**Homework 2 Due: Apr 7, 2016**
Spring 2016
**ECE 598**
**Information-theoretic methods in high-dimensional statistics**
Prof. Yihong Wu

Rules:

- It is mandatory to type your solutions in LaTeX. Email your solution in pdf (file name: `HW2-your name.pdf`) by midnight of the due date to yihongwu@illinois.edu with subject line `Homework XX: your name`.
- Justify your work rigorously. As long as you are able to prove the result or a stronger version, there is no need to follow the hints.

1. (Shannon lower bound) Let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^k$ and $r > 0$. Let $X$ be a $\mathbb{R}^k$-valued random vector with a probability density function $p_X$. Let

$$F(\epsilon) \triangleq \inf_{P_{\hat{X}|X}:\mathbb{E}[\|\hat{X}-X\|^r]\leq\epsilon} I(X;\hat{X})$$

We prove the Shannon lower bound:

$$F(\epsilon) \geq h(X) + \frac{k}{r}\log\frac{k}{\epsilon re} - \log\left(\Gamma\left(\frac{k}{r}+1\right)V_k\right), \tag{1}$$

where $h(X) = \int_{\mathbb{R}^k} p_X(x)\frac{1}{p_X(x)}dx$ is the differential entropy of $X$ and $V_k = \mathrm{vol}(\{x \in \mathbb{R}^k : \|x\| \leq 1\})$ is the volume of the unit $\|\cdot\|$-ball.

(a) Show that

$$F(\epsilon) \geq h(X) - G(\epsilon), \quad G(\epsilon) \triangleq \sup_{P_W:\mathbb{E}[\|W\|^r]\leq\epsilon} h(W).$$

(b) Show that

$$G(\epsilon) = G(1) + \frac{k}{r}\log\epsilon.$$

(c) Show that $0 < V_k < \infty$.

(d) Show that for any $s > 0$,

$$Z(s) \triangleq \int_{\mathbb{R}^k}\exp(-s\|w\|^r)dw = \Gamma\left(\frac{k}{r}+1\right)V_k s^{-\frac{k}{r}}$$

(Hint: $\|\cdot\|$ may not be smooth so you cannot just take derivatives. Use $\int_{\mathbb{R}^k}\exp(-s\|w\|^r)dw = \int_{\mathbb{R}^k}\int_{\|w\|^r}^{\infty} s\exp(-sx)dxdw$ and apply Fubini's theorem.)

(e) Consider the probability density function

$$q_s(w) = \frac{1}{Z(s)}\exp(-s\|w\|^r).$$

Show that for any feasible $W$ such that $\mathbb{E}[\|W\|^r] \leq 1$,

$$h(W) = -D(p\|q_s) + \mathbb{E}\left[\log\frac{1}{q_s(W)}\right] \leq \log Z(s) + s.$$

(f) Optimize over $s > 0$ and conclude that

$$G(1) = -\frac{k}{r} \log \frac{k}{re} + \log \left( \Gamma \left( \frac{k}{r} + 1 \right) V_k \right).$$

(g) Assembling all pieces to conclude (1).

(h) Evaluate (1) for $\|\cdot\|_\infty$ and $r = 2$.

2. (Application of SLB to minimax risk) Consider the $k$-dimensional $n$-sample Gaussian location model: $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, I_k)$ where $\theta \in \Theta \subset \mathbb{R}^k$ and $\Theta$ has nonemtpy interior.

(a) Apply the mutual information method to prove the following minimax lower bound for quadratic loss:

$$R_n^*(\Theta) \triangleq \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta \|\hat{\theta} - \theta\|_2^2 \gtrsim \left( \frac{k}{n} \wedge \text{rad}(\Theta)^2 \right) \left( \frac{\text{vol}(\Theta)}{\text{vol}(B_2)} \right)^{\frac{2}{k}},$$

where $\text{rad}(\Theta)$ denotes the radius of $\Theta$ with respect to the $\ell_2$-norm and $B_2$ denotes an $\ell_2$-ball of the same radius as $\Theta$. (Hint: choose the uniform prior on $\Theta$.)

(b) When $\Theta$ is an $\ell_2$-ball, show that

$$R_n^*(\Theta) \asymp \frac{k}{n} \wedge \text{rad}(\Theta)^2.$$

3. (Maxima of Gaussians) Let $Z_1, \ldots, Z_n$ be distributed as $\mathcal{N}(0,1)$. Note: **Below if you invoke any bound on the Gaussian tail, you have to prove it.**

(a) Show that regardless of the dependence of $Z_i$'s,

$$\mathbb{E}\Big[\max_{i\in[n]} Z_i\Big] \lesssim \sqrt{\log n}. \tag{2}$$

(Hint: union bound $\mathbb{P}[\max Z_i \geq t] \leq n\mathbb{P}[Z_1 \geq t]$.)

