STAT 689: Large-scale and high-dimensional statistical inference **Fall 2024**

Lecture 13

1 Selective inference: constructing confidence intervals

In the past, scientists usually followed the steps below:

- Step 1. Select hypotheses/model/question
- Step 2. Collect data
- Step 3. Perform statistical inference

In modern practice, steps 1 and 2 are reversed. Data is collected before any hypotheses are formulated, and then it is combed through for anything that looks interesting. We need statistical tools that are suited to this new paradigm. In this lecture, we consider the situation where we use our data to select some parameters and then form confidence intervals for the selected parameters. Just as in multiple testing, we must adjust our intervals or else inference will be distorted.

1.1 An example

We begin with an illustrative example. Suppose we have the model

$$Y_i = \beta_0 X_{i,0} + \sum_{j=1}^{10} \beta_j X_{i,j} + \epsilon_i$$

where $\epsilon_i \sim N(0,1)$ and $i = 1, 2, \ldots, 250$. We are interested in a confidence interval for β_0. First, we perform model selection using the BIC criterion, always including X_0. We then use the t-statistic from the selected model to form a 95% confidence interval. In the figure below, we show the nominal distribution of this statistic under the null $\beta_0 = 0$ and the actual distribution (found through simulation) following the selection step.

Due to the discrepancy between these two distributions, the coverage rate of the constructed interval will be 83.5% rather than the desired 95%. The situation only gets worse as the number of predictors increases. For $p = 30$, the coverage rate can become as low as 39%.

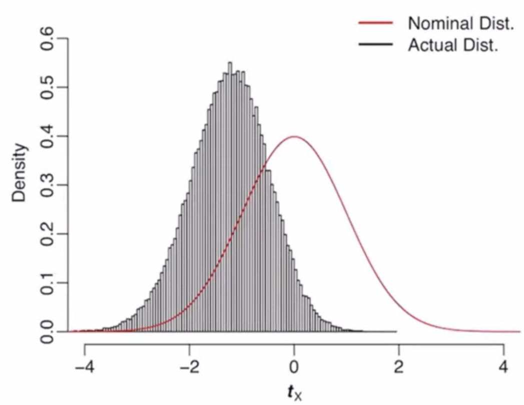

Figure 1: Distribution of the t-statistic, nominal versus actual.

1.2 Soric's Warning

Soric warned about the practice of reporting confidence intervals following some selection procedure in 1989, saying:

> "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 Soric's comment.

- Draw $\theta_i \overset{\text{i.i.d}}{\sim} N(0, 0.04)$ for $i = 1, \ldots, 20$.
- Sample $Z_i \sim N(\theta_i, 1)$ independently.
- Construct 90% marginal confidence intervals for θ_i as $\text{CI}_i = [Z_i - 1.64, Z_i + 1.64]$

A particular instance of this simulation is shown in Figure 2. We see that out of the 20 constructed intervals, about 17 contain the parameter θ_i. However, consider now selecting those "interesting" parameters whose intervals do not contain zero. If we focus on these parameters, we see that out of the four intervals that do not contain zero, only 1 covers the true parameter, far from our expected coverage of 90%.

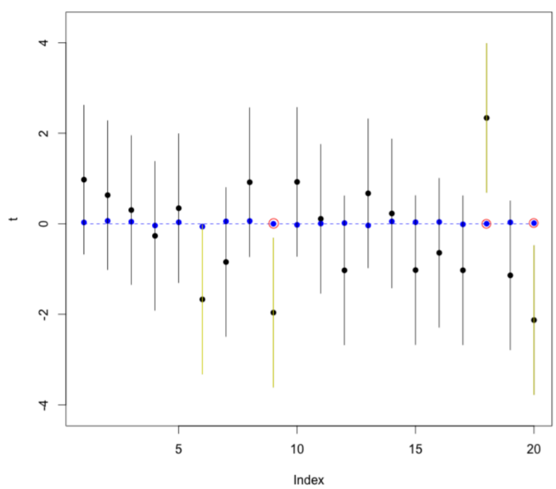

Figure 2: CIs away from zero are selected (indicated in yellow). Red circles indicate selected parameters not covered by corresponding marginal confidence intervals.

