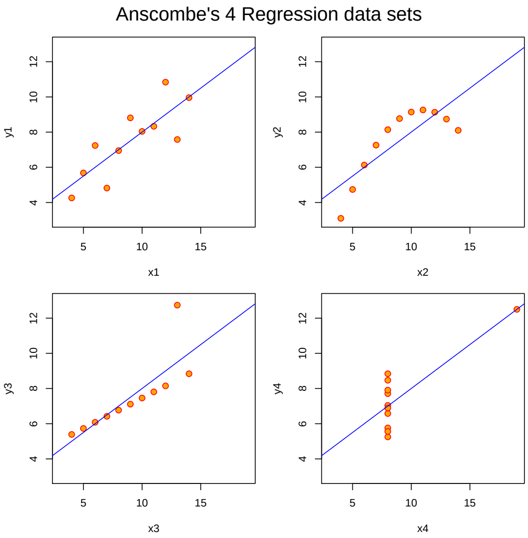
MAT-466: Diagnóstico de influencia en GLM

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Cuarteto de regresiones “idénticas” de Anscombe (1973)



Distancia de Cook (1977)

Basado en el elipsoide de confianza del $(1-\alpha)100\%$ para $\boldsymbol{\beta}$,

$$\frac{(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top \boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})}{ps^2} \leq F_{p,n-p}(1-\alpha).$$

Cook (1977) propuso determinar la influencia de la i-ésima observación, usando

$$D_i = \frac{(\widehat{\boldsymbol{\beta}}_{(i)} - \widehat{\boldsymbol{\beta}})^\top \boldsymbol{X}^\top \boldsymbol{X}(\widehat{\boldsymbol{\beta}}_{(i)} - \widehat{\boldsymbol{\beta}})}{ps^2}, \qquad i = 1, \dots, n.$$

Cook (1977) recomienda comparar D_i con algún percentil de la distribución $F_{p,n-p}$ ($\alpha = 0.10$). Weisberg (1985) sugiere que puede ser más razonable usar $\alpha = 0.50$ y propone que $D_i > 1$ es un indicador de observaciones influyentes.

Evaluando la distancia de Cook

Considere

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{(i)} \\ \boldsymbol{x}_i^\top \end{pmatrix}, \qquad \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_{(i)} \\ Y_i \end{pmatrix}.$$

Luego de un poco de álgebra, tenemos que

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{(i)} &= (\boldsymbol{X}_{(i)}^\top \boldsymbol{X}_{(i)})^{-1} \boldsymbol{X}_{(i)}^\top \boldsymbol{Y}_{(i)} \\ &= \widehat{\boldsymbol{\beta}} - \frac{e_i}{1 - h_i} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_i, \end{aligned}$$

esto permite escribir D_i como

$$D_i = \frac{r_i^2}{p} \Big(\frac{h_i}{1 - h_i} \Big), \qquad i = 1, \dots, n,$$

que es una fórmula bastante agradable desde el punto de vista computacional.

Función de Influencia

Considere que una distribución F es contaminada por una distribución con masa unitaria en $\boldsymbol{z}$, digamos δ_z. Es decir,

$$F(\epsilon) = (1-\epsilon)F + \epsilon\delta_z.$$

Hampel (1968, 1974) propuso medir la influencia de la i-ésima observación mediante

$$\mathrm{IF}(\boldsymbol{z}; F, T) = \lim_{\epsilon \to 0} \frac{T(F(\epsilon)) - T(F)}{\epsilon},$$

donde $T(\cdot)$ es una estadística vector-valorada basada en una muestra aleatoria desde la cdf F.

Versiones empíricas de la IF

Considere el modelo de regresión lineal y sea $\widehat{F}$ la función de distribución empírica basada en la muestra, tomando $T(\widehat{F}) = \widehat{\boldsymbol{\beta}}$, se obtiene la curva de influencia empírica

$$EIC(\boldsymbol{x}, Y) = n(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{x}(Y - \boldsymbol{x}^\top \widehat{\boldsymbol{\beta}}),$$

de este modo, $EIC_i(\boldsymbol{x}_i, Y_i) = n(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{x}_i e_i$.

