1 Outline

Agenda: Selective Inference

1. Confidence Intervals for Selected Parameters
2. False Coverage Rate
3. FCR-Adjusted Confidence Intervals

So far, we have been studying multiple hypothesis testing. Today we will look at the other inference problem, namely, multiple confidence intervals and coverage after selection.

2 Multiple confidence intervals

Consider a classical setting in which we have $n$ parameters $\theta_1, \theta_2, \ldots, \theta_n$, and corresponding statistics $T_1, T_2, \ldots, T_n$. Throughout this lecture, we can keep the following simple example in mind:

$T_i \sim \mathcal{N}(\theta_i, 1), i = \{1, 2, \ldots, n\}$.

In this setting, the notions of marginal and simultaneous coverage are familiar to us:

- Marginal coverage says that $\mathbb{P}(\theta_i \in \text{CI}_i(\alpha)) \geq 1 - \alpha$.
- Simultaneous coverage says that $\mathbb{P}((\theta_1, \theta_2, \ldots, \theta_n) \in \text{CI}(\alpha)) \geq 1 - \alpha$.

In our running example where $T_i \sim \mathcal{N}(\theta_i, 1)$, we can construct marginal confidence intervals via

$$\text{CI}_i(\alpha) = [T_i - z_{1-\alpha/2}, T_i + z_{1-\alpha/2}] .$$

We can also achieve simultaneous coverage via the Bonferroni confidence region

$$\text{CI}(\alpha) = [T_1 - z_{1-\alpha/2n}, T_1 + z_{1-\alpha/2n}] \times \ldots \times [T_n - z_{1-\alpha/2n}, T_n + z_{1-\alpha/2n}] .$$

3 Coverage after selection

We now discuss selection. Oftentimes researchers examine many parameters at once and report confidence for selected ones. However, as noted by Sorić in 1989:
In a large number of 95% confidence intervals, 95% of them contain the population parameter [...] but it would be wrong to imagine that the same rule also applies to a large number of 95% interesting confidence intervals.

We present an example illustrating Sorić’s comment:

- Select $\theta_i \overset{iid}{\sim} N(0, 0.04)$, $i = \{1, 2, \ldots, 20\}$. From now on, they are fixed.
- Assume $T_i \overset{ind}{\sim} N(\theta_i, 1)$, $i = \{1, 2, \ldots, 20\}$
- Construct level $\alpha = 0.1$ marginal confidence intervals $CI_i = [T_i - 1.64, T_i + 1.64]$

A particular instance of the simulations is shown in Figure 1. Out of the 20 constructed intervals, 17 cover their respective parameter pretty much as we would expect. However, consider now selecting those ‘interesting’ parameters whose CIs do not cover 0. In this example, we see that the coverage proportion for selected parameters is 1 out of 4. In 100 simulations, the conditional coverage—the number of times a selected parameter is covered by the marginal CI divided by the number of times that parameter is selected is 0.043.

Figure 1: Confidence intervals and selected parameters. CIs not away from zero are selected (indicated in yellow). Red circles indicate selected parameters not covered by corresponding marginal confidence intervals. We see that the coverage for the selected is 1 out of 4. For all the parameters, it is 17 out of 20.

We see that the marginal confidence intervals may have serious reduced coverage probability after selection. One worthy goal would be to achieve conditional coverage; that is,

$$P(\theta_i \in CI_i(\alpha)|i \in S) \geq 1 - \alpha,$$

where $S$ is the selected set of parameters. However, conditional coverage following any selection rule cannot, in general, be achieved.
Suppose \( \theta_i = 0 \). No matter how we construct a CI for \( i \), selecting \( i \) iff \( \text{CI}_i \not\ni 0 \) results in \( P(\theta_i \in \text{CI}_i | i \in S) = 0 \).

The reason why conditional coverage cannot be achieved is similar to why pFDR = \( \mathbb{E}(\text{FDP} | R > 0) \) cannot be controlled; e.g. under global null, conditional on making a rejection, the pFDR is 1.

4 False coverage rate [Benjamini-Yekutieli, 2005]

In 2005, Benjamini and Yekutieli introduced a relaxation of conditional coverage named false coverage rate (FCR), in a similar sense that FDR is a relaxation to pFDR.

Definition 1. The False coverage rate is defined as

\[
\text{FCR} = \mathbb{E} \left[ \frac{V_{\text{CI}}}{R_{\text{CI}} \lor 1} \right],
\]

where \( R_{\text{CI}} \) is the number of selected parameter and \( V_{\text{CI}} \) the number of constructed confidence intervals not covering.

