Stat 301

$$X_j = \theta_j + \frac{1}{\sqrt{n}} z_j \,, \quad z_j \overset{iid}{\sim} N(0,1). \qquad \Theta_\alpha(R) = \left\{ \theta : \sum_j j^{2\alpha} \theta_j^2 \le R^2 \right\}.$$

$$\sup_{\theta \in \Theta_\alpha(R)} \mathbb{E}_\theta \|\hat{\theta} - \theta\|^2 \le C n^{-\frac{2\alpha}{2\alpha+1}}$$

Question: can we show

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta_\alpha(R)} \mathbb{E}_\theta \|\hat{\theta} - \theta\|^2 \ge c\, n^{-\frac{2\alpha}{2\alpha+1}}$$

Hypothesis Test.

two-point test (simple vs simple)

$$H_0: X \sim P \qquad\qquad H_1: X \sim Q$$

testing function $\phi: X \mapsto \{0, 1\}$

Type-I error $\quad P\phi = \mathbb{E}_{X\sim P}\phi(X)$

Type-II error $\quad Q(1-\phi)$

testing error $\quad P\phi + Q(1-\phi)$

optimal testing error $\quad \inf_{\phi}\left(P\phi + Q(1-\phi)\right).$

def: total variation distance.

$$TV(P,Q) = \sup_{B} |P(B) - Q(B)|.$$

Theorem: $TV(P,Q) = P(p(X) > q(X)) - Q(p(X) > q(X))$

$$= \frac{1}{2}\int |p-q| = 1 - \underbrace{\int \min(p,q)}_{\text{total variantion affinity}}$$

Proof: $A = \{ p(x) > q(x) \}$,

$$TV(P,Q) \geq P(A) - Q(A).$$

$$\forall B \quad |P(B) - Q(B)| = \left| \int_B (p-q) \right|.$$

$$= \left| \int_{B\cap A} (p-q) + \int_{B\cap A^c} (p-q) \right|$$

$$= \left| \int_{B\cap A} (p-q) - \int_{B\cap A^c} (q-p) \right|$$

$$\leq \max\left(\int_{B\cap A} (p-q), \int_{B\cap A^c} (q-p) \right).$$

$$\leq \max\left(\int_A (p-q), \int_{A^c} (q-p) \right)$$

$$= \max\left(P(A) - Q(A), Q(A^c) - P(A^c) \right).$$

$$= P(A) - Q(A).$$

take sup over all B $\Rightarrow$ $TV(P, Q) = \sup_B |P(B) - Q(B)|$

$$\leq P(A) - Q(A).$$

$$\Rightarrow TV(P, Q) = P(A) - Q(A).$$

$$\frac{1}{2}\int |p-q| = \frac{1}{2}\int_{p>q} |p-q| + \frac{1}{2}\int_{p\leq q} |p-q|$$

$$= \frac{1}{2}\int_{p>q} (p-q) + \frac{1}{2}\int_{p\leq q} (q-p).$$

$$= \frac{1}{2}\left(P(A) - Q(A)\right) + \frac{1}{2}\left(Q(A^c) - P(A^c)\right).$$

$$= \frac{1}{2}\left(P(A) - Q(A)\right) + \frac{1}{2}\left(P(A) - Q(A)\right)$$

$$= P(A) - Q(A) = TV(P,Q).$$

$$\int \min(p,q) = \int_{p>q} \min(p,q) + \int_{p\le q} \min(p,q).$$

$$= \int_{p>q} q + \int_{p\le q} p$$

$$= Q(A) + P(A^c) = Q(A) + 1 - P(A).$$

$$= 1 - (P(A) - Q(A)) = 1 - TV(P,Q) \qquad \square$$

Theorem (Neyman-Pearson lemma).

$$\inf_{\phi} \left(P\phi + Q(1-\phi)\right) = 1 - TV(P,Q) = \int \min(p,q)$$

likelihood ratio test

the optimal ϕ is $\phi(x) = \mathbb{1}\{p(x) < q(x)\}$

Proof: $\forall \phi \quad P\phi + Q(1-\phi) = \int p\phi + \int q(1-\phi)$

$$= \int p\phi + q(1-\phi) \geq \int \min(p,q).$$

$$\Rightarrow \inf_{\phi}\left(P\phi + Q(1-\phi)\right) \geq \int \min(p,q).$$

$$\inf_{\phi}\left(P\phi + Q(1-\phi)\right) \leq P(p<q) + Q(p \geq q)$$

$$= 1 - TV(P,Q) = \int \min(p,q)$$

$\square$

Le Cam two-point method.