Through simulation, we find that if $S \subset \{1, \ldots, 20\}$ is the set of parameters whose intervals do not contain zero:

$$P_\theta(\theta_i \in \text{CI}_i | i \in S) \approx 0.043$$

We see that the marginal confidence intervals may have seriously reduced coverage probability after selection.

1.3 Selective Inference Criteria

How should we address this problem? We might try to develop methods that achieve some different notions of coverage, such as those given below:

1. On average over the selected (FCR);
2. Conditional over the selected.

2 Conditional Coverage

Letting S denote the set of selected parameters and $\text{CI}_i(\alpha)$ be the $1-\alpha$ CI, we achieve $1-\alpha$ conditional coverage if:

$$P(\theta_i \in \text{CI}_i(\alpha) | i \in S) \geq 1 - \alpha.$$

In this section, we show that, in general, conditional coverage cannot be achieved. Consider the following situation:

$$Y_i \overset{\text{i.i.d}}{\sim} N(\mu, 1), i = 1, \ldots, 200.$$

As before, given each Y_i, we construct a 95% confidence interval for μ as:

$$\text{CI}_i = [Y_i - 1.96, Y_i + 1.96]$$

Suppose we select those i whose interval does not contain 0, i.e.,

$$S = \{i : 1 \leq i \leq 20, 0 \notin \text{CI}_i\}$$

We plot the conditional coverage $P(\mu \in \text{CI}_i | i \in S)$ as a function of μ in Figure 3. When μ is very small, the conditional coverage is essentially zero. This is expected because when μ is zero, we essentially exclude all intervals that contain μ, attaining zero coverage! When μ is very large, no selection occurs (i.e., $S = \{1, ..., 200\}$), so we recover the marginal coverage rate of 0.95.

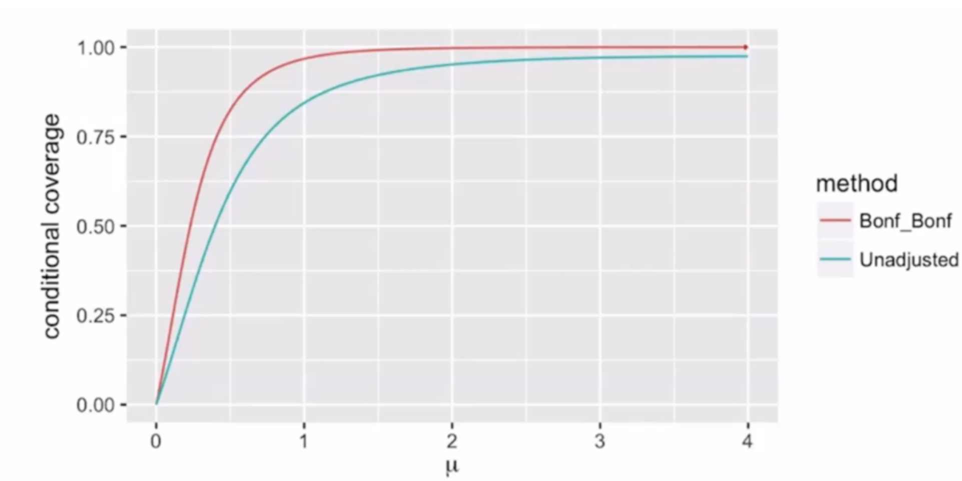

Figure 3: Conditional Coverage achieved by Unadjusted Intervals and Bonferroni Correction.

To address this situation, we might widen the interval to take into account the effect of multiplicity. Suppose we apply the Bonferroni correction so that our confidence intervals take the form

$$\text{CI}_i = \left[Y_i - c\left(1 - \frac{\alpha}{2n}\right), Y_i + c\left(1 - \frac{\alpha}{2n}\right) \right],$$

where $c(1-q)$ is the $1-q$ quantile of $N(0, 1)$. In addition, we use Bonferroni selection so that $S = \{i : p_i \leq \alpha/n\}$ where p_i is the p-value associated with point i. When μ is small, we see that due to our selection rule, we still do not achieve conditional coverage. We conclude that although conditional coverage is highly desirable, it cannot be achieved in general.

