Statistics 210A Lecture 28 Notes

Daniel Raban

December 7, 2021

1 Simultaneous Confidence Bounds for Multiple Hypothesis Testing

1.1 Recap: Multiple testing

Last time, we began discussing multiple hypothesis testing, where $X \sim P_\theta \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with hypotheses $H_1, \dots, H_m$ ($H_i : \theta \in \Theta_{0,i}$). The setup includes individual p-values $p_1(X), \dots, p_m(X)$, rejection set $\mathcal{R}(X)$, and true null set $\mathcal{H}_0$.

The classical approach was to control the **familywise error rate (FWER)**,

$$\mathbb{P}_\theta(\text{any false rejections}).$$

The **Bonferroni correction**, a popular procedure, says to reject H_i if $p_i \leq \alpha/m$ and works under arbitrary dependence. We also learned about a direct improvement, the **closure principle**, with an **intersection null** for $S \subseteq \{1, \dots, m\}$. Here, $H_S : H_i$ true for all $i \in S$, which is equivalent to $\theta \in \bigcap_{i \in S} \Theta_{0,i}$. Then we use some **local test** $\phi_S(X)$ which is valid for H_S. The closed testing procedure rejects H_i if $\phi_S(X) = 1$ for all $S \ni i$.

1.2 Simultaneous upper confidence bounds via closed testing

Definition 1.1. Suppose $g_1(\theta), \dots, g_m(\theta)$ are estimands. Then $C_1(X), \dots, C_m(X)$ are **simultaneous confidence bounds** if

$$\mathbb{P}_\theta(\text{any } g_i(\theta) \notin C_i(X)) \leq \alpha.$$

We can use the closed testing procedure to get an upper confidence bound on the number of null indices in S, $|\mathcal{H}_0 \cap S|$.

Example 1.1. Suppose we are looking at an experiment for the brain, and each voxel i, a tiny region of the brain, corresponds to a null hypothesis H_i (about how the voxel behaves in testing vs control). If we look at a region S of the brain, the scientist gives the subset S, the software will give back $U_S(X)$.

Proposition 1.1. *If we take*

$$U_S(X) = \max_{\phi_{S_0}(x)=0} |S \cap S_0|,$$

we get simultaneous confidence bounds.

Proof.

$$\mathbb{P}_\theta(\text{any } U_S(X) \leq |S \cap \mathcal{H}_0(\theta)|) \leq \mathbb{P}_\theta(\phi_{\mathcal{H}_0}(X) = 1)$$

because the first event is a subset of the other. Indeed, if $\phi_{\mathcal{H}_0}(X) = 0$, then $\mathcal{H}_0$ is going to be one of the S_0 sets we take the max over. In this case,

$$U_S(X) = \max_{\phi_{S_0}(x)=0} |S \cap S_0| \geq |S \cap \mathcal{H}_0(\theta)|. \qquad \square$$

We can get simultaneous confidence bounds for the proportion of null indices by looking at $U_S(X)/|S|$. Goeman, Solari, and other coauthors have developed this procedure in a series of papers.

1.3 Simultaneous confidence intervals for the Gaussian sequence model

Example 1.2 (Gaussian sequence model)**.** Suppose have $X \sim N_d(\theta, I_d)$ with $\theta \in \mathbb{R}^d$ and we want simultaneous confidence intervals for $\theta_1, \ldots, \theta_d$. Let c_α be the upper α quantile of $\max_{i=1,\ldots,d} |X_i - \theta_i|$. Then if we take $C_i(X) = (X_i - c_\alpha, X_i + c_\alpha)$, these are simultaneous confidence intervals for θ_i. Why? If any $\theta_i \notin C_i(X)$, then $|X_i - \theta_i| > c_\alpha$; in particular, $\max_{i=1,\ldots,d} |X_i - \theta_i| > c_\alpha$. In this case, we can show that $c_\alpha = z_{\widetilde{\alpha}_d/2}$, where $\widetilde{\alpha} - d$ is the Šidák correction.

What if we want to make pairwise comparisons? We can deduce a confidence interval for $\theta_i - \theta_j$ from the intervals for θ_i, θ_j.

$$|(\theta_i - \theta_j) - (X_i - X_j)| \leq |X_i - \theta_i| + |X_j - \theta_j|,$$

so we could construct a confidence interval with $2c_\alpha$. But this is not very good. Instead, let c'_α be the upper-α quantile of $\max_{i,j} |(X_i - X_j) - (\theta_i - \theta_j)| = \max_{i,j} |Z_i - Z_j|$, where $Z = X - \theta$; this is something we can directly simulate. Then, let

$$C_{i,j}(X) = (X_i - X_j - c'_\alpha, X_i - X_j + c'_\alpha).$$

This is called **Tukey's Honestly Significant Difference procedure (HSD)**.

