STATS 300C: Theory of Statistics	**Spring 2022**
Lecture 2 — March 31	
Lecturer: Prof. Emmanuel Candès	*Editor: Parth Nobel, Scribe: Debolina Paul*

Warning: *These notes may contain factual and/or typographic errors. They are based on Emmanuel Candès's course from 2018 to 2022, and scribe notes written by John Cherian, Will Fithian, Kenneth Tay, Paulo Orenstein, Stephen Bates, and XY Han.*

2.1 Recap: Needle in a Haystack Problem

In the previous lecture, we introduced Bonferroni's global test. To study its properties further, we considered the following independent Gaussian sequence model:

$$Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1)$$

To precisely characterize the sense in which Bonferroni's global test is optimal, we computed the asymptotic power of the test under a particular null and alternative hypothesis. To this end, we defined the global null as,

$$H_0 = \bigcap_{i=1}^{n} H_{0,i} \qquad \text{where,} \qquad H_{0,i} : \mu_i = 0.$$

Under this testing criterion, the Bonferroni test is determined by the alternate hypothesis, which we denote by H_1, there exists i such that $\mu_i = \mu > 0$.

Then, relying on the fact that $\frac{\max_i Y_i}{\sqrt{2\log n}} \xrightarrow{p} 1$ under the global null,[1] we argued that there exists a sharp detection threshold for the needle in a haystack alternative hypothesis.

2.2 Optimality against sparse alternatives

Given the failure of the Bonferroni global test when $\mu^{(n)}$ is below the decision threshold, we might wonder if there exists some test that has non-negligible asymptotic power in this scenario. However, we show below that this decision threshold cannot be improved using any test of the global null against the "needle in a haystack" alternative. To prove this, we reduce our composite alternative to a simple hypothesis, and show that the optimal test given by the Neyman-Pearson Lemma still does no better than flipping a biased coin.

Bayesian Decision Problem We consider the problem of testing

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

[1] Further discussion of this claim in Appendix A.1.

where π selects a coordinate I uniformly and sets $\mu_I = \mu^{(n)} = (1-\epsilon)\sqrt{2\log n}$ with all other $\mu_i = 0$.

Note that both the null and alternative hypotheses specified above are simple. Thus, we can apply the Neyman-Pearson Lemma, which states that the most powerful test rejects for large values of the likelihood ratio.

Remark: If the coordinate was chosen in a non-random manner, i.e. $I = 1$, the problem above reduces to testing $H_0 : \mu_1 = 0$ against testing $H_1 : \mu_1 > 0$. As we saw in STATS 300A, there is no decision threshold for which this test is powerless. By contrast, we will see that for the "least favorable" alternative defined above, even the optimal likelihood ratio test exhibits the same asymptotic cutoff behavior as the Bonferroni global test.

Proceeding to define the likelihood ratio, we first observe that the densities under the null and alternative, where $\mu^{(n)} = \sqrt{2\log n}$, are given by,

$$f_0(y) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_j^2}$$

$$f_1(y) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_i-\mu)^2} \prod_{j:j\neq i} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_j^2}$$

So, in this case, the likelihood ratio becomes,

$$L = \frac{1}{n}\sum_{i=1}^{n} e^{Y_i\mu - \frac{1}{2}\mu^2}, \qquad \mathbb{E}_{H_0} L = 1$$

Naively, we might think to apply the central limit theorem in order to derive a limiting distribution for L under H_0. However, the central limit theorem (for triangular arrays) cannot be applied here. We shall focus, therefore, on deriving a weaker result.

Proposition 1. Under H_0, if $\mu^{(n)} = (1-\epsilon)\sqrt{2\log n}$, then $L \xrightarrow{p} 1$.

Proof. Proof in Appendix A.3. □

Since L converges in probability to 1, we claim that the likelihood ratio test is asymptotically powerless.

