PRINCIPLES OF STATISTICS

Part II, Michaelmas 2016, QB (email: q.berthet@statslab.cam.ac.uk)

Handout on Regularity Assumptions

I. Dominated convergence theorem and applications

Dominated Convergence Theorem. We need a fact from measure theory. For $\mathcal{X}$ a (measurable) subset of $\mathbb{R}$ (or, more generally, $(\mathcal{X}, \mathcal{A})$ any measurable space), consider a sequence of (measurable) functions $f_n : \mathcal{X} \to \mathbb{R}$ that converges pointwise to $f : \mathcal{X} \to \mathbb{R}$ as $n \to \infty$ and such that $|f_n(x)| \leq g(x), x \in \mathcal{X}$, for some $g : \mathcal{X} \to \mathbb{R}$ such that $E|g(X)| < \infty$, where X is any real random variable taking values in $\mathcal{X}$. Then by the dominated convergence theorem $E|f_n(X) - f(X)| \to 0$ as $n \to \infty$. Particularly if a function $q(x, \theta)$ of two arguments is continuous in $\theta \in \Theta$ for all $x \in \mathcal{X}$ and satisfies $E \sup_{\theta \in \Theta} |q(X, \theta)| < \infty$ then the mapping

$$\theta \mapsto Eq(X, \theta)$$

is continuous too (as can be seen by applying the above result with $f_n(x) = q(x, \theta_n), f(x) = q(x, \theta)$, $g(x) = \sup_{\theta \in \Theta} |q(x, \theta)|$, for any sequence $\theta_n \to \theta$ in Θ).

Upon setting $q(x, \theta) = \log f(x, \theta)$ the following proposition implies that the scaled log-likelihood function $\bar{\ell}_n(\theta) = (1/n)\ell_n(\theta) = (1/n)\sum_{i=1}^n \log f(X_i, \theta)$ converges uniformly in θ to $\ell(\theta) = E \log f(X, \theta)$ almost surely, as was required in the proof of the consistency of the maximum likelihood estimator (under appropriate conditions). Its proof relies on the uniform law of large numbers under bracketing Proposition II.2 from lectures.

Proposition 1 *Suppose Θ is a bounded and closed subset of $\mathbb{R}^p$, and let $q(x, \theta) : \mathcal{X} \times \Theta \to \mathbb{R}$ be continuous in θ for each x and measurable in x for each θ. If $X_1, ..., X_n$ are i.i.d. in $\mathcal{X}$ and if*

$$E \sup_{\theta \in \Theta} |q(X, \theta)| < \infty \tag{1}$$

then, as $n \to \infty$,

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n q(X_i, \theta) - Eq(X, \theta) \right| \to 0 \quad \Pr - a.s. \tag{2}$$

Proof. We apply Proposition II.2 from lectures, so (2) will be proved if we find a suitable bracketing covering for the class of functions

$$\mathcal{H} = \{q(\cdot, \theta) : \theta \in \Theta\},$$

which is done as follows: First define the open balls $B(\theta, \eta) = \{\theta' \in \Theta : \|\theta - \theta'\| < \eta\}$, and define, for *every* $\theta \in \Theta$, the auxiliary brackets

$$u(x, \theta, \eta) = \sup_{\theta' \in B(\theta, \eta)} q(x, \theta')$$

and

$$l(x, \theta, \eta) = \inf_{\theta' \in B(\theta, \eta)} q(x, \theta')$$

so that $l(x,\theta,\eta) \le q(x,\theta') \le u(x,\theta,\eta)$ holds for every x and every $\theta' \in B(\theta,\eta)$. By condition (1) we have

$$E|u(X,\theta,\eta)| < \infty, \quad E|l(X,\theta,\eta)| < \infty \tag{3}$$

for every $\theta \in \Theta$ and every η. Furthermore since $q(\cdot,x)$ is continuous, the suprema in the definition of $u(x,\theta,\eta)$ are attained at points $\theta^u(\theta)$ that satisfy $\|\theta^u(\theta)-\theta\| \le \eta$, and likewise for the infimum in the definition of $l(x,\theta,\eta)$. Hence, again by continuity, $\lim_{\eta\to 0}|u(x,\theta,\eta)-q(x,\theta)| \to 0$ for every x and every $\theta \in \Theta$, and an analogous result holds for the lower brackets. By the dominated convergence theorem and condition (1) we can integrate these limits and conclude

