NMSA459 - Session 4 - 2018/2019

Box-Cox-transformations

$Y_1,\dots,Y_m \overset{iid}{\sim} F$, $Y_1>0$ a.s. Suppose that for some $\lambda\in\mathbb{R}$ and $g_\lambda(y) = \begin{cases} \frac{y^\lambda-1}{\lambda} & y\neq 0 \\ \log \lambda & \lambda=0\end{cases}$ is $g_\lambda(Y_1)\sim N(\mu,\sigma^2)$. Estimate λ using MLE.

Remarks: 1) $\lim\limits_{\lambda\to0}\frac{y^\lambda-1}{\lambda} = \lim\limits_{\lambda\to0}\frac{e^{\lambda\log y}-1}{\lambda} = \lim\limits_{\lambda\to0}\sum_{k=0}^\infty \frac{(\lambda\log y)^k}{k!\,\lambda} - \frac{1}{\lambda} = \lim\limits_{\lambda\to0}\sum_{k=1}^\infty \lambda^{k-1}\frac{(\log y)^k}{k!} = \log y$

2) $\underset{\lambda\neq0}{Y>0}$ a.s. $\Rightarrow$ $g_\lambda(Y) > -1/\lambda$ a.s. and normality cannot hold true exactly.
for $\lambda>0$

In exact inference one supposes that $g_\lambda(Y_1)\sim$ truncated normal $(X|X>-1/\lambda)$ for $X\sim N(\mu,\sigma^2)$. Its density $(\lambda>0)$ is $\varphi\left(\frac{x-\mu}{\sigma}\right)\frac1\sigma I[x>g_\lambda(0)] \Big/ \int_{g_\lambda(0)}^\infty \varphi\left(\frac{x-\mu}{\sigma}\right)\frac1\sigma\, dx$ and depends on μ,σ^2 in a complicated way. At least if μ/σ is quite large, this can be neglected.

3) $g_\lambda(Y_1)\sim N(\mu,\sigma^2) \iff Y_1\sim g_\lambda^{-1}(N(\mu,\sigma^2))$, or its truncated version. This is actually the model we work in. $Y_1,\dots,Y_m \overset{iid}{\sim} g_\lambda^{-1}(N(\mu,\sigma^2))$ with nuisance parameters μ,σ^2 and the parameter of interest λ.

Estimation: $F(y) = P(Y_1\le y) = P(g_\lambda(Y_1)\le g_\lambda(y)) = \Phi\left(\frac{g_\lambda(y)-\mu}{\sigma}\right) \quad / \frac{\partial}{\partial y}$

$$f(y) = \varphi\left(\frac{g_\lambda(y)-\mu}{\sigma}\right)\frac1\sigma\cdot\frac{\partial}{\partial y}g_\lambda(y) = \varphi\left(\frac{g_\lambda(y)-\mu}{\sigma}\right)\frac{y^{\lambda-1}}{\sigma}$$

$$L(\lambda,\mu,\sigma^2) = \prod_{i=1}^m \varphi\left(\frac{g_\lambda(y_i)-\mu}{\sigma}\right)\frac{y_i^{\lambda-1}}{\sigma}$$

$$\ell(\lambda,\mu,\sigma^2) = -\frac m2\log2\pi - \frac m2\log\sigma^2 - \frac1{2\sigma^2}\sum(g_\lambda(y_i)-\mu)^2 + (\lambda-1)\sum\log y_i$$

For the confidence interval for λ, suppose $H_0:\lambda=\lambda_0$ and find estimators of μ and σ^2 under H_0

For λ fixed, clearly $\tilde\sigma^2 = \frac1m\sum(g_\lambda(y_i)-\tilde\mu)^2$ and $\tilde\mu=\frac1m\sum g_\lambda(y_i)$

Likelihood ratio test: $LR = 2(\ell(\hat\lambda,\hat\mu,\hat\sigma^2) - \ell(\lambda_0,\tilde\mu,\tilde\sigma^2)) \overset{as}{\underset{H_0}{\sim}} \chi^2_1$

Profile likelihood $\ell^{(P)}(\lambda) = \ell(\lambda,\tilde\mu,\tilde\sigma^2) = -\frac m2\log2\pi - \frac m2\log\tilde\sigma^2 - \frac m2 + (\lambda-1)\sum\log y_i$