More generally, we may want simultaneous confidence intervals $c_\lambda(X)$ for $\lambda^\top \theta$, there $\lambda \in \mathbb{S}^{d-1}$. Let

$$\begin{aligned} c'_\alpha &= \text{upper-}\alpha \text{ quantile of } \sup_{\lambda \in \mathbb{S}^{d-1}} |\lambda^\top (X - \theta)| \\ &= \text{upper-}\alpha \text{ quantile of } \|X - \theta\|_2 \\ &\sim \chi_d(\alpha). \end{aligned}$$

1.4 Simultaneous confidence intervals in linear regression

Example 1.3 (Linear regression)**.** Suppose we have $\frac{\widehat{\beta}-\beta}{\sigma} \sim N_d(0, (X^\top X)^{-1})$ with $\beta \in \mathbb{R}^d$ and $\frac{\widehat{\sigma}^2}{\sigma^2} \sim \frac{1}{n-d}\chi^2_{n-d}$, and suppose we want simultaneous confidence intervals for $\beta_1, \dots, \beta_d$. Let c_α be the upper-α quantile of $\max_{j=1,\dots,d} |\widehat{\beta}_j - \beta_j|/\widehat{\sigma}$. We can directly simulate

$$\frac{\widehat{\eta} - \beta}{\widehat{\sigma}} = \frac{N_d(0, (X^\top X)^{-1})}{\sqrt{\frac{1}{n-d}\chi^2_{n-d}}}.$$

This has what is known as a **multivariate** t **distribution**. If we want simultaneous confidence intervals for β_i, then we can use

$$C_j = (\widehat{\beta} - j - \widehat{\sigma} c_\alpha, \widehat{\beta} - j + \widehat{\sigma} c_\alpha).$$

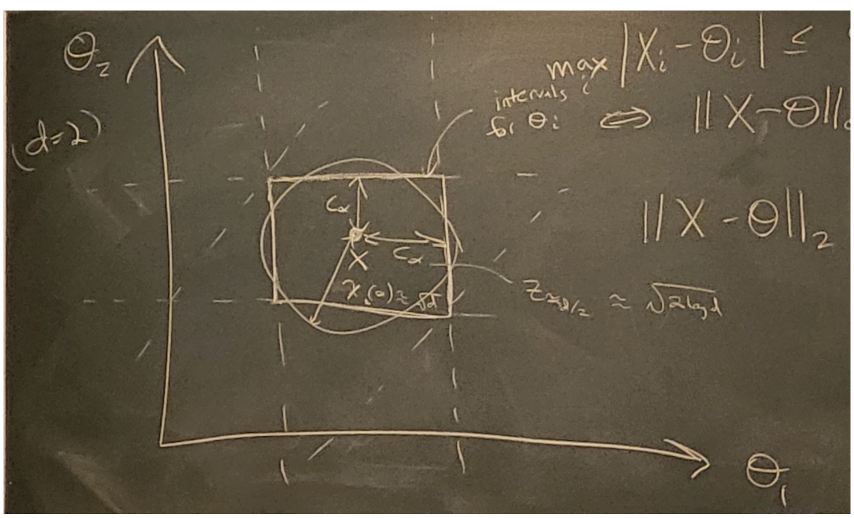
If any $\beta_j \notin C_j$, then, as before,

$$\max_{i=1,\dots,d} |\widehat{\beta}_i - \beta_i| > c_\alpha \widehat{\sigma}.$$

We can do the same procedure with Tukey's HSD, where we let $c'_\alpha = \max_{i,j} |Z_i - Z_j|$ with $Z = (\widehat{\beta} - \beta)/\widehat{\sigma}$ and use the intervals

$$C_{i,j}(X) = (X_i - X_j - c'_\alpha, X_i - X_j + c_\alpha).$$

Observe that $\max_i |X_i - \theta_i| \le c_\alpha \iff \|X_i - \theta\|_\alpha \le C_\alpha$. Alternatively, we could try to control $\|X - \theta\|_2 \le \chi_d(\alpha)$.



Our method involves constructing this rectangle and projecting it onto each of the axes. The naive method of estimating $\theta_i - \theta_j$ from before is projecting in the direction of $\theta_i - \theta_j$; so the projection we use may make a difference.

Example 1.4. Consider testing the global null $H_0 : \theta = 0$. The max test rejects if $\max_| X_i| > c_\alpha \approx \sqrt{2 \log d}$, and the χ^2 test rejects if $\|X\|_2^2 \geq \chi_d^2(\alpha) \approx d + 3\sqrt{d}$. If θ is 1-sparse (only $\theta_1 \neq 0$), then the max test needs $|\theta_1| > \sqrt{2 \log d}$, whereas the χ^2 test needs $|\theta_1 = \|\theta\|_2 \approx d^{1/4}$. If θ is dense, the χ^2 test is vastly more powerful, but if θ is sparse, then the max test is vastly more powerful.

Next time, we will discuss controlling what is known as the false discovery rate.