36-450: Advanced Statistical Theory II **Winter 2025**

Lecture 0: February 03

Lecturer: Matey Neykov *Scribe: Tuorui Peng*

0.1 Example 2: Multinomial Testing

Motivation: We are curious that: given a lottery with d balls, is the lottery fair? That is, is the probability of each ball being drawn equal to $1/d$?

0.1.1 Problem Statement

We have the distribution family $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ for which $\mathbb{P}_\theta$ is supported on $[d] := \{1, 2, \ldots, d\}$, and satisfies

$$\Theta = \Big\{\theta : p_\theta(i) \geq 0, \quad \sum_{i=1}^{d} p_\theta(i) = 1\Big\}$$

And we consider the uniformality test, i.e. the (null) parameter of interest is

$$\{\theta_0\} = \Theta_0 = \big\{\theta : p_\theta(i) = 1/d, \quad \forall i \in [d]\big\}$$

w.r.t. the corresponding alternative:

$$H_0 : p_\theta = p_{\theta_0} = \mathrm{Unif}[d] \longleftrightarrow H_1 : p_\theta \neq p_{\theta_0}$$

and the testing is s.t. the risk is controlled:

$$R_{\hat{\psi},\varepsilon} := \mathbb{P}_0\left(\hat{\psi}_n = 1\right) + \sup_{p_\theta \in H_1} \mathbb{P}_\theta\left(\hat{\psi}_n = 0\right) \leq \eta$$

for which, note that we have the relation between probability of error and the total variation distance d_{TV}, it suffices to control the total variation distance, which would leads to the following form of rejection region represented by ℓ_1 norm:

$$\text{Rejection Region}_\varepsilon = \big\{\theta : \|p_\theta - p_{\theta_0}\|_1 > \epsilon\big\}$$

Goal: We are curious about the (asymptotic) behaviour of the critical value ϵ^*:

$$\varepsilon^* = \inf\{\varepsilon : \inf_{\hat{\psi}} R_{\hat{\psi},\varepsilon} \leq \eta\}$$

0.1.2 Upper Bound Side

0.1.2.1 Challenge

If we can construct an estimator to $\|p_\theta - p_{\theta_0}\|_1$, then a test based on this estimator would be a valid one. But here an unbiased estimator to the $\|p_\theta - p_{\theta_0}\|_1$ is intractable (compared with the previous example of mean hypothesis testing, in which we can access an unbiased estimator to $\|y\|_2^2$). Thus we consider using other related norm to bound it.

0.1.2.2 Roadmap of the upper bound side

1. (Lower) bound ε, i.e. ℓ_1 norm, which further bound $\|p_\theta - p_{\theta_0}\|_2^2$; notice that $\|p_\theta - p_{\theta_0}\|_2^2$ can be easily estimated, so we can construct the test based on its estimator T
2. As required by Neyman-Pearson criterion, we construct the rejection region boundary t_α that can control the type I error α by
$$t_\alpha = \sqrt{\frac{1}{\alpha} var_{\theta_0}(T)}$$
3. The $\|p_\theta - p_{\theta_0}\|_2^2$ bound yields an upper bound on $var_\theta(T)$;
4. then guarentee that
$$\mathbb{E}_{\theta\in\Theta_{H_a}}[T] \geq t_\alpha + \sqrt{\frac{1}{\beta} var_\theta(T)}$$
which further makes sure that the type II error β is controlled, and we have a valid test.

0.1.2.3 Proof of the upper bound

Since we have by Cauchy-Schwarz inequality that $\|p_\theta - p_{\theta_0}\|_2^2$

Denote our data $X = \{X_i\}_{i=1}^n$, $X_i = \{X_{i1}, X_{i2}, \ldots, X_{id}\}$, $X_i \in \{\hat{e}_1, \ldots, \hat{e}_d\}$ where $\hat{e}_j$ is the j-th canonical basis vector in $\mathbb{R}^d$. Then we have the following estimator for $\|p_\theta - p_{\theta_0}\|_2^2$:

Lemma 0.1 *With $\underset{n\times d}{X}$ being the data defined above and $p_{\theta_0} = \mathrm{Unif}[d]$, we have the following U-statistics:*
$$\mathbb{E}_\theta[T] := \mathbb{E}_\theta\left[\binom{n}{2}^{-1} \sum_{i<j} X_i' X_j - \frac{1}{d}\right] = \|p_\theta - p_{\theta_0}\|_2^2 .$$

Proof: Note that

$$\mathbb{E}_\theta\left[X_i X_j\right] = \delta_{ij} + (1-\delta_{ij})\sum_{k=1}^{d} p_\theta(k)^2$$

we have

$$\begin{aligned}
\mathbb{E}_\theta\left[\binom{n}{2}^{-1}\sum_{i<j} X_i' X_j - \frac{1}{d}\right] &= \binom{n}{2}^{-1}\sum_{i<j}\mathbb{E}_\theta\left[X_i X_j\right] - \frac{1}{d}\\
&= \sum_{k=1}^{d} p_\theta(k)^2 - \frac{1}{d}\\
&= \sum_{k=1}^{d}\left(p_\theta(k) - \frac{1}{d}\right)^2\\
&= \left\|p_\theta - p_{\theta_0}\right\|_2^2 .
\end{aligned}$$

■

Lemma 0.2 *For the above U-statistics, we have*

$$\begin{aligned}
var_\theta(T) &= \binom{n}{2}^{-1}\left(\|p_\theta\|_2^2 - \|p_\theta\|_2^4\right) + \binom{n}{2}^{-2} n(n-1)(n-2)\left(\|p_\theta\|_3^3 - \|p_\theta\|_2^4\right)\\
&\asymp \frac{\|p_\theta\|_2^2 - \|p_\theta\|_2^4}{n^2} + \frac{\|p_\theta\|_3^3 - \|p_\theta\|_2^4}{n}
\end{aligned}$$

Proof: Leave as an exercise. ■

Now we can decide the rejection region. By chebyshev's inequality, we have **under** H_0 that $var_{\theta_0}(T) = \binom{n}{2}^{-1}\frac{1}{d}(1-\frac{1}{d})$ and rejection region should take the following form:

$$T > t_\alpha := \sqrt{\frac{1}{\alpha} var_{\theta_0}(T)} = \sqrt{\frac{1}{\alpha}\binom{n}{2}^{-1}\frac{1}{d}(1-\frac{1}{d})} \asymp \frac{1}{n\sqrt{d}}$$

so that type I error $\leq \alpha$. Now it suffices to find the critical rate $\varepsilon \asymp \text{func}(n,d)$ s.t.

type II error $\leq \beta$. We guarentee so by ensuring

$$\begin{aligned}\mathbb{E}_\theta\left[T\right] \geq & t_\alpha + \sqrt{\frac{1}{\beta} var_\theta(T)} \\ \text{i.e. } \mathbb{E}_{\theta\in\Theta_{H_a}}\left[T\right] = & \left\|p_\theta - p_{\theta_0}\right\|_2^2 \\ \geq & t_\alpha + \sqrt{\frac{1}{\beta} var_\theta(T)} \\ \gtrsim & \frac{1}{n\sqrt{d}} + \sqrt{var_\theta(T)} \\ \asymp & \frac{1}{n\sqrt{d}} + \sqrt{\frac{\left\|p_\theta\right\|_2^2 - \left\|p_\theta\right\|_2^4}{n^2} + \frac{\left\|p_\theta\right\|_3^3 - \left\|p_\theta\right\|_2^4}{n}}\end{aligned}$$

and it suffices to upper bound $var_\theta(T)$

Lemma 0.3 *Under some $\theta \in \Theta_{H_a}$, we have*

$$var_\theta(T) \lesssim \frac{\left\|p_\theta\right\|_2^2}{n^2} + \frac{\left\|p_\theta\right\|_3^3 - \left\|p_\theta\right\|_2^4}{n} \tag{1}$$

$$\left\|p_\theta\right\|_2^2 = \left\|p_\theta - p_{\theta_0}\right\|_2^2 + \frac{1}{d} \tag{2}$$

$$\left\|p_\theta\right\|_3^3 - \left\|p_\theta\right\|_2^4 \leq \left\|p_\theta - p_{\theta_0}\right\|_2^3 + \frac{3}{d}\left\|p_\theta - p_{\theta_0}\right\|_2^2 \tag{3}$$

