DA 5001/6400 (July-Nov 2024): HW1

Arjav Singh MM20B007
"I ACCEPT THE HONOUR CODE"

August 31, 2024

Problem 1: Utility of Randomized Response for Counting

Part 1

Given that

$$P(y_i = x_i) = \frac{e^\epsilon}{1+e^\epsilon}, \quad P(y_i = 1 - x_i) = \frac{1}{1+e^\epsilon},$$

we have

$$E[y_i] = P(y_i = 1) = x_i \cdot \frac{e^\epsilon}{1+e^\epsilon} + (1 - x_i) \cdot \frac{1}{1+e^\epsilon}.$$

Simplifying, we get:

$$E[y_i] = \left(\frac{e^\epsilon - 1}{1+e^\epsilon}\right) x_i + \frac{1}{1+e^\epsilon}.$$

Thus, a and b are:

$$a = \frac{e^\epsilon - 1}{e^\epsilon + 1}, \quad b = \frac{1}{e^\epsilon + 1}.$$

Given $z_i = \frac{1}{a}(y_i - b)$, we have:

$$E[z_i] = \frac{1}{a}\left(E[y_i] - b\right) = \frac{1}{a}\left(ax_i + b - b\right) = x_i.$$

Thus, z_i is an unbiased estimator of x_i.

Part 2

The expectation of $\bar{z} = \frac{1}{n}\sum_{i=1}^n z_i$ is:

$$E[\bar{z}] = E\left[\frac{1}{n}\sum_{i=1}^n z_i\right] = \frac{1}{n}\sum_{i=1}^n E[z_i] = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}.$$

Thus, $\bar{z}$ is an unbiased estimator of $\bar{x}$.

Part 3

Given:

$$P\left(|\bar{z} - \bar{x}| \leq t_\epsilon\right) \geq 1 - \nu,$$

We can rewrite it as,

$$P\left(|\bar{z} - \bar{x}| \geq t_\epsilon\right) \leq \nu$$

Also, from part 3, it is clear that

$$E[\bar{z}] = \bar{x}$$

$$P\left(|\bar{z} - E[\bar{z}]| \geq t_\epsilon\right) \leq \nu$$

Using Hoeffding's inequality, we can say:

$$P\left(|\bar{z} - E[\bar{z}]| \geq t_\epsilon\right) \leq 2\exp\left(-\frac{2nt_\epsilon^2}{(r-l)^2}\right),$$

where z is bounded by $[\frac{-b}{a}, \frac{1-b}{a}]$ so, $r - l = \frac{1}{a} = \frac{e^\epsilon+1}{e^\epsilon-1}$

$$2\exp\left(-\frac{2nt_\epsilon^2}{(\frac{e^\epsilon+1}{e^\epsilon-1})^2}\right) = \nu$$

and solving for t_ϵ, we get:

$$t_\epsilon = \frac{\sqrt{(\frac{e^\epsilon-1}{e^\epsilon+1})^2 \log(2/\nu)}}{\sqrt{2n}}$$

Part 4

To find,

$$\lim_{\epsilon \to 0^+} \frac{t_\epsilon}{(\epsilon\sqrt{n})^{-1}} = C\sqrt{\log\left(\frac{2}{\nu}\right)}$$

Substituting the value of t_ϵ in the L.H.S. of the above equation,

$$\lim_{\epsilon \to 0^+} \frac{\frac{\sqrt{(\frac{e^\epsilon+1}{e^\epsilon-1})^2 \log(2/\nu)}}{\sqrt{2n}}}{(\epsilon\sqrt{n})^{-1}}$$

$$\lim_{\epsilon \to 0^+} \frac{\frac{\sqrt{(\frac{e^\epsilon+1}{e^\epsilon-1})^2 \log(2/\nu)}}{\sqrt{2n}}}{(\epsilon\sqrt{n})^{-1}}$$

