Literature Review

Connor Demorest

9/23/2021

Romano et al., "Multiple Testing"

Romano et al. (2010) reviews the notation and vocabulary common in multiple hypothesis testing procedures. He defines MHT as "any instance that involves the simultaneous testing of several hypotheses." If we define $X = X^{(n)}$ as the data generated from any probability distribution P where n refers to the sample size, then we can define H_s as a null hypothesis and H_s' as an alternative hypothesis. A multiple testing procedure (MTP) is a rule that makes a decision about each H_s. A "false discovery" is a rejection of a true null hypothesis. A test of H_s is based on a test statistic is $T_{n,s}$ and a p-value for testing H_s is $\hat{p}_{n,s}$. The "Family-Wise Error Rate" (FWE) is the probability of one or more false discoveries. The FWER is often controlled asymptotically: $\lim sup_{x \to \infty} FWE_P \leq \alpha$ for any P. There are methods for FWE control based on p-values, such as Bonferroni, where H_s is rejected if $\hat{p}_{n,s} \leq \alpha/S$ where there are S tests. Holm is slightly better, which first orders the p-values from lowest to highest, then rejects if the smallest p-value $\hat{p}_{n,(1)} \leq \alpha/S$, then tests $\hat{p}_{n,(2)} \leq \alpha/(S-1)$, until the hypotheses are no longer rejected. Holm is a "step-down" method, testing the most significant hypotheses first, where the Bonferroni is a "Single-step" method. There are also "step-up" methods that start from the least significant hypotheses first. Methods based on p-values typically assume either independence (or other worst case) or a convenient dependence structure, which can lead to losses in power or loss of FWE control if those assumptions aren't satisfied. There is also a lot of work in using the false discovery proportion (FDP), which is the proportion of total rejections that are false. This allows some false discoveries to be found, instead of an α chance of rejecting a single null hypothesis. The False Discovery Rate (FDR) is the expected value of the FDP, $FDR_P = E_P[FDP]$ is controlled so that $FDR \leq \alpha$. This allows us to gain a lot of power when there are many hypotheses.

Storey, "False Discovery Rates"

Storey (2011) describes the different ways we think about false discovery rates. We have the traditional FDR, where $FDR = E[\frac{V}{R \cup 1}] = E[\frac{V}{R}|R > 0]Pr(R > 0)$. V is the number of false discoveries (also: Type I errors, false positives) and R is the number of total discoveries. With FDR, if there are no discoveries (R = 0), the FDR = 0. Two other definitions have been proposed as well: "positive FDR" (pFDR) and "marginal FDR" (mFDR). The pFDR is defined as $pFDR = E[\frac{V}{R}|R > 0]$ and mFDR is $mFDR = \frac{E[V]}{E[R]}$. The mFDR = pFDR = 1 when V = 0. When the number of tests is large, all three are very similar. The two methods to control the FDR are FDR control and FDR estimation. FDR control fixes the acceptable FDR level before data collection, then finds a data dependent threshold rule that asymptotically controls the FDR. FDR estimation fixes the p-value threshold at a value and then forms a point estimate of the FDR. In a Bayesian context, the pFDR exactly equal to $Pr(H_i = 0|P_i \leq t)$ if hypotheses are $H_i \sim iid\ Bernoulli(1 - \pi_0)$ where H_1 is a false null hypothesis, and $P_i|H_i \sim iid\ (1 - H_i)G_0 + H_iG_1$ where G_0 is the null distribution and G_1 is the alternative distribution. A q-value is also called the "Bayesian posterior Type I error rate", and $q - value(p_i) = min_{t \geq p_i} Pr(H_i = 0|P_i \leq t)$ is a Bayesian analogue to a p-value. This is called a "local FDR" (lFDR).

Sarkar, "Controlling the False Discovery Rate"

Sarkar (2008) summarizes the FDR control methods common in 2008 and shows some different proofs or intuitions for them. He proposes that the "False Nondiscovery Rate" (FNR) should be the new frontier

for reseach. He shows that the independence assumption is stronger than necessary for the BH method if they are using a stepup or stepdown version, all that is required is to assume that the p-values are "positive dependent", a more relaxed assumption. He talks about the adaptive methods and relates them to the point estimate methods described by Storey and others. He also talks about methods that have no specific dependence assumption about the p-values.

Sun & Cai, "Oracle and Adaptive Compound Decision Rules for False Discovery Rate Control"

Sun and Cai (2007) discusses using z test statistics instead of p-values as a way to improve the efficiency of the methods to control FDR. These methods asymptotically achieve the same performance as an oracle procedure more efficiently than p-value methods. The z value procedure yield a lower mFNR for the same mFDR level compared to p-value methods. At the same FDR level, the z value methods reject more hypotheses than the p-value methods. The LFDR in this situation is the optimal oracle statistic in the MHT problem where the LFRD(X) controls the mFDR with the smallest mFNR.

Lei and Fithian, "Power of Ordered Hypothesis Testing"

Lei and Fithian (2016) talks about classifying MHT problems into three categories: batch testing, ordered testing, and online testing. Batch refers to situations when the order of hypotheses are irrelevent. Ordered testing are situations where the order of hypotheses encode prior information, usually about the most promising hypotheses (eg. GWAS, LASSO GOF). Online testing is when the ordering of hypotheses imposes a restriction on the selection procedure. The propose a new ordered testing procedure that uses adaptive methods to improve on the way to choose a stopping index of the ordered hypotheses. This improves on BH and other methods in terms of power under certain regularity conditions.

References:

Lei, Lihua, and William Fithian. 2016. "Power of Ordered Hypothesis Testing." In *International Conference on Machine Learning*, 2924–32. PMLR.

Romano, Joseph P, Azeem M Shaikh, Michael Wolf, and others. 2010. "Multiple Testing." *The New Palgrave Dictionary of Economics. Forthcoming.*

Sarkar, Sanat K. 2008. "On Methods Controlling the False Discovery Rate." *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, 135–68.

Storey, John D. 2011. "False Discovery Rate." *International Encyclopedia of Statistical Science* 1: 504–8.

Sun, Wenguang, and T Tony Cai. 2007. "Oracle and Adaptive Compound Decision Rules for False Discovery Rate Control." *Journal of the American Statistical Association* 102 (479): 901–12.