

$$\mathbb{P}\left(-\tau_{n-p}^{1-\alpha/2} \leq \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}} \leq \tau_{n-p)}^{1-\alpha/2} \right) = 1 - \alpha$$

$$\mathbb{P}\left(\hat{\beta}_i - \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}\tau_{n-p)}^{1-\alpha/2} \leq \beta_i \leq \hat{\beta}_i + \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}\tau_{n-p)}^{1-\alpha/2} \right) = 1 - \alpha$$

Por lo tanto un intervalo de confianza al $(1-\alpha)$ % es:

$$\left(\hat{\beta}_i - \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}\tau_{n-p)}^{1-\alpha/2}, \hat{\beta}_i + \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}\tau_{n-p)}^{1-\alpha/2} \right)$$

Donde $\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}$ se le conoce como error estándar del estimador.
$i \in \{0, \ldots, k\}$

Análisis de Varianza La tabla ANOVA nos ayuda a contrastar la hipótesis mas importante de la regresión.

$$H_0 : \beta_1 = \beta_2 = \ldots \beta_k = 0 \quad H_1 : \beta_j \neq 0 \text{ para algún } j \in \{1, \ldots, k\}$$

Var.	S. C.	G. Lib	S.C.M.	F
(SCR)	$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$	$p-1$	$\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{p-1}$	$\frac{(n-p)\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{(p-1)\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
(SCE)	$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$	$n-p$	$\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}$	
(SCT)	$\sum_{i=1}^{n} (y_i - \bar{y})^2$	$n-1$	$\frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n-1}$	

Y se rechaza H_0 al nivel de significancia α si el valor del estadístico F toma valores mas grandes que el cuantíl $F_{(p-1,n-p)}^{1-\alpha}$.

Región de confianza simultaneo para $\underline{\beta}$

Como hemos visto, los intervalos de confianza individuales para los coeficientes tienen el problema que se deben de interpretar de forma independiente, surge entonces la necesidad de crear regiones de confianza que nos de una idea de por donde pueden estar los coeficientes de nuestro modelo. Para resolver este problema requerimos una cantidad pivotal para $\underline{\beta}$. Para ello recordemos lo siguiente:

$$\underline{\hat{\beta}}_{M.V.} = \underline{\hat{\beta}} \sim N_p\left(\underline{\beta}, \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)$$

Entonces por propiedades de la Distribución Normal Multivariada sabemos que:

$$\frac{1}{\sigma}\left(\mathbf{X}^T\mathbf{X}\right)^{\frac{1}{2}}\left(\underline{\hat{\beta}} - \underline{\beta}\right) \sim N_p\left(\underline{0}, \mathbf{I}\right)$$

Por lo tanto:

$$\left(\frac{1}{\sigma}\left(\mathbf{X}^T\mathbf{X}\right)^{\frac{1}{2}}\left(\underline{\hat{\beta}} - \underline{\beta}\right)\right)^T \frac{1}{\sigma}\left(\mathbf{X}^T\mathbf{X}\right)^{\frac{1}{2}}\left(\underline{\hat{\beta}} - \underline{\beta}\right) \sim \chi^2_p$$

Concluimos entonces que:

$$\frac{1}{\sigma^2}\left(\hat{\underline{\beta}}-\underline{\beta}\right)^T\left(\mathbf{X}^T\mathbf{X}\right)\left(\hat{\underline{\beta}}-\underline{\beta}\right)\sim\chi^2_p$$

Desafortunadamente esto no nos sirve como cantidad pivotal pues depende de σ^2 que es desconocida, por lo tanto procedemos a hacer el conciente entre otra χ^2, en este caso utilizaremos el hecho que:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2\sim\chi^2_{n-p}$$

Luego se demuestra que $\hat{\underline{\beta}}$ y $\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$ son independientes, por lo que concluimos:

$$\frac{(n-p)\left(\hat{\underline{\beta}}-\underline{\beta}\right)^T\left(\mathbf{X}^T\mathbf{X}\right)\left(\hat{\underline{\beta}}-\underline{\beta}\right)}{p\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}\sim F_{(p,n-p)}$$

$$\frac{1}{p}\left(\hat{\underline{\beta}}-\underline{\beta}\right)^T\hat{\sigma}^{-2}\left(\mathbf{X}^T\mathbf{X}\right)\left(\hat{\underline{\beta}}-\underline{\beta}\right)\sim F_{(p,n-p)} \tag{9}$$

La cual ya es una cantidad pivotal para $\underline{\beta}$

Región de confianza simultaneo para $\underline{\beta}$

Entonces sabemos que si $F^{1-\alpha}_{(p,n-p)}$ es el cuantíl $(1-\alpha)$ de una distribución $\mathscr{F}$ con $(p, n-p)$ grados de libertad entonces:

$$\mathbb{P}\left(\frac{(n-p)\left(\hat{\underline{\beta}}-\underline{\beta}\right)^T\left(\mathbf{X}^T\mathbf{X}\right)\left(\hat{\underline{\beta}}-\underline{\beta}\right)}{p\sum_{i=1}^n\left(y_i-\hat{y}_i\right)^2}\leq F^{1-\alpha}_{(p,n-p)}\right)=1-\alpha$$

Por lo tanto:

$$\frac{(n-p)\left(\hat{\underline{\beta}}-\underline{\beta}\right)^T\left(\mathbf{X}^T\mathbf{X}\right)\left(\hat{\underline{\beta}}-\underline{\beta}\right)}{p\sum_{i=1}^n\left(y_i-\hat{y}_i\right)^2}\leq F^{1-\alpha}_{(p,n-p)}$$

Es la región de confianza para $\underline{\beta}$. Debe notarse que como función de $\underline{\beta}$ la ecuación:

$$\frac{(n-p)\left(\hat{\underline{\beta}}-\underline{\beta}\right)^T\left(\mathbf{X}^T\mathbf{X}\right)\left(\hat{\underline{\beta}}-\underline{\beta}\right)}{p\sum_{i=1}^n\left(y_i-\hat{y}_i\right)^2}=F^{1-\alpha}_{(p,n-p)}$$

No es más que la expresión de una elipse en el espacio $\mathbb{R}^p$

Inferencia sobre σ^2

Por otro lado la inferencia sobre σ^2 bajo el enfoque clásico se resumen en la cantidad:

$$\frac{\left(\underline{Y}-\underline{\hat{Y}}\right)^t\left(\underline{Y}-\underline{\hat{Y}}\right)}{\sigma^2}=\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sigma^2}\sim\chi^2_{n-p}$$

De aquí por ejemplo se puede construir un intervalos de confianza para σ^2 (no es el óptimo)

$$\left(\frac{1}{\chi^{2\ (1-\alpha/2)}_{n-p}}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2,\frac{1}{\chi^{2\ (\alpha/2)}_{n-p}}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2\right)$$

Inferencia sobre la variable respuesta y

Uno de los usos más importantes de la regresión lineal es poder hacer inferencias sobre el valor de la variable respuesta y cuando las covariables tomaron cierto valor valor digamos $\underline{x}^* = (x_1^*, \ldots, x_k^*)$. La forma natural es hacer la inferencia utilizando la relación funcional entre $\mathbb{E}(Y \mid \underline{x}^*)$ y $\underline{x}^*$ y entonces estimar el valor como:

$$\hat{y}^* = \underline{x}^{*^t} \underline{\hat{\beta}_{MV}} = \underline{x}^{*^t} \left(\mathbf{X}^t\mathbf{X}\right)^{-1} \mathbf{X}^t \underline{Y}$$

Como $\underline{Y}$ es un vector normal-multivariado se sigue que $\hat{y}^*$ sigue una distribución normal

$$\hat{y}^* \sim N_1 \left(\mathbb{E}(y \mid \underline{x}^*), \sigma^2 \underline{x}^{*^t} \left(\mathbf{X}^t\mathbf{X}\right)^{-1} \underline{x}^*\right)$$

Por lo tanto se demuestra que:

$$\frac{\hat{y}^* - \mathbb{E}(y \mid \underline{x}^*)}{\sqrt{\hat{\sigma}^2 \underline{x}^{*t} \left(\mathbf{X}^t\mathbf{X}\right)^{-1} \underline{x}^*}} \sim t_{n-p}$$

La cual es la cantidad de donde se desprende el intervalo de confianza y las pruebas de hipótesis relacionadas para el valor $\mathbb{E}(y \mid \underline{x}^*)$

Inferencia sobre la variable respuesta y
Despejando obtenemos que el Intervalo al $(1-\alpha)\,\%$ de confianza para la respuesta media es:

$$\left(\hat{y}^* - t_{(n-p)}^{(1-\frac{\alpha}{2})}\sqrt{\hat{\sigma}^2 \underline{x}^{*t}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\underline{x}^*}, \hat{y}^* + t_{(n-p)}^{(1-\frac{\alpha}{2})}\sqrt{\hat{\sigma}^2 \underline{x}^{*t}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\underline{x}^*}\right)$$

Sin embargo esto no resuelve el problema sobre hacer inferencias sobre y^*, pues el intervalo anterior está hecho para la respuesta media es decir para $\mathbb{E}(Y \mid \underline{x}^*)$, si deseamos hacer inferencias sobre y^* recordemos entonces que nuestro modelo impone que:

$$y^* = {\underline{x}^*}^t \underline{\beta} + \varepsilon^* \quad \varepsilon^* \sim N\left(0, \sigma^2\right)$$

Haciendo uso de las propiedades distribucionales se prueba entonces que:

$$\frac{\hat{y}^* - y^*}{\sqrt{\hat{\sigma}^2\left(1 + \underline{x}^{*t}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\underline{x}^*\right)}} \sim t_{n-p}$$

De donde finalmente se obtiene un intervalo de **predicción**.

Inferencia sobre la variable respuesta y
Despejando obtenemos que el Intervalo de predicción al $(1-\alpha)\,\%$ para la variable respuesta y^* está dado por

$$\left(\hat{y}^* \pm t_{(n-p)}^{(1-\frac{\alpha}{2})}\sqrt{\hat{\sigma}^2\left(1+\underline{x}^{*^t}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}\underline{x}^*\right)}\right)$$

Debe de observarse que este intervalo incrementa su longitud respecto al intervalo anterior, esto se debe a que este intervalo reconoce dos tipos de incertidumbres, la primera referente a la estimación de los parámetros utilizada en $\hat{y}^* = \underline{x}^{*^t}\underline{\hat{\beta}_{MV}}$ y la segunda referente a la variabialidad presente en la definición del modelo al suponer $\varepsilon \sim N\left(0, \sigma^2\right)$

¿ Como atacaria este problema Thomas Bayes ?

Introducción a la Estadística Bayesiana

El problema en la estadística inferencial puede resumirse en lo siguiente: Supongamos que tenemos $X_1, \ldots, X_n$ m.a. de una distribución F y se nos pide hacer inferencias sobre un nuevo valor X^*. Ciertamente lo anterior es un problema difícil si no conocemos nada de F. Una forma de reducir el problema y darle solución es pensar que $F \in \mathscr{P}$ donde $\mathscr{P}$ es una familia paramétrica de densidades supongamos

$$\mathscr{P} = \{p\,(x|\theta) : \theta \in \Theta\}$$

Es decir en esta primera aproximación **reducimos el problema en hacer inferencias sobre θ.**!!!

Introducción a la Estadística Bayesiana

- Desde el enfoque bayesiano la incertidumbre **inicial** que se tiene sobre θ debe de ser modelada con una medida de probabilidad.
- En este enfoque se crea una variable aleatoria θ que modela esta incertidumbre mediante una densidad inicial $p(\theta)$.
- Es gracias al **teorema de Bayes** que de forma natural nuestro conocimiento inicial se combina con la información muestral para obtener un **densidad final** $p(\theta|\underline{x})$, donde $\underline{x} = (x_1, \ldots, x_n)$ se considera una muestra observada del modelo $p(x|\theta)$

$$\mathbb{P}(A\,|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \Longleftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A\,|B)$$

$$p(\theta|\underline{x}) = \frac{p(\theta, \underline{x})}{p(\underline{x})} = \frac{p(\underline{x}|\theta)\,p(\theta)}{\int p(\theta, \underline{x})\,d\theta} = \frac{p(\underline{x}|\theta)\,p(\theta)}{\int p(\underline{x}|\theta)\,p(\theta)\,d\theta} \propto p(\underline{x}|\theta)\,p(\theta)$$

Introducción a la Estadística Bayesiana

En los problemas de inferencia muchas veces nos vemos en la necesidad de pronosticar el valor de una nueva observación x^* del modelo. Desde el punto de vista bayesiano esto se puede obtener de manera muy natural pues ataca el problema encontrando una densidad predictiva inicial y posteriormente utilizando **Bayes** actualiza esta medida de probabilidad con la información recibida en la muestra.

