INFERÊNCIA BAYESIANA

MODELOS DISCRETOS E CONTINUOS CON y SIN REGRESSORES

H. Bolfarine, USP- Brasil

Brasil, PUCP - Peru

Septiembre, 2012

El principal interes de este minicurso es la presentacion de metodologias en modelos que pueden ser empleados para el tratamiento de datos dicretos y continuos.

Dos de los modelos que mas se usan los modelos normal y binomial.

El modelo binomial es adecuado para datos del tipo de conteo, como el numero de aciertos de un estudiante en un test, el numero de compradores cierto equipo electronico, el numero de caras en n juegos de una moneda, etc.

El modelo normal es adecuado para situaciones de mediciones continuas como altura, nivel de colesterol, pression sistolica, etc.

Comenzamos con la descripcion de elementos basicos de la inferencia bayesiana y maximo verossimil.

Temario

1. Principales conceptos del modelamiento estadistico
2. Modelos de regresion simple
3. Modelos de regresion general
4. Otros modelos de regresion
5. Otras distribuciones para la variable respuesta

Referencia: Inferencia Bayesiana. Daniel Paulino.

1. PRINCIPALES CONCEPTOS DEL MODELAMIENTO ESTADISTICO

Modelo probabilistico: Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ una muestra aleatoria (independiente e identicamente distribuida: i.i.d.) de $X \sim f(x|\theta)$, ddonde $\boldsymbol{X} \in \mathcal{X}$, espacio muestral, $\theta \in \Theta$: espacio parametrico.

Function de densidad (o de probabilidad): Denotada por $f(x|\theta)$, es el modelo probabilistico relacionando el experimento (X) con la cantidad de interes θ. Describe la incertitumbre del estadistico al intentar relacionar θ y el mundo real (X). Note que θ puede ser un vector de parametros. Se puede decir que el estadistico hace espionaje de la naturaleza (X) para obtener informacion acerca de θ.

Funcion de verosimilitud: Sea $\boldsymbol{X} = (X_1, \ldots, X_n)'$ las observaciones. La funcion

$$L(\theta) = f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta),$$

que contiene toda la informacion que $\boldsymbol{X}$ tiene acerca de θ se llama funcion de verosimilitud. Notese que X_i puede ser un vector. Podese ter, por ejemplo, $X_i = (peso, altura)$.

1.1 Estimacion de maxima verosimilitud

Estimador de maxima verosimilitud (EMV): es conocido como el estimador que maximiza la funcion de verosimilitud o equivalientemente para el caso en que el espacio parametrico no depende de θ se puede maximizar el logaritimo de la funcion de verosimilitud.

En el caso de la $U(0, \theta)$, por ejemplo, que no satisface las conditiones de regularidad, se debe proceder a la maximizacion por inspecion de la verosimilitud.

El parametro θ puede se unidimensional o multidimensional.

Una familia importante de distribuciones ddonde se puede tener la situacion en que la condiciones de regularidad se consideran satisfechas es la famila exponencial, ddonde como casos particulares tenemos las importantes distribuiciones: normal, binomial, exponencial, Poisson.

Importantes funciones podem ser obtenidas a partir de la funcion de verosimilitud $L(\theta)$.

Funcion escore:

$$\Lambda(\theta) = \frac{\partial l(\theta)}{\partial\theta} = \sum_{i=1}^{n} \frac{\partial \log f(x_i|\theta)}{\partial\theta}.$$

En casos regulares, la solucion de

$$\Lambda(\theta) = 0$$

produce el EMV. Es decir, si $\hat{\theta}$ es el EMV, para situaciones regulares, se tiene que $\Lambda(\hat{\theta}) = 0$. Se puede mostrar que

$$E[\Lambda(\theta)] = 0,$$

es decir, el valor esperado del escore es nulo.

Informacion esperada (o de Fisher): Para el caso uniparametrico

$$I_F(\theta) = E[-\frac{\partial\Lambda(\theta)}{\partial\theta}] = E\{[\Lambda(\theta)]^2\}.$$

Para la situacion multiparametrica se habla de la matrix de informacion de Fisher.

Informacion observada: es solo la derivada segunda de la log-verosimilitud o derivada primera del escore evaluado en el EMV $\hat{\theta}$, es decir

$$J(\hat{\theta}) = \frac{\partial \Lambda(\theta)}{\partial \theta}\Big|_{\hat{\theta}}.$$

Distribucion asintotica del EMV: El EMV es consistente y asintoticamente normal, esto es,

$$\hat{\theta} \overset{a}{\sim} N(0, I_F(\theta)^{-1}).$$

Entonces un intervalo de confianza aproximado para θ con C.C. $\gamma = 1 - \alpha$ es dado por

$$\hat{\theta} \mp z_\alpha \sqrt{I_F(\hat{\theta})^{-1}},$$

ddonde se calcula z_α en la tabla de la N(0,1) como el punto que deja $1 - \alpha/2$ de area a la izquierda.

Algoritmo de estimacion: En muchas situaciones el EMV no tiene forma analitica (cerrada). Su derivacion tiene que ser hecha numericamente. Un algoritimo numerico puede ser implementado como sigue. Por expansion de Taylor se tiene:

$$0 = \Lambda(\hat{\theta}) = \Lambda(\theta) + \Lambda^*(\theta)(\hat{\theta} - \theta), \tag{1}$$

que se puede escribir

$$\hat{\theta} = \theta - \frac{\Lambda(\theta)}{\Lambda^*(\theta)}, \tag{2}$$

ddonde

$$\Lambda^*(\theta) = \frac{\partial \Lambda(\theta)}{\partial \theta}.$$

Dado asim un valor inicial θ, usando (2) se calcula un nuevo $\hat{\theta}$. Usando este nuevamente como valor inicial se calcula nuevamente el estimador y asi sucessivamente hasta la convergencia. Un algoritimo alternativo se obtiene reemplazando $\Lambda^*(\theta)$ por $I_F(\theta)$. Estos algoritimos son conocidos como de Newton-Raphson.

Teorema de Limite Central para la situacion i.i.d: Sea $X_1, \ldots, X_n$ muestra aleatoria (iid) de X con $E[X] = \mu$ y $Var[X] = \sigma^2 < \infty$.
Entonces para $\bar{X} = \sum_{i=1}^{n} X_i/n$ y $n \to \infty$,

$$\sqrt{n}\frac{(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0,1). \tag{3}$$

o equivalentemente

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathcal{D}} N(\mu, \sigma^2).$$

Notese que se tiene un TLC para el escore

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial log f(X_i|\theta)}{\partial\theta}) \xrightarrow{\mathcal{D}} N(0, I_F(\theta)), \tag{4}$$

porque

$$E[\frac{\partial log f(X_i|\theta)}{\partial\theta}] = 0,$$

y

$$Var[\frac{\partial log f(X_i|\theta)}{\partial\theta}] = E[\frac{\partial log f(X_i|\theta)}{\partial\theta}]^2 = I_F(\theta).$$

Por lo tanto, las variables $\partial log f(X_i|\theta)/\partial\theta$ son iid con media cero y variancia $I_F(\theta)$, que debe ser assumida finita.

Tambien por la ley de los grandes numeros (LGN)

$$\frac{1}{n}[\Lambda^*(\theta)] = \frac{1}{n}[\frac{\partial \Lambda(\theta)}{\partial \theta}] = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 log f(X_i|\theta)}{\partial \theta^2} \xrightarrow{P} -I_F(\theta). \tag{5}$$

Se emplea arriba la Ley debil de grandes numeros mas simple que es del caso iid con varianza finita. Note que en (5) se tiene una media muestral.

De las identidades (1) y (2) se puede escribir

$$\sqrt{n}(\hat{\theta}-\theta) = \sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial log f(X_i|\theta)}{\partial\theta})[-\frac{\partial\Lambda(\theta)}{\partial\theta}]^{-1}.$$

De (4), (5) y (6) se tiene entonces que

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{D} N(0, I_F^{-1}(\theta)).$$

Este resultado se extiende para el caso multiparametrico ddonde $I_F(\theta)$ es la matriz de informacion de Fisher que se puede escribir como

$$I_F(\theta) = E[\frac{\partial\Lambda(\theta)}{\partial\theta\partial\theta'}].$$

Es tambien relevante notar que la varianza asintotica de EMV es el limite inferior de la deseigualdad de la informacion y Por lo tanto no hay estiamdor consistente con menor varianza asintotica.

Ejemplo 1. Suponga

$$X_1, \ldots, X_n \qquad iid \qquad X \sim N(\mu, \sigma^2).$$

Entonces $\theta = (\mu, \sigma^2)'$ de modo que $\theta \in \Theta = R\mathrm{x}R^+$. Note que el parametro es bidimensional para este modelo. en el caso σ^2 se puede tomar $\sigma^2 = 1$ sin perdida de generalidad. La funcion de verosimilitud:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = (\frac{1}{\sigma\sqrt{2\pi}})^n \mathrm{e}^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}.$$

La Funcion de log-verosimilitud (olvidando 2π):

$$l(\theta) \propto -\frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}.$$

La funcion escore es obtenida derivando la log verosimilitud respeto a μ y σ^2. En este caso es dada por la funcion bidimensional

$$\Lambda(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \mu) \\ -\frac{n}{\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^n (x_i - \mu)^2 \end{pmatrix}.$$

Igualando a cero la funcion escore se obtiene los estimadores EMV.
EMV: $\hat{\theta} = (\hat{\mu}, \hat{\sigma}_2)$ maximiza $L(\theta)$, con

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \qquad \text{y} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2.$$

Verifique que $\Lambda(\hat{\theta}) = (0,0)'$.

Diferenciando el escore una vez y tomando esperanza tenemos la matriz de informacion de Fisher

$$I_F(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

Diagonalidad de la informacion implica ortogonalidad (independencia assintotica de los EMV).
Se sabe que

$$\sqrt{n}(\bar{X} - \mu) \sim N(0, \sigma^2)$$

y que para $n \to \infty$,

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{D} N(0, 2\sigma^4),$$

que puede ser usado para construir IC y tests. El ultimo resultado se usa el TLC con la identidad

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2. \tag{6}$$

Se usa el TLC en la variable $Z_i = (X_i - \mu)^2$ que son iid.

Nota. Note que $\hat{\sigma}^2$ es sesgado para σ^2, siendo el estimator insesgado $S^2 = \frac{n\hat{\sigma}^2}{(n-1)}$.

Intervalos de confianza (I.C.) pueden ser construidos para μ y σ^2 basados en el hecho que

$$(7) \qquad \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t(n-1) \qquad \text{y} \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Por ejemplo, para a un nivel de confianza γ y $t_\gamma(n-1)$ el valor correspondiente de la tabla t con n-1 g.l., tenemos que el I.C. es dado por

$$\bar{X} \mp t_\gamma \sqrt{\frac{S^2}{n}}.$$

Para verificar los resultados en (7) se usa la identidad (6) y el hecho de que $Z \sim N(0,1)$ implica que $Z^2 \sim \chi^2_1$.

Aplicacion 1. Datos de Darwin.

Datos correspondientes al crescimiento de plantas bajos dos condiciones experimentales distintas con n=15 diferencias de pares produciendo los valores

$$49, -57, 8, 16, 6, 23, 28, 41, 14, 56, 24, 75, 60, -48, 29.$$

Notese que se tiene dos observaciones muy extremas.

Considerando el modelo normal, es dedir $X \sim N(\theta, \sigma^2)$, lleva a los estimadores puntuales $\bar{x} = 20,84$ y $S^2 = 39,76$.

Con $\gamma = 0,95$ y estimadores por intervalo $(0,39; 41,95)$ para μ, entonces $H_0 : \mu = 0$ (vs $H_0 : \mu \neq 0$) es rechazada al nivel de 5.%

Para σ^2 tenemos el 95% intervalo de confianza $(27,76; 59,72)$. Rechazamos $H_0 : \sigma = 1$ pues el I.C. no contiene el valor uno.

La normal bidimensional

Una extension possible para el modelo arriba es el caso bidimensional o seja $(X, Y)' \sim N_2(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ con densidad

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)}\{\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{(y-\mu_y)}{\sigma_y})\}}.$$

Para este modelo, $\theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$. Un caso particular interesante es la situacion curvada es decir, con $\mu_x = \mu_y = 0$ and $\sigma_x = \sigma_y = 1$ con $\theta = \rho$. En este caso, con una muestra de n observaciones,

$$\hat{\rho}_c = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{P} \rho.$$

Piense en un TLC para $\hat{\rho}_c$.

De acuerdo con el TLC, cuando $n \longrightarrow \infty$,

$$\sqrt{n}(\hat{\rho}_c - \rho) \longrightarrow N(0, Var[XY]),$$

con

$$Var[XY] = E[XY]^2 - (E[XY])^2 = E\{E[X^2Y^2|Y]\} - \rho^2 = (1 + \rho^2).$$

Como no caso binomial variância assintótica depende de ρ o que torna o intervalo bastante instavel. Alternativa - funcion Z de Fisher para estabilizar variância.

Para el caso geral de μ_x,... desconocidos, tenemos los estimadores

$$\hat{\mu}_x = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \hat{\sigma}_x^2 = S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

$$\hat{\rho}_P = \frac{S_{XY}}{S_X S_Y}.$$

Estos estimadores son obtenidos igualando las estadísticas suficientes a sus valores esperados ($E[T] = T$, Lehmann e Casella, 1998, pg 470) pues el espacio parametrico tiene interior no vacÃo. Para $n \to \infty$,

$$\sqrt{n}(S_x^2 - \sigma_x^2) = \sqrt{n}(\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{n} - \sigma^2) - \frac{1}{\sqrt{n}}(\sqrt{n}(\bar{X} - \mu))^2$$

$$\longrightarrow N(0, Var[X - \mu]^2) = N(0, 2\sigma_x^4)$$

Tenemos, conjuntamente, para $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} S_x^2 - \sigma^2 \\ S_y^2 - \sigma_y^2 \\ S_{xy} - \sigma_{xy} \end{pmatrix} \sim N_3(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2\sigma_x^4 & 2\rho^2\sigma_x^2\sigma_y^2 & 2\rho\sigma_x^3\sigma_y \\ & 2\sigma_y^4 & 2\rho\sigma_x\sigma_y^3 \\ & & (1+\rho^2)\sigma_x^2\sigma_y^2 \end{pmatrix}).$$

Para obtener estimadores bayesianos para ρ da normal bivariada se puede entrar diretamente com a verosimilitud que es un caso particular del modelo con errores en las variables de la tesis de Pedro. Tambesn se puede estimar, usando el mismo enfoque, el caso del modelo curvado $N_2(0, 0, 1, 1, \rho)$. Considerar tambem el algoritimo BFGS no R para este ultimo caso.

Un resultado importante es el metodo delta:

Se $\sqrt{n}(\bar{X} - \mu) \longrightarrow N(0, \sigma^2)$,

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \longrightarrow N(0, (g'(\mu))^2\sigma^2).$$

Ejemplo 2.
Suponga

$$X_1, \ldots, X_n \qquad iid \qquad X \sim Bernolli(\theta).$$

Entonces $\theta \in \Theta = (0, 1)$.

La funcion de verosimilitud:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}.$$

En este caso es mas facil maximizar la log-verosimilitud

$$l(\theta) \propto \sum_{i=1}^{n} x_i \log\theta + (n - \sum_{i=1}^{n} x_i)\log(1-\theta).$$

La funcion escore es dada por

$$\Lambda(\theta) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{(n - \sum_{i=1}^n x_i)}{(1-\theta)}$$

con informacion de Fisher

$$I_F(\theta) = \frac{1}{\theta(1-\theta)}.$$

EMV $\hat{\theta}$ que maximiza $L(\theta)$, que es obtenido de la equacion $\Lambda(\theta) = 0$ y es dado por

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i.$$

La informacion de Fisher que se obtiene de la funcion escore y es dada por

$$I_F(\theta) = E[\frac{X}{\theta^2} - \frac{(1-X)}{(1-\theta)^2}] = \frac{1}{\theta(1-\theta)}.$$

La variancia asintotica del estimador es la inversa de la informacion de Fisher/n que es tambien la variancia usual, esto es

$$Var[\bar{X}] = \frac{\theta(1-\theta)}{n}.$$

Es complicado la obtencion de IC para θ. Un IC para grandes muestras con nivel de confianza γ, basados en la aproximacion normal es dado por

$$\bar{X} \mp z_\gamma \sqrt{\frac{\bar{X}(1-\bar{X})}{n}},$$

ddonde z_α es obtido a partir da distribuicion normal como descrito arriba.

Es bastante conocido que el intervalo de confianza es bastante inestable para tamanos de muetra pequenos hay problemas con la probabilidad de cobertura del intervalo aproximado. El intervalo exacto debe ser obtenido en lla posteriori de la distribuction Beta arriba. Possible extension del modelo arriba es para el caso de la comparacion de dos proporciones θ_1 y θ_2. Se puede considerar, por ejemplo, inferencia para la diferenza $\theta_1 - \theta_2$.

Aplicacion 2.
Considere n=20 y la muestra $\boldsymbol{X} = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1)$. de departamentos con presupuestos correctamente formulados a Septiembre de 2010. Ddonde 0 significa presupuesto incorrectamente formulado.

Esto lleva al estimador y variancia estimada

$$\hat{\theta} = \bar{X} = \frac{6}{20} = 0,30, \qquad \hat{Var}[\hat{\theta}] = \frac{\hat{\theta}(1-\hat{\theta})}{n} = \frac{.21}{20} = 0,0105.$$

El Intervalo de confianza basada en el TLC es dado por

$$0,30 \mp 1,96\sqrt{0,0105}, \qquad (0,10; 0,50).$$

Dada la dificuldad de la construcion de IC en el modelo Binomial, vamos abordar el enfoque bayesiano, que envuelve la especificacion de una distribuction a priori y de una function de perdida bajo la cual el estimador es calculado.

La construcion del estimador involucra la minimizacion del riesgo de Bayes que depende de la distribuition a priori para el parametro.

ejemplo. Considere el modelo genestico com cuatro categorias, distribuídas de acuerdo con las Probabilidades:

$$p_1 = \frac{2+\theta}{4}, \quad p_2 = \frac{1-\theta}{4}, \quad p_3 = \frac{1-\theta}{4}, \quad p_4 = \frac{\theta}{4},$$

com $0 \leq \theta \leq 1$. Para una muestra de N animales, sea Y_i del tipo i, $i = 1, \ldots, 4$, com $\sum Y_i = N$. Considere priori $\theta \sim Beta(a, b)$. La distribiucion la posteriori para θ puede ser escrita como

$$\pi(\theta|\boldsymbol{x}) \propto (2+\theta)^{y_1}(1-\theta)^{y_2+y_3+b-1}\theta^{y_4+a-1}.$$

Para $n = 197$, $\boldsymbol{Y} = (125, 18, 20, 34)$, tenemos com $a = b = 1$, o estimador modal $\hat{\theta} = 0.6268$ con

$$\hat{\sigma}^2 = [-\frac{d^2 \log \pi(\theta|\boldsymbol{x})}{d\theta^2}]^{-1}_{\theta=\hat{\theta}} = 1/377.52 = 0.002649.$$

Princípio de verosimilitud

Importante propriedad del metodo de máxima verosimilitud:

El princípio de verosimilitud: Observaciones x e x' em $\mathcal{X}$ (escalares o vectores) son tales que

$$f(x|\theta) = K(x, x')f(x'|\theta), \quad todo \quad \theta,$$

ddonde K no depende de θ. Inferencias sobre θ a partir de x o x' deben ser las mismas, pues las verosimilitudes son las mismas.

Ejemplo. Para una moneda que presenta caras com probabilidad θ se quiere probar $H_0 : \theta = 1/2$ vs $H_1 : \theta > 1/2$. Experimentos relacionados com el lanzamiento de una moneda con probabilidad de caras igual a θ:
E_1: lanzar la moneda 12 veces
E_2: lanzar la moneda hasta que aparezcan tres caras.
Suponga que los dos experimentos llevam al valor $X = 9$, ddonde X: número de coronas observadas.

Para el caso de E_1, $X \sim Binomial(12, \theta)$, con p-valor

$$P_1 = P[X \geq 9|\theta = 1/2] = \binom{12}{9}(\frac{1}{2})^{12} + \cdots + \binom{12}{12}(\frac{1}{2})^{12} = 0.075.$$

Para el caso de E_2, $X \sim BN(3, \theta)$,

$$P_2 = P[X \geq 9|\theta = 1/2] = \binom{11}{9}(\frac{1}{2})^{12} + \binom{12}{10}(\frac{1}{2})^{13} + \binom{13}{11}(\frac{1}{2})^{14} + ... = 0.0325.$$

Para $\alpha = 5\%$, se rechaza para E_2 mas no para E_1. La misma cantidad de informacion lleva a decisiones diferentes.

