STATS 300C: Theory of Statistics	**Spring 2021**
Lecture 12 — May 5, 2022	
Lecturer: Prof. Emmanuel Candès	*Editor: Parth Nobel, Scribe: Sophia Lu*

Warning: *These notes may contain factual and/or typographic errors. They are based on Emmanuel Candès's course from 2018 and 2022, and scribe notes written by Emmanuel Candès, Min Liu, Karthik Rajkumar, and Matthew MacKay.*

12.1 Outline

Agenda: Selective Inference

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

12.2 Confidence Intervals for Selected Parameters

As we have previously discussed, the classical scientific method no longer applies to modern scientific practice. In the past, scientists usually followed the steps below:

1. Select hypotheses/model/question
2. Collect data
3. Perform 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, then form confidence intervals for the selected parameters. Just as in multiple testing, we must adjust our intervals or else inference will be distorted.

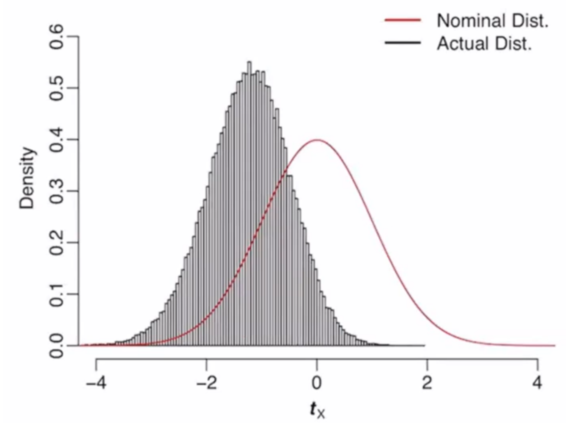
12.2.1 Buja's Example

Here is an illustrative example from A. Buja. Suppose we have the following situation:

$$Y = \beta_0 X_0 + \sum_{j=1}^{p=10} \beta_j X_j + \epsilon \qquad n = 250, \qquad \epsilon \sim \mathcal{N}(0,1)$$

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 Figure 12.1, we show the nominal distribution of this statistic under the null $\beta_0 = 0$ as well as the actual distribution (found through simulation) following the selection step.

Figure 12.1. Distribution of the t-statistic, nominal vs actual



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%.

12.2.2 Soric's Warning

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

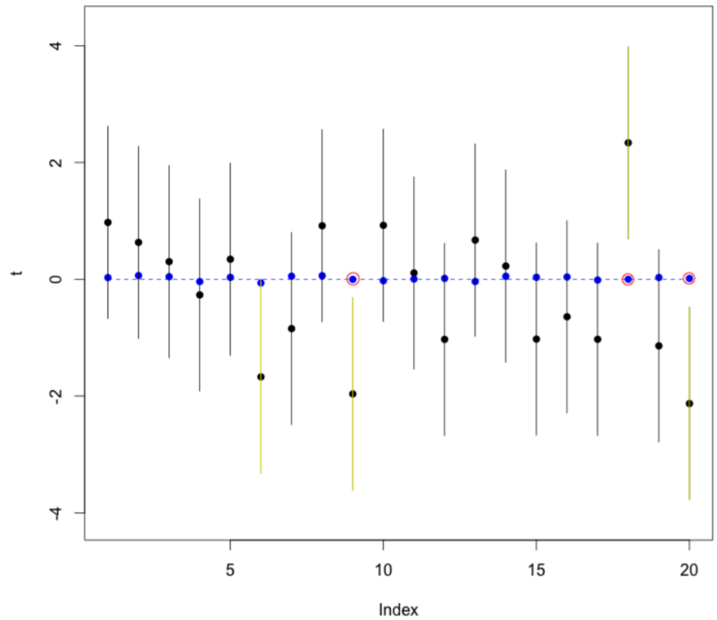
> "In a large number of 95% confidence intervals, 95% of them contain the population parameter [...] but it would be wrong to imagine that the same rule also applies to a large number of 95% interesting confidence intervals."

We present an example illustrating Soric's comment.

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

A particular instance of this simulation is shown in Figure 12.2. We see that out of the 20 constructed intervals, about 17 actually contain the parameter θ_i. However, consider now selecting those "interesting" parameters whose intervals do not contain zero. If we focus our attention on these parameters, we see that out of the 4 intervals which do not contain zero only 1 covers the true parameter, far from our expected coverage of 90%.

Figure 12.2. 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.