$$\inf_{\hat\theta}\sup_{\theta\in\Theta} \mathbb{E}_\theta(\hat\theta-\theta)^2 \geq \inf_{\hat\theta}\sup_{\theta\in\{\theta_1,\theta_2\}} \mathbb{E}_\theta(\hat\theta-\theta)^2 \quad (\theta_1,\theta_2\in\Theta)$$

$$\geq \frac{(\theta_1-\theta_2)^2}{4}\int \min(p_{\theta_1},p_{\theta_2})$$

$$\inf_{\phi}\left(P_{\theta_1}\phi + P_{\theta_2}(1-\phi)\right)$$

Proof:

$$\inf_{\hat\theta}\sup_{\theta\in\{\theta_1,\theta_2\}} \mathbb{E}_\theta(\hat\theta-\theta)^2$$

$$\geq \inf_{\hat\theta}\left(\frac{1}{2}\mathbb{E}_{\theta_1}(\hat\theta-\theta_1)^2+\frac{1}{2}\mathbb{E}_{\theta_2}(\hat\theta-\theta_2)^2\right)$$

$$= \inf_{\hat\theta}\frac{1}{2}\int\left[(\hat\theta-\theta_1)^2 p_{\theta_1} + (\hat\theta-\theta_2)\, p_{\theta_2}\right]$$

$$\geq \inf_{\hat\theta}\frac{1}{2}\int\left[(\hat\theta-\theta_1)^2+(\hat\theta-\theta_2)^2\right]\min(p_{\theta_1},p_{\theta_2})$$

$$\geq \frac{1}{2} \int \frac{(\theta_1-\theta_2)^2}{2} \min(p_{\theta_1}, p_{\theta_2})$$

$(x+y)^2 \leq 2x^2 + 2y^2$

$(\theta_1-\theta_2)^2 = (\theta_1 - \hat{\theta} + \hat{\theta} - \theta_2)^2$

$\leq 2(\hat{\theta}-\theta_1)^2 + 2(\hat{\theta}-\theta_2)^2.$

$$= \frac{(\theta_1-\theta_2)^2}{4} \int \min(p_{\theta_1}, p_{\theta_2})$$

$\square$.

e.g. $X_1, \dots, X_n \overset{iid}{\sim} N(\theta, 1)$. $\theta \in \mathbb{R}$.

$$\inf_{\hat{\theta}} \sup_{\theta\in\mathbb{R}} E_\theta(\hat{\theta}-\theta)^2 \geq \inf_{\hat{\theta}} \sup_{\theta\in\{\theta_1,\theta_2\}} E_\theta(\hat{\theta}-\theta)^2.$$

Choose $\theta_1 = 0$, $\theta_2 = \frac{1}{\sqrt{n}}$. $\quad P_\theta = N(\theta, 1)$

$$\inf_{\hat{\theta}} \sup_{\theta\in\{\theta_1,\theta_2\}} E_\theta(\hat{\theta}-\theta)^2 \geq \frac{1}{4n} \inf_{\phi} \left(P_0^n \phi + P_{\frac{1}{\sqrt{n}}}^n (1-\phi)\right).$$

$$\inf_{\phi}\left(P_0^n\phi + P_{\frac{1}{\sqrt{n}}}^n(1-\phi)\right) \geq P_0^n\left(P_0^n(x) < P_{\frac{1}{\sqrt{n}}}^n(x)\right).$$

$$= P_0^n\left(\prod_{i=1}^n \frac{e^{-\frac{1}{2}\left(X_i-\frac{1}{\sqrt{n}}\right)^2}}{e^{-\frac{1}{2}X_i^2}} > 1\right).$$

$$= P_0^n\left(\sum_{i=1}^n\left[\left(X_i-\frac{1}{\sqrt{n}}\right)^2 - X_i^2\right] < 0\right).$$

$$= P_0^n\left(\sum_{i=1}^n\left(-\frac{2}{\sqrt{n}}X_i+\frac{1}{n}\right) < 0\right).$$