1.2 Estimacion bayesiana

Tenemos entonces como elementos del procedimento Bayesiano:

1. La funcion de verosimilitud:

$$f(\boldsymbol{x}|\theta) = L(\theta) = \prod_{i=1}^{n} f(x_i|\theta),$$

que conecta el estado de la naturaleza (θ) con los datos (X).

2.Distribucion a priori:
$\pi(\theta)$: pdf (o pf) en Θ, expresa distribucion que la naturaleza esta usando para producir θ. Puede tambien representar conocimiento subjectivo estadistico tiene acerca de θ.

3.Funcion de perdida:
$l(\theta, d) : \Theta \mathrm{x} \mathcal{D} \rightarrow R^+$, ddonde $d \in \mathcal{D}$, denotando el espacio de decisiones. Expresa perdida en la que se incurre al usar el estadistico cuando la decision d es tomada y θ es el verdadero estado de la naturaleza.

4. Funcion de riesgo:

$$R(\theta, d) = E[l(\theta, d(\boldsymbol{X}))] = \int_{\mathcal{X}} l(\theta, d(x)) f(\boldsymbol{x}|\theta) d\boldsymbol{x}.$$

De manera general no hay procedimiento com riesgo uniformemente menor. Esto es usualmente el caso de comparacion de dos funciones de decision, una es mejor (menor riesgo) para algunos valores de θ e otra para otros valores de θ. Esto es, graficos de las funciones de riesgo se cruzan.

Ejemplo: Considere X_1, X_2, X_3 iid $N(\theta, 1)$. Con los siguientes estimadores

$$d_1 = \frac{1}{3}(X_1 + X_2 + X_3) \qquad \text{y} \qquad d_2 = \frac{2}{6}X_1 + \frac{1}{6}X_2 + \frac{3}{6}X_3,$$

.

En este caso, d_2 es dominado por d_1 bajo perdida cuadratica.

ejemplo.

$\mathcal{A} = R$,
$L(\theta, d) = (d - \theta)^2$,
$X|\theta \sim N(\theta, 1)$
$\mathcal{D} = \{d_k; \quad d_k(X) = kX\}$
Por lo tanto,

$$R(d_k, \theta) = E[kX - \theta]^2 = \theta^2(1 - k)^2 + k^2.$$

Note que
1) $d_1(X)$ domina $d_k(X)$ para $K > 1$.
2) Para $0 < k < 1$ riesgo se cruzam (ninguno mejor uniformemente)
3) $\sup_\theta R(d_k, \theta) = \sup_\theta\{\theta^2(1-k)^2 + k^2\}$
$= 1, \quad k = 1,$
$= \infty, \quad k \neq 1.$
Como $\sup_\theta R(d_1, \theta) = 1$, $d_1(X) = X$ es minimax.

Usar procedimento minimax es una manera de escoger un procedimiento con propriedades óptimas. Con todo es muchas veces complicado obtener tal estimador. Una alternativa es colocar una distribucion de probabilidad en el espacio paramestrico que lleva entonces a los procedimentos Bayesianos.

5. Riesgo de Bayes para la densidad posterior para priori (o funcion de prob.) $\pi(\theta)$:

$$r_\pi(d) = E_\pi[R(\theta, d)] = \int_\pi [R(\theta, d)]\pi(\theta)d\theta$$
$$= \int_\Theta \{\int_\mathcal{X} l(\theta, d(\boldsymbol{x}))f(\boldsymbol{x}|\theta)d\boldsymbol{x}\}\pi(\theta)d\theta.$$

Estimador de Bayes (regla de Bayes):

$$d_\pi \qquad \text{minimizando} \qquad r_\pi(d).$$

Este problema de minimizacion no siempre tiene solution analitica. Un caso ddonde hay solucion es el caso de la perdida cuadratica.

Teorema 1. *Bajo perdida cuadratica,*

$$l(\theta, d) = (\theta - d)^2,$$

el procedimiento de Bayes es

$$d_\pi = E[\theta|\boldsymbol{X}],$$

esto es, la media la posteriori.

Prueba. Note que

$$f(\boldsymbol{x}|\theta)\pi(\theta) = f(\boldsymbol{x};\theta) = \pi(\theta|\boldsymbol{x})g(\boldsymbol{x}), \quad \text{o} \quad \pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{g(\boldsymbol{x})}$$

ddonde

$$(*) \qquad g(\boldsymbol{x}) = \int_\Theta f(\boldsymbol{x}|\theta)\pi(\theta)d\theta.$$

Se puede escribir entonces

$$r_\pi(d) = \int_\Theta [\int_\mathcal{X} l(\theta - d(\boldsymbol{x}))^2 f(\boldsymbol{x}|\theta) d\boldsymbol{x}] \pi(\theta) d\theta$$

$$= \int_\mathcal{X} [\int_\Theta (\theta - d(\boldsymbol{x}))^2 f(\boldsymbol{x}|\theta)] \pi(\theta) d\pi(\theta)$$

$$= \int_\mathcal{X} [\int_\Theta (\theta - d(\boldsymbol{x}))^2 \pi(\theta|\boldsymbol{x}) d\theta] g(\boldsymbol{x}) d\boldsymbol{x}.$$

Entonces, diferenciando $r_\pi(d)$ con respecto a θ, esto es,

$$\frac{\partial r_\pi(d)}{\partial d} = 0 \quad \to d_\pi(\boldsymbol{x}) = \int_\Theta \theta \pi(\theta|\boldsymbol{x}) d\theta = E[\theta|\boldsymbol{x}].$$

Tambiesn se puede notar que

$$E[l(\theta, d)|x] = E\{[(\theta - E(\theta|x)) - (d - E(\theta|x)]^2|x\}$$

$$= Var[\theta|x] + E[(d - E(\theta|x))^2],$$

lo que muestra que el riesgo es mínimo (riesgo de Bayes) cuando $d = E[\theta|x] = d_B(x)$, siendo el riesgo dado por $r(d_B|x) = Var[\theta|x]$. Resultado puede ser extendido para perdidas

$$l(d, \theta) = a(d - \theta)^2$$

es

$$l(d, \theta) = (d - \theta)'(d - \theta) = \sum_{i=1}^{p}(d_i - \theta_i)^2,$$

para el caso $\theta = (\theta_1, \ldots, \theta_p)'$, $d = (d_1, \ldots, d_p)'$.

Notas.

- En la mayoria de los casos no es simple calcular la media a posteriori. Se puede usar integracion numerica o re-muestreo.
- La situacion de otras funciones de perdida produce otros estimadores de Bayes. Por ejemplo la perdida del valor absoluto produce la mediana de lla posteriori como estimador de Bayes.
- Note tambien que el estimador es apenas un punto. Para mirar algo mas completo tenemos que trabajar con lla posteriori que combina informacion de los datos con la priori.
- El intervalo de maxima probabilidad la posteriori (IMPP) (o en ingles, HPD), es el intervalo de menor longitud y area $\gamma = 1 - \alpha$. Estos intervalos son dificiles de ser obtenidos y requieren procedimentos numericos.
- Otra alternativa es usar remuestreo como el muestreo de Gibbs. Con muestras de lla posteriori se puede usar intervalos empiricos. Los IMPP son tais que la posteriori avaliada nos extermos (θ_1, θ_2) coincidem es decir $\pi(\theta_1|x) = \pi(\theta_2|x)$.

Ejemplo 3. (caso particular Ejemplo 1) Suponga

$$X_1, \dots, X_n \quad iid \quad N(\theta, 1),$$

$$\theta \sim N(a, b),$$

resultando en que

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(x_i-\theta)^2}{2}} = (\frac{1}{2\pi})^{\frac{n}{2}} \mathrm{e}^{-\sum_{i=1}^{n} \frac{(x_i-\theta)^2}{2}}$$

y

$$\pi(\theta) = \frac{1}{\sqrt{2\pi b}} \mathrm{e}^{-\frac{(\theta-a)^2}{2b}}.$$

Note que b en la priori para θ expresa la confianza de que a es un bueno valor a priori para θ.

Pero $g(\boldsymbol{x})$ no depende de θ, entonces sigue que

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{g(\boldsymbol{x})} \propto f(\boldsymbol{x}|\theta)\pi(\theta)$$

$$= (\frac{1}{2\pi})^{\frac{n}{2}} \mathrm{e}^{-\sum_{i=1}^{n} \frac{(x_i-\theta)^2}{2}} \frac{1}{\sqrt{2\pi b}} \mathrm{e}^{-\frac{(\theta-a)^2}{2b}}$$

$$\propto \mathrm{e}^{-\frac{1}{2}(n+1/b)(\theta-\frac{\sum_{i=1}^{n} X_i + a/b}{n+1/b})^2}$$

resultando en que

$$\theta|\boldsymbol{x} \sim N(\frac{\sum_{i=1}^{n} X_i + a/b}{n+1/b}, \frac{1}{n+1/b}).$$

Entonces,

$$d_\pi = E[\theta|\boldsymbol{X}] = \frac{\sum_{i=1}^n X_i + a/b}{n + 1/b},$$

que es el estimador de Bayes, es una combinacion linear entre la media muestral y la informacion a priori, esto es, $c\bar{X}+(1-c)a$, donde $c = 1/(1+1/(b/n))$.

Note que $d_\pi \rightarrow \bar{X}$ quando $n \rightarrow \infty$. El mismo $b \rightarrow \infty$. Tipicamente, la situacion com b grande($b \rightarrow \infty$) es denominando la situacion noninformativa ($b \rightarrow \infty$ es la situacion de una priori vaga "flat prior").

En este caso el intervalo MPP de probabilidad $\gamma = 1 - \alpha$ (simetrico), es dado por

$$\frac{\sum_{i=1}^n X_i + a/b}{n + 1/b} \mp z_\gamma \sqrt{\frac{1}{n + 1/b}}.$$

$$\bar{x} \sim N(\mu, \sigma^2),$$
$$\mu | a, b \sim N(a, b^2)$$

Combinado los exponentes da priori y de lla posteriori tenemos en general que

$$\frac{(\mu - a)^2}{b^2} + \frac{(\mu - \bar{x})^2}{\sigma^2/n}$$
$$= \frac{(\mu - A)^2}{B^2} + \frac{(\bar{x} - a)^2}{\sigma^2/n},$$

ddonde

$$A = \frac{a/b^2 + n\bar{x}/\sigma^2}{1/b^2 + n/\sigma^2} = \frac{c}{1+c}a + \frac{1}{1+c}\bar{x},$$

$$B^2 = \frac{1}{1/b^2 + n/\sigma^2} = \frac{\sigma^2/n}{1+c}, \quad c = \frac{\sigma^2/n}{b^2}.$$

Identidad importante:

$$d_1(z-c_1)^2 + d_2(z-c_2)^2 = (d_1+d_2)(z-c)^2 + \frac{d_1 d_2}{d_1+d_2}(c_1-c_2)^2,$$

ddonde

$$c = \frac{d_1 c_1 + d_2 c_2}{d_1 + d_2}.$$

Aplicacion 3. Para una muestra de datos generados de la $N(0,1)$ con $n = 10$, se tiene
$-1.22, 0.61, -0.71, -0.62, -1.25, 0.95, -0.26, 0.23, 1.60, -0.17,$
llevando a $\sum_{i=1}^n X_i = 0.8$ y entonces al estimador y IMPP para $a = 0$ y $b = 1$,

$$\frac{\sum_{i=1}^n X_i + a/b}{n + 1/b} = \frac{0.8}{11} = 0.075 \qquad y \qquad 0.075 \mp 1.96\sqrt{1/11}.$$

Resultando en $[-0.52, 0.66]$, de modo que $H_0 : \theta = 0$ es rechazada al nivel de 5%.

Se puede muchas veces simplificar los resultados usando una estadistica suficiente, pues siendo $T(X)$ una estatistica suficiente, por critesrio da factorizacion

$$f(x|\theta) = g(t|\theta)h(x),$$

de modo que la poteriori puede ser escrita como

$$\pi(\theta, x) \propto f(x|\theta)\pi(\theta) \propto g(t|\theta)\pi(\theta),$$

de modo que

$$E[g(\theta)|x] = E[g(\theta)|t].$$

ejemplo 3. Continuacion.

$$X|\mu \sim N(\mu, 1)$$

$$\mu \sim N(\mu_0, \tau),$$

com τ conocido. tenemos que $\bar{X}$ es estatística suficiente com

$$\bar{X}|\mu \sim N(\mu, \frac{1}{n}), \quad \hat{\mu}_B = E[\mu|\bar{X}].$$

Distribucion conjunta $(\bar{x}, \mu)$ envuelve calcular

$$E[\mu] = \mu_0, \qquad Var[\mu] = \tau,$$

$$E[\bar{X}] = E\{E[\bar{X}|\mu]\} = E[\mu] = \mu_0.$$

$$Var[\bar{X}] = E[\bar{X} - \mu + \mu - \mu_0]^2 = E[1/n + (\mu - \mu_0)^2] = 1/n + \tau.$$

$$Cov[\bar{X}, \mu] = E\{E[(\bar{X} - \mu_0)(\mu - \mu_0)|\mu]\} = E[\mu - \mu_0]^2 = \tau.$$

entonces,

$$\begin{pmatrix} \bar{X} \\ \mu \end{pmatrix} \sim N_2(\begin{pmatrix} \mu_0 \\ \mu_0 \end{pmatrix}, \begin{pmatrix} \frac{1}{n} + \tau & \tau \\ \tau & \tau \end{pmatrix}).$$

De modo que

$$\mu|\bar{x} \sim N(\mu_0 + \frac{\tau}{1/n}(\bar{x} - \mu_0), \tau - \frac{\tau^2}{1/n + \tau}),$$

de modo que

$$\hat{\mu}_B = E[\mu|\bar{x}] = \mu_0 + \frac{\tau}{\tau + 1/n}(\bar{x} - \mu_0) = \frac{\tau}{1/n + \tau}\bar{x} + \frac{(1/n)\mu_0}{1/n + \tau},$$

de modo que $\hat{\mu}_B \to \mu_0 \ (\bar{x})$, quando $\tau_0 \to 0(\infty)$.

*e*jemplo mais geral. Suposiciones:

$$X|\theta \sim N_p(\theta, \boldsymbol{I}_p),$$

$$\theta \sim N_p(\mathbf{0}, \tau \boldsymbol{I}_p),$$

Estimador de Bayes: $\hat{\theta}_B = E[\theta|\boldsymbol{X}]$.

En este caso tenemos $\boldsymbol{X}' = (X_1, \ldots, X_p)$, $d(\boldsymbol{X}) = (d_1, \ldots, d_p)'$ e $\theta = (\theta_1, \ldots, \theta_p)'$.
Perda: $L(d, \theta) = \sum_{i=1}^{p} (d_i - \theta_i)^2$.
Tambien; $||\boldsymbol{X}||^2 = \sum_{i=1}^{p} X_i^2$.

Distribucion conjunta $(\boldsymbol{X}, \theta)$. tenemos a conjunta

$$\begin{pmatrix} \boldsymbol{X} \\ \theta \end{pmatrix} \sim N_2(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (1+\tau)\boldsymbol{I}_p & \tau\boldsymbol{I}_p \\ \tau\boldsymbol{I}_p & \tau\boldsymbol{I}_p \end{pmatrix}).$$

pois

$$E[\boldsymbol{X}] = E\{E[\boldsymbol{X}|\theta]\} = E[\theta] = \mathbf{0},$$

$$Cov[\boldsymbol{X}] = E[\boldsymbol{X}\boldsymbol{X}'] = E[E\{\boldsymbol{X}\boldsymbol{X}'|\theta\}] = \tau\boldsymbol{I}_p.$$

$$Cov[\boldsymbol{X}, \theta] = E[\boldsymbol{X}\theta'] = \tau\boldsymbol{I}_p.$$

Assim,

$$\theta|\boldsymbol{X} \sim N_p(\frac{\tau}{1+\tau}\boldsymbol{X}, \frac{\tau}{1+\tau}\boldsymbol{I}_p)$$

, de modo que

$$\delta_B(X) = E[\theta|\boldsymbol{X}] = \frac{\tau}{1+\tau}\boldsymbol{X}.$$

El estimador arriba es do tipo contracion (”shrinkage”) em direção ao cero. Considerando o estimador no viesado (EMV) $\delta_1(\boldsymbol{X}) = \boldsymbol{X}$, su EQM es dado por

$$MSE[\delta_1(\boldsymbol{X})] = E[||\boldsymbol{X} - \theta||^2] = E[\sum_{i=1}^{p}(X_i - \theta_i)^2] = p.$$

Por otro lado, para o estimador $\delta_B(\boldsymbol{X})$ tenemos

$$MSE[\delta_B(\boldsymbol{X})] = E[||\delta_B(\boldsymbol{X}) - \theta||^2]$$

$$= E[||\frac{\tau}{1+\tau}\boldsymbol{X} - \theta||^2] = E[||\frac{\tau}{1+\tau}(\boldsymbol{X} - \theta) + \frac{(-1)}{1+\tau}\theta||^2]$$

$$= \frac{p\tau^2}{(1+\tau)^2} + \frac{1}{(1+\tau)^2}||\theta||^2.$$

Note que $MSE[\delta_B(\boldsymbol{X})] \to p$ quando $\tau \to \infty$.

Ejemplo 4. (ejemplo 2)

$$X_1, \ldots, X_n, \qquad iid \qquad \mathrm{B}ernolli(\theta),$$

$$\theta \sim Beta(a, b),$$

de modo que

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i},$$

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1},$$

resultando en que

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{g(\boldsymbol{x})} \propto f(\boldsymbol{x}|\theta)\pi(\theta)$$

$$\propto \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

$$= \theta^{\sum_{i=1}^{n} x_i + a - 1}(1-\theta)^{n-\sum_{i=1}^{n} X_i + b - 1},$$

entonces,

$$\theta|\boldsymbol{x} \sim Bernolli(\sum_{i=1}^{n} x_i + a, n - \sum_{i=1}^{n} x_i + b),$$

resultando en

$$d_\pi = E[\theta|\boldsymbol{X}] = \frac{a^*}{a^* + b^*} = \frac{\sum_{i=1}^{n} X_i + a}{n + a + b},$$

con $a^* = \sum_{i=1}^{n} X_i + a$ and $b^* = n - \sum_{i=1}^{n} X_i + b$. Note que particularmente con $\theta \sim U(0, 1)$ (a=b=1), entonces

$$E[\theta|\boldsymbol{X}] = \frac{\sum_{i=1}^{n} X_i + 1}{n + 2}.$$

En este caso es compleja la obtencion del intervalo de MPP

Tambien tenemos

$$Var[\theta|\boldsymbol{X}] = \frac{a^*b^*}{(a^*+b^*)^2(a^*+b^*+1)}$$

$$= \frac{E[\theta|\boldsymbol{X}](1-E[\theta|\boldsymbol{X}])}{a^*+b^*+1}.$$

Para o caso da proiori $\theta \sim U(0,1)$ (Beta(1,1)), Siendo $t = \sum_{i=1}^n X_i$, $Var[\theta|\boldsymbol{X}] = \frac{(t+1)(n-t+1)}{(n+2)^2(n+3)}$.

Se puede Tambien calcular um intervalo central exato. Siendo $\theta \sim Beta(A, B)$, temps que

$$\phi = \frac{B}{A}\frac{\theta}{1-\theta} \sim F(2A, 2B).$$

Assim, IC central (θ_I, θ_S) com $\gamma = 1 - \alpha$ de es dado por

$$\theta_I = \frac{AF_1}{B + AF_1}, \qquad \theta_S = \frac{AF_2}{B + AF_2},$$

com
$F_1 = F_{\alpha/2}(2A, 2B)$

$F_2 = F_{1-\alpha/2}((2A, 2B)$

Se A=B, simestrico es IMPP.