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

$$\mathbb{P}_\theta(\theta_i \in \mathrm{CI}_i | i \in S) \approx 0.043$$

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

12.2.3 Selective Inference Criteria

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

(A) Simultaneous over all selection rules

(B) Simultaneous over the selected

(C) On average over the selected (FCR)

(D) Conditional over the selected

In this lecture, we shall focus on the criteria (C) and (D).

12.3 Conditional Coverage

Letting S denote the set of selected parameters, we achieve $1 - \alpha$ *conditional coverage* if:

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

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

$$Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1), \quad i = 1, ..., 200$$

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

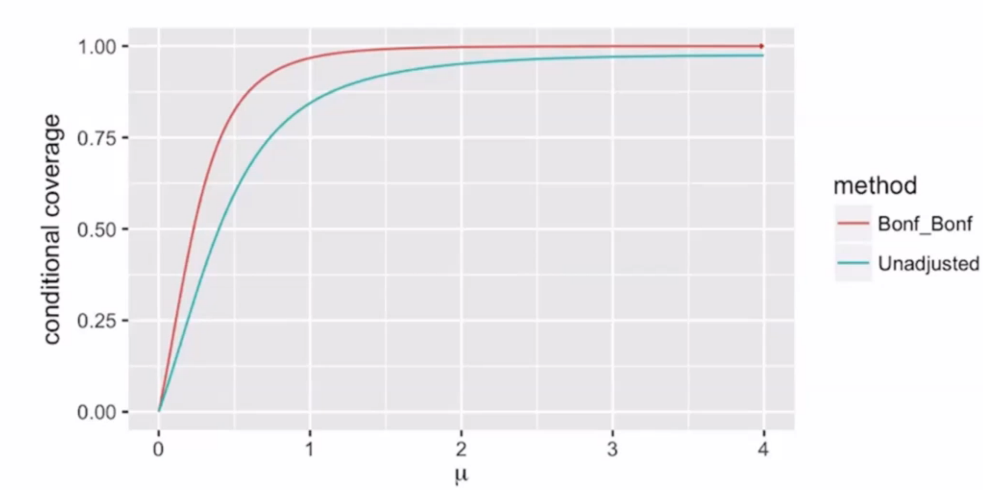
$$\text{CI}_i = [Y_i - 1.64, Y_i + 1.64]$$

Suppose we select those i whose interval does not contain 0, so:

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

We plot the conditional coverage $\mathbb{P}(\mu \in \text{CI}_i | i \in S)$ as a function of μ in Figure 12.3. We see that when μ is very small, the conditional coverage is essentially zero. This is expected- when μ is zero, we exclude all intervals that contain μ, attaining zero coverage! When μ is very large, essentially no selection occurs (i.e. $S = \{1, ..., 200\}$), so we recover the marginal coverage rate of 0.95.

Figure 12.3. Conditional Coverage achieved by Unadjusted Intervals and Bonferroni Selection/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 $\text{CI}_i = [Y_i - z(1 - \frac{\alpha}{n}), Y_i + z(1 - \frac{\alpha}{n})]$. In addition, we use Bonferroni selection, so that $S = \{i : p_i \leq \frac{\alpha}{2n}\}$ where p_i is the p-value associated to point i. When μ is small, we see in Figure 12.3 that due to our selection rule, we still do not achieve conditional coverage.

We conclude that although conditional coverage is highly desirable, in general it cannot be achieved. This is similar to why $\text{pFDR} = \mathbb{E}[\text{FDP} | R > 0]$ cannot be controlled, e.g. under the global null, conditional on making a rejection, $\text{pFDR} = 1$.

12.4 False Coverage Rate

In 2005, Benjamini and Yekutieli introduced the notion of the False Coverage Rate (FCR) [1].

Definition 1. The False Coverage Rate (FCR) is defined as:

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

where R_{CI} is the number of selected parameters and V_{CI} is the number of confidence intervals among those selected which 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 CI's control the FCR since:

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

- With selection, the marginal CI's 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 CI's will guarantee that the FCR is controlled:

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

- Any confidence region $\mathrm{CI}(\alpha)$ achieving *simultaneous* coverage ($\mathbb{P}((\theta_1, ..., \theta_n) \in \mathrm{CI}(\alpha)) \geq 1 - \alpha$) controls the FCR