Profile likelihood is maximized at MLE, $\sup_\lambda \ell^{(P)}(\lambda) = \ell^{(P)}(\hat\lambda) = \sup_{\lambda,\mu,\sigma^2}\ell(\lambda,\mu,\sigma^2)$

Therefore $LR = 2(\sup_\lambda \ell^{(P)}(\lambda) - \ell^{(P)}(\lambda_0)) \overset{as}{\underset{H_0}{\sim}}\chi^2_1$ and confidence interval for λ is

$$\left\{\lambda: \sup_\lambda \ell^{(P)}(\lambda) - \frac{\chi^2(1-\alpha)}{2} \le \ell^{(P)}(\lambda)\right\}$$

In regression analysis: $\binom{X_1}{Y_1}\cdots\binom{X_m}{Y_m} \overset{iid}{\sim} F_{X,Y}$, $\left[\begin{array}{l} E\,g_\lambda(Y_i) = X_i'\beta \text{ given } X_i \\ \operatorname{var} g_\lambda(Y_i) = \sigma^2\end{array}\right]$

$$g_\lambda(Y_i|X_i) \sim \cancel{(g_\lambda(Y_i)|X_i)} \sim N(X_i'\beta,\sigma^2)$$

Profile likelihood: $\ell^{(P)}(\lambda) = -\frac m2\log2\pi - \frac m2\log\tilde\sigma^2 - \frac m2 + (\lambda-1)\sum\log y_i$

for $\tilde\sigma^2 = \frac1m\sum(g_\lambda(y_i) - x_i'\tilde\beta)^2$ and $\tilde\beta = (X'X)^{-1}X'g_\lambda(Y)$

Note that $\ell^{(P)}(\lambda) = c - \frac m2\left(\log\tilde\sigma^2 - \log\left(\prod y_i\right)^{\frac{2(\lambda-1)}{m}}\right) = c - \frac m2\log\left(\frac1m\sum\left(\frac{g_\lambda(y_i)}{(\dot y)^{\lambda-1}} - x_i'\frac{\tilde\beta}{(\dot y)^{\lambda-1}}\right)^2\right)$

with $\dot y = \sqrt[m]{\prod y_i}$ the geometric mean of the response. Thus $\ell^{(P)}(\lambda)$ is in fact sum of squares in a transformed model.

NMSA434 – Ex. session 4 – profile likelihood 2018/2019

Ex: GLM Poisson model

$\binom{X_1}{Y_1}, \dots, \binom{X_n}{Y_n}$ iid, $Y_i | X_i \sim Po(\lambda(X_i))$, $\lambda(x) = e^{\beta_0 + \beta_1 x}$

$$L(\beta) = \prod_{i=1}^{n} \exp\{-\lambda(x_i)\} \frac{\lambda(x_i)^{y_i}}{y_i!} = \exp\{-\sum \lambda(x_i)\} \frac{\prod \lambda(x_i)^{y_i}}{\prod y_i!}$$

$$\ell(\beta) = c - \sum \lambda(x_i) + \sum y_i \log \lambda(x_i) = c - \sum e^{\beta_0 + \beta_1 x_i} + \sum y_i (\beta_0 + \beta_1 x_i)$$

$$\frac{\partial \ell}{\partial \beta} = \begin{pmatrix} -e^{\beta_0} \sum e^{\beta_1 x_i} + \sum y_i \\ -e^{\beta_0} \sum x_i e^{\beta_1 x_i} + \sum x_i y_i \end{pmatrix} \rightsquigarrow e^{\hat\beta_0} = \frac{\sum y_i}{\sum e^{\beta_1 x_i}}$$

$$\hat\beta_0 = \log\left(\frac{\sum y_i}{\sum e^{\beta_1 x_i}}\right)$$

Profile likelihood: $\ell^{(P)}(\beta_1) = \ell((\hat\beta_0(\beta_1), \beta_1)) =$

$$= c - \frac{\sum y_i}{\sum e^{\beta_1 x_i}} \sum e^{\beta_1 x_i} + \sum y_i \left(\log\left(\frac{\sum y_i}{\sum e^{\beta_1 x_i}}\right) + \beta_1 x_i\right)$$

Likelihood ratio test. confidence interval

$$LR = 2\left(\ell(\hat\beta) - \ell(\hat\beta_0(\beta_1), \beta_1)\right) = 2\left(\sup_{\beta_1} \ell^{(P)}(\beta_1) - \ell^{(P)}(\beta_1)\right) \overset{as}{\sim} \chi^2_1$$

Ex: Profile likelihood in $N(\mu, \sigma^2)$

$X_1, \dots X_n \sim N(\mu, \sigma^2)$ μ of interest, σ^2 nuissance

profile log-lik: $\ell_n^{(P)}(\mu) := \sup_{\sigma^2 > 0} \ell_n(\mu, \sigma^2)$ σ^2 is profiled out of the likelihood

maximiser of $\ell_n^{(P)}(\mu)$ is the MLE of μ.

