Stats 300C: Theory of Statistics Spring 2018

Lecture 2 — April 04, 2018

Prof. Emmanuel Candes

Scribe: Paulo Orenstein; edited by Stephen Bates, XY Han

1 Outline

Agenda: Global testing

1. Needle in a Haystack Problem
2. Threshold Phenomenon
3. Optimality of Bonferroni's Global Test

Last time: We introduced Bonferroni's global test. In this lecture, we show that Bonferroni's method is somehow optimal for testing against sparse alternatives. This claim relies on power calculations, which require us to specify alternatives.

In this lecture, we consider an independent Gaussian sequence model:

$$y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1), \quad i = 1, \dots, n.$$

We are interested in the n hypotheses

$$H_{0,i} : \mu_i = 0$$

so that in this case, the global null asserts that all the means μ_i vanish. Under the alternative H_1, some means $\mu_i \neq 0$.

We saw that Bonferroni's method rejects if $\max y_i \geq |z(\alpha/n)|$ in the one-sided case, and if $\max |y_i| \geq |z(\alpha/2n)|$ in the two-sided case. Put another way, Bonferroni rejects the global null hypothesis if the largest y_i is large enough. For the special case where the n tests are mutually independent, we also calculated

$$\mathbb{P}_{H_0}(\text{Type I Error}) := q(\alpha) \approx 1 - e^{-\alpha} \approx \alpha.$$

2 Magnitude of Bonferroni's Threshold

How large is our threshold $t = |z(\alpha/n)|$ (one sided) or $|z(\alpha/2n)|$ (two sided)? If $\phi(t)$ is the standard normal pdf, then we can derive by Markov's Inequality the useful result:

$$\frac{\phi(t)}{t}\left(1 - \frac{1}{t^2}\right) \leq \mathbb{P}(Z > t) \leq \frac{\phi(t)}{t},$$

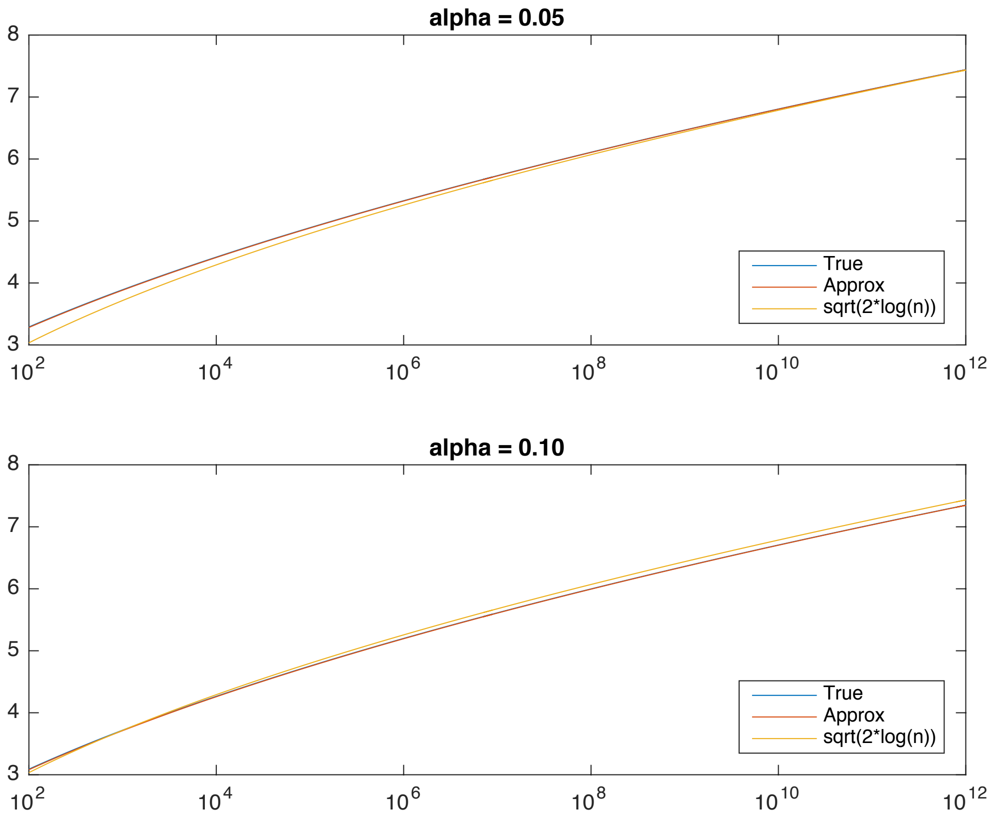

Figure 1: $z(\alpha/n)$, $\sqrt{B\left(1-\frac{\log B}{B}\right)}$ ("Approx") , and $\sqrt{2\log n}$ for $n \in \{10^2, 10^{12}\}$.