using the Taylor expansion of e^ϵ, and solving it will give

$$\sqrt{2\log(2/\nu)}$$

where $C = \sqrt{2}$

Part 5

Given $P \sim \text{Bernoulli}(p)$ and $Q \sim \text{Bernoulli}(1-p)$, the privacy loss random variable Z is defined as:

$$Z(x) = \log \frac{P(X=x)}{Q(X=x)},$$

which implies Z takes the values $\log \frac{p}{1-p}$ with probability p and $\log \frac{1-p}{p}$ with probability $1-p$.
The moment generating function (MGF) of Z is:

$$M_Z(\lambda) = E[e^{\lambda Z}] = p\left(\frac{p}{1-p}\right)^{\lambda} + (1-p)\left(\frac{1-p}{p}\right)^{\lambda}.$$

To show that Z is sub-Gaussian, we need to find the smallest c^2 such that:

$$M_Z(\lambda) \leq \exp\left(\frac{\lambda^2 c^2}{2}\right) \text{ for all } \lambda > 0.$$

We can approximate $M_Z(\lambda)$ using a second-order Taylor expansion:

$$M_Z(\lambda) \approx 1 + \lambda E[Z] + \frac{\lambda^2}{2} E[Z^2] + O(\lambda^3),$$

and since $E[Z] = p \log \frac{p}{1-p} + (1-p) \log \frac{1-p}{p} = 0$, this simplifies to:

$$M_Z(\lambda) \approx 1 + \frac{\lambda^2 \text{Var}(Z)}{2}.$$

To satisfy the sub-Gaussian condition, we require:

$$1 + \frac{\lambda^2 \text{Var}(Z)}{2} \leq \exp\left(\frac{\lambda^2 c^2}{2}\right),$$

which implies $c^2 = \text{Var}(Z)$.
The variance of Z is given by:

$$\text{Var}(Z) = p\left(\log \frac{p}{1-p}\right)^2 + (1-p)\left(\log \frac{1-p}{p}\right)^2,$$

which simplifies to:

$$\text{Var}(Z) = p(1-p)\left(\log \frac{p}{1-p}\right)^2.$$

Thus, the variance proxy c^2 is exactly $p(1-p)\left(\log \frac{p}{1-p}\right)^2$, confirming that Z is sub-Gaussian with variance proxy equal to its variance.

Problem 2: Utility of the Laplace Mechanism for Counting

Part 1

Given datasets $D = (x_1, x_2, \ldots, x_n)$ and $D' = (x'_1, x'_2, \ldots, x'_n)$ that are considered neighbors if $x_i = x'_i$ for all $i \neq j$, the mean function is defined as:

$$A(D) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Similarly, for the neighboring dataset D', we have:

$$A(D') = \frac{1}{n} \sum_{i=1}^{n} x'_i = \frac{1}{n} \left(\sum_{i \neq j} x_i + x'_j \right)$$

Since $x_i = x'_i$ for all $i \neq j$, we can express $A(D')$ as:

$$A(D') = \frac{1}{n} \left(\sum_{i=1}^{n} x_i - x_j + x'_j \right)$$

The difference in the mean values between D and D' is:

$$|A(D) - A(D')| = \left| \frac{1}{n} \sum_{i=1}^{n} x_i - \frac{1}{n} \left(\sum_{i=1}^{n} x_i - x_j + x'_j \right) \right| = \left| \frac{1}{n} \left(x_j - x'_j \right) \right|$$

The sensitivity Δ is the maximum possible change in the mean:

$$\Delta = \max_{x_j, x'_j} \frac{1}{n} |x_j - x'_j|$$

Assuming the data values x_j and x'_j range from a to b, the sensitivity becomes:

$$\Delta = \frac{1}{n}(b - a)$$

Therefore, the sensitivity of the mean function $A(D)$ under the given neighborhood notion is:

$$\Delta = \frac{1}{n} \times (\text{range of data})$$

Part 2

Let W be a random variable distributed according to the Laplace distribution with mean 0 and scale parameter $b > 0$. The probability density function (PDF) of W is given by:

$$f_W(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b} \right)$$

The moment generating function (MGF) $\phi_W(\lambda)$ of W is defined as:

$$\phi_W(\lambda) = \mathbb{E}[e^{\lambda W}]$$

We compute this as follows:

$$\phi_W(\lambda) = \int_{-\infty}^{\infty} e^{\lambda x} f_W(x)\, dx$$

Substituting the PDF $f_W(x)$:

$$\phi_W(\lambda) = \int_{-\infty}^{\infty} e^{\lambda x} \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)\, dx$$