- Densidad predictiva inicial:

$$p\left(x^*\right) = \int p\left(x^*,\theta\right) d\theta = \int p\left(x^*|\theta\right) p\left(\theta\right) d\theta$$

- Densidad predictiva final:

$$\begin{aligned} p\left(x^*|\underline{x}\right) &= \int p\left(x^*,\theta|\underline{x}\right) d\theta = \int \frac{p\left(x^*,\theta,\underline{x}\right)}{p\left(x^*\right)} d\theta = \int \frac{p\left(x^*|\theta,\underline{x}\right) p\left(\underline{x},\theta\right)}{p\left(x^*\right)} d\theta \\ &= \int p\left(x^*|\theta,\underline{x}\right) p\left(\theta|\underline{x}\right) d\theta = \int p\left(x^*|\theta\right) p\left(\theta|\underline{x}\right) d\theta \end{aligned}$$

Introducción a la Estadística Bayesiana

En la última igualdad se utiliza que:

$$p\left(x^{*}|\theta,\underline{x}\right)=p\left(x^{*}|\theta\right)$$

Lo anterior es válido por el supuesto de **independencia condicional** en donde asumimos que $x_1,\ldots,x_n,x^{*}$ son independientes condicionalmente a θ, es decir:

$$p\left(\underline{x},x^{*}|\theta\right)=p\left(x_1,\ldots,x_n,x^{*}|\theta\right)=p\left(x_1|\theta\right)\ldots p\left(x_n|\theta\right)p\left(x^{*}|\theta\right)=p\left(\underline{x}|\theta\right)p\left(x^{*}|\theta\right)$$

Esto se aplica de la siguiente manera:

$$\begin{aligned}
p\left(x^{*}|\theta,\underline{x}\right) &= \frac{p\left(x^{*},\theta,\underline{x}\right)}{p\left(\theta,\underline{x}\right)}=\frac{p\left(\underline{x},x^{*}|\theta\right)p\left(\theta\right)}{p\left(\theta,\underline{x}\right)}\\
&= \frac{p\left(\underline{x}|\theta\right)p\left(x^{*}|\theta\right)p\left(\theta\right)}{p\left(\theta,\underline{x}\right)}=\frac{p\left(\underline{x}|\theta\right)p\left(x^{*}|\theta\right)p\left(\theta\right)}{p\left(\underline{x}|\theta\right)p\left(\theta\right)}\\
&= p\left(x^{*}|\theta\right)
\end{aligned}$$

Introducción a la Estadística Bayesiana: Ejemplo

Sea $x_1, \ldots, x_n$ observaciones del modelo *Bernulli* (θ). Se asume entonces que:

$$p\left(x_i|\theta\right) = \theta^{x_i}\left(1-\theta\right)^{1-x_i}\mathbf{1}_{\{0,1\}}\left(x_i\right)$$

Se desea hacer inferencias sobre θ del cual se desconoce su valor. Supongamos que modelamos esa incertidumbre inicial (antes de observar la muestra) por medio de una densidad Beta tal que $\theta \sim Beta\left(\alpha_0, \beta_0\right)$

$$p\left(\theta\right) \propto \theta^{\alpha_0-1}\left(1-\theta\right)^{\beta_0-1}\mathbf{1}_{(0,1)}\left(\theta\right)$$

Como:

$$p\left(\underline{x}|\theta\right) = \theta^{\sum_{i=1}^{n} x_i}\left(1-\theta\right)^{n-\sum_{i=1}^{n} x_i}\prod_{i=1}^{n}\mathbf{1}_{\{0,1\}}\left(x_i\right)$$

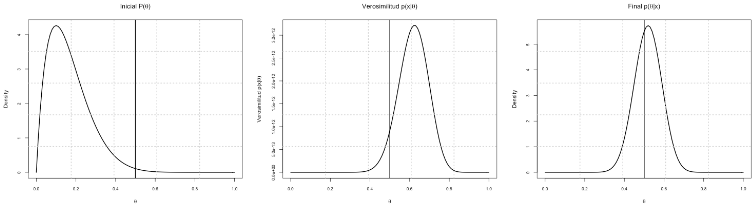
Aplicando mecanicamente la regla de bayes:

$$p\left(\theta|\underline{x}\right) \propto \theta^{\sum_{i=1}^{n} x_i+\alpha_0-1}\left(1-\theta\right)^{n-\sum_{i=1}^{n} x_i+\beta_0-1}\mathbf{1}_{(0,1)}\left(\theta\right)$$

Introducción a la Estadística Bayesiana: Ejemplo

Notemos que:

$$\underbrace{\theta \sim Beta\left(\alpha_0, \beta_0\right)}_{Inicial} \quad + \quad \underbrace{p\left(\underline{x}|\theta\right)}_{muestra} \Rightarrow \underbrace{\theta|\underline{x} \sim Beta\left(\sum_{i=1}^{n} x_i + \alpha_0, n - \sum_{i=1}^{n} x_i + \beta_0\right)}_{Final}$$



Cuando esto ocurre decimos que la familia Beta es conjugada del modelo Bernulli

Introducción a la Estadística Bayesiana: Ejemplo

- En la estadística Bayesiana, con la distribución final $p\left(\theta|\underline{x}\right)$ se hacen las inferencias sobre el parámetro desconocido
- Muchas veces necesitaremos manejar estimaciones puntuales.
- La estadística bayesiana resuelve este problema apoyandose en la **Teoría de la Decisión**.
- Definir una función de perdida que refleje los costos de haber elegido θ^* como estimación de θ.

$$L\left(\theta^*, \theta\right) = \left(\theta^* - \theta\right)^2; \qquad L\left(\theta^*, \theta\right) = \left|\theta^* - \theta\right|;$$

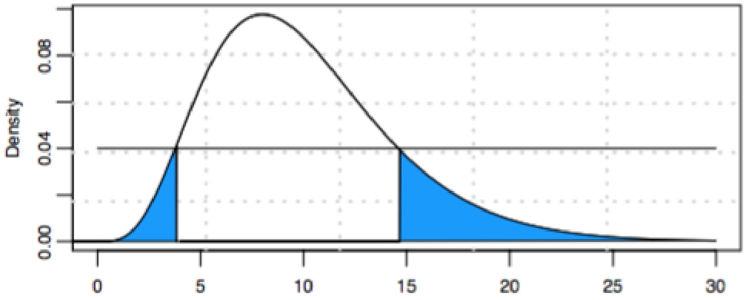
- Se prueba, bajo ciertos axiomas, que la mejor decisón es tomar θ^* que minimice la pérdida esperada.

$$L\left(\theta^*\right) = \mathbb{E}(L\left(\theta^*, \theta\right)) = \int L\left(\theta^*, \theta\right) p\left(\theta|\underline{x}\right) d\theta$$

- Se puede probar que si utiliza $L\left(\theta^*, \theta\right) = \left(\theta^* - \theta\right)^2$, entonces la mejor decisión es tomar como estimador puntual a $\mathbb{E}(\theta)$, mientras que si $L\left(\theta^*, \theta\right) = \left|\theta^* - \theta\right|$ entonces el estimador puntual es la mediana de la distribución final $p\left(\theta|\underline{x}\right)$

Introducción a la Estadística Bayesiana: Ejemplo

- Qué pasa si queremos proporcionar un intervalo?
- En estadística bayesiana, se generan **intervalos de credibilidad** en donde se construyen regiones con probabilidad $1-\alpha$, $\alpha \in (0,1)$ de minima longitud (área)
- Los intervalos son construidos a partir de la distribución final $p\left(\theta|\underline{x}\right)$



Introducción a la Estadística Bayesiana: Ejemplo

- En 1965 (Box-Tiao) definen las **Regiones de máxima probabilidad**. Son regiones de menor tamaño (volumen) que maximizan la probabilidad final.

Definición

Supongamos $\theta \in \mathbb{R}^p$ *y sea* $D \subset \mathbb{R}^P$ *una región en* $\mathbb{R}^p$, D *es llamada la región de Máxima Probabilidad de tamaño* α *si*

- $\mathbb{P}\left(\theta \in D|\underline{x}\right) = 1 - \alpha$
- *Si* $\theta_1 \in D$ *y* $\theta_2 \notin D$ *entonces* $p\left(\theta_1|\underline{x}\right) \geq p\left(\theta_2|\underline{x}\right)$

Se prueba que de todas la regiones de tamaño $1 - \alpha$, D es la de menor volumen y es única si la densidad final $p\left(\theta|\underline{x}\right)$ **no** es uniforme en alguna parte del espacio parametral Θ.

Introducción a la Estadística Bayesiana: Ejemplo

- Otra forma de llevar a cabo inferencia sobre los parámetros es realizar pruebas de hipótesis.

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1;$$

  Con $\Theta_0 \cup \Theta_1 = \Theta$ una partición del espacio parametral.

- Resulta natural decidirnos por la hipótesis que tenga mas probabilidad de ocurrir bajo la distribución final.
- Decido por H_0 si:

$$\mathbb{P}\left(\theta \in \Theta_0 | \underline{x}\right) > \mathbb{P}\left(\theta \in \Theta_1 | \underline{x}\right)$$

- Decido por H_1 si:

$$\mathbb{P}\left(\theta \in \Theta_0 | \underline{x}\right) < \mathbb{P}\left(\theta \in \Theta_1 | \underline{x}\right)$$

- En general este problema también puede plantearse usando **Teoría de la Decisión**.

Introducción a la Estadística Bayesiana: Ejemplo

- Imaginemos que planteamos

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0;$$

- Supongamos que θ es v.a. continua.
- Eso implicaría que:

$$p\left(\theta = \theta_0 | \underline{x}\right) = 0; \quad p\left(\theta \neq \theta_0 | \underline{x}\right) = 1$$

- Bajo el criterio de decidirnos por la hipótesis con mayor probabilidad, siempre estaríamos optando por rechazar H_0. (Por eso en la versión clásica no aceptamos H_0)
- Para atacar este problema se puede utilizar el criterio de Lindley (1965).
- Se genera el **intervalo de credibilidad** para θ al $(1-\alpha)\,\%$, si θ_0 no está en el intervalo se rechaza H_0.

Familias Conjugadas

Definición (Familia Conjugada)

Sea $\wp$ *una familia paramétrica de densidades:*

$$\mathscr{P} = \{p\,(x|\theta) : \theta \in \Theta\}$$

Y sea $\mathscr{F}$ *una familia de distribuciones a priori para* θ, *decimos que* $\mathscr{F}$ *es conjugada de* $\mathscr{P}$ *si para toda* $p(\theta) \in \mathscr{P}$ *se tiene que*
$p\,(\theta|\underline{x}) \propto p\,(\underline{x}|\theta)\,p\,(\theta) \in \mathscr{P}$

Introducción a la Estadística Bayesiana

Algunos Ejemplos:

- $x \sim Poisson\,(x|\lambda) \rightarrow \lambda \sim Gamma\,(\lambda|\alpha_0, \beta_0)$
- $x \sim Binomial\,(x|n^*, \theta) \rightarrow \theta \sim Beta\,(\theta|\alpha_0, \beta_0)$
- $x \sim Multinomial\,(x|p_1, \ldots, p_k, n^*) \rightarrow (p_1, \ldots, p_k) \sim Dir\,(p_1, \ldots, p_k|\alpha_1, \ldots, \alpha_k)$
- $x \sim Geometric\,(x|\theta) \rightarrow \theta \sim Beta\,(\theta|\alpha_0, \beta_0)$
- $x \sim U\,(x|0, \theta) \rightarrow \theta \sim Pareto\,(\theta|\alpha_0, \beta_0)$
- $x \sim Gamma\,(x|\alpha^*, \beta) \rightarrow \beta \sim Gamma\,(\beta|\alpha_0, \beta_0)$
- $x \sim Normal\,(x|\mu, \tau^*) \rightarrow \mu \sim Normal\,(\mu|\mu_0, \tau_0)$, donde: $\tau = \frac{1}{\sigma^2}$ se le conoce como la precisión del modelo.
- $x \sim Normal\left(x|\mu^{*1}, \tau\right) \rightarrow \tau \sim Gamma\,(\tau|\alpha_0, \beta_0)$
- $x \sim Normal\,(x|\mu, \tau) \rightarrow (\mu, \tau) \sim Normal - Gamma\,(\mu, \tau|\mu_0, \tau_0, \alpha_0, \beta_0)$

[1]* Parametros se consideran conocidos

Introducción a la Estadística Bayesiana

¿Cómo determinar una familia conjugada para un modelo $p\,(x|\theta)$?
Una estrategia para determinar una familia conjugada para un modelo $p\,(x|\theta)$ es calcular la verosimilitud y luego ver la verosimilitud como función de θ e identificar el **Kernell** de una distribución conocida.
Ejemplo: Supongamos

$$p\,(x|\theta) = \frac{\theta^x}{x!}e^{-\theta}$$

entonces:

$$p\,(\mathbf{x}|\theta) = \prod_{i=1}^{n} \frac{1}{x_i!}\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}$$

Visto como función de θ

$$p\,(\theta) \propto \theta^{\sum_{i=1}^{n} x_i} e^{-n\theta} \propto \underbrace{\theta^{\alpha_0 - 1} e^{-\beta_0 \theta}}_{Kernell-Gamma}$$

Donde $\alpha_0 = \sum_{i=1}^{n} x_i + 1$ y $\beta_0 = n$. Por lo tanto se propone una densidad Gamma como inicial, la cual se puede probar es conjugada del modelo Poisson

Introducción a la Estadística Bayesiana

Distribuciones No Informativas

- Muchas veces estaremos en condiciones de no tener ningún tipo de información inicial y por tanto tenemos que reflejar esa ignorancia en distribuciones iniciales $p(\theta)$ poco informativas.
- Una estrategia usual es proponer iniciales con mucha varianza (poca precisión) lo que reflejaría esa ignorancia del valor de θ.
- En la literatura se han documentado 2 técnicas que generan distribuciones no informativas
  - Distribución de Jeffreys
  - Análisis de Referencia
- Desafortunadamente estás últimas técnicas en muchas ocasiones generan distribuciones iniciales **impropias**, es decir distribuciones iniciales que no necesariamente integran 1!!!!

Introducción a la Estadística Bayesiana

Imaginemos que estamos en el modelo normal con τ conocida.

$$p\left(x|\mu\right) \propto \tau^{\frac{1}{2}} e^{\frac{\tau}{2}(x-\mu)^2}; \quad p\left(\mu|\mu_0,\tau_0\right) = N\left(\mu|\mu_0,\tau_0\right) \propto \tau_0^{\frac{1}{2}} e^{-\frac{\tau_0}{2}(\mu-\mu_0)^2}$$

Se prueba que bajo esta inicial conjugada se tiene que:

$$p\left(\mu|\underline{x}\right) = N\left(\mu|\mu_1,\tau_1\right)$$

Con $\mu_1 = \frac{\tau_0\mu_0+n\tau\overline{x}}{\tau_0+n\tau}$ y $\tau_1 = n\tau + \tau_0$
Una forma de suponer falta de información es colocar una precisión muy baja en la inicial de μ, esto se reflejaria suponiendo que $\tau_0 \to 0$, si hacemos esto en la final obtendramos que:

$$p\left(\mu|\underline{x}\right) = N\left(\mu|\overline{x},n\tau\right)$$

¿Que inicial necesito para llegar a esta final?

$$p\left(\mu\right) \propto 1$$

Una distribución impropia!!!!!!

Introducción a la Estadística Bayesiana

Regla de Jefreys Una idea que propuso Jeffreys fue colocar como prior lo siguiente:
Pensando $\theta \in \mathbb{R}^p$, entonces:

$$p(\theta) \propto \sqrt{det(\mathbf{I}(\theta))}$$

Donde:

$$(\mathbf{I}(\theta))_{ij} = -\mathbb{E}\left(\frac{\partial^2 \log(p(x|\theta))}{\partial\theta_i\partial\theta_j}\right)$$

La idea de esta propuesta es la siguiente, $\mathbf{I}(\theta)$ mide la información que obtenemos sobre θ por unidad muestral, esta información queda en función de θ que es desconocido, por ejemplo suponiendo $p(x|\theta) = Poi(\theta)$, entonces:

$$\mathbf{I}(\theta) = \frac{1}{\theta} \Rightarrow p(\theta) \propto \theta^{-\frac{1}{2}} \Rightarrow p(\theta|\mathbf{x}) = Ga\left(\sum_i^n x_i + \frac{1}{2}, n\right)$$

Si θ es cercano a 0, entonces la información por unidad muestral es mas grande, Jefreys entonces propone que demos mayor probabilidad a ese tipo de valores de θ para maximizar la información que proporcionará la muestra.