Sea $\widehat{F}_{(i)}$ la cdf muestral con el i-ésimo caso eliminado, y $T(\widehat{F}_{(i)}) = \widehat{\boldsymbol{\beta}}_{(i)}$, entonces

$$EIC_{(i)}(\boldsymbol{x}_i, Y_i) = (n-1)(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{x}_i e_i/(1-h_i)^2.$$

Versiones muestrales de la IF

Una versión muestral de la curva de influencia (SIC) es hallada mediante omitir en límite en la definición de la IF y tomar $F = \widehat{F}$, $T(\widehat{F}) = \widehat{\boldsymbol{\beta}}$ y $\epsilon = -1/(n-1)$.

Evaluando en $\boldsymbol{z}_i = (\boldsymbol{x}_i, Y_i)$ tenemos $(1-\epsilon)\widehat{F} + \epsilon\delta_{z_i} = \widehat{F}_{(i)}$

$$\begin{aligned} SIC_i &= -(n-1)(T(\widehat{F}_{(i)}) - T(\widehat{F})) \\ &= (n-1)(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)}) \\ &= (n-1)\frac{e_i}{1-h_i}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_i e_i. \end{aligned}$$

Típicamente se considera

$$\mathrm{IF}_i = (n-1)^{-1} SIC_i = \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)},$$

como la curva de influencia muestral.

Cook y Weisberg (1980) propusieron normalizar la IF_i como

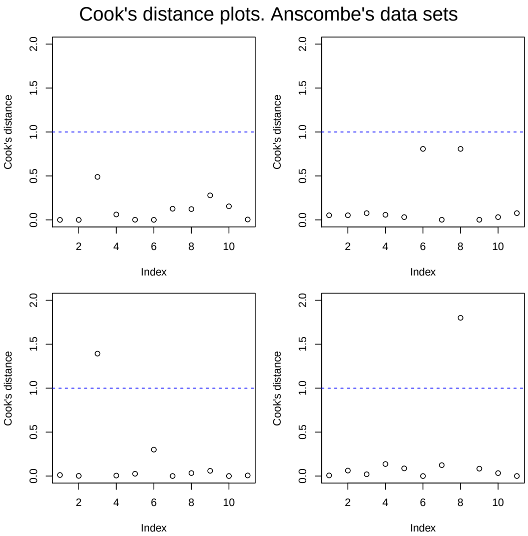
$$D_i(\boldsymbol{M}, c) = \frac{(\mathrm{IF}_i)^\top \boldsymbol{M}(\mathrm{IF}_i)}{c} = \frac{(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)})^\top \boldsymbol{M}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)})}{c},$$

donde $\boldsymbol{M}$ es matriz definida positiva $p \times p$ y $c > 0$ es un factor de escala.

Algunas medidas de influencia:

$\boldsymbol{M}$	c	Medida	Referencia
$\boldsymbol{X}^\top \boldsymbol{X}$	ps^2	D_i	Cook (1977)
$\boldsymbol{X}^\top \boldsymbol{X}$	$ps^2_{(i)}$	$(DFFITS_i)^2$	Welsch y Kuh (1977)

Distancia de Cook: Datos de Anscombe (1973)



Medidas basadas en diferenciación (Pregibon, 1981)

Considere el modelo

$$Y_j = \boldsymbol{x}_j^\top \boldsymbol{\beta} + \epsilon_j, \qquad j = 1, \dots, n,$$

donde var$(\epsilon_i) = \sigma^2/\omega$, $0 \leq \omega \leq 1$ y var$(\epsilon_j) = \sigma^2$, para $j \neq i$. En este caso

$$\widehat{\boldsymbol{\beta}}(\omega) = (\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{Y} = \widehat{\boldsymbol{\beta}} - \frac{(1-\omega)e_i}{1-(1-\omega)h_i}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{x}_i,$$

donde $\boldsymbol{W} = \operatorname{blc}\operatorname{diag}(\boldsymbol{I}_{n-1}, \omega)$.

Note que

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(1) - \widehat{\boldsymbol{\beta}}(0) &= \frac{e_i}{1-h_i}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{x}_i \\ &= \widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)} = \mathrm{IF}_i\,.\end{aligned}$$

Medidas basadas en diferenciación (Pregibon, 1981)

El efecto de la perturbación en la varianza puede ser obtenido como

$$\frac{\partial \widehat{\boldsymbol{\beta}}(\omega)}{\partial \omega} = \frac{e_i}{\{1 - (1 - \omega) h_i\}^2} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_i.$$

En particular,

$$\frac{\partial \widehat{\boldsymbol{\beta}}(\omega)}{\partial \omega}\Big|_{\omega=1} = e_i (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_i,$$

describe un cambio local sobre $\widehat{\boldsymbol{\beta}}(\omega)$ en la solución LS.

