Familia Exponencial

Propiedades
Una distribución pertenece a la familia exponencial si su función de densidad puede expresarse como:

$$f(y;\theta) = s(y)t(\theta)exp[a(y)b(\theta)] = exp[a(y)b(\theta)+c(\theta)+d(y)] \quad (1)$$

con $s(y) = exp[d(y)]$ y $t(\theta) = exp[c(\theta)]$.
La distribución depende de un solo parámetro, a θ se le conoce como parámetro de la familia.

Si

$$a(y) = y$$

se dice que la distribución está expresada en forma CANÓNICA y a

$$b(\theta)$$

se le conoce como el PARÁMETRO NATURAL.
Si hay otros parámetros se les considera como de ruido y se tratan como si fueran conocidos.

$$\int f(y;\theta) = 1$$

Si derivo

$$d/d\theta \int f(y;\theta)dy = 0$$

Si f permite intercambiar el signo de derivación e integración

$$\int df(y;\theta)dy/d\theta = 0$$

Si vuelvo a derivar

$$d/d\theta \int df(y;\theta)dy/d\theta = \int d^2 f(y;\theta)dy/d\theta^2 = 0$$

Usando la expresión (1)

$$df(y;\theta)/d\theta = f(y;\theta)[a(y)b'(\theta) + c'(\theta)] = 0 \tag{2}$$

y ahora integrando

$$\int [a(y)b'(\theta) + c'(\theta)]f(y;\theta)dy = 0$$

Ahora distribuyo la multiplicación

$$b'(\theta)\int a(y)f(y;\theta)dy + c'(\theta)\int f(y;\theta)dy$$

Y tengo una expresión para la esperanza de $E(a(y))$

$$E(a(y)) = -c'(\theta)/b'(\theta) \tag{3}$$

Para llegar a una expresión para la $Var(a(y))$ haremos un desarrollo semejante

$$d^2 f(y;\theta)/d\theta^2 = f(y;\theta)[a(y)b''(\theta)+c''(\theta)]+f'(y;\theta)[a(y)b'(\theta)+c'(\theta)]$$

usando la expresión para $f'(y;\theta)$

$$= f(y;\theta)[a(y)b''(\theta) + c''(\theta)] + f(y;\theta)[a(y)b'(\theta) + c'(\theta)]^2$$

Ahora trabajo el cachito

$$[a(y)b'(\theta)+c'(\theta)]^2 = [b'(\theta)[a(y)+c'(\theta)/b'(\theta)]]^2 = b'(\theta)^2[a(y)-E(a(y))]^2$$

Integrando

$$\int d^2 f(y;\theta)/d\theta^2$$

$$= b''(\theta) \int a(y) f(y;\theta) dy + c''(\theta) \int f(y;\theta) dy$$

$$+[b'(\theta)]^2 \int [a(y) - E(a(y))]^2 f(y;\theta) dy = 0$$

$$Var(a(y)) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3} \qquad (4)$$

Logverosimilitud

Esta es la expresión para la log verosimilitud

$$l(\theta; y) = a(y)b(\theta) + c(\theta) + d(y)$$

Se llama *score statistic* a U

$$U(\theta; y) = dl(\theta; y)/d\theta = a(y)b'(\theta) + c'(\theta) = U$$

U depende de y y se le puede ver como una variable aleatoria, entonces calcularemos su esperanza:

$$E(U) = E(a(y)b'(\theta) + c'(\theta)) = b'(\theta)E(a(y)) + c'(\theta)$$

usando la expresión (3)

$$E(U) = b'(\theta)[-c'(\theta)/b'(\theta)] + c'(\theta) = 0$$

Y para su varianza usamos la expresión (4)

$$
\begin{aligned}
Var(U) = Var(a(y)b'(\theta) + c'(\theta)) &= [b'(\theta)]^2 Var(a(y)) \\
&= b''(\theta) \times c'(\theta)/b'(\theta) - c''(\theta) = \mathcal{I}
\end{aligned}
$$

