Lecture

Evan Ott
UT EID: eao466

November 21, 2016

Note: I was 10 minutes late, so catching up.

Notation

		Decision		
		$\hat{H}_i = 0$	$\hat{H}_i = 1$	Total
Truth	$H_i = 0$	TN	FP	T_0
	$H_i = 1$	FN	TP	T_1
	Total	N	P	M

Sensitivity is the same as “Recall,” “Probability of Detection,” and is equal to TP/T_1

Specificity is the same as “True negative rate” and is equal to TN/T_0

Precision is the same as “Positive predictive value” and is equal to $TP/(TP + FP) = TP/P$

False Positive Rate is F/T_0

False Discovery Rate (FDR) is $FP/(FP + TP) = FP/P$

Mostly, people are looking more at the false discovery rate because it has nice properties when you control it.

Benjamini-Hochberg

See (Benjamini and Hochberg, 1995).

Say we have M p-values for M tests $P_1, \ldots, P_M$. Sort them so that

$$P_{(0)} = 0$$
$$P_{(0)} \leq P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(M)}$$

The testing procedure is a map from $[0, 1]^M \to [0, 1]$ to create the cutoff point.

If we found that our p-values were all more-or-less uniformly distributed between 0 and 1, we probably don’t have any real signals, so we’d need a really tight threshold (close to 0). However, if we have more p-values close to 0, then we’d want a stronger threshold.

So the threshold is

$$T_{BH} = \max \left\{ P_{(i)} : P_{(i)} \leq \alpha \frac{i}{M} \right\}$$

where α is the target FDR

If we were to look at Bonferroni threshold instead, that would be $T_{\text{Bonf}} = \frac{\alpha}{M}$. It’s a bit of an interesting coincidence that the value of the Bonferroni threshold is equal to the slope of the Benjamini- Hochberg.

The conclusion of the Benjamini-Hochberg test is that $\hat{H}_i = 1$ for all $P_i \leq T_{BH}$ and $\hat{H}_i = 0$ otherwise. On average, then, we control the FDR to be, on average α.

Now, this procedure is the “ordinary least squares” of multiple testing, in that it is the standard applied pretty much everywhere.

Let’s look at the “False discovery proportion,” (FDP):

$$
\begin{aligned}
FDP(t) &= \frac{\sum_{i=1}^{M} \mathbb{1}\left\{P_i \le t\right\}(1-H_i)}{\sum_{i=1}^{M} \mathbb{1}\left\{P_i \le t\right\} + \mathbb{1}\left\{\text{all } P_i > t\right\}} \\
FDR(t) &= \mathbb{E}\left[FDP(t)\right]
\end{aligned}
$$

where the expectation is taken under the true data-generating process.

Let's say

$$
\begin{aligned}
H_1, \ldots, H_M &\sim \text{Bernoulli}(a) \\
(P_i | H_i = 0) &\sim \text{Uniform}(0,1) \\
(P_i | H_i = 1) &\sim F
\end{aligned}
$$

where F is some arbitrary CDF on $[0,1]$ stocastically smaller than $\mathrm{U}(0,1)$. Marginally, the p-values follow

$$
P_i \sim G, \qquad G = aF + (1-a)\mathrm{U}(0,1)
$$

Let's create the empirical CDF of the p-values as $\hat{G}$.

$$
\hat{G}(t) = \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}\left\{P_i \le t\right\}
$$

Assuming no ties, this is

$$
\hat{G}(P_{(i)}) = \frac{i}{M}
$$

If instead of taking the actual expectation in order to calculate the FDR, we took the separate expectations of the numerator and denominator, we have

$$
\begin{aligned}
FDR(t) &= \mathbb{E}\left[FDP(t)\right] \\
&\approx \frac{\mathbb{E}\left[\sum_{i=1}^{M} \mathbb{1}\left\{P_i \le t\right\}(1-H_i)\right]}{\mathbb{E}\left[\sum_{i=1}^{M} \mathbb{1}\left\{P_i \le t\right\} + \mathbb{1}\left\{\text{all } P_i > t\right\}\right]} \\
&= \frac{\mathbb{E}\left[\sum_{i=1}^{M} \frac{1}{M}\mathbb{1}\left\{P_i \le t\right\}(1-H_i)\right]}{\mathbb{E}\left[\sum_{i=1}^{M} \frac{1}{M}\mathbb{1}\left\{P_i \le t\right\} + \frac{1}{M}\mathbb{1}\left\{\text{all } P_i > t\right\}\right]} \\
&= \frac{(1-a)t}{G(t) + \frac{1}{M}\left(1-G(t)\right)^M} \approx \frac{(1-a)t}{G(t)}
\end{aligned}
$$

(the expectation of the empirical CDF is the CDF)

Okay, let's tie things together. Benjamini-Hochberg says: $F\hat{D}R(t)$ from $FDR(t) = \frac{(1-a)t}{G(t)}$ with

(1) $a = 0$ (conservative choice)

(2) $G(t) = \hat{G}(t)$

So $\alpha = F\hat{D}R(t) = \frac{t}{\hat{G}(t)}$. If we knew a inherently, we could do better than Benjamini-Hochberg because we wouldn't have to be as conservative. We could also regularize / smooth the empirical CDF to have a better estimate. Doing either of these moves you closer to a Bayesian / empirical Bayesian method.

Bayes Two-Groups Model

In this case, don't use p-values, just use the test statistic.

$$
\begin{aligned}
Z_i | \mu_i &\sim N(\mu_i, \sigma^2) \quad \text{(problem dependent)} \\
\mu_i &\sim \omega \cdot F + (1-\omega)\delta_0
\end{aligned}
$$

In the data for class, $F = N(0, \tau^2)$.

Can write down the MLE given τ^2 and ω, and optimize it, even just numerically. James' code in R will do just that. Finally, just compute the posterior probabilities given those values for all the data points, and can engineer it to be very close to having the same idea as the FDR.

References

[1] 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)* (1995), pp. 289–300. JSTOR: 2346101.