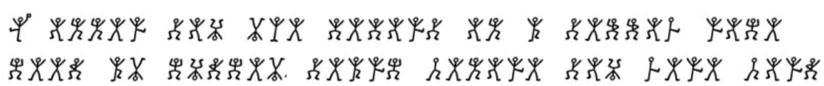
Lecture 9. Dectection Threshold

[1]*School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen)*

1. Optimality of Detection Threshold



For the independent Gaussian sequence model

$$Y_i \sim_{indep.} N(\mu_i, 1), \quad i = 1, \cdots, n.$$

Consider a "Bayesian" decision/testing problem

$$H_0 : \mu_i = 0,\ i = 1, \cdots, n,\ \ v.s.\ \ H_1 : \mu_I = \mu,\ \mu_i = 0,\ i \in \{1, \cdots, n\} \backslash I. \quad (1.1)$$

where I is uniformly distributed on $\{1, \cdots, n\}$. Or we can formulate this decision/testing problem as

$$H_0 : \mu_i = 0,\ i = 1, \cdots, n,\ \ v.s.\ \ H_1 : \{\mu_i\}_{1 \le i \le n} \sim \pi.$$

where π is the distribution who selects a coordinate I uniformly and sets $\mu_I = \mu$ and $\mu_i = 0$ for all other $i \neq I$.

As complicated as this multiple hypothesis testing problem may seems, this setup differs from the previous problem in the important respect that H_0 and H_1 are both simple hypotheses and therefore we can directly apply the Neyman-Pearson Lemma. The UMP test rejects for large values of the likelihood ratio and the densities under the null and the alternative are given by

$$f_0(y) = \prod_{j=1}^{n} \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y_j^2} \right]$$
$$f_1(y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_i - \mu)^2} \prod_{j=1, j \neq i}^{n} \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y_j^2} \right],$$

which leads to the likelihood ratio

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

Since $\{Y_1, \cdots, Y_n\}$ are i.i.d under the null hypothesis, so L would "ideally" converge to a constant according to law of large numbers.

Recall the powerlessness threshold $\mu(n) = (1-\epsilon)\sqrt{2\log n}$ in Bonferroni case. Here we also concentrate on this threshold and investigate the asymptotic behavior of likelihood ratio statistic when taking this $\mu = \mu(n)$. But because $\mu = \mu(n)$ depends on n, so we have to treat it more carefully such as we may need a triangular array argument. For instance, when we check the Lyapunov condition for $q = 3$, we have

$$\frac{1}{\left(\sum_{i=1}^n \mathrm{Var}\left(e^{Y_i\mu-\mu^2/2}\right)\right)^{3/2}} \sum_{i=1}^n \mathbb{E}|e^{Y_i\mu-\mu^2/2}|^3 \to \infty.$$

Therefore, we may target to derive a weaker result compare to asymptotic distribution.

Proposition 1.1. *If $\mu = \mu(n) = (1-\epsilon)\sqrt{2\log n}$, then $L \xrightarrow{p} 1$.*

Proof. Recall that

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

with $X_i = e^{Y_i\mu - \frac{1}{2}\mu^2}$ iid. Assume first $0 < \epsilon < \frac{1}{2}$, take $T_n = \sqrt{2\log n}$, and write

$$\tilde{L} = \frac{1}{n}\sum_{i=1}^n X_i \mathbf{1}_{\{Y_i \le T_n\}},$$

We have

$$P(\tilde{L} \ne L) \le P(\max Y_i > T_n) \to 0,$$

and it suffices to establish that

$$\tilde{L} = \epsilon\sqrt{2\log n} + o_P(1),$$

which, in particular, follows if

1. $E_0(\tilde{L}) = \Phi(\epsilon\sqrt{2\log n})$,
2. $\mathrm{Var}_0(\tilde{L}) = o(1)$.