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

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

Proof. Note that

$$\begin{aligned}
\mathbb{P}_{H_1}(\text{Type II Error}) = \mathbb{P}_{H_1}(L \le T_n(\alpha)) &= \int \mathbf{1}_{\{L\le T_n(\alpha)\}}\, dP_1^n \\
&= \int \mathbf{1}_{\{L\le T_n(\alpha)\}} L\, dP_0^n \\
&= \int \mathbf{1}_{\{L\le T_n(\alpha)\}}\, dP_0^n + \int \mathbf{1}_{\{L\le T_n(\alpha)\}}(L-1)\, dP_0^n \\
&= (1-\alpha) + \int \mathbf{1}_{\{L\le T_n(\alpha)\}}(L-1)\, dP_0^n \\
&\to (1-\alpha).
\end{aligned}$$

The last claim follows from the fact that $L \xrightarrow{p} 1$. We can make this rigorous as follows: let $Z_n = \mathbf{1}_{\{L\le T_n(\alpha)\}}(L-1)$. 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 $\mathbb{E}|Z_n| \to 0$ (this is a simple result that can be checked by hand).

$\square$

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

$$\mathbb{P}_{H_0}\left(\text{Type I Error}\right) + \mathbb{P}_{H_1}\left(\text{Type II Error}\right) \to 1.$$

We further note that the following inequality holds for *any* test,

$$\begin{aligned}
&\mathbb{P}_{H_0}\left(\text{Type I Error}\right) + \sup_{H_1} \mathbb{P}_{H_1}\left(\text{Type II Error}\right) \ge \\
&\qquad\qquad \mathbb{P}_{H_0}\left(\text{Type I Error}\right) + \mathbb{E}_{H_1\sim\pi}\mathbb{P}_{H_1}\left(\text{Type II Error}\right)
\end{aligned}$$

and where the supremum is taken over any alternative hypothesis for $\{\mu_i\}$ in which one coordinate has mean $\mu^{(n)} = (1-\epsilon)\sqrt{2\log n}$. Since we have proven that the limit of the RHS is at least 1, taking a $\liminf$ on both sides of the inequality yields the desired result:

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

In summary, we have proven that there is no test that is asymptotically able to distinguish between the null and alternative hypotheses when the mean of the needle in the haystack, $\mu^{(n)}$, is smaller than the $\sqrt{2\log n}$ threshold.

2.3 χ^2 test

Turning our attention for the time being away from the Bonferroni test, we might also ask ourselves if the Fisher combination test considered in the previous lecture is asymptotically powerful for a different set of alternatives.

The definition of the Fisher combination test statistic ($T_n = -2\sum_{i=1}^n \log p_i$) makes it difficult to directly analyze. However, for the Gaussian model we considered in our power

