Convergência em probabilidade.

Seja $\hat{\theta}_T$ o estimador de θ baseado em uma amostra de tamanho *T*.

Então $\hat{\theta}_T$ converge em probabilidade para θ se:

$$\lim_{T\to\infty}\left(\text{Prob}\left|\hat{\theta}_T-\theta\right|<\varepsilon\right)=1 \qquad \Rightarrow \qquad \lim_{T\to\infty}\left(\text{Prob}(\theta-\varepsilon<\hat{\theta}_T<\theta+\varepsilon)\right)=1$$

Se $\hat{\theta}_T$ converge em probabilidade para θ $\Rightarrow$ plim $\hat{\theta}_T = \theta$

Um estimador $\hat{\theta}_T$ do parâmetro θ é consistente se e somente se plim $\hat{\theta}_T = \theta$

Consistência

se

$$\lim_{T\to\infty} V\left(\hat{\theta}_T\right) = 0$$

e

$$\lim_{T\to\infty} E\left(\hat{\theta}_T\right) = c$$

então o estimador é <u>consistente.</u>

Exemplo:

$$E(\overline{X}) = \mu \quad \text{e} \quad V(\overline{X}) = \frac{\sigma^2}{n}$$

então

$$\lim_{n\to\infty} V\left(\overline{X}\right) = 0$$

e

$$\lim_{n\to\infty} E\left(\overline{X}\right) = \mu$$

$$\therefore \operatorname{plim}\overline{X} = \mu$$

Convergência em distribuição.

Teorema do Limite Central - Seja $Y_1, Y_2, ..., Y_T$ variáveis aleatórias identicamente distribuídas com $E(Y_i)=\beta$ e $Var(Y_i)=\sigma^2$ e seja

$$\hat{\beta}_T = \frac{\sum Y}{T}$$

Então $\sqrt{T}(\hat{\beta}_T - \beta)$ converge em distribuição para uma variável aleatória $N(0,\sigma^2)$.

$$\sqrt{T}(\hat{\beta}_T - \beta) \xrightarrow{d} N(0, \sigma^2)$$

Lembre-se que:

$$E(\hat{\beta}_T) = E(\frac{Y_1 + Y_2 + ... + Y_T}{T}) = \frac{T\beta}{T} = \beta$$

$$Var(\hat{\beta}_T) = Var(\frac{Y_1 + Y_2 + ... + Y_T}{T}) = \frac{T\sigma^2}{T^2} = \frac{\sigma^2}{T}$$

$$\frac{\hat{\beta}_T - \beta}{\sqrt{\frac{\sigma^2}{T}}} = \frac{\sqrt{T}(\hat{\beta}_T - \beta)}{\sigma} \sim N(0,1)$$

$$\sqrt{T}(\hat{\beta}_T - \beta) \sim N(0, \sigma^2)$$

Generalizando p/ vetores de estimadores

Se $\hat{\theta}_T$ é um estimador consistente de θ e $\sqrt{T}(\hat{\theta}_T - \theta)$ converge em distribuição para $N(0, \Sigma)$, então $\hat{\theta}_T$ é dito ter uma distribuição assintótica $N(\theta, (1/T)\Sigma)$

Eficiência assintótica.

Seja θ um vetor de parâmetros e $\hat{\theta}$ e $\tilde{\theta}$ estimadores consistentes tal que,

$$\sqrt{T}(\hat{\theta}-\theta) \xrightarrow{d} N(0,\Sigma)$$

$$\sqrt{T}(\tilde{\theta}-\theta) \xrightarrow{d} N(0,\Omega)$$

então $\hat{\theta}$ é assintoticamente eficiente em relação a $\tilde{\theta}$ se $\Omega - \Sigma$

é positiva semidefinida.

