

# Solutions to Assignment 6

April 20, 2010

Solution 1,2 were obtained from Qirong and 3 was modified based on Stephane Ross.

1. (a) Let  $N = N_{\square}(\epsilon, \mathcal{F}, L_1(P))$ , and let  $C = \{(\ell_1, u_1), \dots, (\ell_N, u_N)\}$  be some  $\epsilon$ -cover for  $\mathcal{F}$  under  $L_1(P)$ . Then for any  $f \in \mathcal{F}$  there is a  $(\ell_j, u_j) \in C$  that brackets  $f$ , and therefore

$$\int |f(x) - \ell_j(x)| dP(x) \leq \int |u_j(x) - \ell_j(x)| dP(x) \leq \epsilon$$

It follows that

$$\begin{aligned} |P_n(f) - P(f)| &\leq |P_n(f) - P_n(\ell_j)| + |P_n(\ell_j) - P(\ell_j)| + |P(\ell_j) - P(f)| \\ &= |P_n(\ell_j) - P(\ell_j)| + \left| \int [f(x) - \ell_j(x)] dP_n(x) \right| + \left| \int [f(x) - \ell_j(x)] dP(x) \right| \\ &\leq |P_n(\ell_j) - P(\ell_j)| + \int |f(x) - \ell_j(x)| dP_n(x) + \int |f(x) - \ell_j(x)| dP(x) \\ &\leq |P_n(\ell_j) - P(\ell_j)| + \epsilon + 2\epsilon \quad (C \text{ is a } 2\epsilon\text{-cover for } \mathcal{F} \text{ under } L_1(P_n)) \\ &\leq \max_{(u_i, \ell_i) \in C} |P_n(\ell_i) - P(\ell_i)| + 3\epsilon \end{aligned}$$

implying

$$\sup_{f \in \mathcal{F}} |P_n(f) - P(f)| \leq \max_{(u_i, \ell_i) \in C} |P_n(\ell_i) - P(\ell_i)| + 3\epsilon$$

Hence for any  $\epsilon$ ,

$$\begin{aligned} \mathbb{P} \left( \sup_{f \in \mathcal{F}} |P_n(f) - P(f)| > 4\epsilon \right) &\leq \mathbb{P} \left( \max_{(u_j, \ell_j) \in C} |P_n(\ell_j) - P(\ell_j)| + 3\epsilon > 4\epsilon \right) \\ &= \mathbb{P} \left( \max_{(u_j, \ell_j) \in C} |P_n(\ell_j) - P(\ell_j)| > \epsilon \right) \\ &= \mathbb{P} \left( \bigcup_{(u_j, \ell_j) \in C} [|P_n(\ell_j) - P(\ell_j)| > \epsilon] \right) \\ &\leq \sum_{j=1}^{N_{\square}(\epsilon, \mathcal{F}, L_1(P))} \mathbb{P}(|P_n(\ell_j) - P(\ell_j)| > \epsilon) \quad (\text{union bound}) \\ &\leq 2N_{\square}(\epsilon, \mathcal{F}, L_1(P)) e^{-2n\epsilon^2/(B-(-B))^2} \\ &= 2N e^{-n\epsilon^2/2B^2} \end{aligned}$$

implying that  $\sup_{f \in \mathcal{F}} |P_n(f) - P(f)| \xrightarrow{P} 0$ .

- (b) Let  $a_1 < a_2 < \dots < a_k$  be such that  $\int_{-\infty}^{a_1} dP(x) = \int_{a_1}^{a_2} dP(x) = \dots = \int_{a_{k-1}}^{a_k} dP(x) = \epsilon$  and  $\int_{a_k}^{\infty} dP(x) \leq$

$\epsilon$ . It follows that  $k = \lceil \frac{1}{\epsilon} \rceil$ . Let  $D = \{(\ell_1, u_1), \dots, (\ell_{k+1}, u_{k+1})\}$ , where

$$\begin{aligned} (\ell_1, u_1) &= (0, I_{(-\infty, a_1]}) \\ (\ell_2, u_2) &= (I_{(-\infty, a_1]}, I_{(a_1, a_2]}) \\ (\ell_3, u_3) &= (I_{(a_1, a_2]}, I_{(a_2, a_3]}) \\ &\vdots \\ (\ell_k, u_k) &= (I_{(a_{k-2}, a_{k-1}]}, I_{(a_{k-1}, a_k]}) \\ (\ell_{k+1}, u_{k+1}) &= (I_{(a_{k-1}, a_k]}, I_{(a_k, \infty)}) \end{aligned}$$

From the definition of  $a_1, \dots, a_k$ , it follows that every  $(\ell_i, u_i) \in D$  is an  $\epsilon$  bracket in  $L_1(P)$ . Furthermore,  $D$  covers  $\mathcal{F}$  under  $L_1(P)$ . Therefore  $N_{[]}(\epsilon, \mathcal{F}, L_1(P)) \leq k + 1 = \frac{C}{\epsilon}$ , where

$$\begin{aligned} \frac{C}{\epsilon} &= k + 1 \\ \frac{C}{\epsilon} &= \left\lceil \frac{1}{\epsilon} \right\rceil + 1 \\ C &= \epsilon \left( \left\lceil \frac{1}{\epsilon} \right\rceil + 1 \right) > 0 \end{aligned}$$

