

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO.

FACULTAD DE CIENCIAS.

ANALISIS DE REGRESION BAYESIANO

T E S I S

QUE PARA OBTENER EL TITULO DE

A C T U A R I O

PRESENTA

EDUARDO CASTAÑO TOSTADO

MAYO 1984.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE.	pág.
1. INTRODUCCION.	1
2. INFERENCIA.	3
3. ANALISIS DE LAS SUPOSICIONES.	15
4. TOMA DE DECISIONES.	26
5. COMENTARIOS FINALES.	35
REFERENCIAS.	37

Capítulo 1. Introducción.

El modelado matemático puede verse como un proceso en el que se desea estudiar un problema específico mediante la construcción de una función (o varias), que describa las relaciones relevantes entre las variables involucradas de acuerdo a ciertos objetivos.

Los modelos estadísticos, como parte de este proceso, son útiles en situaciones de incertidumbre; en este caso, el modelo puede ser construido a partir de una suposición estructural y/o mediante el registro del comportamiento pasado del fenómeno.

Dentro de los modelos estadísticos más útiles está el modelo de regresión, que es construido fundamentalmente en forma estructural, a partir de suponer una función de verosimilitud en la que:

1) se establece una estructura de covarianza en Y (es decir, la matriz de varianzas y covarianzas Σ);

2) dada x una variable regresora se establece la estructura de dependencia de Y con ésta a través de

$$E[Y|x] = \mu(x; \beta)$$

donde β es un vector de parámetros desconocidos.

Así, el análisis de regresión se circunscribe a analizar cómo debe ser $\mu(x; \beta)$ específicamente; este tipo de modelado queda establecido como un problema típicamente estadístico, es decir, que dada una verosimilitud la incertidumbre radica en β y Σ , incertidumbre que es necesario tratar de eliminar para tener una descripción más clara del fenómeno.

El tratamiento que se dará en este trabajo es estrictamente Bayesiano; no se pretende establecer alguna discusión al respecto y se supondrá un

conocimiento al nivel de los libros de Winckler(1972) y Bernardo(1981). Dado este enfoque será necesario encontrar las distribuciones finales relevantes y plantear algunos problemas de decisión. En el capítulo dos se establecerán los resultados fundamentales de la inferencia, usando las suposiciones estándar del modelo. Algunas variantes a estas suposiciones serán presentadas en el siguiente capítulo. El cuarto capítulo plantea situaciones en las que es necesario tomar una decisión. Asimismo, a lo largo del trabajo se reportarán algunos resultados interesantes encontrados en la literatura revisada.

Capítulo 2. Inferencia.

En este capítulo se trabajará en el modelo de regresión lineal desde el enfoque bayesiano; como primer punto, se establecen las suposiciones estándar del modelo de regresión. Posteriormente se procederá a obtener la distribución final de los parámetros desconocidos; esto se hará por dos formas: la primera, que utilizará una distribución inicial de referencia por la regla de Jeffreys y la segunda que trabajará con familias conjugadas para obtener la distribución final. También se derivarán distribuciones específicas. Sin establecer una función de utilidad, se hacen comentarios sobre estimación y se presenta una forma de probar hipótesis. Se presenta lo desarrollado por medio de un ejemplo. Como punto final se presenta una alternativa para hacer inferencias en el modelo lineal.

2.1. Inferencia estándar.

Si el interés es estudiar Y a partir de un conjunto de variables x_1, \dots, x_k las suposiciones estándar del modelo de regresión son:

- 1) $E[y|x] = X\beta$ con β vector de parámetros desconocidos, X $n \times k$ matriz ($n > k$), es decir que la esperanza matemática de Y dada X se expresa como una transformación lineal en β ; en el caso de regresión se supone que X es de rango completo.
- 2) $V[y|x] = \sigma^2 I$, $\sigma^2 > 0$ desconocida; con esto se expresa el hecho de que las entradas del vector Y son no correlacionadas y homoscedásticas.
- 3) Y es un vector aleatorio que se distribuye Normal.

Con estas tres suposiciones se establece la verosimilitud. Nótese que las suposiciones 2 y 3 conjuntamente implican independencia en las entradas de Y . Por lo tanto se establece $Y \sim N(y|X\beta, \sigma^2 I)$, β, σ^2 desconocidos; es necesario remarcar que en ausencia de una de las suposiciones las restantes siguen siendo válidas; este hecho tendrá relevancia en el

siguiente capítulo.

Así, una vez que se tiene una observación del vector Y , la función de verosimilitud $l(\beta, \sigma | Y, X)$ es proporcional a

$$\sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta)\right\} \quad (1)$$

A continuación, ya que la incertidumbre radica en β y σ , se establecen distribuciones iniciales para éstos.

2.1.1. Distribuciones de referencia.

La metodología bayesiana requiere que se establezca una distribución inicial sobre los parámetros desconocidos; en el caso de que el conocimiento inicial sobre ellos sea difuso, o que no se desee que tal conocimiento intervenga, la regla de Jeffreys (1961) establece que si β y $\log \sigma$ son independientes y uniformemente distribuidos, la distribución inicial de referencia es

$$p(\beta, \sigma) \propto \frac{1}{\sigma} \quad \begin{array}{l} -\infty < \beta_i < \infty \\ 0 < \sigma < \infty \end{array} \quad i=1, \dots, k$$

Por medio del teorema de Bayes y (1) la distribución final es

$$p(\beta, \sigma | Y, X) \propto \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta)\right\}$$

Se puede mostrar (Zellner, 1971) que esta distribución puede expresarse de la siguiente forma

$$p(\beta, \sigma | Y, X) \propto \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2} (vS^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}))\right\}$$

con $vS^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})$, $v = n - k$ y $\hat{\beta}$ el estimador de mínimos cuadrados, es decir la solución de las ecuaciones normales $X'X\beta = X'Y$.

A partir de la distribución final conjunta de β y σ se puede mostrar que β condicional a σ se distribuye como una Normal multivariada con vector de medias $\hat{\beta}$ y matriz de varianzas y covarianzas $\sigma^2 (X'X)^{-1}$.

Otra forma de representar el conocimiento inicial sobre los parámetros desconocidos es por medio de familias conjugadas (Raiffa y Schlaifer, 1961);

éstas tienen las siguientes propiedades:

- i) son fáciles de manejar analíticamente y así facilitan la obtención de la distribución final;
- ii) son lo suficientemente flexibles para representar distintos estados en el conocimiento inicial;
- iii) en ciertos casos son propicias para una interpretación.

Denótese por w el inverso multiplicativo de σ^2 ; w y σ^2 en el desarrollo del trabajo se manejarán indistintamente de acuerdo a la conveniencia del punto en donde se involucren.

Se supone que la distribución inicial de β y w es tal que $\beta|w$ sigue una distribución Normal multivariada con media μ y matriz de precisión $w \Sigma$ ($\Sigma_{k \times k}$ y $\mu_{k \times 1}$ conocidos) y w tiene una distribución Gamma con parámetros α y δ (también conocidos), es decir

$$p(\beta, w) \propto w^{k/2} \exp \left\{ -\frac{w}{2} (\beta - \mu)' \Sigma (\beta - \mu) \right\} w^{\alpha-1} \exp \left\{ -\delta w \right\}; \begin{matrix} -\infty < \beta_i < \infty & i=1, \dots, k \\ 0 < w < \infty \end{matrix}$$

Esta distribución es la conjugada de la verosimilitud Normal con vector de medias y precisión desconocida como el siguiente teorema (DeGroot, 1970; pág. 183) lo establece:

Sean z_1, \dots, z_n vectores de una distribución Normal c -variada con vector de medias m desconocido y w R matriz de precisión con R conocida y w desconocida ($w > 0$); si la distribución inicial de m y w es tal que $m|w$ sigue una distribución Normal multivariada con media μ (conocida) y matriz de precisión $w \Sigma$ (Σ conocida), y la marginal de w es una Gamma con parámetros α y δ entonces la distribución final de m y w es como sigue: $m|w$ sigue una distribución Normal multivariada con media $\mu^* = (\Sigma + \delta R)^{-1} (\Sigma \mu + \delta R \bar{z})$ y matriz de precisión $w(\Sigma + \delta R)$, y la mar--

ginal de w es una Gama con parámetros $\alpha^* = \left(\alpha + \frac{c_2}{2}\right)$ y

$$y^* = y + \frac{1}{2} \sum_{i=1}^l (z_i - \bar{z})' R (z_i - \bar{z}) + \frac{1}{2} (\mu^* - \mu)' C (\bar{z} - \mu)$$

donde $\bar{z} = \frac{1}{l} \sum_{i=1}^l z_i$ (Ver demostración en el anexo).

En el caso de regresión se sustituye $(x'x)^{-1}x'y$ por cada z_i , $i=1, \dots, l$; además en este caso, sólo se tiene $l=1$; por otra parte note que en nuestra relación $c=n$ y $R=I_{n \times n}$.

Así, con $\beta_1 = (Z + X'X)^{-1}(Z\mu + X'y)$ y $y^* = y + \frac{1}{2} \{ (y - X\beta_1)' y + (\mu - \beta_1)' Z\mu \}$

se obtiene que $\beta_1 | w \sim N[\beta_1 | \beta_1, \frac{1}{w} (Z + X'X)^{-1}]$,

$$w \sim Ga[w | \alpha + \frac{n}{2}, y^*].$$

También se puede mostrar (DeGroot, 1970) que la marginal de β es una t-Student con $2\alpha + m$ grados de libertad, vector de localización β_1 y matriz de precisión $\frac{2\alpha + n}{2\sigma^2} (Z + X'X)$.

Ahora, en el caso de no información inicial DeGroot (1970) propone que los parámetros en la distribución final se hagan tender a los siguientes valores: C a la matriz 0 (cero), α a $-\frac{k}{2}$ y Y a 0 (cero), obteniéndose

$$\beta_1 | w \sim N[\beta_1 | \hat{\beta}_1, \frac{1}{w} (X'X)^{-1}],$$

$$w \sim Ga[w | \frac{n-k}{2}, \frac{1}{2} (Y - X\hat{\beta}_1)' Y]$$

$$\beta \sim t[\beta | m-k, \hat{\beta}_1, \frac{1}{S^2} (X'X)] \quad \text{con } S^2 = \frac{(Y - X\hat{\beta}_1)' (Y - X\hat{\beta}_1)}{n-k}.$$

Note que en este caso para obtener la distribución final de referencia de β y w no fue necesario suponer como en la regla de Jeffreys la independencia de β y $\log \sigma$, ni que sean uniformemente distribuidas; así, aunque analíticamente las distribuciones finales son idénticas - las suposiciones para generarlas son radicalmente distintas.