Seja $\hat{\theta}$ um estimador consistente de θ tal que,

$$\sqrt{T}(\hat{\theta}-\theta) \xrightarrow{d} N(0,\Sigma)$$

então o estimador $\hat{\theta}$ é assintóticamente eficiente se

$$\Sigma = \lim_{T\to\infty}\left[\frac{1}{T}I(\theta)\right]^{-1}$$

Vimos que: $\hat{\beta}_T = \frac{\sum Y}{T}$ $\hat{\beta}_T \sim N(\beta, \frac{\sigma^2}{T})$ $\sqrt{T}(\hat{\beta}_T - \beta) \sim N(0, \sigma^2)$

$$\ln L = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\frac{\sum_{i=1}^{T}(Y_i - \beta)^2}{\sigma^2}$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{\sum (Y_i - \beta)}{\sigma^2}$$

$$\frac{\partial^2 \ln L}{\partial \beta^2} = -\frac{T}{\sigma^2}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial \beta^2}\right) = \frac{T}{\sigma^2} \qquad \lim_{T\to\infty}\left[\frac{1}{T}I(\theta)\right]^{-1} = \left(\frac{1}{T}\frac{T}{\sigma^2}\right)^{-1} = \sigma^2$$

Portanto, $\hat{\beta}_T$ é assintóticamente eficiente.

O estimador de máxima verossimilhança é consistente, assintóticamente normal, assintóticamente não tendencioso e assintóticamente eficiente, isto é,

$$\sqrt{T}(\hat{\theta}-\theta) \xrightarrow{d} N\left[0, \lim_{T\to\infty}\left[\frac{1}{T} I(\theta)\right]^{-1}\right]$$

O método de máxima verossimilhança.

Seja Y uma variável aleatória com distribuição de Bernoulli, ie,

$$f(y|p) = p^{y}(1-p)^{1-y} \qquad 0 \leq p \leq 1 \qquad y = 0,1$$

com parâmetro p conhecido, que por alguma razão só pode assumir os valores ¼ ou ¾ . Suponha que temos uma amostra aleatória de tamanho T=3 com valores $y_1 = 1$, $y_2 = 1$ e $y_3 = 0$. Como utilizar estes dados para estimar o parâmetro desconhecido p, que nesse caso é p = ¼ ou p = ¾.

O método de máxima verossimilhança escolherá o valor do parâmetro desconhecido (p) que maximiza a probabilidade de aleatoriamente selecionar a amostra já obtida.

A função de densidade de probabilidade conjunta é:

$$f(y_1 = 1, y_2 = 1, y_3 = 0) = f(1,1,0) = \prod_{i=1}^{3} p^{y_i}(1-p)^{1-y_i} = p.p.(1-p)$$

Interpretando a f.d.p. como uma função de p dadas às observações da amostra, esta passa a ser denominada função de verossimilhança indicada por $l(p|\boldsymbol{y})$. A função de verossimilhança é matematicamente idêntica a f.d.p. conjunta da variável aleatória, mas ela é interpretada como uma função do parâmetro desconhecido ao invés dos valores das variáveis aleatórias, os quais pressupõe-se que sejam conhecidos. Assim sendo,

$$l(\frac{1}{4}|y) = 0{,}046$$

e

$$l(\frac{3}{4}|y) = 0{,}141$$

Então, a probabilidade de obtermos os valores da amostra que temos é maximizada escolhendo $\tilde{p} = 3/4$ ao invés de $\tilde{p} = 1/4$, dada que estas são as únicas escolhas que temos. Portanto, $\tilde{p} = 3/4$ é a estimativa de máxima verossimilhança de *p* nesse exemplo.

É claro que raramente teremos um conjunto admissível de parâmetros contendo somente dois elementos. Para a distribuição de Bernoulli, o parâmetro pode assumir qualquer valor no intervalo fechado entre zero e um, ou seja,

$$\Omega = \{ p \,|\, 0 \leq p \leq 1 \}$$

Nesta situação utiliza-se técnicas de cálculo para achar o valor de *p* que maximiza

$$l(p|y) = p.p.(1-p) = p^2 - p^3$$

$$\frac{\partial l(p|y)}{\partial p} = 2p - 3p^2 = 0 \quad \Rightarrow \quad \begin{cases} \tilde{p} = 0 \\ \tilde{p} = 2/3 \end{cases} \qquad \frac{\partial^2 l(p|y)}{\partial p^2} = 2 - 6p$$

Na prática maximiza-se o logaritmo natural da função de verossimilhança ao invés da função de verossimilhança em si. Isto é mais conveniente uma vez que o log da função de verossimilhança é composto por somas ao invés de produtos e a função exponencial é simplificada. Uma vez que o logaritmo natural é uma função monotônica, o máximo da função é atingido para o mesmo valor do parâmetro (ou do conjunto de parâmetros) θ.