También se tiene, $Var(U) = E[(U - 0)^2] = E[U^2] = \mathcal{I}$

Calculando U'

$$U' = dU/d\theta = a(y)b''(\theta) + c''(\theta)$$

y

$$E(U') = b''(\theta)E(a(y)) + c''(\theta) = b''(\theta)[-c'(\theta)/b'(\theta)] + c''(\theta)$$

$$= -Var(U) = -\mathcal{I}$$

La $\mathcal{I}$ es la matriz de información.
En el caso de que U sea univariadad y para n grande $U/\sqrt(\mathcal{I})$ se distribuye aproximadamente como $N(0, \mathcal{I})$ Desde luego cuando generalizamos y θ es un vector de la forma $(\beta_0, \beta_1, \ldots, \beta_p)$, se tiene normalidad en cada una de sus entradas.

Modelo Lineal Generalizado

1. **Componente aleatoria**. La variable respuesta Y_i tiene una distribución que pertenece a la familia exponencial (en forma canónica)

   $$f_i = f(y_i; \theta_i) = exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)]$$

   Todas las Y_i son independientes.
2. **Componente sistemática**. Un vector $\eta_i = (\eta_1, \ldots, \eta_n)$ que se relaciona con las variables explicativas $X = (X_1, \ldots, X_k)$ a través de

   $$\eta_i = x_i' \beta$$
3. **Función liga**. Existe una función g, diferenciable y monótona (y por tanto tiene inversa), tal que $g(\mu_i) = g(E(Y_i)) = x_i' \beta$.

A g se le conoce como función **liga**
y $x_i' = (x_{i1}, \ldots, x_{ik})$ es el vector de variables explicativas X

Estimación

Para cada y_i

$$l_i = y_i b(\theta) + c(\theta_i) + d(y_i)$$

$$E(y_i) = \mu_i = -c'(\theta_i)/b'(\theta_i)$$

$$Var(y_i) = \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{[b'(\theta_i)]^3}$$

$$g(\mu_i) = x_i'\beta = \eta_i$$

La logverosimilitud es

$$l = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i)$$

Para maximizar derivo:

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{n} \left[\frac{\partial l_i}{\partial \beta_j} \right] = \sum_{i=1}^{n} \left[\underbrace{\frac{\partial l_i}{\partial \theta_i}}_{1} \underbrace{\frac{\partial \theta_i}{\partial \mu_i}}_{2} \underbrace{\frac{\partial \mu_i}{\partial \beta_j}}_{3} \right]$$

La derivada parcial 1

$$\frac{\partial l_i}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = y_i b'(\theta_i) - b'(\theta_i)\mu_i = b'(\theta_i)(y_i - \mu_i)$$

La derivada parcial 2

$$\frac{\partial \theta_i}{\partial \mu_i} = 1/\frac{\partial \mu_i}{\partial \theta_i}$$

$$\frac{\partial \mu_i}{\partial \theta_i} = \big[-c'(\theta_i)/b'(\theta_i)\big]' = -\left[\frac{c''(\theta_i)b'(\theta_i) - c'(\theta_i)b''(\theta)}{[b'(\theta_i)^2]}\right]$$

$$= b'(\theta_i)Var(y_i)$$

la derivada parcial 3

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_j}\frac{\partial \eta_j}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_j} x_{ij}$$

Reuniendo esto

$$U_j = \sum_{i=1}^{n} (y_i - \mu_i) b'(\theta_i) * \frac{1}{b'(\theta_i) Var(y_i)} * \frac{\partial \mu_i}{\partial \eta_j} x_{ij} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{Var(y_i)} \frac{\partial \mu_i}{\partial \eta_j} x_{ij}$$

Solo se presentará la matriz de Información $\mathcal{I} = [E(U_j U_k)]$

$$\mathcal{I}_{jk} = \sum_{i=1}^{n} \frac{E[(y_i - \mu_i)^2]}{Var(y_i)^2} x_{ik} x_{ij} \frac{\partial \mu_i}{\partial \eta_j}^2 = \sum_{i=1}^{n} \frac{x_{ik} x_{ij}}{Var(y_i)} \left[\frac{\partial \mu_i}{\partial \eta_j} \right]^2$$

