677 Final

Zengqi Chen

2024-05-05

Main Points of Chapter 15: Large-Scale Hypothesis Testing and FDRs

Chapter 15 of "Computer Age Statistical Inference" by Bradley Efron and Trevor Hastie delves into the advanced topics of large-scale hypothesis testing and the control of false discovery rates (FDRs), which are particularly pertinent in the era of big data. This chapter is crucial for understanding the statistical tools needed to handle complex datasets where multiple hypotheses are tested simultaneously.

Large-Scale Hypothesis Testing

In the context of modern statistical analysis, particularly in fields like genomics, neuroimaging, and other areas where large datasets are common, the need to perform multiple hypothesis tests simultaneously is a significant challenge. The traditional methods of hypothesis testing, which control for the probability of making a Type I error (false positive) in a single test, do not scale well when thousands or even millions of tests are conducted because they lead to a high overall chance of making one or more false discoveries.

Efron and Hastie discuss the adjustment of these methods to handle multiple comparisons without inflating the overall error rate. The authors describe the application of techniques such as the Bonferroni correction and its more sophisticated alternatives, which adjust the threshold for statistical significance according to the number of tests performed.

False Discovery Rates (FDRs)

The concept of False Discovery Rate (FDR), introduced by Benjamini and Hochberg (1995), represents a paradigm shift in handling multiple comparisons. FDR is defined as the expected proportion of false discoveries among the rejected hypotheses. This is mathematically represented as:

$$\text{FDR} = \mathbb{E}\left(\frac{V}{R}\right)$$

where V is the number of false positives, and R is the total number of rejections. If no hypotheses are rejected, R is set to 1 to ensure the formula is well-defined.

Empirical Bayes Methods for Large-Scale Testing

Empirical Bayes methods are particularly highlighted in this chapter for their effectiveness in estimating the parameters used in computing the FDR. These methods use the observed data to estimate the distribution of the test statistics under the null hypothesis and the alternative hypothesis, allowing for a more accurate computation of FDR. The Empirical Bayes approach is often more powerful and flexible compared to traditional methods, especially when the assumptions about the distribution under the null hypothesis may not hold exactly.

Local False Discovery Rates (Local FDRs)

While FDR gives an overall measure of error among multiple tests, local false discovery rates (local FDRs) provide a measure at the individual test level. The local FDR is defined as the probability that a given hypothesis is null given the observed test statistic:

$$\text{local FDR}(t) = \frac{\Pr(\text{Null}|T = t)}{\Pr(\text{Alternative}|T = t)}$$

where T is the test statistic. This measure helps researchers understand the reliability of individual tests within the context of a large dataset.

Choice of Null Distribution

A crucial aspect discussed is the selection of an appropriate null distribution. This selection is fundamental in calculating both FDR and local FDRs. The authors discuss various strategies to estimate this distribution from the data, which is often a critical step in ensuring the robustness of hypothesis testing results.

In summary, Chapter 15 provides a comprehensive treatment of techniques needed to address the challenges posed by large-scale hypothesis testing. It integrates theoretical insights with practical applications, offering readers a deep understanding of both the methodologies and their implications in contemporary statistical practice. This chapter is essential for statisticians and data scientists who deal with the analysis of large-scale data sets where multiple hypothesis testing is unavoidable.

Computational Methods and R Code Examples for Chapter 15

In Chapter 15 of "Computer Age Statistical Inference," Efron and Hastie delve into computational methods vital for understanding and implementing large-scale hypothesis testing and controlling false discovery rates (FDR). This section explains these methods and includes relevant R code examples to demonstrate the application of these techniques in practice.

Estimating False Discovery Rates (FDR)

One of the primary computational challenges in large-scale testing is the estimation of the False Discovery Rate (FDR), which provides a way to control the expected proportion of incorrect rejections among all rejections. The fundamental FDR computation can be formulated as follows:

Given a set of p-values $P_1, P_2, ..., P_m$ from m independent tests, the FDR can be controlled using the Benjamini-Hochberg (BH) procedure, which is designed to ensure that the expected proportion of false positives (incorrect rejections) among the rejected hypotheses does not exceed a pre-specified level α.

