Stats 300C: Theory of Statistics Spring 2018

Lecture 3 — April 06, 2018

Prof. Emmanuel Candes

Scribe: Will Fithian (2012) and E. Candes; edited by Kenneth Tay

1 Outline

Agenda: Global Testing

1. χ^2 Test
2. Detection Thresholds for Small Distributed Effects
3. Comparison of Bonferroni's and χ^2 tests

Global Testing: Recall our independent Gaussian sequence model

$$y_i = \mu_i + z_i,$$

where $z_i \overset{\text{i.i.d.}}{\sim} N(0,1)$, $1 \leq i \leq n$. In vector notation, we can write this as

$$y \sim N(\mu, I)$$

We are testing:

$$\begin{aligned} H_0 &: \mu = 0 \\ H_1 &: \text{at least one } \mu_i \neq 0 \end{aligned}$$

Variation: One-Way Layout:

$$y_i = \tau + \mu_i + z_i,$$

where τ is the grand mean and μ_i are the individual differences. (For identifiability, we usually require $\sum \mu_i = 0$.) Then, H_0 is the hypothesis that all means (treatments) are the same, while H_1 is the hypothesis that at least one is different.

Global Test Statistic: Consider the first model above. A natural test is to reject H_0 if $||y||^2$ is large. In the variation, we would reject if $\sum(y_i - \bar{y})^2$ is large. (If we didn't know the variance σ^2 of y_i, we could estimate it and use an F test.) All of these tests would exhibit similar qualitative behavior.

Goal: Understand when this test is effective.

Note: We saw that the Bonferroni procedure is in some sense as good as it gets for alternatives with only one $\mu_i \neq 0$. We will see that the test above is "optimal" in some some against a different class of alternatives.

2 χ^2 Test

The test statistic for the χ^2 test is

$$T = \sum_{i=1}^{n} y_i^2 = \|y\|^2.$$

Under H_0: $T \sim \chi_n^2$. Thus, the level-α test rejects H_0 when $T > \chi_n^2(1-\alpha)$.

Note that under H_0,

$$T = \sum_{i=1}^{n} z_i^2,$$

with $\mathbb{E}(z_1^2) = 1$ and $\mathrm{Var}(z_1^2) = 2$. Hence, by a CLT approximation, for large n we roughly have

$$\frac{T-n}{\sqrt{2n}} \sim N(0,1),$$

implying that

$$\chi_n^2(1-\alpha) \approx n + \sqrt{2n}z(1-\alpha).$$

Under H_1: T is a non-central χ^2. Here,

$$T = \sum_{i=1}^{n} (\mu_i + z_i)^2, \qquad \begin{aligned} \mathbb{E}\left[(\mu_i + z_i)^2\right] &= \mu_i^2 + 1, \\ \mathrm{Var}\left[(\mu_i + z_i)^2\right] &= 4\mu_i^2 + 2. \end{aligned}$$

Again, for large n we have an approximate normal distribution with

$$\frac{T - (n + \|\mu\|^2)}{\sqrt{2n + 4\|\mu\|^2}} \sim N(0,1).$$

Summary: If we let

$$Z = \frac{T-n}{\sqrt{2n}}$$

be the normalized version of the test statistic and define

$$\theta = \frac{\|\mu\|^2}{\sqrt{2n}}$$

which is, in a sense, the signal to noise ratio (SNR), then we roughly have:

$$\begin{aligned} H_0 &: \ Z \sim N(0,1) \\ H_1 &: \ Z \sim N\left(\theta, 1 + \frac{\theta}{\sqrt{n/8}}\right) \end{aligned}$$

Therefore, the test is easy when $\theta \gg 1$, and hard when $\theta \ll 1$. (E.g. when $\theta = 2$, the power of the test is roughly $P(\mathcal{N}(0,1) > 1.65 - 2) \approx 66\%$.) In other words, the power of the χ^2 test is determined by the relative size of $\|\mu\|^2$ compared to $\sqrt{n}$.