Introducción a la Estadística Bayesiana

Otro Ejemplo Supongamos nuevamente el modelo Normal con τ conocida:

$$p\left(x|\mu\right) \propto \tau^{\frac{1}{2}} e^{\frac{\tau}{2}(x-\mu)^2}$$

Entonces según Jeffreys la prior es:

$$\begin{aligned} p\left(\mu\right) &\propto \sqrt{\mathbf{I}\left(\mu\right)} = \sqrt{-\mathbb{E}\left(\frac{\partial^2}{\partial^2 \mu} \log p\left(x|\mu\right)\right)} \\ &\propto \sqrt{\tau} \propto \tau^{\frac{1}{2}} \propto 1 \end{aligned}$$

Lo último se debe a que τ es conocida y por tanto como función de μ es proporcional a 1, es decir obtenemos la misma inicial (**impropia**) bajo el supuesto de que en la inicial conjugada se hace $\tau_0 \to 0$

Introducción a la Estadística Bayesiana

Una de las desventajas del método de Jeffreys es que genera distribuciones impropias y esto muchas veces puede ocasionar que se obtengan distribuciones finales que también sean impropias, el método basado en el **análisis de referencia** cubre este problema. La idea básica es:

- Dada una familia parámetrica $\{p(x|\theta) : \theta \in \Theta\}$, identificar $p(\theta)$ que sea no informativa en el sentido de que tenga un efecto mínimo en los análisis.
- Utilizar teoría de la información. Medir la cantidad de información total que tenemos del parámetro (muestra + inicial)
- Entoncontrar $p^*(\theta)$ que minimice su aporte a la información total dejando la mayor parte a la información en muestra.
- $p^*(\theta)$ se le conoce como de referencia pues a partir de ella se puede medir cuanta información aportan otras iniciales.
- $p^*(\theta)$ son usualmente impropias, pero tienen la ventaja de generar finales propias!!!!![2]

[2] Ver libro de Bernando y Smith, Bayesian Theory. (2009)

Introducción a la Estadística Bayesiana

En los modelos lineales es común utilizar distribuciones Gaussianas por lo que es necesario estudiar con mas detalle dicha distribución desde el enfoque bayesiano:

$$p\left(x|\mu,\tau\right)=\frac{\tau^{\frac{1}{2}}}{\sqrt{2\pi}}e^{-\frac{\tau}{2}(x-\mu)^2}\propto\tau^{\frac{1}{2}}e^{-\frac{\tau}{2}(x-\mu)^2}$$

Inicial:

$$\begin{aligned}p\left(\mu,\tau|\mu_0,\tau_0,\alpha_0,\beta_0\right) &= \underbrace{\frac{\left(\tau\tau_0\right)^{\frac{1}{2}}}{\sqrt{2\pi}}e^{-\frac{(\tau\tau_0)}{2}(\mu-\mu_0)^2}}_{N(x|\mu_0,\tau\tau_0)}\underbrace{\frac{\beta_0^{\alpha_0}}{\Gamma\left(\alpha_0\right)}\tau^{\alpha_0-1}e^{-\beta_0\tau}}_{Gamma(\tau|\alpha_0,\beta_0)}\mathbf{1}_{(0,\infty)}\left(\tau\right)\\ &\propto \tau^{\alpha_0-\frac{1}{2}}e^{-\tau\left(\frac{\tau_0}{2}(\mu-\mu_0)^2+\beta_0\right)}\mathbf{1}_{(0,\infty)}\left(\tau\right)\end{aligned}$$

A la última densidad se le conoce como **Normal-Gamma**, es una distribución bivariada con 4 parámetros.

Introducción a la Estadística Bayesiana

La distribución Normal-Gamma tiene las siguientes carácteristicas:

- Marginal de μ:

$$
\begin{aligned}
p\left(\mu|\mu_0,\tau_0,\alpha_0,\beta_0\right) &\propto \int_{-\infty}^{\infty} \tau^{\alpha_0-\frac{1}{2}} e^{-\tau\left(\frac{\tau_0}{2}(\mu-\mu_0)^2+\beta_0\right)} \mathbf{1}_{(0,\infty)}\left(\tau\right) d\tau \\
&\propto \int_{0}^{\infty} \tau^{\alpha_0+\frac{1}{2}-1} e^{-\tau\left(\frac{\tau_0}{2}(\mu-\mu_0)^2+\beta_0\right)} d\tau
\end{aligned}
$$

Haciendo $\alpha_0^{'} = \alpha_0 + \frac{1}{2}$ y $\beta_0^{'} = \frac{\tau_0}{2}\left(\mu-\mu_0\right)^2 + \beta_0$, se obtiene el **kernell** de una densidad Gamma y entonces:

$$
\begin{aligned}
p\left(\mu|\mu_0,\tau_0,\alpha_0,\beta_0\right) &\propto \left(\frac{\tau_0}{2}\left(\mu-\mu_0\right)^2+\beta_0\right)^{-\left(\alpha_0+\frac{1}{2}\right)} \\
&\propto \left(1+\frac{\tau_0}{2\beta_0}\left(\mu-\mu_0\right)^2\right)^{-\left(\frac{2\alpha_0+1}{2}\right)}
\end{aligned}
$$

Introducción a la Estadística Bayesiana

$$
\begin{aligned}
p\left(\mu|\mu_0,\tau_0,\alpha_0,\beta_0\right) &\propto \left(1+\frac{\tau_0}{2\beta_0}\left(\mu-\mu_0\right)^2\right)^{-\left(\frac{2\alpha_0+1}{2}\right)} \\
&\propto \left(1+\frac{\frac{\tau_0\alpha_0}{\beta_0}\left(\mu-\mu_0\right)^2}{2\alpha_0}\right)^{-\left(\frac{2\alpha_0+1}{2}\right)}
\end{aligned}
$$

Se le conoce como una distribución t student.

$$
p\left(\mu|\mu_0,\tau_0,\alpha_0,\beta_0\right) = \text{t student}\left(\mu\left|2\alpha_0,\mu_0,\frac{\tau_0\alpha_0}{\beta_0}\right.\right)
$$

Y se puede probar que:

$$
\mathbb{E}(\mu)=\mu_0; \qquad \mathrm{Var}(\mu)=\left(\frac{\tau_0\alpha_0}{\beta_0}\right)^{-1}\frac{2\alpha_0}{2\alpha_0-2}=\frac{\beta_0}{\tau_0\left(\alpha_0-1\right)};\alpha_0>1
$$

Introducción a la Estadística Bayesiana

- Marginal de τ:

$$
\begin{aligned}
p\left(\tau|\mu_0,\tau_0,\alpha_0,\beta_0\right) &\propto \int_{-\infty}^{\infty}\frac{\left(\tau\tau_0\right)^{\frac{1}{2}}}{\sqrt{2\pi}}e^{-\frac{(\tau\tau_0)}{2}(\mu-\mu_0)^2}\tau^{\alpha_0-1}e^{-\beta_0\tau}\mathbf{1}_{(0,\infty)}\left(\tau\right)d\mu \\
&\propto \tau^{\alpha_0-1}e^{-\beta_0\tau}\mathbf{1}_{(0,\infty)}\left(\tau\right)\int_{-\infty}^{\infty}\frac{\left(\tau\tau_0\right)^{\frac{1}{2}}}{\sqrt{2\pi}}e^{-\frac{(\tau\tau_0)}{2}(\mu-\mu_0)^2}d\mu \\
&\propto \tau^{\alpha_0-1}e^{-\beta_0\tau}\mathbf{1}_{(0,\infty)}\left(\tau\right)
\end{aligned}
$$

Lo anterior es el **Kernell** de una densidad *Gamma* de parámetros α_0 y β_0. Por lo tanto:

$$
\mathbb{E}(\tau)=\frac{\alpha_0}{\beta_0};\qquad \text{Var}(\tau)=\frac{\alpha_0}{\beta_0^2}
$$

Introducción a la Estadística Bayesiana

Sabemos que **Normal-Gamma** es conjugada de modelo **Normal** por lo tanto distribución final $(\mu, \tau|\underline{x})$ también debe de pertenecer a la familia **Normal Gamma**, en esta caso se prueba que:

$$p\left(\mu, \tau|\underline{x}\right) = Normal - Gamma\left(\mu, \tau|\mu_1, \tau_1, \alpha_1, \beta_1\right)$$

Donde:

- $\mu_1 = \frac{\tau_0\mu_0 + n\overline{x}}{\tau_0 + n}$ donde $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
- $\tau_1 = \tau_0 + n$
- $\alpha_1 = \alpha_0 + \frac{n}{2}$
- $\beta_1 = \beta_0 + \frac{1}{2}\left(ns^2 + \frac{\tau_0 n(\overline{x} - \mu_0)^2}{\lambda_0 + n}\right)$ donde $s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$

Introducción a la Estadística Bayesiana

Busquemos una generalización del modelo Normal-Gamma, ahora supongamos que tenemos al modelo normal multivariado:

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

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

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

Obs: Debemos tener cuidado cuando marginalizamos y utilizamos la matriz de precisión:

$$\begin{aligned}
\mathbf{x} \sim N_p\left(\mathbf{x}|\mu,\Sigma\right) \quad &\Rightarrow \quad \mathbf{x}^{(1)} \sim N_r\left(\mathbf{x}^{(1)}|\mu^{(1)},\Sigma_{11}\right)\\
\mathbf{x} \sim N_p\left(\mathbf{x}|\mu,\mathbf{P}\right) \quad &\nRightarrow \quad \mathbf{x}^{(1)} \sim N_r\left(\mathbf{x}^{(1)}|\mu^{(1)},\mathbf{P}_{11}\right)\\
\mathbf{x} \sim N_p\left(\mathbf{x}|\mu,\mathbf{P}\right) \quad &\Rightarrow \quad \mathbf{x}^{(1)} \sim N_r\left(\mathbf{x}^{(1)}|\mu^{(1)},\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21}\right)
\end{aligned}$$

Teorema (Inversa de Matriz Particionada)

Sea $\boldsymbol{P} \in R^{p\times p}$ *una matriz cuadrada no singular, tal que es particionada como:*

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22} \end{pmatrix}$$

Donde $\boldsymbol{P}_{11} \in \mathbb{R}^{p_1\times p_1}, \boldsymbol{P}_{22} \in \mathbb{R}^{p_2\times p_2}$ *son ambas cuadradas, tal que* $p_1 + p_2 = p$*. Entonces la inversa de* $\boldsymbol{P}$ *está dado por:*

$$\boldsymbol{P}^{-1} = \begin{pmatrix} \boldsymbol{P}^{11} & \boldsymbol{P}^{12} \\ \boldsymbol{P}^{21} & \boldsymbol{P}^{22} \end{pmatrix}$$

Donde:

$$\begin{aligned} \boldsymbol{P}^{11} &:= \left(\boldsymbol{P}_{11} - \boldsymbol{P}_{12}\boldsymbol{P}_{22}^{-1}\boldsymbol{P}_{21}\right)^{-1} := \left(\boldsymbol{P}_{11.2}\right)^{-1} \\ \boldsymbol{P}^{12} &:= -\left(\boldsymbol{P}_{11.2}\right)^{-1}\boldsymbol{P}_{12}\boldsymbol{P}_{22}^{-1} \\ \boldsymbol{P}^{21} &:= -\boldsymbol{P}_{22}^{-1}\boldsymbol{P}_{21}\left(\boldsymbol{P}_{11.2}\right)^{-1} \\ \boldsymbol{P}^{22} &:= \boldsymbol{P}_{22}^{-1} + \boldsymbol{P}_{22}^{-1}\boldsymbol{P}_{21}\boldsymbol{P}_{11.2}^{-1}\boldsymbol{P}_{12}\boldsymbol{P}_{22}^{-1} \end{aligned}$$

Introducción a la Estadística Bayesiana

La densidad t-multivariada pertenece a la clase de distribuciones elipiticas y tiene densidad dada por:

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

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

$$\mathbf{x}\sim T_{p}\left(\mathbf{x}|d,\mu,\mathbf{P}\right)$$

Observe que haciendo p=1 se obtiene la densidad t univariada.
Recordatorio: Decimos que $X\sim EC\left(\mu,\mathbf{P},g\right)$ (Elliptic Curve) si la densidad está dada por:

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

Observe que el caso de la T-multivarada se obtiene haciendo :

$$g\left(t\right)=\left(1+\frac{t}{d}\right)^{-\left(\frac{d+p}{2}\right)}$$

Introducción a la Estadística Bayesiana

Teorema (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* $\boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{P}^{-1} = \Sigma$; $U^{(p)}$ *es un vector aleatorio en* $\mathbb{R}^P$ *sobre la esfera unitaria y* $R^2 = \left(\boldsymbol{X} - \mu\right)^T \boldsymbol{P}\left(\boldsymbol{X} - \mu\right)$ *es la distribución radial.*
- R^2 *tiene por densidad:*

$$h_{R^2}\left(r\right) \propto r^{\frac{p}{2}-1} g\left(r\right)$$

En el caso de la t-multivariada se tiene que:

$$h_{R^2}\left(r\right) \propto r^{\frac{p}{2}-1}\left(1 + \frac{r}{d}\right)^{-\left(\frac{d+p}{2}\right)} \propto r^{\frac{p}{2}-1}\left(d + r\right)^{-\left(\frac{d+p}{2}\right)}$$

De aquí se puede probar que entonces:

$$p^{-1} R^2 \sim F_{(p,d)}$$

Introducción a la Estadística Bayesiana

Una propiedad importante de la t-multivariada es que la densidad de un subvector también sigue una distribución t-multivarada.

$$\mathbf{x} \sim T_p\left(x|d,\mu,\mathbf{P}\right) \Rightarrow \mathbf{x}^{(1)} \sim T_r\left(\mathbf{x}^{(1)}\left|\mu^{(1)},\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21}\right.\right)$$

Observación Importante: El Hecho de que en la matrix $\Sigma = \mathbf{P}^{-1}$ aparezca un 0 en la posición i, j **no implica** que las variables X_i, X_j sean independientes, en el caso t-multivariado, esto solo implica no correlación, se puede probar que el único caso donde la no correlación implica independencia, es en el la distribución **Normal Multivariada**

Introducción a la Estadística Bayesiana

Estamos en condiciones para presentar la generalización de la Normal-Gamma al caso multivariado.
Diremos que el vector aleatorio $(\mathbf{x}, y)$ con $\mathbf{x} \in \mathbb{R}^p$ y $y \in \mathbb{R}$ sigue una densidad Normal Multivariada-Gamma si su densidad está dada por:

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

Se puede probar que:

$$\mathbb{E}\left(\begin{pmatrix}\mathbf{x}\\ y\end{pmatrix}\right) = \begin{pmatrix}\mu\\ \frac{\alpha}{\beta}\end{pmatrix}$$

$$Moda\left(\begin{pmatrix}\mathbf{x}\\ y\end{pmatrix}\right) = \begin{pmatrix}\mu\\ \frac{\alpha+\frac{p}{2}-1}{\beta}\end{pmatrix}$$