The BH procedure operates as follows:

1. Sort the p-values in ascending order.
2. For each p-value P_i, compute its rank i in the ordered list.
3. Find the largest k such that $P_k \leq \frac{k}{m}\alpha$, where m is the total number of hypotheses tested.

This procedure can be efficiently implemented in R using the `p.adjust` function.

```r
# Generate random p-values for demonstration
set.seed(123)
p_values <- runif(100)

# Adjust p-values using the Benjamini-Hochberg method
adjusted_p_values <- p.adjust(p_values, method = "BH")

# Determine which hypotheses to reject
alpha <- 0.05
```

```
adjusted_p_values <= alpha
```

```
##    [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   [97] FALSE FALSE FALSE FALSE
```

Empirical Bayes for Large-Scale Testing

Empirical Bayes methods are extensively used for estimating parameters that are subsequently used in FDR calculations. These methods involve estimating the distribution of the test statistics under the null hypothesis and combining this with the prior distribution of the parameters.

The key computational aspect here involves modeling the distribution of p-values under the null hypothesis, which is often assumed to be uniform. However, deviations from this uniform distribution in observed p-values can indicate the presence of true effects. Empirical Bayes methods adjust for such deviations by estimating the prior distribution of the test statistics.

Here is an example using the `locfdr` package in R, which computes local FDR values:

```
library(locfdr)


# Simulate some test statistics under null and alternative hypotheses
set.seed(123)
z_scores <- c(rnorm(300), rnorm(100, mean=3))

# Compute local FDR values
fdr_values <- locfdr(z_scores)
```