2.1.2. Otras distribuciones de interés.

A partir de la distribución final de β , w encontrada en la sección anterior por familias conjugadas, pueden derivarse las siguientes distribuciones de interés: la primera es la de $A\beta$ con A matriz $m \times k$; es fácil

ver que si $\beta | w \sim N[\beta | \mu^*, \frac{1}{w} \Sigma^*]$

entonces $A\beta | w \sim N[A\beta | A\mu^*, \frac{1}{w} A\Sigma^*A']$

y $A\beta \sim t[A\beta | 2\alpha+n, A\beta_1, \frac{2\alpha+n}{2\delta_1} (A\Sigma^{*-1}A')^{-1}]$

donde A sea tal que $A\Sigma^{*-1}A'$ sea invertible; análogamente al caso de no información inicial de la sección anterior, las distribuciones finales de referencia en este caso son

$A\beta | w \sim N[A\beta | A\hat{\beta}, \frac{1}{w} A'(X'X)^{-1}A]$ y $A\beta \sim t[A\beta | m-k, A\hat{\beta}, \frac{1}{s^2} [A(X'X)^{-1}A']^{-1}]$.

Si A es un vector de la forma (0,...,1,...,0) es decir un (1) en la posición i-ésima ($i=1, \dots, k$) se obtiene la distribución condicional y marginal para cada elemento de β ; Por otra parte, a A puede expresarse de tal forma que se determinen distribuciones sobre combinaciones lineales de β ; estos dos casos son generalmente de interés.

Otro problema de interés es la obtención de la distribución predictiva de Y; en este caso, dado el comportamiento pasado de Y, se desea saber su comportamiento, a partir de una predicción.

Obviamente, como toda predicción, entre más se aleje de la región muestral ya estudiada menos credibilidad se tendrá.

Denotemos por $p(\tilde{Y} | X, \tilde{X})$ la distribución predictiva del vector q-dimensional \tilde{Y} , donde $\tilde{X}_{q \times k}$ representa las observaciones de las variables regresoras en los q puntos a predecir.

Suponga que $E[\tilde{Y} | \tilde{X}, X, \gamma] = \tilde{X}\beta$, $V[\tilde{Y} | \tilde{X}, X, \gamma] = \sigma^2 I_{q \times q}$

y $\tilde{Y} \sim N[\tilde{Y} | \tilde{X}\beta, \sigma^2 I]$

con \tilde{X} de rango máximo, es decir se está suponiendo la misma estructu-

ra que se consideró en la región muestral ya estudiada. Se sabe que

$$p(\tilde{y}, \beta, \sigma | y, x, \tilde{x}) = p(\tilde{y} | \beta, \sigma, y, x, \tilde{x}) p(\beta, \sigma | y, x)$$

donde

$$p(\tilde{y} | \beta, \sigma, y, x, \tilde{x}) \propto \sigma^{-q} \exp \left\{ -\frac{1}{2\sigma^2} (\tilde{y} - \tilde{x}\beta)' (\tilde{y} - \tilde{x}\beta) \right\}$$

y $p(\beta, \sigma | y, x)$ como en la sección 2.1.1.

De esta manera considerando una distribución inicial de referencia para

β y σ proporcional a σ^{-1} ,

$$p(\tilde{y}, \beta, \sigma | y, x, \tilde{x}) \propto \sigma^{-(n+q+1)} \exp \left\{ -\frac{1}{2\sigma^2} (y-x\beta)'(y-x\beta) + (\tilde{y}-\tilde{x}\beta)'(\tilde{y}-\tilde{x}\beta) \right\}.$$

Como se desea $p(\tilde{y} | y, x, \tilde{x})$ es necesario integrar respecto a σ y a β ;

para encontrar la integral respecto a σ , se completa a una densidad

Gama en σ , resultando

$$p(\tilde{y}, \beta | y, x, \tilde{x}) \propto \left[(y-x\beta)'(y-x\beta) + (\tilde{y}-\tilde{x}\beta)'(\tilde{y}-\tilde{x}\beta) \right]^{-\frac{(v+q)}{2}}.$$

Ahora completando el cuadrado en β para tener una densidad t de student

$$p(\tilde{y} | y, x, \tilde{x}) \propto [y'y + \tilde{y}'\tilde{y} - (y'x + \tilde{y}'\tilde{x})M^{-1}(x'y + \tilde{x}'\tilde{y})]^{-\frac{(v+q)}{2}} \text{ con } M = X'X + \tilde{X}'\tilde{X}.$$

Desarrollando se llega a (Zellner, 1971):

$$p(\tilde{y} | y, x, \tilde{x}) \propto \left[v + (\tilde{y} - \tilde{x}\hat{\beta})' H (\tilde{y} - \tilde{x}\hat{\beta}) \right]^{-\frac{(v+q)}{2}}$$

con $\hat{\beta}$ el estimador de mínimos cuadrados y $H = \frac{1}{S^2} (I - \tilde{X}(X'X + \tilde{X}'\tilde{X})^{-1}X')$

i.e. $\tilde{y} \sim t[\tilde{y} | m-v, \tilde{x}\hat{\beta}, \frac{1}{v}H].$

Hasta aquí se ha encaminado el trabajo a obtener las distribuciones finales relevantes; en la siguiente sección se presenta a nivel inferencial (i.e. sin función de utilidad) la estimación puntual y la estimación por regiones de densidad.

2.1.3. Estimadores y regiones de alta densidad.

Si no se cuenta con una función de utilidad el "estimador" más natural, en el caso de una distribución unimodal es precisamente la moda; en el caso de β dado que su distribución final es una t de Student se estima

β con $\hat{\beta}$; a w se estima con $\frac{n-k-2}{(n-k)S^2}$ y a \tilde{Y} con $\tilde{X}\hat{\beta}$.

En cuanto a la región de alta densidad para β , se recomienda usar elipsoides de probabilidad (Box & Tiao, 1973), los cuales pueden ser obtenidos a partir de la distribución de β ; usando el hecho de que ésta es una función monótona de $(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})/KS^2$ que se distribuye según $F(k, v)$, un elipsoide de alta densidad de $1-\alpha$ es

$$(\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \leq KS^2 F(k, v, \alpha).$$

Para mayores detalles ver la referencia anterior.

2.1.4. Prueba de Hipótesis.

En algunos casos puede surgir la necesidad de saber con qué probabilidad el parámetro β se encuentra en una región del espacio parametral (\mathbb{R}^k). Muchas veces esta región está determinada por restricciones

$$A\beta = b$$

con A $p \times k$, $p < k$ de rango p , b $p \times 1$. Se sabe de las secciones anteriores que

$$\beta \sim t[\beta | m, \mu, T];$$

si denotamos por $S_0 = \{\beta \in \mathbb{R}^k : A\beta = b\}$; el problema es determinar si β está en S_0 .

De la definición de S_0 se tiene que su dimensión es menor que k y por lo tanto $\Pr[\beta \in S_0] = 0$.

DeGroot (1970), resuelve el problema mediante la comparación de las probabilidades contenidas en dos intervalos: uno, muy pequeño de volúmen v con centro en el punto modal de la distribución de β en \mathbb{R}^k , otro, del mismo volúmen, con centro en el punto modal de los que se encuentran en S_0 .

La probabilidad de un intervalo de este estilo puede ser calculada al multiplicar la densidad del punto modal del intervalo por el volúmen

k -dimensional de la región correspondiente; así, si β_1 es el punto modal en \mathbb{R}^k y β_0 el punto modal de los que están en S_0 la comparación se dá en términos de

$$\eta = \frac{p(\beta_1) \cdot v}{p(\beta_0) \cdot v} = \frac{\sup_{\beta \in \mathbb{R}^k} p(\beta)}{\sup_{\beta \in S_0} p(\beta)}$$

Nótese que $\eta \geq 1$.

En el numerador de este cociente se tiene que maximizar una distribución t -multivariada, es decir el problema es encontrar el valor de β donde alcance su supremo

$$c \left(1 + \frac{1}{m} (\beta - \mu)' T (\beta - \mu) \right)^{-\frac{(n+k)}{2}}$$

con $c = \left(\frac{\pi^{(n+k)/2}}{\Gamma(\frac{n+k}{2})} |T|^{1/2} \right) / \left(\frac{\pi^{n/2}}{\Gamma(\frac{n}{2})} (m\pi)^{k/2} \right)$ y $\gamma(\cdot)$ la función gamma.

Entonces, si $\beta = \mu$ la distribución alcanza su supremo, el valor 1.

Para encontrar el supremo del denominador se tiene el siguiente teorema:

Teorema.

Sea S_0 como se definió, sea $Q(\beta) = (\beta - \mu)' T (\beta - \mu)$; entonces

$$\inf_{\beta \in S_0} Q(\beta) = (b - A\mu)' (AT^{-1}A)^{-1} (b - A\mu)$$

Además el punto β_0 donde $Q(\beta)$ alcanza su ínfimo es

$$\beta_0 = \mu + T^{-1}A'(AT^{-1}A)^{-1}(b - A\mu).$$

Demostración:

Sea $x = \beta - \mu$, entonces el problema se transforma en

$$\text{donde } S'_0 = \left\{ x \mid Ax = c, c = b - A\mu \right\}, \quad q(x) = x'Tx.$$

s.a. $Ax = c$

Se demostrará que $x_0 = T^{-1}A'(AT^{-1}A)^{-1}c$ es el valor buscado.

Ya que A es de rango p y T es definida positiva, $AT^{-1}A'$ es de rango p y por lo tanto existe $(AT^{-1}A')^{-1}$.

$$\text{Por otro lado } Ax_0 = AT^{-1}A'(AT^{-1}A')^{-1}c = c$$

y por lo tanto $x_0 \in S'_0$.