Introducción a la Estadística Bayesiana

- Marginal de y

$$
\begin{aligned}
p\left(y|\mu,\mathbf{P},\alpha,\beta\right) &\propto \int_{\mathbb{R}^p} |y\mathbf{P}|^{\frac{1}{2}} e^{-\frac{y}{2}(\mathbf{x}-\mu)^T\mathbf{P}(\mathbf{x}-\mu)} y^{\alpha-1} e^{-y\beta} \mathbf{1}_{(0,\infty)}(y)\, d\mathbf{x} \\
&\propto y^{\alpha-1} e^{-y\beta} \mathbf{1}_{(0,\infty)}(y) \int_{\mathbb{R}^p} |y\mathbf{P}|^{\frac{1}{2}} e^{-\frac{y}{2}(\mathbf{x}-\mu)^T\mathbf{P}(\mathbf{x}-\mu)} d\mathbf{x} \\
&\propto y^{\alpha-1} e^{-y\beta} \mathbf{1}_{(0,\infty)}(y)
\end{aligned}
$$

Por lo tanto $y \sim Ga\left(\alpha,\beta\right)$ de donde concluimos:

$$
\mathbb{E}(y) = \frac{\alpha}{\beta}; \qquad \mathrm{Var}(y) = \frac{\alpha}{\beta^2}; \quad moda\left(y\right) = \frac{\alpha-1}{\beta}
$$

- Marginal de $\mathbf{x}$:

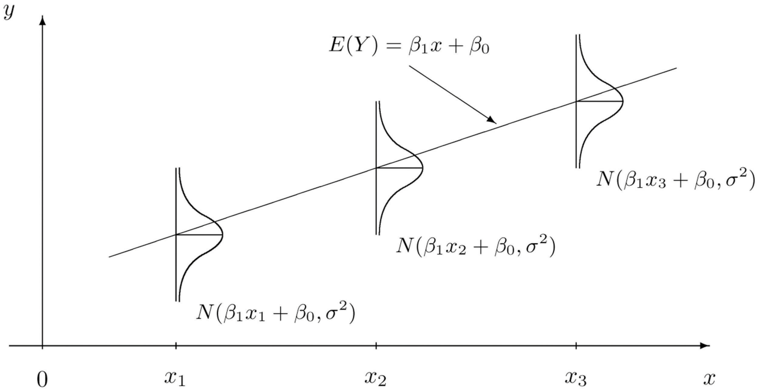
$$
\begin{aligned}
p\left(x|\mu,\mathbf{P},\alpha,\beta\right) &\propto \int_{-\infty}^{\infty} y^{\frac{p}{2}+\alpha-1} e^{-\frac{y}{2}\left((\mathbf{x}-\mu)^T\mathbf{P}(\mathbf{x}-\mu)+2\beta\right)} \mathbf{1}_{(0,\infty)}\left(y\right) dy \\
&\propto \int_{0}^{\infty} \underbrace{y^{\frac{p+2\alpha}{2}-1}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

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

Concluimos que:

- $\mathbb{E}(\mathbf{x}) = \mu$
- $\mathrm{Var}(\mathbf{x}) = \frac{2\alpha}{2\alpha-2}\left(\frac{\alpha}{\beta}\mathbf{P}\right)^{-1} = \frac{\beta}{\alpha-1}\mathbf{P}^{-1}$ (Obs: $\alpha > 1$)

Modelos Lineales Bajo el Enfoque Bayesiano



Regresión Lineal : Enfoque Bayesiano

Sea $\underline{\beta} \in \mathbb{R}^p$ un vector de parámetros, y $\mathbf{X} \in \mathbb{R}^{n\times p}$ una matriz de diseño conocida. Definamos el modelo lineal:

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon} \tag{10}$$

Donde $\underline{\varepsilon} \sim N_n\left(\underline{\varepsilon}|\underline{0}, \tau\mathbf{I}_n\right)$ con $\tau = \sigma^{-2} > 0$. Suponemos una matriz de precisión con correlación 0 entre la variables lo que implica independencia entre las ε_i. De la ecuación (10) concluimos que:

$$\underline{y} \sim N_n\left(\underline{y}|\mathbf{X}\beta, \tau\mathbf{I}_n\right)$$

Objetivo: suponiendo que observamos $\underline{y}$ **inferir** sobre los parámetros $\underline{\beta}$ y $\tau = \sigma^{-2}$

Regresión Lineal : Enfoque Bayesiano

Inferencia es el procedimiento por medio del cual extraemos información sobre lo parámetros $(\underline{\beta}, \tau)$ a partir de una muestra $\underline{y}$. La inferencia se puede resumir en tres técnicas báscas.

- Estimación puntual $(\hat{\underline{\beta}}, \hat{\tau})$
- Estimación Intervalar $\mathbb{P}\left(L < \tau < U\right) = 0.95$
- Pruebas de hipótesis:

$$H_0 : \underline{\beta} = \underline{0} \qquad vs. \qquad H_1 : \underline{\beta} \neq \underline{0}$$

Solucion: Encontrar las distribuciones finales:

$$p\left(\underline{\beta}, \tau | \underline{y}\right) \propto p\left(\underline{y} | \underline{\beta}, \tau\right) p\left(\underline{\beta}, \tau\right)$$

$$p\left(\underline{\beta} | \underline{y}\right) = \int_0^\infty p\left(\underline{\beta}, \tau | \underline{y}\right) d\tau; \qquad p\left(\tau | \underline{y}\right) = \int_{\mathbb{R}^p} p\left(\underline{\beta}, \tau | \underline{y}\right) d\underline{\beta};$$

Tenemos que proponer una distribución conjugada, para ello estudiemos la verosimilitud.

$$\begin{aligned} p\left(\underline{y}|\underline{\beta},\tau\right) &= N_n\left(\underline{y}|\mathbf{X}\underline{\beta},\tau\mathbf{I}_n\right) \\ &\propto \tau^{\frac{n}{2}} e^{-\frac{\tau}{2}\left(\underline{y}-\mathbf{X}\underline{\beta}\right)^T\left(\underline{y}-\mathbf{X}\underline{\beta}\right)} \end{aligned}$$

Pero se puede probar que:

$$\left(\underline{y}-\mathbf{X}\underline{\beta}\right)^T\left(\underline{y}-\mathbf{X}\underline{\beta}\right) = \left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right) + \left(\underline{\beta}-\hat{\underline{\beta}}\right)^T\mathbf{X}^T\mathbf{X}\left(\underline{\beta}-\hat{\underline{\beta}}\right)$$

Donde

$$\hat{\underline{\beta}} = \hat{\underline{\beta}}_{MV} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\underline{y}$$

Entonces:

$$p\left(\underline{y}|\underline{\beta},\tau\right) \propto \tau^{\frac{n}{2}} e^{-\frac{\tau}{2}\left(\left(\underline{\beta}-\hat{\underline{\beta}}\right)^T\mathbf{X}^T\mathbf{X}\left(\underline{\beta}-\hat{\underline{\beta}}\right)+\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)\right)} \tag{11}$$

$$\propto \tau^{\frac{n}{2}} e^{-\frac{\tau}{2}\left(\left(\underline{\beta}-\hat{\underline{\beta}}\right)^T\mathbf{X}^T\mathbf{X}\left(\underline{\beta}-\hat{\underline{\beta}}\right)+\tilde{\beta}_0\right)} \tag{12}$$

Visto como función de $\underline{\beta},\tau$ obtenemos el **Kernell** de una distribución Normal-Multivariada - Gamma !!

Regresión Lineal : Enfoque Bayesiano

Proponemos entonces como prior una distribución Normal-Multivariada - Gamma

$$p\left(\underline{\beta},\tau|\mu_0,\mathbf{P}_0,\alpha_0,\delta_0\right)\propto N_p\left(\underline{\beta}|\mu_0,\tau\mathbf{P}_0\right)Ga\left(\tau|\alpha_0,\delta_0\right)$$

donde $\mathbf{P}_0\in\mathbb{R}^{p\times p}$, $\mu_0\in\mathbb{R}^p$, $\alpha_0,\delta_0\in\mathbb{R}$ son hiperparametros que se deben de ajustar para modelar nuestro conocimiento inicial de $\underline{\beta}$ y τ. Un problema realmente dificil, pero existen técnicas para ayudarnos a establecer los valores!! De las propiedades de esta densidad sabemos que las marginales iniciales son:

$$\underline{\beta}\sim T_p\left(\underline{\beta}\,\middle|\,2\alpha_0,\mu_0,\frac{\alpha_0}{\delta_0}\mathbf{P}_0\right)\quad\Rightarrow\quad\mathbb{E}\left(\underline{\beta}\right)=\mu_0;\quad\mathrm{Var}\left(\underline{\beta}\right)=\frac{\delta_0}{\alpha_0-1}\mathbf{P}_0^{-1}$$

$$\tau\sim Ga\left(\tau|\alpha_0,\delta_0\right)\quad\Rightarrow\quad\mathbb{E}(\tau)=\frac{\alpha_0}{\delta_0};\quad\mathrm{Var}(\tau)=\frac{\alpha_0}{\delta_0^2}$$

Obs: El conocimiento inicial de la correlación de los parámetros está ligado con los parámetros de la distribución inicial de τ

Regresión Lineal : Enfoque Bayesiano

Definida la distribución inicial tenemos todos los ingredientes para obtener la final, usando (11)

$$\begin{aligned} p\left(\underline{\beta},\tau|\underline{y}\right) &\propto p\left(\underline{y}|\underline{\beta},\tau\right)p\left(\underline{\beta},\tau\right) \\ &\propto \tau^{\frac{n}{2}}e^{-\frac{\tau}{2}\left(\left(\underline{\beta}-\hat{\underline{\beta}}\right)^T\mathbf{X}^T\mathbf{X}\left(\underline{\beta}-\hat{\underline{\beta}}\right)+\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)\right)} \\ &\quad \tau^{\frac{p}{2}}e^{-\frac{\tau}{2}\left(\underline{\beta}-\mu_0\right)^T\mathbf{P}_0\left(\underline{\beta}-\mu_0\right)}\tau^{\alpha_0-1}e^{-\delta_0\tau} \\ &\propto \tau^{\frac{n}{2}+\alpha_0-1}e^{-\tau\left(\delta_0+\frac{1}{2}\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)\right)} \\ &\quad \tau^{\frac{p}{2}}e^{-\frac{\tau}{2}\left(\left(\underline{\beta}-\hat{\underline{\beta}}\right)^T\mathbf{X}^T\mathbf{X}\left(\underline{\beta}-\hat{\underline{\beta}}\right)+\left(\underline{\beta}-\mu_0\right)^T\mathbf{P}_0\left(\underline{\beta}-\mu_0\right)\right)} \end{aligned}$$

Se demuestra que:

$$\left(\underline{\beta}-\hat{\underline{\beta}}\right)^T\mathbf{X}^T\mathbf{X}\left(\underline{\beta}-\hat{\underline{\beta}}\right)+\left(\underline{\beta}-\mu_0\right)^T\mathbf{P}_0\left(\underline{\beta}-\mu_0\right)=$$
$$\left(\underline{\beta}-\mu_1\right)^T\mathbf{P}_1\left(\underline{\beta}-\mu_1\right)+\hat{\underline{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\underline{\beta}}+\mu_0^T\mathbf{P}_0\mu_0-\mu_1^T\mathbf{P}_1\mu_1$$

Donde:

- $\mathbf{P}_1=\mathbf{X}^T\mathbf{X}+\mathbf{P}_0;\qquad \mu_1=\left(\mathbf{X}^T\mathbf{X}+\mathbf{P}_0\right)^{-1}\left(\mathbf{X}^T\underline{y}+\mathbf{P}_0\mu_0\right)$

Regresión Lineal : Enfoque Bayesiano

Susituyendo y redistribuyendo:

$$p\left(\underline{\beta},\tau|\underline{y}\right) \propto \tau^{\frac{n}{2}+\alpha_0-1}e^{-\tau\left(\delta_0+\frac{1}{2}\left(\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)+\hat{\underline{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\underline{\beta}}+\mu_0^T\mathbf{P}_0\mu_0-\mu_1^T\mathbf{P}_1\mu_1\right)\right)}$$
$$\tau^{\frac{p}{2}}e^{-\frac{\tau}{2}\left(\underline{\beta}-\mu_1\right)^T\mathbf{P}_1\left(\underline{\beta}-\mu_1\right)}$$

Se puede probar que:

$$\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\hat{\underline{\beta}}\right)+\hat{\underline{\beta}}^T\mathbf{X}^T\mathbf{X}\hat{\underline{\beta}}+\mu_0^T\mathbf{P}_0\mu_0-\mu_1^T\mathbf{P}_1\mu_1 =$$
$$\left(\underline{y}-\mathbf{X}\mu_1\right)^T\left(\underline{y}-\mathbf{X}\mu_1\right)+\left(\mu_1-\mu_0\right)^T\mathbf{P}_0\left(\mu_1-\mu_0\right)$$

Por lo tanto:

$$p\left(\underline{\beta},\tau|\underline{y}\right) \propto \tau^{\frac{n}{2}+\alpha_0-1}e^{-\tau\left(\delta_0+\frac{1}{2}\left(\left(\underline{y}-\mathbf{X}\mu_1\right)^T\left(\underline{y}-\mathbf{X}\mu_1\right)+\left(\mu_1-\mu_0\right)^T\mathbf{P}_0\left(\mu_1-\mu_0\right)\right)\right)}$$
$$\tau^{\frac{p}{2}}e^{-\frac{\tau}{2}\left(\underline{\beta}-\mu_1\right)^T\mathbf{P}_1\left(\underline{\beta}-\mu_1\right)}$$

Regresión Lineal : Enfoque Bayesiano

Finalmente, haciendo:

- $\mathbf{P}_1 = \mathbf{X}^T\mathbf{X} + \mathbf{P}_0$
- $\mu_1 = \left(\mathbf{X}^T\mathbf{X} + \mathbf{P}_0\right)^{-1}\left(\mathbf{X}^T\underline{y} + \mathbf{P}_0\mu_0\right)$
- $\alpha_1 = \frac{n}{2} + \alpha_0$
- $\delta_1 = \delta_0 + \frac{1}{2}\left(\left(\underline{y} - \mathbf{X}\mu_1\right)^T\left(\underline{y} - \mathbf{X}\mu_1\right) + \left(\mu_1 - \mu_0\right)^T\mathbf{P}_0\left(\mu_1 - \mu_0\right)\right)$

Se obtiene que:

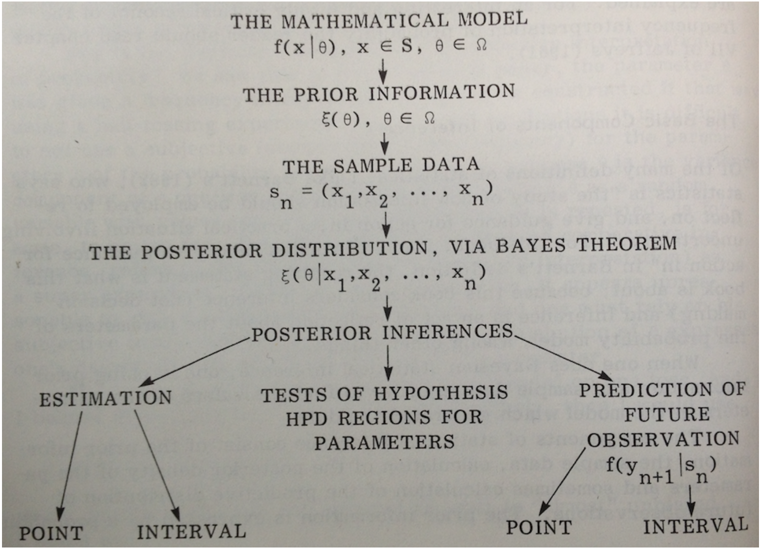
$$p\left(\underline{\beta}, \tau|\underline{y}\right) = N_p\left(\underline{\beta}|\mu_1, \tau\mathbf{P}_1\right)Ga\left(\tau|\alpha_1, \delta_1\right)$$

Entonces sabemos que por propiedades de esta distribución que:

$$p\left(\underline{\beta}|\underline{y}\right) = T_p\left(\underline{\beta}\middle|2\alpha_1, \mu_1, \frac{\alpha_1}{\delta_1}\mathbf{P}_1\right)$$

$$p\left(\tau|\underline{y}\right) = Ga\left(\alpha_1, \delta_1\right) \Rightarrow p\left(\tau^{-1}|\underline{y}\right) = IGa\left(\alpha_1, \delta_1\right)$$

Regresión Lineal :Inferencia



Regresión Lineal :Inferencia sobre τ

Suponiendo que solo estamos interesados en τ o $\sigma^2 = \tau^{-1}$ se procede a marginalizar

$$p\left(\tau|\underline{y}\right) = Ga\left(\tau\,|\alpha_1, \delta_1\right) \Rightarrow p\left(\sigma^2|\underline{y}\right) = IGa\left(\sigma^2\,|\alpha_1, \delta_1\right)$$

- Estimación puntual:

$$\begin{aligned}
\hat{\tau}_{Media} &= \frac{\alpha_1}{\delta_1} = \frac{\frac{n}{2} + \alpha_0}{\delta_0 + \frac{1}{2}\left(\left(\underline{y} - \mathbf{X}\mu_1\right)^T\left(\underline{y} - \mathbf{X}\mu_1\right) + (\mu_1 - \mu_0)^T\,\mathbf{P}_0\,(\mu_1 - \mu_0)\right)} \\
\hat{\tau}_{Moda} &= \frac{\alpha_1 - 1}{\delta_1} = \frac{\frac{n}{2} + \alpha_0 - 1}{\delta_0 + \frac{1}{2}\left(\left(\underline{y} - \mathbf{X}\mu_1\right)^T\left(\underline{y} - \mathbf{X}\mu_1\right) + (\mu_1 - \mu_0)^T\,\mathbf{P}_0\,(\mu_1 - \mu_0)\right)} \\
\hat{\sigma}^2_{Media} &= \frac{\delta_1}{\alpha_1 - 1} = \frac{\delta_0 + \frac{1}{2}\left(\left(\underline{y} - \mathbf{X}\mu_1\right)^T\left(\underline{y} - \mathbf{X}\mu_1\right) + (\mu_1 - \mu_0)^T\,\mathbf{P}_0\,(\mu_1 - \mu_0)\right)}{\frac{n}{2} + \alpha_0 - 1} \\
\hat{\sigma}^2_{Moda} &= \frac{\delta_1}{\alpha_1 + 1} = \frac{\delta_0 + \frac{1}{2}\left(\left(\underline{y} - \mathbf{X}\mu_1\right)^T\left(\underline{y} - \mathbf{X}\mu_1\right) + (\mu_1 - \mu_0)^T\,\mathbf{P}_0\,(\mu_1 - \mu_0)\right)}{\frac{n}{2} + \alpha_0 + 1}
\end{aligned}$$

- Estimación Intervalar: (No tiene forma cerrada, métodos numéricos)

Regresión Lineal :Inferencia sobre $\underline{\beta}$

Suponiendo que solo estamos interesados en $\underline{\beta}$ sabemos que:

$$p\left(\underline{\beta}|\underline{y}\right) = T_p\left(\underline{\beta}\middle|2\alpha_1, \mu_1, \frac{\alpha_1}{\delta_1}\mathbf{P}_1\right)$$

- Estimación Puntual:

$$\hat{\underline{\beta}} \quad = \quad \mu_1 = \left(\mathbf{X}^T\mathbf{X} + \mathbf{P}_0\right)^{-1}\left(\mathbf{X}^T\underline{y} + \mathbf{P}_0\mu_0\right)$$

- Estimación Intervalar. La densidad T-multivariada pertenece a las distribuciones elípticas y por tanto sabemos que:

$$R^2 = \left(\underline{\beta} - \mu_1\right)^T \frac{\alpha_1}{\delta_1}\mathbf{P}_1\left(\underline{\beta} - \mu_1\right) \sim pF_{(p,2\alpha_1)}$$

Por lo tanto

$$\mathbb{P}\left(\frac{1}{p}\left(\underline{\beta} - \mu_1\right)^T \frac{\alpha_1}{\delta_1}\mathbf{P}_1\left(\underline{\beta} - \mu_1\right) \leq F^{1-\alpha}_{(p,2\alpha_1)}\right) = 1 - \alpha$$

Regresión Lineal :Inferencia sobre β_i

Suponiendo que solo estamos interesados en una entrada del vector β_i sabemos que:

$$p\left(\beta_i|\underline{y}\right) = T_1\left(\beta_i \middle| 2\alpha_1, \mu_{1(i+1)}, \frac{1}{c_{(i+1)(i+1)}}\right) \quad i \in \{0,\dots,k\}$$

Donde $c_{jj} = \left(\left(\frac{\alpha_1}{\delta_1}\mathbf{P}_1\right)^{-1}\right)_{jj}$

- Estimación Puntual:

$$\hat{\beta}_i = \mu_{1(i+1)} = \left(\left(\mathbf{X}^T\mathbf{X} + \mathbf{P}_0\right)^{-1}\left(\mathbf{X}^T\underline{y} + \mathbf{P}_0\mu_0\right)\right)_{(i+1)}$$

- Estimación Intervalar.

$$R^2 = \left(\beta_i - \mu_{1(i+1)}\right)\left(\frac{1}{c_{(i+1)(i+1)}}\right)\left(\beta_i - \mu_{1(i+1)}\right) \sim F_{(1,2\alpha_1)}$$

Por lo tanto la región de tamaño $1-\alpha$ está dada por:

$$\left(\beta_i - \mu_{1(i+1)}\right)\left(\frac{1}{c_{(i+1)(i+1)}}\right)\left(\beta_i - \mu_{1(i+1)}\right) \leq F_{(1,2\alpha_1)}^{1-\alpha}$$

$$\left|\beta_i - \mu_{1(i+1)}\right|\left(\frac{1}{\sqrt{c_{(i+1)(i+1)}}}\right) \leq t_{2\alpha_1}^{1-\frac{\alpha}{2}}$$

Regresión Lineal: Inferencia sobre β

Supongamos que estamos interesados en probar:

$$H_0 : \mathbf{A}\underline{\beta} = \mathbf{b} \quad vs \quad H_1 : \mathbf{A}\underline{\beta} \neq \mathbf{b};$$

Con $\mathbf{A} \in \mathbb{R}^{m \times p}$ de rango completo y $\mathbf{b} \in \mathbb{R}^m$
Una forma de probar está hipotesis (Lindley) es construir la región de credibilidad para $\mathbf{A}\underline{\beta}$, es decir encontrar la densidad final el vector aleatorio $\mathbf{A}\underline{\beta}$, luego construir D^* la región de credibilidad de tamaño $1 - \alpha$. Y luego verificar si $\mathbf{b}$ se encuentra en D^*, para decidir si rechazar o no H_0.

Regresión Lineal: Inferencia sobre β

Como:

$$\underline{\beta}|\underline{y} \sim T_p\left(\underline{\beta}\,\middle|\, 2\alpha_1, \mu_1, \frac{\alpha_1}{\delta_1}\mathbf{P}_1\right)$$

Se sigue que (Propiedades de las distribuciones Elipticas)

$$\mathbf{A}\underline{\beta}|\underline{y} \sim T_m\left(\underline{\beta}\,\middle|\, 2\alpha_1, \mathbf{A}\mu_1, \left(\mathbf{A}\left(\frac{\alpha_1}{\delta_1}\mathbf{P}_1\right)^{-1}\mathbf{A}^T\right)^{-1}\right)$$

Entonces:

$$R^2 = (\mathbf{A}\underline{\beta} - \mathbf{A}\mu_1)^T\left(\mathbf{A}\left(\frac{\alpha_1}{\delta_1}\mathbf{P}_1\right)^{-1}\mathbf{A}^T\right)^{-1}(\mathbf{A}\underline{\beta} - \mathbf{A}\mu_1) \sim mF_{(m,2\alpha_1)}$$

La región es:

$$\frac{1}{m}(\mathbf{A}\underline{\beta} - \mathbf{A}\mu_1)^T\left(\mathbf{A}\left(\frac{\alpha_1}{\delta_1}\mathbf{P}_1\right)^{-1}\mathbf{A}^T\right)^{-1}(\mathbf{A}\underline{\beta} - \mathbf{A}\mu_1) \leq F^{1-\alpha}_{(m,2\alpha_1)}$$

Rechazo H_0 si:

$$\frac{1}{m}(\mathbf{b} - \mathbf{A}\mu_1)^T\left(\mathbf{A}\left(\frac{\alpha_1}{\delta_1}\mathbf{P}_1\right)^{-1}\mathbf{A}^T\right)^{-1}(\mathbf{b} - \mathbf{A}\mu_1) > F^{1-\alpha}_{(m,2\alpha_1)}$$

Regresión Lineal: Inferencia sobre β

Una hipótesis muy importante que se pueba en regresión es la siguiente:

$$H_0 : \beta_1 = \ldots = \beta_k = 0 \qquad vs \qquad H_1 : \exists i \text{ tal que } \beta_i \neq 0 \quad i \in \{1, \ldots, k\}$$

Esta prueba la podemos llevar acabo definiendo

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots 0 \\ 0 & 0 & 1 & 0 & \ldots 0 \\ 0 & 0 & 0 & 1 & \ldots 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & 0 & \ldots 1 \end{pmatrix} \in \mathbb{R}^{(p-1)\times p}$$

Lo que entonces equivale a:

$$H_0 : \mathbf{A}\underline{\beta} = \underline{0} \in \mathbb{R}^{p-1} \quad vs \quad H_1 : \mathbf{A}\underline{\beta} \neq \underline{0} \in \mathbb{R}^{p-1}$$

La hipótesis se rechaza si:

$$\frac{1}{p-1} (\mathbf{A}\mu_1)^T \left(\mathbf{A} \left(\frac{\alpha_1}{\delta_1} \mathbf{P}_1 \right)^{-1} \mathbf{A}^T \right)^{-1} (\mathbf{A}\mu_1) > F^{1-\alpha}_{(p-1, 2\alpha_1)}$$

Regresión Lineal: Inferencia sobre y^*

Imaginemos ahora que deseamos hacer inferencia sobre nuevas observaciones del modelo:

$$\mathbf{w} = \mathbf{Z}\underline{\beta} + \mathbf{e}$$

Donde

- $\mathbf{Z} \in \mathbb{R}^{k\times p}$ es una nueva matriz de covariables.
- $\mathbf{e} \sim N_k\left(\mathbf{e}|\underline{0}, \tau\mathbf{I}\right)$
- $\mathbf{w} \sim N_k\left(\mathbf{w}|\mathbf{Z}\underline{\beta}, \tau\mathbf{I}\right)$

Desde el punto de vista Bayesiano, el objetivo es determinar $p\left(\mathbf{w}|\underline{y}\right)$

$$\begin{aligned} p\left(\mathbf{w}|\underline{y}\right) &= \int_0^\infty \int_{\mathbb{R}^p} p\left(\mathbf{w}, \underline{\beta}, \tau|\underline{y}\right) d\underline{\beta} d\tau = \int_0^\infty \int_{\mathbb{R}^p} \frac{p\left(\mathbf{w}, \underline{\beta}, \tau, \underline{y}\right)}{p\left(\underline{y}\right)} d\underline{\beta} d\tau \\ &= \int_0^\infty \int_{\mathbb{R}^p} \frac{p\left(\mathbf{w}|\underline{\beta}, \tau, \underline{y}\right) p\left(\underline{\beta}, \tau, \underline{y}\right)}{p\left(\underline{y}\right)} d\underline{\beta} d\tau \\ &= \int_0^\infty \int_{\mathbb{R}^p} p\left(\mathbf{w}|\underline{\beta}, \tau, \underline{y}\right) p\left(\underline{\beta}, \tau|\underline{y}\right) d\underline{\beta} d\tau \end{aligned}$$

Regresión Lineal: Inferencia sobre y^*

Sin embargo por independencia condicional:

$$\begin{aligned} p\left(\mathbf{w}|\underline{\beta},\tau,\underline{y}\right) &= \frac{p\left(\mathbf{w},\underline{\beta},\tau,\underline{y}\right)}{p\left(\underline{\beta},\tau,\underline{y}\right)} = \frac{p\left(\mathbf{w},\underline{y}|\underline{\beta},\tau\right)p\left(\underline{\beta},\tau\right)}{p\left(\underline{\beta},\tau,\underline{y}\right)} \\ &= \frac{p\left(\mathbf{w}|\underline{\beta},\tau\right)p\left(\underline{y}|\underline{\beta},\tau\right)p\left(\underline{\beta},\tau\right)}{p\left(\underline{\beta},\tau,\underline{y}\right)} = \frac{p\left(\mathbf{w}|\underline{\beta},\tau\right)p\left(\underline{y}|\underline{\beta},\tau\right)p\left(\underline{\beta},\tau\right)}{p\left(\underline{y}|\underline{\beta},\tau\right)p\left(\underline{\beta},\tau\right)} \\ &= p\left(\mathbf{w}|\underline{\beta},\tau\right) \end{aligned}$$

Por lo tanto:

$$\begin{aligned} p\left(\mathbf{w}|\underline{y}\right) &= \int_0^\infty \int_{\mathbb{R}^p} p\left(\mathbf{w}|\underline{\beta},\tau,\underline{y}\right)p\left(\underline{\beta},\tau|\underline{y}\right)d\underline{\beta}d\tau \\ &= \int_0^\infty \int_{\mathbb{R}^p} p\left(\mathbf{w}|\underline{\beta},\tau\right)p\left(\underline{\beta},\tau|\underline{y}\right)d\underline{\beta}d\tau \end{aligned}$$

