HOMEWORK 2 - V1

CSC2547/STA4273 WINTER 2019

University of Toronto

VERSION HISTORY: V0 → V1: CHANGE TITLE (Q1.1), ADD ASSMP ON f (Q2.3)

1. Asymptotics.

1.1. *Delta method-I [10pts].* For a sequence of variables $\{X_n\}_{n\geq 0}$ satisfying

$$\sqrt{n}(X_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \tag{1.1}$$

find the limiting distributions of (a) X_n^2, (b) $\log(|X_n|)$, (c) $1/X_n$, (d) $\exp(X_n)$ (after normalizing appropriately).

1.2. *Delta method-II [10pts].* Let $(X_1, Y_1), ..., (X_n, Y_n)$ be independent, identically distributed two dimensional random vectors with mean (μ_1, μ_2) and covariances given by $\mathrm{Var}(X_1) = \sigma_1^2$ $\mathrm{Var}(Y_1) = \sigma_2^2$, and $\mathrm{Cov}(X_1, Y_1) = \sigma_{12}$. Let f be a function of two variables with continuous gradient which does not vanish in a neighborhood of μ_1, μ_2. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. Find the asymptotic distribution of $f(\bar{X}, \bar{Y})$.

1.3. *Importance Sampling [10pts].* In a sampling problem, we have a function f and a potentially intractable distribution $p(x)$, and we are interested in computing

$$\mu = \mathbb{E}[f(X)] \quad \text{where} \quad X \sim p. \tag{1.2}$$

Assume that p is of the following form

$$p(x) = \frac{1}{Z} h(x)$$

where $h(x)$ can be evaluated; however, the normalizing constant Z is intractable.

In importance sampling, we sample $X_1, X_2, ..., X_n$ i.i.d. random variables from a tractable distribution q, and estimate the expectation (1.2) with

$$\hat{\mu} = \frac{\sum_{i=1}^n w(X_i) f(X_i)}{\sum_{i=1}^n w(X_i)} \quad \text{where} \quad w(x) = \frac{h(x)}{q(x)}. \tag{1.3}$$

(a) (0 pts) Show that $\hat{\mu} \to \mu$.
(b) (5 pts) Find the asymptotic variance of $\hat{\mu}$.
(c) (5 pts) Find the distribution q that minimizes the asymptotic variance. Can you use this distribution in practice?

2. Uniform convergence.

2.1. *ϵ-net over sphere [10pts].* Let $\mathcal{N}_\epsilon$ be an *epsilon*-net over the d-dimensional unit Euclidean sphere $\mathcal{S}^{d-1}$ equipped with the Euclidean metric.

(a) Show that $\forall \epsilon > 0$, $|\mathcal{N}_\epsilon| \leq (1 + 2/\epsilon)^d$.
(b) For $x \in \mathbb{R}^d$, show that $\max_{u \in \mathcal{N}_\epsilon} \langle u, x \rangle \geq (1 - \epsilon)\|x\|_2$.

2.2. *Vector concentration [10pts].* Assume that $X_1, X_2, ..., X_n$ are independent random vectors satisfying $\mathbb{E}[X_i] = \mu \in \mathbb{R}^d$ and $\|X_i\|_2 \leq \kappa$ almost surely. Denoting by $\hat{\mu}$ their sample mean, i.e., $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$, show that $\hat{\mu}$ concentrates around its expectation. That is, find an upper bound on the following probability

$$\mathbb{P}(\|\hat{\mu} - \mu\|_2 \geq \epsilon).$$

Hint: you may find the previous question helpful.

2.3. *More Concentration [10pts].* In a learning task, assume that the features $x \in \mathbb{R}^d$ satisfy $\mathbb{E}[x] = \mu$ and the response $y \in \mathbb{R}$ is given as $y = f(\langle x, \mu \rangle) + \text{noise}$, noise has 0 expectation. Here, f is known, uniformly bounded by $|f| \leq B$ and L-Lipschitz, and the response is not observed. Our objective is to estimate $\mathbb{E}[y]$. For independent observations $x_1, x_2, ..., x_n$ satisfying $\|x_i\|_2 \leq \kappa$ almost surely, we will use

$$\hat{\xi} = \frac{1}{n}\sum_{i=1}^n f(\langle x_i, \hat{\mu} \rangle)$$

for this task where $\hat{\mu}$ denotes the sample mean, i.e., $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$.

Derive the convergence properties of $\hat{\xi}$ in terms of data dimensions using an ϵ-net argument. That is, establish a concentration result for $\hat{\xi}$.