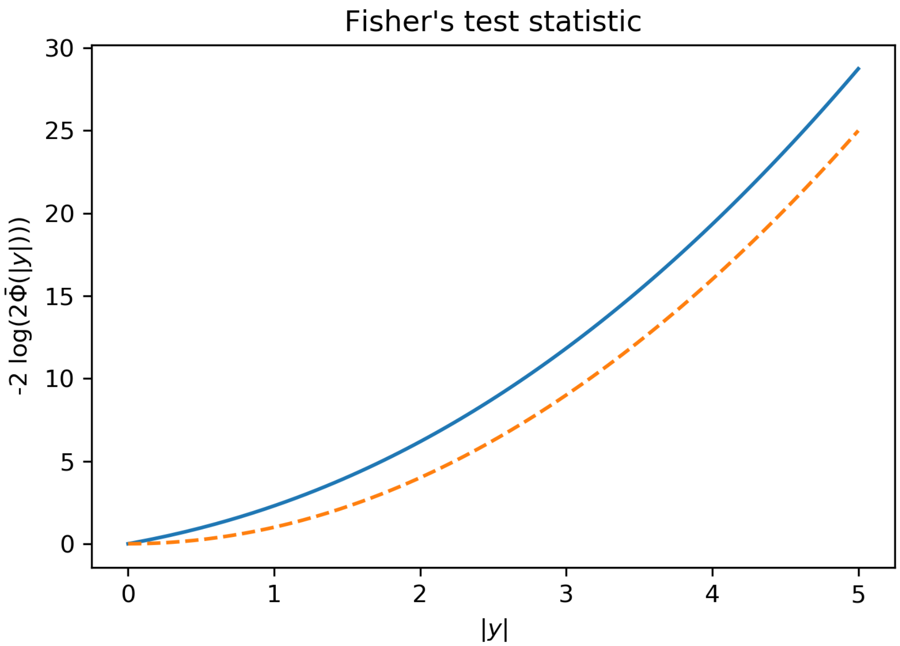

Figure 2.1. Fisher combination test statistic (solid) and $z = |y|^2$ line (dashed)

analysis of the Bonferroni global test, Figure 2.1 demonstrates that each term in the Fisher test statistic is qualitatively similar to the better-understood statistic y^2. This observation motivates our analysis of the χ^2 test. This family of tests is also known as ANOVA, when at least one of the variables concerned is continuous.

For our power analysis of the χ^2 test, we define our null and alternative hypotheses as follows.

$$\begin{aligned} Y_i &= \mu_i + z_i \qquad && z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \\ H_0 &: \mu_i = 0 \text{ for all } i \qquad && H_1 : \text{at least one } \mu_i \neq 0 \end{aligned}$$

Then, the test statistic for the χ^2 test is

$$T_n = \sum_{i=1}^n Y_i^2 = \|Y\|_2^2$$

Under H_0: $T_n \sim \chi_n^2$, and the level-α test rejects H_0 when $T_n > \chi_n^2(1-\alpha)$. Note that under H_0,

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

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

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

implying that

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

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

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

Again applying a CLT approximation for large n, we have that

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

Remark: What is a computationally efficient way to draw samples from the non-central χ_2 distribution with n degrees of freedom? Since the distribution of $\chi_n^2(\|\mu\|_2^2) \stackrel{d}{=} \sum_{i=1}^n (\mu_i + z_i)^2$ is not dependent on how the mass of μ is distributed over its indices, but rather only depends on the value of $\|\mu\|_2^2$ since, $\chi_n^2(\|\mu\|_2^2) \stackrel{d}{=} \chi_{n-1}^2 + (z + \|\mu\|_2)^2$. This is both efficient to compute, and captures the idea that, in this setting, the χ^2 test detects a large deviation from one null hypothesis as well as it detects an equally-sized in ℓ_2 norm small deviations occurring across all or many hypotheses.

2.4 Detection thresholds for small distributed effects

To ascertain a detection threshold for the χ^2 test, we first define two useful quantities: Z_n, which is a normalized test statistic, and θ_n, which we will see is proportional to the signal-to-noise ratio (SNR) for the test.

$$Z_n := \frac{T_n - n}{\sqrt{2n}} \qquad \theta_n := \frac{\|\mu\|^2}{\sqrt{2n}}$$

Applying some elbow grease to the normal approximation above and rearranging terms, we derive the following limiting distributions:

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

The asymptotic power of the χ^2 test is thus determined by θ_n, which measures the relative size of $\|\mu\|^2$ compared to $\sqrt{n}$. In particular, we can see from our expressions for the asymptotic null and alternative distributions above that the χ^2 test is easy when $\theta_n \gg 1$ and hard when $\theta_n \ll 1$. For example, when $\theta_n = 2$, the power of the test is roughly $\mathbb{P}(z_1 > 1.65 - 2) \approx 66\%$.

