Lecture 21 Wald Tests and Related Confidence Sets

Here is a shortcut of large sample LR tests. Suppose for a MLE $\hat{\theta}$ of θ based on $X^n = (X_1, ..., X_n)$ with iid components, we have $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N_k(0, I^{-1}(\theta))$ as $n \to \infty$, where $\theta \in \Theta \subset I\!R^k$ and $I(\theta)$ is the Fisher information (matrix) for X_1. If $I(\hat{\theta}) \xrightarrow{P} I(\theta)$, then $n\ (\hat{\theta} - \theta)\ I(\hat{\theta})\ (\hat{\theta} - \theta)^t \xrightarrow{\mathcal{D}} \chi^2_k$. This suggests the following level α test, called a *Wald test*, for $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$:

Define a test statistic $W_n = W_n(\theta_0) = n\ (\hat{\theta} - \theta_0)\ I(\theta_0)\ (\hat{\theta} - \theta_0)^t$. Reject H_0 if $W_n > c$ with $P(\chi^2_k > c) = \alpha$.

Note:

(a) $I(\theta_0)$ in W_n can be replaced by any consistent estimate of $I(\theta)$, in particular by the *observed Fisher information* $\widetilde{I}(\hat{\theta}) \triangleq \left[\frac{-1}{n}\ \nabla^2 L(\theta|X^n)\right]_{\theta=\hat{\theta}}$. In fact, $\widetilde{I}(\hat{\theta})$ may be preferred because it is usually a byproduct when calculating the MLE $\hat{\theta}$, and the ellipsoid $\mathcal{C} = \left\{\theta \in \Theta :\ n\ (\hat{\theta} - \theta)\ \widetilde{I}(\hat{\theta})\ (\hat{\theta} - \theta)^t \leq c\right\} \subset I\!R^k$ yields an asymptotic level $1 - \alpha$ confidence region for θ.

(b) Note the large sample approximations $I(\theta_0) \approx I(\hat{\theta}) \approx \widetilde{I}(\hat{\theta})$.

(c) See the references, e.g. the book by Bickel and Doksum, for more general versions of Wald tests with a composite H_0.

(d) Wald tests can be extended even further. For illustration, just consider the case with a 1D parameter θ and a CANE T_n, i.e. $a_n(\theta)\ (T_n - \theta) \xrightarrow{\mathcal{D}} N(0, 1)$. Note that usually we have $\frac{a_n(T_n)}{a_n(\theta)} \xrightarrow{P} 1$. Therefore, a Wald test statistic $W_n = a_n(T_n)\ (T_n - \theta_0)$ can be used for testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, or constructing an asymptotic CI for θ.

Example 21.1 Let $X_1, X_2, ...$ be iid samples from a Bernoulli distribution with parameter θ. Denote $\sigma_n = \sigma_n(\theta) = \sqrt{\frac{\theta(1-\theta)}{n}}$. Recall that $\hat{\theta} = \bar{X}$, and $\frac{\hat{\theta}-\theta}{\sigma_n} \xrightarrow{\mathcal{D}} N(0, 1)$. A special feature of this example is $\widetilde{I}(\hat{\theta}) = I(\hat{\theta})$, which can be verified by direct calculation. For testing $H_0 : \theta = \theta_0$, define two statistics:

$$W_n^{(1)} = W_n^{(1)}(\theta_0) = \frac{\hat{\theta} - \theta_0}{\widehat{\sigma}_n}, \quad \text{where } \widehat{\sigma}_n = \sqrt{\frac{\hat{\theta}(1 - \hat{\theta})}{n}};$$

and

$$W_n^{(2)} = W_n^{(2)}(\theta_0) = \frac{\hat{\theta} - \theta_0}{\sigma_n(\theta_0)}.$$

Both $|W_n^{(1)}| > c$ and $|W_n^{(2)}| > c$ give rise to critical regions of level α Wald tests, where $P(N(0,1) > c) = \alpha/2$. However, for a level $1-\alpha$ CI for θ, we need to use $\mathcal{C} = \{\theta \in (0,1) : \ |W_n^{(1)}(\theta)| \le c\} = (\hat{\theta} - c\,\widehat{\sigma}_n,\ \hat{\theta} + c\,\widehat{\sigma}_n)$.

To sum up, a CANE T_n for θ and a consistent estimator for the asymptotic variance of T_n would enable us to construct a Wald test for $H_0 : \theta = \theta_0$ and a CT for θ.