Estimation

December 24, 2013

Point estimation as a decision problem

- General decision problem: choose $\delta(Y)$ to minimize
$$E[\mathcal{L}(\theta, \delta(Y)) \,|\, Y] \, .$$

- Estimation: We want to observe Y, then set δ as close as possible to θ. I.e.,
$$\mathcal{L}(\theta, \delta) = \mathcal{M}(\rho(\theta, \delta)) \, ,$$
where ρ is a metric (a measure of distance between points) and M is an increasing function.

- Examples of ρ: 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 $\mathcal{M}(\rho(\cdot))$.

Estimation with quadratic loss

- $\mathcal{M}(\rho(a)) = a'Wa$, where W p.s.d.

- Regardless of what W is, optimal δ 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

$$
\begin{aligned}
&E[\mathcal{L}(\theta, \delta(Y)) \mid Y] \\
&= E\big[(\theta - \hat{\theta}(Y))'W(\theta - \hat{\theta}(Y)) + 2(\theta - \hat{\theta}(Y))'W\gamma(Y) + \gamma(Y)'W\gamma(Y) \mid Y\big] \\
&\quad = E\big[(\theta - \hat{\theta}(Y))'W(\theta - \hat{\theta}(Y)) \mid Y] + \gamma(Y)'W\gamma(Y) \geq \mathcal{L}(\theta, \hat{\theta}(Y))\,.
\end{aligned}
$$

Estimation with absolute error loss

- $\mathcal{M}\big(\rho(a)\big) = \sum w_i \,|a_i|$

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

- Notice that if θ is β in the SNLM with a conjugate prior, $\hat{\theta} = \check{\theta}$ 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 θ if and only if $E[\delta(Y) \mid \theta] = \theta$ for every θ.

- 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

$$E[(\delta(Y) - \theta)\theta] = E[-(\delta(Y) - \theta)^2] + E[(\delta(Y) - \theta)\delta(Y)]$$
$$= -\operatorname{Var}(\delta(Y) - \theta) < 0, \rightarrow\leftarrow ,$$

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 θ is defined.