Medidas de diagnóstico pueden ser construídas de manera similar que usando procedimientos de eliminación de casos.

Comparando matrices de covarianza

Considere comparar $\text{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ con la matriz de covarianza que resulta de eliminar el i-ésimo caso. Esto lleva a (Belsley, Kuh y Welsch, 1980)

$$
\begin{aligned}
COVRATIO_i &= \frac{\det\{s_{(i)}^2(\boldsymbol{X}_{(i)}^\top\boldsymbol{X}_{(i)})^{-1}\}}{\det\{s^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\}} = \Big(\frac{s_{(i)}^2}{s^2}\Big)^p \frac{\det(\boldsymbol{X}_{(i)}^\top\boldsymbol{X}_{(i)})^{-1}}{\det(\boldsymbol{X}^\top\boldsymbol{X})^{-1}} \\
&= \frac{1}{1-h_i}\Big(\frac{n-p-r_i^2}{n-p-1}\Big)^p,
\end{aligned}
$$

se ha planteado como punto de corte $|COVRATIO_i - 1| > 3p/n$.

Repertorio de medidas de influencia

Existe un repertorio bastante extenso de medidas de influencia, por ejemplo:

Medida	Punto de corte
$D_i = \frac{r_i^2}{p}\left(\frac{h_i}{1-h_i}\right)$	$F_{p,n-p}(1-\alpha)$
$DFFITS_i = \lvert t_i \rvert \sqrt{\frac{h_i}{1-h_i}}$	$2\sqrt{p/n}$
$AK_i = DFFITS_i \sqrt{\frac{n-p}{p}}$	$2\sqrt{(n-p)/n}$
$W_i = DFFITS_i \sqrt{\frac{n-1}{1-h_i}}$	$3\sqrt{p}$
$COVRATIO_i = \frac{1}{1-h_i}\left(\frac{n-p-r_i^2}{n-p-1}\right)^p$	$\lvert COVRATIO_i - 1 \rvert > 3p/n$
$h_i = \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_i$	$2p/n$
$r_i = \frac{e_i}{s\sqrt{1-h_i}}$	$\approx N(0,1)$
$t_i = r_i \sqrt{\frac{n-p-1}{n-p-r_i^2}}$	$\approx t_{n-p-1}$

Belsley, Kuh y Welsch (1981) y Velleman y Welsch (1981) han discutido estratégias para amenizar el cálculo de estas medidas de influencia.

Para modelos de regresión lineal algunas de estas medidas han sido implementadas en software estadístico tal como SAS, SPSS, S-PLUS/R.

En particular, R (o S-PLUS) disponen de las funciones `lm.influence` y `ls.diag` asociadas con las funciones `lm` (o `glm`) y `lsfit`, respectivamente.

La función `lm.influence` dispone de las siguientes medidas:

```
rstandard    rstudent    dffits    dfbetas
covratio    cooks.distance    hatvalues
```

Estas cantidades pueden ser escritas de forma eficiente usando la descomposición QR o SVD.

Distancia de Cook generalizada

Considere $\widehat{\boldsymbol{\theta}}_{(i)}$ la estimación ML de $\boldsymbol{\theta}$ cuando se elimina el i-ésimo caso, es decir $\widehat{\boldsymbol{\theta}}_{(i)}$ maximiza la log-verosimilitud

$$\ell_{(i)}(\boldsymbol{\theta}) = \sum_{j \neq i} \ell_j(\boldsymbol{\theta}),$$

Podemos usar la aproximación

$$\ell_{(i)}(\boldsymbol{\theta}) \approx \ell_{(i)}(\widehat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^\top \boldsymbol{U}_{(i)}(\widehat{\boldsymbol{\theta}}) + \tfrac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^\top \ddot{\ell}_{(i)}(\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}),$$

donde $\boldsymbol{U}_{(i)}(\widehat{\boldsymbol{\theta}}) = \partial \ell_{(i)}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|_{\theta=\widehat{\theta}}$ y $\ddot{\ell}_{(i)}(\widehat{\boldsymbol{\theta}}) = \partial^2 \ell_{(i)}(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top|_{\theta=\widehat{\theta}}$.