(c) Observe that

$$\begin{aligned} F(t) &= \mathbb{P}(X < t) \\ &= \int I_{(-\infty, t]}(x) dP(x) \\ &= \int f_t(x) dP(x) \\ &= P(f_t) \end{aligned}$$

and

$$\begin{aligned} \hat{F}_n(t) &= \frac{1}{n} \sum_{i=1}^n I(X_i < t) \\ &= \mathbb{P}_n(X < t) \\ &= \int I_{(-\infty, t]}(x) dP_n(x) \\ &= \int f_t(x) dP_n(x) \\ &= P_n(f_t) \end{aligned}$$

where  $f_t = I_{(-\infty, t]}$ . Therefore

$$\sup_t \left| \hat{F}_n(t) - F(t) \right| = \sup_{f_t \in \mathcal{F}} |P_n(f_t) - P(f_t)|$$

Now, from part (b)  $D$  is an  $\epsilon$ -cover for  $\mathcal{F} = \{f_t\}$ . Then by part (a),

$$|P_n(f_t) - P(f_t)| \leq \max_{(u_i, \ell_i) \in D} |P_n(\ell_i) - P(\ell_i)| + 3\epsilon$$

for all  $f_t \in \mathcal{F}$ , and therefore

$$\sup_{f_t \in \mathcal{F}} |P_n(f_t) - P(f_t)| \leq \max_{(u_i, \ell_i) \in D} |P_n(\ell_i) - P(\ell_i)| + 3\epsilon$$

Hence

$$\begin{aligned}
\mathbb{P}\left(\sup_t \left|\hat{F}_n(t) - F(t)\right| > 4\epsilon\right) &= \mathbb{P}\left(\sup_{f_t \in \mathcal{F}} |P_n(f_t) - P(f_t)| > 4\epsilon\right) \\
&\leq \mathbb{P}\left(\max_{(u_i, \ell_i) \in D} |P_n(\ell_i) - P(\ell_i)| + 3\epsilon > 4\epsilon\right) \\
&= \mathbb{P}\left(\max_{(u_i, \ell_i) \in D} |P_n(\ell_i) - P(\ell_i)| > \epsilon\right) \\
&= \mathbb{P}\left(\bigcup_{(u_i, \ell_i) \in C} [|P_n(\ell_i) - P(\ell_i)| > \epsilon]\right) \\
&\leq \sum_{i=1}^{N_{\square}(\epsilon, \mathcal{F}, L_1(P))} \mathbb{P}(|P_n(\ell_i) - P(\ell_i)| > \epsilon) \quad (\text{union bound}) \\
&\leq 2N_{\square}(\epsilon, \mathcal{F}, L_1(P)) e^{-2n\epsilon^2/4} \quad (\|\ell_i\|_{\infty} \leq B = 1) \\
&\leq \frac{C}{\epsilon} e^{-n\epsilon^2/2} \\
&\rightarrow 0
\end{aligned}$$

as  $n \rightarrow \infty$ . This implies that  $\sup_t \left|\hat{F}_n(t) - F(t)\right| \xrightarrow{P} 0$ .

2. Markov's inequality gives us the following two bounds on  $P(X > \delta)$ :

$$\begin{aligned}
P(X > \delta) = P(X^k > \delta^k) &\leq \frac{\mathbb{E}[|X|^k]}{\delta^k} \\
P(X > \delta) = P(e^{\lambda X} > e^{\lambda\delta}) &\leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda\delta}}
\end{aligned}$$

where  $k$  and  $\lambda$  are variational parameters to be minimized. Let  $C = \inf_{k=0,1,2,\dots} \frac{\mathbb{E}[|X|^k]}{\delta^k}$ , so for all  $k$  we have

$$\begin{aligned}
C &\leq \frac{\mathbb{E}[|X|^k]}{\delta^k} \\
\delta^k C &\leq \mathbb{E}[|X|^k]
\end{aligned}$$

Thus for any  $\lambda \geq 0$ ,

$$\begin{aligned}
\frac{\lambda^k}{k!} \delta^k C &\leq \frac{\lambda^k}{k!} \mathbb{E}[|X|^k] \\
\sum_{i=0}^{\infty} \frac{\lambda^i \delta^i}{i!} C &\leq \sum_{i=0}^{\infty} \frac{\lambda^i \mathbb{E}[|X|^i]}{i!} \\
\inf_{k=0,1,2,\dots} \frac{\mathbb{E}[|X|^k]}{\delta^k} = C &\leq \frac{\mathbb{E}\left[\sum_{i=0}^{\infty} \frac{\lambda^i |X|^i}{i!}\right]}{\sum_{i=0}^{\infty} \frac{\lambda^i \delta^i}{i!}} = \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda\delta}}
\end{aligned}$$

proving that  $\inf_{k=0,1,2,\dots} \frac{\mathbb{E}[|X|^k]}{\delta^k} \leq \inf_{\lambda \geq 0} \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda\delta}}$  as desired.