3 False Coverage Rate

Benjamini and Yekutieli (2005) introduced the notion of the False Coverage Rate (FCR).

Definition. The False Coverage Rate (FCR) is defined as:

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

where R_{CI} is the number of selected parameters and V_{CI} is the number of confidence intervals among those selected that do not cover.

Here are some properties of the FCR:

- Controlling the FCR is similar to controlling the FDR in testing: it controls the average type I error over the selected.
- Without selection (i.e., $|S| = n$), the marginal CIs control the FCR since:

$$\text{FCR} = \mathbb{E}\left[\frac{\sum_{i=1}^{n} \mathbf{I}\{\theta_i \notin \text{CI}_i(1-\alpha)\}}{n}\right] \leq \alpha.$$

- With selection, the marginal CIs are not guaranteed to control the FCR. We cover this in more detail below.
- In the same way that Bonferroni's procedure controls the FDR, applying the Bonferroni correction to the marginal CIs will guarantee that the FCR is controlled:

$$\begin{aligned}\text{FCR} &= E\left[\frac{V_{CI}}{R_{CI} \vee 1}\right] \\ &\leq P(V_{CI} \geq 1) \\ &\leq P(\theta_i \notin \text{CI}_i(1-\alpha/n) \text{ for some } i) \\ &\leq \sum_{i=1}^{n} P(\theta_i \notin \text{CI}_i(1-\alpha/n)) \\ &\leq \alpha.\end{aligned}$$

- Any confidence region $\text{CI}(\alpha)$ achieving simultaneous coverage

$$P((\theta_1, ..., \theta_n) \in \text{CI}(1-\alpha)) \geq 1-\alpha$$

  controls the FCR because

$$\begin{aligned}\text{FCR} &= E\left[\frac{V_{CI}}{R_{CI} \vee 1}\right] \\ &\leq P(V_{CI} \geq 1) \\ &\leq P(\theta_i \notin \text{CI}(1-\alpha) \text{ for some } i) \\ &\leq \alpha.\end{aligned}$$

To illustrate that marginal CIs can fail to control the FCR, let us return to the example from Section 2. We plot the FCR for the marginal, unadjusted intervals along with the Bonferroni selected intervals in Figure 4.

We see that the unadjusted intervals completely fail to control the FCR when μ is small. In fact, for μ near zero, we have an FCR of 1. In contrast, using the Bonferroni corrected intervals controls the FCR. However, these intervals are extremely wide.

4 FCR Adjusted Confidence Intervals

We now describe a procedure less conservative than applying Bonferroni's correction, which is guaranteed to control the FCR. We assume that associated to our parameters $\theta_1, \ldots, \theta_n$ are test statistics $T_1, \ldots, T_n$, summarized as $T = \{T_1, \ldots, T_n\}$.

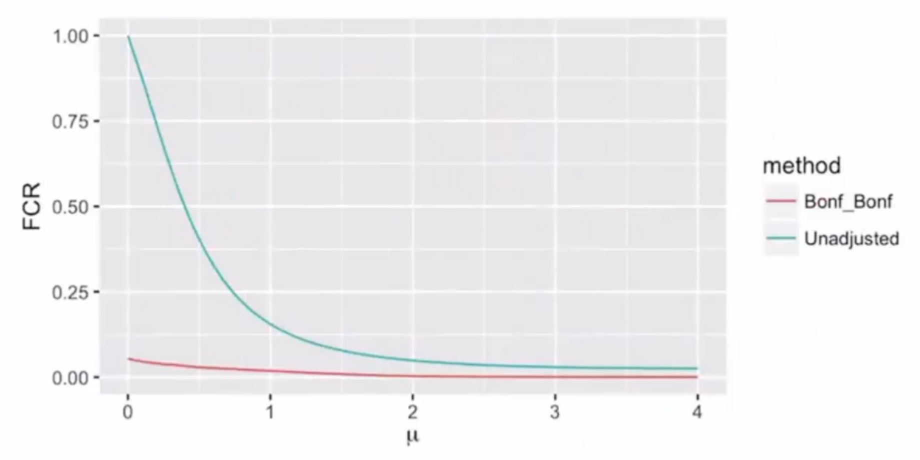

Figure 4: False Coverage Rate achieved by unadjusted intervals and Bonferroni selected and subsequently adjusted intervals.