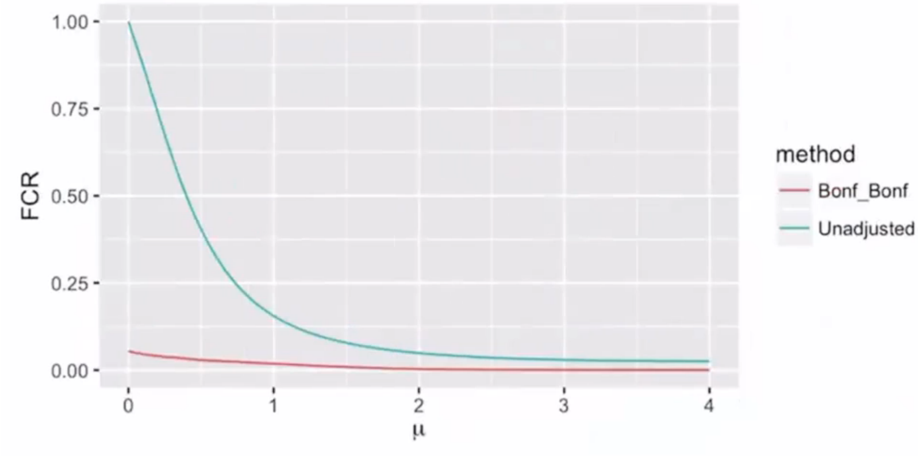
To illustrate that marginal CI's can fail to control the FCR, let us return to the example from Section 12.3. We plot the FCR for the marginal, unadjusted intervals along with the Bonferroni corrected/Bonferroni selected intervals in Figure 12.4.

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.

12.5 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 $\theta_1, ..., \theta_n$ are test statistics $T_1, ..., T_n$, summarized as $T = \{T_1, ..., T_n\}$.

Figure 12.4. False Coverage Rate achieved by unadjusted intervals and Bonferroni selected and subsequently adjusted intervals.



1. Apply any selection rule $\mathcal{S}$ to the statistics T to obtain the selection set $S = \mathcal{S}(T)$
2. For each $i \in S$, compute:

$$R^{(i)} = \min_t \{|\mathcal{S}(T^{(i)}, t)| : i \in \mathcal{S}(T^{(i)}, t)\}$$

   where $T^{(i)} = T \setminus \{T_i\}$. In words, we construct new test statistics vectors T^t by replacing observed value of the i-th entry of T with $t \in \mathbb{R}$. Then, we consider a set (denoted by $\mathcal{A}_i \subseteq \{T^t\}_t$) of all such constructed vectors $\{T^t\}_t$ that results in the selection of the i-th variable. Finally, we find the minimum number of selected variables by applying the selection procedure $\mathcal{S}$ to the constructed test statistics in $\mathcal{A}_i$. Hence $R^{(i)}$ is the smallest number of selections if we keep all $[n] \setminus i$ test statistics fixed at observed values and vary the i-th test statistic in such a way that still results in the selection of the i-th variable. Typically, $R^{(i)}$ is the same as the number of selections. For illustration, see appendix 12.6.

3. The FCR adjusted CI for $i \in S$ is $\mathrm{CI}_i \left(1 - R^{(i)} \frac{\alpha}{n}\right)$

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

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

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

Theorem 1. If the statistics T_i are independent then the FCR of the adjusted confidence intervals of the above procedure is $\leq \alpha$ [1].

Proof. Recall that:

$$\text{FCR} = \mathbb{E}\sum_{i=1}^{n} X_i, \qquad X_i = \frac{\mathbb{I}\left(i \in S, \theta_i \notin \text{CI}_i(R^{(i)}\alpha/n)\right)}{|S| \vee 1}$$

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

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

because $|S| \leq k$, by definition. Now:

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

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

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

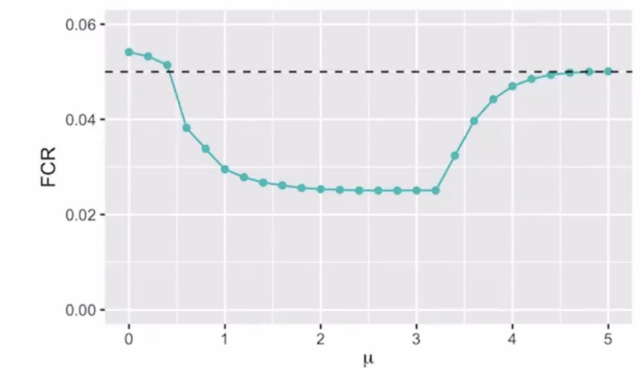
□

Figure 12.5 demonstrates how the procedure performs empirically on the problem of Section 12.3. 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.

To show the procedure is not overly conservative, consider the following example.

- We have n hypotheses, where $H_i : \theta_i = \theta_i^0$ for $i = 1, ..., n$.
- $T_i - \theta_i \sim F_i$ where the F_i are known symmetric distributions.
- We compute p-values for the two-sided test
- We select using the BH(α) procedure.

Figure 12.5. False Coverage Rate achieved by proposed procedure.



