Introduction

Most of the results on this list require the regularity conditions (RC) found in [1]. The appropriate mode of convergence is assumed, i.e. *a.s*, L^p, *c*, *p*, *d*. Recall that except for convergence in distribution, the $X \in R^d$ component extension theorem(s) provide equivalent mode convergence for vectors $X_n \to X \Leftrightarrow X_{i,n} \to X_i, \forall i \in 1, d$, and that a helpful result for convergence in distribution is provided by the Cramér-Wold device, $X_n \xrightarrow{d} X \Leftrightarrow c^T X_n \xrightarrow{d} c^T X \;\; \forall\, c \in R^k$.

Preliminaries

Differentiating under the integral sign

If $f(x,\theta)$, $a(\theta)$ and $b(\theta)$ are differentiable in θ, Leibnitz's rule provides

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx = f(b(\theta),\theta)\frac{d}{d\theta}b(\theta) - f(a(\theta),\theta)\frac{d}{d\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)dx$$

The problem comes when dealing with "improper" integrals, i.e., when the limits go to $\pm\infty$. Using boundedness on the variation of the derivative in θ, the LDC can be applied to the derivative, and then extended uniformly in θ. Let $f(x\,|\,\theta)$ be differentiable at $\theta = \theta_0$.

Under ***RC 0.1***, $\left.\frac{d}{d\theta}\int_{-\infty}^{\infty} f(x\,|\,\theta)dx\right|_{\theta=\theta_0} = \int_{-\infty}^{\infty}\left[\left.\frac{\partial}{\partial\theta} f(x\,|\,\theta)\right|_{\theta=\theta_0}\right]dx$. This is not really uniform in θ. By a version of the MVT, for fixed x and θ_0 and $\|\delta\| \le \delta_0$, we can write $\|\delta\| \cdot \left.\nabla f(x,\theta)\right|_{\theta=\theta_0+\delta^*(x)} = f(x,\theta_0+\delta) - f(x,\theta_0) + o\left(\|\delta\|\right)$ for some $\delta^*(x)$ where $\left\|\delta^*(x)\right\| \le \delta_0$. So RC 0.1 will be satisfied uniformly in $\theta \in \Theta$ by ***RC 0.2***

Finite Sample Size Results

Fisher Information Regularity Conditions (FIRC) – 2, 3, 4 (or 4″)

$I(\theta) \overset{4}{=} E\left[\nabla\ell\nabla\ell^T\right] \overset{4''}{=} -E\left[\nabla^2\ell\right] = -E\left[\nabla\nabla\ell^T\right] = -E[\mathrm{H}_\ell(\theta)] = E[J(\hat{\theta})]$. It really is desired to have conditions under which an estimator can have a minimum variance.

Information Inequality (Fréchet-Cramér-Rao Lower Variance Bound) RC – 2, 3, 4″

For any estimator $W_n(x)$ of $\tau(\theta)$ with $\mathrm{Var}(W_n) < \infty$, and $g(\theta) = E(W_n)$, $\mathrm{Var}(W_n) \ge \dot{g}(\theta)I_n^{-1}(\theta)\dot{g}(\theta)^T$. Recall $A \ge B \Leftrightarrow c^T A \ge c^T B \;\forall\, c \in R^k$. For unbiased $W_n = \hat{\theta}$ estimating $\tau(\theta) = \theta$, this becomes $\mathrm{Var}(\hat{\theta}) \ge I_n^{-1}(\theta)$, which is $\frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}$ in the case of $\theta \in R^1$.

1 – Identifiability
2 - Common Support
3 - Θ Open set
4- Differentiable in θ/Interchangeable
5 - Unique MLE
6 - Fisher Info Exists
7 - Boundedness of l″

Limit Results

***Local Linearization ("Delta Method")** – (No RC per se)*

For sequence of r.k-vecs X_n and constant $c \in R^k$, we just need a distributional convergence result, i.e., $a_n(X_n - c) \xrightarrow{d} Y$ and a Borel function $h : R^k \to R^p$ defined in a neighborhood of c and differentiable at c; then $a_n\left(h(X_n) - h(c)\right) \xrightarrow{d} \nabla h(c)^T Y$. Most applications use $a_n = \sqrt{n}$, $X_n = \hat{\theta}$ or $W_n(x)$, and $c = \theta$ or $E(W_n)$, converging to $Z \sim N_p\left(0, \Sigma\right)$, and they often require all partial derivatives to exist. We know these latter 2 restrictions are not necessary. In cases where $\nabla h(c) = 0$, the conditions for higher order expansions providing non-degenerate limiting distributions are similar.

"A Consistent RLE" (Cramér 1946) RC – 1, 2, 3'

$\exists$ ***a*** root (essentially unique) of the LE ∍ $P\left(\hat{\theta}_n \xrightarrow{p} \theta_0\right) = 1$ as $n \to \infty$.

MLE Consistency RC – 0, 1, 2, 3', 4', 5

If $\hat{\theta}_n$ is the root of the LE, then $P\left(\hat{\theta}_n \xrightarrow{p} \theta_0\right) = 1$ as $n \to \infty$. RC 1-5 are sufficient, but not necessary. If the MLE is of closed form, the finite mean version (Khinchin) of the WLLN is usually easier to work with.

MLE Asymptotic Efficiency (Fisher Efficiency) RC – 0, 1, 2, 3, 4''', 5', 6, 7

For $\hat{\theta}_{ML}$ estimating $\theta_0 \in R^k$, $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N_k\left(0,\ I_1^{-1}(\theta_0)\right)$ as $n \to \infty$. Note $V_{\hat{\theta}} = \left(nI(\theta)\right)^{-1}$ and $V_{\hat{\theta}}^{-\frac{1}{2}} = \sqrt{n} \cdot I_1^{\frac{1}{2}}(\theta)$, so we have consistency since

$$\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} N_d\left(0,\ \tfrac{1}{n} I_1^{-1}(\theta)\right) = 0 \text{ as } n \to \infty \Rightarrow \left(\hat{\theta}_n - \theta\right) \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Extended Results – TBD

These are based on extended conditions developed later to handle anomalous results such as inconsistent MLE's and superefficient estimators.

References

1. Dobelman, *Standard Regularity Conditions for Statistics*, Rice University (2011), available http://www.stat.rice.edu/~dobelman/courses/Regularity.pdf