Computational and Statistical Learning Theory

Problem set 1

Due: October 10th

Please send your solutions to `learning-submissions@ttic.edu`

1. $\{0,1\}$-valued random variables $X_1, \ldots, X_n$ are drawn independently each from Bernoulli distribution with parameter $p = 0.1$. Define $P_n := \mathbb{P}(\frac{1}{n}\sum_{i=1}^n X_i \geq 0.2)$.

   (a) For $n = 1$ to 30 calculate and plot the below in the same plot :

   i. Exact value of P_n (binomial distribution).

   ii. Normal approximation for P_n.

   iii. Hoeffding inequality bound on P_n.

   iv. Bernstein inequality bound on P_n.

   (b) For $n = 30$ to 300 calculate and plot the below in the same plot :

   i. Normal approximation for P_n.

   ii. Hoeffding inequality bound on P_n.

   iii. Bernstein inequality bound on P_n.

2. Consider i.i.d. random variables $X_1, \ldots, X_n$ distributed according to each of the distributions specified below. For each distribution, write down the bound on $\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X]\right| > \epsilon\right)$ implied by each of: (1) Markov inequality (2) Chebyshev inequality (3) Hoeffding inequality and (4) Bernstein inequality (some of the inequalities might not be applicable for some of the distributions)

   (a) $\{1, \ldots, 4\}$-valued random variables X_i such that $\mathbb{P}(X_i = a) \propto a^2$.

   (b) X_i is distributed uniformly on interval $[-1, 1]$.

   (c) X_i drawn from normal distribution $N(1, 4)$.

   (d) X_i drawn from exponential distribution with parameter 2.

3. (a) $X_1, \ldots, X_n$ are i.i.d. random variables in $[0, 1]$ with mean $\mathbb{E}[X_i] \geq \mu$. Show that:

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i = 0\right) \leq e^{-\mu n}$$

   (Hint : for $t > 1$, $(1 - 1/t)^t \leq 1/e$).

(b) $X_1, \ldots, X_n$ are $[0, 1]$-valued random variables drawn i.i.d. from some arbitrary distribution. Using each of the inequalities below, write down the implied bound $\ (n, \hat{}, \ )$ such that, for any $\ > 0$:

$$\mathbb{P}(\mathbb{E}[X] \quad \hat{} + \ (n, \hat{}, \ )) \quad 1$$

where $\hat{} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
(1) Hoeffding inequality
(2) Markov inequality
(3) Bernstein Inequality
(4) The inequality from part (a). Note that in this case we only get a bound when $\hat{} = 0$ and so $\ (n, \hat{}, \ )$ will be of the form:

$$\ (n, \hat{}, \ ) = \begin{cases} 1 & \text{if } \hat{} > 0 \\ \textit{something} & \text{if } \hat{} = 0 \end{cases}$$

(5)(Optional) Chebychev Inequality
Observe the differences in the dependence on when using Markov (or Chebychev) as opposed to the exponential inequalities. Also observe how Bernstein's inequality interpolates nicely between the case $\hat{} = 0$ and the simpler Hoeffding inequality.

4. (Optional) One possible way of approximating binomial distribution $B(n, p)$ is by finding an estimation for the binomial coefficient $\binom{n}{k}$. In this problem, we investigate an approximation approach and compare it to the normal approximation for the binomial distribution.

   (a) Consider the following approximation for the binomial coefficient:

   $$\binom{n}{k} \approx \frac{2^{nH(k/n)}}{\sqrt{n}}$$

   where $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy of x. Use this estimation to approximate the binomial distribution at $B(k, n, p)$ as an exponential function(Find the function $f(k, n, p)$ such that $B(k, n, p) = \exp(f(k, n, p))$. In this problem, we refer to this estimation as entropy approximation.

   (b) i. For $n = 20$, $p = 0.5$ and $k = 0$ to 20 plot the binomial distribution, Normal approximation and entropy approximation in the same plot.
   ii. For $n = 200$, $p = 0.5$ and $k = 0$ to 200 plot the binomial distribution, Normal approximation and entropy approximation in the same plot.
   iii. Describe your results.

   (c) i. For $n = 100$, $p = 0.3$ and $k = 0$ to 100 plot the binomial distribution, Normal approximation and entropy approximation in the same plot.
   ii. For $n = 100$, $p = 0.2$ and $k = 0$ to 100 plot the binomial distribution, Normal approximation and entropy approximation in the same plot.

iii. For $n = 100$, $p = 0.1$ and $k = 0$ to 100 plot the binomial distribution, Normal approximation and entropy approximation in the same plot.

iv. Explain your results.

(d) Let $g_e(k; n; p)$ be a function such that $g_e(k; n; p) = p$ if and only if the error of entropy approximation at for $B(k; n; p)$ is less than the error of the Normal approximation and$g_e(k; n; p) = 0$ otherwise.

i. For $n = 100$ and $k = 0$ to 100 plot four functions $g_e(k; n; 0.1)$, $g_e(k; n; 0.2)$, $g_e(k; n; 0.3)$ and $g_e(k; n; 0.5)$.

ii. Under what conditions do you prefer to use entropy approximation instead of Normal approximation?