Exemplo de como obter estimadores de máxima verossimilhança.

Seja $Y_1, Y_2, ..., Y_T$ uma amostra aleatória de variáveis Bernoulli com,

$$f(y_i|p) = p^{y_i}(1-p)^{1-y_i} \qquad i = 1,...,T \qquad \text{e} \qquad 0 \leq p \leq 1$$

Então o logaritmo da função de máxima verossimilhança é:

$$L = \log l(p|y) = \log\left[\prod_{i=1}^{T} f(y_i|p)\right] = \log\left[p^{y_1}(1-p)^{1-y_1}.p^{y_2}(1-p)^{1-y_2}...p^{y_T}(1-p)^{1-y_T}\right]$$

$$= \log\left[p^{\sum y_i}.(1-p)^{\sum (1-y_i)}\right] = \sum y_i \log p + \sum (1-y_i)\log(1-p)$$

$$\frac{\partial L}{\partial p} = \sum y_i . \frac{1}{p} - \sum (1-y_i). \frac{1}{1-p} = 0$$

$$\sum y_i (1-p) = \sum (1-y_i)\, p \quad \Rightarrow \quad \sum y_i - p\sum y_i = Tp - p\sum y_i$$

Portanto,

$$\tilde{p} = \frac{\sum y_i}{T}$$

$$\frac{\partial^2 L}{\partial p^2} = -\sum y_i . \frac{1}{p^2} - \sum (1-y_i). \frac{1}{(1-p)^2}$$

Estimação por máxima Verossimilhança do modelo linear

$$Y = X\beta + \varepsilon$$

$E(\varepsilon) = 0$ $\quad V(\varepsilon) = \sigma^2 I$ $\quad E(\varepsilon_i \varepsilon_j) = 0$ $\quad \varepsilon$ ~Normal

Se $\varepsilon \sim N(0, \sigma^2 I) \Rightarrow Y \sim N(X\beta, \sigma^2 I)$

Função de densidade.

$$f(Y_t | X_t, \beta, \sigma^2) = \frac{1}{\sqrt{2\Pi\sigma^2}} \cdot \exp\left\{ \frac{-\left(Y_t - X_t'\beta\right)^2}{2\sigma^2} \right\}$$

$$f(Y_1, Y_2 \ldots Y_T) = f(Y_1) f(Y_2) \ldots f(Y_T)$$

$$f(Y | X, \beta, \sigma^2) = \frac{1}{\left(2\Pi\sigma^2\right)^{T/2}} \cdot \exp\left[-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2} \right]$$

Considere o logaritmo da função de verossimilhança.

$$L = \log l\left(\beta, \sigma^2 \middle| Y, X\right) = -\frac{T}{2}\log 2\Pi - \frac{T}{2}\log\sigma^2 - \frac{\left(Y - X\beta\right)'\left(Y - X\beta\right)}{2\sigma^2}$$

Uma vez que o único termo contendo β é $-\frac{\left(Y - X\beta\right)'\left(Y - X\beta\right)}{2\sigma^2}$,

maximizar log *l*com relação a β é equivalente a maximizar

$-\frac{\left(Y - X\beta\right)'\left(Y - X\beta\right)}{2\sigma^2}$ com relação a β , o que é equivalente a

minimizar $S = \left(Y - X\beta\right)'\left(Y - X\beta\right)$ em relação a β .

Portanto, o estimador que minimiza a soma de quadrados dos erros é o mesmo que maximiza a função de verossimilhança.

$$b = \left(X'X\right)^{-1} X'Y$$

$$\text{var}\left(b\right) = E\left(b-\beta\right)^2 = \left(X'X\right)^{-1}\sigma^2$$

Para obter $\hat{\sigma}^2$

$$\frac{\partial \log l}{\partial \sigma^2} = -\frac{T}{2\hat{\sigma}^2} + \frac{1}{2\left(\hat{\sigma}^2\right)^2}\left(Y-X\beta\right)'\left(Y-X\beta\right) = 0$$

$$\hat{\sigma}^2 = \frac{\left(Y-X\beta\right)'\left(Y-X\beta\right)}{T}$$

$$\frac{\partial \log l}{\partial \beta} = -\frac{-2X'(Y - X\beta)}{2\sigma^2} = \frac{X'Y - X'X\beta}{\sigma^2}$$

$$\frac{\partial^2 \log l}{\partial \beta^2} = -\frac{X'X}{\sigma^2}$$

$$\underbrace{-E\left(\frac{\partial^2 \log l}{\partial \beta^2}\right)}_{I(\beta)} = \frac{X'X}{\sigma^2}$$

$$Var(b) = \left[I(\beta)\right]^{-1} = (X'X)^{-1}\sigma^2$$

Estimação Por Máxima Verossimilhança do modelo não linear.