1. Apply any selection rule S to the statistics T to obtain the selection set $S = S(T)$.
2. For each $i \in S$, compute:
$$R^{(i)} = \min_t \{|S(T^{(i)}, t)| : i \in S(T^{(i)}, t)\}$$
where $T^{(i)} = T \backslash \{T_i\}$. In other words, we construct new test statistics vectors T_t by replacing the observed value of the ith entry of T with $t \in \mathbb{R}$. Then, we consider a set (denoted by $A_i \subseteq \{T_t\}_t$) of all such constructed vectors $\{T_t\}_t$ that results in the selection of the ith variable. Finally, we find the minimum number of selected variables by applying the selection procedure S to the constructed test statistics in A_i. Hence $R^{(i)}$ is the smallest number of selections if we keep all $[n]\backslash i$ test statistics fixed at observed values and vary the ith test statistic in such a way that still results in the selection of the ith variable. Typically, $R^{(i)}$ is the same as the number of selections.
3. The FCR adjusted CI for $i \in S$ is
$$\mathrm{CI}_i \left(1 - \frac{R^{(i)}\alpha}{n}\right).$$

Step 2 may appear complex, but it is often the case that $R^{(i)} = R = |S(T)|$ for reasonable selection rules S.

To illustrate the behavior of this procedure, consider two extreme cases.

- If $R_{CI} = n$, then we make no adjustment and simply use the marginal $\mathrm{CI}_i(\alpha)$ intervals.
- If $R_{CI} = 1$, then we make the Bonferroni adjustment to the one confidence interval selected.

Theorem. If the statistics T_i are independent, then the FCR of the adjusted confidence intervals of the above procedure is less or equal to α.

Proof. Recall that:
$$\mathrm{FCR} = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right], \quad X_i = \frac{\mathbf{I}\left\{i \in S, \theta_i \notin \mathrm{CI}_i(1 - R^{(i)}\alpha/n)\right\}}{|S| \vee 1}.$$

It suffices to show that $\mathbb{E}[X_i] \leq \alpha/n$, so that FCR $\leq \alpha$. We have

$$\begin{aligned} X_i &= \sum_{k=1}^{n} \frac{\mathbf{I}\left\{i \in S, \theta_i \notin \mathrm{CI}_i(1 - R^{(i)}\alpha/n), R^{(i)} = k\right\}}{|S|} \\ &\leq \sum_{k=1}^{n} \frac{\mathbf{I}\left\{i \in S, \theta_i \notin \mathrm{CI}_i(1 - k\alpha/n), R^{(i)} = k\right\}}{k} \\ &\leq \sum_{k=1}^{n} \frac{\mathbf{I}\left\{\theta_i \notin \mathrm{CI}_i(1 - k\alpha/n), R^{(i)} = k\right\}}{k} \end{aligned}$$

because $|S| \geq k$ by definition. Now note that

$$\begin{aligned} \mathbb{E}[X_i | T^{(i)}] &\leq \sum_{k=1}^{n} \frac{\mathbf{I}\left\{R^{(i)} = k\right\}}{k} \mathbb{P}(\theta_i \notin \mathrm{CI}_i(1 - k\alpha/n)) \\ &\leq \sum_{k=1}^{n} \frac{\mathbf{I}\left\{R^{(i)} = k\right\}}{k} \frac{k\alpha}{n} \\ &= \frac{\alpha}{n} \sum_{k=1}^{n} \mathbf{I}\left\{R^{(i)} = k\right\} \\ &= \frac{\alpha}{n} \end{aligned}$$

because $R^{(i)}$ takes values in $\{1, ..., n\}$. Naturally

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

Figure 5 demonstrates how the procedure performs empirically on the problem of Section 2. As $\mu \to \infty$, there is no selection, and so $R = n$. Since the marginal coverage is $1 - \alpha = 0.95$, this means the FCR converges to α as $\mu \to \infty$. When $\mu = 0$, any selection will not cover the true parameter. However, with probability $1 - \alpha$, no selection will be made and hence the FCR is α.

