ISyE 8803: Special Topics in Modern Mathematical Data Science
Homework 3

Extra-credit

Please submit electronically directly to Canvas in a PDF file.

1 Local behavior of f-divergences

In this exercise, you will show that f-divergence with a strictly convex potential f locally behaves as the χ^2-divergence (which is a specific f-divergence with a particular potential, to be defined later). Let $f : \mathbb{R}_{++} \to \mathbb{R}$, where $\mathbb{R}_{++}$ is the set of all positive reals, satisfy the following assumptions:

- $f(1) = 0$;
- uniformly bounded third derivative on $\mathbb{R}_{++}$, that is f''' exists on $\mathbb{R}_{++}$ and $\sup_{r>0} |f'''(r)| < \infty$;
- f is strictly convex (and thus by the previous assumption $f''(r) > 0$ for any $r > 0$).

In fact, all common f-divergences, except the TV distance, satisfy these assumptions (including Hellinger, chi-squared, and Kullback-Leibler). Recall that the associated f-divergence between two distributions P, Q on the same space, with densities p, q with respest to a dominating measure μ, is

$$D_f(P||Q) := \mathbb{E}_Q \left[f\left(\frac{dP}{dQ}\right)\right] \quad = \int_{\mathcal{X}} f\left(r(x)\right) q(x) d\mu(x),$$

where $r(x) := \frac{p(x)}{q(x)}$ is the likelihood ratio and $\mathcal{X}$ is the support of μ. Fix P and Q, and consider the segment in-betweenm, that is the family of mixture distributions $P_t := (1-t)Q + tP$ for $t \in [0,1]$.

- Show that as $t \to 0$,
$$D_f(P_t||Q) = (1 + o(1)) \frac{f''(1)}{2} \chi^2(P_t||Q)$$
where $o(1) \to 0$ and $\chi^2(P||Q)$ is the chi-square divergence, i.e. $D_h(P||Q)$ with $h(r) = (1-r)^2$.
- Check that $\chi^2(P_t||Q) = t^2 \chi^2(P||Q)$ and conclude that $D_f(P_t||Q)$ is locally quadratic in t.

Hint: *Consider the 3rd-order Taylor expansion of $f(r)$ at $r = 1$. The 1st-order term must vanish.*

2 Local behavior of KL-divergence

3 Boosting the confidence in binary testing

4 More on quantiles

5 Krein-Milman method I: Sharp constant in Hoeffding's lemma

6 Krein-Milman method II: Gap between the median and mean

7 Sketching a covariance matrix via leverage-scores sampling

8 Dikin walk*