Lecture 7

In this lecture we will explore some aspects of testing nonparametric hypothesis testing in the context of the density model.

In general a goodness of fit type testing problem can be described as follows. Consider observations $Z^{(n)}$ on a measurable space $(\Omega_n, \mathcal{A}_n)_{n\in\mathbb{N}}$. As usual the model for the distribution of $Z^{(n)}$ consists of probability laws $P_\theta^{(n)}$, $\theta \in \Theta$ for some index set Θ. A "statistical hypothesis" H_0 is a subset of Θ. A "statistical test" is a measurable function $\psi_n \equiv \psi_n(Z^{(n)}) : \Omega_n \to \{0,1\}$ that takes value $\psi_n = 0$ to accept H_0 & $\psi_n = 1$ to reject H_0, typically in favour of an alternative hypothesis, H_1 → which is again a subset of Θ disjoint from H_0.

The performance of a test is naturally measured by the sum of its 'Type I' and 'Type II' errors. We formalize this below when (Θ, d) is a metric space.

Definition: Let r_n, ρ_n be sequences of non negative real numbers and let $H_0 \subset \Theta$ be a statistical (null) hypothesis. Also let d be a metric on Θ.
The sequence $\{\rho_n\}_{n\in\mathbb{N}}$ is called the minimax d-separation rate for testing the hypotheses

$\theta \in H_0$ vs $\theta \notin H_1 = H_1(d, r_n) = \{\theta \in \Theta, \inf_{\theta' \in H_0} d(\theta, \theta') \geq r_n\}$

if both the following hold:

(i) For every $\alpha > 0$, there exists a test ψ_n such that for every $n \in \mathbb{N}$ large enough (might be depending on (H))

$$\sup_{\theta \in H_0} \mathbb{E}_{\theta}^{(n)} \psi_n + \sup_{\theta \in H_1(d, \rho_n)} \mathbb{E}_{\theta}^{(n)} (1 - \psi_n) \leq \alpha$$

(ii) For any sequence $r_n = o(\rho_n)$ (i.e $\frac{r_n}{\rho_n} \to 0$ as $n \to \infty$) one has

$$\liminf_{n \to \infty} \inf_{\psi_n} \left[\sup_{\theta \in H_0} \mathbb{E}_{\theta}^{(n)} \psi_n + \sup_{\theta \in H_1(d, r_n)} \mathbb{E}_{\theta}^{(n)} (1 - \psi_n) \right] > 0$$

where the infimum is taken over all measurable functions $\psi_n : \Omega_n \to \{0, 1\}$ of $\mathcal{Z}^{(n)}$.

Remark: In a decision theoretic language, we can define a risk of test for testing H_0 against $H_1(d, r_n)$ as

$$\text{Risk}(\psi_n, H_0, H_1(d, r_n)) = \sup_{\theta \in H_0} \mathbb{E}_{\theta}^{(n)} \psi_n + \sup_{\theta \in H_1(d, r_n)} \mathbb{E}_{\theta}^{(n)} (1 - \psi_n)$$

Then minimax d-separation rate looks for sequence $\{\rho_n\}_{n \geq 1}$ such that $\exists \psi_n$ with Risk $\to 0$ while for any $r_n \ll \rho_n$ Risk $(\psi_n, H_0, H_1(d, r_n))$ stays bounded away from 0 for any test ψ_n

Remark: If the liminf in (ii) equals 1, then H_0 and H_1 are called asymptotically indistinguishable. Also note that the theory of nonparametric hypothesis testing that one can develop in the context of the definition above depends on the space (Θ, d). Finding the correct "ρ_n" is often problem specific. However, a standard way of getting started on proving (ii) for a $r_n \ll \rho_n$ with a conjectured ρ_n is as follows.

Strategy for showing (ii)

For any $\theta_0 \in H_0$ and $\Theta_M = \{\theta_1, \dots, \theta_M\} \subseteq H_1(d, r_n)$ the worst case error for testing against $\{\theta_0\}$ vs Θ_M is definitely smaller than that between H_0 and H_1. Therefore if one cannot test between $\{\theta_0\}$ vs Θ_M then one cannot test between H_0 and H_1. But this reduced problem is similar to a multiple testing problem that is well understood. In particular, fix any $\eta \in (0, 1)$. Then for any test ψ_n,

$$\mathbb{E}_{P_{\theta_0}^{(n)}}(\psi_n) + \sup_{\theta \in \Theta_M} \mathbb{E}_{P_\theta^{(n)}}(1-\psi_n)$$

$$\geq \mathbb{E}_{P_{\theta_0}^{(n)}}(\psi_n) + \int_{\Theta_M} \mathbb{E}_{P_\theta^{(n)}}(1-\psi_n)\, d\pi(\theta) \quad \text{for any prior } \pi \text{ on } \Theta_M.$$

$$= \mathbb{E}_{P_{\theta_0}^{(n)}}(\psi_n) + \mathbb{E}_{P_{\theta_0}^{(n)}}\left[(1-\psi_n)\left(\int_{\Theta_M} \frac{dP_\theta^{(n)}}{dP_{\theta_0}^{(n)}}\, d\pi(\theta)\right)\right] \quad \text{if } P_\theta^{(n)} \ll P_{\theta_0}^{(n)} \ \forall\, \theta \in \Theta_M$$