Para mostrar que x_0 minimiza, primero se tiene

$$q(x_0) = (T^{-1}A'(AT^{-1}A')^{-1}c)' T (T^{-1}A'(AT^{-1}A')^{-1}c)$$

$$q(x_0) = c'(AT^{-1}A')^{-1}c = (b - A\mu)'(AT^{-1}A')^{-1}(b - A\mu). \quad 11$$

Si $q(x_0)$ es menor que $q(x)$ para todo $x \in S_0$, x_0 será el valor donde se minimiza $q(x)$. Ante esto hay dos posibilidades:

si $c=0$, $q(x_0)=0$; pero por ser $q(x)$ una forma definida positiva, entonces

$$q(x)=0 \iff x=0 \text{ con lo que } q(x_0) \leq q(x);$$

si $c \neq 0$, sea U una matriz tal que $U'U=T$; sean $Y=Ux, Y_0=Ux_0$; por la desigualdad de Schwarz

$$(Y'Y_0)^2 \leq (Y'Y)(Y_0'Y_0);$$

pero $(Y'Y_0)^2 = x'A'(AT^{-1}A')^{-1}Ax$

$$(Y'Y_0)^2 = q(x_0)$$

$$(Y'Y) = q(x)$$

además $(Y_0'Y_0) = q(x_0)$

Por lo tanto $q(x_0) \leq q(x) \quad \forall x \in S_0$.

De la misma desigualdad de Schwarz

$$q(x_0) = q(x) \iff \exists \lambda \in \mathbb{R} \cdot \text{ tal que } x_0 = \lambda x$$

pero $Ax=c$ y $Ax_0=c$ con lo que $Ax = Ax_0$; esto implica que

$\lambda = 1, x_0 = x$. Entonces si

$$q(x_0) \leq q(x)$$

y

$$q(x_0) = q(x) \iff x_0 = x$$

se tiene que

$$q(x_0) < q(x) \quad \forall x \in S_0.$$

Esto termina la demostración del teorema.

Así, el cociente de probabilidades toma la forma

$$Y = \left(1 + \frac{1}{m} (b - A\mu)'(AT^{-1}A')^{-1}(b - A\mu)\right)^{\frac{m+k}{2}},$$

y $Y = 1$ si y solo si $\beta \in S_0$; con esto se prueba a nivel inferencial,

es decir sin especificar una función de utilidad, la hipótesis $A\beta = b$;

Como último comentario debe establecerse que este procedimiento no es

el único para hipótesis del tipo $A\beta = b$. De hecho otra posibilidad es

calcular momios con las distribuciones finales; sin embargo el problema $P[\beta \in S_0] = 0$ no se evita, a menos que se dé una probabilidad mayor que cero al conjunto S_0 .

Utilizando regiones de alta densidad también pueden probarse hipótesis al nivel considerado. Si se denota a β_0 un punto en S_0 y se construye una región R_α de alta densidad $(1-\alpha)$, se aceptaría que $A\beta=b$ si β_0 estuviera contenida en R_α . Box y Tiao (1973) mencionan que este enfoque brinda una justificación al análisis de varianza Clásico desde el enfoque Bayesiano. El último enfoque que debe mencionarse es el considerar a la prueba de hipótesis como un proceso de toma de decisiones; esto se comentará en el capítulo cuatro.

2.2. Ejemplo.

En un estudio de la economía mexicana se quiere estudiar el desempleo (y) en términos de la tasa de inflación (x_1) y las importaciones (x_2). Se observaron los nueve primeros meses del año de interés, con los siguientes resultados:

mes	(Y) miles de desempleados	(x_1) tasa de inflación	(x_2) import. mill pesos
1	24	13	20
2	26	6	17
3	28	10	17
4	27	9	16
5	27	9	21
6	29	6	19
7	26	9	19
8	29	10	17
9	27	9	16.

Las suposiciones estándar del modelo de regresión son

$$E[y|x] = X\beta, \quad V[y|x] = \frac{1}{\omega} I \quad y \quad y \sim N[y|x\beta, \frac{1}{\omega} I]$$

La distribución inicial sobre β y ω se supone

$$p(\beta, \omega) = N[\beta | \mu, \frac{1}{\omega} \Sigma^{-1}] \times g_{\alpha}[\omega | \alpha, \delta],$$

y si $\alpha \rightarrow -\frac{k}{2}$, $\Sigma \rightarrow 0$, $\delta \rightarrow 0$ se tiene que

$$p(\beta, \omega) \propto \frac{1}{\omega}$$

Entonces

$$p(\beta, \omega | Y, X) = N[\beta | \hat{\beta}, \frac{1}{\omega} (X'X)^{-1}] g_{\alpha}[\omega | \frac{n-k}{2}, \frac{1}{2} Y'S^2]$$

$$y \quad p(\beta | Y, X) = t[\beta | m-k, \hat{\beta}, \frac{1}{S^2} X'X];$$

de $p(\beta, \omega | Y, X)$ se obtiene

$$p(\beta, \omega, Y, X) = N[\beta | (\frac{333}{10}, -\frac{3}{10}, -\frac{2}{10})', \frac{1}{\omega} (X'X)^{-1}] \quad X'X = \begin{bmatrix} 9 & 81 & 162 \\ 81 & 765 & 1464 \\ 162 & 1464 & 2942 \end{bmatrix}$$

marginalizando $p(\beta, \omega | Y, X)$ se tiene

$$p(\beta | Y, X) = t[\beta | (\frac{333}{10}, -\frac{3}{10}, -\frac{2}{10}), \frac{6}{15} X'X] \quad y \quad p(\omega | Y, X) = g_{\alpha}[\omega | 3, 7.5].$$

Si se desea hacer la prueba de hipótesis $\beta_2 = 0$, es decir que $A = (0, 0, 1)$

y $b = 0$, se tiene que $t = 1.33$ con lo que la hipótesis a nivel inferencial no se acepta.

Ahora, suponga que se desea hacer una predicción en los últimos tres meses del año en base a los siguientes valores:

mes	\tilde{x}_1	\tilde{x}_2
10	10	20
11	12	19
12	10	19

$$\text{Entonces} \quad p(\tilde{y} | Y, X, \tilde{x}) = t(\tilde{y} | 6, (26.3, 25.9, 26.5)', \frac{6}{15} B)$$

donde

$$B = \begin{pmatrix} 0.78 & -0.12 & -0.13 \\ -0.12 & 0.76 & -0.11 \\ -0.13 & -0.11 & 0.89 \end{pmatrix}.$$

Entonces la predicción a nivel inferencial es (26.3, 25.9, 26.5).

2.3. Otra alternativa.

Otro camino para manejar el modelo lineal, lo dan Lindley y Smith (1972) y Smith (1973), en términos de modelos jerárquicos; la idea es expresar la distribución inicial de β en base de hiperparámetros Θ_1 de tal forma que $E[\beta] = A_1 \Theta_1$ con matriz de dispersión C_1 . Este proceso se repite las veces necesarias teniendo como condición que en la última etapa se conozcan la media y la dispersión de la distribución ; por ejemplo, suponiendo que dado β

$$y \sim N[y | X\beta, \Sigma],$$

$$\beta \sim N[\beta | A_1 \Theta_1, C_1]$$

$$\Theta_1 \sim N[\Theta_1 | A_2 \Theta_2, C_2]$$

dado Θ_1

y dado Θ_2

entonces la distribución final de β dados $A_1, A_2, \Sigma, C_1, C_2, \Theta_2, Y$

es $N[\beta | Dd, d]$ donde

$$D^{-1} = X' \Sigma^{-1} X + \{C_1 + A_1 C_2 A_1'\}^{-1}$$

$$d = X' \Sigma^{-1} Y + \{C_1 + A_1 C_2 A_1'\}^{-1} A_1 A_2 \Theta_2$$

Sin embargo, este enfoque no plantea situaciones fáciles de resolver.

Capítulo 3. Análisis de las Suposiciones.

Se analizarán los efectos de las desviaciones en algunas de las suposiciones estándar del modelo de regresión. Primero se verá un enfoque debido a Box y Tiao (1973), sobre la suposición de Normalidad. A continuación se presentará, primero la generalización mediante familias conjugadas, a cualquier matriz de varianzas y covarianzas, y posteriormente algunos casos específicos de ésta matriz.

3.1. El Modelo Probabilístico en el Modelo de Regresión.

Una de las suposiciones fundamentales para llegar a las distribuciones descritas en el capítulo anterior, es que las observaciones se distribuyan según un modelo Normal; sin embargo, ¿qué pasaría si la función de densidad que mejor describe el comportamiento de Y no es la Normal?

En la inferencia Clásica o frecuentista la respuesta es que las distribuciones F son robustas ante desviaciones de la Normalidad.

Un criterio robusto es aquel que ante cambios en las suposiciones se ve poco afectado en la distribución de la estadística de prueba; en este sentido las pruebas clásicas que utilizan distribuciones F son criterios robustos ante No Normalidad. Aún así, Box y Tiao (1973) señalan que no debe confundirse un criterio robusto con una inferencia robusta en la que no se ven afectadas las conclusiones estadísticas ante cambios en las suposiciones; proponen un estudio de la desviación de la Normalidad desde un enfoque particular, en el que no se verifica si las observaciones tienen asociadas una verosimilitud Normal (bondad de Ajuste) sino que se incorpora en el análisis la discrepancia con tal verosimilitud, proponiendo un estudio preventivo más que correctivo.

A partir de los momentos es posible tener una idea aproximada de las carac.

terísticas de una distribución (Kendall y Stuart, 1977). En función del cuarto momento se define el coeficiente de curtosis (kurtosis) que indica la picudez de una distribución. Basándose en este coeficiente Box y Tiao (1973, pág. 151), afirman: "el investigador raramente estará en posición de tener certeza de la forma precisa de toda la distribución...; por el contrario, su opinión podrá ser descrita por una familia de distribuciones centradas en una distribución focal".

