STATISTICAL THEORY – EXAMPLES 1/2

Part III, Michaelmas 2013, RN

Ex. 1 Let Θ be an open interval (possibly equal to $\mathbb{R}$) and let S_n be a sequence of random real-valued continuous functions defined on Θ such that, as $n \to \infty$, $S_n(\theta) \to^P S(\theta)$ $\forall \theta \in \Theta$, where $S : \Theta \to \mathbb{R}$ is nonrandom. Suppose for some $\theta_0 \in \Theta$ and every $\varepsilon > 0$ small enough we have $S(\theta_0 - \varepsilon) < 0 < S(\theta_0 + \varepsilon)$, and that S_n has *exactly one* zero $\hat{\theta}_n$ for every $n \in \mathbb{N}$. Deduce that $\hat{\theta}_n \to^P \theta_0$ as $n \to \infty$.

Ex. 2 Give an example of (possibly random) functions Q_n, Q defined on $\Theta \subset \mathbb{R}$ that have unique maximisers $\hat{\theta}_n, \theta_0$, respectively, such $Q_n(\theta) \to Q(\theta)$ (almost surely) for every $\theta \in \Theta$ as $n \to \infty$, but $\hat{\theta}_n \not\to \theta_0$ (almost surely).

Ex. 3 Derive an analogue of the consistency result for the maximum likelihood estimator for the nonlinear least squares estimator with random design, under the assumptions that the $Y_i = (Z_i, x_i)$ are i.i.d., that $E(y_i|x_i) = g(x_i, \theta_0)$ for some $\theta_0 \in \Theta$. Which further assumptions on g do you need (be as economical as you can)? Show that the general normal linear model

$$Y = X\theta + u$$

with X a $n \times p$ matrix, $\theta \in \mathbb{R}^p$, $u \sim N(0, \sigma^2 I_n), \sigma^2 > 0$, is a special case of the NLS model, and show further that the uniqueness condition for θ_0 is satisfied if the $n \times p$ matrix X has full column rank.

Ex. 4 * Consider the problem of Exercise 3 above, but now with $\Theta = \mathbb{R}$. Assuming that

$$E\left[\inf_{\theta} \left(\frac{\partial g(x_t, \theta)}{\partial \theta}\right)^2\right] > 0,$$

show that one can find a compact set $\Theta^* = [\theta_0 - M, \theta_0 + M]$ such that

$$\inf_{\theta \notin \Theta^*} Q_n(\theta) > Q_n(\theta_0)$$

with probability approaching one, and use this to prove consistency of the NLS estimator. How does the condition on the derivative of g simplify in linear regression where $g(x_i, \theta) = x_i\theta$?

Ex. 5 Formulate mild conditions on $K(\theta)$ such that the conditions for asymptotic normality of maximum likelihood estimators are satisfied for an exponential family of order 1.

Ex. 6 Let $Y_1, \dots, Y_n$ be i.i.d. $N(\mu, \sigma^2)$ distributed. Derive the maximum likelihood estimator for (μ, σ^2) and show that it is asymptotically normal. Calculate the Fisher information and its inverse for this parameter.

Ex. 7 Consider $Y_1, \dots, Y_n$ i.i.d. Poisson random variables with parameter λ. Derive explicit formulas for the MLE and for the likelihood ratio test statistic for testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda \neq \lambda_0$. Deduce the asymptotic distribution of $\sqrt{n}(\hat{\lambda}_n - \lambda)$ directly, and verify that it agrees with what the general asymptotic theory predicts.

Ex. 8 Consider an i.i.d. sample $X_1, \dots, X_n$ arising from the model

$$\{f(\cdot, \theta) : \theta \in \mathbb{R}\}, \quad f(x, \theta) = \frac{1}{2} e^{-|x-\theta|}, x \in \mathbb{R},$$

of *Laplace distributions*. Assuming n to be odd for simplicity, show that the MLE is equal to the sample median. Discuss what happens when n is even. Can you calculate the Fisher information?

Ex. 9 Let $Y_1, \dots, Y_n$ be independent $U(0, \theta)$ random variables, where $\theta \in \Theta \equiv [1, 2]$. Find the maximum likelihood estimator $\hat{\theta}_n$ as well as its distribution function, mean and variance. What is the asymptotic distribution of $n(\theta - \hat{\theta}_n)/\theta$? Why does the standard theory not apply?

Ex. 10 *[Estimation of the Fisher Information.]* Assuming that the regularity conditions for asymptotic normality of maximum likelihood estimators are satisfied, prove that

$$\hat{i}_n = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f(\hat{\theta}_n, Y_i)}{\partial \theta \partial \theta^T} \to^{P_{\theta_0}} i(\theta_0) \quad \text{as } n \to \infty.$$

Ex. 11 *[Confidence Sets and the Wald test.]* Working in the framework and under the regularity conditions for asymptotic normality of maximum likelihood estimators, and using Exercise 10, construct a random set $C_n \in \mathbb{R}^p$ (a 'confidence region') that depends only on α and $Y_1, \dots, Y_n$ such that

$$\lim_n P_{\theta_0}(\theta_0 \in C_n) = 1 - \alpha.$$

If $\hat{\theta}_n$ is the MLE, derive further the asymptotic distribution of the Wald statistic

$$n(\hat{\theta}_n - \theta_0)^T \hat{i}_n (\hat{\theta}_n - \theta_0)$$

under P_{θ_0}, and use it to design an asymptotic level α test for the null hypothesis $H_0 : \theta = 0$ against $H_1 : \theta \in \Theta, \theta \neq 0$.

Ex. 12 The following result is known as Hoeffding's inequality: If $X_1, ..., X_n$ are mean zero independent random variables taking values in $[b_i, c_i]$ for constants $b_i < c_i$, then

$$\Pr\left\{\sum_{i=1}^{n} X_i > u\right\} \le \exp\left(-\frac{2u^2}{\sum_{i=1}^{n}(c_i - b_i)^2}\right) \tag{1}$$

of which an obvious consequence is (why?)

$$\Pr\left\{\left|\sum_{i=1}^{n} X_i\right| > u\right\} \le 2\exp\left(-\frac{2u^2}{\sum_{i=1}^{n}(c_i - b_i)^2}\right). \tag{2}$$

Provide a proof of this inequality. [You may find it useful to first prove the auxiliary result $E(\exp\{vX_i\}) \le \exp\{v^2(b_i - a_i)^2/8\}$ for $v > 0$, and then use Markov's inequality in conjuction with a bound for the moment generating function of $v \sum X_i$.]