SNR: If we had started with a model in which the noise variance is σ^2 as in

$$y_i = \mu_i + \sigma z_i, \qquad i = 1, \dots, n \tag{1}$$

where the z_i's are as before, then we would see that the detection power depends sensitively on

$$\theta = \sqrt{\frac{n}{2}} \cdot \frac{\|\mu\|^2}{\sigma^2 n}$$

This is because the model (1) is equivalent to $y_i = \mu_i/\sigma + z_i$, $i = 1, \dots, n$. Therefore, if we define the SNR as

$$\text{SNR} = \frac{\text{total signal power}}{\text{total expected noise power}} = \frac{\|\mu\|^2}{\sigma^2 n},$$

we can see that

$$\theta \propto \text{SNR}$$

with a constant of proportionality equal to $\sqrt{n/2}$. We now assume $\sigma = 1$ without loss of generality.

A natural question arises: **when $\theta \ll 1$, is there a test that does better than the χ^2 test?** To show that the answer is no, we use the strategy employed last lecture to show the optimality of the Bonferroni test: introduce a simpler "Bayesian" decision problem, and show that even in this setting, the optimal test given by the Neyman-Pearson Lemma is powerless.

Bayesian Problem:

$$\begin{aligned} H_0 &: \mu = 0 \\ H_1 &: \mu \sim \pi_\rho \end{aligned}$$

where π_ρ distributes mass uniformly on the sphere of radius ρ.

Likelihood ratio: We introduce some notation: let $\mu = \rho u$, where u is uniformly distributed on the unit sphere, and let π is the uniform distribution on the sphere. We have

$$L = \int_{S^{n-1}} \frac{e^{-\frac{1}{2}||y - \rho u||^2}}{e^{-\frac{1}{2}||y||^2}} \, \pi(du) = \int_{S^{n-1}} e^{-\frac{1}{2}\rho^2 + \rho u^T y} \, \pi(du).$$

We will show that if $\theta_n = \frac{\rho^2}{\sqrt{2n}} \to 0$ as $n \to \infty$, then $\text{Var}_0(L) \to 0$. Because $\mathbb{E}_0(L) = 1$, we have that $L \xrightarrow{p} 1$.

As in last lecture, this implies that $\mathbb{P}_1(\text{Type II Error}) = \mathbb{E}_0(\mathbf{1}_{\{L \le T_n\}} L) \to 1 - \alpha$, i.e. we can do no better than a coin toss (we have no power).

Useful Relationship: If $y \sim N(0, I)$, then

$$\mathbb{E}\left(e^{a^T y}\right) = e^{||a||^2/2},$$

which is the mgf of a Gaussian random vector. Then

$$\begin{aligned}
\mathbb{E}_0(L^2) &= \mathbb{E}_0\left[\int\int e^{-\rho^2/2+\rho u^T y} e^{-\rho^2/2+\rho v^T y}\,\pi(du)\,\pi(dv)\right] \\
&= \mathbb{E}_0\left[\int\int e^{-\rho^2+\rho(u+v)^T y}\,\pi(du)\,\pi(dv)\right] \\
&= e^{-\rho^2}\int\int e^{\rho^2||u+v||^2/2}\,\pi(du)\,\pi(dv) \\
&= \int\int e^{\rho^2 u^T v}\,\pi(du)\,\pi(dv),
\end{aligned}$$

where the third equality uses the mgf and the fourth uses $u^T u = v^T v = 1$. By spherical symmetry, we can fix $v = e_1 = (1, 0, \dots, 0)$ to obtain

$$\mathbb{E}_0(L^2) = \int e^{\rho^2 u_1}\,\pi(du),$$

with $u = (u_1, \dots, u_n)$ uniform on S^{n-1}. Using the Taylor approximation

$$e^{\rho^2 u_1} = 1 + \rho^2 u_1 + \frac{\rho^4 u_1^2}{2} + \cdots,$$

we have

$$\begin{aligned}
\mathbb{E}e^{\rho^2 u_1} &= 1 + \mathbb{E}[\rho^2 u_1] + \mathbb{E}\left[\frac{\rho^4 u_1^2}{2}\right] + \cdots \\
&= 1 + 0 + \frac{\rho^4}{2n} + 0 + O\left(\frac{\rho^8}{n^2}\right),
\end{aligned}$$

which is to say

$$\mathbb{E}_0 L^2 = 1 + \theta_n^2 + O(\theta_n^4) \to 1$$

when $\theta_n = \frac{\rho^2}{\sqrt{2n}} \to 0$.