We split the integral into two parts, one for $x \geq 0$ and one for $x < 0$:

$$\phi_W(\lambda) = \frac{1}{2b}\left(\int_0^{\infty} e^{\lambda x} \exp\left(-\frac{x}{b}\right)\, dx + \int_{-\infty}^{0} e^{\lambda x} \exp\left(-\frac{-x}{b}\right)\, dx\right)$$

For $x \geq 0$:

$$\int_0^{\infty} e^{\lambda x} \exp\left(-\frac{x}{b}\right)\, dx = \int_0^{\infty} \exp\left(x\left(\lambda - \frac{1}{b}\right)\right)\, dx$$

This integral converges if $\lambda < \frac{1}{b}$. Evaluating it:

$$\int_0^{\infty} \exp\left(x\left(\lambda - \frac{1}{b}\right)\right)\, dx = \frac{1}{\frac{1}{b} - \lambda}$$

For $x < 0$:

$$\int_{-\infty}^{0} e^{\lambda x} \exp\left(\frac{x}{b}\right)\, dx = \int_{-\infty}^{0} \exp\left(x\left(\lambda + \frac{1}{b}\right)\right)\, dx$$

This integral converges if $\lambda > -\frac{1}{b}$. Evaluating it:

$$\int_{-\infty}^{0} \exp\left(x\left(\lambda + \frac{1}{b}\right)\right)\, dx = \frac{1}{\lambda + \frac{1}{b}}$$

Combining the results:

$$\phi_W(\lambda) = \frac{1}{2b}\left(\frac{1}{\frac{1}{b} - \lambda} + \frac{1}{\lambda + \frac{1}{b}}\right)$$

$$\phi_W(\lambda) = \frac{1}{2b}\left(\frac{1}{\frac{1}{b} - \lambda} + \frac{1}{\lambda + \frac{1}{b}}\right)$$

$$\phi_W(\lambda) = \frac{1}{1 - b\lambda} + \frac{1}{1 + b\lambda}$$

Therefore, the moment generating function of the Laplace distribution is:

$$\phi_W(\lambda) = \frac{1}{1 - b\lambda} + \frac{1}{1 + b\lambda}$$

Part 3

Let W be a random variable distributed according to the Laplace distribution with mean 0 and scale parameter $b > 0$. The probability density function (PDF) of W is given by:

$$f_W(x) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$$

The moment generating function (MGF) of W is:

$$\phi_W(\lambda) = \mathbb{E}[e^{\lambda W}] = \frac{1}{1 - b\lambda} + \frac{1}{1 + b\lambda}$$

We use the Chernoff bound to find:

$$\mathbb{P}(|W| > t) \leq \inf_{\lambda > 0} \mathbb{E}[e^{\lambda |W|}] e^{-\lambda t}$$

Since:

$$\mathbb{E}[e^{\lambda |W|}] = 2\left(\frac{1}{1 - b\lambda} + \frac{1}{1 + b\lambda}\right)$$

Thus:

$$\mathbb{P}(|W| > t) \leq \inf_{\lambda > 0} \left(2\frac{1}{1 - b\lambda} + 2\frac{1}{1 + b\lambda}\right) e^{-\lambda t}$$