Proceeding,

$$\begin{aligned}
\mathbb{E}_0(\tilde{L}) &= \mathbb{E}_0\left(\frac{1}{n}\sum_{i=1}^n X_i \mathbf{1}_{\{Y_i \le T_n\}}\right) \\
&= \int_{-\infty}^{T_n} e^{\mu z - \mu^2/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz \\
&= \int_{-\infty}^{T_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2}}\,dz
\end{aligned}$$

$$= \Phi(T_n - \mu) = \Phi(\epsilon\sqrt{2\log n}),$$

where Φ is the cumulative distribution function of the standard normal distribution. Furthermore,

$$\begin{aligned}\mathrm{Var}_0(\tilde{L}) &= \frac{1}{n}\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \mathbf{1}_{\{Y_i \le T_n\}}\right) \\ &\le \frac{1}{n}\mathbb{E}_0(X_1^2\mathbf{1}_{\{Y_1\le T_n\}}) = \frac{1}{n}\int_{-\infty}^{T_n} e^{-\mu^2+2\mu z}\phi(z)dz = \frac{1}{n}e^{\mu^2}\Phi(T_n - 2\mu)\end{aligned}$$

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

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

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

Proposition 1.2. *Set threshold $T_n(\alpha)$ such that $P_0(L \ge T_n(\alpha)) = \alpha$. Then for the likelihood ratio test,*

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

Proof.

$$\begin{aligned}P(\text{Type II Error}) &= P_1(L \ge T_n(\alpha)) = \int \mathbb{1}(L \le T_n(\alpha))dP_1 \\ &= \int \mathbb{1}(L \le T_n(\alpha))LdP_0 \\ &= \int \mathbb{1}(L \le T_n(\alpha))dP_0 + \int \mathbb{1}(L \le T_n(\alpha))(L-1)dP_0 \\ &= (1-\alpha) + \int \mathbb{1}(L \le T_n(\alpha))(L-1)dP_0 \\ &\approx (1-\alpha).\end{aligned}$$

Therefore,

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

□

The last claim follows from the fact that $L \xrightarrow{p} 1$. We can make this rigorous as follows: let $Z_n = \mathbf{1}_{\{L \le T_n(\alpha)\}}(L-1)$. 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 then gives that $\mathbb{E}[|Z_n|] \to 0$.

Conclusion: If $\mu^{(n)} = (1-\epsilon)\sqrt{2\log n}$, 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 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. That is, for any test

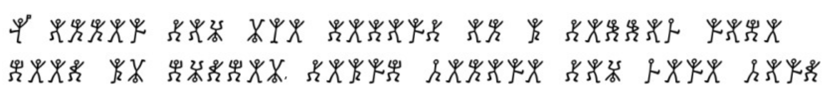
$$\liminf_{n\to\infty} \left[\mathbb{P}_{H_0}(\text{Type I Error}) + \sup_{H_1} \mathbb{P}(\text{Type II Error})\right] \geq 1,$$

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.

2. Global Testing



Global Testing: Recall our independent Gaussian sequence model

$$y_i = \mu_i + z_i,$$

where z_i are i.i.d. $N(0,1)$ for $1 \leq i \leq n$. In vector notation, we can write this as $y \sim N(\theta, I)$.

We are testing:

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

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.

Notice that, 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.

2.1. 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, we have $T \approx \chi_n^2$. Thus, the level-α test rejects H_0 when $T > \chi_n^2(1-\alpha)$, where $\chi_n^2(1-\alpha)$ is the $(1-\alpha)-$th quantile of χ^2 distribution with degree of freedom n.

Note that under H_0,

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

with $\mathbb{E}(z_1) = 1$ 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

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

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

$$T = \sum_{i=1}^{n} (\mu_i + z_i)^2,$$

with $\mathbb{E}[(\mu_i + z_i)^2] = \mu_i^2 + 1$, and $\mathrm{Var}[(\mu_i + z_i)^2] = 4\mu_i^2 + 2$. Again, for large n we have an approximate normal distribution with

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

To summary up, If we let

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

be the normalized version of the test statistic and define

$$\theta = \frac{\sum_{i=1}^{n} \mu_i^2}{\sqrt{2n}}$$

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

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

Therefore, the test is easy when $\theta \ll 1$, and hard when $\theta \gg 1$. (For example, 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 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

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

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

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

This is because the model is equivalent to $y_i = \frac{\mu_i}{\sigma} + z_i$, $i = 1, \dots, n$. 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 $\theta \propto$ SNR with a constant of proportionality equal to $\sqrt{n/2}$. So 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 same strategy as before 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.