Considere o modelo

$$Y = f\left(X, \beta\right) + \varepsilon \qquad\qquad \varepsilon \sim N\left(0, \sigma^2 I\right)$$

A função de verossimilhança para este modelo é:

$$L\left(\beta, \sigma^2 \middle| Y, X\right) = \frac{1}{\left(2\Pi\sigma^2\right)^{T/2}} \exp\left\{ - \frac{\left[Y - f\left(X, \beta\right)\right]' \left[Y - f\left(X, \beta\right)\right]}{2\sigma^2} \right\}$$

$$= \frac{1}{\left(2\Pi\sigma^2\right)^{T/2}} \exp\left\{ - \frac{S\left(\beta\right)}{2\sigma^2} \right\}$$

$$\ln L\left(\beta,\sigma^2\middle|Y,X\right)=-\frac{T}{2}\ln 2\Pi-\frac{T}{2}\ln\sigma^2-\frac{S(\beta)}{2\sigma^2}$$

$$\frac{\partial \ln L}{\partial \sigma^2}=-\frac{T}{2}\frac{1}{\sigma^2}+\frac{S(\beta)}{2\left(\sigma^2\right)^2}=0$$

$$\tilde{\sigma}^2=\frac{S(\beta)}{T}$$

$$\frac{\partial \ln L}{\partial \beta}=-\frac{1}{2\tilde{\sigma}^2}2(Y-f(X,\beta))'\left(-\frac{\partial f(X,\beta)}{\partial \beta}\right)=0$$

que é igual a condição de primeira ordem ao minimizar a soma de quadrado dos erros $S(\beta)$.

Propriedades do estimador:

Seja $\tilde{\theta}$ o estimador de máxima verossimilhança de $\theta = \left(\beta, \sigma^2\right)$, então

$$\sqrt{T}\left(\tilde{\theta} - \theta\right) \xrightarrow{d} N\left(0, \left[\frac{1}{T} I(\theta)\right]^{-1}\right)$$

$$I(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial\theta\partial\theta'}\right) = -E\begin{bmatrix} \dfrac{\partial^2 \ln L}{\partial\beta\partial\beta'} & \dfrac{\partial^2 \ln L}{\partial\beta\partial\sigma^2} \\ \dfrac{\partial^2 \ln L}{\partial\sigma^2\partial\beta'} & \dfrac{\partial^2 \ln L}{\partial\left(\sigma^2\right)^2} \end{bmatrix}$$

$$\frac{\partial \ln L}{\partial \beta} = -\frac{1}{2\sigma^2} 2(Y - f(X,\beta))' \left(-\frac{\partial f(X,\beta)}{\partial \beta} \right)$$

$$\frac{\partial^2 \ln L}{\partial \beta^2} = \sigma^{-2} \cdot \left(-\frac{\partial f(X,\beta)}{\partial \beta}' \frac{\partial f(X,\beta)}{\partial \beta} \right) + \sigma^{-2} \cdot (\underbrace{Y - f(X,\beta)}_{\varepsilon}) \left(-\frac{\partial^2 f(X,\beta)}{\partial \beta \partial \beta'} \right)$$

$$\frac{\partial^2 \ln L}{\partial \left(\sigma^2\right)^2} = \frac{T}{2\sigma^4} - \frac{2(Y - f(X,\beta))'(Y - f(X,\beta))}{2\left(\sigma^2\right)^3} = \frac{T}{2\sigma^4} - \frac{2T\sigma^2}{2\sigma^6} = -\frac{T}{2\sigma^4}$$

Portanto,

$$I(\theta) = \begin{bmatrix} \sigma^{-2} \left[Z(\beta)' Z(\beta) \right] & 0 \\ 0 & \frac{T}{2\sigma^4} \end{bmatrix}$$

Algoritmos comumente usados para obter estimativas de máxima verossimilhança.

