Example Sheet 2

Topics in Statistical Theory
Part III, Lent 2016 - Quentin Berthet*

Exercise 1

For $\theta \in (0,1)$, let $\mathbf{P}_\theta = \mathsf{Bin}(\theta, n)$. Show that

$$\mathsf{KL}(\mathbf{P}_\theta, \mathbf{P}_{\theta'}) \geq Cn(\theta - \theta')^2 \,.$$

Exercise 2

For $\mu \in \mathbf{R}^d$ and $\Sigma \in \mathbf{R}^{d\times d}$ such that $\Sigma \succ 0$, let $\mathbf{P}_{\mu,\Sigma} = \mathcal{N}(\mu, \Sigma)$. Show that

$$\mathsf{KL}(\mathbf{P}_{\mu_0,\Sigma_0}, \mathbf{P}_{\mu_1,\Sigma_1}) = \frac{1}{2}\left(\mathbf{Tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1 - \mu_0)^\top \Sigma_1^{-1}(\mu_1 - \mu_0) - d + \log\left(\frac{\det(\Sigma_1)}{\det(\Sigma_0)}\right)\right) \,.$$

Exercise 3

(a) Let $\mathbf{P}_n$ (resp. $\mathbf{Q}_n$) be the distribution of n independent samples from distribution $\mathbf{P}$ (resp. $\mathbf{Q}$). For all $n \leq n'$ show that

$$d_{\mathsf{TV}}(\mathbf{P}_n, \mathbf{Q}_n) \leq d_{\mathsf{TV}}(\mathbf{P}_{n'}, \mathbf{Q}_{n'}) \,.$$

(b) Let f be a measurable function from $\mathcal{X}$ to $\mathcal{Y}$. For any distribution $\mathbf{P}$ on $\mathcal{X}$, let $\mathbf{P}_f$ be the distribution on $\mathcal{Y}$ of $f(X)$, when $X \sim \mathbf{P}$. Show that

$$d_{\mathsf{TV}}(\mathbf{P}_f, \mathbf{Q}_f) \leq d_{\mathsf{TV}}(\mathbf{P}, \mathbf{Q}) \,.$$

Exercise 4

Let $\mathcal{H}_d = \{0,1\}^d$. For $k \leq d/8$, let $\mathcal{H}_{d,k}$ be defined as

$$\mathcal{H}_{d,k} = \{\omega \in \mathcal{H}_d : \mathbf{1}^\top \omega = k\} \,,$$

i.e. $\mathcal{H}_{d,k}$ is the set of elements of the discrete hypercube with exactly k ones. We recall that the Hamming distance ρ is defined as

$$\rho(\omega, \omega') = \sum_{i=1}^d \mathbf{1}\{\omega_i \neq \omega_i'\} \,.$$

Show that there exists a set $\mathcal{P} \subset \mathcal{H}_{d,k}$ such that for all $\omega \neq \omega' \in \mathcal{P}$, we have $\rho(\omega, \omega') \geq k/2$ and $\log|\mathcal{P}| \geq ck\log(d/k)$, for a universal constant $c > 0$.

*q.berthet@statslab.cam.ac.uk

Exercise 5

Let $\theta^* \in \mathbf{R}^d$, $X \in \mathbf{R}^{n\times d}$, and

$$y = X\theta^* + z\,,$$

where z is such that $\mathbf{E}[z] = 0$ and $\mathbf{E}[zz^\top] = I_n$. For $\lambda > 0$, let $\hat{\theta}$ be the *ridge-regression* estimator defined as

$$\hat{\theta} \in \operatorname*{argmin}_{\theta \in \mathbf{R}^d} \|y - X\theta\|_2^2 + \lambda\|\theta\|_2^2\,.$$

(a) Show that $\hat{\theta}$ is uniquely determined, and give a closed-form expression for it.

(b) Let X be such that $X^\top X \succ 0$. Compute the bias and variance (mean ℓ_2 norm squared) of $\hat{\theta}$.

Exercise 6

For all $v \in \mathbf{R}^d$ and $\theta > 0$, let $\mathbf{P}_v = \mathcal{N}(0, I_d + \theta vv^\top)$.

(a) What are the eigenvalues and associated eigenspaces of $\Sigma = I_d + \theta vv^\top$?

(b) For $X_1, \ldots, X_n \sim \mathbf{P}_v$, let $\hat{\Sigma}$ be the *empirical covariance matrix* deined as

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top$$

The estimator $\hat{v}$ is defined as

$$\hat{v} \in \operatorname*{argmax}_{u \in \mathcal{B}_2^d} u^\top \hat{\Sigma} u\,.$$

Show that with probability $1-\delta$, it holds that

$$\|\hat{v}\hat{v}^\top - vv^\top\|_F \le C_\theta\Big(\frac{d + \log(1/\delta)}{n} + \sqrt{\frac{d+\log(1/\delta)}{n}}\Big)\,,$$

for a specific constant C_θ.

(c) For $|v|_0 \le k$, let $\hat{v}$ be defined by

$$\hat{v} \in \operatorname*{argmax}_{\substack{u \in \mathcal{B}_2^d \\ |u|_0 \le k}} u^\top \hat{\Sigma} u\,.$$

Show that with probability $1-\delta$, it holds that

$$\|\hat{v}\hat{v}^\top - vv^\top\|_F \le C'_\theta\Big(\frac{k\log(ed/k) + \log(1/\delta)}{n} + \sqrt{\frac{k\log(ed/k)+\log(1/\delta)}{n}}\Big)\,,$$

for a specific constant C_θ.