2.2. 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.

Notice this is a simple hypothesis problem, therefore, to apply Neyman-Pearson Lemma, we look at the likelihood ratio statistic. But before that, we introduce some notation: let $\mu = \rho u$, where u is uniformly distributed on the unit sphere, and let π be 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 $\mathrm{Var}_0(L) \to 0$. Because $\mathbb{E}_0(L) = 1$, we have that $L \xrightarrow{p} 1$, this implies that $\mathbb{P}_1$ (Type II Error) $= \mathbb{E}_0(1_{\{L \le T_n\}} L) \to 1 - \alpha$, i.e. we can do no better than a coin toss (we have no power).

One of the useful relationship worth mentioning is, if $y \sim N(0,1)$, then

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

which is the moment generating function 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 form of mgf. and the fourth uses the fact that $u^T u = v^T v = 1$. By spherical symmetry, we can fix $v = e_1 = (1, 0, \cdots, 0)$ to obtain

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

with $u = (u_1, \cdots, u_n)$ 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 $\theta_n = \frac{\rho^2}{\sqrt{2n}} \to 0$. Thus proving that the likelihood ratio test has no power if $\frac{||\mu||^2}{\sqrt{2n}} \to 0$ 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* 3.1. Consider $n^{1/4}$ of the μ_i's are equal to $\sqrt{2logn}$. (E.g. when $n = 10^6, n^{1/4} \approx 32$ and $\sqrt{2logn} \approx 5.3$.) 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* 3.2. Consider $\sqrt{2n}$ of the μ_i's are equal to 3. The χ^2 test has almost full power. Meanwhile, 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 these nulls, the largest y_i has size $\approx \sqrt{2logn}$ while among the true signals, the largest y_i has size $\approx 3 + \sqrt{2log\sqrt{2n}}$. 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
χ^2 test	Powerful	Weak
Bonferroni test	Weak	Powerful

We present some further 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,i.e., $(|z(\alpha/(2n))| = 5.45)$ 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 of 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, χ^2 test (and similarly Fisher's test) would be almost powerless, as $\theta = ||\mu||^2/\sqrt{2n} = 0.084 \ll 1$.

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

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

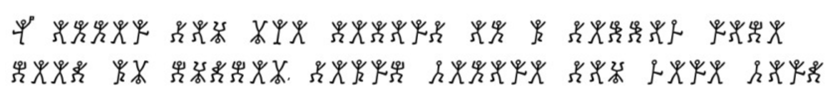
For the other alternative, the power of Bonferroni is roughly

$$\begin{aligned} \mathbb{P}_{H_1}(\max |y_i| > |z(\alpha/(2n))|) &\leq \mathbb{P}(\max_{i \leq k} |y_i| > |z(\alpha/(2n))|) \\ &\quad + \mathbb{P}(\max_{i > k} |z_i| > |z(\alpha/(2n))|) \approx 0.066. \end{aligned}$$

Also $\theta = ||\mu||^2/\sqrt{2n} = 2.05$. Hence, Bonferroni's approach has almost no power while the χ^2 test and Fisher's test should have significant power. Numerically,

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

4. Comparison of Bonferroni and other Global Tests



Recall Simes procedure, which rejects the global null when

$$\min_{1 \leq i \leq n} \{p_{(i)} \cdot n/i\} \leq \alpha.$$

The Simes procedure is strictly less conservative than Bonferroni.

