CSC 696H-001 Fall'24: Homework 1

Kwang-Sung Jun

Due: September 28 11:59pm

Instructions for reporting your answers – not following these will result in deducted grading points:

- Show all work along with answers.
- Place your final answer into an 'answer box' that can be easily identified (unless the answer is a proof). Your final answer should not be based on symbols that you defined (okay to define it in the answer box). Your answer should be in the most explicit and simplified form.
- The grading will also be based on the clarity. There should be no room for interpretation about what you are writing. Otherwise, I will assume that they are wrong. The same goes with undefined symbols.
- Each subproblem's answer must be separated clearly from that of the other subproblems; e.g., do not mix up answers for subproblem 1.(a) with that of subproblem 1.(b). Please write down the subproblem number explicitly like 4.(a) followed by the answer.
- If you get stuck, you can post your questions on Piazza. Try posing your questions to be generic, so you maintain the academic integrity while promoting discussion among your classmates.
- You are encouraged to discuss the homework questions with your classmates, but the discussions should only be at a high level, and you should write your solutions in your own words. For every question you have had discussions on, please mention explicitly whom you have discussed with; otherwise it may be counted as academic integrity violation.
- There will be no late days. Late homework result in zero credit. Not even one minute. It is a good idea to set yourself up your own deadline like one day before it is due.
- For detailed homework policies, please read the course syllabus carefully, available on the course website.
- Each subproblem (i.e., Problem X.Y) is worth 10 points unless noted otherwise.

Submission instruction:

- Submit homework via gradescope. You can hand-write your answers and scan them to make it a PDF, or type up your answers as pdf using LaTeX. If you use your phone camera, I recommend using TurboScan (smartphone app) or similar ones to avoid uploading a slanted image or showing the background. Make sure you rotate it correctly.
- Watch the video and follow the instruction for the submission: `https://youtu.be/KMPoby5g_nE`
- **Report the code as part of the answer as texts.** You should also submit the code to a separate submission entry in gradescope 'HW4 - code' as well.

Problem 1

Consider the exponential distribution (Wikipedia has a nice summary of this distribution). If X follows the exponential distribution with the rate parameter λ, then its (uncentered) moment generating function is $\mathbb{E}[\exp(tX)] = \frac{\lambda}{\lambda - t}, \forall t < \lambda$. Please review the definition of subgamma distribution and find the smallest subgamma parameter c and ν^2 for the exponential distribution. Justify your answer. (Hint: $1 + x \le \exp(x), \forall x$)

Problem 2 (20pt)

This problem is about anytime the Hoeffding's inequality. Please solve this problem after we learn the anytime confidence bound.

Let $\hat{\mu}_t = \frac{1}{t}\sum_{s=1}^t X_s$ where X_s is an i.i.d. $(\sigma^2 = 1)$-sub-Gaussian with mean μ. Recall that we have the following inequalities.

$$\text{(fixed time)} \qquad \text{Fix } t \ge 1.\ \mathbb{P}\left(\mu - \hat{\mu}_t \ge \sqrt{\frac{2}{t}\ln(1/\delta)}\right) \le \delta$$

$$\text{(anytime)} \qquad \mathbb{P}\left(\exists t \ge 1, \mu - \hat{\mu}_t \ge \sqrt{\frac{2}{t}\ln(4t^2/\delta)}\right) \le \delta$$

Your classmate claims that in fact the fixed time deviation bound above actually works for all time t throughout, with probability at least $1 - \delta$. Let us empirically show that she is wrong! Let $N = 10,000$ and $\delta = 0.1$. Draw N Gaussian random variables from mean 0 and variance 1; call them $X_1, \ldots, X_N$. With those, compute $\{\hat{\mu}_t\}_{t=1}^N$. Record whether there exists $t \in [1, N]$ such that $\hat{\mu}_t$ that cross the deviation $\sqrt{(2/t)\ln(1/\delta)}$. Let $Y = 1$ if this was true and 0 otherwise. Now, repeat this 100 times with a fresh set of random samples. Denote by $Y_1, \ldots, Y_{100}$ those binary values. If your friend is correct, we must be seeing that the average of $\{Y_i\}_{i=1}^{100}$ is around δ or less.

- Use your favorite programming language to perform the simulation.
- Report the average of $\{Y_i\}$ for both the fixed time version and anytime version.
- Pick 10 of the random trials and plot $t \cdot \hat{\mu}_t$ and the both deviation bounds multiplied by t, all three of them in one plot with x-axis being t. (Multiplying t is merely to improve the visual).

Problem 3

In this problem, we will perform experiments with the Catoni estimator. As a toy problem, consider the following linear model

$$y = \langle x, \theta^* \rangle + \eta$$

where $x \in \mathbb{R}^d$ is the feature vector, $\theta^* \in \mathbb{R}^d$ is the unknown parameter, and η is an independent standard normal random variable (zero-mean). We will test PopArt discussed in the lecture. Although it was original designed for the sparse linear model where d is very large, we will consider $d = 2$ for simplicity. Suppose x is drawn from the following distribution

$$x = \begin{cases} x^{(1)} := \binom{1}{-\tau} & \text{with probability } 1 - 2\varepsilon \\ x^{(2)} := \binom{1}{\tau} & \text{with probability } \varepsilon \\ x^{(3)} := \binom{0}{1} & \text{with probability } \varepsilon \end{cases}$$

Let $\varepsilon = \tau^2$. Define $Q = \mathbb{E}[xx^\top]$.

(a) Compute Q^{-1}.

(b) Evaluate $(Q^{-1})_{11}$ and $R_i := |e_1^\top Q^{-1} x^{(i)}|, \forall i \in [3]$. Report which i becomes much larger than $(Q^{-1})_{11}$ (call this i^*). Also, report the ratio $\frac{R_{i^*}}{(Q^{-1})_{11}}$ and explain how this scales with τ. You must see that there exists such an i^*, and this implies that the Catoni estimator step of PopArt is necessary to perform well in the nonasymptotic regime.

(c) (30pts) Implement PopArt and the PopArt without Catoni (call it PopArt$^-$) procedure (the one that simply averages out the per-data-point estimator $\theta^{(i)} = Q^{-1} x_i y_i$ where (x_i, y_i) is the i-th data point). Generate $n = 50$ data points. Estimate the first coordinate of θ^* with Catoni and Catoni$^-$. For Catoni, use the variance parameter of $\sigma^2 = ((1+\tau)^2 + 1)(Q^{-1})_{ii}$ for estimating the i-th coordinate, along with $\delta = 0.05$. Repeat this for $m = 10,000$ times (if this takes too long, feel free to reduce the repetition, but not less than $1,000$). Report the code. Plot the two histogram of the estimators in one plot (e.g., `https://datavizpyr.com/overlapping-histograms-with-matplotlib-in-python/`). Next, plot the empirical CDF of the two estimators in one plot (e.g., `https://www.geeksforgeeks.org/how-to-make-ecdf-plot-with-seaborn-in-python/`) State your observations on how the distribution of the two estimators differ.

Problem 4

How much time did it take you to complete this homework?