IT for Statistics and Learning
2023

Assignment 7
Assigned: Thursday, Dec 21, 2023
Due: Friday, Jan 12, 2024 M. Skoglund

Problem 7.1: Prove that if Z is σ^2-sub-Gaussian then for any $\lambda \geq 0$,

$$E[e^{\lambda Z^2}] \leq \frac{1}{\sqrt{[1-2\sigma^2\lambda]^+}}$$

where $[x]^+ = \max\{x, 0\}$, and where we get equality for $Z \sim \mathcal{N}(0, \sigma^2)$

Problem 7.2: Let $X^n = (X_1, \ldots, X_n)$ be drawn iid $\sim P$ from $\mathcal{X}$. Let $\{\phi_t\}_{t\in\mathcal{T}}$ be a collection of mappings $\phi_t : \mathcal{X} \to \mathbb{R}$ such that $\phi_t(X_i)$ is σ^2-sub-Gaussian for each t and i. Let T be any random variable with values in $\mathcal{T}$ (for example uniform over $\mathcal{T}$ if $\mathcal{T}$ is an interval in $\mathbb{R}$), and not necessarily independent of X^n. Let

$$M_n(X^n; T) = \frac{1}{n}\sum_{i=1}^{n} \phi_T(X_i)$$

and $M(t) = E_P[\phi_t(X)]$ for any fixed $t \in \mathcal{T}$. Use the result in Prob. 7.1 and the Donsker–Varadhan lemma to show that

$$E[(M_n(X^n;T) - M(T))^2] \leq \frac{1}{\lambda}\left(I(X^n;T) - \frac{1}{2}\log[1 - 2\lambda\sigma^2/n]^+\right)$$

Problem 7.3: Show that for squared Hellinger distance,

$$H^2(P,Q) = \sup_X \left(E_P[X] - E_Q\left[\frac{X}{1-X}\right]\right)$$

Problem 7.4: Let $\theta = \sigma^2$ for

$$f_\theta(x_1, x_2) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}(x_1^2 + x_2^2)\right]$$

and consider approximating the Gaussian pdf

$$g(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{x_1^2}{2\sigma_1^2} - \frac{x_2^2}{2\sigma_2^2}\right)$$

with f_θ by minimizing $D(f_\theta\|g)$ over $\theta = \sigma^2$. Compare the result with what you obtain by instead minimizing $D(g\|f_\theta)$ and comment on the nature of the two different solutions.

Problem 7.5: Consider the variational Bayes example presented in Lec. 7, where

$$f(\theta|x^n) = \frac{f_\theta(x^n)\pi(\theta)}{\int f_\theta(x^n)\pi(\theta)d\theta}, \quad f(x^n) = \int f_\theta(x^n)\pi(\theta)d\theta$$

for $X^n = (X_1, \ldots, X_n)$ iid $\sim \mathcal{N}(\mu, \sigma^2)$ and with $\pi(\theta)$, $\theta = (\mu, \sigma)$, such that

$$f_{\theta|X^n=x^n}(\theta|x^n)f(x^n) = \frac{1}{\sigma_\mu\sigma(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{n(\mu-\bar{x})^2 + S}{2\sigma^2}\right)$$

For $q(\theta) = q(\mu)q(\sigma)$, derive necessary conditions for minimizing $D(q(\theta)\|f_\theta(x^n)\pi(\theta))$ over $q(\mu)$ for a fixed $q(\sigma)$, and vice versa.