Conclusion: The LR test has *no power* if $\dfrac{||\mu||^2}{\sqrt{2n}} \to 0$ as $n \to \infty$.

3 Comparison between Bonferroni's and χ^2 tests

The regimes in which Bonferroni and χ^2 are effective are completely different.

Example 1: $n^{1/4}$ of the μ_i's are equal to $\sqrt{2\log n}$. (E.g. when $n = 10^6$, $n^{1/4} \approx 32$ and $\sqrt{2\log n} \approx 5.3$.) In this set-up, the Bonferroni test has full power, but because

$$\theta_n = \frac{n^{1/4} 2\log n}{\sqrt{2n}} \to 0,$$

the χ^2 test has no power.

Example 2: $\sqrt{2n}$ of the μ_i's are equal to 3. The χ^2 test has (almost) full power. The Bonferroni test has no power, because when n is large (large number of tests) it's very likely that the smallest p-value comes from a null μ_i, not a true signal. An intuitive argument is as follows: among the nulls, the largest y_i has size $\approx \sqrt{2\log n}$ while among the true signals, the largest y_i has size $\approx 3 + \sqrt{2\log\sqrt{2n}}$. If n is large, the former value is larger.

We can summarize our conclusions thus far in a table:

	Small, distributed effects	**Few strong effects**
ANOVA	Powerful	Weak
Bonferroni	Weak	Powerful

Numerical illustration: Let $n = 10^6$ and $\alpha = 0.05$, and consider Bonferroni's, χ^2 and Fisher's combination global tests for the following alternatives :

- **Sparse strong effects**: μ_i is the same as the Bonferroni Threshold($|z(\alpha/(2n))| \doteq 5.45$) for $1 \le i \le 4$ and 0 otherwise.
- **Distributed weak effects**: μ_i is 1.1 for $1 \le i \le k = 2400$ and 0 otherwise.

In the sparse setting, the power Bonferroni's method can be approximated as follows :

$$1 - \mathbb{P}_{H_1}(\max|y_i| \le |z(\alpha/(2n))|) \approx 1 - (\mathbb{P}(|y_1| \le \mu_1))^4 \approx 1 - 1/16 = 0.9375$$

On the other hand, χ^2 (and similarly Fisher) would be almost powerless, as $\theta = ||\mu||^2/\sqrt{2n} \doteq 0.084 \ll 1$.

A numerical estimate of the power for these tests with 500 trials is as expected:

$$\text{Bonferroni} = 95.0\%$$
$$\text{Chi-sq} = 5.6\%$$
$$\text{Fisher} = 6.0\%.$$

For the other alternative, the power of Bonferroni is roughly

$$\mathbb{P}_{H_1}(\max|y_i| > |z(\alpha/(2n))|) \le \mathbb{P}(\max_{i\le k}|y_i| > |z(\alpha/(2n))|) + \mathbb{P}(\max_{i>k}|z_i| > |z(\alpha/(2n))|) \approx 0.066.$$

Also $\theta = ||\mu||^2/\sqrt{2n} \doteq 2.05$. Hence, Bonferroni has almost not power while the χ^2 and Fisher's test should have significant power. Numerically,

$$\text{Bonferroni} = 6.0\%$$
$$\text{Chi-sq} = 68.8\%$$
$$\text{Fisher} = 63.4\%.$$

Next week: Can we introduce a method that has the best of both worlds? Is there a single statistic which is powerful whichever world we are actually in?