• *Example* 4.1. Consider $n = 2$, thus Simes' test rejects if $p_{(1)} \leq \alpha/2$ or $p_{(2)} \leq \alpha$. Below we plot the rejection regions of each test in Figure.1. We can easily check that in this case, the size of Bonferroni is $\alpha - \alpha^2/4$, while the size of Simes is α. Nevertheless, Simes still tends to look at lower p-values, since higher p-values are unlikely to be less or equal to $\alpha \frac{i}{n}$.

Theorem 4.2. *Under H_0 and independence of the p_i, the Simes test statistics $T_n \sim U(0,1)$, and thus the Simes test rejects H_0 if $T_n \leq \alpha$.*

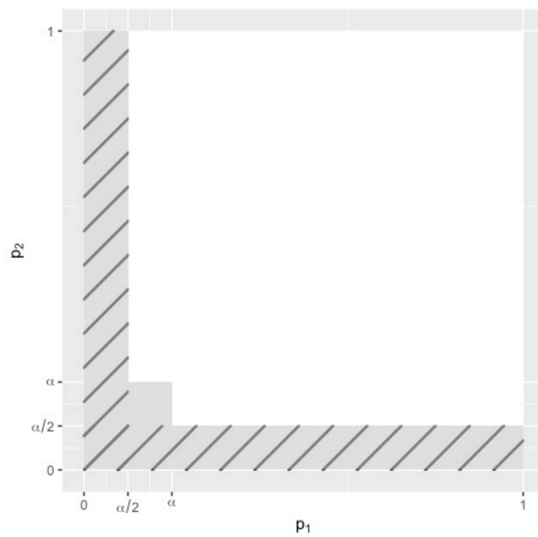


Figure 1: Striped area is the rejection region of Bonferroni approach and shadow areea is the rejection region of Simes approach.

Proof. We adopt the mathematical induction approach. Clearly the theorem results hold for $n = 1$. Assume that it is true for $n-1$, i.e., $T_{n-1} \sim U(0, 1)$. Notice that the density of $p_{(n)}$ is

$$f(t) = nt^{n-1}$$

for $t \in [0, 1]$. Then

$$\begin{aligned}\mathbb{P}(T_n \leq \alpha) &= \int_0^1 \mathbb{P}(T_n \leq \alpha | p_{(n)} = t) f(t) dt \\ &= \int_0^\alpha f(t)dt + \int_\alpha^1 \mathbb{P}(T_n \leq \alpha | p_{(n)} = t) f(t) dt.\end{aligned}$$

We first handle the second integral. Conditional on $p_{(n)} = t$, the other p values are independently uniform on $U(0, t)$, so if we divide them by t, we can apply the inductive hypothesis, once we observe:

$$\min_{1 \leq i \leq n-1} p_{(i)} \frac{n}{i} \leq \alpha \quad \Leftrightarrow \quad \min_{1 \leq i \leq n-1} \frac{p_{(i)}}{t} \cdot \frac{n-1}{i} \leq \frac{\alpha}{t} \cdot \frac{n-1}{n}$$

Therefor, $\mathbb{P}(T_n \leq \alpha | p_{(n)} = t) = \frac{\alpha}{t} \cdot \frac{n-1}{n}$ for $t \geq \alpha$. Then

$$\begin{aligned}\mathbb{P}(T_n \leq \alpha) &= \int_0^\alpha nt^{n-1} dt + \int_\alpha^1 \frac{\alpha}{t} \cdot \frac{n-1}{n} nt^{n-1} dt \\ &= \alpha^n + \alpha \int_\alpha^1 (n-1)t^{n-2} dt = \alpha^n + \alpha[1 - \alpha^{n-1}] = \alpha.\end{aligned}$$

Thus finishes the proof. As a summary, we see that the Simes procedure is powerful for a single strong effect, but has moderate power for many mild effects. □

5. Tests Based on Empirical CDF's

Define empirical CDF of $p_1, \cdots, p_n$ as

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{i : p_i \leq t\}$$

Under the global null H_0 we have that

$$\mathbb{E}(\hat{F}_n(t)) = t.$$

Moreover, if we assume that p_i's are independent, $n\hat{F}_n(t)$ is a binomial random variable with parameter t. Now, the idea is that, under the global null, $\hat{F}_n(t)$ should be around t. Hence, we can measure the distance between what we observe and what we expect and reject if the difference is large.