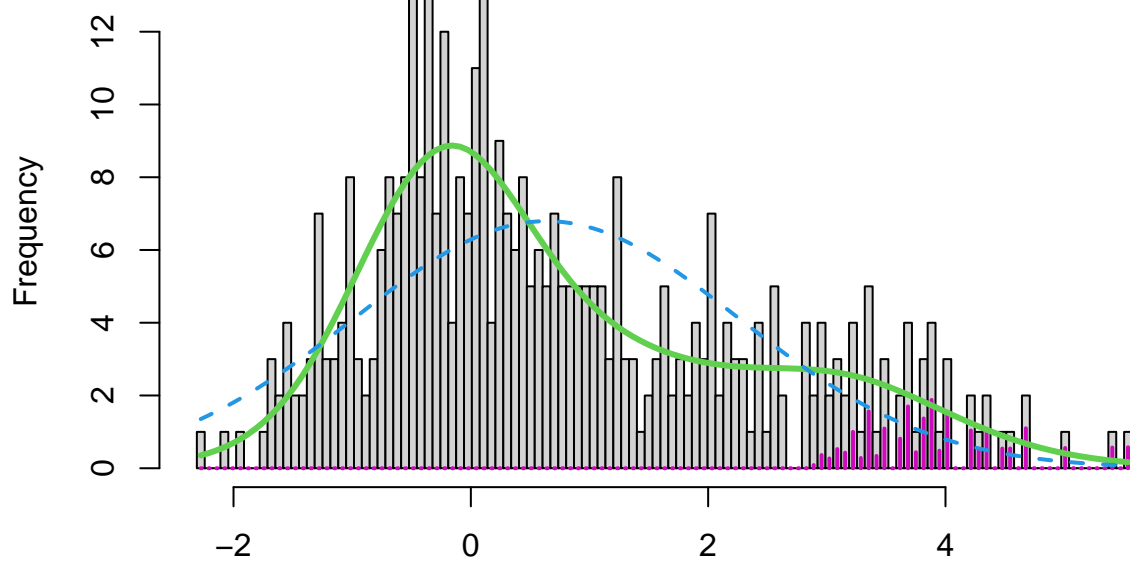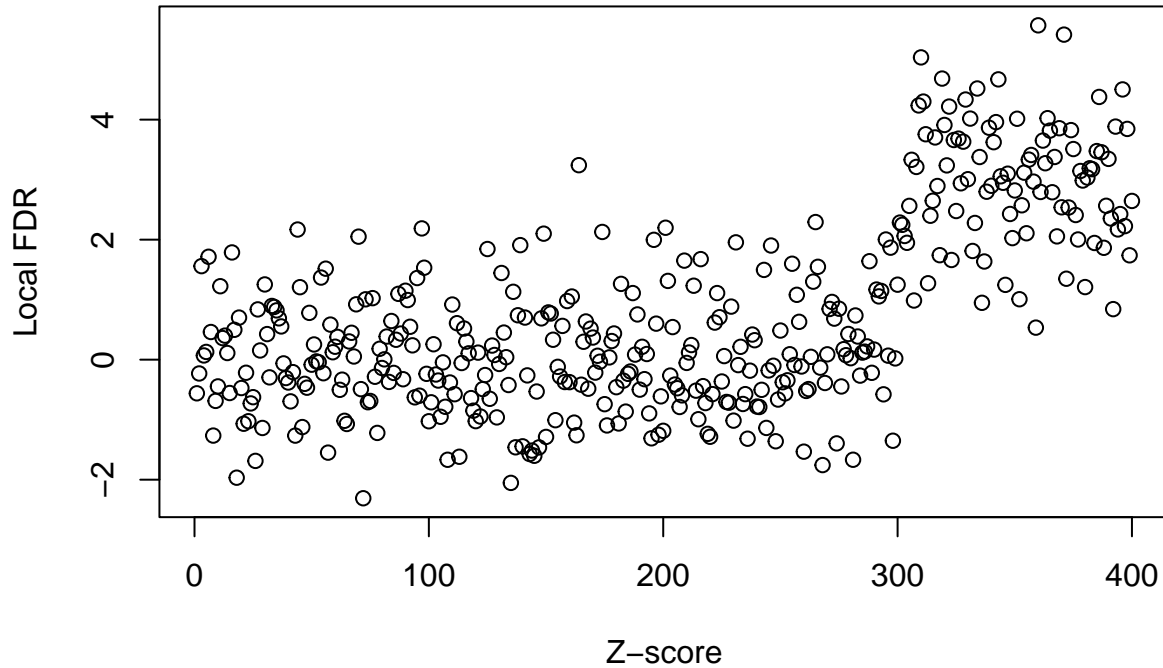
MLE: delta: 0.637 sigma: 1.622 p0: 1.005
CME: delta: −3.37 sigma: 2.656 p0: 1.61

```r
# Plot local FDR values
plot(z_scores, fdr_values$lfdr, main="Local FDR vs. Z-scores", xlab="Z-score", ylab="Local FDR")
```

Local FDR vs. Z–scores

Selection of Null Distribution

Selecting an appropriate null distribution is crucial for accurately computing FDR and local FDRs. Computational methods involve using techniques like bootstrap or smoothing methods to estimate the null distribution from the data. The choice of the null distribution impacts the entire process of hypothesis testing, especially in how the p-values are interpreted.

For example, if the null distribution is incorrectly specified as too broad, then even significant deviations might be considered as noise, leading to a high number of false negatives. On the other hand, a too-narrow null distribution might lead to too many false positives.

In summary, the computational methods discussed in Chapter 15 are essential for effective large-scale hypothesis testing. These methods leverage modern statistical software capabilities, allowing researchers to handle large datasets and perform multiple hypothesis tests with confidence in the integrity of their findings.

Mathematical Underpinnings of Chapter 15: Large-Scale Hypothesis Testing and FDRs

In Chapter 15 of "Computer Age Statistical Inference," the mathematical foundation focuses on the probabilistic frameworks and statistical theories underlying large-scale hypothesis testing and the control of False Discovery Rates (FDRs). This section discusses the mathematical concepts and derivations that are essential for understanding these statistical methodologies.