En este sentido la distribución que proponen, llamada potencia exponencial (exponential power), es de la forma

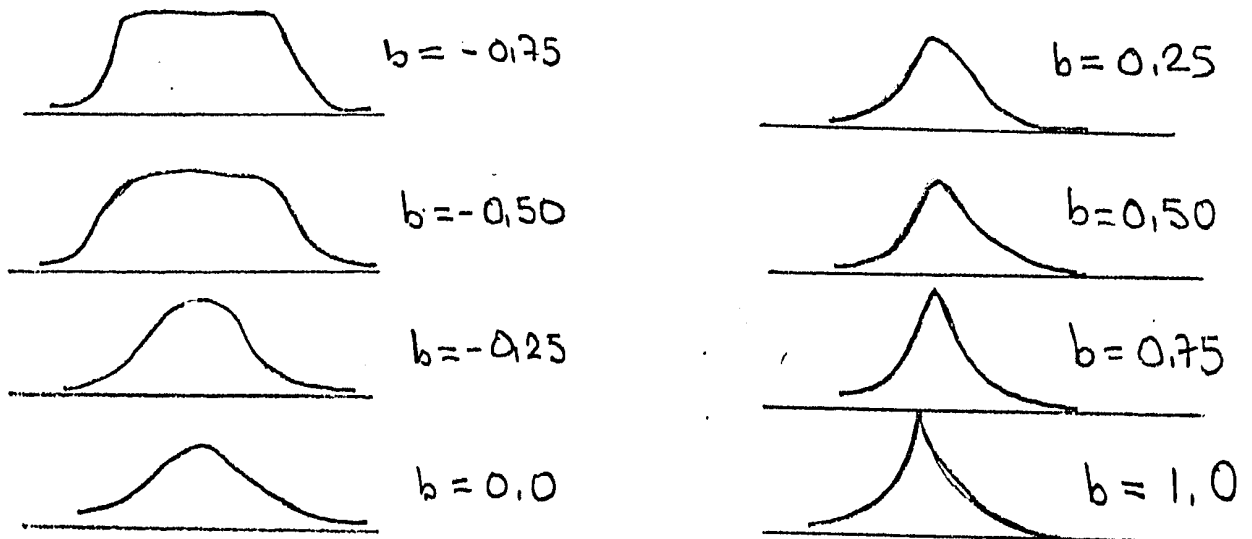
$$p(y_i | \beta, \sigma, b, x) \propto w(b) \sigma^{-1} \exp \left\{ -c(b) \left| \frac{y_i - X_{(i)} \beta}{\sigma} \right|^{\frac{2}{1+b}} \right\}$$

donde $x'_{(i)}$ es el i -ésimo renglón de la matriz de diseño,

$$\sigma > 0, \quad b \in (-1, 1], \quad \sigma \in \cdot$$

$$c(b) = \left\{ \frac{\Gamma\left[\frac{3}{2}(1+b)\right]}{\Gamma\left[\frac{1}{2}(1+b)\right]} \right\}^{\frac{1}{1+b}}, \quad w(b) = \frac{\left\{ \Gamma\left[\frac{3}{2}(1+b)\right] \right\}^{\frac{1}{2}}}{(1+b) \left\{ \Gamma\left[\frac{1}{2}(1+b)\right] \right\}^2}$$

El coeficiente b tiene la interpretación de una medida de curtosis de la distribución (ver figura tomada de Box y Tiao, 1973, pág. 159)



Si $b=0$ se recupera la Normalidad, cuando $b=1$ y b tiende a -1 , Box y Tiao mencionan que la distribución es una doble exponencial y una distribución uniforme en $(x_{(i)}\beta - \sqrt{3}\sigma, x_{(i)}\beta + \sqrt{3}\sigma)$ respectivamente.

Suponiendo que cada Y_i tiene la distribución potencia exponencial, la verosimilitud resulta ser

$$L(\beta, \sigma, b | Y, X) \propto w(b) \sigma^{-n} \exp \left\{ -c(b) \sum_{i=1}^n \left| \frac{y_i - x_{(i)}\beta}{\sigma} \right|^{\frac{2}{1+b}} \right\}$$

Surgen así dos posibilidades:

- 1) b conocido,
- 2) b desconocido.

En el caso (1), si $p(\beta, \sigma) \propto \sigma^{-1}$ la distribución final es

$$p(\beta, \sigma | b, Y, X) \propto \sigma^{-(n+1)} \exp \left\{ -c(b) \sum_{i=1}^n \left| \frac{y_i - x_{(i)}\beta}{\sigma} \right|^{\frac{2}{1+b}} \right\}$$

por lo que la distribución final de β es

$$p(\beta | b, Y, X) \propto \int_0^\infty \sigma^{-(n+1)} \exp \left\{ -\frac{c(b)}{\sigma^{\frac{2}{1+b}}} \left(\sum_{i=1}^n |y_i - x_{(i)}\beta|^{\frac{2}{1+b}} \right) \right\} d\sigma$$

Haciendo $z = \frac{c(b)}{\sigma^{\frac{2}{1+b}}}$ e integrando respecto a esta nueva variable, se tiene

$$p(\beta | b, Y, X) = J(b) M(\beta)^{-\frac{n}{2}(1+b)}$$

con $M(\beta) = \sum_{i=1}^n |y_i - x_{(i)}\beta|^{\frac{2}{1+b}}$ y $J(b) = \int_{\mathbb{R}^k} M(\beta)^{-\frac{n}{2}(1+b)} d\beta$

Box y Tiao demuestran que $p(\beta | b, Y, X)$ es tal que:

- a) Es una función monótona de $M(\beta)$.
- b) Es una función continua, y para $-1 < b < 1$ diferenciable y unimodal.
- c) Si $b=0$ $M(\beta) = (vS^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta}))$ con lo que $\beta \sim t(\beta | v, \hat{\beta}, \frac{1}{2} X' X)$.

d) $p(\beta | b, Y, X)$ no toma una forma simple; pero puede ser aproximada por una distribución t en el caso $-0.5 < b < 0.5$.

En el caso (2), suponen que b es independiente de β y σ , es decir,

$$p(\beta, \sigma, b | Y, X) \propto p(b) \sigma^{-1} L(\beta, \sigma, b | Y, X)$$

Ahora

$$p(\beta, b | Y, X) \propto p(b) w(b)^n \int_0^\infty \sigma^{-(n+1)} \exp \left\{ -\frac{c(b) M(\beta)}{\sigma^{\frac{2}{1+b}}} \right\} d\sigma$$

Sea $z = \frac{c(b)}{\sigma^{\frac{2}{1+b}}}$ entonces la expresión anterior se expresa como

$$p(\beta, b | Y, X) \propto p(b) \omega(b)^m \left(\frac{1+b}{2}\right)^{\int_0^{\infty} \frac{(m+1)(1+b) - (1+b)z}{z} dz} \exp\{-Az\} dz$$

con $A = a(b)M(\beta)$.

Así,

$$p(\beta, b | Y, X) \propto p(b) \omega(b)^m \frac{2}{1+b} \left(\gamma[\alpha] \cdot A^{-\alpha} \right),$$

con $\alpha = \frac{m(1+b)}{2}$

Integrando respecto a β

$$p(b | Y, X) \propto p(b) \frac{\gamma\left[\frac{3}{2}(1+b)\right]^{\frac{m}{2}}}{(1+b)^m \gamma\left[\frac{1}{2}(1+b)\right]^{\frac{2m}{2}}} \left(\frac{1+b}{2}\right) \gamma\left[\frac{m(1+b)}{2}\right] \frac{\gamma\left[\frac{1}{2}(1+b)\right]^{\frac{m}{2}}}{\gamma\left[\frac{3}{2}(1+b)\right]^{\frac{m}{2}}} J(b).$$

Ahora, si se supone que $p(b)$ es Uniforme en $(-1, 1)$, se tiene que

$$p_u(b | Y, X) \propto \delta\left[\frac{1}{2}(1+b)+1\right] \delta\left[\frac{1}{2}(1+b)+1\right]^{-n} J(b).$$

En general $p(b | Y, X) \propto p(b) p_u(b | Y, X)$, donde $p(b)$ es la densidad inicial que contenga la información del investigador. Box y Tiao sugieren que una distribución para el parámetro b que resulta ser adecuada para estudiar los efectos de las desviaciones de la Normalidad es

$$p(b) = u (1-b^2)^{a-1}, \quad -1 < b \leq 1, \quad a \geq 1, \quad u = \frac{\Gamma[2a] \Gamma[a]^2}{\Gamma[2a+1]} \cdot 2^{-(2a-1)}$$

Esta función tiene moda única en $b=0$. Por otro lado, es fácil ver que si $x = \frac{1+b}{2}$ entonces $x \sim \text{be}(x | a, a)$, de donde $E[x] = \frac{1}{2}$ y por lo tanto $E[b] = 0$.

De esta forma la información proveniente de la muestra es utilizada para cuantificar las desviaciones; ésto no sucede en la estadística Clásica en la que la información es usada para "probar" Normalidad y que en ausencia de resultados significativos no es aprovechada directamente en el análisis.

La distribución final de β se obtiene entonces

$$p(\beta | Y, X) = \int p(\beta | Y, X, b) p(b) db$$

Obviamente en la obtención de $p(\beta | Y, X)$ tendrá que influir el valor que se asigna al parámetro a en $p(b)$; así, cuando $a=1$ la distribución $p(b)$ es uniforme, cuando $a \neq 1$ $p(b)$ es una distribución simétrica con moda en $b=0$ y para valores muy grandes de a esta distribución se irá concentrando alre-

dedor de $b=0$, llegando en el límite a que aproximadamente sea una distribución delta, y así la Normalidad se recupera. Este análisis de sensibilidad es posible llevarlo a cabo en $p(\beta|b, y, x)$, en donde si a cambios en b esta distribución experimenta fuertes cambios se debe mantener sin eliminar b ya que es más informativa que β pues maneja explícitamente las desviaciones en cuanto a curtosis de la Normalidad.

Cabe señalar que este desarrollo es enteramente aplicable en el caso de que el modelo estadístico no sea lineal en β (Box y Tiao, 1973, pág. 186-193)

3.2. Las Suposiciones respecto a la Matriz de Precisión de Y.

En todo lo desarrollado anteriormente, la matriz de precisión se ha supuesto igual a wI , donde w es un parámetro desconocido mayor que cero, sin embargo puede haber situaciones en la que la forma de la matriz de precisión sea distinta, por ejemplo:

- 1) de la forma wR , donde R es una matriz conocida simétrica y definida positiva pero distinta a la identidad,
- 2) de la forma wR , con R desconocida.

En ambos casos el problema está resuelto teóricamente en base a familias conjugadas (DeGroot, 1970). Para el caso (1) se usa como familia conjugada la distribución Normal Gamma. Si

$$p(y|\beta, w, x) = N[y | x\beta, \frac{1}{w} R^{-1}]$$

y
$$p(\beta, w) = p(\beta|w) p(w)$$

con
$$p(\beta|w) = N[\beta | \mu, \frac{1}{w} \Sigma^{-1}]$$

$$p(w) = g_a[w | \alpha, \eta]$$

entonces
$$p(\beta|w, y, x) = N[\beta | (\Sigma + R)^{-1} (\Sigma \mu + R y), \frac{1}{w} (\Sigma + R)^{-1}]$$

y
$$p(w|y, x) = g_a[w | \frac{\alpha+k}{2}, \frac{\alpha+1}{2} [(\Sigma + R)^{-1} (\Sigma \mu + R y) - \mu] \Sigma^{-1} (\Sigma + R)^{-1}]$$

Para la demostración de este teorema véase el Apéndice.