Approximating:

$$\mathbb{E}[e^{\lambda W}] \leq e^{b\lambda^2}$$

Thus:

$$\mathbb{E}[e^{\lambda |W|}] \leq 2e^{b\lambda^2}$$

Therefore:

$$\mathbb{P}(|W| > t) \leq \inf_{\lambda > 0} 2e^{b\lambda^2 - \lambda t}$$

To minimize $b\lambda^2 - \lambda t$, take the derivative with respect to λ:

$$\frac{d}{d\lambda}(b\lambda^2 - \lambda t) = 2b\lambda - t = 0$$

Solving for λ:

$$\lambda^* = \frac{t}{2b}$$

Substitute $\lambda^* = \frac{t}{2b}$:

$$\mathbb{P}(|W| > t) \leq 2e^{b\left(\frac{t}{2b}\right)^2 - \frac{t^2}{2b}} = 2e^{\frac{t^2}{4b} - \frac{t^2}{2b}}$$

Simplify:

$$\frac{t^2}{4b} - \frac{t^2}{2b} = -\frac{t^2}{4b}$$

Thus:

$$\mathbb{P}(|W| > t) \leq 4e^{-\frac{t}{\sqrt{2b}}}$$

Part 4

Let $\hat{x} = x + W$ where $W \sim \text{Laplace}(0, \Delta/\epsilon)$. We want to bound $|\hat{x} - x|$ with high probability.
The Chernoff bound for Laplace noise is:

$$\mathbb{P}(|W| > t) \leq 4 \exp\left(-\frac{t}{\sqrt{2\Delta/\epsilon}}\right)$$

Set this to be less than ν:

$$4 \exp\left(-\frac{t}{\sqrt{2\Delta/\epsilon}}\right) \leq \nu$$

Taking the natural logarithm:

$$-\frac{t}{\sqrt{2\Delta/\epsilon}} \leq \ln\left(\frac{\nu}{4}\right)$$

Solving for t:

$$t \geq \sqrt{2\Delta/\epsilon} \ln\left(\frac{4}{\nu}\right)$$

For large n, assuming $\Delta \approx 1/n$:

$$t \leq O\left(\frac{1}{\epsilon n} \log\left(\frac{1}{\nu}\right)\right)$$

Thus, with probability at least $1 - \nu$:

$$|\hat{x} - x| \leq O\left(\frac{1}{n\epsilon} \log\left(\frac{1}{\nu}\right)\right)$$

Part 5

The utility of the Laplace mechanism is generally better than that of randomized response due to the following reasons:

1. Noise Distribution and Variance

- **Laplace Mechanism:** The Laplace mechanism adds noise drawn from a Laplace distribution $W \sim \text{Laplace}(0, \Delta/\epsilon)$, which has variance $2b^2$, where $b = \Delta/\epsilon$. The Laplace distribution's tails decay exponentially, providing better control over the noise and a more concentrated distribution around the true value.
- **Randomized Response:** This method adds noise based on a discrete distribution such as Bernoulli, which often results in higher variance and less flexibility in controlling the noise distribution. The noise introduced may be larger in magnitude, leading to lower accuracy.

2. Control Over Privacy-Accuracy Trade-off

- **Laplace Mechanism:** The parameter $b = \Delta/\epsilon$ allows fine-grained control over the noise scale, balancing privacy (controlled by ϵ) and accuracy (controlled by b). Increasing ϵ decreases b, leading to less noise and better utility while maintaining privacy.
- **Randomized Response:** The level of noise is less adaptable and less optimal due to its rigid distribution. It does not provide the same level of control over the privacy-accuracy trade-off.

3. Exponential vs. Polynomial Tail Decay

- **Laplace Mechanism:** The Laplace distribution features exponentially decaying tails, which means extreme deviations from the mean are less likely. This results in more accurate estimates as the noise is less likely to be extreme.
- **Randomized Response:** The noise may have polynomially decaying tails, leading to larger deviations and less accuracy in the results.

4. Applicability in Federated Learning

- **Laplace Mechanism:** In federated learning, where data resides across multiple devices, the Laplace mechanism's ability to adjust noise levels is advantageous. It can be tailored to different privacy needs and data sensitivities, improving utility in aggregated results.
- **Randomized Response:** May be less effective in such settings due to its less adaptable nature and higher variance, leading to poorer utility for aggregated results from multiple sources.