Proof:

1. Trivial by Lemma 0.2.
2. We have

$$\left\|p_\theta\right\|_2^2 = \sum_{i=1}^d p_\theta(i)^2 = \sum_{i=1}^d \left(p_\theta(i) - p_{\theta_0}(i) + \frac{1}{d}\right)^2 = \left\|p_\theta - p_{\theta_0}\right\|_2^2 + \frac{1}{d}$$

3. By the above we have $\left\|p_\theta\right\|_2^2 \geq \frac{1}{d}$. Substituting this into the formula we have

$$\begin{aligned}\left\|p_\theta\right\|_3^3 - \left\|p_\theta\right\|_2^4 \leq & \left\|p_\theta\right\|_3^3 - \frac{1}{d^2} \\ = & \sum_{i=1}^d p_\theta(i)^3 - \frac{1}{d^2} \\ = & \sum_{i=1}^d \left(p_\theta(i) - p_{\theta_0}(i) + \frac{1}{d}\right)^3 - \frac{1}{d^2} \\ = & \left\|p_\theta - p_{\theta_0}\right\|_3^3 + \frac{3}{d}\left\|p_\theta - p_{\theta_0}\right\|_2^2 \\ \leq & \left\|p_\theta - p_{\theta_0}\right\|_2^3 + \frac{3}{d}\left\|p_\theta - p_{\theta_0}\right\|_2^2 \qquad (2)\end{aligned}$$

where in the last step we utilize the relation between ℓ_2 and ℓ_3 norms.

■

Putting the three together we have the desired upper bound that:

$$\begin{aligned} var_\theta(T) &\lesssim \frac{\|p_\theta\|_2^2}{n^2} + \frac{\|p_\theta\|_3^3 - \|p_\theta\|_2^4}{n} \\ &\lesssim \frac{\|p_\theta - p_{\theta_0}\|_2^2 + \frac{1}{d}}{n^2} + \frac{\|p_\theta - p_{\theta_0}\|_2^3 + \frac{3}{d}\|p_\theta - p_{\theta_0}\|_2^2}{n} \end{aligned}$$

combined with the condition for $\mathbb{E}_\theta[T]$, the optimal optimal rate of $\|p_\theta - p_{\theta_0}\|_2^2$ should be chosen s.t.

$$\begin{aligned} \|p_\theta - p_{\theta_0}\|_2^2 &\gtrsim \frac{1}{n\sqrt{d}} + \sqrt{\frac{\|p_\theta - p_{\theta_0}\|_2^2 + \frac{1}{d}}{n^2} + \frac{\|p_\theta - p_{\theta_0}\|_2^3 + \frac{3}{d}\|p_\theta - p_{\theta_0}\|_2^2}{n}} \\ &\asymp \frac{1}{n\sqrt{d}} + \frac{\|p_\theta - p_{\theta_0}\|_2}{n} + \frac{\|p_\theta - p_{\theta_0}\|_2^{3/2}}{\sqrt{n}} + \frac{\|p_\theta - p_{\theta_0}\|_2}{\sqrt{nd}} \\ \Rightarrow \|p_\theta - p_{\theta_0}\|_2 &\gtrsim \max\left\{\frac{1}{n}, \frac{1}{nd}, \frac{1}{n^{1/2}d^{1/4}}\right\} \end{aligned}$$

combined with the relation between ℓ_2 and ℓ_1 norms that $\|\cdot\|_2 \geq \|\cdot\|_1/\sqrt{d}$, we get the condition that $\|\cdot\|_1$ (i.e. ε) should satisfy:

$$\sqrt{d}\,\|p_\theta - p_{\theta_0}\|_2^2 \geq \|p_\theta - p_{\theta_0}\|_1 \geq \varepsilon \geq \max\left\{\frac{d^{1/2}}{n}, \frac{1}{nd^{1/2}}, \frac{d^{1/4}}{n^{1/2}}\right\} \tag{0.1}$$