En el caso (2), se usa la familia conjugada Normal Wishart; véase la demos-

tracción general de este teorema en el Apéndice.

Como el número de parámetros involucrados en la matriz Σ puede ser grande, es poco realista suponer que se tiene conocimiento específico sobre su comportamiento, por lo que es razonable el utilizar una distribución inicial de referencia en estos casos.

Bernardo (1979) y Geisser y Cornfield (1963) trabajan en este sentido; el segundo trabajo muestra que en el caso de una verosimilitud Normal con vector de medias μ y matriz de varianzas y covarianzas Σ la distribución de referencia está dada por

$$p(\mu, \Sigma) \propto |\Sigma|^{-\frac{(p+1)}{2}}$$

Villegas (1969) trabaja en la obtención de una distribución de referencia para Σ en el caso Normal.

3.3. Representaciones Especiales para _____.

Se presentan a continuación situaciones específicas en las que

3.3.1. Autocorrelación de orden 1.

Suponga que $y_t = x_t' \beta + u_t$

con $u_t = \rho u_{t-1} + \varepsilon_t \quad t=1, \dots, m, \quad -\infty < \rho < \infty.$

donde $E[\varepsilon_t] = 0$.

Entonces, se tiene un modelo autocorrelacionado de primer orden.

Así, $E[y_t] = x_t' \beta + E[u_t] = x_t' \beta + \rho E[u_{t-1}]$

$$E[y_t] - \rho E[y_{t-1}] = (x_t - \rho x_{t-1})' \beta$$

y por lo tanto $E[y_t] = \rho E[y_{t-1}] + (x_t - \rho x_{t-1})' \beta \quad t=1, \dots, m.$

Nótese que las expresiones anteriores forman un sistema iterativo en el que es necesario manejar y_0 para resolverlo, entonces supóngase que

$$y_0 \sim N[y_0 | x_0' \beta + M, \sigma^2],$$

donde M es un parámetro desconocido; a partir de esto se obtiene la expres-

sión explícita de $E[y_t]$ como

$$E[y_t] = \rho^t (x_0' \beta + M) + \sum_{j=1}^t \rho^{t-j} (x_j - x_{j-1})' \beta \quad t=1, \dots, n.$$

Entonces la verosimilitud es

$$l(\beta, \sigma, M, \rho | y_n, y_0) \propto \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} \left[(y_0 - x_0' \beta - M)^2 + \sum_{t=1}^n (y_t - E[y_t])^2 \right] \right\}.$$

Ahora, si se supone que $\beta, \rho, \log \sigma$ y M son uniforme e independientemente

distribuidos, ésto es $p(\beta, \rho, \sigma, M) \propto \frac{1}{\sigma}$.

la distribución final respectiva resulta ser

$$p(\beta, \rho, \sigma, M | y, y_0, x, x_0) \propto \sigma^{-(n+2)} \exp \left\{ -\frac{1}{2\sigma^2} \left[(y_0 - x_0' \beta - M)^2 + \sum_{t=1}^n (y_t - E[y_t])^2 \right] \right\}$$

Integrando sobre M , $p(\beta, \rho, \sigma | y, y_0, x, x_0) \propto \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n (y_t - E[y_t])^2 \right\}$.

Generalmente el interés estará centrado en β y ρ , marginalizando con respecto a σ , $p(\beta, \rho | y_n, y_0, x, x_0) \propto \left\{ \sum_{t=1}^n (y_t - E[y_t])^2 \right\}^{-\frac{n}{2}}$.

Esta distribución permite hacer inferencias sobre β y ρ .

3.3.2. Observaciones No Homoscedásticas.

En esta sección se presentan los aspectos relevantes de dos enfoques que se hallaron durante la revisión bibliográfica.

Zellner (1971), estudia la situación en la que tiene dos varianzas durante la observación del vector Y , es decir

$$y_i \sim N(y_i | x_i' \beta, \sigma_1^2) \quad i=1, \dots, m_1$$

$$y_j \sim N(y_j | x_j' \beta, \sigma_2^2) \quad j=m_1+1, \dots, m_1+n_2$$

En este caso la verosimilitud es

$$l(\beta, \sigma_1, \sigma_2 | Y, X) \propto \sigma_1^{-m_1} \sigma_2^{-n_2} \exp \left\{ -\frac{1}{2\sigma_1^2} (Y_1 - X_1 \beta)' (Y_1 - X_1 \beta) - \frac{1}{2\sigma_2^2} (Y_2 - X_2 \beta)' (Y_2 - X_2 \beta) \right\}$$

donde $Y_1' = (y_1, \dots, y_{m_1})$ y $Y_2' = (y_{m_1+1}, \dots, y_{m_1+n_2})$.

Supone una distribución inicial de referencia $p(\beta, \sigma_1, \sigma_2) \propto \sigma_1^{-1} \sigma_2^{-1}$ con $\beta_i \in (-\infty, \infty) \quad i=1, \dots, k$, $\sigma_j \in (0, \infty) \quad j=1, 2$.

Así, $p(\beta, \sigma_1, \sigma_2 | Y, X) \propto \sigma_1^{-(m_1+1)} \sigma_2^{-(n_2+1)} \exp \left\{ -\frac{1}{2\sigma_1^2} (Y_1 - X_1 \beta)' (Y_1 - X_1 \beta) - \frac{1}{2\sigma_2^2} (Y_2 - X_2 \beta)' (Y_2 - X_2 \beta) \right\}$.

Completando a una función de distribución Gamma en σ_1 e integrando y procediendo de la misma forma respecto a σ_2 se tiene que

$$p(\beta | Y, X) \propto \left[1 + \frac{(\beta_1 - \hat{\beta}_1)' X_1' X_1 (\beta_1 - \hat{\beta}_1)}{v_1 S_1^2} \right]^{-\frac{1}{2}(v_1 + k)} \left[1 + \frac{(\beta_2 - \hat{\beta}_2)' X_2' X_2 (\beta_2 - \hat{\beta}_2)}{v_2 S_2^2} \right]^{-\frac{1}{2}(v_2 + k)} \quad 21$$

que es una distribución llamada Doble-t.

El otro enfoque es desarrollado por Box y Tiao (1973), bajo el mismo tratamiento del punto 3.1, es decir, a partir de la familia potencia exponencial desarrollan el estudio del caso en que se tienen varias varianzas. Suponga que se tienen k varianzas, entonces la verosimilitud es

$$L(\beta, \sigma, b | Y, X) = w(b)^m \prod_{i=1}^k \sigma_i^{-n_i} \left[\frac{-\frac{1}{2} c(b) m_i s_i(b, \beta_i)}{\sigma_i^{2/1+b}} \right]$$

donde

$s_i(b, \beta_i) = \frac{1}{m_i} \sum_{j=1}^{m_i} |y_{ij} - x'_{ij} \beta_i|^{2/1+b}$, $\beta = (\beta_1, \dots, \beta_k)$, $\sigma = (\sigma_1, \dots, \sigma_k)$, $Y = (Y_1, \dots, Y_k)$
 $Y_i = (Y_{i1}, \dots, Y_{in_i})$, $i=1, \dots, k$, $m = \sum_{i=1}^k m_i$, x'_{ij} el i-ésimo renglón del j-ésimo grupo correspondiente a σ_j^2 , y $c(b)$ y $w(b)$ como antes.

si $p(\sigma_i) \propto \frac{1}{\sigma_i}$

$$p(\sigma | \beta, b, Y, X) = \prod_{i=1}^k p(\sigma_i | \beta_i, b, Y_i, X_i)$$

$$\text{con } p(\sigma_i | \beta_i, b, Y_i, X_i) \propto \sigma_i^{-(n_i+1)} \exp \left\{ -\frac{1}{2} \frac{c(b) m_i s_i(b, \beta_i)}{\sigma_i^{2/1+b}} \right\}$$

Proponen el estudio de $\phi = (\phi_1, \dots, \phi_{k-1})$ donde $\phi_i = 2(\log \sigma_i - \log \sigma_k)$.

Así, el estudio se centra en

$$p(\phi | \beta, b, Y, X) = C \prod_{i=1}^{k-1} U_i^{\frac{1}{2} n_i (1+b)} \left(1 + \sum_{i=1}^{k-1} U_i \right)^{-\frac{n}{2} (1+b)}$$

donde

$$-\infty < \phi_i < \infty \quad i=1, \dots, k-1$$

$$U_i = \frac{n_i s_i(b, \beta_i)}{m_k s_k(b, \beta_k)} \exp \left\{ \frac{-\phi_i}{1+b} \right\} \quad i=1, \dots, (k-1).$$

3.4. Comentarios Generales.

Otros trabajos hallados en la literatura que versan sobre las suposiciones del modelo de regresión son:

El de Zellner (1976), en el que presenta el caso de una distribución t en lugar de suponer Normalidad; su análisis es realizado frecuentista y Bayesianamente. En el segundo caso cuando supone una distribución inicial de referencia obtiene la distribución ya obtenida en el caso de Normalidad. Oman (1978), estudia la multicolinealidad de la matriz de diseño X ; parte de las distintas soluciones clásicas al problema de estimación de β (Stein, Ridge Regression, componentes Principales, etc.), y encuentra una distribución inicial una distribución final cuya moda sea el respectivo estimador. Belsley, Kuh y Wesch (1980), mencionan que el problema de colinealidad es un problema de diseño más que un problema estadístico y en particular de conocimiento inicial: es decir su solución es estrictamente numérica.