1) Algoritmo de Newton-Raphson

$$\beta_{n+1} = \beta_n - \left[\frac{\partial^2 \ln L}{\partial\beta\partial\beta'}\right]^{-1}\Bigg|_{\beta_n} \frac{\partial \ln L}{\partial\beta}\Bigg|_{\beta_n} = \beta_n - \left[\frac{-1}{2\sigma^2}\cdot\frac{\partial^2 S}{\partial\beta\partial\beta'}\right]^{-1}\Bigg|_{\beta_n} \frac{-1}{2\sigma^2}\frac{\partial S}{\partial\beta}\Bigg|_{\beta_n}$$

$$= \beta_n - \left[\frac{\partial^2 S}{\partial\beta\partial\beta'}\right]^{-1}\Bigg|_{\beta_n} \frac{\partial S}{\partial\beta}\Bigg|_{\beta_n}$$

2) Método de Score

$$\beta_{n+1} = \beta_n - \left[E\left(\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} \right) \right]^{-1} \Bigg|_{\beta_n} \frac{\partial \ln L}{\partial \beta} \Bigg|_{\beta_n} = \beta_n - \left[-\frac{1}{\sigma^2} Z(\beta_n)' Z(\beta_n) \right]^{-1} \left[-\frac{1}{2\sigma^2} \frac{\partial S}{\partial \beta} \right] \Bigg|_{\beta_n}$$

$$= \beta_n - \frac{1}{2} \left[Z(\beta_n)' Z(\beta_n) \right]^{-1} \frac{\partial S}{\partial \beta} \Bigg|_{\beta_n}$$

3) Algoritmo de Berndt, Hall, Hall, e Haussman (BHHH algoritmo)

$$\beta_{n+1} = \beta_n + \left[\sum_{t=1}^{T} \left(\frac{\partial \ln L_t}{\partial \beta} \right) \left(\frac{\partial \ln L_t}{\partial \beta'} \right) \right]^{-1} \Bigg|_{\beta_n, \sigma_n^2} \frac{\partial \ln L}{\partial \beta} \Bigg|_{\beta_n}$$

Testes Assintóticos

Testa restrições não lineares.

Ex: $H_0 : \theta_1 \theta_2 = 2$

1) Teste da razão de verossimilhança.

Considere $L(\theta)$ a função de verossimilhança e $\hat{\theta}$ o estimador de max. verossimilhança de θ, e seja θ_0 o estimador de max. verossilhança restrito então:

$$\frac{L(\theta_0)}{L(\hat{\theta})} \leq 1$$

Aplicando logaritmo,

$$\ln L(\theta_0) - \ln L(\hat{\theta}) \leq 0$$

ou

$$2\left[\ln L(\hat{\theta}) - \ln L(\theta_0)\right] \geq 0$$

$$LR = 2\left[\ln L(\hat{\theta}) - \ln L(\theta_0)\right] \sim \chi^2(J)$$

onde *J* é o número de restrições, ou a dimensão do vetor θ_0 .

2) Teste de Wald

Considere J restrições, $g(\theta)$ é um vetor $J\times 1$

$$H_0 : g(\theta) = 0$$

$$W = g\left(\hat{\theta}_J\right)' \left(\hat{Var}(\hat{g})\right)^{-1} g\left(\hat{\theta}_J\right) \sim \chi^2_J$$

$$\hat{Var}(\hat{g}) = \left(\frac{\partial \hat{g}}{\partial \hat{\theta}}\right)' \hat{\text{var}}(\hat{\theta}) \left(\frac{\partial \hat{g}}{\partial \hat{\theta}}\right)$$

Exemplo: $Y = \alpha + \beta X_t + \varepsilon_t$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \qquad u_t \sim N\left(0, \sigma_u^2\right)$$
$$E\left(u_t, \varepsilon_{t-1}\right) = 0$$

$$Y_t - \rho Y_{t-1} = \alpha\left(1-\rho\right) + \beta\left(X_t - \rho X_{t-1}\right) + u_t$$

$$Y_t = \alpha\left(1-\rho\right) + \beta X_t - \beta\rho X_{t-1} + \rho Y_{t-1} + u_t$$