SNR: If we had started with a model in which the noise variance is defined to be σ^2, i.e.

$$Y_i = \mu_i + \sigma z_i \qquad i = 1, \dots, n$$

then by the same argument as above, we would see that the detection power depends sensitively on

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

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 $\theta_n \propto \text{SNR}$ with a constant of proportionality equal to $\sqrt{n/2}$.

Though this test does not have a sharp decision threshold, a natural question arises: **when $\theta_n \ll 1$, is there a test that is asymptotically more powerful than the χ^2 test?** To show that the answer is no, we use the strategy employed above for the Bonferroni test: we first introduce a pair of simple hypotheses using a "Bayesian" alternative, and show that in this setting, the most powerful (likelihood ratio) test is powerless as $\theta_n \to 0$.

Decision Problem:

$$H_0 : \mu = \mathbf{0}_n$$
$$H_1 : \mu \sim \pi_\rho^{(n)}$$

where $\pi_\rho^{(n)}$ distributes mass uniformly on the sphere (in $\mathbb{R}^n$) of radius $\rho^{(n)}$.

Likelihood ratio: By way of the Neyman-Pearson Lemma, we note that the optimal test for this pair of hypotheses rejects for large values of the likelihood ratio. To derive this statistic and its limiting distribution under the null, we introduce some notation. Let $\mu = \rho^{(n)}u$ where u is uniformly distributed on the unit sphere denoted by S^{n-1}. Let π_n be the uniform distribution on this sphere. We have then that

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

In this setting, $\theta_n := \frac{(\rho^{(n)})^2}{\sqrt{2n}}$ measures the signal-to-noise ratio for our test, and as expected, when $\theta_n \to 0$, our test becomes powerless. This claim is summarized by the following proposition.

Proposition 3. Under H_0, if $\theta_n \to 0$, then $L \xrightarrow{P} 1$ and the Neyman-Pearson test is asymptotically no better than flipping a biased coin.

Proof. Proof in Appendix A.5. □

Heuristically, this proposition leads us to conclude that the χ^2 test is a "good" test. It is only asymptotically powerless for vanishing SNR, and in this setting, *no test* can do better in the limit.

2.5 Comparison of Bonferroni's and χ^2 tests

We turn now to a more practical comparison of the two global tests; namely, we provide a qualitative characterization of the sets of alternatives for which the Bonferroni and χ^2 tests are most powerful.

Before we present this summary, we consider two examples in which the tests have very different power characteristics.

Example 1: $n^{1/4}$ of the μ_i's are equal to $\sqrt{2\log n}$. To make this concrete, if $n = 10^6$, $n^{1/4} \approx 32$ and $\sqrt{2\log n} \approx 5.3$. In this set-up, our asymptotic results suggest that the Bonferroni test will have (approximately) full power, but because

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

the χ^2 test has (approximately) 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, however, because when n is large, it's very likely that the smallest p-value comes from a null μ_i rather than 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.

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

- **Sparse strong effects**: μ_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**: μ_i is 1.1 for $1 \leq i \leq k = 2400$ and 0 otherwise.
- **Distributed moderate effects**: μ_i is 1.3 for $1 \leq i \leq k = 2400$ and 0 otherwise.

Sparse strong effects: In the sparse setting, the asymptotic power of Bonferroni's method can be approximated using the inclusion-exclusion principle. Let A denote the event that we reject because at least one of $Y_1, \ldots, Y_4$ exceeds the rejection threshold, and let B denote the event that we reject because one of the "noise" variables exceeds the Bonferroni threshold. Then,

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\
&\approx \underbrace{\left(1 - \left(\frac{1}{2}\right)^4\right)}_{=0.9375} + q(\alpha) - \left(1 - \left(\frac{1}{2}\right)^4\right) q(\alpha) \\
&\approx 0.94
\end{aligned}$$