Ex. do Paulino (pg. 147):

$\theta|x \sim Beta(9,3)$,
$F_1 = F_{0.025}(18,6) = 0.3105, \qquad F_2 = F_{0.975}(18,6) = 5.202,$

que leva a

$\theta_I = \frac{9*0.3105}{3+9*0.3105} = 0.482,$
$\theta_S = \frac{9*5.202}{3+9*5.202} = 0.940$

***A**plicacion 4.* Para los datos de la Aplicacion 2 y priori $\theta \sim U(0,1)$, tenemos que

$$d_\pi = \frac{\sum_{i=1}^{n} X_i + 1}{n+2} = \frac{7}{22} = 0,3177.$$

$$Var[\theta|\boldsymbol{X}] = \frac{0.3177(1-0.3177)}{23} = 0.0094.$$

Usando el programa BUGS, el 95% IMPP para θ es dado por $(0,1474; 0,5208)$. Note que los intervalos son poco diferentes. En la inferencia Bayesiana hay la influencia de la priori uniforme.

Intervalo simestrico: $\theta|x \sim Beta(7, 15)$
$F_1 = F_{0.025}(14, 30) = 0.3991$

$F_2 = F_{0.975}(14, 30) = 2.2554$
Assim,

$\theta_I = \frac{7*0.3991}{15+7*0.3991} = 0.157$
$\theta_S = \frac{7*2.2554}{15+7*2.2554} = 0.513.$
IC com $\gamma = 0.95$: [0.157,0.513].

Prevision de observaciones :Interes por datos amuestrales futuros. Se tiene interes en preveer Y (puede ser vector) de $X|\theta \sim f(x|\theta)$. Usando los datos observados y la posteriori acumulada (x es priori $\pi(\theta)$) tenemos lla posteriori de Y

$$\pi(y|x) = \int f(y|x,\theta)\pi(\theta|x)d\theta.$$

En el caso en que X es indepediente de Y dado θ, $f(y|x,\theta) = f(y|\theta)$.

ejemplo. $X_1, \ldots, X_n$ a.a. $X \sim Bernolli(\theta)$. $t = \sum_{i=1}^n X_i$, numero de sucesos.

$\theta \sim Beta(a, b)$.

$\theta|x \sim Beta(A, B)$, com $A = a + t$, $B = b + n - t$.

$$p(y|x) = \int_0^1 \theta^{y-1}(1-\theta)^{1-y}\frac{\theta^{A-1}(1-\theta)^{B-1}}{B(A,B)}$$
$$= \frac{B(A+y, B-y+1)}{B(A,B)}, \quad y = 0, 1$$
$$= \frac{a+t}{a+b+n} I_{\{1\}}(y) + \frac{b+n-t}{a+b+n} I_{\{0\}}(y).$$

Se en una sucesion de n ensayos identicos suceso ocurre n veces, qual es la probabilidad de sucesso en el ensayo n+1, com a=b=1?

$$p(y|x) = \frac{n+1}{n+2}.$$

Como queda el caso de m ensayos futuros?

Para $u = \log(\theta/(1-\theta))$, tenemos la priori (imprópria)

$$p(u) = 1, \qquad -\infty < u < \infty.$$

Considerando la transformacion

$$\theta = \frac{e^u}{1+e^u},$$

com Jacobiano,

$$J = |\frac{\partial \theta}{\partial u}|^{-1} = \theta^{-1}(1-\theta)^{-1},$$

resultado en la priori

$$p(\theta) \propto \theta^{-1}(1-\theta)^{-1},$$

que es una priori $Beta(0,0)$ (imprópria). Novik and Hall (1965, JASA).

Posteriori:

$$\pi(\theta|y) = \frac{1}{B(y, n-y)} \theta^{y-1}(1-\theta)^{n-y-1},$$

de modo que

$$E[\theta|y] = \frac{y}{n},$$

$$Var[\theta|y] = \frac{y(n-y)}{n^2(n-1)}.$$

Note que EB coincide com EMV. Priori menos informativa que priori uniforme.

El caso Poisson.

$Y \sim Poisson(\theta)$; $f(y|\theta) = \frac{e^{-\theta}\theta^{y}}{y!}$.
$\theta \sim Gama(\alpha, \beta)$; $\pi(\theta) = \frac{\beta^{\alpha}\theta^{\alpha-1}e^{-\theta\beta}}{\Gamma(\alpha)}$

Poteriori:

$$\pi(\theta|y, \alpha, \beta) \propto \theta^{y+\alpha-1}e^{-\theta(\beta+1)} \sim Gama(y + \alpha, \beta + 1).$$

Prevision de observacion futura:

$$f(y|x) = \int_0^\infty f(y|x,\theta)\pi(\theta|x)d\theta$$

$$= \int_0^\infty \frac{\theta^y \mathrm{e}^{-\theta}}{y!} \frac{(1+\beta)^{x+\alpha}\theta^{x+\alpha-1}\mathrm{e}^{-\theta(\beta+1)}}{\Gamma(x+\alpha)} d\theta$$

$$= \binom{y+x+\alpha-1}{y} (\frac{\beta+1}{\beta+2})^{x+\alpha} (\frac{1}{\beta+2})^y,$$

que es uma distribucion binomial negativa.

Caso Normal. $X_1, \ldots, X_n$ a.a. $X \sim N(\mu, \sigma^2)$.

$\mu \sim N(\mu_0, \tau^2)$

$\mu | x \sim N(A, B^2)$

$Y_1, \ldots, Y_M$ futuro

$$f(\bar{y}|\mu)\pi(\mu|x) \propto \mathrm{e}^{-\frac{1}{2}\{\frac{(\mu-\bar{y})^2}{\sigma^2/m}+\frac{(\mu-A)^2}{B^2}\}} \propto \mathrm{e}^{-\frac{1}{2}\{\frac{(\mu-C)^2}{2V^2}\frac{1}{2}\frac{(\bar{y}-A)^2}{(B^2+\sigma^2/m)}\}},$$

donde

$$V^2 = (\frac{m}{\sigma^2} + \frac{1}{B^2})^{-1} \qquad C = V^2(\frac{\bar{y}}{\sigma^2/m} + \frac{A}{B^2}).$$

Asi,

$$\bar{Y}|x \sim N(A, B^2 + \sigma^2/m).$$

HPD 95%:

$$[A \mp 1.96\sqrt{B^2 + \sigma^2/m}].$$

Ejemplo 5: situacion normal general.
En el estudio anterior del caso normal (ejemplo 3), la varianza σ^2 de la distribucion es considerada conocida. Nosotros generalizamos ahora considerando la situacion σ^2 desconocida. El modelo considerado es

$$X_1, \ldots, X_n \sim N(\mu, \sigma^2),$$

con $\theta = (\mu, \tau^2)$, $\tau^{-1} = \sigma^2$.

a) Considerando priori dependiente (conjugada):

$$\pi(\mu, \tau) = \pi(\mu|\tau)\pi(\tau),$$

con

$$\mu|\tau \sim N(\mu_0, \tau),$$

$$\tau \sim Gamma(\frac{\delta_0}{2}, \frac{\gamma_0}{2}),$$

de modo que lla posteriori conjunta es dada por

$$\pi(\theta|\boldsymbol{x}) \propto \tau^{n/2}\mathrm{e}^{-\frac{\tau\sum_{i=1}^{n}(x_i-\mu)^2}{2}}\tau^{1/2}\mathrm{e}^{\frac{-\tau(\mu-a)^2}{2}}\tau^{\frac{\delta_0}{2}-1}\mathrm{e}^{-\frac{\gamma_0\tau}{2}}.$$

Es possible obtener posteriores marginales cerradas para μ y τ obtenidas por

$$\pi(\mu|\boldsymbol{x}) = \int_0^\infty \pi(\theta|\boldsymbol{x})d\tau \qquad \text{y} \qquad \pi(\tau|\boldsymbol{x}) = \int_0^\infty \pi(\theta|\boldsymbol{x})d\mu.$$

Usando la distribucion Gamma invertida

$X \sim Gama(a, b)$ si la densidad es de la forma

$$f(x|a, b) = (\Gamma[a])^{-1} b^a x^{a-1} \mathrm{e}^{-bx}, \qquad x > 0.$$

Tenemos $E[X] = a/b$, $Var[X] = a/b^2$. Caso particular

$$X \sim Gama(k/2, 1/2) \longrightarrow X \sim \chi_k^2.$$

$X \sim GI(a, b)$ (GI: gama invertida) se densidad es da forma

$$f(x|a, b) = (\Gamma[a])^{-1} b^a (1/x)^{a+1} \mathrm{e}^{-b/x}, \qquad x > 0,$$

com $E[X] = b/(a-1)$ e $Var[X] = b^2/[(a-1)^2(a-2)], \qquad a > 2.$
Caso $a = k/2$, $b = 1/2$ implica na chi cuadrado invertida, χ_k^{-2}.Se puede Tambien definir raices cuadradas de la Gamma y de la chi cuadrado (chi) inversa.

Família t-Student univariadas $t_k(\lambda, \delta)$ es una familia de distribuciones unimodais simetricas com tres parametros: Localizacion (λ: moda o mediana), escala ($\delta > 0$) es el número de grado de libertad (k). La funcion de densidad es dada por

$$f(x|\lambda, \delta) = (B(k/2, 1/2))^{-1}(\sqrt{k}\delta)^{-1}\{1 + \frac{(x-\lambda)^2}{k\delta^2}\}^{-\frac{k+1}{2}},$$

con $E[X] = \lambda$, $k > 1$ e $Var[X] = k\delta^2/(k-2)$, $k > 2$. Se usa la notacion $t_k(\lambda, \delta)$ ($k = 1$ define a Cauchy). Note que

$$B(k/2, 1/2) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(k/2)\Gamma(1/2)},$$

es que $\Gamma(1/2) = \sqrt{\pi}$.

Se puede tambien obtener la distribucion t-Student como una mezcla de normales com la distribucion Gamma inversa en la escala, es decir

$$X|\lambda, v, W \sim N(\lambda, W/v),$$

$$W \sim GI(a, b)$$

implica que

$$X|\lambda, v, a, b \sim t_{2a}(\lambda, \sqrt{\frac{b}{av}}).$$

Considerando a parametrizacion $a = k/2$, $b = (k/2)c$ se tiene que

$$X|\lambda, k, c \sim t_k(\lambda, \delta),$$

donde $\delta = \sqrt{(c/v)}$. Núcleo de la funcion de densidad:

$$f(x, w|\lambda, v, a, b) \propto w^{-(a+1+1/2)} \mathrm{e}^{-\frac{1}{w}[b+\frac{v}{2}(x-\lambda)^2]},$$

que es denominada NORMAL-GAMA inversa e denotada por $(X, W) \sim NGI(\lambda, v, a, b)$.

Dados: $X_1, \ldots, X_n$ a.a. $X \sim N(\mu, \sigma^2)$. En terminos das estatisticas suficientes $\bar{x}$, $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/k$, k=n-1,

Verosimilitud:

$$L(\mu, \sigma^2|\boldsymbol{x}) \propto (\sigma^2)^{-n/2} \mathrm{e}^{\{-\frac{n}{2\sigma^2}(\mu-\bar{x})^2 - \frac{ks^2}{2\sigma^2}\}}.$$

Priori conjugada: $(\mu, \sigma^2) \sim NGI(a, v, c, d)$, es decir,

$$\mu|\sigma^2 \sim N(a, \sigma^2/v)$$

$$\sigma^2 \sim GI(c, d),$$

que leva a

$$\mu \sim t_{2c}(a, \sqrt{d/(vc)})$$

$$\sigma^2|\mu \sim Gama(c + 1/2, d + \frac{v}{2}(\mu - a)^2).$$

Núcleo de la priori:

$$\pi(\mu,\sigma^2) \propto (\sigma^2)^{-(c+1+1/2)} \mathrm{e}^{-\frac{1}{\sigma^2}[d+\frac{v}{2}(\mu-a)^2]},$$

Núcleo de la posteriori:

$$\pi(\mu,\sigma^2|\boldsymbol{x}) \propto (\sigma^2)^{-1/2} \mathrm{e}^{\{-\frac{1}{2}Q(\mu,\sigma^2)\}}$$
$$(\sigma^2)^{-((n/2)+c+1)} \mathrm{e}^{\{-\frac{d+ks^2/2}{\sigma^2}\}},$$

donde

$$\begin{aligned} Q(\mu,\sigma^2) &= \frac{v}{\sigma^2}(\mu-a)^2 + \frac{n}{\sigma^2}(\mu-\bar{x})^2 \\ &= \frac{n+v}{\sigma^2}(\mu-A)^2 + \frac{nv}{\sigma^2(n+v)}(a-\bar{x})^2, \end{aligned}$$

donde $A = (n\bar{x}+va)/(n+v)$.

Posteriori conjunta es tal que

$$\mu|\sigma^2, x \sim N(A, \sigma^2/(n+v)),$$

$$\sigma^2|x \sim GI(C, D),$$

com

$$C = \frac{n+2c}{2}, \qquad D = d + ks^2/2 + \frac{nv}{2(n+v)}(a-\bar{x})^2,$$

es decir,

$$\mu, \sigma^2|x \sim NGI(A, n+v, C, D).$$

de donde segue que

$$\mu|x \sim t_{n+2c}(A, \sqrt{\frac{D}{C(n+v)}}),$$

$$\sigma^2|\mu, x \sim GI(C+1/2, D + \frac{n+v}{2}(\mu - A)^2).$$

Priori de Jeffreys: asumiendo independência a priori entre μ e σ^2

$$\pi(\mu,\sigma^2) \propto I_F(\sigma^2)^{1/2} = (\sigma^2)^{-1}, \qquad -\infty < \mu < \infty, \qquad \sigma^2 > 0,$$

que segue da conjugada NGI tomando $v \longrightarrow 0$, $c \longrightarrow -1/2$, $d \longrightarrow 0$.
Posteriori conjunta es tal que

$$(\mu,\sigma^2|x) \sim NGI(\bar{x}, n, k/2, ks^2/2),$$

es decir

$$\mu|\sigma^2, x \sim N(\bar{x}, \sigma^2/n),$$

$$\sigma^2|x \sim GI(k/2, ks^2/2)$$

$$\mu|x \sim t_k(\bar{x}, s\sqrt{n}), \qquad \frac{\mu-\bar{x}}{s/\sqrt{n}}|x \sim t_k(0,1), \qquad (\frac{\mu-\bar{x}}{s/\sqrt{n}})^2|x \sim F_{(1,k)},$$

Se puede tambien usar as estatísticas suficientes:

$$\bar{y} \sim N(\mu, \frac{\sigma^2}{n}),$$

$$s^2 \sim Gama(\frac{n-1}{2}, \frac{(n-1)\tau}{2}),$$

com $\tau = 1/\sigma^2$.

Predcion de novas observaciones:

Para predizer a mesdia $\bar{Y}$ dos pesos de m sacos a encher, com a máquina nas mismas condiç oes, a correspondiente densidad preditiva la posteriori es a mistura das densidads normais

$$\bar{Y}|\mu, \sigma^2, x \sim N(\mu, \sigma^2/n),$$

por lla posteriori NGI

$$(\mu, \sigma^2)|x.$$

Dados (Paulino, pg. 172): 16.1,16.0,15.8,15.7,15.9,15.8,16.1,15.7,15.8 16.0,16.2,15.8, 16.0, 16.0,15.9

$$\bar{X} = 15.92, \qquad s^2 = 0.02314.$$

Considerando priori de Jeffreys arriba:

$$(\mu, \sigma^2)|x \sim NGI(15.92, 15, 7, 0.162),$$

$$\mu|x \sim t_{14}(15.92, 0.0392),$$

pois $ks^2/2 = 0.162$ e $s/\sqrt{n} = 0.0392$. O HPD de 95% es dado por

$$[\bar{X} \mp t_{14}(0.025)s/\sqrt{n}] = [15.836, 16.0004],$$

de modo que ao nível de 5% no se rejeita $H_0 : \mu = 16$.

Tambien,

$$\sigma^2|x \sim GI(k/2, ks^2/2) = GI(7, 0.162) \longrightarrow 0.3234/\sigma^2|x \sim \chi^2_{14}.$$

A hipotesis $H_0 : \sigma < 0.15$ (o $\sigma^2 < 0.0225$) o

$$P(0.3234/\sigma^2 > 14.37) = P(\chi^2_{14} > 14.37) = 0.422,$$

de modo que

$$O(H_0, H_1|x) = 0.422/0.578 = 0.73.$$

Mostra que a hipotesis do d.p. ser maior que 150g es 1.37 vezes mais provável a posteiori que H_0.
Simulando da $.3234/\sigma^2|x \sim \chi^2_{14}$ 200 observaciones tomar 5a. menor e 5a. maior observaciones para formar o IC ($\gamma = 0.95$) e entonces de $\sigma^2|x \sim .3234/\chi^2_{14}$, apos ordenadas central $[0.01, 0.05]$.

Testes de hipotesiss

hipotesiss: $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$ Usamos a quantidade:

$$O(H_0, H_1|x) = \frac{P(H_0|x)}{P(H_1|x)},$$

que pode ser denominada como chance la posteriori de vantagens a posteriori a favor de H_0. Define-se entonces o factor de Bayes a favor de H_0 (o contra H_1) como

$$B(x) = \frac{O(H_0, H_1|x)}{O(H_0, H_1)},$$

donde $O(H_0, H_1)$ denota a evidência a priori em favor de H_0. Para comparar dois modelos M_0 e M_1 pose-se considerar a quantidade

$$p(x_i|M_i) = \int f_i(x|\theta_i)h_i(\theta_i)d\theta_i,$$

definindo entonces

$$B(x) = \frac{p(x|M_1)}{p(x|M_2)}.$$

ejemplo. $X_1, \ldots, X_n$ a.a. $X \sim N(\mu, \sigma^2)$ com σ^2 conocido. es de interesse testar as hipotesiss $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$. Considera-se priori no informativa i.e., $P(H_0) = P(H_1)$. Tambien,

$$P[H_0|x] = P[\mu \leq \mu_0|x] = P[\frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \bar{x}}{\sigma/\sqrt{n}}|x] = \Phi(\frac{\mu_0 - \bar{x}}{\sigma/\sqrt{n}}).$$

entonces,

$$O(H_0, H_1|x) = \frac{\Phi(\sqrt{n}(\mu_0 - \bar{x})/\sigma)}{1 - \Phi(\sqrt{n}(\mu_0 - \bar{x})/\sigma)}.$$

O teste UMP tem nível crítico

$$P = P[\bar{X} \geq \bar{x}|\mu = \mu_0] = 1 - \Phi(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}) = P[H_0|x].$$

ejemplo:
Siendo $n = 4$, $\bar{x} = 106$, $\sigma^2 = 400$, $\mu_0 = 100$, tenemos

$$P[\mu \leq 100|x] = \Phi(\frac{100 - 106}{10}) = 0.274 = P[\bar{X} \geq 106|\mu = 100].$$

entonces,

$$O(H_1, H_0|x) = (\frac{.274}{.726})^{-1} = 2.653.$$

Assim H_0 es duas vezeds mais provável la posteriori que H_1. O teste clássico afirma que no há evidência suficiente para rejeitar H_0. $\hat{\alpha} = P[\bar{X} \geq 106|\mu = 100] = P[Z \geq 0.6] = .27$.

b) considerando prioris independientes:
Consideramos ahora la situacion

$$X_1, \ldots, X_n \sim N(\mu, \sigma^2),$$

$$\mu \sim N(\mu_0, \tau_0), \qquad \tau \sim Gama(\frac{\delta_0}{2}, \frac{\gamma_0}{2}),$$

llevando a la siguiente posteriori

$$\pi(\theta|\boldsymbol{X}) \propto \tau^{n/2} \mathrm{e}^{-\frac{\tau \sum_{i=1}^{n} (x_i - \mu)^2}{2}} \tau_0^{1/2} \mathrm{e}^{\frac{-\tau_0 (\mu - \mu_0)^2}{2}} \tau^{\frac{\delta_0}{2} - 1} \mathrm{e}^{-\frac{\gamma_0 \tau}{2}}.$$

En este caso no es posible encontrar analiticamente lla posteriori marginal para μ y τ (o σ^2). Es posible, contudo, encontrar densidades condicionales, las cuales, pueden ser usadas para generar muestras de las distributiones condicionales. La condicional para μ es dada por

$$\pi(\mu|\boldsymbol{X},\tau) \propto \mathrm{e}^{-\frac{(n\tau+\tau_0)}{2}(\mu-\frac{n\bar{x}\tau+\mu_0\tau_0}{n\tau+\tau_0})^2},$$

de modo que

$$\mu|\boldsymbol{X},\tau \sim N(\frac{n\bar{x}\tau+\mu_0\tau_0}{n\tau+\tau_0};\frac{1}{n\tau+\tau_0}).$$

La condicional para τ es tal que

$$\pi(\tau|\boldsymbol{X},\mu) \propto \tau^{\frac{n+\delta_0}{2}-1} \mathrm{e}^{-\frac{\tau}{2}[(n-1)s^2+n(\bar{x}-\mu)^2+\gamma_0]},$$

esto es,

$$\tau|\boldsymbol{X},\mu \sim Gama(\frac{n+\delta_0}{2}; \frac{(n-1)s^2+n(\bar{x}-\mu)^2+\gamma_0}{2}),$$

de las cuales muestras para las densidades marginales pueden ser producidas. Se debe inicializar el algoritimo com valor inicial para τ, generar μ de la condicional dado τ. Para el μ generado, generar τ de su condicional y asa sucesivamente.

