Conditional Hypothesis Testing

1 author:

Kun Joo Michael Ang
Bloomberg LP

Conditional Hypothesis Testing

Kun Joo Michael Ang

July 16, 2019

Abstract

When testing multiple hypotheses, conventional techniques used for reducing false positives require all tests to be pre-specified and do not account for correlation between p-values. This makes them incompatible with sequential modelling techniques, where models are built one-at-a-time and future models benefit from the insight of previous testing. We propose here a technique for adjusting future tests to incorporate prior information and show that this reduces to replacing the likelihood function with the conditional likelihood. A numerical algorithm is also developed that uses Monte Carlo integration to efficiently compute conditional acceptance regions from conditional sizes.

Keywords— multiple hypothesis testing, conditional hypothesis testing, conditional likelihood, Monte Carlo integration

1 Introduction

When multiple hypothesis tests are performed on the same data, the risks of encountering a False Positive(FP) result increase. In statistical literature, the two most popular techniques for lowering FP rates are the Bonferroni Correction [1] and the Benjamini-Hochberg procedure [2]. Both of these ensemble tests involve performing N hypothesis tests simultaneously, then shrinking the size of each test to minimize either the False-Discovery Rate or the Family-Wise Error Rate. While both mathematically elegant and easily implemented, we argue that neither test has enough flexibility to account for how data is modelled in practice.

It is quite common for hypothesis tests to be developed sequentially, often in pursuit of a model that sufficiently explains the data to a specified degree of accuracy. A statistician first analyses the data and then uses his subject matter expertise to propose a likely model. If attempts to fit the data to the model fail (at some chosen significance level), the statistician returns to the data and again uses his expertise to develop a different model, perhaps a more complicated one that controls for specific effects that the former model failed to account for. The sequential nature of this form of modelling means future models are constructed in full knowledge of and conditional on the failure of initial testing. Later tests are conducted only if prior modelling fails to capture sufficient variation in the data. Because we cannot know ahead of time how many tests will be conducted, neither correction technique is fully compatible. While one could naively increase N retroactively before applying either correction, this can be quite problematic. Hypothesis tests are often highly correlated when models are nested or similar in design. The Bonferroni correction is known to over-correct in such cases and reduces statistical power. Both corrections are also fundamentally incompatible with the situations described, as they are statements about FPs on a family of hypotheses constructed before testing, and we are interested in only the current hypothesis, subject to the information available during the time of testing.

The proposed solution is best described as Conditional Hypothesis Testing. The technique modifies the likelihood function of the model to incorporate information about prior testing, then constructs acceptance regions based on the conditional likelihood. This method of correction ignores family-wise errors in lieu of constructing tests with accurate significance levels, subject to any information gained up to the testing times. It should most clearly be used in situations where the hypotheses were not all developed simultaneously prior to testing, and future models have gained the insight of previous ones.

There are a few other minor inconveniences to the Bonferroni Correction and the Benjamini-Hochberg that Conditional Hypothesis Testing avoids. Both correction techniques impose restrictions on individual significance levels α_i. In the Bonferroni correction, the sum of significance levels is bounded above by the family significance level and in the Benjamini-Hochberg, acceptance/rejection is controlled by a single family-wise error rate. In practice, we might demand a better fit (smaller test size) for models with additional degrees of freedom or additional complexity. Conditional Hypothesis Testing allows for this asymmetry by modifying only the likelihood function, not α_i. Finally, neither technique correctly adjusts for correlation between tests. Acceptance regions will have a size error increasing in correlation between p-values.

In the next sections, we will derive the mathematics behind the Conditional Hypothesis Testing technique and remark on sufficient conditions for the existence of the conditional likelihood function. Taking it further, we will also develop a numerical algorithm for efficiently computing the conditional likelihood functions and constructing the conditional acceptance regions. This is useful in the situations where the acceptance regions do not have a closed-form expressions, likelihood functions are difficult to integrate, and acceptance regions are difficult to construct from the conditional likelihood.