Note that we have a trivial bound that $\|p_\theta - p_{\theta_0}\|_1 \leq 2 = \Theta(1)$, so the term that would eventually take effect in equation 0.1 is the term $\frac{d^{1/4}}{n^{1/2}}$, which gives that optimal rate:

$$\varepsilon \gtrsim \frac{d^{1/4}}{n^{1/2}}$$

0.1.3 Lower Bound Side

For lower bound side, we conversely consider that we have lower bound of type I and type II error, which suffices to upper bound the total variation distance d_{TV} noticing the following relation:

$$\mathbb{P}_0\left(\hat{\psi}_n = 1\right) + \sup_{p_\theta \in H_1} \mathbb{P}_\theta\left(\hat{\psi}_n = 0\right) \geq (1 - d_{\mathrm{TV}}(p_\theta, p_{\theta_0})) \gtrsim \text{const} \Rightarrow d_{\mathrm{TV}}(p_\theta, p_{\theta_0}) \leq c < 1$$

Note that by Jensen's inequality we have relation $d_{\text{TV}} \leq \frac{1}{2}\sqrt{\chi^2}$ so it suffices to upper bound the χ^2 divergence as

$$\chi^2(p_{\theta_0}^{\otimes n}, p_\theta^{\otimes n}) \lesssim c$$

with $\theta_0 \sim \pi_{\theta_0}$, $\theta \sim \pi_\varepsilon$.

We construct the following priors (in which WLOG we take d to be even, if not using $(d+1)/2$ and the magnitude should be the same):

$$\begin{aligned}
\pi_{\theta_0} :=& \text{dirac}(p_{\theta_0}) \\
\pi_\varepsilon =:& \text{Unif}(\left\{ P_\zeta : \; p_\zeta(i) = \frac{1 + (-1)^i \zeta_{\lceil i/2 \rceil} \cdot 3\varepsilon}{d} \right\}_{\zeta \in \{\pm 1\}^{d/2}})
\end{aligned}$$

Remark: i.e. π_ε is the uniform distribution over $\{p_{\theta_\zeta}(i)\}$ vectors looks like:

$$\underset{d\times 1}{p_{\theta_\zeta}} = \frac{1}{d} + \frac{3\varepsilon}{d}\big(\underbrace{+1, -1}_{\text{pair 1}}, \underbrace{-1, +1}_{\text{pair 2}}, \dots, \underbrace{+1, -1}_{\text{pair d/2}}\big)$$

in which each "pair" can only take $(+1, -1)$ or $(-1, +1)$. This construction ensures that $\left\| p_{\theta_\zeta} - p_{\theta_0} \right\|_1 = 3\varepsilon = \Theta(\varepsilon)$.

Then:

$$\begin{aligned}
\chi^2(\mathbb{E}_{\theta \sim \pi_\varepsilon}\left[p_\theta^{\otimes n}\right] \| p_{\theta_0}^{\otimes n}) + 1 \overset{(i)}{=}& \mathbb{E}_{\zeta, \tilde{\zeta} \sim \text{Unif}(\{\pm\}^{d/2})}\left[\mathbb{E}_{X_1^n \overset{i.i.d.}{\sim} p_{\theta_0}}\left[\frac{\mathbb{P}_\zeta^{\otimes n}\mathbb{P}_{\tilde{\zeta}}^{\otimes n}}{(\mathbb{P}_{\theta_0}^{\otimes n})^2}\right]\right] \\
\overset{(ii)}{=}& \mathbb{E}_{\zeta, \tilde{\zeta} \sim \text{Unif}(\{\pm\}^{d/2})}\left[\underbrace{\mathbb{E}_{X \sim p_{\theta_0}}\left[\frac{\mathbb{P}_\zeta \mathbb{P}_{\tilde{\zeta}}}{(\mathbb{P}_{\theta_0})^2}\right]}_{(*)}{}^n\right]
\end{aligned}$$

where (i) according to Ingster-Suslina's method, (ii) is due to the tensorization property of χ^2 divergence: $\chi^2(\prod_{i=1}^n P_i \| \prod_{i=1}^n Q_i) + 1 = \prod_{i=1}^n (\chi^2(P_i \| Q_i) + 1)$. Now we turn to the term $(*)$, which can be further computed as:

$$\begin{aligned}
(*) = \mathbb{E}_{X \sim p_{\theta_0}}\left[\frac{\mathbb{P}_\zeta \mathbb{P}_{\tilde{\zeta}}}{(\mathbb{P}_{\theta_0})^2}\right] =& \mathbb{E}_{X \sim p_{\theta_0}}\left[\frac{(\mathbb{P}_\zeta(x) - \mathbb{P}_{\theta_0}(x))(\mathbb{P}_{\tilde{\zeta}}(x) - \mathbb{P}_{\theta_0}(x))}{\mathbb{P}_{\theta_0}^2(x)} + 1\right] \\
=& \sum_{x=1}^d \frac{(\mathbb{P}_\zeta(x) - \mathbb{P}_{\theta_0}(x))(\mathbb{P}_{\tilde{\zeta}}(x) - \mathbb{P}_{\theta_0}(x))}{1/d^2} + 1 \\
=& \sum_{x=1}^d \frac{(-1)^x \zeta_{\lceil x/2 \rceil} \cdot 3\varepsilon}{d} \cdot \frac{(-1)^x \tilde{\zeta}_{\lceil x/2 \rceil} \cdot 3\varepsilon}{d} \cdot d + 1 \\
=& \frac{18\varepsilon^2}{d}\zeta'\tilde{\zeta} + 1
\end{aligned}$$

substituting this back to the previous equation we have:

$$
\begin{aligned}
\chi^2(\mathbb{E}_{\theta\sim\pi_\varepsilon}\left[p_\theta^{\otimes n}\right] \| p_{\theta_0}^{\otimes n}) + 1 =& \mathbb{E}_{\zeta,\tilde{\zeta}\sim\text{Unif}(\{\pm\}^{d/2})}\left[\underbrace{\mathbb{E}_{X\sim p_{\theta_0}}\left[\frac{\mathbb{P}_\zeta\mathbb{P}_{\tilde{\zeta}}}{(\mathbb{P}_{\theta_0})^2}\right]}_{(*)}{}^n\right] \\
=& \mathbb{E}_{\zeta,\tilde{\zeta}\sim\text{Unif}(\{\pm\}^{d/2})}\left[\left(\frac{18\varepsilon^2}{d}\zeta'\tilde{\zeta}+1\right)^n\right] \\
\leq& \mathbb{E}_{\zeta,\tilde{\zeta}\sim\text{Unif}(\{\pm\}^{d/2})}\left[\exp\left[\frac{18n\varepsilon^2}{d}\zeta'\tilde{\zeta}\right]\right] \\
=& \prod_{i=1}^{d/2}\mathbb{E}_{\zeta_i,\tilde{\zeta}_i'\sim\text{Unif}(\pm)}\left[\exp\left[\frac{18n\varepsilon^2}{d}\zeta_i\tilde{\zeta}_i\right]\right] \\
=& \cosh\left[\frac{18n\varepsilon^2}{d}\right]^{d/2} \\
\leq& \exp\left[\frac{162n^2\varepsilon^4}{d^2}\right]^{d/2} \\
=& \exp\left[\frac{81n^2\varepsilon^4}{d}\right] < c < \Theta(1)
\end{aligned}
$$

To ensure the condition we require $n^2\varepsilon^4/d \lesssim 1$, i.e.

$$\varepsilon \lesssim \frac{d^{1/4}}{n^{1/2}}$$

which is a matching lower bound to the upper bound side.

0.1.4 Conclusion

Thus we have the optimal rate of ε as:

$$\varepsilon^* \asymp \frac{d^{1/4}}{n^{1/2}}$$

or equivalently

$$n^* \asymp \frac{\sqrt{d}}{\varepsilon^2}$$

Remark: We would notice that this gives the same rate as gaussian location model ($\varepsilon^* \asymp d^{1/4}/n^{1/2}$), which is an interesting result.

0.2 Other Reference

An alternative proof see *Lecture notes on Information-theoretic methods for high-dimensional statistics* by Yihong Wu, Chapter 24.3, page 146.