On the other hand, the χ^2 test would be almost powerless, as $\theta_n = \|\mu\|^2/\sqrt{2n} \approx 0.084 \ll 1$. A numerical estimate of the power for these tests with 500 trials yields the expected result:

$$\text{Bonferroni} \approx 94\%$$
$$\text{Chi-sq} \approx 6\%$$

Distributed weak effects: For the alternative of distributed weak effects, the power of the Bonferroni test is roughly

$$\mathbb{P}_{H_1^n}\left[\max|Y_i| > |z(\alpha/(2n))|\right] \le \mathbb{P}_{H_1}\left[\max_{i\le k}|Y_i| > |z(\alpha/(2n))|\right] + \mathbb{P}\left[\max_{i>k}|z_i| > |z(\alpha/(2n))|\right] \approx 0.066.$$

By contrast, $\theta_n = \|\mu\|^2/\sqrt{2n} \doteq 2.05$. So, while the Bonferroni test has almost no power, the χ^2 test can detect the alternative more often than not. Numerically,

$$\text{Bonferroni} = 6\%$$
$$\text{Chi-sq} = 66\%$$

Distributed moderate effects : For the "moderate effects" scenario listed above, $\theta_n = \|\mu\|^2/\sqrt{2n} \approx 2.8$. Here, the χ^2 test is very powerful, but the Bonferroni test remains nearly powerless. Numerically,

$$\text{Bonferroni} = 8\%$$
$$\text{Chi-sq} = 89\%$$

Simulating the weak distributed effect experiment confirms that the practical challenge for the Bonferroni cutoff is that the largest values of Y_i tend to come from the noise variables. Figure 2.2 thus captures why the Bonferroni global test is incapable of detecting weak distributed effects. We can summarize our conclusions in the following table:

	Small, distributed effects	**Few strong effects**
ANOVA (Analysis of Variance)	Powerful	Weak
Bonferroni	Weak	Powerful

Next week: Can we introduce a method that has the best of both worlds? Is there a single test that is powerful for any alternative?

A Appendix

A.1 Concentration of Gaussian Maxima

How large is our threshold $t_n = |z(\alpha/n)|$ (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:

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

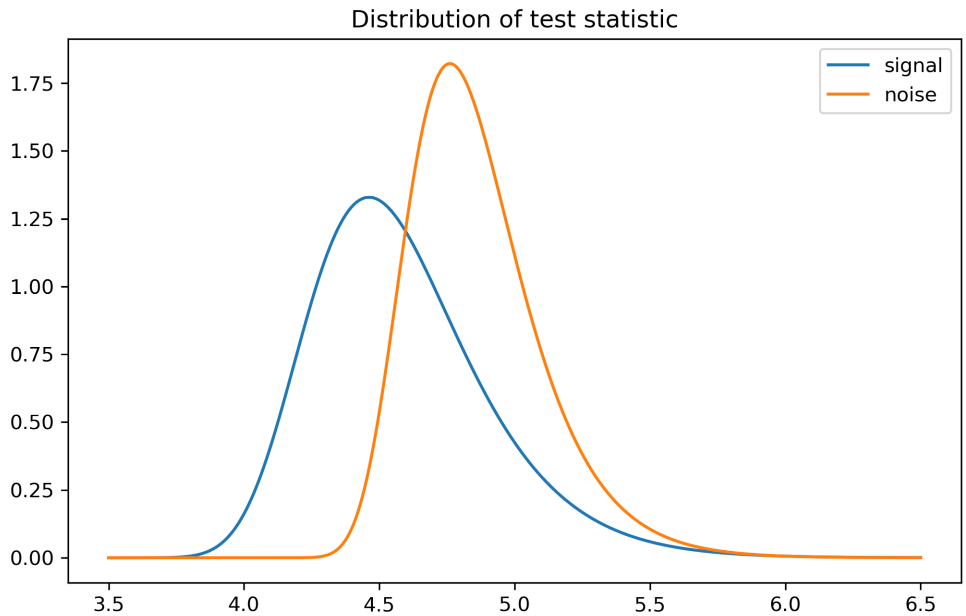

Figure 2.2. Distribution of the maximum Y_i observed for the "distributed weak effects" example.

where $Z \sim \mathcal{N}(0,1)$. 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.