Si $-\ddot{\ell}_{(i)}(\widehat{\boldsymbol{\theta}})$ es definida positiva, la aproximación cuadrática es maximizada en

$$\widehat{\boldsymbol{\theta}}_{(i)}^{1} = \widehat{\boldsymbol{\theta}} + \{-\ddot{\ell}_{(i)}(\widehat{\boldsymbol{\theta}})\}^{-1} \boldsymbol{U}_{(i)}(\widehat{\boldsymbol{\theta}}),$$

que es una aproximación de un paso para $\widehat{\boldsymbol{\theta}}_{(i)}$ (Cook y Weisberg, 1982).

Notando que

$$\mathbf{0} = \boldsymbol{U}(\widehat{\boldsymbol{\theta}}) = \frac{\partial \ell_{(i)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\theta=\widehat{\theta}} + \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\theta=\widehat{\theta}},$$

podemos escribir $\boldsymbol{U}_{(i)}(\widehat{\boldsymbol{\theta}}) = -\boldsymbol{U}_i(\widehat{\boldsymbol{\theta}})$.

De este modo

$$\widehat{\boldsymbol{\theta}}^{1}_{(i)} - \widehat{\boldsymbol{\theta}} = -\{-\ddot{\ell}_{(i)}(\widehat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{U}_i(\widehat{\boldsymbol{\theta}}).$$

Una aproximación adicional, lleva a la distancia de Cook generalizada

$$\begin{aligned} GD_i^1 &= (\widehat{\boldsymbol{\theta}}^{1}_{(i)} - \widehat{\boldsymbol{\theta}})^\top\{-\ddot{\ell}(\widehat{\boldsymbol{\theta}})\}^{-1}(\widehat{\boldsymbol{\theta}}^{1}_{(i)} - \widehat{\boldsymbol{\theta}}) \\ &= \boldsymbol{U}_i^\top(\widehat{\boldsymbol{\theta}})\{-\ddot{\ell}(\widehat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{U}_i(\widehat{\boldsymbol{\theta}}). \end{aligned}$$

Desplazamiento de verosimilitudes

Sea $\ell(\boldsymbol{\theta})$ la log-verosimilitud de los datos completos. Cook y Weisberg (1982) definieron el desplazamiento de verosimilitudes como

$$LD_i = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{(i)})\},$$

o, usando una aproximación de un paso

$$LD_i^1 = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}^{\,1}_{(i)})\}.$$

Las medidas LD_i y LD_i^1 pueden ser interpretadas en términos de la región de confianza asintótica

$$\{\boldsymbol{\theta} : 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta})\} \leq \chi_q^2(1-\alpha)\},$$

donde $\chi_q^2(1-\alpha)$ es un valor cuantil de la distribución chi-cuadrado con q grados de libertad y q denota la dimensión de $\boldsymbol{\theta}$.

Relación con la distancia de Cook

Considere la siguiente aproximación usando la expansión de Taylor de $\ell(\widehat{\boldsymbol{\theta}}_{(i)})$ en torno de $\widehat{\boldsymbol{\theta}}$

$$\ell(\widehat{\boldsymbol{\theta}}_{(i)}) \approx \ell(\widehat{\boldsymbol{\theta}}) + (\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}})^\top \boldsymbol{U}(\widehat{\boldsymbol{\theta}}) + \tfrac{1}{2}(\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}})^\top \ddot{\ell}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}}),$$

como $\boldsymbol{U}(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$, tenemos que

$$LD_i \approx (\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}})^\top \{-\ddot{\ell}(\widehat{\boldsymbol{\theta}})\}^{-1}(\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}}) = GD_i.$$

Se puede obtener una aproximación diferente substituyendo la matriz de información observada $-\ddot{\ell}(\widehat{\boldsymbol{\theta}})$ por la matriz de información esperada.

Cook (1986) sugiere estudiar el efecto de introducir pequeñas perturbaciones en el modelo (o datos).