2 Conditional Hypothesis Testing

Let X be a set of observations and $i = 1, 2, .., N$ be the indices of an ordered set of hypotheses tests conducted sequentially on X. I_i are indicator functions of the success of each test, α_i its (conditional) size and C_i its acceptance region.

For $i = 2, 3, ..., N$, define the conditional hypothesis tests as

$$H_0 : (\theta \in \Theta_0 | I_1, I_2, ..., I_{i-1})$$

$$H_a : (\theta \in \Theta \backslash \Theta_0 | I_1, I_2, ..., I_{i-1})$$

where the model parameter and domain are as in the original hypothesis tests. Each conditional test is constructed recursively from the results of previous tests. Both the model $f_i(I_1, I_2, ..., I_{i-1})$ and the conditional significance level $\alpha_i(f_i, I_1, I_2, ..., I_{i-1})$ can be functions of prior testing and results. Note that even if α_i is independent of prior information, past-dependence is embedded within the Conditional Hypothesis Test and α_i does not have the same interpretation as the α_i in Benjamini-Hochberg or the Bonferroni correction, or any unconditional test in general.

Next, we can rewrite the conditioning as

$$H_0 : \left(\theta \in \Theta_0 \,\middle|\, X \in D_i\right) \quad H_a : \left(\theta \in \Theta \backslash \Theta_0 \,\middle|\, X \in D_i\right)$$

$$\text{where} \quad D_i = \bigcap_{k=1}^{i-1} \left[I_k \cdot C_k + (1 - I_k) \cdot C_k' \right]$$

Letting $f_i(\theta, X)$ denote the likelihood functions of the models in test i, then the conditional likelihood is

$$f_i(\theta, X \big| X \in D_i) = \begin{cases} 0 & \text{if } f_i(\theta, X) = 0 \text{ or } X \notin D_i \\ \frac{f_i(\theta, X) \mathbb{1}_{\{X \in D_i\}}}{\int_{D_i} f_i(\theta, U) dU} & \text{otherwise} \end{cases}$$

To ensure the conditional likelihood is always well-defined, we make the assumption that D_i has strictly positive measure in f_i for all $\theta \in \Theta$ whenever $f_i(\theta, X)\mathbb{1}_{\{X \in D_i\}}$ is non-zero. With these conditional distributions as functions of θ, we can construct the acceptance region $C_i(\alpha_i)$. In practice, this is difficult to do for general Θ, Θ_0, and optimal acceptance regions (uniformly most powerful) are not guaranteed to exist, but many of the results from classical hypothesis testing will still apply.

For example, for nested models, we can compute the likelihood ratio as before,

$$\Lambda = -2 \log \frac{\sup_{\theta \in \Theta_0} f_i(\theta, X | X \in D_i)}{\sup_{\theta \in \Theta} f_i(\theta, X | X \in D_i)}$$

asymptotically perform the Likelihood Ratio Test and the results still hold in the conditional case.

3 Algorithm Implementation

Our technique has two practical challenges. First, it is necessary to keep track of the domain D_i over time. As is the intersection of multiple acceptance/rejection regions of former tests, it is likely that no explicit analytical expression for D_i exists. Second, in cases where the acceptance regions $C_i(\alpha_i)$ are easy to compute in the unconditional case, it is not clear what the equivalent regions are in the conditional case, in particular when we insist on properties such as a uniformly most powerful test or a concave domain.

We propose the following algorithm that eliminates both problems, presented in the special case where the null hypothesis is always a point estimate $\Theta_0 = \{\theta_0\}$ and we are fitting models to explain our data X. Let $\hat{C}_i(\alpha_i)$ denote the acceptance region of test i if it were the only test to be conducted. This will be called the unconditional acceptance region and the unhatted version the conditional acceptance region.

1. For the first test, conduct a regular hypothesis test. Here $\hat{C}_1 = C_1$ and if $X \in C_1$, we have found a fitted model and can terminate.