Informacion de Fisher para distribucion Gama

$$
\begin{aligned}
f(x|\tau) &\propto \tau^{a-1}\mathrm{e}^{-b\tau},\\
\log(f(x|\tau)) &\propto (a-1)\log(\tau) - b\tau,\\
\frac{\partial \log(f(x|\tau))}{\partial\tau} &\propto \frac{(a-1)}{\tau} - b,\\
\frac{\partial^2 \log(f(x|\tau))}{\partial\tau^2} &\propto \frac{(a-1)}{\tau^2},
\end{aligned}
$$

de modo que priori de Jeffeys es dada por

$$\pi(\tau) \propto I_F(\tau)^{1/2} = \frac{1}{\tau}.$$

c) Considerando priori no informativa:
Consideramos ahora la situacion previa com priori no informativa, esto es,

$$X_1, \ldots, X_n \sim N(\mu, \sigma^2),$$

$$\pi(\mu, \sigma^2) \propto \frac{1}{\tau},$$

que puede ser obtenida da situacion previa considerando $(\tau_0 \rightarrow 0, \gamma_0 = \delta_0 = 0)$. Entonces, la posteriori conjunta es dada por

$$\pi(\theta|\boldsymbol{X}) \propto \tau^{n/2} \mathrm{e}^{-\tau \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2}} \tau^{-1}$$

$$= \tau^{\frac{n}{2}-1} \mathrm{e}^{-\frac{\tau}{2} \sum_{i=1}^{n} (x_i-\mu)^2}.$$

Entonces, usando la identidad

$$\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2=(n-1)s^2+n(\mu-\bar{x})^2,$$

la densidad posteriori marginal de μ es dada por

$$\pi(\mu|\boldsymbol{x})=\int_0^\infty \pi(\theta|\boldsymbol{x})d\tau=\int_0^\infty \tau^{\frac{n}{2}-1}\mathrm{e}^{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2}d\tau$$

$$=\int_0^\infty \tau^{\frac{n}{2}-1}\mathrm{e}^{-\frac{\tau}{2}[(n-1)s^2+n(\mu-\bar{x})^2]}d\tau \propto [(n-1)s^2+n(\mu-\bar{x})^2]^{-n/2}$$

$$\propto [1+\frac{n}{(n-1)s^2}(\mu-\bar{x})^2]^{-\frac{n-1+1}{2}},$$

ddonde

$$\mu|\boldsymbol{X}\sim t(n-1,\bar{x},\frac{s^2}{n}). \tag{7}$$

El siguiente resultado fue utilizado:

$$\int_0^\infty x^{p-1}\mathrm{e}^{-ax}dx = a^{-p}\Gamma(p).$$

De (7), sigue que

$$\frac{\mu - \bar{x}}{s/\sqrt{n}} \sim t(n-1),$$

de modo que el intervalo de 95% MPP para μ es dado por

$$\bar{x} \mp t_{n-1}(0,975)\frac{s}{n},$$

ddonde $P[t_{n-1} > t_{n-1}(0,975)] = 0,025.$

Applicacion 5. Para los datos de Darwin, considerando una priori no informativa, la media de la densidad a posteriori (estimador de Bayes) es dado por $\hat{\mu}_B = \bar{x} = 20,84$ y el intervalo 95% MPP es dado por $(0,39;41,95)$ muy similar al del caso clasico. Para σ^2, tenemos $\hat{\sigma^2} = 39.02$, com IC com $\gamma = 0.95$, $[27.42, 58.07]$.

Por otro lado, la densidad la posteriori para el parametro $\tau = \sigma^{-2}$ es dado por

$$\pi(\tau|\boldsymbol{x}) = \int_{-\infty}^{\infty} \pi(\theta|\boldsymbol{x}) d\mu$$

$$\propto \tau^{\frac{n}{2}-1} \mathrm{e}^{-\frac{\tau}{2}\frac{(n-1)s^2\tau}{2}} \int_{-\infty}^{\infty} \tau^{\frac{n}{2}-1} \mathrm{e}^{-n\frac{\tau}{2}(\mu-\bar{x})^2} d\mu$$

$$= \tau^{\frac{n}{2}-\frac{1}{2}-1} \mathrm{e}^{\frac{-(n-1)s^2\tau}{2}},$$

llevando a

$$\tau|\boldsymbol{X} \sim Gamma(\frac{n-1}{2}, \frac{(n-1)s^2}{2}),$$

o, equivalentemente,

$$(n-1)s^2\tau|\boldsymbol{x} \sim \chi^2_{n-1},$$

y consecuentemente el estimador (puntual) de Bayes para τ es

$$E[\tau|\boldsymbol{X}] = \frac{1}{s^2},$$

y el intervalo de credibilidad de 95% es dado por

$$(\frac{\chi_{n-1}(0,025)}{(n-1)s^2}; \frac{\chi_{n-1}(0,975)}{(n-1)s^2}),$$

donde

$$P[\chi_{n-1} < \chi_{n-1}(0,025)] = P[\chi_{n-1} > \chi_{n-1}(0,975)] = 0,025.$$

Se puede obtener el intervalo HPD (aproximado) por simulacion. Generar 200 (o 400) $z = \chi^2_{n-1}$ y entonces transformar

$$\tau = z/((n-1)s^2),$$

transformando-se entonces de volta para $\sigma^2 = 1/\tau$. Ordenar estes valores y obtener los cuantiles 2.5% e 97.5% para un intervalo HPD de 95% de confianza.

Practica 1. Estimacion en la pratica usando R y WinBugs

Comparacion de dos poblaciones normales.

tenemos las
$X_1, \ldots, X_{n1}$ a.a. $X \sim N(\mu_1, \sigma_1^2)$,
$Y_1, \ldots, Y_{n2}$ a.a. $Y \sim N(\mu_2, \sigma_2^2)$,
Caso 1: $\sigma_1 = \sigma_2 = \sigma^2, \theta = (\mu_1, \mu_2, \sigma^2)'$ (Variâncias iguales)
verosimilitud:

$$L(\theta|x, y) \propto (\sigma^2)^{-(n_1+n_2)/2} \mathrm{e}^{-\frac{1}{2\sigma^2} ks^2}$$
$$\mathrm{e}^{-\frac{1}{2\sigma^2}[n_1(\mu_1-\bar{x})^2+n_2(\mu_2-\bar{y})]},$$

donde

$$k = n_1 + n_2 - 2, \qquad s^2 = \frac{1}{k}[\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{j=1}^{n_2}(y_j - \bar{y})^2].$$

Priori (no informativa): (Jeffreys independiente)

$$\pi(\theta) \propto \frac{1}{\sigma^2}, \qquad \sigma^2 > 0.$$

Posteriori:

$$\pi(\theta|x,y) \propto (\frac{1}{\sigma^2})^{-1/2} \mathrm{e}^{-\frac{n_1}{2\sigma^2}(\mu_1-\bar{x})^2}$$
$$(\frac{1}{\sigma^2})^{-1/2} \mathrm{e}^{-\frac{n_2}{2\sigma^2}(\mu_2-\bar{y})^2}$$
$$(\sigma^2)^{-(\frac{k}{2}+1)} \mathrm{e}^{-\frac{1}{2\sigma^2}ks^2}.$$

Por lo tanto,

$$\mu_1 \sim N(\bar{x}, \sigma^2/n_1)$$

$$\mu_2 \sim N(\bar{y}, \sigma^2/n_2)$$

$$\sigma^2|x,y \sim GI(k/2, ks^2/2), \qquad \longrightarrow ks^2/\sigma^2|x,y \sim \chi^2_k.$$

De las distribuiciones a posteriori arriba tenemos para $\lambda = \mu_1 - \mu_2$ que

$$\lambda|\sigma^2, x, y \sim N(\bar{x} - \bar{y}, \sigma^2(n_1^{-1} + n_2^{-2})),$$

e

$$\sigma^2|x, y \sim GI(k/2, ks^2/2),$$

de modo que

$$\lambda|x, y \sim t_k(\bar{x} - \bar{y}, s\sqrt{n_1^{-1} + n_2^{-2}}),$$

o equivalentemente

$$\frac{\lambda - (\bar{x} - \bar{y})}{s\sqrt{n_1^{-1} + n_2^{-1}}}|x, y \sim t_k,$$

Dados A.C.:
131, 125, 131, 119, 136, 138, 139, 125, 131, 134
129, 134, 126, 132, 141, 131, 135, 132, 139, 132
126, 135, 134, 128, 130, 138, 128, 127, 131, 124
Dados D.C.:
137, 136, 128, 130, 138, 126, 136, 126, 132, 139
143, 141, 135, 137, 142, 139, 138, 137, 133, 145
138, 131, 143, 134, 132, 137, 129, 140, 147, 136

A.C.: $\bar{x} = 131.37$, $s_1^2 = 26.31$, $n_1^{-1} + n_2^{-1} = 1/15$
D.C.: $\bar{y} = 136.17$, $s_2^2 = 28.63$, $s^2 = 27.47$.
Assim,

$$\mu_1 - \mu_2 | x, y \sim\sim t_{58}(-4.80, 1.35),$$

pois

$$\bar{x} - \bar{y} = -4.80, \qquad d.p. = \sqrt{(58/56)s^2/15} = 1.35,$$

ddonde $\nu/(\nu - 2) = 58/56$. HPD $\gamma = 0.95\, t_{58}(.975) = 1.96$

$$-4.8 \mp 1.96 * 1.35 = [-7.50, -2.09].$$

que exclui o valor $\lambda = 0$. entonces $H_1 : \lambda \neq 0$ es aceita ao nível de 5%.
Ademas,

$$58s^2/\sigma^2 \sim \chi_{58}, \quad \longrightarrow \sigma^2 \sim 58s^2/\chi_{58}$$

de modo que um HPD aproximado para σ^2 (Generando 200 χ_{58}) es dado por $[17.75, 37.95]$.

Caso 2. Variâncias desiguales.

En este caso $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma^2)'$.

verosimilitud:

$$L(\theta|x,y) \propto (\sigma_1^2)^{-n_1/2} \mathrm{e}^{-\frac{1}{2\sigma_1^2}[ks_1^2+n_1(\mu_1-\bar{x})^2]} (\sigma_2^2)^{-n_2/2} \mathrm{e}^{-\frac{1}{2\sigma_2^2}[ks_2^2+n_2(\mu_2-\bar{y})]}$$

Priori: $\pi(\theta) \propto \sigma_1^{-2}\sigma_2^{-2}$.
Problema conocido como de Behrens-Fisher. Sin solucion exacta. Se puede usar muestrador de Gibbs. Calcula-se a condicionales:

$$\mu_1|\sigma_1, x, y \sim N(\bar{x}, \sigma_1^2/n_1),$$

$$\mu_2|\sigma_2, x, y \sim N(\bar{y}, \sigma_2^2/n_2),$$

$$\sigma_1^2|\mu_1, x, y \sim GI(k_1/2, k_1 s_1^2 + n_1(\mu_1 - \bar{x})^2)$$

$$\sigma_2^2|\mu_2, x, y \sim GI(k_2/2, k_2 s_2^2 + n_2(\mu_2 - \bar{x})^2)$$

Generando muestra de μ_i calcula-se muestra para $\lambda = \mu_1 - \mu_2$. Se puede entonces obtener HPDs.

Comparacion de Variâncias.

Siendo $\Psi = \sigma_1^2/\sigma_2^2$, tenemos que

$$\frac{s_2}{s_1^2}\Psi|x, y \sim F_{k_1,k_2}.$$

Se puede obtener HPDs por simulacion.
En el ejemplo arriba, $s_1^2 = 26.31$ e $s_2^2 = 28.63$

$$(28.63/26.31)\Psi|x, y \sim F_{29,29} \longrightarrow \Psi \sim (26.31/28.63)F_{29,29},$$

que leva ao HPD $[0.49, 1.89]$ indicando que $H_0 : \sigma_1^2 = \sigma_2^2$ no deve ser rejeitada ao nivel de 5%.

Medidas da altura nasal

A.C.

49, 48, 50, 44, 54, 56, 48, 48, 51, 51, 50, 53, 51, 50, 51

54, 50, 53, 50, 49, 51, 47, 53, 50, 49, 55, 53, 48, 54, 46

D.C.

50, 49, 57, 52, 47, 52, 58, 45, 55, 54, 51, 54, 56, 53, 52

47, 51, 54, 50, 47, 46, 44, 54, 55, 52, 57, 52, 48, 48, 51

A.C.: $\bar{x} = 50.53$, $s_1^2 = 7.64$

D.C.: $\bar{y} = 51.37$, $s_2^2 = 13.83$

Ademas, $P[\sigma_1^2 < \sigma^2|x] = 185/200 = 0.93$.

$$(13.83/7.64)\Psi|x,y \sim F_{29,29} \longrightarrow \Psi \sim (7.64/13.83)F_{29,29},$$

que leva ao HPD: [0.29,1.0]. tenemos entonces poca evidencia em favor da igualdade de variâncias.
Ademas, $P[\sigma_1^2 < \sigma^2|x] = 185/200 = 0.93$.

Aproximacion t-Student para distribucion B.F.
Siendo

$$\tau = \frac{\lambda - (\bar{x} - \bar{y})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_1}}},$$

$$\tau | x, y \sim t_b(0, a),$$

donde

$$a = \sqrt{(b-2)c_1/b}, \qquad b = 4 + c_1^2/c_2,$$

$$c_1 = \frac{k_1}{k_1 - 2} sen^2 u + \frac{k_2}{k_2 - 2} cos^2 u,$$

$$c_2 = \frac{k_1^2}{(k_1-2)^2(k_1-4)} sen^4 u + \frac{k_2^2}{(k_2-2)^2(k_2-4)} cos^4 u,$$

En el ejemplo arriba, $u = arctan(s_1/s_2) = 36.6^0$, que leva a

$$\tau \sim t_{50.2}(0, 1.02),$$

equivalente a

$$\lambda | x, y \sim t_{50.2}(\bar{x} - \bar{y}, 1.02\sqrt{(s_1^2 + s_2^2)/30}),$$

com $\bar{x} - \bar{y} = -0.83$, $1.02\sqrt{(s_1^2 + s_2^2)/30} = 0.86$. HPD (por simulacion) 95% para λ: [-2.72,0.79].

Extension para ANOVA (efectos fixos):

Modelo:

$$Y_{ij} = \mu_i + e_{ij},$$

$$eij \sim N(0, \tau), \qquad i = 1, \ldots, p, \qquad j = 1, \ldots, n.$$

Priori:Com $\tau = 1/\sigma^2$

$$\mu_i \sim N(\mu_{i0}, \tau_{\mu_i}),$$

$$\tau \sim Gamma(\frac{\delta_0}{2}, \frac{\gamma_0}{2}).$$

Posteriori: Para $\theta = (\mu_1, \ldots, \mu_p, \tau)$

$$\pi(\theta|\boldsymbol{y}) \propto \tau^{np/2} \mathrm{e}^{-\frac{\tau}{2}\sum_{i=1}^{p}\sum_{j=1}^{n}(y_{ij}-\mu_i)^2} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{p}(\mu_i-\mu_{i0})^2\tau_{\mu_i}}$$

$$\tau^{\frac{\delta_0}{2}-1}\mathrm{e}^{-\frac{\tau\gamma_0}{2}}.$$

no se pode obtener posterioris marginales em forma fechada. No entanto Se puede obtener las posterioris condicionales para μ_i e τ, que son dadas por

$$\mu_i|\boldsymbol{y},\tau \sim N(\frac{\tau\sum_{i=1}^{p} y_{ij} + \mu_{i0}\tau_{\mu_i}}{n\tau + \tau_{\mu_i}}, n\tau + \tau_{\mu_i}),$$

Para τ, tenemos que

$$\tau|\mu_1,\ldots,\mu_p,\boldsymbol{y} \sim Gama(\frac{pn+\delta_0}{2}, \frac{\sum_{i=1}^{p}\sum_{j=1}^{n}(y_{ij}-\mu_i)^2+\gamma_0}{2}).$$

Algorítimo: $\tau^{(0)}$, $\mu_i^{(1)}$, $\tau^{(1)}$,...

ANOVA com efectos aleatórios

Modelo

$$Yij = \mu + a_i + e_{ij},$$

con a_i independientes e independiente de e_{ij} para todo i e j e

$$a_i \sim N(0, \sigma_a^2),$$

$$e_{ij} \sim N(0, \sigma_e^2).$$

Representaciones:

1. $Y_{ij}|\mu_i \sim N(\mu_i, \sigma^2)$
$\mu_i \sim N(\mu, \sigma_a^2), \qquad i = 1, \ldots, p, \qquad j = 1, \ldots, n.$
2. $Y_{ij}|a_i \sim N(\mu + a_i, \sigma^2)$
$a_i \sim N(0, \sigma_a^2), \qquad i = 1, \ldots, p, \qquad j = 1, \ldots, n.$

Marginalmente:

$$E[Y_{ij}] = E[E[Y_{ij}|a_i]] = \mu,$$

$$Var[Y_{ij}] = \sigma^2 + \sigma_a^2,$$

$$Cov(Y_{ij}, Y_{ik}) = \sigma_a^2,$$

de modo que $Corr(Y_{ij}, Y_{ik}) = \frac{\sigma_a^2}{\sigma^2+\sigma_a^2}$.

entonces,

$\boldsymbol{V}_i = \boldsymbol{I}_n\sigma^2 + \boldsymbol{J}_n\sigma_a^2$, com

$$|\boldsymbol{V}_i| = (\sigma^2)^{n-1}(\sigma^2 + n\sigma_a^2),$$
$$\boldsymbol{V}_i^{-1} = \boldsymbol{I}_n\frac{1}{\sigma^2} + \boldsymbol{J}_n\frac{\sigma_a^2}{\sigma^2(\sigma^2+n\sigma_a^2)},$$

de modo que a funcion verosimilitud pode ser escrita como

$$L(\mu, \sigma^2, \sigma_a^2) \propto (\frac{1}{\sigma^2})^{\frac{p(n-1)}{2}} (\frac{1}{\sigma^2 + n\sigma_a^2})^{\frac{p}{2}}$$

$$\mathrm{e}^{-\frac{1}{2}\sum_{=1}^{p}(\boldsymbol{y}_i - \mu \mathbf{1}_n)' \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \mu \mathbf{1}_n)}$$

$$= (\frac{1}{\sigma^2})^{\frac{p(n-1)}{2}} (\frac{1}{\lambda})^{\frac{p}{2}} \mathrm{e}^{-\frac{1}{2}[\frac{1}{\sigma^2} SSD - \frac{n}{\lambda} SSE - \frac{np}{\lambda}(\bar{y}_{..} - \mu)^2]},$$

que lleva a los MLEs:

$$\hat{\mu} = \bar{y}_{..}, \qquad \hat{\lambda} = \frac{SSE}{p} = \frac{n\sum_{i=1}^{p}(\bar{y}_{i.} - \bar{y}_{..})}{p}.$$

$$\hat{\sigma}^2 = \frac{SSD}{p(n-1)} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2}{p(n-1)},$$

$$\hat{\sigma}_a^2 = \frac{\hat{\lambda} - \hat{\sigma}^2}{n} = \frac{\frac{SSE}{p} - \frac{SSD}{p(n-1)}}{n}.$$

Note que $\hat{\sigma}_a^2$ pode ser negativo.

B.I. complicada en el modelo marginal.