$\ell_n(\mu, \sigma^2) = c - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2$ μ fixed, maximized in $\hat\sigma^2(\mu) = \frac{1}{n} \sum (x_i - \mu)^2$ and

$$\ell_n^{(P)}(\mu) = c - \frac{n}{2} \log \hat\sigma^2(\mu) - \frac{n}{2}$$

for a test of $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$

$$LR_n = 2\left(\log L_n(\hat{\mu}, \hat{\sigma}^2(\hat{\mu})) - \log L_n(\mu_0, \hat{\sigma}^2(\mu_0))\right) =$$

$$= 2\left(\sup_{\substack{\mu \in \mathbb{R} \\ \sigma^2 > 0}} \ell_n(\mu, \sigma^2) - \sup_{\sigma^2 > 0} \ell_n(\mu_0, \sigma^2)\right) =$$

$$= 2\left(\ell^{(p)}(\hat{\mu}) - \ell^{(p)}(\mu_0)\right) \underset{H_0}{\overset{as}{\sim}} \chi^2_1$$

and profile likelihood can be used to test in the same way as a LR test with nuissance parameters.

NMST 434
Ex. session 4

(28) 24) MLE in AR(1) model - dependent observations

$X_t = \psi + \varphi X_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} N(0,\sigma^2)$ independent of $X_{t-1} \qquad |\varphi|<1$

stationary process $\quad EX_t = \psi + \varphi EX_{t-1} + 0 \Rightarrow EX_t = \frac{\psi}{1-\varphi} =: \mu$

$\operatorname{var} X_t = \varphi^2 \operatorname{var} X_{t-1} + \sigma^2 \Rightarrow \operatorname{var} X_t = \sigma^2/1-\varphi^2$

$\operatorname{cov}(X_t, X_{t-1}) = \operatorname{cov}(X_{t-1}\cdot\varphi, X_{t-1}) = \varphi \cdot \operatorname{var} X_{t-1} = \varphi\sigma^2/1-\varphi^2$

$\Theta = (\psi, \varphi, \sigma^2)$

In general the joint density is given by

$p_\theta(x_1 \dots x_m) = p_\theta(x_1)\, p_\theta(x_2|x_1)\, p_\theta(x_3|x_1,x_2) \dots p_\theta(x_m|x_1, x_2 \dots x_{m-1})$

$L(\theta) = p_\theta(x_1) \prod_{j=2}^m p_\theta(x_j|x_1 \dots x_{j-1}) \underset{\wedge}{=} p_\theta(x_1) \prod_{j=2}^m p_\theta(x_j|x_{j-1})$

$x_j | x_{j-1} \sim N(\psi + \varphi x_{j-1}, \sigma^2)$

$$= \underbrace{p_\theta(x_1)}_{L_1(\theta)} \underbrace{(2\pi\sigma^2)^{-m/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=2}^m (x_j - \psi - \varphi x_{j-1})^2\right\}}_{L_2(\theta)}$$

$L_1(\theta)$ is usually ignored in computations by considering $x_1 = 0$ a.s. Though, marginally $X_1 \sim N\left(\frac{\psi}{1-\varphi}, \frac{\sigma^2}{1-\varphi^2}\right)$ and this brings non-linearities into computation, can be solved only numerically. We work with ~~marginal~~ conditional likelihood $X_2 \dots X_m | X_1$

$\ell(\theta) = c - \frac{m}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum(x_j - \psi - \varphi x_{j-1})^2$

solutions as in linear models $\Rightarrow$ define $\quad \beta = \begin{pmatrix}\psi\\ \varphi\end{pmatrix} \quad X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{m-1}\end{pmatrix}$