Sea $\boldsymbol{\omega} \in \Omega$ un vector de perturbación $q \times 1$. Considere el modelo perturbado

$$\mathcal{M} = \{f(\boldsymbol{Y}; \boldsymbol{\theta}, \boldsymbol{\omega}) : \boldsymbol{\omega} \in \Omega\},$$

donde $f(\boldsymbol{Y}; \boldsymbol{\theta}, \boldsymbol{\omega})$ representa la función de densidad de $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \dots, \boldsymbol{Y}_n^\top)^\top$ perturbada por $\boldsymbol{\omega}$ y función de log-verosimilitud perturbada

$$\ell(\boldsymbol{\theta}, \boldsymbol{\omega}) = \log f(\boldsymbol{Y}; \boldsymbol{\theta}, \boldsymbol{\omega}).$$

Suponga también que existe un vector de no perturbación $\boldsymbol{\omega}_0$, tal que

$$\ell(\boldsymbol{\theta}, \boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta}).$$

Influencia Local (Cook, 1986)

Sea $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ el estimador máximo verosímil de $\boldsymbol{\theta}$ bajo el modelo perturbado $\mathcal{M}$.

La comparación entre $\widehat{\boldsymbol{\theta}}$ y $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ es una manera de determinar la influencia de una perturbación particular $\boldsymbol{\omega}$.

Si la "diferencia" entre $\widehat{\boldsymbol{\theta}}$ y $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ es pequeña conforme $\boldsymbol{\omega}$ varia en Ω. Entonces el análisis es estable.

El procedimiento es suficientemente flexible, pues permite perturbar tanto los datos como el modelo.

EX UMBRA IN SOLEM

Esquemas de perturbación habituales

Para fijar ideas, suponga el modelo de regresión $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathsf{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I})$.

1. Perturbación de respuesta: considere substituir $\boldsymbol{Y}$ por
$$\boldsymbol{Y}(\boldsymbol{\omega}) = \boldsymbol{Y} + \boldsymbol{\omega}, \qquad \boldsymbol{\omega}_0 = \mathbf{0}.$$
2. Perturbación de predictores: donde para el j-ésimo regresor,
$$\boldsymbol{X}(\boldsymbol{\omega}) = \boldsymbol{X} + \boldsymbol{\omega}\boldsymbol{d}_j^\top, \qquad \boldsymbol{\omega}_0 = \mathbf{0},$$
donde $\boldsymbol{d}_j$ tiene un 1 en la j-ésima posición y ceros en las restantes.
3. Perturbación de casos: los disturbios aleatorios son reemplazados por
$$\boldsymbol{\epsilon}(\boldsymbol{\omega}) \sim \mathsf{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{W}), \qquad \boldsymbol{W} = \operatorname{diag}(\boldsymbol{\omega}), \qquad \boldsymbol{\omega}_0 = \mathbf{1}_n.$$

Comparar $\widehat{\boldsymbol{\theta}}$ con $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ conforme $\boldsymbol{\omega}$ varia en Ω no es necesariamente una tarea simple.

Cook (1986) sugirió usar el desplazamiento de verosimilitudes

$$LD(\boldsymbol{\omega}) = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}))\},$$

para medir la influencia de una perturbación sobre el MLE de $\boldsymbol{\theta}$.

$LD(\boldsymbol{\omega})$ es siempre no negativo y mide la diferencia entre $\widehat{\boldsymbol{\theta}}$ y $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$ mediante usar los contornos de la log-verosimilitud $\ell(\boldsymbol{\theta})$ para el modelo no perturbado.

La log-verosimilitud del modelo perturbado $\ell(\boldsymbol{\theta}, \boldsymbol{\omega})$ sólo es utilizada para proveer el estimador $\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})$, no interviene en el cálculo de $LD(\boldsymbol{\omega})$.

La idea de influencia local (Cook, 1986) está interesada en caracterizar el comportamiento local de $LD(\boldsymbol{\omega})$ en torno de $\boldsymbol{\omega}_0$.

Cook (1986) mostró que la curvatura normal $C_h(\boldsymbol{\theta})$ usada para caracterizar $LD(\boldsymbol{\omega}_0 + a\boldsymbol{h})$ en torno de $a = 0$ con $\|\boldsymbol{h}\| = 1$ asume la forma

$$C_h(\boldsymbol{\theta}) = 2\boldsymbol{h}^\top\{-\ddot{\boldsymbol{F}}\}\boldsymbol{h},$$

donde $\ddot{\boldsymbol{F}} = \partial^2\ell(\boldsymbol{\theta}(\boldsymbol{\omega}))/\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^\top|_{\omega=\omega_0}$.