Se puede sin embargo MCMC. Con priori

$$\tau_a = 1/\sigma_a^2 \sim Gama(c, d),$$

$$\pi(\mu, \sigma^2, \sigma_a^2) = \pi(\sigma^2)\pi(\sigma_a^2)\pi(\mu),$$

com

$$\pi(\mu) = 1, \qquad -\infty < \mu < \infty,$$

$$\tau = 1/\sigma^2 \sim Gama(a, b),$$

$$\tau_a = 1/\sigma_a^2 \sim Gama(c, d),$$

tenemos a posterir conjunta:

$$\pi(\boldsymbol{a}, \mu, \tau, \tau_a) \propto \tau^{np/2} \mathrm{e}^{-\frac{\tau}{2}\sum_{i=1}^{p}\sum_{i=1}^{n}(y_{ij}-\mu-a_i)^2} \tau_a^{\frac{p}{2}} \mathrm{e}^{-\tau_a \frac{\sum_{i=1}^{p} a_i^2}{2}}$$

$$.\tau^{\frac{a-1}{2}} \mathrm{e}^{-\frac{\tau b}{2}} \tau_a^{\frac{c-1}{2}} \mathrm{e}^{-\frac{\tau_a d}{2}}.$$

Se puede entonces calcular las condicionales:

$$\mu|\tau, \tau_a^2, \boldsymbol{a}, \boldsymbol{y} \sim N(\frac{\sum_{i=1}^p \sum_{j=1}^n (y_{ij} - a_i)^2}{np}, np\tau)$$

$$a_i|\mu, \tau, \tau_a^2, \boldsymbol{a}_{(-i)}, \boldsymbol{y} \sim N(\frac{n\tau}{n\tau + \tau_a}(\bar{y}_{i.} - \mu), n\tau + \tau_a)$$

$$\tau|\mu, \tau, \tau_a, \boldsymbol{a}, \boldsymbol{y} \sim Gama(\frac{np + a}{2}, \frac{b + \sum i = 1^p \sum_{j=1}^n (y_{ij} - \mu - a_i)^2}{2},$$

$$\tau_a|\mu, \tau, \tau_a, \boldsymbol{a}, \boldsymbol{y} \sim Gama(\frac{p + c}{2}, \frac{\sum_{i=1}^p a_i^2 + d}{2})$$

2. MODELOS DE REGRESION SIMPLE

Modelo relaciona pares de observaciones $(y_1, x_1), ..., (y_n, x_n)$ por el uso de la relacion

$$y_i = \alpha + \beta x_i + e_i,$$

donde la variable (aleatoria) y_i es tipicamente conocida como variable respuesta, x_i la variable controlada (variable fijada o covariable), y e_i es el termino de error asociado con el modelo, $i = 1, \dots, n$. Una suposicion comun considera que e_i es independiente de e_j, $i \neq j$, y que

$$e_i \sim N(0, \sigma^2), \qquad i = 1, \dots, n.$$

Este modelo es conocido como modelo homoscedastico de regression simple (MRS), es decir, varianza constante. Note que en este caso el modelo depende de tres parametros, esto es, $\theta = (\alpha, \beta, \sigma^2)$.

2.1 Inferencia de MV para MRS

Bajo los supuestos, la funcion de verosimilitud es dada por

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2}$$

con la funcion log-verosimilitud dada por

$$l(\theta) = \log L(\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2}.$$

Maximizacion de la funcion de verosimilitud (o log-verosimilitud) sigue de la solucion que siguen de las funciones escore

$$\frac{\partial l(\theta)}{\partial \alpha} = 0, \qquad \frac{\partial l(\theta)}{\partial \beta} = 0, \qquad \frac{\partial l(\theta)}{\partial \sigma^2} = 0,$$

que lleva a los siguientes estimadores

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \qquad \hat{\beta} = \frac{S_{xy}}{S_x^2} = r^2 \frac{S_y^2}{S_{xy}}$$

y

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2,$$

ddonde $r = S_{xy}/S_x S_y$,

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}),$$

and $\bar{x} = \sum_{i=1}^{n} x_i/n$. Similarmente para S_y^2.

Los estimadores $\hat{\alpha}$ y $\hat{\beta}$ son de menor varianza en la clase de los estimadores insesgados (UMVUE).
El estimador $\hat{\sigma}^2$ es sesgado para σ^2. Un estimador insesgado sigue reemplazando n por $n-2$ en el denominador de $\hat{\sigma}^2$.
Note que probar $H_0 : \beta = 0$ es equivalente a probar $H_0 : \rho_{xy} = 0$, ddonde ρ_{xy} es la correlacion entre Y e X. La correlacion ρ_{xy} es estimada por r arriba.

Inferencia para β segue del hecho que

$$\hat{\beta} \sim N(\beta, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}),$$

conjuntamente con

$$\frac{\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2}{\sigma^2} \sim \chi^2_{n-2},$$

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{n\hat{\sigma}^2}{n-2}}} \sim t_{n-2},$$

que puede usarse para probar hipotesis y construir intervalos de confianza para β. Se retiramos la suposicion de normalidad, resultados valen en la classe de los estimadores lineales y normalidade sigue del TLC.

Distribucion t-multivariada
$\boldsymbol{Z} = (Z_1, \ldots, Z_p)'$ t-Student p-variada se fdp

$$f(z|\theta) = c(k, p, \Delta)\{1 + \frac{(z-\lambda)'\Delta^{-1}(z-\lambda)}{k}\}^{-\frac{k+p}{2}},$$

$Z \sim t_p(k; \lambda, \Delta)$. Representação como mistura:

$$Z|\lambda, \Delta, Y \sim N_p(\lambda, Y\Delta),$$

$$Y \sim GI(a, b),$$

que leva a $Z \sim t_p(2a; \lambda, \frac{b}{a}\Delta)$.

Siendo $a = k/2$ e $b = 1/2$, entonces dado Y,

$$(Z-\lambda)'\Delta^{-1}(Z-\lambda)/Y \sim \chi^2_p.$$

Marginalmente,

$$U = \frac{k}{p}(Z-\lambda)'\Delta^{-1}(Z-\lambda)/Y \sim \chi^2_p \qquad \frac{(Z-\lambda)'(\Delta^{-1}/Y)(Z-\lambda)}{Y^{-1}/p} \sim F_{p,k},$$

com densidad

$$f_U(u) = \frac{(p/k)^{k/2}}{B(\frac{k}{2},\frac{p}{2})}u^{p/2}(1+\frac{p}{k}u)^{-\frac{k+p}{2}}, \qquad u > 0.$$

Transformaciones lineales de um vector t-Student no altera su tipo distribucional. es decir,

$$\boldsymbol{Z}|Y \sim N_p(\lambda, \Delta Y) \longrightarrow \boldsymbol{AZ}|Y \sim N_r(\lambda^*, \Delta^*)$$

$$Y \sim Gama(k/2, ck/2)$$

$$\boldsymbol{AZ} \sim t_r(\lambda^*, c\Delta^*)$$

donde

$$\lambda^* = \boldsymbol{A}\lambda, \qquad \Delta^* = \boldsymbol{A}\Delta.$$

Distribuiciones la posteriori.

Priori no informativa (Jeffreys): independência a priori entre β e σ^2:

$$\pi(\beta, \sigma^2) \propto (\sigma^2)^{-1}.$$

$$\pi(\beta, \sigma^2|\boldsymbol{y}) \propto (\sigma^2)^{-(\frac{k}{2}+1)} \mathrm{e}^{-\frac{ks^2}{2\sigma^2}}$$
$$(\sigma^2)^{-(\frac{p}{2})} \mathrm{e}^{-\frac{1}{2}[(\beta-\hat{\beta})' \boldsymbol{X}' \boldsymbol{X} (\beta-\hat{\beta})]}.$$

entonces,

$$\beta|\sigma^2, \boldsymbol{y} \sim N_p(\hat{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$$
$$\sigma^2|\boldsymbol{y} \sim GIk/2, ks^2/2).$$

Por lo tanto, β e σ^2 son conjuntamente distribuidos de acuerdo con la Normal Multivariada - Gamma Inversa.

Segue-se entonces de los resultados arriba sobre t-Student Multivariada que la posteriori de β es uma distribucion t-Student p-variada

$$\beta|\boldsymbol{y}, \boldsymbol{X} \sim t_p(k, \hat{\beta}, s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}),$$

implicando que

$$(\beta - \hat{\beta})'\boldsymbol{X}'\boldsymbol{X}(\beta - \hat{\beta})/(ps^2)|\boldsymbol{y}, \boldsymbol{X} \sim F_{p,k}.$$

Tambien,

$$\sigma^2|\beta, y \sim GI(\frac{n}{2}, \frac{ks^2 + (\beta - \hat{\beta})'\boldsymbol{X}'\boldsymbol{X}(\beta - \hat{\beta})}{2}).$$

Las distribuiciones marginales univariadas son expressdas por

$$\beta_j | y \sim t_1(k; \hat{\beta}_j, s^2 d_j) \longrightarrow \frac{(\beta_j - \hat{\beta}_j)^2}{s^2 d_j} | y \sim F_{1,k},$$

$j = 1, \ldots, p$. Para combinaciones lineales do tipo $a'\beta$, tenemos

$$a'\beta | y \sim t_1(k; a'\beta, s^2 a'(\boldsymbol{X}'\boldsymbol{X})^{-1} a) \longrightarrow ....$$

Para x_0 com correspondiente reposta Y_0

$$E[Y_0 | \beta, \sigma^2, x_0] = x_0'\beta \sim t_1(k; x_0'\hat{\beta}, s^2 x_0'(X'X)^{-1} x_0).$$

Infêrencia Marginal

$$\frac{ks^2}{\sigma^2}|y \sim \chi^2_k, \qquad k = n - p,$$

$$\beta|y \sim t_P(k; \hat{\beta}, s^2(X'X)^{-1}),$$

es decir,

$$\pi(\beta|y) \propto [1 + \frac{(\beta - \hat{\beta})'X'X(\beta - \hat{\beta})}{ks^2}]^{-\frac{k+p}{2}},$$

com

$$var[\beta|y] = \frac{ks^2}{k-2}(X'X)^{-1} = (X'X)^{-1}E[\sigma^2|y].$$

HPD para β_j com C.C. $\gamma = 1 - \alpha$

$$\hat{\beta}_j \mp t_{\alpha/2}(k)s\sqrt{d_j},$$

donde d_j es o j-essimo elemento diagonal de $(X'X)^{-1}$.

Y: contagem no Geiger
X: altura, n=39.
Modelo: $Y_j = \alpha + \beta x_j + e_j$, $e_j \sim N(0, \sigma^2)$
Y: 11.0,12.7,...
X: 213.4, 426.8,....
Priori: $\pi(\beta_0, \beta_1, \sigma^2) \propto (\sigma^2)^{-1}$.
tenemos entonces la posteriori

$$\beta_1|y \sim t_1(37, 0.0062, 0.00000032)$$

$$\beta_0|y \sim t_1(37, 8.55, 0.82).$$

tenemos Tambien que

$$Var[\beta|y] = \frac{ks^2}{k-2}(X'X)^{-1}, \qquad k = n - p, \qquad p = 2,$$

que leva a

$$E[\beta_1|y] = 0.0062, \qquad \sqrt{Var[\beta_1|y]} = 0.00058,$$

$$E[\beta_0|y] = 8.55, \qquad \sqrt{Var[\beta_0|y]} = 0.93,$$

com HPD a 95%: $[0.0051; 0.0074]$, indicando forte evidência contra $H_0 : \beta_1 = 0$.

2.2 Inferencia Bayesiana para MRS

Tenemos que especificar prioris para los parametros α, β y $\tau = 1/\sigma^2$. Consideramos prioris conjuntas, esto es,

$$\alpha \sim N(\alpha_0, \sigma^2_\alpha),$$

$$\beta \sim N(\beta_0, \sigma^2_\beta),$$

$$\sigma^2 \sim Gamma(\frac{\delta_0}{2}, \frac{\gamma_0}{2}),$$

con α_0, β_0, δ_0 y γ_0 conocidos.

La combinacion de la verosimilitud y priori lleva a lla posteriori para θ:

$$\pi(\theta) \propto \frac{1}{\sigma^n} \mathrm{e}^{-\frac{1}{2\sigma^2}(y_i-\alpha-\beta x_i)^2}$$

$$.\mathrm{e}^{-\frac{(\alpha-\alpha_0)^2}{\sigma^2_\alpha}} \mathrm{e}^{-\frac{(\beta-\beta_0)^2}{\sigma^2_\beta}} \tau^{\frac{\delta_0}{2}-1} \mathrm{e}^{-\frac{\gamma_0 \tau}{2}}.$$

Como en el caso del modelo de localizacion-escala, no es possible obtener forma cerrada para la expression para la densidads posteriores marginales. Despuas de extensas manipulaciones algebraicas sigue que las densidads condicionales marginales para α, β y τ son dadas por

$$\alpha|\beta, \tau, \boldsymbol{X}, \boldsymbol{Y} \sim N(\frac{\tau \sum_{i=1}^{n}(y_i - \beta x_i) + \alpha_0 \tau_\alpha}{n\tau + \tau_\alpha}; \frac{1}{n\tau + \tau_\alpha}),$$

$$\beta|\alpha, \tau, \boldsymbol{X}, \boldsymbol{Y} \sim N(\frac{\tau \sum_{i=1}^{n}(y_i - \alpha)x_i + \beta_0 \tau}{\sum_{i=1}^{n} x_i^2 + \tau_\beta}; \frac{1}{\sum_{i=1}^{n} x_i^2 + \tau_\beta})$$

and

$$\tau|\alpha, \beta, \boldsymbol{X}, \boldsymbol{Y} \sim Gamma(\frac{n + \delta_0}{2}; \frac{\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 + \gamma_0}{2}).$$

El algoritmo puede empezar com valores iniciales $\beta^{(0)}$ y $\tau^{(0)}$ para β y τ y generando $\alpha^{(1)}$ de la densidad conditional para α y entonces generar $\beta^{(1)}$ y usando $\alpha^{(1)}$ and $\tau^{(0)}$ y generar $\tau^{(1)}$ de la condicional para τ usando $\beta^{(1)}$ y $\alpha^{(1)}$, completando asa un ciclo do algoritimo.
Una priori no informativa sigue tomando

$$\pi(\alpha, \beta, \tau) \propto \frac{1}{\tau}.$$

Esta situacion puede ser discutida en el caso general.

Modelo Lineal General:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e},$$

com $\boldsymbol{Y}$, nx1, $\boldsymbol{X}$, nxp,

$$\boldsymbol{e} \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$$

. tenemos entonces que

$$\boldsymbol{Y}|\beta, \sigma^2 \sim N_n(\boldsymbol{X}\beta, \sigma^2 \boldsymbol{I}_n)$$

com fdp

$$f(\boldsymbol{y}|\beta, \sigma^2) \propto (\sigma^2)^{-n/2} \mathrm{e}^{-\{\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X}\beta)'(\boldsymbol{Y}-\boldsymbol{X}\beta)\}}$$

$$= (\sigma^2)^{-n/2} \mathrm{e}^{-\{\frac{1}{2\sigma^2}[ks^2+(\beta-\hat{\beta})'\boldsymbol{X}'\boldsymbol{X}(\beta-\hat{\beta})]\}}.$$

com $ks^2 = (\boldsymbol{y} - \boldsymbol{X}\hat{\beta})'(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})$ a soma dos quadrados residual, $k = n - p$ e $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, estimativa de mínimos quadrados.

Caso particular: Análisis de Variância

$\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$, com $Y_{ij} \sim N(\mu_i, \sigma^2)$, $i = 1, \ldots, m$ com $\boldsymbol{X}$ diagonal com elementos $\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_m}$, vectores de "uns". tenemos entonces

$$\hat{\beta} = \begin{pmatrix} 1/n_1 & 0 & 0 \\ 0 & 1/n_2 & 0 \\ 0 & 0 & 1/n_3 \end{pmatrix} \begin{pmatrix} \sum_j y_{1j} \\ \sum_j y_{2j} \\ \sum_j y_{3j} \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \end{pmatrix}$$

Podemos combinar ANOVA fixa com regresion.

Resultado=ANOCOVA

$$y_{ij} = \mu_i + \beta x_{ij} + e_{ij},$$

$i = 1, \ldots, I, j = 1, \ldots, J.$

Com as suposiciones usuais:

$$e_{ij} \sim N(0, \sigma^2).$$

Practica 2. Modelo de regresion lineal simple en la practica usando R y WinBugs

3. EL MODELO DE REGRESION GENERAL: Perspectiva bayesiana

El modelo linear general en notation matricial es denotado por

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e},$$

ddonde

$$\boldsymbol{e} \sim N_p(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n), \tag{8}$$

ddonde $\boldsymbol{Y} = (y_1, \ldots, y_n)'$, un nx1-vector, $\boldsymbol{X}$ es a nxp-matriz de covariables, β es un px1-vector de parametros desconocido y $\boldsymbol{I}_n$ es la n-dimensional matriz identidad.

El vector de parametros es $\theta = (\beta', \sigma^2)'$.
La suposicion (8) indica que el modelo es homocedastico, esto es, las variancias son iguales.

a) Considerando prioris no informativas
Considerando las prioris

$$\pi(\beta, \sigma^2) \propto \frac{1}{\tau},$$

llevando a la siguiente distribucion la posteriori

$$\pi(\theta|\boldsymbol{X}, \boldsymbol{Y}) \propto \tau^{\frac{n}{2}-1} \mathrm{e}^{-\frac{\tau}{2}(\boldsymbol{Y}-\boldsymbol{X}\beta)'(\boldsymbol{Y}-\boldsymbol{X}\beta)}$$

$$= \tau^{\frac{n}{2}-1} \mathrm{e}^{-\frac{\tau}{2}[\boldsymbol{Y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{Y}+(\beta-\hat{\beta})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}-\boldsymbol{\beta})]},$$

ddonde $\boldsymbol{M} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Denotando

$$s^2 = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{Y}/(n-p)$$

e integrando con respecto a τ, sigue que la densidad posteriori marginal de β es dada por

$$\pi(\beta|\boldsymbol{X},\boldsymbol{Y}) \propto \left[1+\frac{1}{s^2(n-p)}(\beta-\hat{\beta})'(\boldsymbol{X}'\boldsymbol{X})(\beta-\hat{\beta})\right]^{-\frac{n-p+p}{2}},$$

esto es,

$$\beta|\boldsymbol{X},\boldsymbol{Y} \sim t_p((n-p),\hat{\beta},s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}).$$

Por otra parte, integrando lla posteriori conjunta con respecto a β, obtenemos lla posteriori marginal que es dada por

$$\pi(\tau|\boldsymbol{X},\boldsymbol{Y}) \propto \tau^{\frac{n-p}{2}}\mathrm{e}^{-\frac{\tau}{2}[(n-p)s^2]},$$

que es,

$$\tau|\boldsymbol{Y},\boldsymbol{X} \sim Gamma(\frac{n-p}{2},\frac{(n-p)s^2}{2}).$$

Entonces es posible en este caso hacer inferencia marginal para β and σ^2.

b) Considerando prioris informativas

Bajo normalidad, se puede mostrar que,

$$\boldsymbol{Y}|\boldsymbol{X} \sim N(\boldsymbol{X}\beta, \sigma^2\boldsymbol{I}_n),$$

$$\beta \sim N(\beta_0, \boldsymbol{\Sigma}_\beta),$$

$$\tau \sim Gamma(\frac{\delta_0}{2}, \frac{\gamma_0}{2}),$$

com la matriz $\boldsymbol{\Sigma}_\beta$ conocida. No hay forma cerrada para las expresiones de las densidades marginales posteriores son obtenidos de manera que las tacnicas MCMC como el muestreo de Gibbs se puede utilizar. Posterioris condicionales eston dadas por

$$\beta|\tau, \boldsymbol{Y}, \boldsymbol{X} \sim N(\boldsymbol{A}^{-1}(\boldsymbol{X}'\boldsymbol{Y}\tau + \boldsymbol{\Sigma}_0^{-1}\beta_0), \boldsymbol{A}^{-1}),$$

$$\tau|\beta, \boldsymbol{X}, \boldsymbol{Y} \sim Gamma(\frac{n+\delta_0}{2}, \frac{(\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta) + \gamma_0}{2}),$$

ddonde $\boldsymbol{A} = \boldsymbol{X}'\boldsymbol{X}\tau + \boldsymbol{\Sigma}_0^{-1}$.
Una extension del modelo arriba se puede obtener tomandose prioris informativas dadas por

$$\beta \sim N(\beta_0, \boldsymbol{\Sigma}_0),$$

$$\tau \sim Gamma(\frac{\delta_0}{2}, \frac{\gamma_0}{2}).$$

Practica 3. Modelo de regresion general en la practica usando R y WinBugs

4. OTROS MODELO DE REGRESION

4.1 El modelo de intercepto nulo.

Una alternativa al modelo de regresion simple se puede obtener de las suposiciones

$$y_i = \beta x_i + e_i,$$

con

$$e_i \sim N(0, \sigma^2 x_i),$$

$i = 1, \ldots, n$. Note que el modelo es heterocedastico pues la varianza no es constante.