Esta expresión es así de sencilla debido a que las observaciones son independientes $(E[(y_i - \mu_i)(y_s - \mu_s)] = 0)$

Método de Newton-Raphson

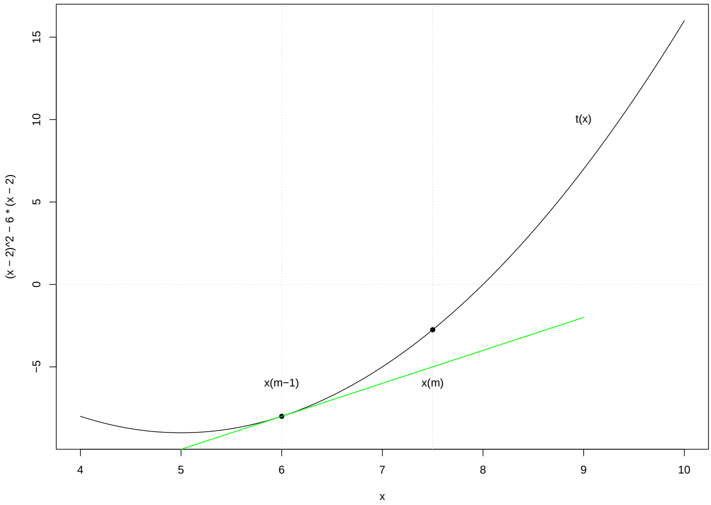
Para cualquier función $t(x)$ si se desea encontrar el punto x tal que $t(x) = 0$ se puede hacer lo siguiente: Si se considera que la distancia entre x^{m-1} y x^m es pequeña

$$\left[\frac{dt}{dx}\right]_{x=x^{m-1}} = t'(x^{m-1}) = \frac{t(x^m) - t(x^{m-1})}{x^m - x^{m-1}}$$

Si x^m es la solución a $t(x) = 0$ entonces

$$t'(x^{m-1}) = \frac{0 - t(x^{m-1})}{x^m - x^{m-1}} \Rightarrow$$

$$x^m = x^{m-1} - \frac{t(x^{m-1})}{t'(x^{m-1})}$$



t(x)
x(m−1)
x(m)
(x − 2)^2 − 6 * (x − 2)
x
−5
0
5
10
15
4
5
6
7
8
9
10

Entonces regresando al tema de estimación de β, lleva a encontrar el cero de U, y se conoce esto como método *score* y la podemos estimar con:

$$\hat{\beta}^{m} = \hat{\beta}^{m-1} - \frac{U^{m-1}}{U'^{m-1}} \approx \hat{\beta}^{m-1} - \frac{U^{m-1}}{E(U'^{m-1})} = \hat{\beta}^{m-1} + \frac{U^{m-1}}{\mathcal{I}^{m-1}}$$

También es válido usar

$$\mathcal{I}^{m-1}\hat{\beta}^{m} = \mathcal{I}^{m-1}\hat{\beta}^{m-1} + U^{m-1} \tag{5}$$

Basta entonces tener una solución inicial $\hat{\beta}^{(0)} = (\hat{\beta}_o^{(0)}, \hat{\beta}_1^{(0)}, \cdots, \hat{\beta}_p^{(0)})$y después iterar con esta fórmula que depende de U y de $\mathcal{I}$

Para cualquier modelo GLM las fórmulas para U_j y $\mathcal{I}_{jk}$ son:

$$U_j = \sum_{i=1}^{n} \frac{x_{ij}(y_i - \mu_i)}{Var(y_i)}(\frac{\partial \mu_i}{\partial \eta_i})$$

y

$$\mathcal{I}_{jk} = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{Var(y_i)}(\frac{\partial \mu_i}{\partial \eta_i})^2$$

otra forma de escribir la información es:

$$\mathcal{I}_{jk} == X'WX$$

donde W es una matriz diagonal de $n \times n$ con

$$w_{ii} = \frac{1}{Var(y_i)} \left[\frac{\partial \mu_i}{\partial \eta_j}\right]^2$$