Es posible notar que $-\ddot{\boldsymbol{F}} = \boldsymbol{\Delta}^\top\{-\ddot{\ell}(\widehat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{\Delta}$, donde

$$\ddot{\ell}(\widehat{\boldsymbol{\theta}}) = \frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\Big|_{\theta=\widehat{\theta}} \qquad \text{y} \qquad \boldsymbol{\Delta} = \frac{\partial^2\ell(\boldsymbol{\theta},\boldsymbol{\omega})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top}\Big|_{\theta=\widehat{\theta},\omega=\omega_0}.$$

Influencia Local (Cook, 1986)

Cook (1986) sugirió prestar atención a la dirección de máxima curvatura $\boldsymbol{h}_{\max}$.

$\boldsymbol{h}_{\max}$ indica cómo se debe perturbar el modelo para obtener el mayor cambio local en el desplazamiento de verosimilitudes.

La máxima curvatura $C_{h_{\max}}(\boldsymbol{\theta})$ está dada por el mayor valor propio de $\ddot{\boldsymbol{F}}$ (en valor absoluto) y la dirección de mayor curvatura $\boldsymbol{h}_{\max}$ está dada por el vector propio asociado.

Otras direcciones también pueden ser de interés, por ejemplo $\boldsymbol{h}_{\text{2nd}}$. Lesaffre y Verbeke (1998) sugieren realizar el gráficos de índices de $C_i = C_{h_i}(\boldsymbol{\theta})$, donde $\boldsymbol{h}_i$ es un vector con un 1 en la posición i-ésima y ceros en las restantes.

En ocasiones se tiene interés en determinar la influencia de un subconjunto $\boldsymbol{\theta}_1$ de $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$. En este caso,

$$LD(\boldsymbol{\omega}) = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_1(\boldsymbol{\omega}), \widehat{\boldsymbol{\theta}}_2(\widehat{\boldsymbol{\theta}}_1(\boldsymbol{\omega})))\},$$

donde $\widehat{\boldsymbol{\theta}}_1(\boldsymbol{\omega}), \widehat{\boldsymbol{\theta}}_2(\widehat{\boldsymbol{\theta}}_1(\boldsymbol{\omega}))$ es el MLE de $\boldsymbol{\theta}_2$ en el modelo perturbado, para $\boldsymbol{\theta}_1$ fijado. Considere las matrices

$$\ddot{\ell}(\boldsymbol{\theta}) = \begin{pmatrix} \ddot{\ell}_{\theta_1\theta_1}(\boldsymbol{\theta}) & \ddot{\ell}_{\theta_1\theta_2}(\boldsymbol{\theta}) \\ \ddot{\ell}_{\theta_2\theta_1}(\boldsymbol{\theta}) & \ddot{\ell}_{\theta_2\theta_2}(\boldsymbol{\theta}) \end{pmatrix} \qquad \text{y} \qquad \boldsymbol{B}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\ell}^{-1}_{\theta_2\theta_2}(\boldsymbol{\theta}) \end{pmatrix}.$$

Entonces, la curvatura normal en la dirección $\boldsymbol{h}$ para $\boldsymbol{\theta}_1$ asume la forma

$$C_h(\boldsymbol{\theta}_1) = -2\boldsymbol{h}^\top \boldsymbol{\Delta}^\top \{\ddot{\ell}^{-1}(\boldsymbol{\theta}) - \boldsymbol{B}_{22}\}\boldsymbol{\Delta}\boldsymbol{h}.$$

Considere la función de influencia $T(\boldsymbol{\omega})$. La forma general para la curvatura normal es

$$C_h(\boldsymbol{\theta}) = \frac{\boldsymbol{h}^\top \ddot{\boldsymbol{\Gamma}} \boldsymbol{h}}{\boldsymbol{h}^\top (\boldsymbol{I} + \dot{\boldsymbol{\Gamma}} \dot{\boldsymbol{\Gamma}}^\top) \boldsymbol{h} (1 + \dot{\boldsymbol{\Gamma}}^\top \dot{\boldsymbol{\Gamma}})^{1/2}},$$

donde $\dot{\boldsymbol{\Gamma}} = \partial \widehat{T}(\boldsymbol{\omega}) / \partial \boldsymbol{\omega} |_{\omega = \omega_0}$ y $\ddot{\boldsymbol{\Gamma}} = \partial^2 \widehat{T}(\boldsymbol{\omega}) / \partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^\top |_{\omega = \omega_0}$.