Apéndice

Teorema 1

Suponga que $\underline{X}_1, \dots, \underline{X}_n$ es una m.a. de una distribución Normal multivariada para la cual el vector de medias \underline{M} tiene un valor desconocido y la matriz de precisión es de la forma $W r$ donde r es una matriz definida positiva especificada, W 0 es desconocida. Suponga también que la distribución apriori de M y W es tal

$$M|W=w \sim N[m|\mu, wZ] \quad \mu \in \mathbb{R}^k, \quad Z_{k \times k} \text{ simétrica}$$

positiva definida,

$$W \sim ga[w|\alpha, \beta] \quad \alpha > 0, \beta > 0.$$

Entonces la distribución conjunta de M y W es tal que

$$M|W=w \sim N[\mu^*, wZ^*] \quad \text{con } \mu^* = (Z+nr)^{-1}(Z\mu+nr\bar{x}), Z^* = (Z+nr)$$

$$y \quad W \sim ga[w|\alpha + \frac{nk}{2}, \beta + \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})'(x_i - \bar{x}) + \frac{1}{2} (\mu^* - \mu)' Z (\bar{x} - \mu)].$$

Demostración:

$$\text{la verosimilitud } l(m, w | x_1, \dots, x_n) \propto |r|^{\frac{n}{2}} w^{\frac{kn}{2}} \exp\left\{-\frac{w}{2} \sum_{i=1}^n (x_i - m)' r (x_i - m)\right\}$$

$$\text{la distribución inicial } p(m, w) \propto |Z|^{\frac{1}{2}} w^{\frac{k}{2}} \exp\left\{-\frac{w}{2} (m - \mu)' Z (m - \mu)\right\} w^{\alpha-1} e^{-\beta w}.$$

Desarrollando

$$\sum (x_i - m)' r (x_i - m) = n(m - \bar{x})' r (m - \bar{x}) + \sum (x_i - \bar{x})' r (x_i - \bar{x})$$

$$n(m - \bar{x})' r (m - \bar{x}) + (m - \mu)' Z (m - \mu)$$

$$= [m - (Z+nr)^{-1}(Z\mu+nr\bar{x})]' (Z+nr) [m - (Z+nr)^{-1}(Z\mu+nr\bar{x})]$$

$$- (Z\mu+nr\bar{x})' (Z+nr) (Z\mu+nr\bar{x}) + n\bar{x}' r \bar{x} + \mu' Z \mu$$

$$= (m - \mu^*)' (Z+nr) (m - \mu^*) + (\mu^* - \mu)' Z (\bar{x} - \mu)$$

ya que
$$-\mu^{*'}(\zeta\mu + nr\bar{x}) + n\bar{x}'r\bar{x} + \mu'\zeta\mu = (\mu^* - \mu)' \zeta (\bar{x} - \mu)^2 +$$

Teorema 2.

Suponga que X_1, \dots, X_n es una m.a. de una distribución Normal multivariada con un vector de medias M desconocido y una matriz de precisión R desconocida. Suponga también que la distribución conjunta inicial de M y R es como sigue: la distribución condicional de M cuando $R = r$ es una Normal multivariada con media μ y matriz de precisión νr tal que $\mu \in \mathbb{R}^k$ y $\nu > 0$, y la marginal de R es una distribución Wishart con α grados de libertad y matriz de precisión ζ tal que $\alpha > k-1$ y ζ es una matriz simétrica positiva definida. Entonces la distribución final de M y R cuando $X_i = x_i$ ($i=1, \dots, n$) es como sigue: La condicional de M

cuando $R = r$ es una Normal multivariada con media μ^* y matriz de precisión $(\nu+n)r$ donde $\mu^* = \frac{\nu\mu + n\bar{x}}{\nu+n}$ y la marginal de R es una Wishart con $\alpha+n$ grados de libertad y matriz de precisión ζ^* , donde $\zeta^* = \zeta + \frac{\nu n}{\nu+n} (\mu - \bar{x})(\mu - \bar{x})'$,

$$S = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

Demostración:

$$\begin{aligned} \sum_{i=1}^n (x_i - m)' r (x_i - m) &= n(\bar{x} - m)' r (\bar{x} - m) + \sum_{i=1}^n (x_i - \bar{x})' r (x_i - \bar{x}) \\ &= m(\bar{x} - m)' r (\bar{x} - m) + \sum_{i=1}^n \text{tr} (x_i - \bar{x})' r (x_i - \bar{x}) \\ &= m(\bar{x} - m)' r (\bar{x} - m) + \sum_{i=1}^n \text{tr} (x_i - \bar{x})' (x_i - \bar{x}) r \\ &= m(\bar{x} - m)' r (\bar{x} - m) + \text{tr} \left\{ \left[\sum_{i=1}^n (x_i - \bar{x})' (x_i - \bar{x}) \right] r \right\} \end{aligned}$$

Por otro lado

$$\nu(m - \mu)' r (m - \mu) + n(m - \bar{x})' r (m - \bar{x})$$

$$= (\nu+n) \left(m' r m - \frac{2m' r (\nu\mu + n\bar{x})}{\nu+n} \right) + \nu (\mu' r \mu) + n (\bar{x}' r \bar{x}) \quad \text{25}$$

$$= (\nu+n) \left[\left(m - \frac{\nu\mu + n\bar{x}}{\nu+n} \right)' r \left(m - \frac{\nu\mu + n\bar{x}}{\nu+n} \right) \right]$$

$$- \frac{1}{\nu+n} (\nu\mu + n\bar{x})' r (\nu\mu + n\bar{x})$$

$$= (\nu+n) (m - \mu^*)' r (m - \mu^*) - \frac{n\nu}{\nu+n} (\mu - \bar{x})' r (\mu - \bar{x})$$

$$= (\nu+n) (m - \mu^*)' r (m - \mu^*) - \text{tr} \left[\frac{n\nu}{\nu+n} (\mu - \bar{x}) (\mu - \bar{x})' r \right].$$

Entonces

$$\xi(m, r) \mathcal{L}(m, r | x_1, \dots, x_n) \propto |r|^{1/2} \exp \left\{ -\frac{\nu}{2} (m - \mu)' r (m - \mu) \right\}$$

$$\times |r|^{(\alpha-k-1)/2} \exp \left\{ -\frac{1}{2} \text{tr}(Zr) \right\}$$

$$\times |r|^{n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^m (x_i - m)' r (x_i - m) \right\}$$

$$= |r|^{1/2} \exp \left\{ -\frac{\nu+n}{2} (m - \mu^*)' r (m - \mu^*) \right\} |r|^{(\alpha+n-k-1)/2}$$

$$\times \exp \left\{ -\frac{1}{2} (\text{tr} S r + \text{tr} \left[\frac{n\nu}{\nu+n} (\mu - \bar{x}) (\mu - \bar{x})' r + \text{tr} Z \right]) \right\} \quad +.$$

Capítulo 4. Tomä de Dec-isiones.

En este capítulo se presentan dos problemas de decisión en el caso de contar con un modelo de regresión: predicción el primero, calibración el segundo; la idea central está basada en el trabajo de Lindley (1968).

En este artículo se dice que para usar en algún problema el modelado matemático son necesarias dos cosas:

- a) que la estructura del problema esté bien definida,
- b) que se tenga claridad en el uso del modelo seleccionado.

Del punto (a) se supondrá inicialmente que ya se realizó un experimento de regresión, que junto con el conocimiento inicial del experimentador (H_0) conformará el estado de información actual (H); del punto (b) se mostrará que dependiendo del uso del modelo el análisis se modifica, y que dado ésto, el uso de la estructura de teoría de decisiones es adecuado.

4.1. Predicción.

Suponga que en H se tiene un experimento de regresión de tal forma que: una columna de la matriz de diseño es constante (regresión homogénea), denote por x_1 la variable asociada a tal columna. Además suponga que las suposiciones estándar del modelo se cumplen.

Ahora, se considera otra observación en las variables independientes denotada por x ($x \in \mathbb{R}^r$), y la variable a predecir Y .

Sobre la estructura probabilística del problema se hacen las dos suposiciones siguientes:

1. Si β es el vector de parámetros de la regresión ($\beta \in \mathbb{R}^r$) y x es el vector de observaciones en las variables independientes consideradas en H

$$p(\beta, x | H) \approx p(\beta | H) p(x | H)$$

es decir, que son independientes. Esencialmente ésta es una suposición sobre H , como se verá más adelante.

2. Y es una variable aleatoria en \mathbb{R} con densidad $p(y|\beta, x, H)$ que no depende de H , es decir, $p(y|\beta, x, H) = p(y|\beta, x)$.

Además, esta densidad tiene las propiedades siguientes:

$$E[Y|\beta, x] = \beta'x$$

$$V[Y|\beta, x] = \sigma^2 \quad \sigma^2 > 0 \quad \text{conocida (*).}$$

Esta suposición establece que dados β y x la regresión es conocida; con esto no habrá cuestionamientos sobre las suposiciones estándar del modelo.

El problema: ¿qué variables de las r originales hay que observar para hacer una predicción? Denote por I el subconjunto de índices (s) de las variables sí observadas y por J el complemento ($r-s$); si x denota a la observación en las r variables entonces x_I denota a las variables observadas y x_J su complemento.

Sobre la estructura de decisión supóngase:

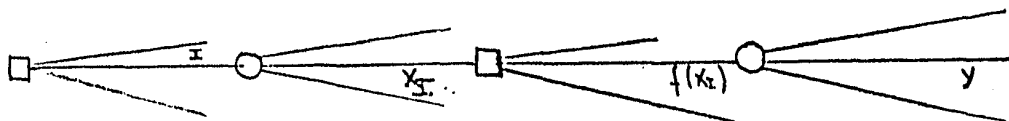
3. El espacio de decisiones es $\{d | d = (I, f(\cdot))\}$ donde $f(\cdot)$ es la función predictora de Y .

4. La función de pérdidas es

$$\{y - f(x_I)\}^2 + C_I \quad C_I > 0$$

con C_I el costo de observar $x_i, i \in I$.

Así, el árbol de decisión es:



es decir, que una vez seleccionado I se observa x_I y se predice Y con $f(x_I)$; con el valor verdadero de Y se introduce el costo de la predicción.

La solución bayesiana es minimizar la pérdida esperada en los nodos cuadrados y tomar esperanzas en los nodos circulares, procediendo de izquierda a derecha. Así, el problema es encontrar I tal que sea mínimo el valor de

$$\sigma^2 + \text{tr}[V[\beta]V[x]] + E[x]'V[\beta]E[x] + E[\beta]'V[x_J]E[\beta] + C_I$$

que equivale a minimizar sobre β $E[\beta]'V[x_J]E[\beta] + C_I$
 con $V[x_J] = E\left\{\left[x - E[x|x_I]\right]\left[x - E[x|x_I]\right]'\right\}$. Por otra parte el mejor predictor de Y es $E[\beta]'E[x|x_I]$.

Para hacer más manejables las expresiones anteriores se introducen las siguientes suposiciones:

5. La distribución $p(x|H)$ tiene regresiones lineales (homogéneas), es decir, $E[x_J|x_I] = B_{JI}x_I$ con B_{JI} matriz $(r-s) \times s$.