(b) In the remainer of this problem we assume that $Z_i$'s are independent. Show that

$$\mathbb{E}\Big[\max_{i\in[n]} Z_i\Big] \asymp \mathbb{E}\Big[\max_{i\in[n]} |Z_i|\Big] \asymp \sqrt{\log n}.$$

(Hint: $\mathbb{P}[\max Z_i \geq t] = 1 - \mathbb{P}[Z_1 \leq t]^n$. You cannot use Sudakov minorization because its proof uses this very fact.)

(c) (Bonus) Show that

$$\mathbb{E}\Big[\max_{i\in[n]} Z_i\Big] = \sqrt{2\log n}(1 + o(1)).$$

(d) (Bonus$^2$) Show that

$$\mathbb{E}\Big[\max_{i\in[n]} Z_i\Big] = \sqrt{2\log n} + o(1).$$

4. (Metric entropy) Recall that $N(T, \|\cdot\|, \epsilon)$ and $M(T, \|\cdot\|, \epsilon)$ denote the covering and packing number of a subset $T$.

(a) We say $(x_1, \ldots, x_N)$ is an *$\epsilon$-net* of $T$ if each $x_i \in T$ and $T \subset \cup_{i=1}^{n} B(x_i, \epsilon)$. (Note the difference with $\epsilon$-covering which are not required to lie in the set). Show that a maximal $\epsilon$-packing forms an $\epsilon$-net.

(b) Let $S^{n-1}$ denote the unit sphere in $\mathbb{R}^n$. Show that its packing number satisfies

$$M(S^{n-1}, \|\cdot\|_2, \epsilon) \leq \left(1 + \frac{2}{\epsilon}\right)^n. \tag{3}$$

(c) Show that the bound (3) is not tight by finding out the exact covering number when $n = 1$.

(d) The above observation the correct exponent in (3) should be $n-1$. Indeed, for any $\epsilon < 1$, prove that

$$\left(\frac{1}{\epsilon}\right)^{n-1} \leq M(S^{n-1}, \|\cdot\|_2, \epsilon) \leq 2n \left(1 + \frac{2}{\epsilon}\right)^{n-1}. \tag{4}$$

(Hint: revisit the volume argument. Do not blindly apply it since volume of the sphere is zero).

5. (Random matrix) Let $A$ be an $m \times n$ matrix of iid $\mathcal{N}(0,1)$ entries. Denote its operator norm by $\|A\|_{\text{op}} = \max_{v \in S^{n-1}} \|Av\|$, which is also the largest singular value of $A$.

(a) Show that

$$\|A\|_{\text{op}} = \max_{u \in S^{m-1}, v \in S^{n-1}} \left\langle A, uv' \right\rangle. \tag{5}$$

(b) Let $\mathcal{U} = \{u_1, \ldots, u_M\}$ and $\mathcal{V} = \{v_1, \ldots, v_M\}$ be an *$\epsilon$-net* for the spheres $S^{m-1}$ and $S^{n-1}$ respectively. Show that

$$\|A\|_{\text{op}} \leq \frac{1}{(1-\epsilon)^2} \max_{u \in \mathcal{U}, v \in \mathcal{V}} \left\langle A, uv' \right\rangle.$$

(c) Use (2) and (3) to conclude that

$$\mathbb{E}[\|A\|] \lesssim \sqrt{n} + \sqrt{m} \tag{6}$$

(d) By choosing $u$ and $v$ in (5) smartly, show a matching lower bound and conclude that

$$\mathbb{E}[\|A\|] \asymp \sqrt{n} + \sqrt{m} \tag{7}$$

(e) Use Sudakov minorization to prove a matching lower bound. (Hint: use (4)).

6. (Maurey's empirical method) Let $H$ be an inner product space with the norm defined by the inner product $\|x\| \triangleq \sqrt{\langle x, x \rangle}$. Let $T \subset H$ be a finite set and denote the radius of $T$ by

$$r = \text{rad}(T) = \inf_{y \in H} \sup_{x \in T} \|x - y\|.$$

Denote the convex hull of $T$ by $\text{co}(T)$.

(a) Use Maurey's empirical method discussed in the lecture to prove the following dimension-free bound on the cover number

$$\mathcal{N}(\text{co}(T), \|\cdot\|, \epsilon) \leq \binom{|T| + \lceil \frac{r^2}{\epsilon^2} \rceil - 1}{\lceil \frac{r^2}{\epsilon^2} \rceil - 1}. \tag{8}$$

(b) Show that for any $T$ (not necessarily finite), $\text{rad}(\text{co}(T)) = \text{rad}(T)$.

(c) Conclude that (8) holds with equality when $\epsilon = r$.