$Y = (X_2, \dots X_m)^T$. MLE solutions $\hat\beta$ maximize $\quad (Y - X\beta)'(Y-X\beta)$

$$\Rightarrow \hat\beta = \begin{pmatrix}\hat\psi\\ \hat\varphi\end{pmatrix} = (X'X)^{-1}X'Y = \begin{pmatrix} m-1 & \sum x_{j-1} \\ \sum x_{j-1} & \sum x_{j-1}^2\end{pmatrix}^{-1} \begin{pmatrix}\sum x_j \\ \sum x_j x_{j-1}\end{pmatrix}$$

$$\hat\sigma^2 = \frac{1}{m}\sum(x_j - \hat\psi - \hat\varphi x_{j-1})^2.$$

asymptotics: $\quad \frac{\partial \ell}{\partial \psi} = \frac{1}{\sigma^2}\sum(x_j - \psi - \varphi x_{j-1})$

$\frac{\partial \ell}{\partial \varphi} = \frac{1}{\sigma^2}\sum(x_j - \psi - \varphi x_{j-1})x_{j-1}$

$\frac{\partial \ell}{\partial \sigma^2} = -\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(x_j - \psi - \varphi x_{j-1})^2$

$\frac{\partial^2 \ell}{\partial \psi^2} = -\frac{1}{\sigma^2}\cdot(m-1) \qquad \frac{\partial^2 \ell}{\partial\psi\partial\varphi} = -\frac{1}{\sigma^2}\sum x_{j-1} \qquad \frac{\partial^2 \ell}{\partial\psi\partial\sigma^2} = -\frac{1}{\sigma^4}\sum\overbrace{(x_j - \psi - \varphi x_{j-1})}^{\varepsilon_t}$

$\frac{\partial^2 \ell}{\partial\varphi^2} = -\frac{1}{\sigma^2}\sum x_{j-1}^2 \qquad \frac{\partial^2 \ell}{\partial\varphi\partial\sigma^2} = -\frac{1}{\sigma^4}\sum(x_j - \psi - \varphi x_{j-1})x_{j-1}$

$\frac{\partial^2 \ell}{\partial(\sigma^2)^2} = +\frac{m}{2\sigma^4} - \frac{\sum(x_j - \psi - \varphi x_{j-1})^2}{\sigma^6}$

$$\lim_{m\to\infty} \frac{1}{m} J_m(\theta) = \lim_{m\to\infty} \frac{1}{m} \begin{pmatrix} \frac{(m-1)}{\sigma^2} & \frac{(m-1)\mu}{\sigma^2} & 0 \\ \frac{(m-1)\mu}{\sigma^2} & \frac{m-1}{\sigma^2}\cdot\left(\frac{\sigma^2}{1-\varphi^2}+\mu^2\right) & 0 \\ 0 & 0 & -\frac{m}{2\sigma^4}+\frac{(m-1)}{\sigma^4} \end{pmatrix}$$

$$= \begin{pmatrix} 1/\sigma^2 & \mu/\sigma^2 & 0 \\ \mu/\sigma^2 & \left(\frac{\sigma^2}{1-\varphi^2}+\mu^2\right)\frac{1}{\sigma^2} & 0 \\ 0 & 0 & \frac{\cancel{m}\,1}{2\sigma^4} \end{pmatrix} =: \overline{I}(\theta)$$

$$\sqrt{m}\left(\begin{pmatrix}\hat{\gamma}\\ \hat{\varphi}\\ \hat{\sigma}^2\end{pmatrix} - \begin{pmatrix}\gamma\\ \varphi\\ \sigma^2\end{pmatrix}\right) \xrightarrow{d} N_3\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \overline{I}(\theta)^{-1}\right)$$

special case for $\gamma = 0$, i.e. $X_t = \varphi X_{t-1} + Z_t$ $\quad \mu = 0 \Rightarrow$ centred process

and for $\theta = (\varphi, \sigma^2)'$ $\quad \hat{\varphi} = \frac{1}{\sum x_{j-1}^2} \sum x_j x_{j-1}$ $\quad \hat{\sigma}^2 = \frac{1}{m}\sum (x_j - \hat{\varphi} x_{j-1})^2$

$$\sqrt{m}\left(\begin{pmatrix}\hat{\varphi}\\ \hat{\sigma}^2\end{pmatrix} - \begin{pmatrix}\varphi\\ \sigma^2\end{pmatrix}\right) \xrightarrow{d} N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} 1/1-\varphi^2 & 0 \\ 0 & 1/2\sigma^4\end{pmatrix}^{-1}\right) = \left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} 1-\varphi^2 & 0 \\ 0 & 2\sigma^4\end{pmatrix}\right).$$