Estimation

December 24, 2013
**Point estimation as a decision problem**

- General decision problem: choose $\delta(Y)$ to minimize

  $$E[\mathcal{L}(\theta, \delta(Y)) \mid Y].$$

- Estimation: We want to observe $Y$, then set $\delta$ as close as possible to $\theta$. I.e.,

  $$\mathcal{L}(\theta, \delta) = M(\rho(\theta, \delta)),$$

  where $\rho$ is a metric (a measure of distance between points) and $M$ is an increasing function.

- Examples of $\rho$: Ordinary Euclidean distance $\sqrt{\sum(a_i^2)}$, uniform $\max(|a_i|)$, absolute deviation $\sum(|a_i|)$.

- There are different versions of estimation, depending on the form of $M(\rho(\cdot))$. 
Estimation with quadratic loss

\[ M(\rho(a)) = a'Wa, \] where \( W \) p.s.d.

Regardless of what \( W \) is, optimal \( \delta \) is \( \delta(Y) = \hat{\theta}(Y) = E[\theta \mid Y] \).

Proof: Suppose \( \delta(Y) = \hat{\delta}(Y) + \gamma(Y) \), with \( \gamma(Y) \) not zero. Then

\[
E[\mathcal{L}(\theta, \delta(Y)) \mid Y] \\
= E[(\theta - \hat{\theta}(Y))'W(\theta - \hat{\theta}(Y)) + 2(\theta - \hat{\theta}(Y))'W\gamma(Y) + \gamma(Y)'W\gamma(Y) \mid Y] \\
= E[(\theta - \hat{\theta}(Y))'W(\theta - \hat{\theta}(Y)) \mid Y] + \gamma(Y)'W\gamma(Y) \geq \mathcal{L}(\theta, \hat{\theta}(Y)).
\]
Estimation with absolute error loss

\[ M(\rho(a)) = \sum w_i |a_i| \]

Regardless of the \(w_i\)’s, it is optimal to choose \(\delta(Y) = \bar{\theta}\) such that \(P[\theta_i < \bar{\theta}_i | Y] = .5\), i.e. \(\bar{\theta}\) is the element-wise median of \(\theta | Y\).

Notice that if \(\theta\) is \(\beta\) in the SNLM with a conjugate prior, \(\hat{\theta} = \bar{\theta}\) and both are OLS applied to the data augmented by dummy observations.
Non-Bayesian approaches to estimation

- Consider candidate $\delta(Y)$’s, check the properties of the distribution of $\delta(Y) \mid \theta$. There are a variety of properties for this distribution that are considered desirable.

- $\delta(Y)$ is an **unbiased** estimator for $\theta$ if and only if $E[\delta(Y) \mid \theta] = \theta$ for every $\theta$.

- While this sounds kind of reasonable, it is important to note that it is quite different from the criterion for an optimal estimator under quadratic loss, which is instead $E[\theta \mid Y] = \delta(Y)]$. Indeed, if $E[\delta(Y)^2] < \infty$, it is impossible that an unbiased estimator coincides with $E[\theta \mid Y]$. 
• Proof: $E[(\delta(Y) - \theta)\theta] = 0$, by definition of conditional expectation and $E[\delta(Y) \mid \theta] = \theta$. But

$$E[(\delta(Y) - \theta)\theta] = E[-(\delta(Y) - \theta)^2] + E[(\delta(Y) - \theta)\delta(Y)]$$

$$= -\text{Var}(\delta(Y) - \theta) < 0, \quad \rightarrow$$

unless of course $\theta \equiv \delta(Y)$ so there is no estimation error at all.

• Note that this result does depend on there being a proper prior, so that the unconditional expectation of $\theta$ is defined.