where $Z \sim N(0,1)$. That is, for large t, $\frac{\phi(t)}{t}$ is a good approximation to the normal tail probability (Gaussian quantile). Roughly speaking, then,

$$\mathbb{P}(Z > t) = \alpha/n \text{ “} \Longleftrightarrow \text{”} \frac{\phi(t)}{t} \approx \alpha/n.$$

Holding α fixed, then, we can show that for large n,

$$\begin{aligned} |z(\alpha/n)| &\approx \sqrt{2\log n}\left[1 - \frac{1}{4}\frac{\log\log n}{\log n}\right] \\ &\approx \sqrt{2\log n}. \end{aligned}$$

Hence, the quantiles grow like $\sqrt{2\log n}$, with a small correction factor. Figure 1 plots $z(\alpha/n)$ and $\sqrt{2\log n}$. Notice that Bonferroni then basically amounts to rejecting when $\max|y_i| > \sqrt{2\log n}$.

One remarkable fact about all of this is that there is (asymptotically) no dependence on α. That is, whatever α we use, our rejection threshold for $\max|y_i|$ is asymptotic to $\sqrt{2\log n}$. This is a consequence of the fact that, under H_0,

$$\frac{\max|y_i|}{\sqrt{2\log n}} \xrightarrow{p} 1.$$

In other words, the first order term $\sqrt{2\log n}$ asymptotically dominates the terms containing α.

For finite samples, it is of course possible to develop approximations to $z(\alpha/n)$ which are a both more accurate than $\sqrt{2\log n}$. Set

$$B = 2\log(n/\alpha) - \log(2\pi) = 2\log(n/\alpha) - 1.8379.$$

Then

$$|z(\alpha/n)| \approx \sqrt{B\left(1 - \frac{\log B}{B}\right)}.$$

Figure 1 shows that this approximation is nearly indistiguishable from $z(\alpha/n)$, even for modest values of n.

3 Sharp Detection Threshold for the "Needle in a Haystack"

Asymptotic Power: Consider a sequence of problems with $n \to \infty$. How powerful is Bonferroni, or, put it another way, what is the limiting power $\mathbb{P}_{H_1}(\max y_i > z(\alpha/n))$?

Needle in a Haystack Problem: To answer the question above, we need to specify alternative hypotheses. The *needle in a haystack* problem is this: under the alternative, *one* $\mu_i = \mu > 0$. We don't know which one.

For the needle in the haystack problem, we shall see that the answer to the power question depends very sensitively on the limiting ratio $\frac{\mu^{(n)}}{\sqrt{2\log n}}$, where $\mu^{(n)} > 0$ is the value of the single nonzero mean. (The (n) in the superscript captures the dependence of $\mu^{(n)}$ on n). There are two cases.

1. Asymptotic full power above threshold: Suppose $\mu^{(n)} > (1+\varepsilon)\sqrt{2\log n}$. Then, assuming without loss of generality that $\mu_1 = \mu^{(n)}$,

$$\begin{aligned}
\mathbb{P}_{H_1}(\max y_i > |z(\alpha/n)|) &\geq \mathbb{P}(y_1 > |z(\alpha/n)|) \\
&= \mathbb{P}(z_1 > |z(\alpha/n)| - \mu^{(n)}) \\
&\to 1
\end{aligned}$$

In the second to last step, we use the fact that $y_1 = z_1 + \mu^{(n)}$ where z_1 follows $N(0,1)$.

2. Asymptotic powerlessness below threshold: Suppose $\mu^{(n)} < (1-\varepsilon)\sqrt{2\log n}$. Then

$$\begin{aligned}
\mathbb{P}_{H_1}(\max y_i > |z(\alpha/n)|) &\leq \mathbb{P}(y_1 > |z(\alpha/n)|) + \mathbb{P}(\max_{i>1} y_i > |z(\alpha/n)|) \\
&\leq \mathbb{P}(z_1 > |z(\alpha/n)|) - \mu^{(n)}) + \mathbb{P}(\max_{i>1} z_i > |z(\alpha/n)|) \\
&\to 0 + q(\alpha) \\
&\approx \alpha.
\end{aligned}$$

This is a bad test because we can obtain the same level and power by flipping a biased coin that rejects α of the time.

Conclusion: We effectively see that $\sqrt{2\log n}$ constitutes a sharp detection threshold. When $\frac{\mu^{(n)}}{\sqrt{2\log n}} = 1+\varepsilon$, we always detect the needle $\mu_1 > 0$. We can even achieve

$$\mathbb{P}_{H_0}(\text{Type I Error}) \to 0 \quad \text{and} \quad \mathbb{P}_{H_1}(\text{Type II Error}) \to 0$$

if we use $\sqrt{2\log n}$ instead of $z(\alpha/n)$ as our threshold. In other words, asymptotically we make no mistakes.

