Casi-verosimilitud (Quasi-Likelihood)

El concepto de parámetro de dispersión es también usado en el método de *quasi-likelihood* (QL). La técnica QL es usada para estimar coeficientes de regresión sin especificar completamente la distribución de los datos observados.

<u>Introducción</u>

Recuerde que la función de score de un MLG está dada por:

$$U_j = \sum_{i=1}^{n} \left[\frac{y_i - \mu_i}{\text{Var}[Y_i]} \cdot \frac{\partial \mu_i}{\partial \beta_j} \right],$$

$j = 1, \ldots, p$.

Teniamos la relación $\text{Var}[Y_i] = \phi V(\mu_i)$, donde ϕ es el parámetro de dispersión y V es la *función de varianza* que describe cómo la varianza de Y_i se relaciona a su media. Entonces, en la forma matricial,

$$\mathbf{U}(\boldsymbol{\beta}; \mathbf{y}) = \mathbf{D}' \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}) / \phi, \qquad (1)$$

donde $\mathbf{D}$ es la matriz $n \times p$ con entradas $(i,j)^{th}$ $\partial \mu_i / \partial \beta_j$, $\mathbf{V}$ es la i^{th} matriz diagonal

$n \times n$ con entradas $V(\mu_i)$, $\boldsymbol{y} = (y_1, \ldots, y_n)$, y $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$.

La *función quasi-score* es idéntica a (1). Podemos usar la funcipon quasi-score de la misma forma que usabamos la verdadera función score para estimar los parámetros.

La diferencia clave entre estos dos métodos de estimación es que para EMV, escogemos la *distribución* de las Y_i's, y derivamos la función score (incluyendo la varianza de Y_i's) como una consecuencia de la distribución.

Para estimaciones por QL, comenzamos con la función score general (1) y escogemos la *función de varianza* de las Y_i's.

Note que los métodos de estimación por QL y MV también requieren que se especifique la relación entre la media y las variables explicativas, es decir:

$$g(\mu_i) = \left(\sum_{j=1}^{p} x_{ij}\beta_j \right),$$

entonces podemos calcular $\mathbf{D}$ en (1).

En general, la función quasi-score puede no corresponder con la verdadera función de verosimilitud, en cuyo caso los $\hat{\beta}_j$'s *no* son los EMVs.

Sin embargo, si la función QL si corresponde a una verdadera verosimilitud para alguna distribución conocida, entonces los estimadores por QL estimators son idénticos a los EMVs basados en esta distribución.

Motivación

Note que, independiente de la elección de $\mathbf{V}$ y g en (1),

1.
$$\begin{aligned} \mathrm{E}[\mathbf{U}(\boldsymbol{\beta};\mathbf{Y})] &= \mathrm{E}[\mathbf{D}'\mathbf{V}^{-1}(\mathbf{Y}-\boldsymbol{\mu})/\phi] \\ &= \mathbf{D}'\mathbf{V}^{-1}\mathrm{E}[\mathbf{Y}-\boldsymbol{\mu}]/\phi \\ &= \mathbf{0} \end{aligned}$$

2. Usando la propiedad de que para una matriz de constantes $\mathbf{A}$, y un vector de variables aleatorias, $\mathbf{Y}$, $\mathrm{Var}[\mathbf{A}\mathbf{Y}] = \mathbf{A}\mathrm{Var}[\mathbf{Y}]\mathbf{A}'$:

$$\begin{aligned} \mathrm{Var}[\mathbf{U}(\boldsymbol{\beta};\mathbf{Y})] &= \mathrm{Var}[\mathbf{D}'\mathbf{V}^{-1}(\mathbf{Y}-\boldsymbol{\mu})/\phi] \\ &= \frac{1}{\phi^2}(\mathbf{D}'\mathbf{V}^{-1})(\phi\mathbf{V})(\mathbf{D}'\mathbf{V}^{-1})' \\ &= \frac{1}{\phi}\mathbf{D}'\mathbf{V}^{-1}\mathbf{V}(\mathbf{V}^{-1}\mathbf{D}) \\ &= \frac{1}{\phi}\mathbf{D}'\mathbf{V}^{-1}\mathbf{D}. \end{aligned}$$

