Stat210A: Theoretical Statistics **Lecture Date: October 23, 2014**

Asymptotic Confident Intervals and M-Estimation

Lecturer: Michael I. Jordan *Scribe: Sören Künzel*

1 Asymptotic Confident Intervals

Definition 1. Let δ_0 and δ_1 be statistics. The random interval (δ_0, δ_1) is called $(1-\alpha)$ confident interval for $g(\theta)$ is

$$\mathcal{P}_\theta(g(\theta) \in (\delta_0, \delta_1)) \geq 1-\alpha \text{ for all } \theta \in \Omega$$

Note that for a Frequentist $g(\theta)$ is fix and the interval is the random object.

Definition 2. A variable, which depends on both the data and the paramter, but whose distribution is independent of the parameter is called Pivot.

The following example shows how to use a Pivot to construct confident intervals.

Example 3. Let $X_1, X_2, ... \overset{iid}{\sim} N(\mu, \sigma^2)$. And consider

$$S^2 = \frac{1}{n-1}\sum_i (X_i - \bar{X})^2 \quad \text{and}$$

$$V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Furthermore let $\chi^2_{p,\nu}$ denote the pth quantile of a χ^2_ν random variable. Then it follows that

$$\mathbb{P}_{\sigma^2}(V \geq \chi^2_{\frac{\alpha}{2},n-1}) = \mathbb{P}_{\sigma^2}(V \leq \chi^2_{1-\frac{\alpha}{2},n-1}) = \frac{\alpha}{2}.$$

And hence

$$1-\alpha = \mathbb{P}_{\sigma^2}\left(\chi^2_{1-\frac{\alpha}{2},n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2},n-1}\right) = \mathbb{P}_{\sigma^2}\left(\sigma^2 \in \left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)\right).$$

1.1 Asymptotic Confidence Intervals

Suppose

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N\left(0, 1/I(\theta)\right).$$

Then

$$\sqrt{nI(\theta)}(\hat{\theta}_n - \theta) \Rightarrow N\left(0, 1\right).$$

Let z_p denote the pth quantile of $N(0,1)$. Then

$$\mathbb{P}_\theta(\sqrt{nI(\theta)}|\hat{\theta}_n - \theta| \leq z_{\frac{\alpha}{2}}) \to 1-\alpha.$$

It is often difficult to calculate the Fisher Information. In the following we will discuss strategies to approximate the Fischer Information:

1. We can use $I(\hat{\theta}_n)$ instead: If $I(\theta)$ is continuous, then
$$\sqrt{\frac{I(\hat{\theta}_n)}{I(\theta)}} \xrightarrow{p} 1.$$
Thus using Slutsky's's theorem we can conclude that
$$\sqrt{nI(\hat{\theta}_n)}(\hat{\theta}_n - \theta) = \sqrt{\frac{I(\hat{\theta}_n)}{I(\theta)}}\sqrt{nI(\theta)}(\hat{\theta}_n - \theta) \Rightarrow N(0,1).$$

2. We can also use the results from empirical process theory: Remember that $l(\theta) = \sum_{i=1}^n \log f_\theta(X_i)$. Thus by the law of large numbers
$$\frac{-l''(\hat{\theta}_n)}{n} \xrightarrow{p} I(\theta).$$
And thus again by Slytsky's theorem,
$$\sqrt{-l''(\hat{\theta}_n)}(\hat{\theta}_n - \theta) \Rightarrow N(0,1).$$

3. Another method are **profile regions**: Expand $l_n(\theta)$ in a Taylor series
$$l_n(\theta) = l_n(\hat{\theta}_n) + \frac{1}{2} l_n''(\theta_n^*)(\theta - \hat{\theta}_n)^2.$$
Here θ_n^* is a random variable between θ and $\hat{\theta}_n$. By rearranging this equation we get
$$2l_n(\hat{\theta}_n) - 2l_n(\theta) = \left[\sqrt{-l_n''(\theta_n^*)}(\theta - \hat{\theta}_n)\right]^2 \Rightarrow \chi_1^2$$
Now note that for $Z \sim N(0,1)$
$$\mathbb{P}(Z^2 \leq z_{\alpha/2}^2) = \mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$
And thus
$$\mathbb{P}_\theta(2l_n(\hat{\theta}_n) - 2l_n(\theta) \leq z_{\alpha/2}^2) \to 1 - \alpha$$
This identity can now be used to calculate an asymptotic $(1-\alpha)$ confidence interval for θ.

1.2 Credible regions

Credible regions are a bayesian interpretation of confidence intervals: For a posterior distribution $p(\theta|x)$ a $(1-\alpha)$ an interval (r_0, r_1) is called credible interval if
$$\mathbb{P}(\theta \in (r_0, r_1)|x) \geq 1 - \alpha.$$

2 M-estimation

Definition 4. Let $X_1, X_2, ... \overset{\text{iid}}{\sim} Q$ and $\rho(x)$ be convex with $\rho(x) \to \infty$ if $x \to \pm\infty$. Then we define an M-estimator T_n to be a random variable, which minimizes
$$H(t) = \sum_{i=1}^n \rho(X_i - t).$$

Typical examples for $\rho(x)$ are $-\log f(x)$, x^2 and $|x|$.
Furthermore if ρ' exists and is continuous, we define $\psi = \rho'$ and

$$\bar{w}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \psi(X_i - t).$$

Then we can derive the root of $\bar{w}_n(t)$ to obtain an M-estimator.

Theorem 5. *Define* $\lambda(t) = \mathbb{E}\psi(X - t)$.

- *If* $\lambda(t)$ *is finite for all* $t \in \mathbb{R}$, *and* $\lambda(t)$ *has a unique root* c, *then* $T_n \xrightarrow{p} c$.
- *Moreover weak convergence for* T_n *holds under some regularity conditions:*

$$\sqrt{n}(T_n - c) \Rightarrow N(0, v(\psi, Q)).$$

where

$$v(\psi, Q) = \frac{\mathbb{E}\psi^2(X - c)}{(\lambda'(c))^2}.$$

2.1 Robustness (to outliers)

Let $X \sim Q = (1 - \epsilon)N(\theta, 1) + \epsilon Q^*$. We want to minimize

$$\sup_{\theta \in C_\epsilon} v(\psi, Q).$$

Theorem 6 (Theorem 9.34 in Keener)**.** *There exists* $Q_0 = (1 - \epsilon)N(\theta, 1) + \epsilon Q_0^*$ *and there exists* ψ_0 *s.t.*

$$\sup_{Q \in C_\epsilon} v(\psi_0, Q) = v(\psi_0, Q_0) = \inf_{\psi} v(\psi, Q_0).$$

$\psi_0 = \rho_0'$ *with*

$$\rho_0(t) = \begin{cases} \frac{1}{2}t^2 & |t| \le k \\ k|t| - \frac{1}{2}k^2 & |t| > k \end{cases}$$

and Q_0^* *has support on* $[-k, k]^c$.