Donde :

$$p\left(\mathbf{w}|\underline{\beta},\tau\right) = N_k\left(\mathbf{w}|\mathbf{Z}\underline{\beta},\tau\mathbf{I}\right); \qquad p\left(\underline{\beta},\tau|\underline{y}\right) = NG\left(\underline{\beta},\tau|\mu_1,\mathbf{P}_1,\alpha_1,\delta_1\right)$$

Regresión Lineal: Inferencia sobre y^*

$$p\left(\mathbf{w}|\underline{y}\right) \propto \int_0^\infty \int_{\mathbb{R}^p} \tau^{\frac{k}{2}} e^{-\frac{\tau}{2}\left(\mathbf{w}-\mathbf{Z}\underline{\beta}\right)^T\left(\mathbf{w}-\mathbf{Z}\underline{\beta}\right)} \tau^{\frac{p}{2}} e^{-\frac{\tau}{2}\left(\underline{\beta}-\mu_1\right)^T \mathbf{P}_1 \left(\underline{\beta}-\mu_1\right)} \tau^{\alpha_1-1} e^{-\tau\delta_1} d\underline{\beta} d\tau$$

En el integrando se puede completar un kernell de una Normal-Multivariada en $\underline{\beta}$, cual integra 1 por lo que finalmente nos quedamos con lo siguiente:

$$\begin{aligned}
p\left(\mathbf{w}|\underline{y}\right) &\propto \int_0^\infty \underbrace{\tau^{\frac{k}{2}} e^{-\frac{\tau}{2}(\mathbf{w}-\mathbf{Z}\mu_1)^T\left(\mathbf{I}-\mathbf{Z}\mathbf{P}_1^{-1}\mathbf{Z}\right)^{-1}(\mathbf{w}-\mathbf{Z}\mu_1)}}_{N_k\left(\mathbf{w}\middle|\mathbf{Z}\mu_1,\tau\left(\mathbf{I}+\mathbf{Z}\mathbf{P}_1^{-1}\mathbf{Z}^T\right)^{-1}\right)} \underbrace{\tau^{\alpha-1} e^{-\tau\delta_1}}_{Ga(\tau|\alpha_1,\delta_1)} d\tau \\
&\propto \int_0^\infty NG\left(\mathbf{w},\tau|\mathbf{Z}\mu_1,\left(\mathbf{I}+\mathbf{Z}\mathbf{P}_1^{-1}\mathbf{Z}^T\right)^{-1},\alpha_1,\delta_1\right) d\tau \\
&\propto T_k\left(\mathbf{w}\middle|2\alpha_1,\mathbf{Z}\mu_1,\frac{\alpha_1}{\delta_1}\left(\mathbf{I}+\mathbf{Z}\mathbf{P}_1\mathbf{Z}^T\right)^{-1}\right)
\end{aligned}$$

Regresión Lineal: Distribuciones no Informativas

Supongamos que deseamos aplicar la teoría Bayesiana pero no contamos con información inicial, la solución que se plantea es entonces utilizar distribuciones iniciales no informativas, veamos como se ve la inicial de **Jeffreys**:
Obtengamos la matriz de informacón de Fisher $I\left(\underline{\beta},\tau\right)$

$$p\left(\underline{y}|\underline{\beta},\tau\right)\propto\tau^{\frac{n}{2}}e^{-\frac{\tau}{2}\left(\underline{y}-\mathbf{X}\underline{\beta}\right)^{T}\left(\underline{y}-\mathbf{X}\underline{\beta}\right)}$$

Entonces:

$$\log\left(p\left(\underline{y}|\underline{\beta},\tau\right)\right)\propto\frac{n}{2}\log\left(\tau\right)-\frac{\tau}{2}\left(\underline{y}-\mathbf{X}\underline{\beta}\right)^{T}\left(\underline{y}-\mathbf{X}\underline{\beta}\right)$$

Obtenemos la primera derivada (vector gradiente)

$$\left(\begin{array}{c}\frac{\partial\log\left(p\left(\underline{y}|\underline{\beta},\tau\right)\right)}{\partial\underline{\beta}}\\ \frac{\partial\log\left(p\left(\underline{y}|\underline{\beta},\tau\right)\right)}{\partial\tau}\end{array}\right)_{(p+1)\times1}=\left(\begin{array}{c}\tau\mathbf{X}^{T}\left(\underline{y}-\mathbf{X}\underline{\beta}\right)\\ \frac{n}{2\tau}-\frac{1}{2}\left(\underline{y}-\mathbf{X}\underline{\beta}\right)^{T}\left(\underline{y}-\mathbf{X}\underline{\beta}\right)\end{array}\right)_{(p+1)\times1}$$

Regresión Lineal: Distribuciones no Informativas

Obtenemos la segunda derivada (matriz)

$$\begin{pmatrix} \frac{\partial^2 \log\left(p\left(\underline{y}|\underline{\beta},\tau\right)\right)}{\partial \underline{\beta}^2} & \frac{\partial^2 \log\left(p\left(\underline{y}|\underline{\beta},\tau\right)\right)}{\partial \tau \partial \underline{\beta}} \\ \frac{\partial^2 \log\left(p\left(\underline{y}|\underline{\beta},\tau\right)\right)}{\partial \underline{\beta} \partial \tau} & \frac{\partial^2 \log\left(p\left(\underline{y}|\underline{\beta},\tau\right)\right)}{\partial \tau^2} \end{pmatrix} = \begin{pmatrix} -\tau \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\left(\underline{y} - \mathbf{X}\underline{\beta}\right) \\ \left(\underline{y} - \mathbf{X}\underline{\beta}\right)^T \mathbf{X} & -\frac{n}{2\tau^2} \end{pmatrix}$$

Por lo tanto:

$$\mathbf{I}\left(\underline{\beta},\tau\right) = -\mathbb{E}\left(\begin{pmatrix} -\tau \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\left(\underline{y} - \mathbf{X}\underline{\beta}\right) \\ \left(\underline{y} - \mathbf{X}\underline{\beta}\right)^T \mathbf{X} & -\frac{n}{2\tau^2} \end{pmatrix}\right) = \begin{pmatrix} \tau \mathbf{X}^T\mathbf{X} & 0 \\ 0 & \frac{n}{2\tau^2} \end{pmatrix}$$

Por lo tanto la inicial es:

$$\begin{aligned} p\left(\underline{\beta},\tau\right) &\propto \sqrt{det\left(\mathbf{I}\left(\underline{\beta},\tau\right)\right)} \propto \sqrt{det\left(\tau\mathbf{X}^T\mathbf{X}\right) det\left(\frac{n}{2\tau^2}\right)} \propto \sqrt{\tau^{p-2} det\left(\mathbf{X}^T\mathbf{X}\right)\frac{n}{2}} \\ &\propto \tau^{\frac{p-2}{2}}\sqrt{det\left(\mathbf{X}^T\mathbf{X}\right)\frac{n}{2}} \propto \tau^{\frac{p-2}{2}} \end{aligned}$$

Obtenemos una inicial impropia !!!!!

Regresión Lineal: Distribuciones no Informativas

Ojo: Cuando trabajamos con distribuciones iniciales impropias es muy importante verificar que la final si sea propia.
Aplicando mecánicamente el teorema de Bayes con la inicial de **Jeffreys** obtenemos:

$$
\begin{aligned}
p\left(\underline{\beta},\tau|\underline{y}\right) &\propto p\left(\underline{y}|\underline{\beta},\tau\right)p\left(\underline{\beta},\tau\right)\\
&\propto \tau^{\frac{n}{2}}e^{-\frac{\tau}{2}\left(\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)+\left(\underline{\beta}-\underline{\hat{\beta}}\right)^T\mathbf{X}^T\mathbf{X}\left(\underline{\beta}-\underline{\hat{\beta}}\right)\right)}\tau^{\frac{p-2}{2}}\\
&\propto \underbrace{\tau^{\frac{p}{2}}e^{-\frac{\tau}{2}\left(\underline{\beta}-\underline{\hat{\beta}}\right)^T\mathbf{X}^T\mathbf{X}\left(\underline{\beta}-\underline{\hat{\beta}}\right)}}_{N_p\left(\underline{\beta}|\underline{\hat{\beta}},\tau\mathbf{X}^T\mathbf{X}\right)}\underbrace{\tau^{\frac{n}{2}-1}e^{-\frac{\tau}{2}\left(\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)\right)}}_{Ga\left(\tau|\frac{n}{2},\frac{1}{2}\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)\right)}
\end{aligned}
$$

Con la inicial de **Jeffreys** obtenemos que la final es propia!!!!

$$
p\left(\underline{\beta},\tau|\underline{y}\right)=NG\left(\underline{\beta},\tau\,\middle|\,\underline{\hat{\beta}},\mathbf{X}^T\mathbf{X},\frac{n}{2},\frac{1}{2}\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)\right)
$$

Regresión Lineal: Distribuciones no Informativas

Recordemos que bajo una inicial informativa los parámetro finales son:

- $\mathbf{P}_1 = \mathbf{X}^T\mathbf{X} + \mathbf{P}_0$
- $\mu_1 = \left(\mathbf{X}^T\mathbf{X} + \mathbf{P}_0\right)^{-1}\left(\mathbf{X}^T\underline{y} + \mathbf{P}_0\mu_0\right)$
- $\alpha_1 = \frac{n}{2} + \alpha_0$
- $\delta_1 = \delta_0 + \frac{1}{2}\left(\left(\underline{y} - \mathbf{X}\mu_1\right)^T\left(\underline{y} - \mathbf{X}\mu_1\right) + \left(\mu_1 - \mu_0\right)^T\mathbf{P}_0\left(\mu_1 - \mu_0\right)\right)$

En el caso de la inicial de Jeffreys, los parámetros de la final son:

- $\mu_1 = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\underline{y}\right) = \underline{\hat{\beta}}_{MV}$
- $\mathbf{P}_1 = \mathbf{X}^T\mathbf{X}$
- $\alpha_1 = \frac{n}{2}$
- $\delta_1 = \frac{1}{2}\left(\underline{y} - \mathbf{X}\mu_1\right)^T\left(\underline{y} - \mathbf{X}\mu_1\right) = \frac{n-p}{2}\hat{\sigma}^2$

La distribución de Jeffreys se obtiene haciendo en la inicial $\mathbf{P}_0 \to \mathbf{0}$, $\alpha_0 \to 0$, $\delta_0 \to 0$

Inferencia sobre τ: Distribución Inicial de Jeffreys

Suponiendo que solo estamos interesados en τ o $\sigma^2 = \tau^{-1}$ se procede a marginalizar

$$p\left(\tau|\underline{y}\right) = Ga\left(\tau \left| \frac{n}{2}, \frac{n-p}{2}\hat{\sigma}^2\right.\right) \Rightarrow p\left(\sigma^2|\underline{y}\right) = IGa\left(\sigma^2 \left| \frac{n}{2}, \frac{n-p}{2}\hat{\sigma}^2\right.\right)$$

- Estimación puntual:

$$\begin{aligned}
\hat{\tau}_{Media} &= \frac{\alpha_1}{\delta_1} = \frac{n}{(n-p)\hat{\sigma}^2} = \hat{\sigma}^{-2}_{M.V.} \\
\hat{\tau}_{Moda} &= \frac{\alpha_1 - 1}{\delta_1} = \frac{n-1}{(n-p)\hat{\sigma}^2} \\
\hat{\sigma}^2_{Media} &= \frac{\delta_1}{\alpha_1 - 1} = \frac{n-p}{n-2}\hat{\sigma}^2 \\
\hat{\sigma}^2_{Moda} &= \frac{\delta_1}{\alpha_1 + 1} = \frac{n-p}{n+2}\hat{\sigma}^2
\end{aligned}$$

- Estimación Intervalar: (No tiene forma cerrada, métodos numéricos)

Inferencia sobre $\underline{\beta}$: Distribución Inicial de Jeffreys

Suponiendo que solo estamos interesados en $\underline{\beta}$ sabemos que:

$$p\left(\underline{\beta}|\underline{y}\right) = T_p\left(\underline{\beta}\middle|n, \underline{\hat{\beta}}_{M.V.}, \frac{n}{n-p}\hat{\sigma}^{-2}\mathbf{X}^T\mathbf{X}\right)$$

- Estimación Puntual:

$$\underline{\hat{\beta}} \quad = \quad \mu_1 = \underline{\hat{\beta}}_{M.V.}$$

- Estimación Intervalar. La densidad T-multivariada pertenece a las distribuciones elípticas y por tanto sabemos que:

$$R^2 = \left(\underline{\beta} - \underline{\hat{\beta}}_{M.V.}\right)^T \frac{n}{n-p}\hat{\sigma}^{-2}\mathbf{X}^T\mathbf{X}\left(\underline{\beta} - \underline{\hat{\beta}}_{M.V.}\right) \sim pF_{(p,n)}$$

  Por lo tanto

$$\mathbb{P}\left(\frac{1}{p}\left(\underline{\beta} - \underline{\hat{\beta}}_{M.V.}\right)^T \frac{n}{n-p}\hat{\sigma}^{-2}\mathbf{X}^T\mathbf{X}\left(\underline{\beta} - \underline{\hat{\beta}}_{M.V.}\right) \leq F^{1-\alpha}_{(p,n)}\right) = 1-\alpha$$

Distribución Inicial de Referencia

Se puede probar que otra forma la inicial de referencia está dada por:

$$p\left(\underline{\beta},\tau\right) \propto \tau^{-1}$$

Aplicando mecánicamente la regla de Bayes obtenemos:

$$\begin{aligned}
p\left(\underline{\beta},\tau|\underline{y}\right) &\propto p\left(\underline{y}|\underline{\beta},\tau\right)p\left(\underline{\beta},\tau\right)\\
&\propto \tau^{\frac{n}{2}}e^{-\frac{\tau}{2}\left(\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)+\left(\underline{\beta}-\underline{\hat{\beta}}\right)^T\mathbf{X}^T\mathbf{X}\left(\underline{\beta}-\underline{\hat{\beta}}\right)\right)}\tau^{-1}\\
&\propto \underbrace{\tau^{\frac{p}{2}}e^{-\frac{\tau}{2}\left(\underline{\beta}-\underline{\hat{\beta}}\right)^T\mathbf{X}^T\mathbf{X}\left(\underline{\beta}-\underline{\hat{\beta}}\right)}}_{N_p\left(\underline{\beta}|\underline{\hat{\beta}},\tau\mathbf{X}^T\mathbf{X}\right)}\underbrace{\tau^{\frac{n}{2}-\frac{p}{2}-1}e^{-\frac{\tau}{2}\left(\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)\right)}}_{Ga\left(\tau|\frac{n-p}{2},\frac{1}{2}\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)\right)}
\end{aligned}$$

