# 1: Convex compact parameter space

Let $\mathcal{P} = \left\{ P_\theta : \theta \in \Theta \right\}$ be a statistical model with $\Theta \subseteq \mathbb{R}^d$ a convex compact set, and $\Theta$ is not a singleton ($\Theta$ contains at least two points). Let $\Theta^\varepsilon = \left\{ \theta : d(\theta, \Theta) \leq \varepsilon \right\}$, where $d(\theta, \Theta) \equiv \inf \left\{ v \in \Theta : \|v - \theta\|_2 \right\}$.

(a) Consider the case of square loss $L(\hat{\theta}, \theta) = \|\hat{\theta} - \theta\|_2^2$. Assume that (for some $\varepsilon, \delta > 0$) $P_\theta(\hat{\theta}(X) \not\in \Theta^\varepsilon) > \delta$ for all $\theta \in \Theta$. Prove that $\hat{\theta}(\cdot)$ cannot be minimax optimal.

(b) Keeping to the square loss, consider now the linear model $P_\theta = N(D\theta, \sigma^2 I_n)$, where $D \in \mathbb{R}^{n \times d}$ is a known design matrix, of rank $d$, and $\sigma^2 > 0$ is known noise variance. Prove that no affine estimator (i.e. no estimator of the form $\hat{\theta}(y) = My + \theta_0$) can be minimax optimal.

(c) Produce a counter-example showing that the conclusion at point (a) does no longer hold if $\Theta$ is not convex.

(d) Consider the case $d = 1$, $\Theta = [\theta_{\text{min}}, \theta_{\text{max}}]$, and assume that $L$ is continuous, with $a \mapsto L(a, \theta)$ is strictly decreasing for $a < \theta$, and strictly increasing for $a > \theta$. Assume that (for some $\varepsilon, \delta > 0$) $P_\theta(\hat{\theta}(X) \not\in \Theta^\varepsilon) > \delta$ for all $\theta \in \Theta$, and that the risk function $\theta \mapsto R(\hat{\theta}; \theta)$ is continuous. Prove that $\hat{\theta}$ cannot be minimax optimal.

What can you conclude if $a \mapsto L(a, \theta)$ is decreasing (but not necessarily strictly decreasing) for $a < \theta$ and increasing (but not necessarily strictly decreasing) for $a > \theta$.

# 2: On the minimax estimator of a binomial parameter

Let $X \sim P_\theta = \text{Binom}(n, \theta)$, where $\theta \in \Theta = [0, 1]$, and we consider the square loss $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$. Recall that a minimax estimator is given by

$$\hat{\theta}_{\text{MM}}(X) = \frac{\sqrt{n}}{1 + \sqrt{n}} \cdot \frac{X}{n} + \frac{1}{1 + \sqrt{n}} \cdot \frac{1}{2}.$$  \hspace{1cm} (1)

We know already that this is Bayes optimal with respect to the prior distribution $Q = \text{Beta}(\sqrt{n}/2, \sqrt{n}/2)$.

(a) Consider the case $n = 1$. Construct a two points prior $Q = q\delta_{\theta_1} + (1 - q)\delta_{\theta_2}$ whose Bayes optimal estimator coincides with $\hat{\theta}_{\text{MM}}$. 

1
(b) Show that, for any $n$, there exists a prior supported on $m$ number of points for some integer $m$, whose Bayes estimators coincides with $\hat{\theta}_M$.

[You can assume that the linear system $\sum_{i=0}^m q(i/m)^k = \int \theta^k Q(d\theta)$, $k \in \{0, \ldots, n + 1\}$ has a solution $q = (q_0, \ldots, q_m) \geq 0$ for $m$ large enough. (Here $Q = \text{Beta}(\sqrt{n}/2, \sqrt{n}/2)$.)]

# 3: Minimax estimation of sparse vectors

Let $\Theta \subseteq \mathbb{R}^d$ and consider estimation with a loss $L : A \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ upper bounded by $L_0$: $\sup_{a \in A, \theta \in \Theta} L(a, \theta) \leq L_0$.

(a) Prove that, for any probability distribution $Q$ on $\mathbb{R}^d$,

$$R_M(\Theta) \geq R_0(Q) - L_0 Q(\Theta^c),$$

(2)

where $Q(\Theta^c) = \int_{\Theta^c} Q(d\theta)$ is the probability of $\Theta^c$ under $Q$, and $R_0(Q) = \int_{\mathbb{R}^d} R(A; \theta) Q(d\theta)$. (Here we assume that $P_\theta$ is not only defined for $\theta \in \Theta$, but for any $\theta \in \mathbb{R}^d$.)

