STAT 689: Large-scale and high-dimensional statistical inference **Fall 2024**

Lecture 2

1 Fisher's combination tests (L_2 tests)

As we have shown in the previous lecture, the Bonferroni's method (or equivalently the maximum test) is optimal for detecting sparse and strong signals (the "needle in a haystack" alternative). It is interesting to ask if a similar result can be established for Fisher's combination test for detecting dense and weak signals.

Instead of considering the original definition of the Fisher's combination test which is hard to analyze, we consider the Gaussian location model and a L_2 type test. More precisely, we consider the model

$$z_i = \xi_i + \mu_i, \quad \xi_i \sim^{\text{i.i.d}} N(0,1), \quad i = 1, 2, \dots, n,$$

and the hypotheses

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = 0 \quad \text{versus} \quad H_a : \mu_i \neq 0 \text{ for some } 1 \leq i \leq n.$$

We consider the L_2 test defined as

$$T_n = \|\mathbf{z}\|_2^2 = \sum_{i=1}^n z_i^2.$$

The L_2 type test (or χ^2 test as $T \sim \chi_n^2$ under H_0) is powerful for detecting weak and dense signals. When $n \to +\infty$, we can consider the test after centering and standardization:

$$Z_n := \frac{T_n - n}{\sqrt{2n}} = \frac{\sum_{i=1}^n \xi_i^2 - n}{\sqrt{2n}} \to^d N(0,1).$$

Therefore, we reject H_0 if

$$Z_n \geq z_{1-\alpha},$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of $N(0,1)$.

Under either H_0 or H_a, we have

$$T_n = \sum_{i=1}^n (\xi_i + \mu_i)^2 = \sum_{i=1}^n \xi_i^2 + \|\boldsymbol{\mu}\|^2 + 2\sum_{i=1}^n \xi_i \mu_i,$$

where $\|\boldsymbol{\mu}\|^2 = \sum_{i=1}^n \mu_i^2$.

Exercise 2.1: Assuming $\|\boldsymbol{\mu}\| = o(\sqrt{n})$, prove that $\sum_{i=1}^n \xi_i \mu_i / \sqrt{n} = o_p(1)$ and in this case, we have

$$\frac{T_n - (n + \|\boldsymbol{\mu}\|^2)}{\sqrt{2n}} \to^d N(0,1).$$

More generally, without any assumption on $\|\boldsymbol{\mu}\|$, we have

$$\frac{T_n - (n + \|\boldsymbol{\mu}\|^2)}{\sqrt{2n + 4\|\boldsymbol{\mu}\|^2}} \to^d N(0,1).$$

Note: use the fact $E[\xi_i^3] = 0$ in the proof.

2 Power function for L_2 tests

Our first goal is to derive a detection threshold for the L_2 test. We first define $\theta_n = \|\boldsymbol{\mu}\|^2/\sqrt{2n}$ which is proportional to the signal-to-noise (SNR) ratio. The power function of T_n is given by

$$\begin{aligned}\beta(\theta_n) :=& P_{H_a}\left(\frac{T_n - n}{\sqrt{2n}} \geq z_{1-\alpha}\right)\\ =& P_{H_a}\left(\frac{T_n - (n + \|\boldsymbol{\mu}\|^2)}{\sqrt{2n + 4\|\boldsymbol{\mu}\|^2}} \geq \frac{\sqrt{2n}z_{1-\alpha} - \|\boldsymbol{\mu}\|^2}{\sqrt{2n + 4\|\boldsymbol{\mu}\|^2}}\right)\\ \approx& P_{H_a}\left(Z \geq \frac{\sqrt{2n}z_{1-\alpha} - \|\boldsymbol{\mu}\|^2}{\sqrt{2n + 4\|\boldsymbol{\mu}\|^2}}\right)\\ =& 1 - \Phi\left(\frac{z_{1-\alpha} - \theta_n}{\sqrt{1 + \theta_n/\sqrt{n/8}}}\right),\end{aligned}$$

where $Z \sim N(0,1)$ and the approximation to due **Exercise 1**. We have the following cases:

- $\theta_n = o(1)$, $\beta(\theta_n) \to \alpha$;
- $\theta_n \to c < \infty$, we have $\beta(\theta_n) \to 1 - \Phi(z_{1-\alpha} - c)$;
- $\theta_n \to +\infty$, $\beta(\theta_n) \to 1$.