El lado derecho de 5 es un vector y puede escribirse como

$$X'Wz$$

donde

$$z_i = \sum_{k=1}^{p} x_{ik}\beta_k^{(m-1)} + (y_i - \mu_i)(\frac{\partial \eta_i}{\partial \mu_i})$$

y con μ_i y $\frac{\partial \mu_i}{\partial \eta_i}$ evaluadas en $\beta^{(m-1)}$.

finalmente 5 puede escribirse como

$$X'WX\beta^{(m)} = X'Wz$$

Y tiene la forma de un sistema de ecuaciones normales para un modelo lineal obtenido por mínimos cuadrados ponderados, excepto porque z y W dependen de β.
La mayoría de los paquetes utiliza un algoritmo basado en 5.

Ejemplo del modelo logístico

La variable y_i se distribuye como binomial $B(n_i, p_i)$, de ahí que $E(y_i) = n_i p_i$; la función liga es el logit de p_i, esto es:

$$\eta_i = \log\left(\frac{p_i}{(1-p_i)}\right) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}$$

Para estimar a β se maximiza la verosimilitud

$$L(\beta) = \prod_{i=1}^{n} \left(\begin{array}{c} n_i \\ p_i \end{array} \right) p_i^{y_i} (1-p_i)^{n_i - y_i}$$

ésta depende de p_i que a su vez depende de β

$$\log(L(\beta)) = \sum_{i=1}^{n} \{\log \left(\begin{array}{c} n_i \\ p_i \end{array} \right) + y_i \log(p_i) + (n_i - y_i) \log(1 - p_i)\}$$

$$= \sum_{i=1}^{n} \{\log \left(\begin{array}{c} n_i \\ p_i \end{array} \right) + y_i \log(\frac{p_i}{1 - p_i}) + n_i \log(1 - p_i)\}$$

usando que

$$e^{\eta_i} = \frac{p_i}{1 - p_i}$$

$$1 + e^{\eta_i} = 1 + \frac{p_i}{1 - p_i} = \frac{1}{1 - p_i}$$

$$\log(1 + e^{\eta_i}) = -\log(1 - p_i)$$

se tiene que

$$= \sum_{i=1}^{n} \{\log \left(\begin{array}{c} n_i \\ p_i \end{array} \right) + y_i \eta_i - n_i \log(1 + e^{\eta_i})\}$$

Como $\eta_i = \sum_{i=1}^{n} \beta_j x_{ji}$ con $x_{0i} = 1$ y calculando la derivada se tiene:

$$U_j = \partial \log(L(\beta))/\partial\beta_j = \partial(l(\beta))/\partial\beta_j = \sum_{i=1}^{n} y_i x_{ij} - \sum_{i=1}^{n} \frac{n_i x_{ij} e^{\eta_i}}{(1 + e^{\eta_i})}$$

$$U_j = \sum_{i=1}^{n} (y_i x_{ij} - n_i p_i x_{ij}) = \sum_{i=1}^{n} (x_{ij}(y_i - n_i p_i)) = \sum_{i=1}^{n} x_{ij}(y_i - \mu_i)$$

Ahora para calcular $\mathcal{I}_{jk} = E(U_j U_k)$

$$E(U_j U_k) = E(\sum_{i=1}^{n} x_{ij}(y_i - \mu_i) * \sum_{i=1}^{n} x_{ik}(y_i - \mu_i)) =$$

Usando el hecho que las observaciones son independientes, es decir que $E[(y_i - \mu_i)(y_s - \mu_s)] = 0$ finalmente se tiene:

$$E(U_j U_k) = \sum_{i=1}^{n} x_{ij} x_{ik} E(y_i - \mu_i)^2 = \sum_{i=1}^{n} x_{ij} x_{ik} Var(y_i)$$

$$= \sum_{i=1}^{n} x_{ij} x_{ik} n_i p_i (1 - p_i) = \mathcal{I}_{jk}$$

Ya se tienen las expresiones de U y de $\mathcal{I}$ para poder usar el método de Newton Raphson.