Functions of Estimators

This handout is a technical aside in our course, to fill in one detail about continuous functions of an estimator, and consistency. It will not be tested on exams in this level of generality.

In order to simplify the notation and setting only consider a single real estimator $\hat{\theta}_n$ (not vector valued) and a function g which maps a subset of reals to reals.

Suppose that $\hat{\theta}_n$ is consistent, that is $\hat{\theta}_n \to \theta$ in probability as $n \to \infty$. Now fix a specific value of θ, say θ_0.

Suppose that g is continuous at θ_0. Here the student should recall the ϵ, δ definition of continuity. This means that for a given value of $\epsilon > 0$ there is a value δ (in general $\delta = \delta(\epsilon)$, that is it depends on ϵ) so that

$$\text{if } |\theta - \theta_0| \leq \delta \text{ then } |\mathrm{g}(\theta) - \mathrm{g}(\theta_0)| \leq \epsilon \ .$$

Our goal is to consider the r.v.s $g(\hat{\theta}_n)$.

We have already seen such examples. One was $X_1, \ldots, X_n$ iid exponential parameter λ. Then $\bar{X}_n$ is an unbiased estimator and the sequence of the $\bar{X}_n$ is also a consistent estimator of $\frac{1}{\lambda}$. We then consider the estimator $\hat{\lambda}_n = g(\bar{X}_n)$ where g is the function given by

$$g(x) = \frac{1}{x}$$

and $g : R^+ \mapsto R^+$.

Theorem 1 *Suppose that $\hat{\theta}_n$ is a sequence of consistent estimators for θ, and that g is continuous at θ. Consider the r.v.s $\hat{\psi}_n = g(\hat{\theta}_n)$ as estimators of $\psi = g(\theta)$. Then $\hat{\psi}_n$ converges in probability to ψ.*

Proof :

Now consider the events, for the corresponding ϵ and δ above :

$$A_n = \left\{ |\hat{\theta}_n - \theta_0| \leq \delta \right\}$$

and

$$B_n = \left\{ |g(\hat{\theta}_n) - g(\theta_0)| \leq \epsilon \right\}$$

By the property of g being continuous at θ_0 then if the outcome of the experiment falls into the event A_n, this outcome must also fall into the event B_n. Why? If the outcome of the experiment has the property

$$|\hat{\theta}_n - \theta_0| \leq \delta$$

then it must also have the property that

$$|g(\hat{\theta}_n) - g(\theta_0)| \leq \epsilon \ .$$

Therefore $A_n \subset B_n$, and hence $B_n^c \subset A_n^c$. Therefore

$$\begin{aligned} P(|g(\hat{\theta}_n) - g(\theta_0)| > \epsilon) &= P(B_n^c) \\ &\leq P(A_n^c) \\ &= P(|\hat{\theta}_n - \theta_0| > \delta) \to 0 \end{aligned}$$

as $n \to \infty$.

Since $\epsilon > 0$ is arbitrary, then $P(|g(\hat{\theta}_n) - g(\theta_0)| > \epsilon) \to 0$ as $n \to \infty$ for all $\epsilon > 0$. Therefore $g(\hat{\theta}_n)$ satisfies the definition of convergence in probability to $g(\theta_0)$.

Thus $g(\hat{\theta}_n)$ is a consistent estimator of $g(\theta_0)$.

End of Proof

The middle inequality above is one of the properties we obtained from the Axioms of Probability.

If g is continuous for all parameters θ, then the above argument will hold for every possible θ. In this if $\hat{\theta}_n$ is consistent for θ (no matter what the value of θ happens to be in the parameter space) then $g(\hat{\theta}_n)$ is consistent for $g(\theta)$.