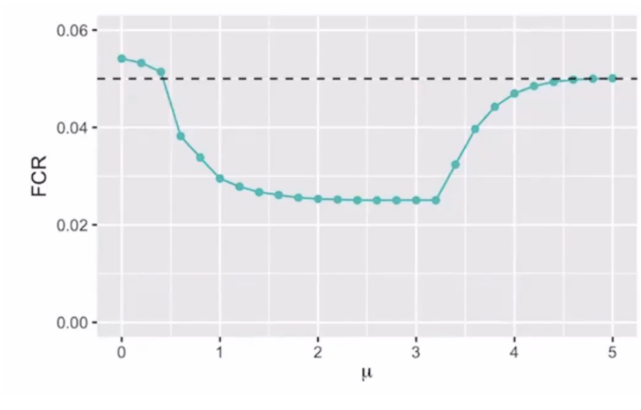

Figure 5: False Coverage Rate achieved by proposed procedure.

4.1 An issue

There are, however, some issues with the FCR controlling procedure. Consider the following setting: $n = 10,000$, and

$$\mu_i = \begin{cases} 0 & 1 \leq i \leq 9,000, \\ \overset{iid}{\sim} N(3,1) & 9,001 \leq i \leq 10,000, \end{cases}$$

$$z_i \overset{ind}{\sim} N(\mu_i, 1).$$

Consider performing selection using the one-sided Bejamini-Hochberg BH(q) procedure at level q. Using the FCR controlling procedure at level $\alpha = 0.05$, the realized FCR is $18/610 \approx 0.03 < 0.05$. The plot below shows the FCR-adjusted 95% CIs.

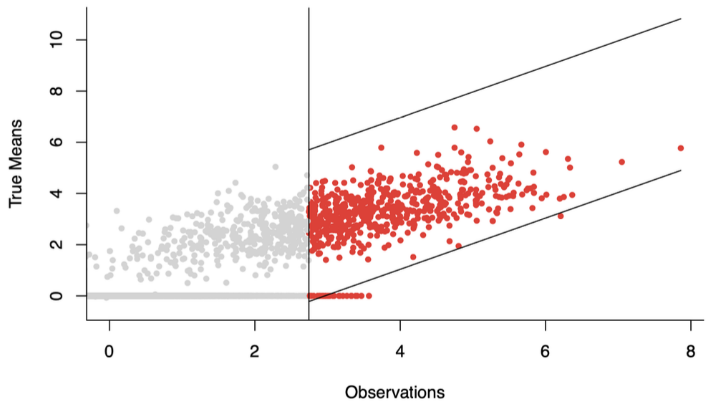

Figure 6: FCR-adjusted confidence intervals.

Observing the slope of the confidence intervals does not seem right. Intuitively, the FCR-adjusted CIs should extend downwards due to the selection bias, yet the FCR procedure produced CIs that are too wide upwards, failing to adequately capture the regression effect. Consequently, Daniel Yekutieli proposed the eBayes procedure. The plot below shows the CIs produced by the eBayes procedure.

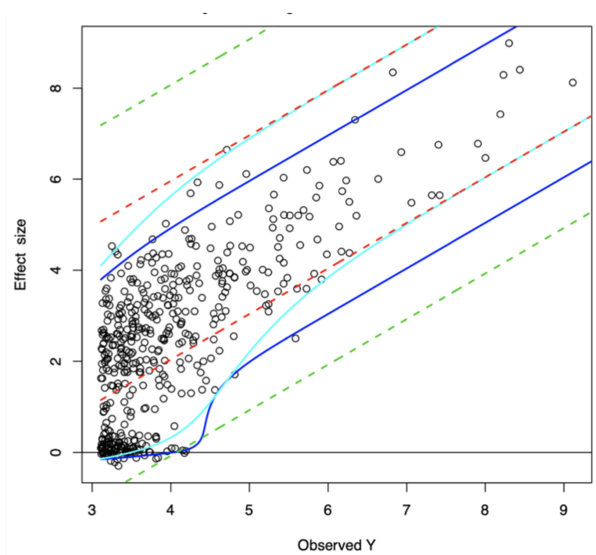

Figure 7: eBayes-adjusted confidence intervals.