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 Y_{t-1} + u_t$$

$$\beta_1 = \alpha\left(1-\rho\right)$$
$$\beta_2 = \beta$$
$$\beta_3 = -\beta\rho \qquad \beta_2 \cdot \beta_4 + \beta_3 = 0$$
$$\beta_4 = \rho$$

$$W = \frac{\left(\hat{\beta}_2\hat{\beta}_4 + \hat{\beta}_3\right)^2}{\hat{Var}(\hat{g})} \qquad \hat{Var}(\hat{g}) = \left(\frac{\partial \hat{g}}{\partial \hat{\beta}}\right)' \hat{\text{var}}(\hat{\beta}) \left(\frac{\partial \hat{g}}{\partial \hat{\beta}}\right)$$

$$g(\beta) = \beta_2\beta_4 + \beta_3 \qquad \frac{\partial g}{\partial \beta}' = \left[\frac{\partial g}{\partial \beta_1} \frac{\partial g}{\partial \beta_2} \frac{\partial g}{\partial \beta_3} \frac{\partial g}{\partial \beta_4}\right] = \begin{bmatrix} 0 & \hat{\beta}_4 & 1 & \hat{\beta}_2 \end{bmatrix}$$

$$\hat{Var}(\hat{g}) = \begin{bmatrix} 0 & \hat{\beta}_4 & 1 & \hat{\beta}_2 \end{bmatrix} \cdot \begin{bmatrix} Var(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1\hat{\beta}_2) & \text{cov}(\hat{\beta}_1\hat{\beta}_3) & \text{cov}(\hat{\beta}_1\hat{\beta}_4) \\ \text{cov}(\hat{\beta}_1\hat{\beta}_2) & Var(\hat{\beta}_2) & \text{cov}(\hat{\beta}_2\hat{\beta}_3) & \text{cov}(\hat{\beta}_2\hat{\beta}_4) \\ \text{cov}(\hat{\beta}_1\hat{\beta}_3) & \text{cov}(\hat{\beta}_2\hat{\beta}_3) & Var(\beta_3) & \text{cov}(\hat{\beta}_3\hat{\beta}_4) \\ \text{cov}(\hat{\beta}_1\hat{\beta}_4) & \text{cov}(\hat{\beta}_2\hat{\beta}_4) & \text{cov}(\hat{\beta}_3\hat{\beta}_4) & Var(\beta_4) \end{bmatrix} \begin{bmatrix} 0 \\ \hat{\beta}_4 \\ 1 \\ \hat{\beta}_2 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_4 \text{cov}(\hat{\beta}_1\hat{\beta}_2) + \text{cov}(\hat{\beta}_1\hat{\beta}_3) + \hat{\beta}_2 \text{cov}(\hat{\beta}_1\hat{\beta}_4) & \hat{\beta}_4 Var(\hat{\beta}_2) + \text{cov}(\hat{\beta}_2\hat{\beta}_3) + \hat{\beta}_2 \text{cov}(\hat{\beta}_2\hat{\beta}_4) \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta}_4 \text{cov}(\hat{\beta}_2\hat{\beta}_3) + Var(\hat{\beta}_3) + \hat{\beta}_2 \text{cov}(\hat{\beta}_3\hat{\beta}_4) & \hat{\beta}_4 \text{cov}(\hat{\beta}_2\hat{\beta}_4) + \text{cov}(\hat{\beta}_3\hat{\beta}_4) + \hat{\beta}_2 Var(\hat{\beta}_4) \end{bmatrix} \begin{bmatrix} 0 \\ \hat{\beta}_4 \\ 1 \\ \hat{\beta}_2 \end{bmatrix}$$

$$= \hat{\beta}_4^2 Var\left(\hat{\beta}_2\right) + \hat{\beta}_4 \operatorname{cov}\left(\beta_2 \beta_3\right) + \hat{\beta}_4 \hat{\beta}_2 \operatorname{cov}\left(\hat{\beta}_2 \hat{\beta}_4\right) + \beta_4 \operatorname{cov}\left(\beta_2 \beta_3\right) + Var\left(\hat{\beta}_3\right) +$$

$$+ \hat{\beta}_2 \operatorname{cov}\left(\beta_3 \beta_4\right) + \hat{\beta}_4 \hat{\beta}_2 \operatorname{cov}\left(\beta_2 \beta_4\right) + \hat{\beta}_2 \operatorname{cov}\left(\beta_3 \beta_4\right) + \hat{\beta}_2^2 Var\left(\hat{\beta}_4\right)$$