(Λ = the inner integral)

$$\geq \mathbb{E}_{P_{\theta_0}^{(n)}}(\psi_n) + \mathbb{E}_{P_{\theta_0}^{(n)}}\left[(1-\psi_n)(1-\eta)\, I(\Lambda \geq 1-\eta)\right]$$

$$\geq (1-\eta)\, P_{\theta_0^{(n)}}(\Lambda \geq 1-\eta) \geq \left(1 - \frac{E_{P_{\theta_0^{(n)}}} |\Lambda - 1|}{\eta}\right)(1-\eta)$$

$$\geq \left(1 - \frac{\sqrt{E_{P_{\theta_0^{(n)}}} (\Lambda-1)^2}}{\eta}\right)(1-\eta)$$

$$= \left(1 - \sqrt{\chi^2\left(P_\pi^{(n)}, P_{\theta_0^{(n)}}\right)/\eta}\right)(1-\eta) \quad (*)$$

where $\chi^2(P_\pi^{(n)}, P_{\theta_0^{(n)}}) = \int \left(\frac{dP_\pi^{(n)}}{dP_{\theta_0^{(n)}}}\right)^2 dP_{\theta_0}^{(n)} - 1$

and $\frac{P(A)}{\pi(A)} = \int \frac{dP}{P}$ for any measurable set $A \subseteq \Omega_n$,

$P_\pi^{(n)}(A) = \int_{\Theta_M} P_\theta^{(n)}(A)\, d\pi(\theta)$ is the convex hull of the probability measures $P_{\theta_1}^{(n)}, \ldots, P_{\theta_M}^{(n)}$ with the convex combination determined by π.

The idea is now obvious. Choose $\{\theta_1, \ldots, \theta_M\} \subseteq H_1(d, r_n)$ such that $\chi^2(P_\pi^{(n)}, P_{\theta_0^{(n)}})$ stays bounded. Indeed one naturally often takes $\theta_1, \ldots, \theta_M$ such that $d(\theta_0, \theta_j) = r_n$ $j = 1, \ldots, M$ to make the alternatives the most difficult possible. Such a Θ_M is referred to as asymptotically least favorable.

Note that $(*)$ also implies the following proposition.

Proposition: For any $\Theta_M \subseteq H_1(d_n, r_n)$ and $\theta_0 \in H_0$ and $\eta \in (0,1)$ and any prior π on Θ_M

$$\inf_{\psi_n} \left[\sup_{\theta \in H_0} \mathbb{E}_{P_\theta^{(n)}} \psi_n + \sup_{\theta \in H_1(d_n, r_n)} \mathbb{E}_{P_\theta^{(n)}} (1 - \psi_n) \right]$$

$$\geq \inf_{\psi_n} \left[\mathbb{E}_{P_{\theta_0}^{(n)}} \psi_n + \sup_{\theta \in \Theta_M} \mathbb{E}_{P_\theta^{(n)}} (1 - \psi_n) \right]$$

$$\geq (1-\eta) \left[1 - \sqrt{\chi^2(P_\pi^{(n)}, P_{\theta_0}^{(n)})/\eta} \right]$$

where $P_\pi^{(n)}(A) = \int_{\Theta_M} P_\theta^{(n)}(A)\, d\pi(\theta)$

Remark: Indeed the above calculation reduces the problem to simple hypothesis listing of $H_0 = \{P_{\theta_0}^{(n)}\}$ vs $H_1 = \{P_\pi^{(n)}\}$ and by Neyman-Pearson lemma, the test that minimizes $P_{\theta_0}^{(n)} \psi_n + P_\pi^{(n)}(1 - \psi_n)$ over all tests ψ_n is δ given by the procedure that rejects when

$$\Lambda(Z^{(n)}) := \int_{\Theta_M} \frac{dP_\theta^{(n)}}{dP_{\theta_0}^{(n)}} d\pi(\theta) > 1 \qquad \text{(check this fact)}$$

The total error of this test can now be bounded controlled as above.

With this background we are ready to consider the goodness of fit testing problem in the density model on $([0,1], \mathcal{B}([0,1]))$.

Goodness of Fit Testing in L_2 for the density Model

Let $X_1, X_2, \ldots, X_n$ be iid samples from a density η on $\left([0,1], \mathcal{B}[0,1]\right)$. Consider the hypothesis testing problem

$$H_0: \eta \equiv 1 \quad \text{vs} \quad H_1 := \{\eta \in \mathcal{P}(\alpha, M, B) : \|\eta - 1\|_2 \geq r_n\}$$

where $\|\cdot\|_2$ stands for the L_2 norm on $[0,1]$. Problems of this kind is often called a goodness of fit problem since we are trying to understand whether the uniform distribution on $[0,1]$ fits the data better compared to set of "rougher" densities.

Let us put ourselves in the context of the general discussion above. Here $Z^{(n)} = (X_1, \ldots, X_n)$, $\Omega_n = [0,1]^n$ and $\mathcal{A}_n = \mathcal{B}[0,1]^n$. Take $\Theta = \mathcal{P}(\alpha, M, B)$ with $B > 1$, and for any $\theta \in \Theta$ identify θ with η. Then $P_\eta^{(n)}(A) = \int_A \prod_{i=1}^n \eta(x_i)\, dx_i$ for any $\eta \in \Theta$.
Also consider $\Theta \subseteq L_2[0,1]$ and thus $(\Theta, \|\cdot\|_2)$ is a metric space
The testing problem is then,

$$H_0 = \{\eta_0 \equiv 1\} \quad \text{vs} \quad H_1(d, r_n) := \{\eta \in \Theta : \inf_{\eta' \in H_0} d(\eta, \eta') = \|\eta' - 1\|_2 \geq r_n\}$$

Therefore to solve this problem we need to address both (i) (Upper Bound) and (ii) (Lower Bound) of our Definition of minimax optimal d-separation.