Probability Framework for Multiple Testing

Large-scale hypothesis testing involves handling multiple hypotheses simultaneously, which fundamentally depends on understanding the behavior of Type I (false positive) and Type II (false negative) errors across many tests. The key mathematical challenge is to control the overall error rates, which can dramatically increase with the number of hypotheses tested.

Bonferroni Correction

A basic approach to control the family-wise error rate (FWER) in multiple testing is the Bonferroni correction. The correction is based on the union bound from probability theory:

$$\text{FWER} = \Pr(\text{at least one false positive}) \leq \sum_{i=1}^{m} \Pr(\text{false positive in test } i)$$

Given individual tests are performed at significance level α, for m independent tests, the Bonferroni correction suggests testing each hypothesis at a significance level of α/m to ensure that the FWER does not exceed α. This is overly conservative, especially as m becomes large, but it provides a simple and rigorous control of Type I error.

False Discovery Rates (FDR)

The concept of FDR is a more sophisticated approach that allows a certain proportion of false positives among the rejected hypotheses, which is more appropriate for large-scale testing scenarios. Introduced by Benjamini and Hochberg, the FDR is mathematically defined and controlled as follows:

$$\text{FDR} = \mathbb{E}\left(\frac{V}{\max(R, 1)}\right)$$

where V is the number of false positives and R is the total number of rejected hypotheses. The expectation is over the random selection of hypotheses being true or false under the null hypothesis.

The FDR controlling procedure by Benjamini and Hochberg involves sorting the p-values and selecting a cutoff point. The largest k where:

$$P_k \leq \frac{k}{m}\alpha$$

This threshold ensures that the expected proportion of incorrect rejections is controlled at the level α.

Empirical Bayes and Local FDR

Empirical Bayes methods are used for estimating the distribution parameters necessary for FDR calculations. The local FDR, an extension of the FDR concept, gives the probability that a specific null hypothesis is true given the observed data:

$$\text{local FDR}(t) = \Pr(\text{Null} \mid T = t)$$

This is typically estimated using the ratio of the density of the test statistic under the null hypothesis to its density under the alternative hypothesis. These densities are often estimated using smoothing techniques or parametric models fitted to the data.

Mathematical Modeling of Null and Alternative Hypotheses

The accurate modeling of the null and alternative hypotheses is critical. The null hypothesis typically assumes no effect (e.g., zero mean difference), while the alternative suggests some non-zero effect. The distributions of test statistics under both hypotheses need to be well approximated to accurately compute p-values and, subsequently, FDRs.

Summary

The mathematical underpinnings of Chapter 15 provide a robust framework for understanding the complexities of large-scale hypothesis testing. The transition from simple Bonferroni corrections to more nuanced techniques like FDR and empirical Bayes methods reflects a sophisticated evolution in statistical thinking, crucial for dealing with modern scientific data analysis challenges. These methodologies not only offer more power and flexibility but also ensure that statistical inference remains reliable even when faced with vast datasets.

Historical Context of Chapter 15: Large-Scale Hypothesis Testing and FDRs

Chapter 15 of "Computer Age Statistical Inference" covers important developments in the statistical field, particularly focusing on large-scale hypothesis testing and the introduction of false discovery rates (FDR). This section explores the historical context of these concepts, tracing their evolution and the pivotal moments that shaped modern statistical practices in handling multiple comparisons.

Early Developments in Multiple Testing

The challenge of multiple testing has its roots in early 20th-century statistics, where the need to control errors across multiple hypothesis tests first emerged. Initially, the focus was primarily on controlling the family-wise error rate (FWER), which ensures that the probability of making one or more Type I errors does not exceed a pre-set level across all tests.

The simplest approach to control FWER is the Bonferroni correction, derived from Boole's inequality, which was widely used due to its simplicity and stringent control over Type I errors:

$$\text{FWER} \leq \sum_{i=1}^{m} P(\text{Type I error for test } i)$$

However, as the number of tests increased with advances in data collection and generation, the Bonferroni correction became impractically conservative, leading to a high rate of Type II errors (false negatives).

