Lecture course *Statistics II*
Winter semester 2016/17
Ruprecht-Karls-Universität Heidelberg

Prof. Dr. Jan JOHANNES
Xavier LOIZEAU

Exercise sheet 3

Exercise 1. (Lemma §3.2.2) For each $n \in \mathbb{N}$ let X_n and Y_n be r.v.'s defined on a common probability space $(\Omega_n, \mathscr{A}_n, \mathbb{P}_n)$. Show the following statements.

(a) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, then $(X_n, Y_n) \xrightarrow{d} (X, c)$.

(b) $X_n \xrightarrow{d} X$ if and only if $\liminf_{n\to\infty} \mathbb{E} f(X_n) \geqslant \mathbb{E} f(X)$ for any non-negative and continuous function f (not necessarily bounded). (4 points)

Exercise 2. (Le Cam's first lemma, §3.2.3) Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be probability measures on measurable spaces $(\Omega_n, \mathscr{A}_n)$ for each $n \in \mathbb{N}$. Consider the following statements:

(a) If $d\mathbb{P}_n/d\mathbb{Q}_n \xrightarrow{d} U$ under $\mathbb{Q}_n$ along a sub-sequence, then $\mathbb{E}\mathbb{1}_{\{U>0\}} = 1$.

(b) If $d\mathbb{Q}_n/d\mathbb{P}_n \xrightarrow{d} V$ under $\mathbb{P}_n$ along a sub-sequence, then $\mathbb{E}V = 1$.

Show that the statement (a) implies (b). (4 points)

Exercise 3. Let the assumptions of Theorem §2.3.2 be satisfied. Assume another estimator $\check{\theta}_n$ of θ_o satisfying $\left\|\check{\theta}_n - \theta_o\right\| = O_{\mathbb{P}}(n^{-1/2})$. Consider the following up date of this estimator:

$$\tilde{\theta}_n = \check{\theta}_n - \ddot{M}_n(\check{\theta}_n)^{-1}\dot{M}_n(\check{\theta}_n).$$

Show that $\widetilde{\theta}_n = \widehat{\theta}_n + o_{\mathbb{P}}(n^{-1/2})$ and that $\sqrt{n}(\widetilde{\theta}_n - \theta_o)$ has the same asymptotic normal limit as $\sqrt{n}(\widehat{\theta}_n - \theta_o)$. (4 points)

Exercise 4. (Exponential frailty model) Let X, Y be r.v.'s which are conditionally independent given an unobserved r.v. Z, that is, for $\lambda, \theta > 0$ hold

$$\begin{aligned} Z &\sim \mathfrak{E}\mathrm{xp}(\lambda) \\ X, Y\,|Z &\sim \mathfrak{E}\mathrm{xp}(Z) \cdot \mathfrak{E}\mathrm{xp}(\theta Z). \end{aligned}$$

Consequently, $f(z) = \lambda \exp(-\lambda z)$ is the density of Z and $f(x, y|z) = z\exp(-zx) \cdot z\theta\exp(-z\theta y)$ is the conditional joint density of (X, Y) given Z. In the sequel let $\lambda = 2$. Using the software environment R (see r-project.org) we intend to analyse the MLE of θ by working through the four steps below. For the computational art you may use the template „exp_frailty_model_gaps.R" downloadable on the course website and fill in the gaps. Fell free, of course, to write the code by yourself.

(1) Show that the joint density of (X, Y) is given by $f(x, y) = 2\lambda\theta(x + \theta y + \lambda)^{-3}$.
Hint: You may use that $f(x, y) = \int_0^\infty f(x, y, z)dz = \int_0^\infty f(x, y|z)f(z)dz$.

(2) Simulate 1 000, 2 000, 5 000, 10 000, 50 000, 100 000 i.i.d. copies of the exponential frailty model with $\lambda = 2$ and $\theta = 0.5$. Compute in each case the MLE and find a suitable way of visualising the values of these six estimators in relation to the true value.
Hint: To maximise the log-likelihood you may want to use the R-function nlm *by implementing a function* Q_n *which returns the negative log-likelihood in dependence of* θ *(which needs to be the first argument of* Q_n*). If your function* Q_n *needs further arguments you can just add these arguments to your call of* nlm*. Moreover,* nlm *requires an initial value for* θ*, this value doesn't matter, choose 5 for example. If* Q_n *expects three arguments, i.e.,* $Q_n(\theta, v_1, v_2)$*, then use the following call* $nlm(Q_n, 5, v_1, v_2)$*. Keep in mind that* nlm *requires* Q_n *to handle vectors in its first argument, the output in this case should be a vector of the same length where each entry is the output of* Q_n *if* Q_n *was called with the corresponding element of the input vector. Finally, for* nlm *to run without warnings you need that* Q_n *returns always a number, so you need to detect input that would create* Inf *or* NaN *and handle it appropriately.*

(3) Focus on the case $n = 1000$ only. Simulate 1 000 i.i.d. samples of the exponential frailty model (above) and compute the MLE. Repeat this 20 000 times and plot the histogram of $\sqrt{n}(\theta_0 - \hat{\theta})$. Estimate also the variance and add a plot of a normal density with the corresponding variance to the histogram. What do you think? Is the approximation by a normal density good? What do think about the variance?

(4) Focus on the case $n = 1000$ only. Simulate 500 i.i.d. samples of the above model and compute the Wald test statistic. Repeat this 10 000 times and plot a histogram of the test statistics together with the density of a chi squared distribution with the correct number of degrees of freedom. Repeat this simulation scheme (10 000 simulations of 500 samples) but this time with a different value for θ. In each case check whether the Wald test of level $\alpha = 0.05$ would reject the hypothesis $\theta = 0.5$ and compute the percentage of rejections. Do this for different values of θ and visualise your results in a plot.

(5) **Bonus (+2 points):** Research in the internet what the exponential frailty model is good for. Is the distribution of Z always assumed to be exponential? (8 points)

Handing in on **Friday, November 18, 2016** in **fixed groups of two**.