La funcion de verosimilitud con $\tau = 1/\sigma^2$ y $\theta = (\beta, \tau)$ es dada por

$$L(\theta) \propto \tau^{\frac{n}{2}} \mathrm{e}^{-\frac{\tau}{2}\sum_{i=1}^{n} \frac{(y_i - \beta x_i)^2}{x_i}},$$

la que lleva al EMVs de β y σ^2

$$\hat{\beta} = \frac{\bar{y}}{\bar{x}}$$

y

$$\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \beta x_i)^2}{x_i}.$$

Considerando prioris

$$\pi(\beta, \sigma^2) \propto \tau^{-1},$$

tenemos que lla posteriori conjunta para $\theta = (\beta, \sigma)$

$$\pi(\theta|\boldsymbol{X}, \boldsymbol{Y} \propto \tau^{\frac{n}{2}-1} \mathrm{e}^{-\frac{\tau}{2}\sum_{i=1}^{n} \frac{(y_i-\beta x_i)^2}{x_i}}.$$

Se puede mostrar que

$$\sum_{i=1}^{n} \frac{(y_i - \beta x_i)^2}{x_i} = \sum_{i=1}^{n} x_i(\beta - \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i})^2 + \sum_{i=1}^{n} \frac{y_i^2}{x_i} - \frac{(\sum_{i=1}^{n} y_i)^2}{\sum_{i=1}^{n} x_i},$$

de modo que la posteriori conjunta se puede escribir como

$$\pi(\theta|\boldsymbol{X},\boldsymbol{Y}) \propto \tau^{\frac{n}{2}-1}\mathrm{e}^{-\frac{\tau}{2}\left[\sum_{i=1}^{n} x_i(\beta-\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i})^2+\sum_{i=1}^{n}\frac{y_i^2}{x_i}-\frac{(\sum_{i=1}^{n} y_i)^2}{\sum_{i=1}^{n} x_i}\right]}$$

de modo que la densidad posteriori marginal de β puede ser escrita como

$$\pi(\beta|\boldsymbol{X},\boldsymbol{Y}) \propto \left\{1+(n-1)\frac{\sum_{i=1}^{n} x_i(\beta-\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i})^2}{\sum_{i=1}^{n}\frac{y_i^2}{x_i}-\frac{(\sum_{i=1}^{n} y_i)^2}{\sum_{i=1}^{n} x_i}(n-1)}\right\}^{-\frac{n-1+1}{2}},$$

es decir,

$$\beta|\boldsymbol{X},\boldsymbol{Y}) \sim t(\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i};\sum_{i=1}^{n}\frac{y_i^2}{x_i}-\frac{(\sum_{i=1}^{n} y_i)^2}{(n-1)\sum_{i=1}^{n} x_i};n-1).$$

Lla posteriori marginal de β es dada por

$$\pi(\tau|\boldsymbol{X},\boldsymbol{Y}) \propto \tau^{\frac{n}{2}-1}\mathrm{e}^{-\frac{\tau}{2}(\sum_{i=1}^{n}\frac{y_i^2}{x_i}-\frac{(\sum_{i=1}^{n}y_i)^2}{\sum_{i=1}^{n}x_i})},$$

de modo que

$$\tau|\boldsymbol{X},\boldsymbol{Y} \sim Gamma(\frac{n-1}{2},\frac{\sum_{i=1}^{n}\frac{y_i^2}{x_i}-\frac{(\sum_{i=1}^{n}y_i)^2}{\sum_{i=1}^{n}x_i}}{2}).$$

Modelos (aditivo) com errores en las variables:

$$y_i = \alpha + \beta x_i + e_i,$$
$$X_i = x_i + u_i,$$
$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \sim N_3(\begin{pmatrix} 0 \\ 0 \\ \mu_x \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ & \sigma_u^2 & 0 \\ & & \sigma_x^2 \end{pmatrix}),$$

$i = 1, \ldots, n$. vector de parâmetros $\theta = (\mu_x, \alpha, \beta, \sigma_x^2, \sigma_u^2, \sigma_e^2)$.

Distribucion conjunta de (y_i, X_i) (observables):

$$\begin{pmatrix} y_i \\ X_i \end{pmatrix} \sim N_2(\begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix}, \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_e^2 & \beta\sigma_x^2 \\ & \sigma_x^2 + \sigma_u^2 \end{pmatrix}),$$

$i = 1, \ldots, n.$

Modelo no es identificável, es decir existe $\theta_1 \neq \theta_2$ que levam a misma verosimilitud, es decir, $L(\theta_1) = L(\theta_2)$. Ademas, N_2 só depende de 5 parâmetros e θ envuelve 6 paramêtros. son necesarias condiciones adicionais.

1. σ_u^2 conocido

$$\hat{\beta}_c = \frac{S_{XY}}{S_X^2 - \sigma_u^2}.$$

tenemos as $n \to \infty$:

$$\hat{\beta}_{MQ} = \frac{S_{XY}}{S_X^2} \longrightarrow k_x \beta, \qquad \hat{\beta}_c \longrightarrow \beta.$$

, $K_x = \sigma_x^2/(\sigma_x^2 + \sigma_u^2)$. Logo, $\hat{\beta}_{MQ}$ consistente e $\hat{\beta}_c$ no es consistente.

2. $\sigma_e^2/\sigma_u^2 = k$ conocido $(\sigma_u^2 = \sigma_u^2 = \sigma^2)$

$$\hat{\beta}_c = \frac{S_y^2 - S_X^2 + \sqrt{(S_y^2 - S_X^2)^2 + 4S_{XY}^2}}{2S_{XY}}.$$

Complicado obtener distribuiciones asintoticas para estimadores consistentes.

3. Modelo heterocedástico I:

$$y_i = \alpha + \beta x_i,$$
$$X_i = x_i + u_i,$$

com

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \sim N_3(\begin{pmatrix} 0 \\ 0 \\ \mu_x \end{pmatrix}, \begin{pmatrix} \sigma_{ei}^2 & 0 & 0 \\ & \sigma_{ui}^2 & 0 \\ & & \sigma_x^2 \end{pmatrix}),$$

$i = 1, \ldots, n$. Como observado

$$\begin{pmatrix} y_i \\ X_i \end{pmatrix} \sim N_2(\begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix}, \begin{pmatrix} \beta^2\sigma_x^2 + \sigma_{ei}^2 & \beta\sigma_x^2 \\ & \sigma_x^2 + \sigma_{ui}^2 \end{pmatrix}),$$

$i = 1, \ldots, n$.

Usando a forma da densidad N_2, podemos escribir a log-verosimilitud como

$$l(\theta) \propto \prod_{i=1}^{n} -\log(\sigma_{ui}^2\sigma_{ei}^2 + \beta^2\sigma_{ui}^2\sigma_x^2 + \sigma_{ei}^2\sigma_x^2)/2$$

$$+((\sigma_x^2 + \sigma_{ui}^2)(y_i - \alpha - \beta\mu_x)^2 + (\beta^2\sigma_x^2 + \sigma_{ei}^2)(X_i - \mu_x)^2$$
$$-2\beta\sigma_x^2(y_i - \alpha - \beta\mu_x)(X_i - \mu_x))/(2(\sigma_{ui}^2\sigma_{ei}^2 + \beta^2\sigma_{ui}^2\sigma_x^2 + \sigma_{ei}^2\sigma_x^2.$$

Modelo funciona no Winbugs.

Modelo heterocedástico II:

$$y_i = \alpha + \beta x_i,$$
$$X_i = x_i + u_i,$$

com

$$\begin{pmatrix} e_i \\ u_i \\ x_i \end{pmatrix} \sim N_3(\begin{pmatrix} 0 \\ 0 \\ \mu_x \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ & \sigma_u^2 & 0 \\ & & \sigma_{xi}^2 \end{pmatrix}),$$

$i = 1, \ldots, n$. Como observado

$$\begin{pmatrix} y_i \\ X_i \end{pmatrix} \sim N_2(\begin{pmatrix} \alpha + \beta\mu_x \\ \mu_x \end{pmatrix}, \begin{pmatrix} \beta^2\sigma_{xi}^2 + \sigma_e^2 & \beta\sigma_{xi}^2 \\ & \sigma_{xi}^2 + \sigma_u^2 \end{pmatrix}),$$

$i = 1, \ldots, n$.

Usando la forma de la densidad N_2, podemos escribir la log-verosimilitud como

$$l(\theta) \propto \prod_{i=1}^{n} - \log(\sigma_u^2\sigma_e^2 + \beta^2\sigma_u^2\sigma_{xi}^2 + \sigma_e^2\sigma_{xi}^2)/2$$

$$+((\sigma_{xi}^2 + \sigma_u^2)(y_i - \alpha - \beta\mu_x)^2 + (\beta^2\sigma_{xi}^2 + \sigma_e^2)(X_i - \mu_x)^2$$
$$-2\beta\sigma_{xi}^2(y_i - \alpha - \beta\mu_x)(X_i - \mu_x))/(2(\sigma_u^2\sigma_e^2 + \beta^2\sigma_u^2\sigma_{xi}^2 + \sigma_e^2\sigma_{xi}^2.$$

Modelo funciona en el Winbugs.

Modelos de regression binarios

Nos consideramos agora o modelo de regression probit definido parav respostas binarias, es decir, $y_i = 0$ se falla y $y_i = 1$ se successo. Modelo sigue considerando

$$p_i = P[y_i = 1] = \Phi(\boldsymbol{z}_i'\beta),$$

de modo que la funcion de log-likelihood puede ser escrita como

$$l(\theta) \propto \sum_{i=1}^{n} (log(p_i) - log(1 - p_i)) + log(1 - p_i)$$

$$= \sum_{i=1}^{n} \frac{(y_i - \Phi(\boldsymbol{z}_i'\beta))\phi(\boldsymbol{z}_i'\beta)}{\Phi(\boldsymbol{z}_i'\beta)(1 - \Phi(\boldsymbol{z}_i'\beta))},$$

que puede ser maximizada numaricamente. Un caso particular segue cuando

$$\boldsymbol{z}_i = (1, x_i)',$$

llevando al modelo de regresion binario simples.

Para utilizar el enfoque Bayesiano distribuciones a priori deben ser especificadas para β, esto es, $\beta \sim N(\beta_0, \sigma^2_\beta$, con los dos parametros desconocidos. Otros enlace que puede ser usada es el enlace COMPLEMENTARY LOG-LOG LINK (cloglog) definido por la funcion

$$p_i = 1 - \mathrm{e}^{-\mathrm{e}^{\alpha+\beta x_i}}.$$

La inferencia Bayesiana puede ser implementada directamente usando WINBUGS.

Una extension de este modelo es el modelo de Teoraa de la Respuesta al Item (TRI). Esta involucrada con la situacion ddonde n estudiantes hacen un test de "k" preguntas, con repuesta (Y_{ij}) correcta, codificada con "1" o incorrecta, codificada con "0". Es decir, Y_{ij} es la respuesta del estudiante i a la pregunta j, $i = 1, \ldots, n$, $j = 1, \ldots, k$. Tenemos entonces como datos una matrix nxk de zeros y unos. Podemos considerar el modelo

$$p[Y_{ij} = 1] = F(a_j(\theta_i - b_j)),$$

ddonde a_j y b_j correspdonden, respectivamente, a la discriminacion y dificuldad de la pregunta $j = 1, \ldots, k$ y θ_i correspdonde a la habilidad del estudiante $i = 1, \ldots, n$. La funcion de enlace puede ser la normal, esto es, $F = \Phi$. En este caso tenemos el modelo TRI probito. Se puede ajustar este modelo empleando el programa WinBUGS. El interas mayor esta en la estimacion de las habilidades. Una posibilidad es tomar prioris normales para los parametros del modelo.

Practica 4. Otros modelo de regresion en la practica usando R y WinBugs

5. OTRAS DISTRIBUCIONES PARA LA RESPUESTA

5.1 El modelo t de Student

Una variable aleatória X se dice distribuada de acuerdo con la distribuicion t de Student con parametro de localizacion μ y parametro de escala σ, denotado por $\sim t(\mu, \sigma, \nu)$, si su funcion de densidad es dada por

$$f(x|\theta) = \frac{c(\nu)}{\sqrt{2\pi}\sigma}\left(1 + (\frac{x-\mu}{\sqrt{\nu}\sigma})^2\right)^{-\frac{\nu+1}{2}},$$

ddonde

$$c(\nu) = \frac{\Gamma(\frac{(\nu+1)}{2})}{\Gamma(\frac{\nu}{2})}.$$

Se puede mostrar que la distribucion t de Student puede ser obtenida como mezcla de las distribuciones normal y gamma.

Proposicion 1. *Si*

$$X|\lambda \sim N(\mu, \sigma^2/\lambda),$$

y

$$\lambda \sim Gamma(\frac{\nu}{2}, \frac{\nu}{2}),$$

then $X \sim t(\mu, \sigma, \nu)$.

Prueba. Note que la distribuicion de X puede ser obtenida como

$$f(x|\theta) = \int_0^\infty f(x,\lambda)d\lambda = \int_0^\infty f(x|\lambda)g(\lambda)d\lambda$$

$$= \int_0^\infty \frac{\lambda^{1/2}}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{\lambda}{2}(\frac{x-\mu}{\sigma})^2} c^*(\nu)\lambda^{\lambda/2-1}\mathrm{e}^{-\frac{\lambda\nu}{2}}d\lambda$$

$$= \frac{c(\lambda)}{\sqrt{2\pi}\sigma}\int_0^\infty \lambda^{\frac{\lambda+1}{2}-1}\mathrm{e}^{-\frac{\lambda}{2}\left\{(\frac{x-\mu}{\sigma})^2+\nu\right\}}d\lambda$$

$$= \frac{c(\nu)}{\sqrt{2\pi}\sigma}\left(1+(\frac{x-\mu}{\sqrt{\nu}\sigma})^2\right)^{-\frac{\nu+1}{2}},$$

where

$$c^*(\nu)^{-1}\int_0^\infty \lambda^{\lambda/2-1}\mathrm{e}^{-\frac{\lambda\nu}{2}}d\lambda = \frac{1}{c^*(\nu)}$$

y

$$c(\nu) = \frac{\Gamma(\frac{(\nu+1)}{2})}{\Gamma(\frac{\nu}{2})},$$

concluyendo la prueba.

Para calcular el estimador de maxima verosimilitud tenemos que maximizar la funcion de log-verosimilitud que tiene que ser maximizada numaricamente. Una alternativa es usar el algoritimo EM. Empezamos escribiendo la verosimilitud completa que es dada por

$$L_c(\theta) = \prod_{i=1}^{n} f(x_i, \lambda_i) = \prod_{i=1}^{n} \frac{\lambda^{1/2}}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{\lambda_i}{2}(\frac{x_i-\mu}{\sigma})^2} c^*(\nu)\lambda_i^{\nu/2-1}\mathrm{e}^{-\frac{\lambda_i \nu}{2}},$$

y entonces la etapa-E,

$$\lambda_i | x_i \sim Gamma(\frac{\nu+1}{2}; \frac{1}{2}[\left\{(\frac{x_i-\mu}{\sigma})^2 + \nu\right\}],$$

de modo que

$$\tilde{\lambda}_i = \frac{\nu+1}{(\frac{x_i-\mu}{\sigma})^2 + \nu},$$

$i = 1, \ldots, n.$

La verosimilitud completa es maximizada tomando

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \tilde{\lambda}_i x_i}{\sum_{i=1}^{n} \tilde{\lambda}_i},$$

y

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \tilde{\lambda}_i (x_i - \mu)^2.$$

Para implementar el algoritmo EM sugerir valores iniciales para μ y σ, calcular $\tilde{\lambda}_i$, re calcular μ y σ y entonces $\tilde{\lambda}_i$ y asa por de lante.

Para implementar la inferencia Bayesiana, se puede considerar la jerarquaa

$$y_i|\theta, \lambda_i \sim N(\mu, \frac{\sigma^2}{\lambda_i}),$$

$$\lambda_i \sim Gamma(\frac{\nu}{2}, \frac{\nu}{2}),$$

$$\mu \sim N(\mu_0, \tau_\mu),$$

$$\tau = \sigma^{-2} \sim Gamma(\frac{\delta_0}{2}, \frac{\gamma_0}{2}).$$

La distribucion la posteriori conjunta para la estrutura arriba se puede escribir como

$$\pi(\theta, \lambda) \propto \prod_{i=1}^{n} \frac{\lambda^{1/2}}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{\lambda_i}{2}(\frac{x_i-\mu}{\sigma})^2} \lambda_i^{\nu/2-1} \mathrm{e}^{-\frac{\lambda_i \nu}{2}}$$

$$\mathrm{e}^{-\frac{\tau_\mu}{2}(\mu-\mu_0)^2} \tau^{\delta_0/2-1} \mathrm{e}^{-\frac{\tau\gamma_0}{2}}.$$

Entonces, la densidad la posteriori condicional es dada por

$$\mu|\tau, \lambda, \boldsymbol{y} \sim N(\frac{\tau \sum_{i=1}^{n} y_i \lambda_i + \mu_0 \tau_\mu}{\tau \sum_{i=1}^{n} \lambda_i^2 + \tau \mu}, \tau \sum_{i=1}^{n} \lambda_i^2 + \tau \mu),$$

$$\tau|\mu, \lambda, \boldsymbol{y} \sim Gamma(\frac{n + \delta_0}{2}, \frac{\sum_{i=1}^{n} (y_i - \mu)^2 \lambda_i + \gamma_0}{2}),$$

$$\lambda_i|\mu, \tau, \boldsymbol{y} \sim Gamma(\frac{n + \nu}{2}, \tau (y_i - \mu)^2 + \nu),$$

$i = 1, \ldots, n$, ddonde $\lambda = (\lambda_1, \ldots, \lambda_n)$.

5.2 La distribuicion normal-assimetrica

Version estandard: $X \sim SN(0, 1, \lambda) = SN(\lambda)$ si la densidad es dada por

$$f(x) = 2\phi(x)\Phi(\lambda x).$$

La funcion de verosimilitud para la muestra de tamaao n, $X_1, \ldots, X_n$,

$$L(\lambda) \propto \prod_{i=1}^{n} \Phi(\lambda x_i),$$

con log-verosimilitud

$$l = \log L(\sigma) \propto \sum_{i=1}^{n} \log \Phi(\lambda x_i),$$

y funcion escore dada

$$\Lambda(\lambda) = l'(\lambda) = \frac{\partial l(\lambda)}{\partial \lambda} = \sum_{i=1}^{n} \frac{\phi(\lambda x_i)}{\Phi(\lambda x_i)}.$$

La ecuacion de verosimilitud es dada por

$$\sum_{i=1}^{n} \frac{\phi(\lambda x_i)}{\Phi(\lambda x_i)} = 0.$$

Note que si $x_i > 0 (< 0)$, para todo i, la verosimilitud es ilimitada y EMV para $\lambda = \infty$.

Esta ecuacion no tiene solucion analitica. Tiene que usarse matodos numaricos. Una extension para el modelo de localizacion-escala con $\theta = (\mu, \sigma, \lambda)$ es dada por

$$L(\theta) = \prod_{i=1}^{n} \frac{2}{\sigma} \phi(\frac{y_i - \mu}{\sigma}) \Phi(\lambda(\frac{y_i - \mu}{\sigma})),$$

que tambion debe ser maximizada numaricamente.