However, when $\frac{\mu^{(n)}}{\sqrt{2\log n}} = 1-\varepsilon$, with $q(\alpha) = 1-e^{-\alpha} \approx \alpha$ being the asymptotic size, Bonferroni's global test gives

$$\mathbb{P}(\text{Type I Error}) \to q(\alpha) \quad \text{and} \quad \mathbb{P}(\text{Type II Error}) \to 1-q(\alpha),$$

that is, it does no better than flipping a coin.

Can we do better than Bonferroni? When $\mu^{(n)} = (1-\varepsilon)\sqrt{2\log n}$, we saw that, roughly,

$$P_{H_0}(\text{Type I Error}) \le \alpha \quad \text{and} \quad P_{H_1}(\text{Type II Error}) \ge 1-\alpha,$$

so we are doing no better than flipping a biased coin that disregards the actual data. This is in fact true for *any* test in this scenario. To see this, we first reduce our composite hypothesis to a simple one, and then we show that even the optimal test given by the Neyman-Pearson Lemma does no better than flipping a coin.

4 Optimality of Detection Threshold

"Bayesian" Decision Problem: Consider

$$\begin{aligned} &H_0 : \mu_i = 0 \text{ for all } i \\ &H_1 : \{\mu_i\} \sim \pi \end{aligned}$$

where π selects a coordinate I uniformly and sets $\mu_I = \mu$, with all other $\mu_i = 0$.

This setup differs from the previous problem in the important respect that H_0 and H_1 are both *simple hypotheses* and we can now apply Neyman-Pearson.

The optimal test rejects for large values of the likelihood ratio. The densities under the null and the alternative are given by

$$\begin{aligned} f_0(y) &= \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_j^2}, \\ f_1(y) &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_i-\mu)^2} \prod_{j:j\neq i} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_j^2}. \end{aligned}$$

After cancellations, the likelihood ratio is given by

$$L = \frac{f_1}{f_0} = \frac{1}{n}\sum_{i=1}^{n} e^{y_i\mu - \frac{1}{2}\mu^2}.$$

Properties of L under H_0: Writing $X_i = e^{y_i\mu - \frac{1}{2}\mu^2}$, we have that under H_0 the X_i are iid and

$$L = \frac{1}{n}\sum_{i=1}^{n} X_i;$$

this is a sample average with mean $\mathbb{E}X_1$ and variance $\frac{1}{n}\mathrm{Var}X_1$.

First impulse: We would like to apply the CLT; however, because μ is not fixed but rather $\mu^{(n)} = (1-\varepsilon)\sqrt{2\log n} \to \infty$, we would need a triangular array argument.

The (sufficient but not necessary) Lyapunov condition, for instance, is violated for $q = 3$:

$$\frac{1}{[\sum_i \mathrm{Var}(X_i)]^{3/2}}\sum_i \mathbb{E}|X_i|^3 \to \infty$$

as $n \to \infty$. We shall, therefore, focus on deriving a weaker result.

Proposition 1. *If $\mu = (1-\varepsilon)\sqrt{2\log n}$, then $L \xrightarrow{p} 1$.*

Proof. Proof provided at the end of the notes. $\square$

This already hints at the fact that the likelihood test cannot do very well. But before we formally prove this (this will not be done in this lecture), we skip to the punchline.

Proposition 2. *Set threshold $T_n(\alpha)$ such that $\mathbb{P}_0(L \geq T_n(\alpha)) = \alpha$. Then*

$$\lim_{n\to\infty} \mathbb{P}(\text{Type II error}) = 1 - \alpha.$$

Proof. Note that

$$\begin{aligned}
\mathbb{P}(\text{Type II Error}) = \mathbb{P}_1(L \leq T_n(\alpha)) &= \int \mathbf{1}_{\{L\leq T_n(\alpha)\}}\, dP_1 \\
&= \int \mathbf{1}_{\{L\leq T_n(\alpha)\}} L\, dP_0 \\
&= \int \mathbf{1}_{\{L\leq T_n(\alpha)\}}\, dP_0 + \int \mathbf{1}_{\{L\leq T_n(\alpha)\}}(L-1)\, dP_0 \\
&= (1-\alpha) + \int \mathbf{1}_{\{L\leq T_n(\alpha)\}}(L-1)\, dP_0 \\
&\approx (1-\alpha).
\end{aligned}$$

The last claim follows from the fact that $L \xrightarrow{p} 1$. We can make this rigorous as follows: let $Z_n = \mathbf{1}_{\{L\leq T_n(\alpha)\}}(L-1)$. First, $Z_n \xrightarrow{p} 0$. Second, Because $L \xrightarrow{p} 1$, $T_n(\alpha)$ is uniformly bounded, and hence so is Z_n. The bounded convergence theorem [1, Section 13.6] then gives that $\mathbb{E}|Z_n| \to 0$ (this is a simple result that can be checked by hand).