Además, podemos calcular:

$$\begin{aligned}
\frac{\partial U_j}{\partial \beta_k} &= \sum_{i=1}^{n} \frac{\partial}{\partial \beta_k}\left[\frac{Y_i-\mu_i}{\mathrm{Var}[Y_i]}\cdot\frac{\partial \mu_i}{\partial \beta_j}\right] \\
&= \sum_{i=1}^{n} \frac{-\frac{\partial \mu_i}{\partial \beta_k}\cdot \mathrm{Var}[Y_i]-(Y_i-\mu_i)\left(\frac{\partial}{\partial \beta_k}\mathrm{Var}[Y_i]\right)}{\mathrm{Var}^2[Y_i]}\frac{\partial \mu_i}{\partial \beta_j}+\frac{Y_i-\mu_i}{\mathrm{Var}[Y_i]}\cdot\frac{\partial \mu_i^2}{\partial \beta_j \partial \beta_k}
\end{aligned}$$

entonces

$$\begin{aligned}
\mathrm{E}\left[\frac{\partial U_j}{\partial \beta_k}\right] &= \frac{-\frac{\partial \mu_i}{\partial \beta_k}\cdot \mathrm{Var}[Y_i]\frac{\partial \mu_i}{\partial \beta_j}}{\mathrm{Var}^2[Y_i]} \\
&= \frac{-\frac{\partial \mu_i}{\partial \beta_k}\cdot\frac{\partial \mu_i}{\partial \beta_j}}{\phi V(\mu_i)}
\end{aligned}$$

y por lo tanto

$$\begin{aligned}
\mathrm{E}\left[\frac{\partial \mathbf{U}(\boldsymbol{\beta};\mathbf{Y})}{\partial \boldsymbol{\beta}}\right] &= -\frac{1}{\phi}\mathbf{D}'\mathbf{V}^{-1}\mathbf{D} \\
&= -\mathrm{Var}[\mathbf{U}(\boldsymbol{\beta};\mathbf{Y})].
\end{aligned}$$

En otras palabras, la función quasi-score tiene estas dos propiedades importantes de la verdadera función de score.

Por estas razones la función quasi-score se comporta de forma muy similar a la verdadera función score.

Sea

$$Q(\boldsymbol{\beta};\mathbf{y}) = \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(t)} dt.$$

Note que

$$\frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta};\mathbf{y}) = \mathbf{U}(\boldsymbol{\beta};\mathbf{y}).$$

Por esta razón, la función Q es llamada de *quasi-likelihood*, o, de forma más precisa, la *log quasi-likelihood*.

Resulta que Q se comporta de forma similar a la función de log-verosimilitud en un ajuste paramétrico.

PUNTO CLAVE: Podemos usar la función quasi-score para estimar $\boldsymbol{\beta}$ de la misma forma que usamos la función score bajo un enfoque paramétrico. En particular, podemos encontrar el valor $\hat{\boldsymbol{\beta}}_{QL}$ que maximiza Q definiendo $\mathbf{U}(\hat{\boldsymbol{\beta}}_{QL};\mathbf{y}) = 0$ y resolviendo para $\hat{\boldsymbol{\beta}}_{QL}$. **Sin embargo, Q puede o no corresponder a una fdp conjunta válida** $Y_1, \ldots, Y_n$.

Cuándo usar QL

Para permitir sobre-o sub dispersión relativa a las distribuciones Poisson o binomial.

En este caso, deberiamos usualmente comenzar con una función de varianza de una Poisson ($V(\mu) = \mu$) o de una binomial ($V(\mu) = \mu(1 - \mu)$), como una extensión preliminar del MLG original.

Luego puede estimarse el parámetro de dispersión ϕ, como una forma de evaluar la cantidad de sobre o sub dispersión.

Propiedades asintóticas de los estimadores QL

Recuerde que para un MLG con vector de coeficientes de regresión $\boldsymbol{\beta}$, el EMV $\boldsymbol{\beta}$es asintóticamente distribuido como

$$\hat{\boldsymbol{\beta}}_{ML} \sim \mathrm{N}(\boldsymbol{\beta}, \mathcal{J}^{-1}),$$

donde $\mathcal{J}$ es la matriz de información.

Resulta que, bajo ciertas condiciones de regularidad, el estimador QL de $\boldsymbol{\beta}$ tiene propiedades asintóticas como el EMV de $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}_{QL} \sim \mathrm{N}(\boldsymbol{\beta}, \mathcal{J}^{-1}).$$

Considere un MLG con función de enlace g, predictor lineal η, distribución con función de varianza $V(\mu)$, y parámetro de dispersión $\phi = 1$. (Es decir, piense en los casos Poisson y Binomial).