La distribuicion normal asimatrica (“skew-normal”) tiene un problema de regularidad. Su matriz de informacion de Fisher es singular en $\lambda = 0$.
La informacion de Fisher en $\lambda = 0$ es dada por

$$I_F(\mu, \sigma, 0) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 & \sqrt{\frac{2}{\pi}}\frac{1}{\sigma} \\ 0 & \frac{2}{\sigma^2} & 0 \\ \sqrt{\frac{2}{\pi}}\frac{1}{\sigma} & 0 & \frac{2}{\pi} \end{pmatrix}.$$

Resultado importante: Representacion Estocastica de la normal asimatrica:

Dado V_1 y V_2 iid $N(0, 1)$ entonces

$$X = \delta|V_1| + \sqrt{1 - \delta^2}V_2,$$

con

$$\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}},$$

es tal que $X \sim SN(\lambda)$.
Puede usarse esta representacion para implementar el muestreo de Gibs o el algoritimo EM. Se puede entonces escribir el modelo jerarquicamente como

$$y_i|t_i \sim N(\delta t_i, 1 - \delta^2),$$

$$t_i \sim HN(0, 1).$$

El enfoque Bayesiano necesita priori para λ que puede ser

$$\lambda \sim U(-0.995, 0.995).$$

Se ingresa con la jerarquica directamente en en WinBugs (ver programa). Para la situacion localizacion-escala se necesita considerar prioris para μ y σ.

Practica 5. Otros distribuciones de la respuesta en la practica usando R y Win-Bugs

5.3 La distribucion power-normal

Version estandard: $Z \sim PN(0,1,\alpha) = SN(\alpha)$ si la densidad es dada por

$$f_\phi(z;\alpha) = \alpha\phi(z)\{\Phi(z)\}^{\alpha-1}.$$

Natese que esta densidad es obtenida del modelo de alternativas de Lehmann (1956) con funcion de distribucion

$$F_\Phi(x;\alpha) = \{\Phi(\lambda x)\}^{\alpha}.$$

Una situacion mas general es obtenida quando se reemplaza la funcion de distribucion de la normal por una distribucion cualquiera $F(.)$. Para el caso normal el modelo tiene asimetria en (-0.6115,0.9007) y curtosis en el intervalo (1.7170,4.3556). Estos valores son obtenidos de la expresion

$$E[Z^n] = \alpha \int_0^1 \Phi(z)^{-n} z^{n-1} dz,$$

que se evalua numaricamente.

La function de verosimilitud para el modelo estandard y muestra de tamano n, $Z_1, \ldots, Z_n$,

$$L(\alpha) = \alpha^n \prod_{i=1}^{n} \phi(z_i)\{\Phi(z_i)\}^{\alpha-1},$$

con log-verosimilitud

$$l(\alpha) = \log L(\alpha) \propto n \log \alpha + (\alpha - 1) \sum_{i=1}^{n} \log \Phi(z_i),$$

y function escore dada

$$\Lambda(\alpha) = l'(\alpha) = \frac{\partial l(\alpha)}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log \Phi(z_i).$$

Despues de igualarse a cero el escore, obtenemos el estimador de maxima verosimilitud de α que es dado por

$$\hat{\alpha} = -\frac{n}{\sum_{i=1}^{n} \log \Phi(z_i)}.$$

La Hessiana es entonces dada por

$$\frac{\partial \Lambda(\alpha)}{\partial \alpha} = -\frac{n}{\alpha^2},$$

que lleva a la informacion para una muestra de tamaao n,

$$I_F(\alpha) = \frac{n}{\alpha^2}.$$

Entonces para $n \to \infty$ tenemos que

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{D} N(0, \alpha^2).$$

Una extension importante es la situacion localizacion-escala que se obtiene de la transformacion $X = \mu + \sigma Z$, con fdp

$$f_\phi(x;\mu,\sigma,\alpha) = \frac{\alpha}{\sigma}\phi(\frac{x-\mu}{\sigma})\{\Phi(\frac{x-\mu}{\sigma})\}^{\alpha-1}.$$

Para este modelo se puede mostrar que la matriz de informacion de Fisher para $\theta = (\mu,\sigma,\alpha = 1)$ es dada por

$$I_F(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 & 0.903192 \\ 0 & \frac{2}{\sigma^2} & -0.595636 \\ 0.903197 & -0.595636 & 1 \end{pmatrix}.$$

Se puede verificar que

$$|I_F(\theta)| = 0.013688/\sigma^4.$$

Entonces para este modelo la matriz de informacion no es singular en el punto de simetrica.

Para inferencia Bayesiana para el parametro α se puede por ejemplo tomar una priori uniforme en (0,10). La implementacion se puede hacer en el Winbugs.

BIMODAL (SKEW) MODELS

Based on the doctoral thesis:

Martinez, G. (2011). Extensões do modelo alfa-potência. IME-USP.

Bimodal skew-normal model:

$$f(x|\lambda) = c_\lambda \phi(x) \Phi(\lambda|x|).$$

Kim (2005) showed model is bimodal (symmetric) for $\lambda > 0$. No information matrix seems available.

Kim, H.J. (2005). On a class of two-piece skew normal distribution. Statistics, 39, 6, 537-553.

ALTERNATIVE APPROACH:

Bayesian analysis for the model

$$f(x|\theta) = \frac{2}{\sigma}(\frac{(1+\alpha y^2)}{1+\alpha})\phi(y)\Phi(\lambda y),$$

$$x \in R, \quad \mu \in R, \quad \lambda \in R, \quad \sigma > 0, \quad \alpha > 0.$$

Reference:

Elal-Oliveiro, D., Gomez, H.W., Quintana, F.A. (2009). Bayesian modeling using a class of bimodal skew-elliptical distributions. JSPI, 139,1484-1492.

Log-bimodal skew-normal model

$$f(x|\theta) = \frac{2}{\sigma x}(\frac{1+\alpha y^2}{1+\alpha})\phi(y)\Phi(\lambda y),$$

where

$$y = \frac{log(y) - \mu}{\sigma},$$

$$x > 0, \quad \mu \in R, \quad \lambda \in R, \quad \sigma > 0, \quad \alpha > 0$$

Maximum likelihood estimation presented in

Bolfarine, H. Gomez, H.W. and Rivas, L.I. (2011). A log-bimodal-skew-normal model. A geochemical application. Jornal of Chemometrics, 23,1-4.

An asymmetric extension of Kim (2005) was developed by Arnold et al. (2009) which is given by the cdf

$$f(x|\lambda, \beta) = 2c_\lambda \phi(x)\Phi(\lambda|x|)\Phi(\beta x).$$

Information matrix was derived and shown to be singular at $\lambda = \beta = 0$.

Arnold, B., Gomez, H. and Salinas, H. (2009). On multiple constraints skewed models. Statistics, 33, 279-293.

New proposal (Martinez et a., 2011): Bimodal (symmetric) power-normal (PN) model (Gupta and Gupta, 2008, TEST)

$$f(z|\alpha) = \alpha c_\alpha \phi(z)\{\Phi(|z|)\}^{\alpha-1},$$

$\alpha > 0$, with

$$c_\alpha = \frac{2^{\alpha-1}}{2^\alpha - 1}.$$

Location-scale extension follows by taking $X = \psi + \eta Z$.

Fisher information matrix for location-scale model given by

$$I_F = \begin{pmatrix} 1/\eta^2 & 0 & a_{01}/\eta \\ & 2/\eta^2 & a_{11}/\eta \\ & & (1 + 2(log2)^2 \end{pmatrix}$$

It can be shown that

$$|I_F| = 2.808/\eta^4.$$

To make the above bimodal model asymmetric, we can use the idea in Arnold et. al. (2009) and define the bimodal asymmetric power-normal modeld (Martinez et al., 2011) as

$$f(z|\alpha) = 2\alpha c_\alpha \phi(z)\{\Phi(|z|)\}^{\alpha-1}\Phi(\beta z),$$

$\alpha > 0$, $z \in R$, with

$$c_\alpha = \frac{2^{\alpha-1}}{2^\alpha - 1}.$$

Location scale version follows by taking $X = \psi + \eta Z$. Likelihood has to be maximized numerically. Or use the Fisher information matrix, presented next.

Fisher information matrix for location-scale model given by

$$I_F = \begin{pmatrix} 1/\eta^2 & 0 & \sqrt{\frac{2}{\pi}}/\eta & a_{01}/\eta \\ & 2/\eta^2 & 0 & a_{11}/\eta \\ & & 2/\pi & 0 \\ & & 0 & (1+2(log2)^2 \end{pmatrix}$$

It can be shown that

$$|I_F| = -0.2999/\eta^4 \neq 0.$$

Hence can test normality, i.e., $H_0 : \alpha = 1.0$, $\beta = 0$, using ordinary likelihood ratio statistics.

POWER SKEW-NORMAL MODEL

Azzalini (1985), developed a general structure for a skew-symmetric density function

$$\varphi(z;\lambda) = 2f(z)G(\lambda z), \quad z, \lambda \in \mathbb{R},$$

where f is a pdf symmetric arond cero and G is an absolutely continuos distribution function symmetric arond cero and λ is the parameter controlling asymmetry. For $f = \phi$ and $G = \Phi$, the pdf and cdf of the standard normal distribution, the skew-normal distribution with density function follows

$$\varphi(z;\lambda) = 2\phi(z)\{\Phi(\lambda z)\}, \quad z \in \mathbb{R},$$

which we denote by $Z \sim SN(\lambda)$.

Alternative model presenting high degrees of asymmetry and kurtosis we consider for $\alpha \in \mathbb{R}^+$, the pdf given by

$$\varphi_F(z;\alpha) = \alpha f(z)\{F(z)\}^{\alpha-1}, \qquad z \in \mathbb{R}, \;\; \alpha \in \mathbb{R}^+,$$

denoted $Z \sim AP(\alpha)$. A particular case (PN model) follows when $F = \Phi$, normal cdf (Gupta and Gupta, 2008, TEST).

The power-skew-normal model

Generalizations for models above were proposed by Barakrishnan (2002) and later by Gupta and Gupta (2004), among others. The model considered in Balakrishnan (2002) is related to the family of pdfs given by

$$g_{\boldsymbol{B}}(x;\lambda,a)=k(\lambda,a)\left[\Phi(\lambda x)\right]^{a-1}\phi(x),\quad x\in\mathbb{R},$$

$a>0$, and $\lambda\in\mathbb{R}$), which for $\lambda=1$ reduces to the pdf of the PN model (Gupta and Gupta, 2008) and for $a=2$ reduces to the pdf of the skew-normal model (Azzalini, 1985). Properties of model above were were studied by Gupta and Gupta (2008).

In Gupta and Gupta (2004) a closely related model is proposed with pdf given by

$$\varphi(z;\lambda) = \frac{\phi(z)\{\Phi(\lambda z)\}^n}{C_n(\lambda)}, \quad z \in \mathbb{R},$$

was studied, where $n \in \mathbb{Z}^+, \ \ \lambda \in \mathbb{R}$, and

$$C_n(\lambda) = \int_{-\infty}^{\infty} \phi(z)\{\Phi(\lambda z)\}^n dz.$$

Hence, for $n = 1$ the skew-normal distribution follows (Azzalini, 1985) and, hence, $C_1(\lambda) = 1/2, \ \lambda \in \mathbb{R}$. For $\lambda = 1$, power-normal (PN) distribution (Gupta and Gupta, 2008) follows for $\alpha \in \mathbb{Z}^+$, $C_n(1) = C_\alpha(1) = 1/\alpha, \ \forall n \in \mathbb{Z}^+$.

One difficulty presented by models above is the normalizing constant which is difficult to obtain and complicated to deal with in maximization procedures, a problem magnified in the location-scale situation.

Suggested by the shape parameters α and λ, which model asymmetry and kurtosis we propose a generalization of the models proposed by SN and PN models introducing the power-skew-normal model (Martinez et al., 2011) with density function given by

$$\phi_{\alpha,\lambda}(x) = \alpha\phi_{\lambda}(x)\left\{\Phi_{\lambda}(x)\right\}^{\alpha-1}, \quad x \in \mathbb{R},\ \ \lambda \in \mathbb{R}\ \ e\ \ \alpha \in \mathbb{R}^{+},$$

where

$$\phi_{\lambda}(x) = 2\phi(x)\Phi(\lambda x)$$

and

$$\Phi_{\lambda}(x) = \int_{-\infty}^{x} \phi_{\lambda}(t)dt.$$

Notation: $PSN(\lambda, \alpha)$.

Special cases of the model occurs for $\alpha = 1$, so that the skew-normal model $\phi_{\alpha=1,\lambda}(x) = \varphi(z;\lambda)$, follows.

For $\lambda = 0$ the model with pdf $\phi_{\alpha,\lambda=0}(x) = \varphi(z;\alpha)$, that is,power-normal (PN) model follows. Ordinary standard normal model is also a special case which follows by taking $\alpha = 1$ and $\lambda = 0$, that is $\phi_{\alpha=1,\lambda=0}(x) = \phi(x)$. α and λ affect both distribution asymmetry and kurtosis and hence the model proposed seems more flexible than the models by Azzalini (1985) and Gupta and Gupta (2008).

With $X \sim PSN(\alpha)$, moments of the random variable X have closed form, but under a variable change the r-th moment of the random variable X can be written as

$$E(X^r) = \alpha \int_0^1 [\Phi_\lambda^{-1}(x)]^r x^{\alpha-1} dx, \tag{0.1}$$

where Φ_λ^{-1} is the inverse of the function Φ_λ. This expectation agrees with the expected value of the function $[\Phi_\lambda^{-1}(x)]^r$ with random variable X distributed beta with parameters α e 1.

For values of λ and α between 0.1 and 100, asymmetry and kurtosis coefficients $\sqrt{\beta_1}$, β_2, for $X \sim PSN(\alpha)$ are in the intervals [-1.4676,0.9953) and [1.4672,5.4386] respectively. Intervals contain the corresponding intervals for the skew-normal model, given by (-0.9953,0.9953) and [3,3.8692) respectively, and for the PN, given by [-0.6115,0.9007] and [1.7170,4.3556], respectively. This illustrates the fact that the skew-power-normal family contains models with greater (and smaller) asymmetry than both skew-normal and the power-normal model.

The Fisher information matrix at $\alpha = 1.0$ and $\lambda = 0$, for the location-scale model is

$$I_F = \begin{pmatrix} 1/\eta^2 & 0 & \sqrt{\frac{2}{\pi}}/\eta & 1/\eta \\ & 2/\eta^2 & 0 & 0 \\ & & 2/\pi & \sqrt{1/2} \\ & & & 1 \end{pmatrix}$$

Notice that this information matrix is not singular.

Application: Roller data. Available for download at the internet. Interest in testing the hypotheses

1) $H_{01} : \alpha = 1.0$ vs $H_{02} : \alpha \neq 1.0$,

using $\Delta_1 = \frac{L_{SN}}{L_{PSN}}$ leads to $-2log(\Delta_1) = 26.34$.

2) $H_{11} : \lambda = 0$ vs $H_{12} : \lambda \neq 0$,

using $\Delta_1 = \frac{L_{PN}}{L_{PSN}}$ leads to $-2log(\Delta_1) = 54.92$.

Both null are rejected at the 5% level (Critical value: 3.84).

LIMITED OR CENSORED REGRESSION (Gomez, Martinez, Bolfarine)

Motivation:

Researchers are often confronted with data for which the recorded continuos response variable has a lower bond (considered here to be cero withot loss of generality) and takes on this bondary value for a sizeable fraction of sample observations. One of the circumstances in which this is the case is when the observed values are true zeroes. An example is provided by hosehold expenditures on a particular type of durable good over a certain period of time in which many hoseholds do not purchase the good and hence their actual expenditure is cero.

Another situation is where the response variable y is antibody concentration, which is measured by any one of a number of laboratory techniques, the choice will depend on the particular antigen(s) of interest, the method of sample collection, the desired scaling, and the available technology. Regardless of the technique, there is always a concentration value T below which an exact measurement cannot be reported, although the particular T will be a function of the assay that is employed. In general T is assumed to be a known constant.
When data from an assay are left-censored, the lower detection limit is known and may be used to substitute a value for the censored observation, that is the value T.

Conventional regression analysis may not be an adequate method for making inference in the presence of such point mass at the low end of the measurement scale. One can resort to binary regression analysis (eg.probit or logit techniques) if one is willing to restrict attention to probabilities of limit and nonlimit responses. Given such a framework, those binary response models represent adequate statistical models. However they can be less efficient if the goal is to extract information regarding continuos response. A more efficient analysis follows by combining continuos information with binary data.

Another approach to deal with the zeroes is to treat them a latent (unobserved) continuos observations that have been left censored. This idea was popularized by Tobin (1956) and the resulting model is typically refereed to as the tobit model. Formally, the tobit model can be formulated as

$$y_i = \begin{cases} 0 & \text{if } w_i \leq 0, \\ w_i; & \text{otherwise,} \end{cases}$$

where the latent variable $w_i = \mathbf{x}_i'\beta + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma^2)$, $i = 1, \ldots, n$. Therefore, we denote observed otcomes by y_i, $i = 1, \ldots, n$, the values of the k explanatory variables for the i-th observation by $\mathbf{x}_i \in R^k$ and the regression parameter by $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ and the i-th residual term by ϵ_i.

The log likelihood function for the Tobit model (for T=c) for the PN model can be written as

$$\ell(\theta;Y) = \alpha\sum_i (1-I_i)\ln\left[F\left(\frac{c-\xi}{\eta}\right)\right]+$$
$$\sum_i I_i\left\{\ln(\alpha)-\ln(\eta)+\ln\left(f\left(\frac{y_i-\xi}{\eta}\right)\right)+(\alpha-1)\ln\left(F\left(\frac{y_i-\xi}{\eta}\right)\right)\right\}$$

donde

$$I_i = \begin{cases} 1, & \text{se } y_i > c, \\ 0, & \text{se } y_i \leq c. \end{cases}$$

Tobit power model

Martinez, Bolfarine, Gomez (2012). The alpha power-tobit model. Communication in Statistics (to appear).

Althogh the interpretation of the negative latent observation can be slightly ambiguos, the Tobit specification is usually appropriate for the situation in which the sample proportions of zeros is roghly equivalent to the left tail area of the assumed parametric distribution. Such a specification has interpretation difficulties, however, when there is an excess of limit responses above what wold be implied by the tail area of the relevant parametric distribution. When such is the case a more general model than the Tobit model is required. An extension of the Tobit model to the more general situation is considered in Muños (2009), where the normal distribution is replaced by the Student-t distribution.

Generally speaking, Bernolli/$f^+(y)$ mixture has been considered in

Cragg(1971):

$$g(y_i) = q_i I_i + (1 - q_i) f^+(y_i)(1 - I_i).$$

Molton and Halsey (1995):

$$g(y_i) = [q_i + (1 - q_i)F^+(T)]I_i + (1 - q_i) f^+(y_i)(1 - I_i),$$

where $I_i = 0$ if $y_i > T$ and $I_i = 1$, otherwise. The situation where truncation point is $T = 0$ is straightforward.

