Spring 2020

CREST OFPR - Information Theory and Stats

Final Exam

Due: Thur, Jan 23, 2020 (email to `yp@mit.edu`)

`Prof. Y. Polyanskiy`

1 Exercises

NOTE: Each exercise is 10 points. Only 3 exercises per assignment will be graded. If you submit more than 3 solved exercises please indicate which ones you want to be graded.

1 Let $\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$ be the Gaussian distribution on $\mathbb{R}^n$ with mean $\mathbf{m} \in \mathbb{R}^n$ and covariance matrix $\mathbf{\Sigma}$.

1. Under what conditions on $\mathbf{m}_0, \mathbf{\Sigma}_0, \mathbf{m}_1, \mathbf{\Sigma}_1$ is

$$D(\ \mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1) \ \|\ \mathcal{N}(\mathbf{m}_0, \mathbf{\Sigma}_0)\ ) < \infty$$

2. Compute $D(\mathcal{N}(\mathbf{m}, \mathbf{\Sigma}) \| \mathcal{N}(0, \mathbf{I}_n))$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix.
3. Compute $D(\ \mathcal{N}(\mathbf{m}_1, \mathbf{\Sigma}_1) \ \|\ \mathcal{N}(\mathbf{m}_0, \mathbf{\Sigma}_0)\ )$ for non-singular $\mathbf{\Sigma_0}$. (Hint: think how Gaussian distribution changes under shifts $\mathbf{x} \mapsto \mathbf{x} + \mathbf{a}$ and non-singular linear transformations $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$. Apply data-processing to reduce to previous case.)

2 Let X be distributed according to the exponential distribution with mean $\mu > 0$, i.e., with density $p(x) = \frac{1}{\mu} e^{-x/\mu} \mathbf{1}_{\{x \geq 0\}}$. Let $a \in \mathbb{R}$. Compute the divergence $D(P_{X+a} \| P_X)$.

3 (Information bottleneck) Let $X \to Y \to Z$ where Y is a discrete random variable taking values on a finite set $\mathcal{Y}$. Prove that

$$I(X; Z) \leq \log |\mathcal{Y}|.$$

4 Let $D_{SKL}(P||Q) = D(P\|Q) + D(Q\|P)$ be an f-divergence, and define $I_{SKL}(X;Y) = D_{SKL}(P_{X,Y} \| P_X P_Y)$. Now consider three random variables X, A, B s.t. $A \perp\!\!\!\perp B | X$ (i.e. $A \leftarrow X \to B$). Show that

$$I_{SKL}(X; A, B) = I_{SKL}(X; A) + I_{SKL}(X; B)\,.$$

(Hint: first show $I_{SKL}(X;Y) = \sum_{x,x'} P_X(x) P_X(x') D(P_{Y|X=x} \| P_{Y|X=x'})$. Then notice $D(P_{A,B|X=x} \| P_{A,B|X=x'}$ $D(P_{A|X=x} \| P_{A|X=x'}) + D(P_{B|X=x} \| P_{B|X=x'})$.)

5 Let $I_{TV}(X;Y) = \mathrm{TV}(P_{X,Y}, P_X P_Y)$. Let $X \sim \mathrm{Bern}(1/2)$ and conditioned on X generate A and B independently setting them equal to X or $1 - X$ with probabilities $1 - \delta$ and δ, respectively (i.e. $A \leftarrow X \to B$). Show

$$I_{TV}(X; A, B) = I_{TV}(X; A) = |1 - 2\delta|\,,$$

and the second observation of X is "uninformative" (in the I_{TV} sense).

6 For a family of probability distributions $\mathcal{P}$ and a functional $T : \mathcal{P} \to \mathbb{R}$ define its χ^2-modulus of continuity as

$$\delta_{\chi^2}(t) = \sup_{P_1, P_2 \in \mathcal{P}} \{T(P_1) - T(P_2) : \chi^2(P_1 \| P_2) \leq t\}\,.$$

When the functional T is affine, and continuous, and $\mathcal{P}$ is compact[1] Polyanskiy and Wu (2017) show that

$$\frac{1}{7}\delta_{\chi^2}(1/n)^2 \le \inf_{\hat{T}_n} \sup_{P\in\mathcal{P}} \mathbb{E}_{X_i \overset{iid}{\sim} P}(T(P) - \hat{T}_n(X_1,\ldots,X_n))^2 \le \delta_{\chi^2}(1/n)^2 \,. \tag{1}$$

Consider the following bio-statistics problem: In i-th mouse a tumour develops at time $A_i \in [0,1]$ with $A_i \overset{iid}{\sim} \pi$ where π is a pdf on $[0,1]$ bounded by $\frac{1}{2} \le \pi \le 2$. For each i the existence of tumour is checked at another random time $B_i \overset{iid}{\sim} \mathrm{Unif}[0,1]$ with $B_i \perp\!\!\!\perp A_i$. Given observations $X_i = (1\{A_i \le B_i\}, B_i)$ one is trying to estimate $T(\pi) = \pi[A \le 1/2]$. Show that

$$\inf_{\hat{T}_n} \sup_{\pi} \mathbb{E}\left[(T(\pi) - \hat{T}_n(X_1,\ldots,X_n))^2\right] \asymp n^{-2/3} \,.$$

[1] Both under the same, but otherwise arbitrary topology on $\mathcal{P}$