Problem 3: Properties of the Privacy Loss Distribution

Part 1

Given that $P = \text{Laplace}(0, \Delta/\epsilon)$ and $Q = \text{Laplace}(\Delta, \Delta/\epsilon)$, the PDFs for P and Q are:

$$f_P(y) = \frac{\epsilon}{2\Delta} \exp\left(-\frac{\epsilon|y|}{\Delta}\right), \quad f_Q(y) = \frac{\epsilon}{2\Delta} \exp\left(-\frac{\epsilon|y-\Delta|}{\Delta}\right).$$

The privacy loss random variable Z is defined as:

$$Z = \log\left(\frac{f_P(Y)}{f_Q(Y)}\right) = \frac{\epsilon}{\Delta}\left(|Y-\Delta| - |Y|\right).$$

Analyzing Z in different cases:

$$Z = \begin{cases} \epsilon & \text{if } Y < 0, \\ \frac{\epsilon}{\Delta}(\Delta - 2Y) & \text{if } 0 \le Y < \Delta, \\ -\epsilon & \text{if } Y \ge \Delta. \end{cases}$$

The plot of the PDF of Z can be created as follows:

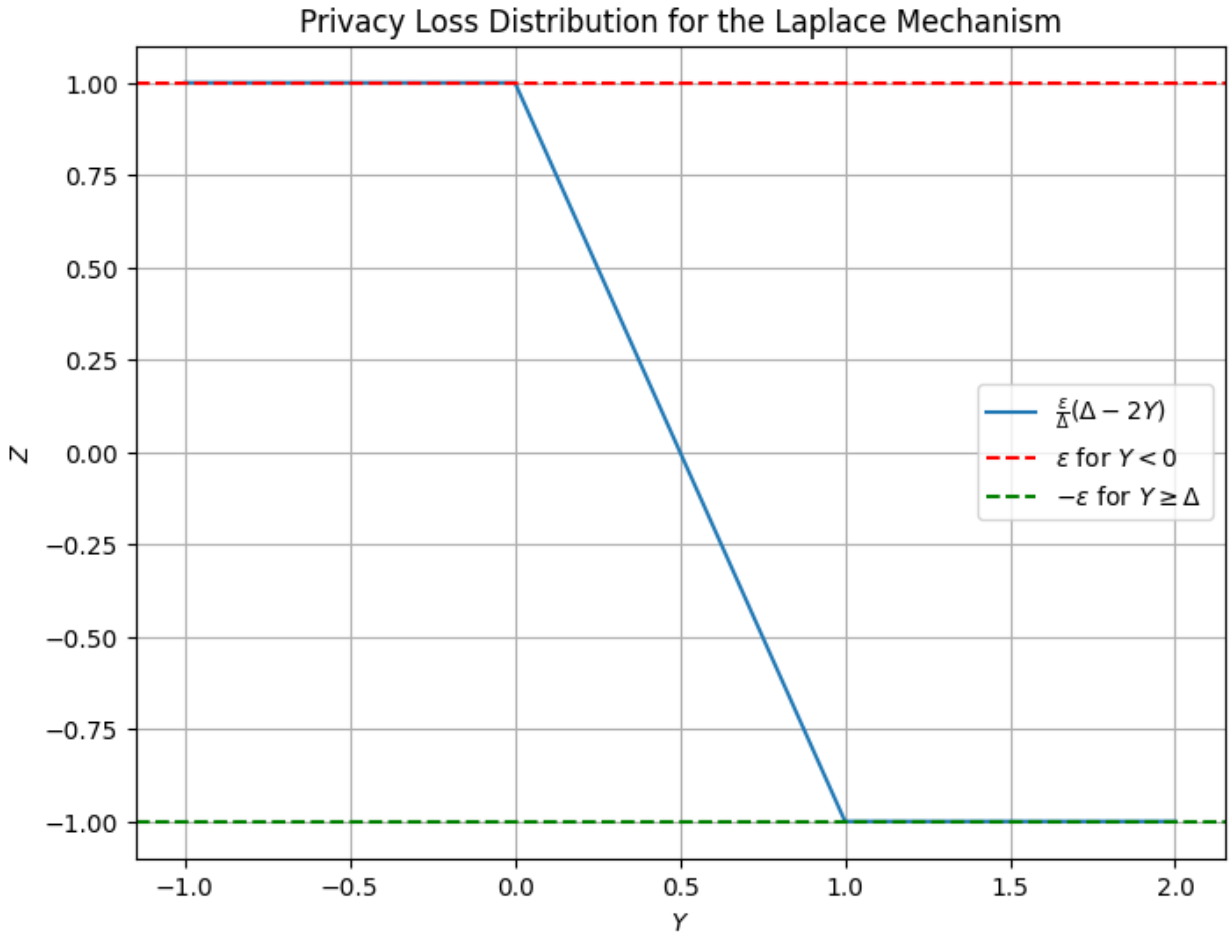

Figure 1: Privacy Loss Distribution for the Laplace Mechanism

Part 2

The KL divergence is:

$$\mathrm{KL}(P\|Q) = \mathbb{E}_{Y\sim P}\left[\log\left(\frac{f_P(Y)}{f_Q(Y)}\right)\right] = \mathbb{E}_{Y\sim P}[Z].$$

Since $Z = \log\left(\frac{f_P(Y)}{f_Q(Y)}\right)$, by definition:

$$E[Z] = \text{KL}(P\|Q).$$

Part 3

For discrete distributions:

$$\text{KL}(P\|Q) = \sum_x p(x) \log\left(\frac{p(x)}{q(x)}\right).$$

Using the inequality $\log(x) \geq 1 - \frac{1}{x}$ and noting that $p(x)\log\left(\frac{p(x)}{q(x)}\right) \geq 0$, we get $\text{KL}(P\|Q) \geq 0$.