Given two integers $1 \leq k \leq d$, and a real number $M \geq 0$, define the set of vectors

$$\Theta(d, k; M) = \left\{ \theta \in \{0, +M, -M\}^d : \|\theta\|_0 \leq k \right\},$$

(3)

where $\|\theta\|_0 = |\text{supp}(\theta)|$ is the number of non-zero entries in $\theta$. We are interested in the minimax error for the Gaussian location model with this parameters space $\Theta = \{P_\theta : \theta \in \Theta(d, k; M)\}$, action space $\mathbb{R}^d$, and square loss $L(\hat{\theta}, \theta) = \|\hat{\theta} - \theta\|_2^2$. We will denote this minimax risk by $R_M(d, k; M)$.

(b) Prove that, in determining the minimax error, we can restrict ourselves to estimators that take values in $A = B^d(0; M\sqrt{k}) = \{\theta \in \mathbb{R}^d : \|\theta\|_2 \leq M\sqrt{k}\}$. Further, we can replace the square loss by $\tilde{L}(\hat{\theta}, \theta) = \min(\|\hat{\theta} - \theta\|_2^2, 4M^2k)$.

(c) Prove that there exists a least favorable prior $Q_*$, and that it can be taken of the form

$$Q_* = \sum_{\ell=0}^k p_\ell Q_\ell$$

(4)

where $p = (p_\ell)_{0 \leq \ell \leq k}$ is a probability distribution over $\{0, 1, \ldots, k\}$, and $Q_\ell$ is the uniform distribution over vectors in $\Theta(d, k; M)$ with $\|\theta\|_0 = \ell$.

[Hint: Note that this claim is equivalent to $Q_*(\{\theta_1\}) = Q_*(\{\theta_2\})$, for any $\theta_1, \theta_2 \in \Theta(d, k; M)$ with $\|\theta_1\|_0 = \|\theta_2\|_0$.

Computing the Bayes risk for the prior $Q_*$ described above is somewhat intricate. We thus consider a simpler prior $Q_{M, \varepsilon}$. Under $Q_{M, \varepsilon}$ the coordinates of $\theta$ are independent with $Q_{M, \varepsilon}(\{\theta_i = M\}) = Q_{M, \varepsilon}(\{\theta_i = -M\}) = \varepsilon/2$, and $Q_{M, \varepsilon}(\{\theta_i = 0\}) = 1 - \varepsilon$. Equivalently $Q_{M, \varepsilon} = q_{M, \varepsilon} \times \cdots \times q_{M, \varepsilon}$, where $q_{M, \varepsilon}$ is the three points distribution $q_{M, \varepsilon} = (1-\varepsilon)\delta_0 + (\varepsilon/2)\delta_M + (\varepsilon/2)\delta_{-M}$.

(d) Prove that

$$R_M(d, k; M) \geq \tilde{R}_a(Q_{M, \varepsilon}) - 4M^2k \mathbb{P}\left(\text{Binom}(d, \varepsilon) > k\right),$$

(5)

where $\tilde{R}_a$ is the Bayes risk for the loss function $\tilde{L}$. 

2
Setting $\varepsilon = (k/d)(1 - \eta)$, it is possible to show (for instance by Bernstein inequality [BLM13]) that

$$\mathbb{P}\left(\text{Binom}(d, \varepsilon) > k\right) \leq e^{-k\eta^2/4}.$$  

(6)

Let $R_B$ denote the Bayes risk for the square loss. It is also possible to show that

$$\tilde{R}_B(Q_{M,\varepsilon}) \geq R_B(Q_{M,\varepsilon}) - (M^2 + 1) o_\eta(k),$$  

(7)

where $o_\eta(k)$ is a quantity such that $\lim_{k \to \infty} o_\eta(k)/k = 0$ for any $\eta > 0$.

(e) Prove that the above implies

$$R_M(d, k; M) \geq d R_B(q_{M,\varepsilon}) - (M^2 + 1) o_\eta(k).$$  

(8)

where $R_B(q_{M,\varepsilon})$ is the Bayes risk for the one-dimensional problem of estimating $\theta \sim q_{M,\varepsilon}$ from $X = \theta + Z$, $Z \sim \mathcal{N}(0, 1)$.

**Optional**

This question will not be graded and is mainly food for thought:

- Continuing from the previous problem, what is the behavior of $R_B(q_{M,\varepsilon})$ with $\varepsilon$ and $M$? What are the consequences for $R_M(d, k; M)$? Of particular interest is the regime $\varepsilon \ll 1$ (corresponding to $k \ll d$).

**References**