Esta suposición permite expresar $V[x_J]$ en términos de $V[x]$; se puede mostrar

$$\text{que } V[x_J] = V_{JJ} - B_{JI}V_{IJ}$$

$$\text{donde } V[x] = \begin{bmatrix} V_{II} & V_{IJ} \\ V_{JI} & V_{JJ} \end{bmatrix}$$

Si V_{II} tiene inversa

$$V[x_J] = V_{JJ} - V_{JI}V_{II}^{-1}V_{IJ}$$

Entonces la expresión a minimizar es

$$E[\beta_J]'(V_{JJ} - V_{JI}V_{II}^{-1}V_{IJ})E[\beta_J] + C_I.$$

Las siguientes son suposiciones sobre H .

6. H contiene n observaciones independientes Y_1, \dots, Y_n en los valores x_i'

($i=1, \dots, n$), y la densidad $p(y_i|\beta, x_i)$ es Normal, donde $x_i' = (x_{i1}, \dots, x_{ir})'$.

Con ésto, $E[Y_i|\beta, x_i] = x_i\beta$ $V[Y_i|x_i, \beta] = \sigma^2 I_n$.

Además se supondrá que las x_i' ($i=1, \dots, n$) ocurrieron al azar y son tales que

7. Las x_i' son variables aleatorias independientes idénticamente distribuidas Normal multivariada. La nueva observación se considera independiente de las anteriores pero proveniente de la misma población. Los parámetros de esta distribución se suponen además independientes de β .

Estas suposiciones significan que H contiene entre otras cosas un experimento aleatorio de regresión.

La siguiente suposición es sobre el conocimiento inicial.

8. Dado H_0 , es decir la información inicial antes de H , β se distribuye uniforme sobre \mathbb{R}^r ; la distribución del vector de medias de la distribución Normal de x también es uniforme en \mathbb{R}^r ; y por último, la distribución de la matriz de varianzas y covarianzas de x es inversamente proporcional a su determinante, siendo todas estas distribuciones independientes.

Así, combinando 6 y 8 (Lindley, 1965, 8.3):

$$E[\beta | H] = (X'X)^{-1} X'Y$$

$$V[\beta | H] = (X'X)^{-1} \sigma^2.$$

Ahora, $p(x|H)$ es (Ando y Kaufman, 1965) una distribución t multivariada tal que

$$E[x | H] = \sum_{i=1}^m x_i / n$$

$$V[x | H] = W W' \frac{(m+1)}{m(m-3)} \quad \text{con } W'W = (x_{ij} - x_{.j})^2_{i,j}$$

Lindley trabaja con $V[x|H] = W W' n^{-1}$ y además señala que la suposición 5 es satisfecha por esta distribución de x . Así se tiene el siguiente resultado:

Teorema. Bajo las suposiciones 1-8 anteriormente señaladas, la solución óptima en predicción es escoger I tal que

$$\min_I \{ R(J:I) n^{-1} + C_I \}$$

y predecir Y con $f(x_I) = E[\beta]' E[x | x_I]$

donde $R(J:I) n^{-1} = E[\beta_J]' V[x_J] E[\beta_J]$.

$$V[x_J] = V_{JJ} - V_{JI} V_{II}^{-1} V_{IJ}.$$

Puede demostrarse que $R(J:I) n^{-1}$ es la diferencia en suma de cuadrados de residuales de ajustar con los elementos de $I(R_I)$, menos la correspondiente al ajuste con todas las variables (R) , i.e.

$$R(J:I) n^{-1} = R_I - R.$$

La pérdida esperada es, bajo este procedimiento (Lindley, 1968)

$$\sigma^2 \left(1 + \frac{I}{n}\right) + \min_I \{R(J:I)\bar{m}^{-1} + C_I\}.$$

Este procedimiento es comparando con dos métodos clásicos. El primero de ellos consiste en seleccionar I de tal manera que las variables restantes x_j no contribuyan significativamente a la regresión. La estadística de prueba es

$$\frac{R(J:I)}{\sigma^2} \sim \chi^2_{(r-s)}$$

El procedimiento de Lindley usa prácticamente la misma estadística; la diferencia radica en lo que sirve de comparación, ya que el primer método compara con χ^2_{α} donde χ^2 es el cuantil de orden alfa de la distribución Ji-cuadrada con $r-s$ g.l., mientras que el segundo con nc_I .

El segundo método clásico es de la Cs de Mallows, cuya estadística es

$$C_S \sigma^2 \bar{m}^{-1} = R(J:I)\bar{m}^{-1} + 2s\sigma^2 \bar{m}^{-1} + R\bar{m}^{-1} - \sigma^2$$

Si el costo para cada variable es igual a $2\sigma^2 \bar{m}^{-1}$, se tiene

$$C_S \sigma^2 \bar{m}^{-1} = \{R(J:I)\bar{m}^{-1} + C_I\} + R\bar{m}^{-1} - \sigma^2$$

y entonces este procedimiento y el de Lindley son equivalentes salvo transformaciones lineales.

4.2. Calibración.

Dado Y en un nivel Y_0 predeterminado se requiere seleccionar I y x_I que cumplan con $Y_0 = f(x_I)$.

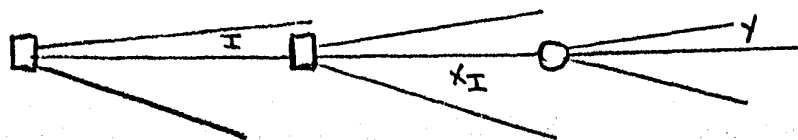
Se retienen las suposiciones 1 y 2 de 4.1 y además:

3'. El espacio de decisiones es $\{d \mid d = (I, x_I)\}$.

4'. La función de pérdidas es $(Y - Y_0)^2 + c(x_I)$, con $c(x_I) \geq 0$.

Nótese que el costo $c(x_I)$ no sólo depende de I sino también de los valores que toman las variables seleccionadas.

En este caso el árbol de decisión queda representado de la siguiente manera



Primero, se decide en que variables se va a calibrar y después en que valores; finalmente con Y da la estructura de pérdidas.

El problema es entonces, procediendo en forma análoga a 4.1, minimizar la siguiente expresión
$$\sigma^2 + \text{Tr} \{ V(\beta) V(x_j | x_I) \} + E[x | x_I]' V(\beta) E[x | x_I] + E[\beta]' V(x_j | x_I) E[\beta] + \{ E[\beta]' E[x | x_I] - y_0 \}^2 + C(x_I)$$

Para poder trabajar esta expresión, se supone que $V(x_j | x_I) = V(x_j)$, es decir que no depende de x_I , y además:

5'. La distribución $p(x|H)$ tiene regresiones lineales homocedásticas, es decir

$$E[x | x_I] = A_I x_I \quad A_I \text{ rxs.}$$

El problema es minimizar la siguiente expresión:

$$\min_I \sigma^2 + \text{Tr} [V(\beta) V(x_j)] + x_I' A_I' V(\beta) A_I x_I + E[\beta] V(x_j) E[\beta] + \{ E[\beta] A_I x_I - y_0 \}^2 + C(x_I).$$

Tomando sólo los términos que involucran I, x_I se tiene la forma cuadrática

$$x_I' A_I' M(\beta) A_I x_I - 2 y_0 E[\beta]' A_I x_I + C(x_I)$$

con $M(\beta) = V(\beta) + E[\beta] E[\beta]'$.

9. La función de costo $c(x_I)$ es cuadrática en x_I , es decir,

$$C(x_I) = x_I' C_I x_I + 2 C_I' x_I + C_I$$

Así, el problema es minimizar la siguiente expresión:

$$\min_I y_0^2 + x_I' (A_I' M(\beta) A_I + C_I) - 2 (y_0 - E[\beta]' A_I - C_I') x_I + C_I$$

El mínimo de esta forma cuadrática no es fácil de encontrar dado que x_I se ha considerado constante. para simplificar supóngase que

$$E[x_i | H] = 0 \quad i > 1 \quad E[y | H] = 0.$$

De esta forma, puede demostrarse que $E[\beta | H]$ y A_I es tal que su primera columna y su primer renglón son ceros salvo la entrada (1,1) con valor uno.

Adicionalmente,

10. $\text{COV}(\beta_i, \beta_j)$ es cero, $i, j > 1$.

Entonces, el problema es minimizar

$$\sigma^2(1+n^{-1}) + \frac{\sigma^2}{tn} + R(I:I)m^{-1} + C_I + \eta_0^2 / ((L + R(I:O)) / \sigma^2)$$

donde $t=r-s$, $R(I:O)$ es la reducción en suma de cuadrados debida a las variables en I .

4.3 Otros Problemas.

En el Capítulo dos se mencionó que el proceso para probar hipótesis estadísticas puede considerarse como una toma de decisión ; en este sentido, se involucran dos decisiones, (1) aceptar H_1 y (2) rechazar H_1 en favor de H_2 . Obviamente existe la posibilidad de cometer errores; las pérdidas asociadas a éstos debe ser cuantificadas y así determinar la región de rechazo para después minizar la pérdida esperada. Si $L(1)$ denota la pérdida al rechazar H_1 incorrectamente, $L(2)$ la correspondiente a H_2 y se supone que las pérdidas son cero cuando la decisión es la correcta, las pérdidas esperadas son:

para H_1 : $L(2)P(H_2)$; para H_2 : $L(1)P(H_1)$. Entonces se toma H_1 si y sólo si $L(2)P(H_2) < L(1)P(H_1)$.

Por otra parte, como se puede observar en el 4.1 y 4.2, seleccionar variables involucra seleccionar un modelo anidado en uno más general. En la literatura respecto a selección de modelos, se pueden encontrar dos posturas generales: la llamada "Bayesiana" y la basada en teoría de Información.

El problema que atacan es: dados J modelos representados por las densidades $f_1(\cdot | \theta_1), \dots, f_J(\cdot | \theta_J)$ para la explicación de un vector aleatorio y dadas n observaciones ¿cuál es el mejor modelo?.

La postura del criterio "Bayesiano" es como sigue:

se supone un modelo compuesto de

- 1) $f_k(\cdot | \theta_k)$ distribución de los datos.
- 2) $p_k(\theta_k)$ distribución inicial de θ_k (el vector de parámetros del modelo k).