5.1. Tukey's Second-Level Significance Testing

Second-Level Significance Testing: Define the Higher Criticism Statistic

$$HC_n^* = \max_{0 \leq t \leq 1} \frac{\hat{F}_n(t) - t}{t(1-t)/n}$$

The difference between this test and Anderson-Darling statistic is that this uses a maximum value rather than a (squared) average.
Define statistics $HC_n(t)$ as

$$HC_n(t) = \frac{\hat{F}_n(t) - t}{t(1-t)/n} - \frac{\#\{\text{significance of level 1}\} - nt}{\sqrt{nt(1-t)}}$$

then HC_n^* scans across significance levels for departure from H_0. Hence, a large value of HC_n^* indicates signficance of an overall body of tests.

6. Sparse Mixtures

Original Model: We have independent statistics X_i distributed as

$$H_{0,i} : X_i \sim N(0, 1)$$

$$H_{1,i} : X_i \sim N(\mu_i, 1), \qquad \mu_i > 0$$

Here we consider a framework in which we are interested in possibilities within H_1 with a small fraction of non-null hypotheses. Rather than directly saying that there are some amount of nonzeromeans under H_1, we assume that our samples follow a mixture of $N(0, 1)$ and $N(\mu, 1)$ with μ fixed, resulting in the following:

Simple Model with Equal Means:

$$\begin{aligned} H_0 &: X_i \overset{i.i.d}{\sim} N(0, 1) \\ H_1 &: X_i \overset{i.i.d}{\sim} (1-\varepsilon)N(0, 1) + \varepsilon N(\mu, 1) \end{aligned}$$

Put another way, there are about $n\varepsilon$ non-nulls under H_1. The likelihood ratio for this model is then

$$L = \prod_{i=1}^{n} [(1-\varepsilon) + \varepsilon e^{\mu X_i - \mu^2/2}].$$

In Ingster (2000) and Jin (2003), they considered the dependence scheme of ε and μ on n as

$$\begin{aligned} \varepsilon_n &= n^{-\beta}, \frac{1}{2} < \beta < 1 \\ \mu_n &= \sqrt{2r \log n}, 0 < r < 1 \end{aligned}$$

to carry out asymptotic analysis. This automatically incorporates the settings of "the needle in a haystack problem" corresponds to $\beta = 1$ and $r = 1$; the small distributed effects case corresponds to $\beta = 1/2$.

In this more general case, Ingster (2000) and Jin (2003) found that there is a threshold curve for r of the form

$$\rho^*(\beta) = \begin{cases} \beta - \frac{1}{2} & \frac{1}{2} < \beta < \frac{3}{4} \\ (1 - \sqrt{1-\beta})^2 & \frac{3}{4} \le \beta \le 1 \end{cases}$$

such that

1. If $r > \rho^*(\beta)$, we can adjust the Neyman-Pearson test to achieve

$$\mathbb{P}_0(\text{Type I Error}) + \mathbb{P}_1(\text{Type II Error}) \to 0$$

2. If $r < \rho^*(\beta)$, then for any test

$$lim \quad inf_n \mathbb{P}_0(\text{Type I Error}) + \mathbb{P}_1(\text{Type II Error}) \ge 1$$

More interestingly, it has been proved for $r > \rho^*(\beta)$ in the sparse mixture setting, the higher criticism statistic with a proper threshold has full power asymptotically. This is interesting because the HC statistic does not need knowledge of ε and / or μ.

References

Ingster, Y. I. (2000). Adaptive chi-square tests. Journal of Mathematical Sciences, 99, 1110-1119.
Jin, J. (2003). Detecting and estimating sparse mixtures. Stanford University.
Williams, D. (1991). Probability with martingales. Cambridge university press.