1 Outline

Agenda: Global Testing

1. $\chi^2$ Test
2. Detection Thresholds for Small Distributed Effects
3. Comparison of Bonferroni’s and $\chi^2$ tests

Global Testing: Recall our independent Gaussian sequence model

$$y_i = \mu_i + z_i,$$

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

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

We are testing:

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

Variation: One-Way Layout:

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

where $\tau$ is the grand mean and $\mu_i$ 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 $\sigma^2$ 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 sense against a different class of alternatives.

2 \( \chi^2 \) Test

The test statistic for the \( \chi^2 \) test is

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

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

Note that under \( H_0 \),

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

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

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

implying that

\[
\chi^2_n(1 - \alpha) \approx n + \sqrt{2n}z(1 - \alpha).
\]

Under \( H_1 \): \( T \) is a non-central \( \chi^2 \). Here,

\[
T = \sum_{i=1}^{n} (\mu_i + z_i)^2, \quad \mathbb{E}[(\mu_i + z_i)^2] = \mu_i^2 + 1, \quad \text{Var}[(\mu_i + z_i)^2] = 4\mu_i^2 + 2.
\]

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:

\[
H_0 : Z \sim N(0, 1)
\]

\[
H_1 : Z \sim N\left(\theta, 1 + \frac{\theta}{\sqrt{n/8}}\right)
\]
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 $P(N(0, 1) > 1.65 - 2) \approx 66\%$.) In other words, the power of the $\chi^2$ test is determined by the relative size of $\parallel \mu \parallel^2$ compared to $\sqrt{n}$.

**SNR:** If we had started with a model in which the noise variance is $\sigma^2$ as in

$$y_i = \mu_i + \sigma z_i, \quad i = 1, \ldots, n$$

(1)

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

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

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

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

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 $\chi^2$ 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:**

$$H_0 : \mu = 0$$

$$H_1 : \mu \sim \pi_\rho$$

where $\pi_\rho$ distributes mass uniformly on the sphere of radius $\rho$.

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

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

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 $E_0(L) = 1$, we have that $L \xrightarrow{P} 1$.

As in last lecture, this implies that $\mathbb{P}_1(\text{Type II Error}) = \mathbb{E}_0(1_{\{L \leq T_n\}}) \to 1 - \alpha$, i.e. we can do no better than a coin toss (we have no power).
Useful Relationship: If $y \sim N(0, I)$, then 
\[
\mathbb{E}(e^{a^T y}) = e^{||a||^2/2},
\]
which is the mgf of a Gaussian random vector. Then
\[
\mathbb{E}_0(L^2) = \mathbb{E}_0 \left[ \int \int e^{-\rho^2/2 + \rho u^T y} e^{-\rho^2/2 + \rho v^T y} \pi(du) \pi(dv) \right]
= \mathbb{E}_0 \left[ \int \int e^{-\rho^2 + \rho(u+v)^T y} \pi(du) \pi(dv) \right]
= e^{-\rho^2} \int \int e^{\rho^2 ||u+v||^2/2} \pi(du) \pi(dv)
= \int \int e^{\rho^2 u^T v} \pi(du) \pi(dv),
\]
where the third equality uses the mgf and the fourth uses $u^T u = v^T v = 1$. By spherical symmetry, we can fix $v = e_1 = (1, 0, \ldots, 0)$ to obtain
\[
\mathbb{E}_0(L^2) = \int e^{\rho^2 u_1} \pi(du),
\]
with $u = (u_1, \ldots, u_n)$ 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
\[
\mathbb{E}e^{\rho^2 u_1} = 1 + \mathbb{E}[\rho^2 u_1] + \mathbb{E}\left[\frac{\rho^4 u_1^2}{2}\right] + \cdots
= 1 + 0 + \frac{\rho^4}{2n} + 0 + O\left(\frac{\rho^8}{n^2}\right),
\]
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 $\frac{||\mu||^2}{\sqrt{2n}} \to 0$ as $n \to \infty$.

3 Comparison between Bonferroni’s and $\chi^2$ tests

The regimes in which Bonferroni and $\chi^2$ are effective are completely different.
**Example 1:** \( n^{1/4} \) of the \( \mu_i \)'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

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

the \( \chi^2 \) test has no power.

**Example 2:** \( \sqrt{2n} \) of the \( \mu_i \)'s are equal to 3. The \( \chi^2 \) 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 \( \mu_i \), 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:

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>Small, distributed effects</th>
<th>Few strong effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powerful</td>
<td>Weak</td>
<td></td>
</tr>
</tbody>
</table>

**Numerical illustration:** Let \( n = 10^6 \) and \( \alpha = 0.05 \), and consider Bonferroni’s, \( \chi^2 \) and Fisher’s combination global tests for the following alternatives:

- **Sparse strong effects:** \( \mu_i \) is the same as the Bonferroni Threshold(|\( z(\alpha/(2n)) | \approx 5.45 \)) for \( 1 \leq i \leq 4 \) and 0 otherwise.

- **Distributed weak effects:** \( \mu_i \) is 1.1 for \( 1 \leq i \leq k = 2400 \) and 0 otherwise.

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

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

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

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

- Bonferroni = 95.0%
- Chi-sq = 5.6%
- Fisher = 6.0%

For the other alternative, the power of Bonferroni is roughly

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

Also \( \theta = ||\mu||^2/\sqrt{2n} \approx 2.05 \). Hence, Bonferroni has almost not power while the \( \chi^2 \) and Fisher’s test should have significant power. Numerically,

- Bonferroni = 6.0%
- Chi-sq = 68.8%
- 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?