Some properties of the FCR:

- It is similar to FDR in that it controls type I error over the selected parameters.
- Without selection, i.e. \( |S| = n \), the marginal CI's control the FCR since

\[
\text{FCR} = \mathbb{E} \left[ \sum_{i=1}^{n} \frac{1}{n} \{ \theta_i \not\in \text{CI}_i(\alpha) \} \right] \leq \alpha.
\]

- With selection, marginal CI's will not generally control the FCR. To see this, set \( \alpha = 0.05 \), \( n = 1000 \), \( T_i \overset{\text{ind}}{\sim} N(\theta_i, 1) \) and \( \theta_i = 0, \forall i \). Let the selection rule be \( i \in S \iff |T_i| \geq 2 \), then \( P(|S| \geq 1) \approx 1 \),

\[
\text{FCR} \approx \mathbb{E}\left[\text{FCP} | |S| > 0\right] + \mathbb{E}\left[\text{FCP} | |S| = 0\right] P(|S| = 0)
\]

\[
\approx \mathbb{E}\left[\text{FCP} | |S| > 0\right] P(|S| > 0)
\]

\[
\approx 1.
\]

The last \( \approx \) holds because if we select \( i \) then \( |T_i| \geq 2 \), since the half width of the confidence interval is 1.96, \([T_i - z_{1-\alpha/2}, T_i + z_{1-\alpha/2}]\) cannot cover the true parameter \( \theta_i = 0 \).

- Bonferroni’s CI’s do control FCR in the same way that Bonferroni’s procedure controls the FDR.

5 FCR adjusted CIs

We now discuss an FCR controlling procedure.
(i) Apply selection rule \( \mathcal{S}(T) \)

(ii) For each \( i \in (S) \), find \( R_{\min}(T^{(i)}) = \min_{t} \{|\mathcal{S}(T^{(i)}, t)\} : i \in \mathcal{S}(T^{(i)}, t)\} \), where \( T^{(i)} = T \setminus \{T_i\} \).

(iii) FCR adjusted CI for \( i \in S \) is \( \text{Cl}_i(R_{\min}(T^{(i)}) \alpha/n) \)

Usually \(|\mathcal{S}(T^{(i)}, t)|\) such that \( i \in \mathcal{S}(T^{(i)}, t) \) is just \(|\mathcal{S}(T)|\) so that in this case \( R_{\min}(T^{(i)}) = R_{\text{Cl}} \) and adjusted CI is \( \text{Cl}_i(R_{\text{Cl}} \alpha/n) \). In our previous example, \( R = 4, n = 20 \) and \( \alpha = 0.1 \), so the adjusted CI's are \( \text{Cl}_i(0.1 \times \frac{1}{2}) \). So we see that they are wider than marginal CI's.

Some special cases:

- \( R_{\text{Cl}} = n \), no adjustment
- \( R_{\text{Cl}} = 1 \), Bonferroni adjustment

**Theorem 2.** If \( T_i \)'s are independent, then for any selection procedure, the FCR of adjusted CI's obey \( \text{FCR} \leq \alpha \). (There also are extensions to PRDS statistics).

A different way to read this theorem is that BH(q) can actually detect effect direction through confidence intervals. It offers directional FDR control.

**Proof.** Recall that

\[
\text{FCR} = \mathbb{E} \sum_{i=1}^{n} X_i, \quad X_i = \frac{1\{i \in S, \theta_i \notin \text{CI}_i(\alpha R_{(i)}/n)\}}{|S|},
\]

where \( R_{(i)} \) is a shorthand for \( R_{\min}(T^{(i)}) \). It suffices to show that \( \mathbb{E}X_i \leq \alpha/n \) so that \( \text{FCR} \leq \alpha \).

We have

\[
X_i = \sum_{k=1}^{n} \frac{1\{i \in S, \theta_i \notin \text{CI}_i(\alpha k/n), R_{(i)} = k\}}{|S|} \leq \sum_{k=1}^{n} \frac{1\{i \in S, \theta_i \notin \text{CI}_i(\alpha k/n), R_{(i)} = k\}}{k} \leq \sum_{k=1}^{n} \frac{1\{\theta_i \notin \text{CI}_i(\alpha k/n), R_{(i)} = k\}}{k}
\]

because \(|S| \leq k\), by definition. Now

\[
\mathbb{E}[X_i|T^{(i)}] \leq \sum_{k=1}^{n} \frac{1\{R_{(i)} = k\}}{k} \mathbb{P}(\theta_i \notin \text{CI}_i(\alpha k/n)) \leq \sum_{k=1}^{n} \frac{1\{R_{(i)} = k\}}{k} \times \frac{\alpha k}{n} = \frac{\alpha n}{n} \sum_{k=1}^{n} 1\{R_{(i)} = k\}
\]

\[
= \frac{\alpha}{n}
\]

4
because $R(i)$ takes values in $1,\ldots,n$. Naturally,

$$
\mathbb{E}[X_i] = \mathbb{E}[\mathbb{E}[X_i | T^{(i)}]] \leq \alpha/n.
$$

6 Selection via multiple testing

Now we investigate whether the above procedure is overly conservative and study a particular example:

- We have $n$ hypotheses $H_i : \theta_i = \theta_i^0, \ i = 1,2,\ldots,n$
- $T_i - \theta_i \sim F_i$, where $F_i$’s are known and symmetric
- We compute p-values for two sided test
- Select with BH($\alpha$)

The marginal confidence intervals are $CI_i(\gamma) = \{\theta_i : |T_i - \theta_i| \leq F_i^{-1}(1 - \gamma/2)\}$

**Theorem 3.** *Suppose $R_{\min}(T^{(i)}) = R_{CI}$ a.s. and the $T_i$’s are independent. Then FCR $\geq \alpha/2$.*

This shows that the FCR controlling procedure is not overly conservative.