$$\lim_{\eta\to 0} E|u(X,\theta,\eta) - q(X,\theta)| \to 0 \text{ and } \lim_{\eta\to 0} E|l(X,\theta,\eta) - q(X,\theta)| \to 0$$

for every $\theta \in \Theta$. Consequently, for $\varepsilon > 0$ arbitrary and every $\theta \in \Theta$ we can find $\eta := \eta(\varepsilon,\theta)$ small enough such that

$$E|u(X,\theta,\eta) - l(X,\theta,\eta)| \le E|u(X,\theta,\eta) - q(X,\theta)| + E|q(X,\theta) - l(X,\theta,\eta)| < \varepsilon. \tag{4}$$

The open balls $\{B(\theta,\eta(\varepsilon,\theta))\}_{\theta\in\Theta}$ constitute an open cover of the compact set Θ in $\mathbb{R}^p$, so by compactness there exists a finite subcover with centers $\theta_1,...,\theta_N$, $j = 1,...,N$ (the Heine-Borel theorem from Analysis Ib). The functions $q(\cdot,\theta')$ for $\theta' \in B(\theta_j,\eta(\varepsilon,j))$ are bracketed between $u_j := u(\cdot,\theta_j,\eta(\varepsilon,j))$ and $l_j := l(\cdot,\theta_j,\eta(\varepsilon,j))$, $j = 1,...,N$, so that (3) and (4) complete the proof. ■

We note that above proof is valid, with only notational changes, for X_i taking values in a general measurable space and for Θ any compact metric space.

II. Regularity conditions for asymptotic normality of the MLE

Assumption B. Consider a model

$$\{f(\cdot,\theta), \theta \in \Theta\}, \quad \Theta \subseteq \mathbb{R}^p,$$

of pdf/pmf's on $\mathcal{X} \subseteq \mathbb{R}^k$ such that $f(x,\theta) > 0$ for all $x \in \mathcal{X}$ and all $\theta \in \Theta$, and such that $\int_{\mathcal{X}} f(x,\theta)dx = 1$ for every $\theta \in \Theta$.

Let $\theta_0 \in \Theta$ be a fixed ('true') value, and assume
1) that θ_0 is an interior point of Θ,
2) that there exists an open set U satisfying $\theta_0 \in U \subseteq \Theta$ such that $f(x,\theta)$ is, for every $x \in \mathcal{X}$, twice continuously differentiable w.r.t. θ on U,
3) The $p \times p$ matrix $E_{\theta_0}[\partial^2 \log f(X,\theta_0)/\partial\theta\partial\theta^T]$ is nonsingular and

$$E_{\theta_0}\left\|\frac{\partial \log f(X,\theta_0)}{\partial\theta}\right\|^2 < \infty,$$

4) there exists a compact ball $K \subset U$ (with nonempty interior) centered at θ_0 s.t.

$$E_{\theta_0} \sup_{\theta\in K}\left\|\frac{\partial^2 \log f(X,\theta)}{\partial\theta\partial\theta^T}\right\| < \infty,$$

$$\int_{\mathcal{X}} \sup_{\theta\in K}\left\|\frac{\partial f(x,\theta)}{\partial\theta}\right\| dx < \infty \text{ and } \int_{\mathcal{X}} \sup_{\theta\in K}\left\|\frac{\partial^2 f(x,\theta)}{\partial\theta\partial\theta^T}\right\| dx < \infty.$$

5) Suppose the MLE $\hat\theta_n$ in the model $\{f(\theta,\cdot);\theta \in \Theta\}$ based on the sample $X_1,\dots,X_n$ exists and is consistent, i.e., $\hat\theta_n \to^{P_{\theta_0}} \theta_0$ as $n \to \infty$.