TOPICS IN STATISTICAL THEORY – EXAMPLES 1/2

Part III, Lent 2019, RN (email: r.nickl@statslab.cam.ac.uk)

Ex. 1 i) Show that the collection of functions $\{e_k = \exp\{2\pi i k \cdot\} : k \in \mathbb{Z}\}$ forms an orthonormal basis of the Hilbert space $L^2([0,1])$ equipped with inner product $\langle f, g \rangle = \int f(x)\overline{g(x)}dx$. [Hint: You may use the Stone-Weierstrass theorem and that the space $C([0,1])$ of continuous functions on $[0,1]$ is dense in $L^2([0,1])$.]

ii) Let $\alpha \in \mathbb{N}$ and consider a function $f \in L^2([0,1])$ such that $D^j f \in L^2, D^j f(0) = D^j f(1)$, for all $0 \le j \le \alpha$. If $K \in \mathbb{N}$ and $f_K = \sum_{|k| \le K} \langle f, e_k \rangle e_k$, show that

$$\|f_K - f\|^2_{L^2([0,1])} \le cK^{-2\alpha}$$

for some constant c independent of K. Now let $\theta_k = \langle f, e_k \rangle$ and suppose you are given noisy observations $Y_k = \theta_k + (g_k/\sqrt{n})$ with $g_k \sim^{i.i.d.} N(0,1), n \in \mathbb{N}$. Let $\hat{f}_K = \sum_{|k| \le K} Y_k e_k$. Prove that for a choice of $K = K_n \to \infty$ to be specified, there exists a constant C independent of n such that

$$E\|\hat{f}_K - f\|^2_{L^2([0,1]} \le Cn^{-2\alpha/(2\alpha+1)}.$$

Ex. 2 Let X and Y be independent identically distributed centred normal random vectors in $\mathbb{R}^n$. For $\theta \in [0, 2\pi]$, define $X(\theta) = X \sin\theta + Y \cos\theta$ and let $X'(\theta) = (d/d\theta)X(\theta)$. Show that the random vectors (X, Y) and $(X(\theta), X'(\theta))$ have the same distribution in $\mathbb{R}^n \times \mathbb{R}^n$

Ex. 3 Let X be a Gaussian random variable in $\mathbb{R}^d$, and let $V \subseteq \mathbb{R}^d$ be any linear subspace. Show

$$\Pr(X \in V) \in \{0, 1\}.$$

Ex. 4 Let $X \sim N(0, I_d)$ be a standard Gaussian vector in $\mathbb{R}^d$ and let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function such that $\|\nabla f\| \le 1$ almost everywhere, where $\|\cdot\|$ is the Euclidean norm. Prove that

$$\Pr(|f(X) - Ef(X)| > t) \le 2\exp\Big\{-\frac{2t^2}{\pi^2}\Big\} \quad \forall t > 0.$$

Ex. 5 Let $(g_i : i \in \mathbb{N})$ be an i.i.d. sequence of $N(0,1)$ variables. Show that

$$\lim_{n\to\infty} \frac{E\max_{i \le n}|g_i|}{\sqrt{2\log n}} = 1.$$

Ex. 6 i) Let B_r be the ball of radius r in a d-dimensional normed linear space V whose norm $\|\cdot\|_V$ induces the metric d_V. Show that its covering numbers satisfy

$$d\log(r/2\epsilon) \le \log N(B_r, d_V, \epsilon) \le d\log(3r/\epsilon), \quad \forall 0 < \epsilon < r.$$

ii) Consider the Sobolev ellipsoid in sequence space $\ell_2(\mathbb{N})$ given by

$$\Theta(\alpha, B) = \left\{\theta \in \ell_2(\mathbb{N}) : \sum_{k\in\mathbb{N}} k^{2\alpha}\theta_k^2 \le B^2\right\}.$$

Show that for all $0 < \epsilon < B$ there exists a constant K such that

$$\log N(\Theta(\alpha, B), d_{\ell_2}, \epsilon) \le K(B/\epsilon)^{1/\alpha},$$

where d_{ℓ_2} is the metric induced by the norm $\|\theta\|^2_{\ell_2} = \sum_k \theta_k^2$. [Hint: Re-arrange the index set $\{1, \dots, K, \dots\}$ into a suitable dyadic scheme and use Part i).]

Ex. 7 Suppose two probability measures P, Q have a common dominating measure μ on a measure space $(\mathcal{X}, \mathcal{A})$, and are absolutely continuous with respect to each other. Write $p = dP/d\mu$ and $q = dQ/d\mu$, and denote by $K(P, Q) = E_P[\log(dP/dQ)]$ the Kullback-Leibler divergence between P and Q. Show that

$$E_P \left| \log \frac{p}{q}(X) \right| \leq K(P, Q) + \sqrt{2K(P, Q)}$$

Ex. 8 (Hoeffding's inequality) Let $X_1, \dots, X_n$ be independent random variables such that

$$EX_i = 0, \;\; X_i \in [a_i, b_i], a_i < b_i, \quad \forall\; i = 1, \dots, n.$$

Prove that, for every $t > 0$,

$$\Pr\left(\Big| \sum_{i=1}^n X_i \Big| > t \right) \leq 2 \exp\left\{ - \frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}.$$

Ex. 9 * (Sparse Varshamov-Gilbert inequality). Let $1 \leq k \leq d$ be integers such that $d \geq 32k$ and equip the hypercube $\{0, 1\}^d$ with the Hamming metric ρ. Show that there exists vectors $\omega_1, \dots, \omega_M$ in $\{0, 1\}^d$ that are all k-sparse (i.e., have at most k non-zero entries) and for which a) $\rho(\omega_i, \omega_j) \geq k/2$ for all $i \neq j$, and b) $\log M \geq (k/9) \log(1 + (d/2k))$.