En influencia local usando el desplazamiento de verosimilitudes se tiene

$$\dot{\boldsymbol{\Gamma}} = 2 \partial \ell(\widehat{\boldsymbol{\theta}}_\omega) / \partial \boldsymbol{\omega} |_{\omega = \omega_0} = \mathbf{0}.$$

Fung y Kwan (1993) mostraron que $C_h(\boldsymbol{\theta})$ no es invariante bajo transfor- maciones de escala de $T(\boldsymbol{\omega})$.

Curvatura normal conformal (Poon y Poon, 1999)

Poon y Poon (1999) propusieron la curvatura normal conformal

$$B_h(\boldsymbol{\theta}) = \frac{1}{\boldsymbol{h}^\top(\boldsymbol{I} + \dot{\boldsymbol{\Gamma}}\dot{\boldsymbol{\Gamma}}^\top)\boldsymbol{h}} \frac{\boldsymbol{h}^\top \ddot{\boldsymbol{\Gamma}} \boldsymbol{h}}{\{\operatorname{tr} \ddot{\boldsymbol{\Gamma}}^2\}^{1/2}}\bigg|_{\omega=\omega_0}.$$

En particular, usando el desplazamiento de verosimilitudes

$$B_h(\boldsymbol{\theta}) = \frac{\boldsymbol{h}^\top \boldsymbol{\Delta}^\top \{-\ddot{\ell}(\boldsymbol{\theta})\}^{-1} \boldsymbol{\Delta} \boldsymbol{h}}{\sqrt{\operatorname{tr}\{\boldsymbol{\Delta}^\top \{-\ddot{\ell}(\boldsymbol{\theta})\}^{-1} \boldsymbol{\Delta}\}^2}}\bigg|_{\theta=\widehat{\theta},\omega=\omega_0}.$$

Observación:

Considere una reparametrización $\phi : \Omega \to \Theta$ tal que la matriz Jacobiana de ϕ es no singular. Una matriz $\boldsymbol{M}$ se dice conformal si existe $\tau > 0$ tal que $\boldsymbol{M}\boldsymbol{M}^\top = \tau \boldsymbol{I}$. De este modo, una reparametrización es dicha conformal en $\boldsymbol{\omega}_0$ si su matriz Jacobiana en $\boldsymbol{\omega}_0$ es conformal.

Propiedades:

(a) La curvatura normal conformal (Poon y Poon, 1999) es invariante bajo reparametrizaciones conformales.

(b) Para cualquier dirección $\boldsymbol{h}$, $B_h(\boldsymbol{\theta})$ satisface que $0 \leq |B_h(\boldsymbol{\theta})| \leq 1$.

Considere la descomposición espectral

$$\ddot{\boldsymbol{F}} = \sum_{i=1}^{q} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^{\top},$$

donde $\{\lambda_i, \boldsymbol{u}_i\}_{i=1}^{q}$ son los valores y vectores propios de $\ddot{\boldsymbol{F}}$ con $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iq})^{\top}$. Se tiene que $\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots \lambda_q = 0$, sea

$$\widetilde{\lambda}_i = \frac{\lambda_i}{\sum_{k=1}^{r} \lambda_k}, \qquad \text{y} \qquad \sum_{i=1}^{r} \widetilde{\lambda}_i = 1.$$

Para $\boldsymbol{h}_j$ un vector con un 1 en la j-ésima posición y ceros en las restantes. Tenemos,

$$B_{h_j}(\boldsymbol{\theta}) = \sum_{i=1}^{r} \lambda_i u_{ij}^2.$$

Bench-mark para la curvatura normal

Zhu y Lee (2001) definieron $M(0)_j = B_{h_j}(\boldsymbol{\theta})$ y notaron que

$$\overline{M}(0) = 1/q.$$

Ellos propusieron declarar la j-ésima observación como influyente si $B_{h_j}(\boldsymbol{\theta})$ es mayor que el punto de corte

$$\overline{M}(0) + 2\,SM(0),$$

donde $SM(0)$ es el error estándar muestral de $\{M(0)_k, k = 1, \ldots, q\}$.

Una nota sobre cálculo

El cálculo de $\boldsymbol{h}_{\max}$ involucra la descomposición espectral de la matriz

$$\ddot{\boldsymbol{F}} = \boldsymbol{\Delta}^\top \{-\ddot{\ell}(\widehat{\boldsymbol{\theta}})\}^{-1} \boldsymbol{\Delta},$$

que puede ser de alto costo computacional.