Digamos que estima $\boldsymbol{\beta}$ de dos formas: a) usando EMV, y b) usando el enfoque QL con función de enlace g, predictor lineal η, función varianza $V(\mu)$, y parámetro de dispersión ϕ.

Bajo estos supuestos, las funciones de score y quasi-score pueden ser diferentes sólo en terminos de los valores de ϕ.

Sin embargo, como, en ambos casos ML y QL resolvemos

$$\mathbf{U}(\hat{\boldsymbol{\beta}}; \mathbf{y}) = \hat{\mathbf{D}}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \hat{\mu})/\phi = 0,$$

vemos que $\hat{\boldsymbol{\beta}}$ no depende de ϕ. En otras palabras, $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{QL} = \hat{\boldsymbol{\beta}}_{ML}$, luego las estimaciones son idénticas.

Por otro lado, las predicciones de μ_i son idénticas si usamos EMV o QL.

Ejemplo: Distribución de Poisson

Considere las observaciones independientes $Y_1, \ldots, Y_n$. Digamos que escoge, $V(\mu_i) = \mu_i$ y $g(\mu_i) = \log \mu_i = \Sigma_{j=1}^{p} x_{ij}\beta_j = \eta_i$ y obtiene las estimaciones por QL de $\boldsymbol{\beta}$ resolviendo $\mathbf{U} = 0$. Luego, esta estimación será idéntica a la que halló por EMVs $\boldsymbol{\beta}$ bajo el supuesto de que $Y_i \sim$Poisson(e^{η_i}).

Entonces ... *Cuál es la diferencia entre las estimaciones de QL y MV?!!*

Note que

$$\mathcal{J} \equiv \mathrm{Var}[\mathbf{U}(\boldsymbol{\beta}; \mathbf{Y})] = \frac{1}{\phi}\mathbf{D'V}^{-1}\mathbf{D}.$$

Por lo tanto,

$$\mathcal{J}^{-1} = \phi\left\{\mathbf{D'V}^{-1}\mathbf{D}\right\}^{-1}.$$

Como el valor de ϕ puede diferir en ambos modelos (EMV y QL), podemos tener $\mathrm{Var}[\hat{\boldsymbol{\beta}}_{QL}] \neq \mathrm{Var}[\hat{\boldsymbol{\beta}}_{ML}]$.

Por ejemplo, si tenemos datos de conteo con sobredispersion relativa a la distribucion de Poisson (tal que $\phi > 1$), la varianza de $\hat{\boldsymbol{\beta}}_{QL}$ sera mayor que la varianza de $\hat{\boldsymbol{\beta}}_{ML}$ (asumiendo un MLG Poisson).

Por el contrario , si tenemos datos con subdispersion (tal que $\phi < 1$), the variance of la varianza de $\hat{\boldsymbol{\beta}}_{QL}$ sera menor que la varianza de $\hat{\boldsymbol{\beta}}_{ML}$ (asumiendo un MLG Poisson).

En cualquiera de estos casos, la varianza de $\hat{\boldsymbol{\beta}}_{ML}$ no es precisa dado que surge del supuesto que $\phi = 1$. Bajo estas circustancias, el enfoque ML seria preferible.

Por otro lado, note que estimamos $\mathcal{J}^{-1}$ por

$$\hat{\mathcal{J}}^{-1} = \hat{\phi}\left\{\hat{\mathbf{D}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{D}}\right\}^{-1}.$$

Si nuestra *estimacion* de ϕ no es muy precisa, la estimacion de $\mathrm{Var}[\hat{\boldsymbol{\beta}}_{QL}]$ puede ser inprecisa tambien.

Estimación del parámetro de dispersión

Si ϕ no es conocido, se puede estimar usando el método de momentos, es decir,

$$\hat{\phi} = \frac{X^2}{n-p},$$

donde X^2 es la estadística de Pearson, n es el tamaño de muestra, y p es el número de coeficientes de regresión.

Para obtener las estimaciones por QL en R, se usa la familia `quasi` en el comando `glm`. La familia `quasi` tiene dos argumentos: `link` y `variance`.

El argumento `link` puede ser

```
logit, probit, cloglog, identity,
inverse, log, "1/mu^2", or sqrt.
```

El argumento `variance` puede ser

```
constant, mu(1-mu), mu,
mu^2, or mu^3.
```