3) $\pi(k)$ la distribución inicial para el modelo k .

$$k=1, \dots, J$$

Entonces $p(\theta_k, k | x) = p_k(\theta_k | x) \pi(k | x)$

donde $p_k(\theta_k | x) = \frac{f_k(x | \theta_k) p_k(\theta_k)}{f(x | k)}$

$$\pi(k | x) = \frac{f(x | k) \pi(k)}{f(x)}, \quad f(x | k) = \int f_k(x | \theta_k) p_k(\theta_k) d\theta_k,$$

$$f(x) = \sum_{k=1}^J f(x | k) \pi(k).$$

Sin embargo la especificación de $p_k(\theta_k)$ no es siempre un problema fácil de resolver; el uso de distribuciones de referencia puede llevar a paradojas (Akaike, 1981). Además en el caso de modelos que tengan distintas dimensiones en sus vectores de parámetros plantea problemas de interpretación de las probabilidades finales para cada modelo (Atkinson, 1978).

Por otra parte, la selección de modelos utilizando teoría de información, se basa en la medida de la información de Kullback-Liebler (1951) que resulta ser el negativo de la entropía. En este caso suponga que $f(\cdot | \theta^0)$ es el modelo verdadero de Y y que $f(\cdot | \theta)$ es una aproximación a $f(\cdot | \theta^0)$ donde θ está sujeto a ciertas restricciones ajenas a θ^0 .

Para discriminar entre los dos modelos usando n observaciones independientes Y_1, \dots, Y_n se utiliza

$$\ln[\theta^0, \theta] = E \left[\sum_{i=1}^n \left[\log \frac{f(Y_i | \theta^0)}{f(Y_i | \theta)} \right] \right].$$

es decir, el valor esperado del logaritmo del cociente de verosimilitudes.

La suposición inicial de contar con un modelo verdadero plantea dificultades.

Otro enfoque sumamente interesante es el de Bernardo (1984), en el que se plantea el problema de selección de un modelo como un proceso de toma de decisiones secuenciales:

i) selección de un modelo probabilístico en Y ,

ii) diseño de la matriz de datos,

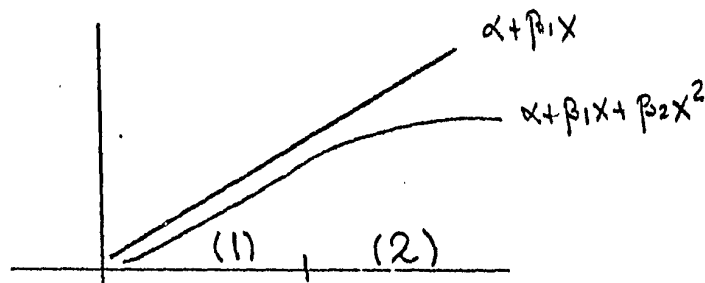
iii) inferencia (obtención de las distribuciones finales de interés),

iv) selección de variables relevantes al objetivo.

Este enfoque no ha sido explorado por el autor de este trabajo; curiosamente parece reunir elementos de teoría de información como del enfoque estrictamente Bayesiano.

Por último, durante la revisión bibliográfica se encontraron trabajos que tratan sobre la discriminación de modelos, es decir, métodos que intentan determinar la región muestral en donde la diferencia entre varios modelos (los relevantes al caso), sea más marcada. A los interesados se remite a Atkinson y Cox (1974), Atkinson y Fedorov (1975a, 1975b), Fedorov (1972).

Por ejemplo, si se tienen dos modelos que posiblemente explican un fenómeno: $E(Y|X) = \alpha + \beta_1 X$, $E(Y|X) = \alpha + \beta_1 X + \beta_2 X^2$; la discriminación intenta encontrar los puntos x donde la diferencia entre estos dos modelos sea evidente; viendo la gráfica siguiente, la región 1 no discriminaría bien entre los dos modelos mientras que la 2 sí lo haría.



Sin embargo, no fueron estudiadas a profundidad estas referencias por lo que se invita al lector interesado a profundizar en ellas.

(*) Lindley (1968) menciona que él cree que cuando es desconocida no habrá mayor problema sustituyendo Normales por distribuciones t .

Capítulo 5. Comentarios Finales.

El objetivo de este trabajo, a fin de cuentas resultó ser el proveer de una introducción a algunos aspectos del análisis de regresión Bayesiano, a los interesados en éste tema. Se espera que cumpla con tal objetivo.

Por otra parte, hablando en términos generales, es de notar que la estadística Bayesiana plantea una estructura general para resolver problemas estadísticos; además, la filosofía es preventiva más que correctiva. Esto es un cambio de enfoque muy importante; por ejemplo, si no hay Normalidad se incorpora al análisis la desviación sin intentar hacer transformaciones o utilizar resultados asintóticos.

Sin embargo, a nivel práctico, el autor encontró dificultades para realizar un trabajo más consistente a falta de programas de cómputo adecuados (integración numérica en varias variables, optimización); este aspecto, para tener mayor seriedad, parece fundamental desarrollarlo.

Por la parte inferencial, parece que es necesario trabajar en el procedimiento de prueba de Hipótesis; en ninguna de las soluciones planteadas se tiene plena satisfacción.

En cuanto al procedimiento de selección de variables propuesto por Lindley, es necesario investigar sobre un procedimiento operativo en la búsqueda de I. Como último comentario, debe mencionarse que la motivación inicial de estudiar este tema fué consecuencia de revisar los textos de Box y Tiao (1973) Zellner (1971) y Lindley (1965). En estos libros, el análisis de regresión es tratado sólo inferencialmente, es decir, sin considerar una estructura de decisión. Una cualidad de la estadística Bayesiana, a diferencia de la Clásica, es que se basa en una metodología única para tomar decisiones en ambiente de incertidumbre. Partiendo de éste hecho, si se está dispuesto a aceptar tal posición, el análisis de regresión visto bayesianamente debería estar

inscrito en la metodología general, es decir, deberían especificarse los elementos de un problema de decisión. Sin embargo, en la literatura revisada no se encontró lo deseado y ésto llevar a cuestionar si realmente el análisis de regresión es una técnica ad-hoc dentro de la estructura Bayesiana. El plantear la regresión como un problema (o varios) de toma de decisiones no es sencillo; se espera que en un trabajo futuro se desarrolle; definitivamente, la estadística Bayesiana implica mayores dificultades en su análisis, sin embargo no debe negarse como posibilidad.

Referencias:

- 1) Ando, A. y Kaufman, G.M. (1965). Bayesian analysis of the independent multinomial process-neither mean and precision known. JASA, 60, 347-358.
- 2) Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. B.N. Petrov and Csáki, eds., Proceedings of the 2nd international symposium on information theory (Akademiai Kiadó, Budapest) 267-281.
- 3) Akaike, H., (1974). A new look at the statistical model identification, IEEE Transactions on Automatic Control AC-19, 716-723.
- 4) Atkinson, A.C. (1978). Posterior probabilities for choosing a regression model. Biometrika 65, 39-48.
- 5) Atkinson, A.C. (1981). Likelihood ratio, posterior odds and information criteria May Vol. 16. Journal of Econometrics. North-Holland.
- 6) Atkinson, A.C. y D.R. Cox (1974). Planning experiments for discriminating between models (with discussion). JRSS B 36, 321-348.
- 7) Atkinson, A.C. y V.V. Fedorov (1975a). The Design of experiments for discriminating between two rival models. Biometrika 62, 57-70.
- 8) Atkinson, A.C. y V.V. Fedorov (1975b). Optimal Design: Experiments for discriminating between several models. Biometrika 62, 289-303.
- 9) Bernardo, J.M. (1979). Reference Posterior distribution for Bayesian Inference. JRSS B , 113, 147.
- 10) Bernardo, J.M. (1981). Bioestadística, una perspectiva bayesiana. Vicens Vives S.A.
- 11) Bernardo, J.M. (1984). Notas del curso impartido en la Facultad de Ciencias (laboratorio de Estadística) de la UNAM.
- 12) Box, G.E.P. G.C. Tiao (1973). Bayesian Inference in Statistical Analysis. Addison Wesley.

- 13) Chow, G. C. (1981). A comparison of the Information and Posterior criterion for model selection. May Vol. 16 Journal of Econometrics, 21-33, North Holland
- 14) De Groot, M. H. (1970). Optimal Statistical Decisions. McGraw Hill.
- 15) Fedorov, V. V. (1972). Theory of Optimal Experiments. Academic Press.
- 16) Geisser, S. y Cornfield, J. (1963). Posterior Distributions for Multivariate Normal Parameters. JRSS B25, 368-376.
- 17) Jeffreys, H. Theory of Probability (3rd ed.) Oxford: Clarendon 1961 & 1966.
- 18) Kendall, M. G. y A. Stuart (1977). The Advanced Theory of Statistics, Vol. 1 New York: Hafner Press.
- 19) Kullback, S. y R. A. Liebler (1951). Information and sufficiency. Ann. Math. Stat., Vol. 2, pp 79-86.
- 20) Leamer, E. (1978). Specifications Searches. Wiley New York.
- 21) Lindley, D. V. (1965). An Introduction to Probability and Statistics from a Bayesian Viewpoint (2 vol.). Cambridge: Cambridge University Press.
- 22) Lindley, D. V. (1968). The choice of variables in Multiple Regression. JRSS B 30, 31-66 (with discussion).
- 23) Lindley, D. V. y A. F. M. Smith (1972). Bayes estimates for the linear model (with discussion). JRSS B 34, 1-41.
- 24) Oman, S. D. (1978). A bayesian comparison of some estimators used with multicollinear data. Comm. in Stat. Theory and Methods A7(6), 517-534.
- 25) Raiffa, H. y R. Schlaiffer (1961). Applied Statistical Decision Theory. Cambridge, Mass: the MIT Press.
- 26) Smith, A. F. F. (1973). A general bayesian linear model. JRSS 62, 407-416.
- 27) Villegas, C. (1969). On the a priori distribution of the covariance matrix. The ANN. of Math. Stat., Vol. 40, No. 3. 1098-1099.

- 28) Winckler, R.L. (1972). Introduction to Bayesian Inference and Decision. New York: Holt, Reinhart & Winston.
- 29) Zellner, A. (1971). An introduction to Bayesian Inference in Econometrics. New York: Wiley.
- 29) Zellner, A. (1976). Bayesian and Non-Bayesian Analysis of the regression Model with Multivariate Student-t Errors terms. JASA 71 No. 354, Theory and Methods, 400-405.