Models for Haiti data (Molton and Halsey, 1995)

Y: antibody concentration

EZ (vaccine strain,0: Schwarz, 1: Edmonston-Zagreb); HI (vaccine dose, 0: medium, 1: high); and FEM (sex; 0: male,1:female). The constant term is denoted INT

Lower Detection Limit: .1 (IU)
Observations below LDL: 26.1%
Expected with Tobit: 1.1%

MODEL A: Withot censored data and covariates in the limited response and in the point distribution located at cero;

$$g(y_i) = f^+(y_i; \mu_2, \sigma) \tag{0.2}$$

MODEL B: With covariates, withot censored data and limited response with covariates in limited response, withot censure and covariates in the point distribution located at cero;

$$g(y_i) = f^+(y_i; \mu_{i2}, \sigma) \tag{0.3}$$

$$\mu_{i2} = \beta_{02} + \beta_{12}EZ + \beta_{22}HI + \beta_{32}FEM \tag{0.4}$$

$$\tag{0.5}$$

MODEL C: with censored data, withot covariates in limited response and in the point distribution located at cero;

$$g(y_i) = \left(1 - q_i + q_i F^+(T; \mu_2, \sigma)\right)^{I_i} \times \left(q_i f^+(y_i; \mu_2, \sigma)\right)^{1-I_i} \tag{0.6}$$

$$q_i = H(\mu_1) \tag{0.7}$$

Models

MODEL D: with censored data and covariates in limited response, withot covariates in the point distribution located at cero;

$$g(y_i) = \left(1 - q_i + q_i F^+(T; \mu_{2i}, \sigma)\right)^{I_i} \times \left(q_i f^+(y_i; \mu_{i2}, \sigma)\right)^{1-I_i} \tag{0.8}$$
$$q_i = H(\mu_1) \tag{0.9}$$
$$\mu_{i2} = \beta_{02} + \beta_{12} EZ + \beta_{22} HI + \beta_{32} FEM \tag{0.10}$$

MODEL E: with censored data, withot covariates in limited response and with covariates in the point distribution located at cero;

$$g(y_i) = \left(1 - q_i + q_i F^+(T; \mu_2, \sigma)\right)^{I_i} \times \left(q_i f^+(y_i; \mu_2, \sigma)\right)^{1-I_i} \tag{0.11}$$
$$q_i = H(\mu_{1i}) \tag{0.12}$$
$$\mu_{i1} = \beta_{01} + \beta_{11} EZ + \beta_{21} HI + \beta_{31} FEM \tag{0.13}$$

MODEL F: with censure, covariates in limited response and in the point distribution located at cero;

$$g(y_i) = \left(1 - q_i + q_i F^+(T; \mu_{2i}, \sigma)\right)^{I_i} \times \left(q_i f^+(y_i; \mu_{i2}, \sigma)\right)^{1-I_i} \tag{0.14}$$

$$q_i = H(\mu_{i1}) \tag{0.15}$$

$$\mu_{i1} = \beta_{01} + \beta_{11}EZ + \beta_{21}HI + \beta_{31}FEM \tag{0.16}$$

$$\mu_{i2} = \beta_{02} + \beta_{12}EZ + \beta_{22}HI + \beta_{32}FEM \tag{0.17}$$

$$\tag{0.18}$$

MODEL G: with censored data, covariates in limited response and in the point distribution located at cero, particular model

$$g(y_i) = \left(1 - q_i + q_i F^+(T; \mu_{2i}, \sigma)\right)^{I_i} \times \left(q_i f^+(y_i; \mu_{2i}, \sigma)\right)^{1-I_i} \tag{0.19}$$

$$q_i = H(\mu_{i1}) \tag{0.20}$$

$$\mu_{i1} = \beta_{01} + \beta_{11}EZ + \beta_{21}HI \tag{0.21}$$

$$\mu_{i2} = \beta_{02} + \beta_{32}FEM \tag{0.22}$$

$$\tag{0.23}$$

Models

where

- H^{-1} is a link function. For example, Logit and Probit;
- $f^{+}()$ and $F^{+}()$ correspond to a probability density function and cumulative density function with positive support associated with the function to limit responses. For example, log-normal and truncated normal distributions;
- μ_{i1}, and μ_1 correspond to the parameter of location including or not covariates associated with the point distribution located at cero;
- μ_{i2} and μ_2 correspond to the parameter of location including or not covariates associated with the function to limit responses;
- σ is the correspond scale parameter associated with $f^{+}()$

2.1 THE LOG-ALPHA-POWER DISTRIBUTION

The distribution of the lifetime of an equipment and the concentration of a chemical element in soil (blood or water) samples is typically described by the log-normal distribution. In many of these situations, however, the asymmetry of the distribution as well as its kurtosis can be above or below the expected for the log-normal model. As an extension of the Azzalini (1985) model to positive data, Mateu-Figueras et al. (2003-2004) examines the log-skew-normal distribution, for which the log-normal (LN) distribution is a special case. Chai and Bailey (2008) extended the log-skew-normal (LSN) model to the analysis of continuos data with a discrete response.

However, due to singularity of the Fisher information matrix, one can not test for log-normality using the likelihood ratio statistics.
As an alternative to the skew normal model to fit data from such experimental situations, we consider the log-power-normal (LPN) distribution, which contains as a special case the log-normal model. The advantage of such model is that it has an extra shape parameter which makes it more flexible for fitting data coming from experiments such as the ones described above.

We say that a random variable Y, with support in R^+, follows a univariate log-alpha-power distribution with parameter α, that we denote $Y \sim LAP(\alpha)$, if the transformed variable $X = \log(Y) \sim AP(\alpha)$.
The pdf for the random variable $Y \sim LAP(\alpha)$ can be written as

$$\varphi_Y(y;\alpha) = \frac{\alpha}{y} f(\log(y)) \left\{F(\log(y))\right\}^{\alpha-1}, \quad y \in \mathbb{R}^+, \qquad (0.24)$$

where F is an absolutely continuos distribution function with density function $f = dF$. We refer to this model as the standard log-alpha-power distribution.

In the special case where $f = \phi(\cdot)$ and $F = \Phi(\cdot)$, the density and distribution functions of the standard normal distribution, respectively, the standard log-power-normal distribution follows, with density function given by

$$\varphi_Y(y;\alpha) = \frac{\alpha}{y}\phi(\log(y))\left\{\Phi(\log(y))\right\}^{\alpha-1}, \quad y \in \mathbb{R}^+, \tag{0.25}$$

denoted by $Y \sim LPN(\alpha)$. Its cumulative distribution function can be written as

$$\mathcal{F}_Y(y;\alpha) = \left\{\Phi(\log(y))\right\}^{\alpha}, \quad y \in \mathbb{R}^+. \tag{0.26}$$

According to (0.26), the inversion method can be used for generating from a random variable with distribution $LPN(\alpha)$. That is, if $U \sim U(0,1)$, the uniform distribution, then random variable $Y = e^{\Phi^{-1}(U^{1/\alpha})}$ is distributed according to the LPN distribution with parameter α.
Let $X \sim PN(\xi, \eta, \alpha)$, where $\xi \in \mathbb{R}$ is a location parameter and $\eta \in \mathbb{R}^+$ is a scale parameter. Hence, transformation $X = \ln(Y)$ leads to the location-scale log-power-normal model. We use the notation $Y \sim LPN(\xi, \eta, \alpha)$.

In the particular case of $\alpha = 1$, that is, $Z = \frac{\log(Y) - \xi}{\eta} \sim N(0,1)$, it can be shown (after lengthy algebraic manipulations) that the Fisher information matrix for $\theta = (\xi, \eta, \alpha)'$ is given by

$$I(\theta) = \begin{pmatrix} 1/\eta^2 & 0 & a_{01}/\eta \\ 0 & 2/\eta^2 & a_{11}/\eta \\ a_{01}/\eta & a_{11}/\eta & 1 \end{pmatrix},$$

where $a_{kj} = E\{z^k(\phi(z)/\Phi(z))^j\}$ for $k = 0, 1, 2, 3$ and $j = 1, 2$, which agrees with the Fisher information matrix for the power-normal distribution (Pewsey et al., 2012).

Hence, using numerical procedures, it can be shown that

$$|I(\theta)| = [2 - (a_{11}^2 + 2a_{01}^2)]/\eta^4 \neq 0,$$

so that the Fisher information matrix is not singular at $\alpha = 1.0$. The complete Fisher information matrix is available from the authors upon request. Hence, for n large,

$$\hat{\theta} \xrightarrow{A} N_3(\theta, I(\theta)^{-1}),$$

meaning that $\hat{\theta}$, the maximum likelihood estimator of θ, is consistent and asymptotically normally distributed with $I(\theta)^{-1}$ as the large sample variance.

This important feature allows testing (with the LPN model) log-normality (that is, $H_0 : LPN = LN$), using the ordinary large sample property of the likelihood ratio statistics, which states that in large sample it follows a chi-square distribution. This is not, however, the case with the LSN distribution, for which the Fisher information matrix is singular. Another important difference between the two models (LSN and LPN) concerns asymmetry and kurtosis ranges. While the kurtosis range is wider for the LSN model, the asymmetry range is wider for the LPN model. Such differences allow one to select a more appropriate model.

Haiti Data:

We test now the hypothesis of no difference between the Bernolli/log-power-normal model with the traditional Bernolli/log-normal (Molton and Halsey, 1995) by testing the hypothesis

$$H_0: \ \alpha = 1 \ \text{VS} \ H_1: \ \alpha \neq 1$$

using the likelihood ratio statistics, that is,

$$\Lambda = \frac{L_{Bernolli-lognormal}(\hat{\theta})}{L_{Bernolli-LPSN}(\hat{\theta})}$$

we have after replacing MLEs for each model that

$$-2\log(\Lambda) = -2(-557.9150 + 544.6657) = 13.2493,$$

which is greater than the 5% critical value 3.84. Hence, the combination Bernolli-log-normal is clearly rejected in favor of the more general Bernolli/log-power-normal model.

Model	Method	Bernolli Component				Lognormal Component			
		INT	EZ	HI	FEM	INT	EZ	HI	FEM
A	Clas					-0.979			
	Bay					-0.981			
B	Clas					-1.287	0.340	0.182	0.115
	Bay					-0.932	0.203	0.097	0.114
C	Clas	1.198				-0.273			
	Bay	1.227				-0.285			
D	Clas	1.178				-0.327	-0.109	-0.037	0.290
	Bay	1.226				-0.361	-0.083	-0.025	0.277
E	Clas	0.732	0.843	0.431	-0.166	-0.274			
	Bay	0.813	0.950	0.445	-0.244	-0.305			
F	Clas	0.765	0.932	0.433	-0.281	-0.304	-0.192	-0.063	0.329
	Bay	0.910	1.112	0.439	-0.425	-0.353	-0.199	-0.055	0.339
G	Clas	0.648	0.830	0.426		-0.404			0.279
	Bay	0.678	0.893	0.440		-0.421			0.266

Parameter estimates from fits of one and two component mixture models by considering Classical and Bayesian approach to Haiti data.
EZ (vaccine strain,0: Schwarz, 1: Edmonston-Zagreb); HI (vaccine dose, 0: medium, 1: high); and FEM (sex; 0: male,1:female). The constant term is denoted INT.

Model	$-2 \times loglik$	df	DIC	pD
A	1115.830	328	136.600	1.89
B	111.180	325	-580.5	5.17
C	1079.320	327	101.800	2.7
D	1075.620	324	104.500	5.79
E	1068.720	324	95.560	5.08
F	1063.360	321	99.470	9.07
G	1065.810	324	93.840	5.42

Model comparison for Haiti data by considering a Classical and Bayesian Criteria

The Bayesian estimator for λ using extension of Bazan's program for log-skew-normal is $\hat{\lambda} = 17.01$ with $CI = (7.95; 33.47)$ using prior $U(0, 100)$. Hence strong indication that asymmetric model fits data better.

MODELO DE REGRESSAO BETA

Modelo de regeresson beta es usada quando a variável resposta Y esta no intervalo $(0, 1)$.
Sabemos que distribucion $beta(\alpha, \beta)$

$$f(y|\alpha, \beta) = \frac{\Gamma(\alpha, \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} I_{(0,1)}(y),$$

$\alpha > 0$. $\beta > 0$.
Propriedades:

$$E[Y] = \frac{\alpha}{\alpha+\beta}, \quad Var[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Usando parametrizacao natural fica complicado construir modelo de regresson.

Dada muestra $Y_1, \ldots, Y_n$, tenemos a funcion de verossimilhança

$$L(\alpha, \beta|\mathbf{Y}) = \prod_{i=1}^{n} \frac{\Gamma(\alpha, \beta)}{\Gamma(\alpha)\Gamma(\beta)} y_i^{\alpha-1}(1 - y_i)^{\beta-1} I_{(0,1)}(y),$$

$\alpha > 0, \beta > 0.$
Complicado definir priori conjugada e verosimilitud complicada de se tratar.

Transformacion Ferrari/Cribari-Neto

$$\mu = \frac{\alpha}{\alpha + \beta} \qquad \phi = \alpha + \beta,$$

que leva a funcion de densidad de Probabilidad

$$f(y|\mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\phi\mu)\Gamma(\phi(1-\mu))} y^{\mu\phi - 1}(1-y)^{\phi(1-\mu)-1} I_{(0,1)}(y),$$

com

$$E[Y] = \mu \qquad \mathrm{a}nd \qquad Var[Y] = \frac{\mu(1-\mu)}{1+\phi}.$$

Componentes sistemáticos:

$$g_1(\mu_i) = \mathbf{x}_i'\beta, \qquad g_2(\phi_i) = \mathbf{z}_i'\alpha$$

Priori: $\pi(\beta, \alpha)$

ejemplo: Avaliar contribucion do nível de renda e número de pessoas em uma família na proporcion de renda familiar gasta com alimentacion.
Seja
Y_i : Proporcion renda gasta com alimentacion i-essima família

Modelo: $Y_i \sim Beta(\mu_i, \phi)$

Sistemática: $\log(\frac{\mu_i}{1-\mu_i}) = \beta_0 + \beta_1 x_{1i}^c + \beta_2 x_{2i}^c,$

x_{1i} : nível de renda (c: centrada)

x_{2i}: número de pessoas (c: centrada)

Priori: $\beta \sim N_3(\mathbf{0}, k\mathbf{I}_3)$ and $\phi \sim N(0,k)I_{(0,\infty)}$, $k = 10^3$.

Conditional predictive ordinate (CPO)

Among the variety of methodologies to compare several models for a given data set, the conditional predictive ordinate (CPO) approach is one of the most used. This statistics is based on a cross validation criterion to compare the models. For a detailed discussion on properties of the CPO_i statistics (for the i-th observation) and application to some models, see [**?**].

For the $i-$th observation, the CPO_i can be written as:

$$CPO_i = p\left(y_i|y_{(-i)}\right) = \int_{\theta\in\Theta} f\left(y_i|\theta\right)\pi\left(\theta|y_{(-i)}\right)d\theta$$

$$= \left\{\int_{\theta\in\Theta} \frac{\pi\left(\theta|\mathbf{y}\right)}{f\left(y_i|\theta\right)}d\theta\right\}^{-1},$$

where $p\left(y_i|y_{(-i)}\right)$ represents the predictive distribution of y_i given the other observations in the sample and $y_{(-i)} = \left(y_1, y_2, \cdots, y_{i-1}, y_{i+1}, \cdots, y_n\right)$.

These values are an indicator of the likelihood value of each observation given all the other observations, and therefore low values of CPO_i must correspond to poorly fitting observations. To prove (4), notice that

$$
\begin{aligned}
CPO_i = p\left(y_i|\mathbf{y}_{(-i)}\right) &= \frac{p\left(\mathbf{y}\right)}{p\left(\mathbf{y}_{(-i)}\right)} \\
&= \frac{1}{\int_{\theta\in\Theta} \frac{f\left(\mathbf{y}_{(-i)}|\theta\right)\pi(\theta)\pi(\theta|\mathbf{y})}{p(\mathbf{y})\pi(\theta|\mathbf{y})} d\theta} \\
&= \frac{1}{\int_{\theta\in\Theta} \frac{1}{f\left(y_i|\mathbf{y}_{(-i)},\theta\right)} \pi\left(\theta|\mathbf{y}\right) d\theta} \\
&= \frac{1}{\int_{\theta\in\Theta} \frac{1}{f(y_i|\theta)} \pi\left(\theta|\mathbf{y}\right) d\theta} \\
&= \left\{ \int_{\theta\in\Theta} \frac{\pi\left(\theta|\mathbf{y}\right)}{f\left(y_i|\theta\right)} d\theta \right\}^{-1}.
\end{aligned}
$$

For the proposed model a closed form of the (CPO), is not available. However, a Monte Carlo estimate of the CPO_i can be obtained by using a single MCMC sample from the posterior distribution $\pi(\theta|\mathbf{y})$. Then, a Monte Carlo approximation of the CPO_i - see [**?**] - for a sample of size M is given by:

$$\widehat{CPO}_i = \left\{ \frac{1}{M} \sum_{m=1}^{M} \frac{1}{f(y_i|\theta_m)} \right\}^{-1}.$$

A summary statistic of CPO_i values is the log-marginal psudo likelihood: $LMPL = \sum_{i=1}^{n} Log\left(\widehat{CPO}_i\right)$, so that the larger is the value of CPO, the better is the fit of the model under consideration.

cadenas de Markov em tempo discreto:

cadenas de Markov: Sequência de variables aleatórias com a propriedade de Markov:

$$P[X_{n+1} = x | X_1 = x_1, \ldots, X_n = x_n] = P[X_{n+1} = x | X_n = x_n].$$

cadena es estacionária se

$$P[X_{n+1} = x | X_n = y] = P[X_n = x | X_{n-1} = y].$$

Tambien se puede definir MC de orden m, donde o estado futuro depende dos m estados pasados. Para $m = 1$ tenemos a MC usual.

Probabilidad de ir do estado i para o estado j em n pasos:

$$p_{ij}^{(n)} = P[X_n = j | X_0 = i].$$

Para transcion a um passo tenemos

$$p_{ij} = P[X_1 = j | X_0 = i].$$

Para cadenas estacionárias, para todo k,

$$p_{ij}^{(n)} = P[X_{k+n} = j | X_k = i]$$

e

$$p_{ij} = P[X_{k+1} = j | X_k = i].$$

Equa¸oes de Chapman-Kolmogorov:

$$p_{ij}^{(n)} = \sum_{r \in S} p_{ir}^{(k)} p_{rj}^{n-k},$$

donde S son os estados da cadena.
Distribucion Marginal ($P(X_n = x)$): es as distribucion sobre os estados no tempo n:

$$P[X_n = j] = \sum_{r \in S} p_{rj} P[X_{N-1} = r] = \sum_{r \in S} p_{rj}^{(n)} P[X_0 = r]$$

Se

$$P[X_n = j | X_0 = i] = p_{ij}^{(n)} > 0,$$

j es acessível a partir de i.

cadena es irredutível se todos os estados se comunicam.

Si el proceso es estacionário, proceso descrito por la matriz $P = (p_{ij})$ (matriz de Probabilidades de transicion), entonces vector Π es llamado distribucion estacionária (o medida invariante) si los elementos π_{ij} de Π son não negativos suman 1 e satisfacen:

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}.$$

Podemos usar a ecuacion arriba na forma

$$\pi = P\pi,$$

P a matriz de transcion.

Função de transcion no muestrador de Gibbs para $\theta = (\theta_1, \theta_2)$:

$$k(\theta^{(t-1)}, \theta^{(t)}) = f_{\theta_1|\theta_2}(\theta_1^{(t)}|\theta_2^{(t-1)}) f_{\theta_2|\theta_1}(\theta_2^{(t)}|\theta_1^{(t)}).$$

Posteriori conjunta:

$$f(\theta_1, \theta_2) = f(\theta)$$

Invariância da cadena generada por el muestrador de Gibbs:

$$\int f(\theta_1^{(t-1)}, \theta_2^{(t-1)}) k(\theta^{(t-1)}, \theta^{(t)}) d\theta_1^{(t-1)} d\theta_2^{(t-1)} =$$

$$= \int f(\theta_1^{(t-1)}, \theta_2^{(t-1)}) d\theta_1^{(t-1)} f_{\theta_1|\theta_2}(\theta_1^{(t)}|\theta_2^{(t-1)}) f_{\theta_2|\theta_1}(\theta_2^{(t)}|\theta_1^{(t)}) d\theta_2^{(t-1)}$$

$$= \int f(\theta_2^{(t-1)}) f_{\theta_1|\theta_2}(\theta_1^{(t)}|\theta_2^{(t-1)}) f_{\theta_2|\theta_1}(\theta_2^{(t)}|\theta_1^{(t)}) d\theta_2^{(t-1)}$$

$$= f(\theta_1^{(t)}, \theta_2^{(t)}).$$

Con respecto al ultimo integrando,

$$f(\theta_2^{(t-1)}) f_{\theta_1|\theta_2}(\theta_1^{(t)}|\theta_2^{(t-1)}) =$$

$$= \frac{f(\theta_2^{(t-1)}) f(\theta_1^{(t)}, \theta_2^{(t-1)})}{f(\theta_2^{(t-1)})} = f(\theta_1^{(t)}, \theta_2^{(t-1)}).$$

ejemplo. Considere o modelo genestico com cuatro categorias, distribuídas de acordo com as Probabilidades:

$$p_1 = \frac{2+\theta}{4}, \quad p_2 = \frac{1-\theta}{4}, \quad p_3 = \frac{1-\theta}{4}, \quad p_4 = \frac{\theta}{4},$$

com $0 \leq \theta \leq 1$. Para uma muestra de N animais, seja Y_i do tipo i, $i = 1, \ldots, 4$, com $\sum Y_i = N$. Considere priori $\theta \sim Beta(a, b)$. A distribcion la posteriori para θ pode ser escrita como

$$\pi(\theta|\boldsymbol{x}) \propto (2+\theta)^{y_1}(1-\theta)^{y_2+y_3+b-1}\theta^{y_4+a-1}.$$

Para $n = 197$, $\boldsymbol{Y} = (125, 18, 20, 34)$, tenemos com $a = b = 1$, o estimador modal $\hat{\theta} = 0.6268$ com

$$\hat{\sigma}^2 = [-\frac{d^2 \log \pi(\theta|\boldsymbol{x})}{d\theta^2}]^{-1}_{\theta=\hat{\theta}} = 1/377.52 = 0.002649.$$