$X_1,\cdots,X_n$ iid $N(0,1)$

$$= P_0^n\left(\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i}_{\sim N(0,1)} > \frac{1}{2}\right) = \mathbb{P}\left(N(0,1) > \frac{1}{2}\right)$$

$$\Rightarrow \inf_{\hat{\theta}}\sup_{\theta\in\mathbb{R}} \mathbb{E}(\hat{\theta}-\theta)^2 \geq \frac{P\left(N(0,1)>\frac{1}{2}\right)}{4}\cdot\frac{1}{n}.$$

e.g. $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Unif}(0, \theta) \quad \theta \in [0,1]$

$$\inf_{\hat\theta} \sup_{\theta\in[0,1]} \mathbb{E}_\theta (\hat\theta - \theta)^2 \asymp n^{-?}$$

upper bound: $\hat\theta = \max_{1\le i\le n} X_i$

$$F(t) = \mathbb{P}(\hat\theta \le t) = \prod_{i=1}^n \mathbb{P}(X_i \le t) = \left(\frac{t}{\theta}\right)^n$$

$$f(t) = \theta^{-n} n t^{n-1}, \quad t \in (0, \theta).$$

$$\mathbb{E}_\theta (\hat\theta - \theta)^2 = \int_0^\theta (t-\theta)^2 \theta^{-n} n t^{n-1} dt$$

$$= \theta^{-n} n \int_0^\theta (t^2 + \theta^2 - 2t\theta) t^{n-1} dt$$

$$= \theta^{-n} n \left(\left. \frac{t^{n+2}}{n+2} \right|_0^\theta + \theta^2 \left. \frac{t^n}{n} \right|_0^\theta - 2\theta \left. \frac{t^{n+1}}{n+1} \right|_0^\theta \right)$$

$$= \theta^{-n} n \left(\frac{\theta^{n+2}}{n+2} + \frac{\theta^{n+2}}{n} - \frac{2\theta^{n+2}}{n+1} \right).$$

$$= \theta^2 n \left(\frac{1}{n+2} - \frac{1}{n+1} + \frac{1}{n} - \frac{1}{n+1} \right).$$

$$= \theta^2 n \left(\frac{-1}{(n+2)(n+1)} + \frac{1}{n(n+1)} \right).$$

$$= \frac{\theta^2 n}{n+1} \left(\frac{1}{n} - \frac{1}{n+2} \right) = \frac{\theta^2 n}{n+1} \frac{2}{n(n+2)}$$

$$= \frac{2\theta^2}{(n+1)(n+2)}$$

$$\sup_{\theta\in[0,1]} \mathbb{E}_\theta(\hat\theta-\theta)^2 = \frac{2}{(n+1)(n+2)} = O(n^{-2}).$$

Lower bound: $\inf_{\hat\theta}\sup_{\theta\in[0,1]} \mathbb{E}_\theta(\hat\theta-\theta)^2 \geq \inf_{\hat\theta}\sup_{\theta\in\{\theta_1,\theta_2\}} \mathbb{E}_\theta(\hat\theta-\theta)^2$

$$\theta_1 = 1, \quad \theta_2 = 1-\frac{1}{n}. \qquad P_\theta = \mathrm{Unif}(0,\theta).$$

$$\inf_{\hat\theta}\sup_{\theta\in\{\theta_1,\theta_2\}} \mathbb{E}_\theta(\hat\theta-\theta)^2 \geq \frac{1}{4n^2}\int P_1^n \wedge P_{1-\frac{1}{n}}^n$$

inequality: $\int p\wedge q \geq \frac{1}{2}\Big(\underbrace{\int\sqrt{pq}}_{\text{Hellinger affinity}}\Big)^2.$

$$\int P_1^n \wedge P_{1-\frac{1}{n}}^n \geq \frac{1}{2}\left(\int\sqrt{P_1^n P_{1-\frac{1}{n}}^n}\right)^2$$

$$= \frac{1}{2}\left(\int\sqrt{\prod_{i=1}^n P_1(x_i)\prod_{i=1}^n P_{1-\frac{1}{n}}(x_i)}\right)^2$$

$$= \frac{1}{2}\left(\int \sqrt{p_1(x_1)\, p_{1-\frac{1}{n}}(x_1)}\right)^{2n}. \qquad p_\theta(x) = \frac{1}{\theta}\mathbb{1}\{x\in(0,\theta)\}$$

$$= \frac{1}{2}\left(\int_0^{\theta_2} \sqrt{\frac{1}{\theta_1}\cdot\frac{1}{\theta_2}}\right)^{2n}.$$

$$= \frac{1}{2}\left(\sqrt{\frac{\theta_2}{\theta_1}}\right)^{2n} = \frac{1}{2}\left(1-\frac{1}{n}\right)^{n} \geq \frac{1}{8}$$

$$\forall n \geq 2.$$

$$\Rightarrow \inf_{\hat\theta}\sup_{\theta\in[0,1]} E_\theta(\hat\theta-\theta)^2 \geq \frac{1}{32n^2}.$$