$\square$

Conclusion: If $\mu^{(n)} = (1-\varepsilon)\sqrt{2\log n}$, then the *optimal test* has

$$\mathbb{P}(\text{Type I Error}) + \mathbb{P}(\text{Type II Error}) \to 1.$$

Broad Conclusion: Let's think back to the original problem, with $H_1 : \mu_i > 0$ for *one* i, a composite of n alternatives.

We have shown today that the *average* type II error (Bayes risk) of any level-α procedure is no better than $1-\alpha$, from which it of course follows that the worst-case error (minimax risk) is no better either: i.e. for any test

$$\liminf \left[\mathbb{P}_{H_0}(\text{Type I Error}) + \sup_{H_1} \mathbb{P}(\text{Type II Error})\right] \geq 1$$

where the sup is taken over all alternatives in which one coordinate has mean $\mu^{(n)} = (1-\epsilon)\sqrt{2\log n}$.

In this regime, the Bonferroni procedure is optimal for testing the global null. Asymptotically, it is able to perfectly differentiate between the null and alternative hypothesis when $\mu^{(n)}$ is larger than the $\sqrt{2\log n}$ threshold, and we have just shown that *no test* is able to do better in minimax risk than a coin flip when $\mu^{(n)}$ is smaller than the $\sqrt{2\log n}$ threshold.

5 Proof of the proposition 1

Recall the statement of Proposition 1: If $\mu = (1-\varepsilon)\sqrt{2\log n}$, then $L \xrightarrow{p} 1$.

Proof. Recall

$$L = \frac{1}{n}\sum_{i=1}^{n} X_i$$

with $X_i = e^{y_i\mu - \frac{1}{2}\mu^2}$ iid.

Assume first $0 < \epsilon < 1/2$, take $T_n = \sqrt{2\log n}$, and write

$$\tilde{L} = \frac{1}{n}\sum_{i=1}^{n} X_i\, \mathbf{1}_{\{y_i \leq T_n\}}.$$

We have

$$\mathbb{P}(\tilde{L} \neq L) \leq \mathbb{P}(\max y_i \geq T_n) \to 0,$$

and it suffices to establish that

$$\tilde{L} = \Phi(\varepsilon\sqrt{2\log n}) + o_{P_0}(1)$$

which in particular follows if

1. $\mathbb{E}_0(\tilde{L}) = \Phi(\varepsilon\sqrt{2\log n})$
2. $\mathrm{Var}_0(\tilde{L}) = o(1)$

Proceeding,

$$\begin{aligned}
\mathbb{E}_0(\tilde{L}) = \mathbb{E}_0\left[X_1 \mathbf{1}_{\{y_1 \le T_n\}}\right] &= \int_{-\infty}^{T_n} e^{\mu z - \mu^2/2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\
&= \int_{-\infty}^{T_n} \frac{1}{\sqrt{2\pi}} e^{-(z-\mu)^2/2}\, dz \\
&= \Phi(T_n - \mu) \\
&= \Phi(\varepsilon\sqrt{2\log n}).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathrm{Var}_0(\tilde{L}) = \frac{1}{n}\mathrm{Var}\left(X_1 \mathbf{1}_{\{y_1 \le T_n\}}\right) \le \frac{1}{n}\mathbb{E}_0\left[X_1^2 \mathbf{1}_{\{y_1 \le T_n\}}\right] &= \frac{1}{n}\int_{-\infty}^{T_n} e^{-\mu^2} e^{2\mu z} \phi(z)\, dz \\
&= \frac{1}{n} e^{\mu^2} \Phi(T_n - 2\mu).
\end{aligned}$$

Since $\Phi(T_n - 2\mu) \le \phi(2\mu - T_n)$, this gives

$$\begin{aligned}
\mathrm{Var}_0(\tilde{L}) \le \frac{1}{n} e^{\mu^2} \phi(2\mu - T_n) &= \frac{1}{n} e^{(1-\varepsilon)^2 T_n^2} \frac{1}{\sqrt{2\pi}} e^{-(1-2\varepsilon)^2 T_n^2/2} \\
&= \frac{1}{\sqrt{2\pi} n} e^{(1-2\varepsilon^2)T_n^2/2} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\varepsilon^2 T_n^2} \\
&\to 0.
\end{aligned}$$

This proves the result for $0 < \epsilon < 1/2$. The claim for $1 > \epsilon > 1/2$ is even simpler since $\exp(\mu^2)/n$ converges to zero in this case. $\square$

References

[1] D. Williams. *Probability with martingales*, Cambridge University Press, 1991.