Con la inicial de **Referencia** obtenemos que la siguiente final!!!!

$$p\left(\underline{\beta},\tau|\underline{y}\right) = NG\left(\underline{\beta},\tau\left|\underline{\hat{\beta}},\mathbf{X}^T\mathbf{X},\frac{n-p}{2},\frac{1}{2}\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)^T\left(\underline{y}-\mathbf{X}\underline{\hat{\beta}}\right)\right.\right)$$

Regresión Lineal: Distribuciones no Informativas

Recordemos que bajo una inicial informativa los parámetro finales son:

- $\mathbf{P}_1 = \mathbf{X}^T\mathbf{X} + \mathbf{P}_0$
- $\mu_1 = \left(\mathbf{X}^T\mathbf{X} + \mathbf{P}_0\right)^{-1}\left(\mathbf{X}^T\underline{y} + \mathbf{P}_0\mu_0\right)$
- $\alpha_1 = \frac{n}{2} + \alpha_0$
- $\delta_1 = \delta_0 + \frac{1}{2}\left(\left(\underline{y} - \mathbf{X}\mu_1\right)^T\left(\underline{y} - \mathbf{X}\mu_1\right) + \left(\mu_1 - \mu_0\right)^T\mathbf{P}_0\left(\mu_1 - \mu_0\right)\right)$

En el caso de la inicial de Jeffreys, los parámetros de la final son:

- $\mu_1 = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{X}^T\underline{y}\right) = \underline{\hat{\beta}}_{MV}$
- $\mathbf{P}_1 = \mathbf{X}^T\mathbf{X}$
- $\alpha_1 = \frac{n-p}{2}$
- $\delta_1 = \frac{1}{2}\left(\underline{y} - \mathbf{X}\mu_1\right)^T\left(\underline{y} - \mathbf{X}\mu_1\right) = \frac{n-p}{2}\hat{\sigma}^2$

La distribución de Jeffreys se obtiene haciendo en la inicial $\mathbf{P}_0 \to \mathbf{0}$, $\alpha_0 \to -\frac{p}{2}$, $\delta_0 \to 0$

Inferencia sobre τ: Distribución Inicial de Referencia

Suponiendo que solo estamos interesados en τ o $\sigma^2 = \tau^{-1}$ se procede a marginalizar

$$p\left(\tau|\underline{y}\right) = Ga\left(\tau \left| \frac{n-p}{2}, \frac{n-p}{2}\hat{\sigma}^2\right.\right) \Rightarrow p\left(\sigma^2|\underline{y}\right) = IGa\left(\sigma^2 \left| \frac{n-p}{2}, \frac{n-p}{2}\hat{\sigma}^2\right.\right)$$

- Estimación puntual:

$$\begin{aligned}
\hat{\tau}_{Media} &= \frac{\alpha_1}{\delta_1} = \frac{1}{\hat{\sigma}^2} = \hat{\sigma}^{-2} \\
\hat{\tau}_{Moda} &= \frac{\alpha_1 - 1}{\delta_1} = \frac{n-p-1}{(n-p)\hat{\sigma}^2} \\
\hat{\sigma}^2_{Media} &= \frac{\delta_1}{\alpha_1 - 1} = \frac{n-p}{n-p-2}\hat{\sigma}^2 \\
\hat{\sigma}^2_{Moda} &= \frac{\delta_1}{\alpha_1 + 1} = \frac{n-p}{n-p+2}\hat{\sigma}^2
\end{aligned}$$

- Estimación Intervalar: (No tiene forma cerrada, métodos numéricos)

Inferencia sobre $\underline{\beta}$: Distribución Inicial de Referencia

Suponiendo que solo estamos interesados en $\underline{\beta}$ sabemos que:

$$p\left(\underline{\beta}|\underline{y}\right) = T_p\left(\underline{\beta}\,\middle|\, n-p, \hat{\underline{\beta}}_{M.V.}, \hat{\sigma}^{-2}\mathbf{X}^T\mathbf{X}\right)$$

- Estimación Puntual:

$$\hat{\underline{\beta}} \quad = \quad \mu_1 = \hat{\underline{\beta}}_{M.V.}$$

- Estimación Intervalar. La densidad T-multivariada pertenece a las distribuciones elípticas y por tanto sabemos que:

$$R^2 = \left(\underline{\beta} - \hat{\underline{\beta}}_{M.V.}\right)^T \hat{\sigma}^{-2}\mathbf{X}^T\mathbf{X}\left(\underline{\beta} - \hat{\underline{\beta}}_{M.V.}\right) \sim pF_{(p,n-p)}$$

(Obs: Compare con la ecuación (9)) Por lo tanto

$$\mathbb{P}\left(\frac{1}{p}\left(\underline{\beta} - \hat{\underline{\beta}}_{M.V.}\right)^T \hat{\sigma}^{-2}\mathbf{X}^T\mathbf{X}\left(\underline{\beta} - \hat{\underline{\beta}}_{M.V.}\right) \leq F^{1-\alpha}_{(p,n-p)}\right) = 1-\alpha$$

Ejemplo:

$$\underline{y} = \mathbf{X}\beta + \varepsilon; \quad \mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n\times 1} \quad \varepsilon \sim N_n\left(\mathbf{0}, \tau \mathbf{I}\right); \beta \in \mathbb{R}$$

Bajo estas condiciones estaríamos en el caso en que $\underline{y} = \left(y_1, \ldots y_n\right)^T$ es muestra aleatoria de $N\left(\beta, \tau\right)$. Utilizando la teoría Bayesiana de modelos lineales, nos gustaría hacer inferencias sobre los parámetros $\left(\beta, \tau\right)$. Suponiendo inicial Informativa tendramos:

$$p\left(\beta, \tau\right) = NG\left(\beta, \tau | \mu_0, p_0, \alpha_0, \beta_0\right)$$

Los parámetros de la final son entonces:

- $\mathbf{P}_1 = \mathbf{X}^T\mathbf{X} + \mathbf{P}_0 = n + p_0$
- $\mu_1 = \left(\mathbf{X}^T\mathbf{X} + \mathbf{P}_0\right)^{-1}\left(\mathbf{X}^T\underline{y} + \mathbf{P}_0\mu_0\right) = \left(n + p_0\right)^{-1}\left(\sum_{i=1}^n y_i + p_0\mu_0\right)$
- $\alpha_1 = \frac{n}{2} + \alpha_0$
- $\delta_1 = \delta_0 + \frac{1}{2}\left(\left(\underline{y} - \mathbf{X}\mu_1\right)^T\left(\underline{y} - \mathbf{X}\mu_1\right) + \left(\mu_1 - \mu_0\right)^T\mathbf{P}_0\left(\mu_1 - \mu_0\right)\right)$
  $= \delta_0 + \frac{1}{2}\left(\sum_{i=1}^n\left(y_i - \frac{\sum_{i=1}^n y_i + p_0\mu_0}{n + p_0}\right)^2 + p_0\left(\frac{\sum_{i=1}^n y_i + p_0\mu_0}{n + p_0} - \mu_0\right)^2\right)$

Utilizando la incial de Jeffreys $p_0 \to 0$, $\alpha_0 \to 0$, $\delta_0 \to 0$

- $\mathbf{P}_1 = n + p_0 = n$
- $\mu_1 = (n + p_0)^{-1} \left(\sum_{i=1}^n y_i + p_0 \mu_0 \right) = \overline{y} = \hat{\mu}_{M.V.}$
- $\alpha_1 = \frac{n}{2} + \alpha_0 = \frac{n}{2}$
- $\delta_1 = \delta_0 + \frac{1}{2} \left(\sum_{i=1}^n \left(y_i - \frac{\sum_{i=1}^n y_i + p_0 \mu_0}{n + p_0} \right)^2 + p_0 \left(\frac{\sum_{i=1}^n y_i + p_0 \mu_0}{n + p_0} - \mu_0 \right)^2 \right)$
  $= \frac{n-1}{2} \frac{\sum_{i=1}^n (y_i - \overline{y})^2}{n-1} = \frac{n-1}{2} \hat{\sigma}^2$

Utilizando la incial de Referencia $p_0 \to 0$, $\alpha_0 \to -\frac{1}{2}$, $\delta_0 \to 0$

- $\mathbf{P}_1 = n + p_0 = n$
- $\mu_1 = (n + p_0)^{-1} \left(\sum_{i=1}^n y_i + p_0 \mu_0 \right) = \overline{y} = \hat{\mu}_{M.V.}$
- $\alpha_1 = \frac{n}{2} + \alpha_0 = \frac{n-1}{2}$
- $\delta_1 = \delta_0 + \frac{1}{2} \left(\sum_{i=1}^n \left(y_i - \frac{\sum_{i=1}^n y_i + p_0 \mu_0}{n + p_0} \right)^2 + p_0 \left(\frac{\sum_{i=1}^n y_i + p_0 \mu_0}{n + p_0} - \mu_0 \right)^2 \right)$
  $= \frac{n-1}{2} \frac{\sum_{i=1}^n (y_i - \overline{y})^2}{n-1} = \frac{n-1}{2} \hat{\sigma}^2$

Estimaciones para τ
Usando **Jeffreys**:

$$p\left(\tau|\underline{y}\right) = Ga\left(\tau \left| \frac{n}{2}, \frac{n-1}{2}\hat{\sigma}^2\right.\right) \Rightarrow \mathbb{E}(\tau|\underline{y}) = \frac{n}{(n-1)\hat{\sigma}^2}; Moda = \frac{n-2}{(n-1)\hat{\sigma}^2}$$

$$p\left(\sigma^2|\underline{y}\right) = IGa\left(\tau \left| \frac{n}{2}, \frac{n-1}{2}\hat{\sigma}^2\right.\right) \Rightarrow \mathbb{E}(\sigma^2|\underline{y}) = \frac{(n-1)\hat{\sigma}^2}{n-2}; Moda = \frac{(n-1)\hat{\sigma}^2}{n+2}$$

Usando **Referencia**:

$$p\left(\tau|\underline{y}\right) = Ga\left(\tau \left| \frac{n-1}{2}, \frac{n-1}{2}\hat{\sigma}^2\right.\right) \Rightarrow \mathbb{E}(\tau|\underline{y}) = \frac{1}{\hat{\sigma}^2}; Moda = \frac{n-3}{(n-1)\hat{\sigma}^2}$$

$$p\left(\sigma^2|\underline{y}\right) = IGa\left(\tau \left| \frac{n-1}{2}, \frac{n-1}{2}\hat{\sigma}^2\right.\right) \Rightarrow \mathbb{E}(\sigma^2|\underline{y}) = \frac{(n-1)\hat{\sigma}^2}{n-3}; Moda = \frac{(n-1)\hat{\sigma}^2}{n+3}$$

Estimaciones para β
Usando **Jeffreys**:
Estimación Puntual:

$$p\left(\beta|\underline{y}\right) = T_1\left(\beta \middle| n, \overline{y}, \frac{n^2}{(n-1)\hat{\sigma}^2}\right) = T_1\left(\beta \middle| n, \overline{y}, \frac{n}{\hat{\sigma}^2_{M.V.}}\right)$$

$$\Rightarrow \mathbb{E}\left(\beta|\underline{y}\right) = \overline{y}; \quad \mathrm{Var}\left(\beta|\underline{y}\right) = \left(\frac{\frac{(n-1)}{2}\hat{\sigma}^2}{\frac{n}{2}-1}\right)\frac{1}{n} = \frac{\hat{\sigma}^2_{M.V.}}{n-2}$$

Estimación Intervalar:
Sabemos que:

$$\left(\beta - \overline{y}\right)\frac{n}{\hat{\sigma}^2_{M.V.}}\left(\beta - \overline{y}\right) = \left(\beta - \overline{y}\right)^2\frac{n}{\hat{\sigma}^2_{M.V.}} \sim F_{(1,n)} \Rightarrow \left(\beta - \overline{y}\right)\sqrt{\frac{n}{\hat{\sigma}^2_{M.V.}}} \sim t_{(n)}$$

Entonces:

$$\mathbb{P}\left(\overline{y} - t_{(n)}^{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\sigma}^2_{M.V.}}{n}} \leq \beta \leq \overline{y} + t_{(n)}^{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\sigma}^2_{M.V.}}{n}}\right) = 1-\alpha$$

Es un intervalo de credibilidad al $1-\alpha$ de probabilidad

Estimaciones para β
Usando **Referencia**:
Estimación Puntual:

$$p\left(\beta|\underline{y}\right) = T_1\left(\beta\,\middle|\, n-1, \overline{y}, \frac{n}{\hat{\sigma}^2}\right)$$

$$\Rightarrow \mathbb{E}\left(\beta|\underline{y}\right) = \overline{y}; \quad \mathrm{Var}\left(\beta|\underline{y}\right) = \left(\frac{\frac{(n-1)}{2}\hat{\sigma}^2}{\frac{n-1}{2}-1}\right)\frac{1}{n} = \frac{\hat{\sigma}^2_{M.V.}}{n-3}$$

Estimación Intervalar:
Sabemos que:

$$\left(\beta - \overline{y}\right)\frac{n}{\hat{\sigma}^2}\left(\beta - \overline{y}\right) = \left(\beta - \overline{y}\right)^2\frac{n}{\hat{\sigma}^2} \sim F_{(1,n-1)} \Rightarrow \left(\beta - \overline{y}\right)\sqrt{\frac{n}{\hat{\sigma}^2}} \sim t_{(n-1)}$$

Entonces:

$$\mathbb{P}\left(\overline{y} - t_{(n-1)}^{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\sigma}^2}{n}} \leq \beta \leq \overline{y} + t_{(n-1)}^{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\sigma}^2}{n}}\right) = 1-\alpha$$

Es un intervalo de credibilidad al $1-\alpha$ de probabilidad (Coincide con el intervalo clásico de confianza!!!)

Estimaciones para y^* (nueva observación)

Supongamos que queremos pronosticar un nuevo valor $w = y^*$. Entonces sabemos:

$$p\left(w|\underline{y}\right) = T_1\left(w \middle| 2\alpha_1, \mathbf{Z}\mu_1, \frac{\alpha_1}{\delta_1}\left(\mathbf{I} + \mathbf{Z}\mathbf{P}_1^{-1}\mathbf{Z}^T\right)^{-1}\right)$$

Donde en nuestro caso usando distribuciones no informativas tenemos que: $\mathbf{Z} = \mathbf{I} = 1$; $\mathbf{P}_1 = n$; $\mu = \overline{y}$; $\delta_1 = \frac{n-1}{2}\hat{\sigma}^2$. Por lo tanto:

$$p\left(w|\underline{y}\right) = T_1\left(w \middle| 2\alpha_1, \overline{y}, \frac{2\alpha_1}{(n-1)\hat{\sigma}^2}\left(1 + \frac{1}{n}\right)^{-1}\right)$$

Usando **Jeffreys**; ($\alpha_1 = \frac{n}{2}$)

$$p\left(w|\underline{y}\right) = T_1\left(w \middle| n, \overline{y}, \frac{n}{(n-1)\hat{\sigma}^2}\left(1 + \frac{1}{n}\right)^{-1}\right)$$

Usando **Referencia**; ($\alpha_1 = \frac{n-1}{2}$)

$$p\left(w|\underline{y}\right) = T_1\left(w \middle| n-1, \overline{y}, \frac{1}{\hat{\sigma}^2}\left(1 + \frac{1}{n}\right)^{-1}\right)$$

En ambas casos, la estimación puntal es $\overline{y}$.