The result suggests that the Chernoff bound  $P(X > \delta) \leq \inf_{\lambda \geq 0} \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda\delta}}$  is not tight. In particular, the other bound  $P(X > \delta) \leq \inf_{k=0,1,2,\dots} \frac{\mathbb{E}[|X|^k]}{\delta^k}$  is tighter.

3. (a) The minimax risk is defined by

$$\inf_{\hat{\theta}} \max_{\theta \in \{a, -a\}} \mathbb{E}_{Y \sim N(\theta, 1)}(L(\theta, \hat{\theta}(Y))) = \inf_{\hat{\theta}} \max_{\theta \in \{a, -a\}} P(\hat{\theta}(Y) \neq \theta | \theta)$$

By Le Cam's Bound, this is lower bounded by  $\frac{\|P_1 \wedge P_2\|}{2} = \frac{1}{2} \int \min(p_1(x), p_2(x)) dx$  where here  $p_1$  is the normal density for  $N(a, 1)$  and  $p_2$  is the normal density for  $N(-a, 1)$ . For such  $p_1, p_2$ , and assuming  $a > 0$ , then  $p_1(x) < p_2(x)$  for  $x < 0$  and  $p_2(x) < p_1(x)$  for  $x > 0$  so we have:

$$\begin{aligned} & \int_{-\infty}^{\infty} \min(p_1(x), p_2(x)) dx \\ &= \int_{-\infty}^0 p_1(x) dx + \int_0^{\infty} p_2(x) dx \\ &= P(X \leq 0 | a) + (1 - P(X \leq 0 | -a)) \\ &= \Phi(-a) + (1 - \Phi(a)) \\ &= 2\Phi(-a) \end{aligned}$$

where  $\Phi(x)$  is the cumulative density function of the standard normal distribution. Hence we obtain that the minimax risk is lower bounded by  $\Phi(-a)$ .

We can show that  $\Phi(-a)$  is exactly the expression for the minimax risk by showing that there exist an estimator that achieves this minimax risk. Consider the estimator  $\hat{\theta}(Y) = a$  if  $Y \geq 0$ ,  $\hat{\theta}(Y) = -a$  otherwise (again assuming  $a > 0$ ). Then for this estimator we have:

$$\begin{aligned} & \max_{\theta \in \{-a, a\}} \mathbb{E}_{Y \sim N(\theta, 1)}(L(\theta, \hat{\theta}(Y))) \\ &= \max_{\theta \in \{-a, a\}} P(\hat{\theta}(Y) \neq \theta | \theta) \\ &= \max(1 - \Phi(a), \Phi(-a)) \\ &= \max(\Phi(-a), \Phi(-a)) \\ &= \Phi(-a) \end{aligned}$$

Hence we conclude the minimax risk for this problem is exactly  $\Phi(-a)$ .

- (b) Here we can apply directly Fano's inequality to get a lower bound on the minimax risk. Fano's inequality states that if  $\max_{j \in \{1, 2, \dots, d\}} P_{\xi_j}(Z \neq j) \geq 1 - \frac{\beta + \log 2}{\log d}$  for any estimator  $Z$  of which density in  $\xi_1, \dots, \xi_d$  generated the observation, where  $\beta = \max_{j \neq k} K(P_{\xi_j}, P_{\xi_k})$  for  $K$  the Kullback-Leibler distance. Here for any  $j \neq k$ , we have:

$$\begin{aligned} & K(P_{\xi_j}, P_{\xi_k}) \\ &= \frac{1}{2} \int \frac{1}{(2\pi)^{d/2}} \exp(-(x - \xi_j)^T(x - \xi_j)/2) [(x - \xi_k)^T(x - \xi_k) - (x - \xi_j)^T(x - \xi_j)] dx \\ &= a \int \frac{1}{(2\pi)^{d/2}} \exp(-(x - \xi_j)^T(x - \xi_j)/2) [x_j - x_k] \\ &= a \mathbb{E}_{\xi_j}(x_j - x_k) \\ &= a^2 \end{aligned}$$

So we conclude  $\beta = a^2$  and hence that the minimax risk  $\inf_{\hat{\xi}} \max_{\xi \in \Theta} P_{\xi}(\hat{\xi} \neq \xi) \geq 1 - \frac{a^2 + \log 2}{\log d}$ . If we pick  $a = \sqrt{\frac{\log(d/4)}{2}}$  (for  $d > 4$ ) then the minimax risk is at least  $\frac{1}{2}$ .

Since  $a > 0$ , an obvious estimator is to pick  $\hat{\xi}(X) = \xi_j$  where  $j = \arg \max_j X_j$ . We claim it is minimax estimator since the risk is constant w.r.t.  $\theta$  by symmetry and it is easy to show that  $\hat{\xi}(X) = \xi_j$   $j = \arg \max_j X_j$  is Bayes Estimator with respect to uniform prior.