S&DS 602: Homework 4

Due Wednesday, September 25 at 2PM, via Gradescope

1. Let $X_1, \ldots, X_n$ be any independent random variables, and let $X_{(1)} \geq X_{(2)} \geq \ldots \geq X_{(n)}$ be their ordered values. Show that for any $1 \leq k \leq n/2$,

$$\mathrm{Var}[X_{(k)}] \leq k\,\mathbb{E}[(X_{(k)} - X_{(k+1)})^2].$$

2. Let $X_1, \ldots, X_n$ be i.i.d. random points uniformly distributed in the unit square $[0,1]^2$.

   (a) Show that for an absolute constant $C > 0$, $\mathbb{E}[\min_{i=2}^n \|X_1 - X_i\|_2^2] \leq C/n$.

   (b) Let $f(X_1, \ldots, X_n)$ be the total length of the shortest path that visits each point exactly once and returns to the starting point. (This is known as the solution to the *traveling salesman problem.*) Show that for an absolute constant $C > 0$ and any $n \geq 2$,

$$\mathrm{Var}\, f(X_1, \ldots, X_n) \leq C.$$

3. (van Handel 2.9) Let $X_1, \ldots, X_n$ be i.i.d. standard exponential random variables (having density $e^{-x}\mathbf{1}\{x \geq 0\}$ on the real line) and let $f : \mathbb{R}^n \to \mathbb{R}$ be any bounded function with bounded first partial derivatives. Show the Poincaré inequality

$$\mathrm{Var}[f(X_1, \ldots, X_n)] \leq 4\mathbb{E}[\|\nabla f(X_1, \ldots, X_n)\|_2^2]$$

   [Hint: Use the identity $\int_0^\infty g(x)e^{-x}\,dx = g(0) + \int_0^\infty g'(x)e^{-x}\,dx$.]

4. (PAC-Bayes) (a) Let X be a random variable taking value in any Euclidean space $\mathcal{X}$, and let $f_\theta : \mathcal{X} \to \mathbb{R}$ be a family of uniformly bounded functions indexed by a parameter $\theta \in \Theta$. Define $\psi(\theta) = \log \mathbb{E}[e^{f_\theta(X)}]$. For any prior distribution P on Θ and any $x \in \mathcal{X}$, show that

$$\sup_{Q \ll P} \Big[\mathbb{E}_{\theta \sim Q}[f_\theta(x) - \psi(\theta)] - D_{\mathrm{KL}}(Q\|P)\Big] \leq \log \mathbb{E}_{\theta \sim P} e^{f_\theta(x) - \psi(\theta)}$$

   with the supremum taken over all distributions Q on Θ absolutely continuous with respect to P. Show that this implies

$$\mathbb{E}_X \sup_{Q \ll P} \left[\exp\Big(\mathbb{E}_{\theta \sim Q}[f_\theta(X) - \psi(\theta)] - D_{\mathrm{KL}}(Q\|P)\Big)\right] \leq 1$$

   where the outer expectation is over X.

(b) Let $X_1, \ldots, X_n$ be i.i.d. random variables, and let $g_\theta : \mathbb{R} \to \mathbb{R}$ be a family of uniformly bounded functions for which $\mathbb{E} g_\theta(X_i) = 0$ and $g_\theta(X_i)$ is σ^2-subgaussian for all $\theta \in \Theta$.

Fix any $\lambda > 0$, $\delta \in (0,1)$, and prior distribution P on Θ. Show that with probability at least $1 - \delta$ over $X_1, \ldots, X_n$, for any $Q \ll P$,

$$\mathbb{E}_{\theta \sim Q} \left[\frac{1}{n} \sum_{i=1}^{n} g_\theta(X_i) \right] \leq \frac{\lambda \sigma^2}{2n} + \frac{1}{\lambda} D_{\mathrm{KL}}(Q \| P) + \frac{1}{\lambda} \log \frac{1}{\delta}.$$