Similar to the analysis for Bonferroni's test, we consider the following problem: $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_a : \boldsymbol{\mu} \sim \pi_\rho^{(n)}$, where $\pi_\rho^{(n)}$ is a uniform distribution on the sphere with radius ρ. In other words, $\boldsymbol{\mu} = \rho\mathbf{u}$, where $\mathbf{u}$ is uniformly distributed on the unit sphere $\mathcal{S}^{n-1}$. By the Neyman-Pearson Lemma, we know that the optimal test for this pair of hypotheses rejects for large values of the likelihood ratio. Under H_a, the distribution of $\mathbf{z}$ after integrating out $\boldsymbol{\mu}$ is given by

$$\frac{1}{(2\pi)^{n/2}}\int_{\mathcal{S}^{n-1}} \exp(-\|\mathbf{z} - \rho\mathbf{u}\|^2/2)d\pi_1^{(n)}(\mathbf{u}).$$

Thus, the likelihood ratio statistic is equal to

$$\begin{aligned}L_n(\mathbf{z}) =& \frac{\int_{\mathcal{S}^{n-1}} \exp(-\|\mathbf{z} - \rho\mathbf{u}\|^2/2)d\pi_1^{(n)}(\mathbf{u})}{\exp(-\|\mathbf{z}\|^2/2)}\\ =& \int_{\mathcal{S}^{n-1}} \exp(-\rho^2\|\mathbf{u}\|^2/2 + \rho\mathbf{z}^\top\mathbf{u})d\pi_1^{(n)}(\mathbf{u})\\ =& \int_{\mathcal{S}^{n-1}} \exp(-\rho^2/2 + \rho\mathbf{z}^\top\mathbf{u})d\pi_1^{(n)}(\mathbf{u}).\end{aligned}$$

Claim: When $\rho^2/\sqrt{2n} \to 0$, $L_n \to^p 1$ under H_0. In this case, the Neyman-Pearson test is asymptotically no better than flipping a biased coin.

Let P and Q be two measures on a measurable space $(\Omega, \mathcal{A})$. We say that Q is absolutely continuous with respect to P if $P(A) = 0$ implies that $Q(A) = 0$ for every measurable set A; this is denoted by $Q \ll P$. We now consider an asymptotic version of the absolute continuity. Let $(\Omega_n, \mathcal{A}_n)$ be measurable spaces, each equipped with a pair of probability measures P_n and Q_n.

Definition. The sequence Q_n is contiguous with respect to the sequence P_n if $P_n(A_n) \to 0$ implies that $Q_n(A_n) \to 0$ for every sequence of measurable sets A_n.

The name "contiguous" is standard but perhaps conveys a wrong image. 'contiguity" suggests sequences of probability measures living next to each other, but the correct image is "on top of each other" in the limit.

Step I: We first state a version of LeCam's first Lemma; see Lemma 6.4 of van der Vaart's book, *Asymptotic Statistics*.

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

1. Q_n is contiguous with respect to P_n.
2. If $dQ_n/dP_n \rightsquigarrow^{P_n} V$ along a subsequence, then $E[V] = 1$.

In the second statement above, you shall think of $dQ_n/dP_n : \Omega_n \to [0, \infty)$ as a random variable and study its law under P_n.

To apply this Lemma, we let Q_n be the distribution under H_a and P_n be the distribution under H_0. Then $L_n = dQ_n/dP_n$. When $L_n \to^p 1$ under H_0, Condition 2 is fulfilled, and hence by LeCam's first Lemma, Q_n is contiguous with respect to P_n. We next state LeCam's third Lemma; see Theorem 6.6 of van der Vaart's book.

LeCam's third Lemma. Let P_n and Q_n be sequences of probability measures on measurable spaces $(\Omega_n, \mathcal{A}_n)$. Let $T_n : \Omega_n \to \mathbb{R}^k$ be a sequence of random vectors. Suppose Q_n is contiguous with respect to P_n and

$$(T_n, dQ_n/dP_n) \rightsquigarrow^{P_n} (T, V).$$

Then $T_n \rightsquigarrow^{Q_n} \widetilde{T}$, where $\widetilde{T}$ is a random vector with the distribution

$$P(\widetilde{T} \in B) = E[\mathbf{1}\{T \in B\} V].$$

In our case, we let T_n be some test statistic, where T_n converges to T in distribution under the null. As $V = 1$, $P(\widetilde{T} \in B) = P(T \in B)$, which indicates that T_n has the same limiting distribution under both the null and the alternative. Thus, it must be asymptotically powerless.

