Section 14: Confidence regions

ARE 210

December 5, 2017

The section notes are available on the section Github at github.com/johnloeser/are210 in the "section14" folder.

1 Definitions

- Definitions
    - Let $X = \{X_i\}_{i=1}^n \sim P \in \mathbf{P}$, where $X \in \mathcal{X}$, $\nu(P) \in \mathcal{N}$ parameter of interest
    - $(1-\alpha)100\%$ **confidence region for** $\nu(P)$ is a mapping $S : \mathcal{X} \to 2^{\mathcal{N}}$ such that $P_P[\nu(P) \in S(X)] \geq 1 - \alpha$
    - The **confidence coefficient** of S is $\inf_{P \in \mathbf{P}}[\nu(P) \in S(X)]$
    - The **probability of coverage** of S is $P_P[\nu(P) \in S(X)]$
    - Assume $\nu(P)$ is scalar, then $S(X)$ is a $(1-\alpha)100\%$ **lower (upper) confidence bound** if $P_P[\nu(P) > S(X)] \geq 1 - \alpha$ $(P_P[\nu(P) < S(X)] \geq 1 - \alpha)$.
    - Suppose $L(X)$ is a $(1-\alpha_1)100\%$ lower confidence bound and $U(X)$ is a $(1-\alpha_2)100\%$ upper confidence bound. Then $S(X) = \mathbf{1}\{\nu(P) \in [L(X), U(X)]\}$ is a $(1-\alpha_1-\alpha_2)100\%$ confidence region.
- Duality
    - Let δ_v be a level-α test of the null $H_v : \nu(P) = v$. Define $C_v = \{X : \delta_v(X) = 1\}$ and $A_v = \mathcal{X} \setminus C_v$. Then let $S(X) \equiv \{v \in \mathcal{N} : X \in A_v\}$, then $S(X)$ is a $(1-\alpha)100\%$ confidence region for $\nu(P)$.
    - Suppose $X \sim P_\theta$, where $\theta \in \Theta \subset \mathbf{R}$, and the family is monotone likelihood ratio in $T(X) \sim F_\theta(t)$, where F is continuous in (θ, t). Implicitly define $\theta_L(\alpha, t)$ by $F_{\theta_L(\alpha,t)}(t) = 1 - \alpha$, then $P_\theta[\theta > \theta_L(\alpha, T(X))] \geq 1 - \alpha$.
- Pivots
    - $R(X, \nu(P))$ is a pivot if and only if the distribution of $R(X, \nu(P))$ does not depend on P
    - Let $C(X) = \{v \in \mathcal{N} : R(X, v) \in A\}$, and $P_P[R(X, \nu(P)) \in A] = P[R(X, \nu(P)) \in A] \geq 1 - \alpha$

2 Useful tips

- The duality result is particularly useful for constructing confidence regions for permutation or bootstrap based tests, but generally the intuition is key - to construct a $(1 - \alpha)100\%$ confidence region, just invert a family of level-α tests.
- The most common example of pivots we have is statistics with a known distribution (e.g. $\frac{\overline{X} - \mu}{s^2}$ has a t-distribution when $X_i \sim N(\mu, \sigma^2)$). In the same way that these statistics are useful for constructing tests, they are useful for constructing confidence regions.
- I omit asymptotic definitions, but they follow the patterns for asymptotic definitions for tests.

3 Practice questions

1. Prove the result on the duality of $(1 - \alpha)100\%$ confidence regions and level-α tests.

4 Review

We observe iid data $\{X_i\}_{i=1}^n$, where $X_i \sim P_\theta$, with $\theta \in \Theta$. Represent θ as the solution to a minimum distance problem, $\theta = \arg\min_{b \in \Theta} Q_0(b)$, where $Q_0(b) \equiv g_0(b)^{\mathrm{T}} W g_0(b)$. Examples include m-estimation (where $g_0(b)$ is the expectation of the score), NLLS (where $g_0(b) = \mathbf{E}[(Y_i - \Phi(X_i, \theta)) \frac{\partial \Phi(X_i, \theta)}{\partial \theta}]$), and GMM (where $g_0(b)$ is the population moment condition). We next need to construct finite sample equivalents, $Q_n(b)$ and $g_n(b)$. Frequently, $g_n(b)$ will be a sample average of moment conditions (e.g. score, NLLS FOC, ...). In this cases, it will often turn out that the estimator implied by the optimization problem is $\sqrt{n}$-consistent.

In other cases, we construct our estimator directly. We can start by looking for complete and sufficient statistics for θ. The sample score is typically a complete and sufficient statistic. We can apply the factorization theorem to find sufficient statistics. When P_θ is exponential family, we can find complete and sufficient statistics from factorization. Unbiased estimators we can sometimes find by analyzing population moments (e.g. kth moment of X_i, covariances, CDF, ...). The expectation of an unbiased estimator conditional on a complete and sufficient statistic will be UMVUE. Generally, asymptotically, MLE is as good as we can do. In many cases other estimators

will be asymptotically equivalent to MLE. In special cases, multiple approaches yield the same estimator (e.g. OLS).

Typically, our estimators will be $\sqrt{n}$-consistent, so $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0, V)$. V typically has a nice form for MLE and efficient-GMM, and a less nice sandwich form otherwise. The formula is relatively simple in the NLLS and OLS cases and easy to interpret. For interpretation in particular problems, it's useful to remember here that the quadratic form $X^{\mathrm{T}} V^{-1} X$, where $X \sim N(0, V)$ can roughly be interpreted as summing the squares of the X, normalizing each by its residual variance.

Once we find a finite sample or asymptotic distribution of our estimator, we find a way to construct a pivotal or asymptotically pivotal quantity. For example, often the Wald statistic $n(A(\hat{\theta} - \theta))^{\mathrm{T}}(A^{\mathrm{T}} \hat{V} A)^{-1}(A(\hat{\theta} - \theta)) \xrightarrow{d} \chi_r^2$, where r is the rank of $A^{\mathrm{T}} V A$. We can use this to construct a hypothesis test that rejects when this pivotal quantity is large, i.e. when $\hat{\theta}$ is further from θ than we would "normally" expect to see. This test can be inverted to construct a $(1 - \alpha)100\%$ confidence region for θ. In finite samples, using a t-test or F-test instead of the normal/χ^2 asymptotic approximation often performs better. Bootstrap and permutation based inference are often robust to certain forms of bias in other tests.