The marginal confidence intervals will be:

$$\mathrm{CI}_i(\gamma) = \{\theta_i : |T_i - \theta_i| \leq F_i^{-1}(1 - \gamma/2)\}$$

Theorem 2. Suppose $R^{(i)} = R_{\mathrm{CI}}$ almost surely and the T_i are independent. Then FCR $\geq \alpha/2$.

This result shows that the FCR controlling procedure is not overly conservative as FCR $\geq \alpha/2$ in the setting above.

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

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

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

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.

Observe from the slope of the confidence intervals do 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 [2]. The plot below shows the CIs produced by the eBayes procedure. For further information on the eBayes procedure, see reference [2]

Figure 12.6. FCR-adjusted confidence intervals.

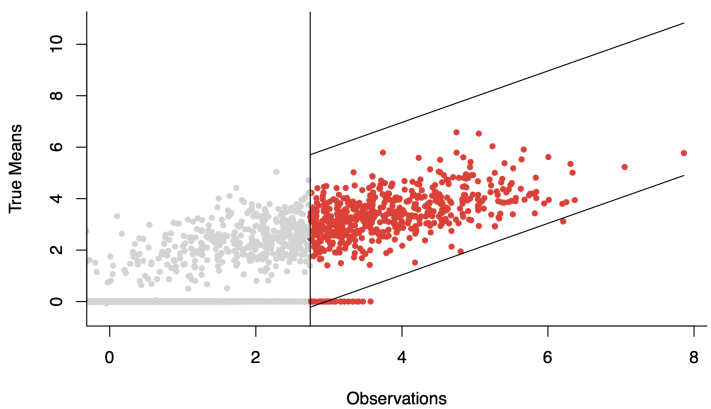
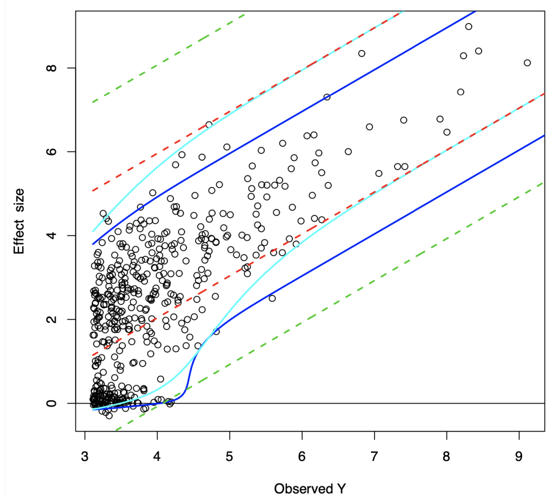


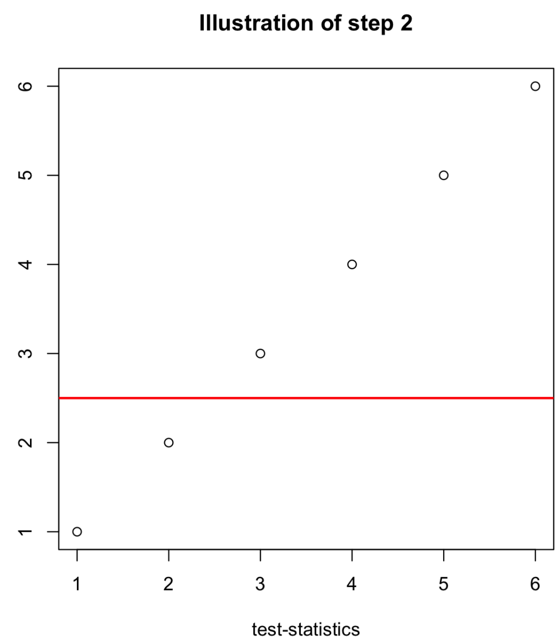
Figure 12.7. eBayes-adjusted confidence intervals.



12.6 Appendix

Here we provide a concrete illustration to computing $R^{(i)}$ as outlined in 2. Consider the example in the figure below. We first sort the test statistics; then we select all test statistics that are equal or larger than 2.5 in value. Consider the 4-th largest test statistic. The smallest value that it can take while still resulting in its selection is 2.5, resulting in 4 total selections still. Hence we find that $R^{(4)} = 4$.

Figure 12.8. Example for computing $R^{(i)}$ in step 2 of 2



Bibliography

[1] Yoav Benjamini and Daniel Yekutieli. False discovery rate–adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association*, 100(469):71–81, 2005.

[2] Daniel Yekutieli. Adjusted bayesian inference for selected parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):515–541, 2012.