Una alternativa es considerar $\{-\ddot{\ell}(\widehat{\boldsymbol{\theta}})\} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top$ y hacer $\boldsymbol{B} = \boldsymbol{\Delta}^\top \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1/2}$. Luego, realizamos una descomposición valor singular de $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$. Entonces,

$$\ddot{\boldsymbol{F}} = \boldsymbol{B}\boldsymbol{B}^\top = \boldsymbol{U}\boldsymbol{D}^2\boldsymbol{U}^\top = \sum_{j=1}^{r} \delta_j \boldsymbol{u}_j \boldsymbol{u}_j^\top,$$

donde $\boldsymbol{U} = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_r) \in \mathbb{R}^{q \times r}$, es ortogonal por columnas y $\boldsymbol{D} = \operatorname{diag}(\delta_1^{1/2}, \dots, \delta_r^{1/2})$. Note que

$$\boldsymbol{h}_{\max} = \boldsymbol{u}_1, \qquad \text{y} \qquad B_h(\boldsymbol{\theta}) = \frac{\boldsymbol{h}^\top \boldsymbol{U}\boldsymbol{D}^2\boldsymbol{U}^\top \boldsymbol{h}}{\operatorname{tr} \boldsymbol{D}^2}.$$

Billor y Loynes (1993) consideran la siguiente función de influencia

$$LD^*(\boldsymbol{\omega}) = -2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}), \boldsymbol{\omega})\}.$$

Se desea caracterizar la superficie $(\boldsymbol{\omega}, LD^*(\boldsymbol{\omega}))$ para perturbaciones del tipo $\boldsymbol{\omega} = \boldsymbol{\omega}_0 + a\boldsymbol{h}$ donde $\boldsymbol{\omega}_0$ es la perturbación nula, $\boldsymbol{h}$ es un vector unitario y el escalar a determina la magnitud de la perturbación.

En este caso la primera derivada provee información relevante del comportamiento local de $LD^*(\boldsymbol{\omega})$ en torno de $\boldsymbol{\omega}_0$.

La influencia local de una perturbación es medida usando la pendiente en la dirección $\boldsymbol{h}$, denotada por S_h

$$S_h = \frac{\partial LD^*(\boldsymbol{\omega})}{\partial a}\Big|_{a=0}.$$

Note que

$$S_h = \frac{\partial LD^*(\boldsymbol{\omega})}{\partial a}\Big|_{a=0} = \boldsymbol{h}^\top \frac{\partial LD^*(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\Big|_{a=0},$$

usando la regla de la cadena, se tiene que

$$S_h = \boldsymbol{h}^\top \Big\{ \frac{\partial \ell(\widehat{\boldsymbol{\theta}}(\boldsymbol{\omega}), \boldsymbol{\omega})}{\partial \boldsymbol{\omega}} + \frac{\partial \ell(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\theta}} \frac{\partial \widehat{\boldsymbol{\theta}}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \Big\}\Big|_{\theta=\widehat{\theta}, \omega=\omega_0}.$$

Finalmente, esto reduce a

$$S_h = 2\boldsymbol{h}^\top \frac{\partial \ell(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\Big|_{\theta=\widehat{\theta}, \omega=\omega_0},$$

y la dirección de máxima pendiente está dada por

$$\boldsymbol{h}_{\max} = 2\, \frac{\partial \ell(\widehat{\boldsymbol{\theta}}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\Big|_{\omega=\omega_0}.$$

Cook, R.D. (1977).
Detection of influential observation in linear regression.
Technometrics **19**, 15-18.

Cook, R.D. and Weisberg, S. (1982).
Residuals and Influence in Regression.
Chapman and Hall, New York.

Cook, R.D. (1986).
Assessment of local influence.
Journal of the Royal Statistical Society, Series B **48**, 133-169.

Billor, N., and Loynes, R.M. (1993).
Local influence: A new approach.
Communications in Statistics - Theory and Methods **22**, 1595-1611.

Poon, W., and Poon, Y.S. (1999).
Conformal normal curvature and assessment of local influence.
Journal of the Royal Statistical Society, Series B **61**, 51-61.

Zhu, H., Ibrahim, J.G., Lee, S., and Zhang, H. (2007).
Perturbation selection and influence measures in local influence analysis.
The Annals of Statistics **35**, 2565-2588.