The Introduction of False Discovery Rates

The concept of False Discovery Rate (FDR) was introduced by Yoav Benjamini and Yosef Hochberg in their seminal 1995 paper. They proposed a less conservative approach that directly addressed the challenges posed by modern large-scale testing environments. FDR is defined as the expected proportion of false positives among the rejected hypotheses, fundamentally changing how multiple comparisons are handled.

The original FDR controlling procedure, now known as the Benjamini-Hochberg (BH) procedure, was a landmark in statistical methodology because it allowed researchers to handle the multiplicity of tests without severely compromising power. The BH procedure orders p-values and finds the largest number, k, such that:

$$P_k \leq \frac{k}{m}\alpha$$

This procedure was a significant shift from controlling the probability of any false positives (FWER) to controlling the expected proportion of false positives (FDR).

Empirical Bayes and Modern Developments

Another major development discussed in Chapter 15 is the application of Empirical Bayes methods in large-scale testing. These methods, which estimate the distribution of test statistics under the null hypothesis using observed data, represent a fusion of classical and Bayesian statistics. The introduction of local FDRs, which assess the probability that a specific null hypothesis is true given the observed test statistic, further refined the resolution of statistical tests:

$$\text{local FDR}(t) = \Pr(\text{Null} \mid T = t)$$

Empirical Bayes methods provide a more nuanced approach to handling large datasets, where the assumptions of traditional methods often do not hold.

Summary

The historical context of the methodologies discussed in Chapter 15 illustrates a significant evolution in the field of statistics. From the conservative Bonferroni correction to the more nuanced and powerful FDR and Empirical Bayes methods, statistical practice has adapted to the increasing complexity and scale of data in modern research. These developments reflect a broader trend in statistics towards more adaptive and data-driven methodologies, enabling researchers to make more informed decisions in the face of uncertainty.

Statistical Practice Implications of Chapter 15: Large-Scale Hypothesis Testing and FDRs

Chapter 15 of "Computer Age Statistical Inference" significantly influences statistical practice by introducing and elaborating on methodologies for large-scale hypothesis testing and False Discovery Rates (FDR). The broader implications for statistical practice involve not only the adoption of these methods across various scientific disciplines but also a paradigm shift in how data analysis is approached in the context of multiple comparisons.

Enhanced Accuracy and Power in Hypothesis Testing

The shift from traditional methods such as the Bonferroni correction to more sophisticated techniques like the FDR and local FDR allows statisticians to achieve a balance between minimizing Type I errors (false positives) and maximizing the test's power. This balance is crucial for scientific fields where making correct decisions based on hypothesis testing can have significant implications, such as in medical research and policy making.

Data-Driven Adaptive Methods

The introduction of Empirical Bayes and FDR-based methods has ushered in more adaptive and data-driven approaches. These methods leverage the data itself to inform decisions about which hypotheses to reject, allowing for more flexible responses to the data's inherent characteristics. This adaptability is especially beneficial in high-dimensional data environments, such as genomics and large-scale social sciences experiments, where the relationships between variables can be complex and not well-suited to traditional fixed threshold methods.

Expansion of Statistical Methodology Applications

The methodologies discussed in Chapter 15 have enabled statistical analyses that were previously not possible due to the limitations of earlier techniques. Researchers can now undertake more ambitious projects, such as large-scale A/B testing in tech industries, genome-wide association studies in genetics, and real-time fraud detection in finance, all of which rely heavily on the ability to make accurate inferences when testing thousands or millions of hypotheses.

References

- **Benjamini, Y., & Hochberg, Y. (1995).** "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing."

- **Efron, B., & Hastie, T. (2016).** "Computer Age Statistical Inference."

- **Storey, J.D. (2002).** "A direct approach to false discovery rates."

- **Tusher, V.G., Tibshirani, R., & Chu, G. (2001).** "Significance analysis of microarrays applied to the ionizing radiation response."