2. For further tests, $C_i(\alpha_i)$ are constructed as follows.

   (a) For a large value of N, simulate N realizations of X from the null hypothesis density $f_i(\theta_0, X)$

   (b) Approximate $\int_{D_i} f_i(\theta, X) dX \approx \frac{1}{N} \sum_{k=1}^{N} f_i(\theta_0, X_k)\mathbb{1}_{\{X_k \in D_i\}}$. It will be shown later that $\mathbb{1}_{\{X_k \in D_i\}}$ can be easily computed.

(c) Start with the unconditional acceptance region $\hat{C}_i(\alpha_i)$ and compute $\int_{\hat{C}_i(\alpha_i)} f_i(\theta_0, X|X \in D_i)dX$. This can again be done by Monte-Carlo simulation as in (b).

(d) If the computed value in (c) is less than $1 - \alpha_i$, enlarge the acceptance region by recomputing (c) with $\alpha_i' < \alpha_i$. If the converse is true, apply $\alpha_i' > \alpha_i$ to shrink the acceptance region.

(e) Once the conditional density integrates to $1-\alpha_i$ in $\hat{C}_i(\alpha_i')$, set $C_i(\alpha_i) = \hat{C}_i(\alpha_i')$. This acceptance region will have exactly size α_i in the conditional hypothesis test.

3. If at any time $X \in C_i$, terminate immediately and declare $f_i(\theta_0)$ the fitted model.

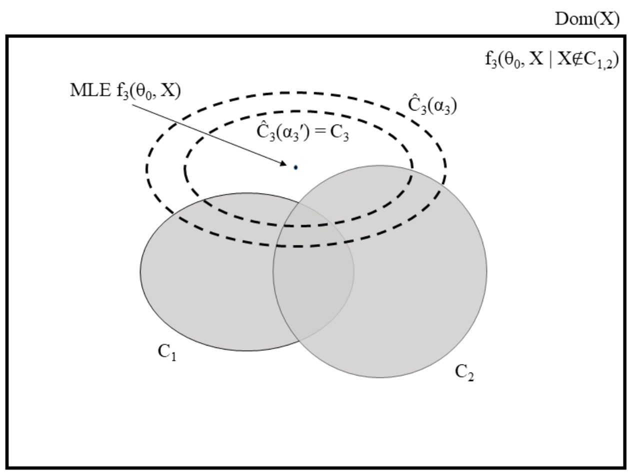

Figure 1: Example Conditional Acceptance Region for Test 3

This algorithm has several advantages. All the conditional acceptance regions $C_i(\alpha_i)$ have an representation $\hat{C}_i(\alpha_i')$, so given any observation X we can easily compute $\mathbb{1}_{\{X \in C_i\}}$ by performing the equivalent unconditional test. (2b) is established by the relationship $\mathbb{1}_{\{X \in D_i\}} = \prod_{k=1}^{i-1} \mathbb{1}_{\{X \notin C_i\}}$. Because $\mathbb{1}_{\{X \in D_i\}}$ is cheap to compute for any X, it spares us from having to keep D_i in analytic form since we can always recover the density integral by numerical simulation. The algorithm can also be extended to the case where we continue testing despite having passed previous tests by simply modifying the corresponding indicator function.

Unfortunately, although this computationally convenient method preserves the size of each hypothesis test, it cannot guarantee power. Even when the unconditional acceptance region C_i is uniformly-most-powerful or at least not strictly dominated, there is no guarantee that $\hat{C}_i$ will not be uniformly dominated by another test in the conditional setting. In situations where power cannot be compromised, we recommend constructing the acceptance regions from areas with the highest conditional likelihood. However, this will compromise the algorithm's simplicity and can quickly become numerically intractable.

4 Conclusions

This paper considers existing correction techniques for testing multiple hypotheses. Analysing their shortcomings in dealing with sequential hypothesis testing, we propose a general technique that creates acceptance regions with the correct conditional power level. Because computing acceptance regions is computationally expensive under this regime, we propose a more efficient algorithm that produces tractable acceptance regions with correctly specified significance at the potential cost of power.

References

[1] C. E. Bonferroni *Teoria statistica delle classi e calcolo delle probabilit.* Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 1936.

[2] Yoav Benjamini and Yosef Hochberg *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), vol. 57, no. 1, 1995, pp. 289300