**Stat 300A Theory of Statistics**

# Homework 5

*Andrea Montanari* *Due on November 7, 2018*

- Solutions should be complete and concisely written. Please, use a separate sheet (or set of sheets) for each problem.
- We will be using Gradescope (`https://www.gradescope.com`) for homework submission (you should have received an invitation) - no paper homework will be accepted. Handwritten solutions are still fine though, just make a good quality scan and upload it to Gradescope.
- You are welcome to discuss problems with your colleagues, but should write and submit your own solution.

## # 1: A function denoising problem

Let $\boldsymbol{\theta}$ be a discrete function sampled on a regular grid in $[0,1]$. Namely, for $n \in \mathbb{N}$, we let $\varepsilon = 1/n$, and

$$\boldsymbol{\theta} = \big(\theta(0), \theta(\varepsilon), \theta(2\varepsilon), \dots, \theta((n-1)\varepsilon)\big) \in \mathbb{R}^n\,. \tag{1}$$

We observe noisy measurements of this function $y_k = \theta(k\varepsilon) + z_k$, where $(z_k)_{k\le n} \sim_{iid} \mathsf{N}(0,\sigma^2)$, and are interested in estimating $\boldsymbol{\theta}$ with respect to the normalized square loss $L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2/n$.

We define the discrete derivative by letting $\Delta\theta(k\varepsilon) = [\theta((k+1)\varepsilon) - \theta(k\varepsilon)]/\varepsilon$ for $k \in \{0,\dots,n-2\}$, and $\Delta\theta((n-1)\varepsilon) = [\theta(0) - \theta((n-1)\varepsilon)]/\varepsilon$ (periodic boundary conditions). We consider the following parameter class

$$\Theta(R,n) = \Big\{\boldsymbol{\theta} : \sum_{k=0}^{n-1}\theta(k\varepsilon) = 0,\ \sum_{k=0}^{n-1}\varepsilon\big(\Delta\theta(k\varepsilon)\big)^2 \le R\Big\}\,. \tag{2}$$

(*a*) Give an expression for the linear minimax risk $R_{\rm L}(\Theta(R,n))$.

[Hint: It might be convenient to use the discrete Fourier transform of $\boldsymbol{\theta}$.]

(*b*) Can you apply Pinsker's theorem and show that the linear minimax risk is close to the overall minimax risk $R_{\rm M}(\Theta(R,n))$? Justify your answer and state explicitly any eventual condition that you are imposing on $R$, $n$.

## # 2: A simple application of Le Cam's method

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable probability density function, and assume that there exists another density function $g : \mathbb{R}^d \to \mathbb{R}$, and a constant $M$ such that, for all $\boldsymbol{x} \in \mathbb{R}^d$

$$\big\|\nabla f(\boldsymbol{x})\big\|_2 \le M\, g(\boldsymbol{x})\,. \tag{3}$$

We will denote by $\mathsf{P}_{\boldsymbol{\theta}}$ the probability distribution of $\boldsymbol{X} = \boldsymbol{\theta} + \boldsymbol{W}$ where $\boldsymbol{W} \sim f(\,\cdot\,)$ is noise with density $f$.

(*a*) Prove that, for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}^d$,

$$\big\|\mathsf{P}_{\boldsymbol{\theta}_1} - \mathsf{P}_{\boldsymbol{\theta}_2}\big\|_{\rm TV} \le \frac{M}{2}\,\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2\,. \tag{4}$$

(*b*) Consider the problem of estimating $\boldsymbol{\theta} \in \Theta \equiv \mathbb{R}^d$ from data $\boldsymbol{X} \sim \mathsf{P}_{\boldsymbol{\theta}}$ under the square loss $L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2$. Use the previous result to derive a lower bound on the minimax risk.

[Hint: It is sufficient to consider two priors $\mathsf{Q}_1$, $\mathsf{Q}_2$ given by Dirac's deltas.]

(*c*) Apply this lower bound to the case of Gaussian noise, namely to the case of $f$ the density of the Gaussian distribution $\mathsf{N}(0, \sigma^2 \mathrm{I}_d)$. How does the result compare with the actual minimax risk?

## # 3: Some properties of distances between distributions

(*a*) Let $\mathsf{P} = \mathsf{P}_1 \times \mathsf{P}_2 \times \cdots \times \mathsf{P}_n$ and $\mathsf{Q} = \mathsf{Q}_1 \times \mathsf{Q}_2 \times \cdots \times \mathsf{Q}_n$ be two product-form distributions (where, for each $i \le n$, $\mathsf{P}_i$, $\mathsf{Q}_i$ are probability measures on the same space $\mathcal{X}_i$). Show that

$$\big\|\mathsf{P} - \mathsf{Q}\big\|_{\mathrm{TV}} \le \sum_{i=1}^{n} \big\|\mathsf{P}_i - \mathsf{Q}_i\big\|_{\mathrm{TV}} \,. \tag{5}$$

[Hint: Start with $n = 2$. It is fine to assume that the $\mathcal{X}_i$'s are finite sets.]

(*b*) Prove that there cannot be a reverse Pinsker inequality. Namely, there does not exist any function $f : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ with $f(t) > 0$ for $t > 0$ such that, for any two distributions $\mathsf{P}, \mathsf{Q}$.

$$D(\mathsf{P}\|\mathsf{Q}) \le f(\|\mathsf{P} - \mathsf{Q}\|_{\mathrm{TV}}) \,. \tag{6}$$

(*c*) Assume that $\mathsf{P}$ and $\mathsf{Q}$ are probability distributions over a finite set $\mathcal{X}$, with probability mass functions $\mathsf{p}$, $\mathsf{q}$, and assume $\mathsf{q}(x) \ge \mathsf{q}_{\min} > 0$ for all $x \in \mathcal{X}$. Prove that there exists $g : \mathbb{R}_{\ge 0} \times \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ with $g(t, s) > 0$ for $t, s > 0$ such that, for any two probability mass functions $\mathsf{p}, \mathsf{q}$, we have

$$D(\mathsf{P}\|\mathsf{Q}) \le g(\|\mathsf{P} - \mathsf{Q}\|_{\mathrm{TV}}, \mathsf{q}_{\min}) \,. \tag{7}$$

We would like the function $g$ to be such that $\lim_{z \to 0} g(z; \mathsf{q}_{\min}) = 0$ for any $\mathsf{q}_{\min} > 0$. Give an explicit expression for the function $g$.

[Hint: Write $D(\mathsf{P}\|\mathsf{Q}) = \mathbb{E}_{\mathsf{Q}}(X \log X - X + 1)$, for $X = \frac{\mathrm{d}P}{\mathrm{d}Q}$.]

## Optional

Can you suggest different priors $\mathsf{Q}_1$, $\mathsf{Q}_2$ to improve the lower bound in problem 2?