Step II: To complete the proof, we now show that $L_n \to^p 1$ under H_0. Clearly, $E_0[L_n]$=1. We aim to show that $\mathrm{var}(L_n|H_0) \to 0$ or equivalently $E_0[L_n^2] - 1 \to 0$. As $\mathbf{z} \sim N(0, \mathbf{I})$ under the null, we have $E_0[\exp(\mathbf{z}^\top \mathbf{t})] = \exp\left(\|\mathbf{t}\|^2/2\right)$ and thus

$$\begin{aligned}
E_0[L_n^2] &= \int_{\mathcal{S}^{n-1}} \int_{\mathcal{S}^{n-1}} E \exp(-\rho^2 + \rho \mathbf{z}^\top \mathbf{u} + \rho \mathbf{z}^\top \mathbf{v}) d\pi_1^{(n)}(\mathbf{u}) d\pi_1^{(n)}(\mathbf{v}) \\
&= \int_{\mathcal{S}^{n-1}} \int_{\mathcal{S}^{n-1}} \exp(-\rho^2 + \rho^2 \|\mathbf{u} + \mathbf{v}\|^2/2) d\pi_1^{(n)}(\mathbf{u}) d\pi_1^{(n)}(\mathbf{v}) \\
&= \int_{\mathcal{S}^{n-1}} \int_{\mathcal{S}^{n-1}} \exp(\rho^2 \mathbf{u}^\top \mathbf{v}) d\pi_1^{(n)}(\mathbf{u}) d\pi_1^{(n)}(\mathbf{v}) \\
&= \int_{\mathcal{S}^{n-1}} \int_{\mathcal{S}^{n-1}} \exp(\rho^2 \mathbf{u}^\top \mathbf{v}) d\pi_1^{(n)}(\mathbf{u}) d\pi_1^{(n)}(\mathbf{v}).
\end{aligned}$$

Given $\mathbf{v}$, we can consider a rotation $\mathbf{B}$ such that $(\mathbf{B}\mathbf{u})^\top \mathbf{v} = u_1$ and $\mathbf{B}\mathbf{u}$ still follows a uniform distribution on the sphere. Thus we have

$$\begin{aligned}
E_0[L_n^2] &= \int_{\mathcal{S}^{n-1}} \int_{\mathcal{S}^{n-1}} \exp(\rho^2 \mathbf{u}^\top \mathbf{v}) d\pi_1^{(n)}(\mathbf{u}) d\pi_1^{(n)}(\mathbf{v}) \\
&= \int_{\mathcal{S}^{n-1}} \exp(\rho^2 u_1) d\pi_1^{(n)}(\mathbf{u}) \\
&= E[\exp(\rho^2 u_1)].
\end{aligned}$$

Using the Taylor expansion and the fact that $E[u_1^2] = 1/n$, we have

$$E_0[L_n^2] = 1 + E[\rho^2 u_1 + \rho^4 u_1^2/2 + \rho^4 u_1^3/6 + \cdots] = 1 + \frac{\rho^4}{2n}(1 + o(1)) = 1 + o(1).$$

3 Comparison of Bonferroni's and L_2 tests

We now provide a qualitative characterization of the sets of alternatives for which the Bonferroni and L_2 tests are most powerful. Recall the in theory

- Bonferroni test's power is determined by $\max_i |\mu_i|$.
- L_2 test's power is determined by $\theta_n = \sum_{i=1}^n \mu_i^2/\sqrt{2n}$.

We consider two examples in which the tests have very different power characteristics.

Example 1. Suppose $n^{3/8}$ components of $\boldsymbol{\mu}$ are equal to $1.1\sqrt{2\log(n)}$. Then we have $\theta_n = 2.42n^{3/8}\log(n)/\sqrt{2n} \to 0$. Hence the L_2 test is powerless. In contrast, as $\max_i |\mu_i| = 1.1\sqrt{2\log(n)}$, Bonferroni test has power approaching one.

Example 2. Consider the case where $\sqrt{2n}$ components of $\boldsymbol{\mu}$ are equal to a. In this case, $\theta_n = a^2$. For large n, the L_2 test's power is approximately $1 - \Phi\left(z_{1-\alpha} - a^2\right)$. With $\alpha = 0.05$ and $a = 2$, the power is about 99%. On the other hand, Bonferroni's test is asymptotically powerless because $\max_i |\mu_i| = a$.