Estimación intervalar
Usando **Jeffreys**; ($\alpha_1 = \frac{n}{2}$)

$$(w - \overline{y}) \frac{n}{(n-1)\hat{\sigma}^2} \left(1 + \frac{1}{n}\right)^{-1} (w - \overline{y}) \sim F_{(1,n)}$$

Lo que implica que:

$$(w - \overline{y}) \sqrt{\frac{n}{(n-1)\hat{\sigma}^2} \left(1 + \frac{1}{n}\right)^{-1}} = (w - \overline{y}) \sqrt{\hat{\sigma}^{-2}_{M.V.} \left(1 + \frac{1}{n}\right)^{-1}} \sim t_{(n)}$$

Por lo tanto:

$$\mathbb{P}\left(\overline{y} - t_{(n)}^{1-\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2_{M.V.} \left(1 + \frac{1}{n}\right)} \leq w \leq \overline{y} + t_{(n)}^{1-\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2_{M.V.} \left(1 + \frac{1}{n}\right)}\right) = 1 - \alpha$$

Estimación intervalar
Usando **Referencia**; ($\alpha_1 = \frac{n-1}{2}$)

$$(w-\overline{y})\,\frac{1}{\hat{\sigma}^2}\left(1+\frac{1}{n}\right)^{-1}(w-\overline{y}) \sim F_{(1,n-1)}$$

$$(w-\overline{y})\sqrt{\hat{\sigma}^{-2}\left(1+\frac{1}{n}\right)^{-1}} \sim t_{(n-1)}$$

Por lo tanto:

$$\mathbb{P}\left(\overline{y}-t_{(n-1)}^{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}\right)} \leq w \leq \overline{y}+t_{(n-1)}^{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}\right)}\right) = 1-\alpha$$

Obs: Se obtiene el mismo intervalo clásico, pero con diferente interpretración. !!!! Ver (10)

Ejemplo Numérico

Programa en R.

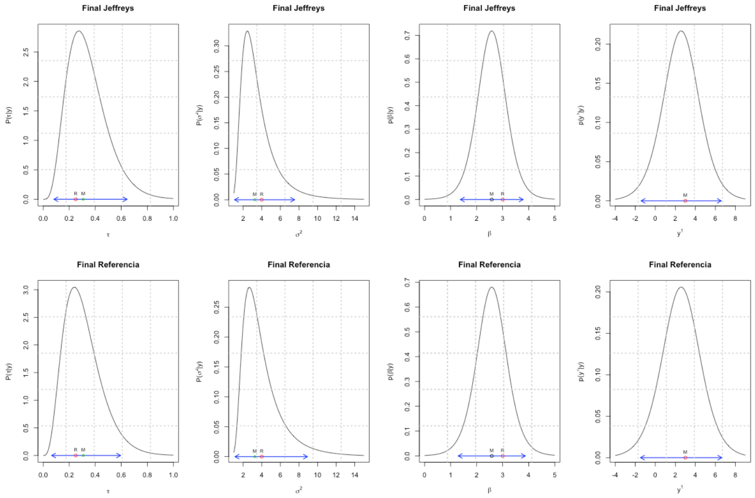
```
beta=3
tau=1/4
n=10
y=rnorm(n,3,tau^(-1/2))
4.3701841 2.9888996 1.4447173 6.7514057 2.2457418
2.0906260 1.7329556 1.4992820 0.3678763 2.3153768
#Estimacion \tau (Jeffreys)
    Media      Moda         MV       REAL
0.3417447 0.2733957 0.3075702 0.2500000
#Estimacion \sigma^2 (Jeffreys)
   Media     Moda        MV     REAL
3.657702 2.438468 3.251290 4.000000
#Estimacion \beta (Jeffreys)
  Media     Moda        MV     REAL
2.580707 2.580707 2.580707 3.000000
```

Ejemplo Numérico

Programa en R.

```
beta=3
tau=1/4
n=10
y=rnorm(n,3,tau^(-1/2))
4.3701841 2.9888996 1.4447173 6.7514057 2.2457418
2.0906260 1.7329556 1.4992820 0.3678763 2.3153768
#Estimacion \tau (Referencia)
    Media       Moda         MV       REAL
0.3075702 0.2392213 0.3075702 0.2500000
#Estimacion \sigma^2 (Referencia)
   Media      Moda        MV      REAL
4.180230 2.250893 3.251290 4.000000
#Estimacion \beta (Referencia)
   Media      Moda        MV      REAL
2.580707 2.580707 2.580707 3.000000
```

Ejemplo Numérico



Propuesta de una distribución a Priori

Muchas veces contaremos con información que queremos sea reflejada en una distirbución inicial. Esto lo debemos hacer ajustando los hiperparámetros correctos $(\mu_0, \mathbf{P}_0, \alpha_0, \delta_0)$. Desafortunadamente muchas veces será dificil que el experto en la materia pueda ofrecernos esta información directamente. Para dar solución a este problema Winkler (1977) propone lo siguiente:
Supongamos que tenemos el modelo lineal

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}; \quad \mathbf{X} \in \mathbb{R}^{n\times p}; \quad \underline{\beta} \in \mathbb{R}^p; \quad \underline{y}, \underline{\varepsilon} \in \mathbb{R}^n; \underline{\varepsilon} \sim N_n\left(\underline{0}, \tau\mathbf{I}\right)$$

El cual describe el experimento que estaremos estudiando. Ahora consideremos el modelo:

$$\underline{z} = \mathbf{X}^*\underline{\beta} + \underline{\varepsilon}^*; \quad \mathbf{X}^* \in \mathbb{R}^{m\times p}; \quad \underline{\beta} \in \mathbb{R}^p; \quad \underline{y}, \underline{\varepsilon}^* \in \mathbb{R}^m; \underline{\varepsilon} \sim N_m\left(\underline{0}, \tau\mathbf{I}\right)$$

Este último modelo tiene los mismos parámetros $\left(\underline{\beta}, \tau\right)$, pero suponemos que viene de un experimento anterior o bien $\underline{z}$ ya fue proporcionado gracias a la información de un experto que "subjetivamente"pronostica los distintos valores de $\underline{z}$ en las distintas covariables $\mathbf{X}^*$, una opción válida de hecho podría ser suponer $\mathbf{X}^* = \mathbf{X}$.

Propuesta de una distribución a Priori

Dado que $\underline{z}$ los suponemos observado y asumiendo una inicial no informativa, digamos la de **Referencia**, entonces sabemos que los parámetros finales son:

- $\mu_1 = \left(\mathbf{X}^{*T}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*T}\underline{z} = \mu_0$
- $\mathbf{P}_1 = \mathbf{X}^{*T}\mathbf{X}^{*} = \mathbf{P}_0$
- $\alpha_1 = \frac{n}{2} = \alpha_0$
- $\delta_1 = \frac{1}{2}\left(\left(\underline{z} - \mathbf{X}^{*}\mu_1\right)^T\left(\underline{z} - \mathbf{X}^{*}\mu_1\right)\right) = \delta_0$

Winkler propone usar esto parámetros finales del modelo hipotético $\underline{z} = \mathbf{X}^{*}\underline{\beta} + \underline{\varepsilon}^{*}$ como los parámetros iniciales del modelo que vamos a estudiar $\underline{y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}$;.

De esta manera los parámetros propuestos reflejan el conocimiento inicial del experto, o bien la información de un experimento anterior.

Ejemplo: Propuesta de una distribución a Priori

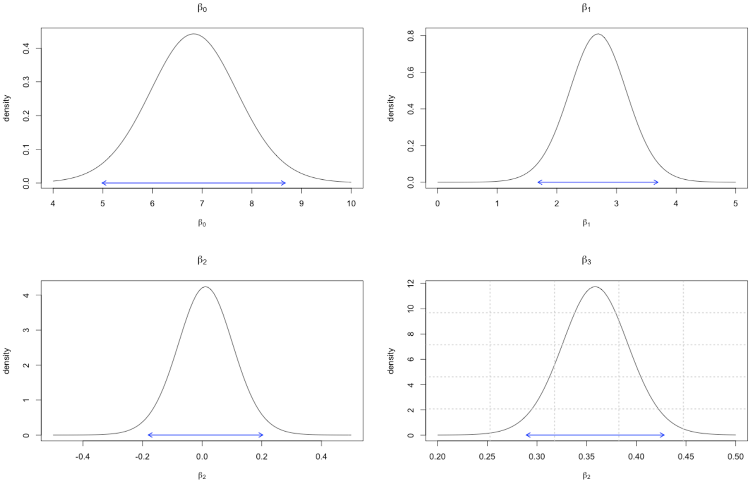
Programa en R.

```
#Simulamos de un modelo lineal (polinomial)
#Y=7+2.5*x+0.35*x^3+eps; eps \sim  N(0,10)
set.seed(1)
x=runif(26,-5,5)
#Simulamos observaciones de la variable respuesta
eps<-rnorm(26,0,sqrt(10))
y.sim=7+2.5*x+0.35*x^3+eps
#Ajustamos el modelo
#Y=beta0+beta1*x+beta2*x^2+beta*x^2+eps;  eps \sim N(0,\tau)
#Consideramos las primeras 6 observaciones como provenientes de otro experiment
z=y.sim[1:6]
X.star=cbind(rep(1,6),x[1:6],x[1:6]^2,x[1:6]^3)
#Las ultimas 20 obsevaciones se consideran del nuevo modelo
y=y.sim[7:26]
X=cbind(rep(1,20),x[7:26],x[7:26]^2,x[7:26]^3)
```

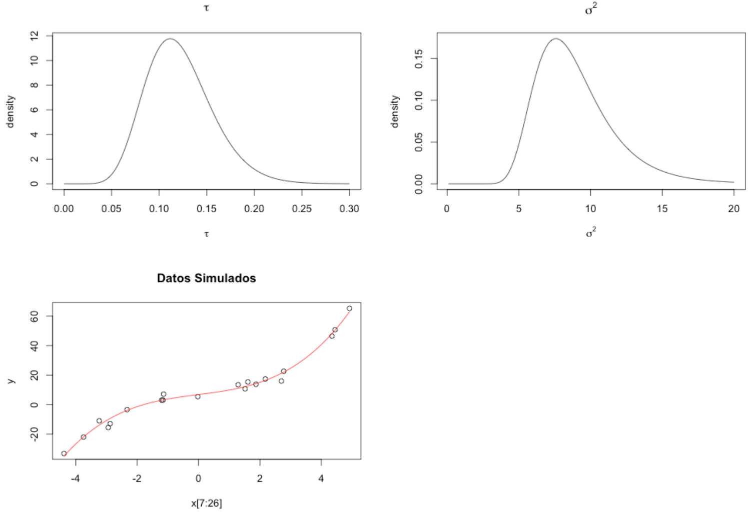
Ejemplo: Propuesta de una distribución a Priori

```
#Proponemos iniciales
m.0=solve(t(X.star)%*%X.star)%*%t(X.star)%*%z
P.0=t(X.star)%*%X.star
alpha.0=6/3
delta.0=1/2*t((z-X.star%*%m.0))%*% (z-X.star%*%m.0)
> m.0
          [,1]
[1,] 7.2457567
[2,] 3.6883241
[3,] 0.1181727
[4,] 0.2619451
> P.0
          [,1]       [,2]       [,3]       [,4]
[1,]  6.000000   2.187653   49.09877   90.10469
[2,]  2.187653  49.098771   90.10469  641.95921
[3,] 49.098771  90.104695  641.95921 1826.63761
[4,] 90.104695 641.959211 1826.63761 9500.51764
> alpha.0
[1] 2
> delta.0
         [,1]
[1,] 35.93982
```

Ejemplo: Distribuciones Finales



Ejemplo: Distribuciones Finales



¿Que es JAGS?

"Just Another Gibbs Sampler"

JAGS (Plummer, 2013)

Es un programa para el análisis de modelos Bayesianos usando Cadenas de Markov Monte Carlo

¿Que es JAGS?

JAGS fue escrito motivado para:

- Tener un motor para lenguaje BUGS que corra en Unix (Mac y Windows)
- Sea extendible
- Plataforma para experimentos con ideas de modelación Bayesiana

Corriendo un modelo en JAGS

Para obtener muestras de las distribuciones posteriores de los parámetros, JAGS realiza 5 pasos:

1. Definición del modelo
2. Compilación
3. Inicialización
4. Adaptacióin y burn-in
5. Monitoreo

Otros estado de análisis se realizan fuera de JAGS, por ejemplo, diagnísticos de convergencia.

1. Definición del Modelo

Existen dos partes en la definición del modelo en JAGS: El modelo y los Datos.

Descripción del modelo. El modelo es definido en un archivo de texto usando el lenguaje BUGS.

Ej RLS (Regresión Lineal Simple)

```
model {
for (i in 1:N){
Y[i] ~ dnorm(mu[i], tau)
mu[i] <- alpha + beta * x[i]
}
alpha ~ dnorm(0.0, 1.0E-4)
beta ~ dnorm(0.0, 1.0E-4)
tau ~ dgamma(1.0E-3, 1.0E-3)
sigma <- 1.0/sqrt(tau)

}
```

Datos

Los datos de pueden ser dados en un archivo por separado o directamente en R.
Por ejemplo:

- En .txt

```
"x"<- c(1, 2, 3, 4, 5)
"Y" <- c(1, 3, 3, 3, 5)
"N" <- 5
```

- En R:

```
x<- c(1, 2, 3, 4, 5)
Y <- c(1, 3, 3, 3, 5)
N <- 5
datos<-list("Y","x","N")
```

cont..

- **2. Compilación**: Verifica si no hay errores de sintaxis
- **3. Inicialización:**
  - El usuario puede fijar los valores iniciales.
  - Si no se especifican los valores iniciales, un "valor típico" es obtenido de la distribución a priori (media, mediana o moda)
- **4. Burn-in**
- **5. Monitoreo:** Un objeto que registra los valores de los parámetros en cada iteración. (p.e Trace monitor)

Artículos, libros, software

- Plummer, Martyn. (2013).
  JAGS Version 3.3.0 User Manual
- Plummer, Martyn y Northcott, Bill (2013).
  JAGS Version 3.4.0 Installation Manual