$$\hat{Var}\left(\hat{g}\right) = Var\left(\beta_2 \beta_4 + \beta_3\right) = \hat{\beta}_4^2 Var\left(\hat{\beta}_2\right) + Var\left(\hat{\beta}_3\right) + \hat{\beta}_2^2 Var\left(\hat{\beta}_4\right) +$$

$$2 \cdot \hat{\beta}_4 \operatorname{cov}\left(\hat{\beta}_2, \hat{\beta}_3\right) + 2 \hat{\beta}_2 \hat{\beta}_4 \operatorname{cov}\left(\hat{\beta}_2, \hat{\beta}_4\right) + 2 \hat{\beta}_2 \operatorname{cov}\left(\hat{\beta}_3 \hat{\beta}_4\right)$$

3) Teste do Multiplicador de Lagrange

Dado que $\dfrac{\partial \ln L(\theta_0)}{\partial(\theta)} \sim N(0,\Sigma)$

$$LM = \frac{\partial \ln L(\theta_0)}{\partial(\theta)'} \hat{\Sigma}^{-1} \frac{\partial \ln L(\theta_0)}{\partial(\theta)'} \sim \chi^2(J)$$

$$\Sigma = -E\left(\frac{\partial^2 \ln L}{\partial\theta\partial\theta'}\right) = I(\theta_0)$$

Exemplo:

Seja $Y_1, Y_2, ..., Y_T$ uma amostra aleatória retirada de uma população com distribuição Normal com média β e variância 1, então

$$\ln L(\beta) = -\frac{T}{2}\log(2\Pi) - \frac{1}{2}\sum_{i=1}^{T}(Y_i - \beta)^2$$

$$\frac{\partial \ln L(\beta)}{\partial \beta} = \sum (Y_i - \beta) = 0 \quad \rightarrow \quad \hat{\beta} = \frac{\sum Y_i}{T}$$

$$\frac{\partial^2 \ln L(\beta)}{\partial \beta^2} = -T$$

$$I(\beta) = -E\left(\frac{\partial^2 \ln L(\beta)}{\partial \beta^2}\right) = T$$

Testar a hipótese $H_0 : \beta = \beta_0$

$$LR = 2\left[-\frac{T}{2}\ln 2\Pi - \frac{1}{2}\sum (Y_i - \hat{\beta})^2 + \frac{T}{2}\ln 2\Pi + \frac{1}{2}\sum (Y_i - \beta_0)^2 \right] = \sum (Y_i - \beta_0)^2 - \sum (Y_i - \hat{\beta})^2$$

$$LR = \sum Y_i^2 - 2\sum Y_i \beta_0 + T\beta_0^2 - \sum Y_i^2 + 2\sum Y_i \hat{\beta} - T\hat{\beta}^2 = -2T\hat{\beta}\beta_0 + T\beta_0^2 + 2T\hat{\beta}\hat{\beta} - T\hat{\beta}^2$$

$$LR = T\hat{\beta}^2 - 2T\hat{\beta}\beta_0 + T\beta_0^2 = T(\hat{\beta} - \beta_0)^2$$

$$W = (\hat{\beta} - \beta_0)^2 I(\hat{\beta}) = T(\hat{\beta} - \beta_0)^2$$

$$LM = \left(\frac{\partial \ln L(\beta_0)}{\partial \beta} \right)^2 \left[I(\beta_0) \right]^{-1}$$

$$\frac{\partial \ln L(\beta_0)}{\partial \beta} = \quad \sum (Y_i - \beta_0) = \sum Y_i - T\beta_0 = T(\hat{\beta} - \beta_0)$$

$$LM = \left[T(\hat{\beta} - \beta_0)\right]^2 \frac{1}{T} = T(\hat{\beta} - \beta_0)^2$$

Neste caso os testes são numericamente idênticos porque a função de máxima verossimilhança é quadrática.

Berndt e Savin (1977) mostraram que em amostras pequenas e em modelos lineares, LM≤LR≤W para os mesmos dados e testando as mesmas restrições. Portanto, pode haver conflito de resultados usando um teste ou outro.