A.2 Triangular Array CLT Failure

Writing $X_i = e^{y_i\mu^{(n)} - \frac{1}{2}(\mu^{(n)})^2}$, 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 $\frac{1}{n}\mathrm{Var}X_1$.

We would like to apply the CLT; however, because $\mu^{(n)} = (1-\varepsilon)\sqrt{2\log n} \to \infty$, we would need a triangular array argument.

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

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

as $n \to \infty$. Though this is not a proof that the CLT fails, it does suggest that we might take the alternative approach discussed in the main text of these notes.

A.3 Convergence of L for the Sparse Alternative

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

Proof. Recall

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

with $X_i = e^{y_i \mu^{(n)} - \frac{1}{2}(\mu^{(n)})^2}$ i.i.d.

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

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

We have

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

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 $\exp\left((\mu^{(n)})^2\right)/n$ converges to zero in this case. $\square$

A.4 Asymptotic Powerlessness of LRT (Proof via Contiguity)

We argued in the main text of these notes that when $L \xrightarrow{P} 1$, the Bonferroni and χ^2 tests become asymptotically powerless. Here, we describe an alternative proof of this claim that relies on two contiguity results proven in STATS300B.

First, we state a simplified version of LeCam's first lemma.

Lemma 1. Let P_n and Q_n be sequences of probability measures on measurable spaces $(\Omega_n, \mathcal{A}_n)$. Then, the following statements are equivalent:

1. $Q_n \triangleleft P_n$

2. If $dQ_n/dP_n \overset{P_n}{\rightsquigarrow} V$ along a subsequence, then $\mathbb{E}[V] = 1$.

Taking $Q_n = P_1^n$ and $P_n = P_0^n$, we see that $L \xrightarrow{P} 1$ implies condition 2, and we conclude that $P_1^n \triangleleft P_0^n$, namely, P_1^n is contiguous with respect to P_0^n. We can now apply a generalized version of LeCam's third lemma to obtain the desired result.

Theorem 1. Let P_n and Q_n be sequences of probability measures on measurable spaces $(\Omega_n, \mathcal{A}_n)$ and let $X_n : \Omega_n \to \mathbb{R}^k$ be a sequence of random vectors. Suppose that $Q_n \triangleleft P_n$ and

$$\left(X_n, \frac{dQ_n}{dP_n} \right) \overset{P_n}{\rightsquigarrow} (X, V)$$

Then, $X_n \overset{Q_n}{\rightsquigarrow} L$ where L is any random variable whose law satisfies $\mathbb{P}(L \in B) = \mathbb{E}[\mathbf{1}_B(X)V]$ for arbitrary Borel set B.

Let X_n be the test statistic T_n (which we know to have some limiting distribution T under the null), and let $P_n = P_0^n$ and $Q_n = P_1^n$ as above. We conclude that T_n converges to T under both the null and alternative hypotheses since the limit law for T_n under the alternative can be re-expressed as $\mathbb{E}[\mathbf{1}_B(X) \cdot (1)] = \mathbb{P}(T \in B)$. Since the limiting distributions under the null and alternative sequences are identical, our test must be asymptotically powerless.

A.5 Convergence of L for the Distributed Alternative

We use some shorthand in this argument for brevity's sake. Namely, for the derivation, we drop the n-index corresponding to the sequence of hypotheses and refer to an expectation under the null using the subscript 0 rather than H_0.

Useful Relationship: If $y \sim \mathcal{N}(\mathbf{0}_n, \mathbb{I}_n)$, 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 $v = e_1 = (1, 0, \ldots, 0)$ 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

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

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 $\frac{\|\mu\|^2}{\sqrt{2n}} \to 0$ as $n \to \infty$.