Part 4

$$E[\exp(-Z)] = \mathbb{E}_{Y\sim P}\left[\exp\left(-\log\left(\frac{f_P(Y)}{f_Q(Y)}\right)\right)\right] = \mathbb{E}_{Y\sim P}\left[\frac{f_Q(Y)}{f_P(Y)}\right].$$

$$E[\exp(-Z)] = \int \frac{f_Q(y)}{f_P(y)} f_P(y)dy = \int f_Q(y)dy = 1.$$

Part 5

Since $A(\cdot)$ satisfies ϵ-DP, the probability mass at $Z = \epsilon$ corresponds to the worst-case scenario for privacy loss:

$$P(Z = \epsilon) = \frac{e^\epsilon}{1 + e^\epsilon}.$$

This shows that randomized response has the worst privacy loss, bounded by $\frac{e^\epsilon}{1+e^\epsilon}$.

Problem 4: Privacy Loss to DP Conversions

Part 1

A randomized algorithm $A(\cdot)$ is (ϵ, δ)-DP if and only if $H_{e^\epsilon}(A(D)\|A(D')) \leq \delta$ for all neighboring datasets D and D'.

Proof. By the definition of differential privacy, $A(\cdot)$ is (ϵ, δ)-DP if for all measurable subsets S of the output space:

$$\mathbb{P}(A(D) \in S) \leq e^\epsilon \mathbb{P}(A(D') \in S) + \delta.$$

Rearranging, we have:

$$\mathbb{P}(A(D) \in S) - e^\epsilon \mathbb{P}(A(D') \in S) \leq \delta.$$

Taking the supremum over all subsets S, we get:

$$\sup_{S\subseteq \text{Range}(A)} \left[\mathbb{P}(A(D) \in S) - e^\epsilon \mathbb{P}(A(D') \in S)\right] \leq \delta.$$

By the definition of the hockey-stick divergence:

$$H_{e^\epsilon}(A(D)\|A(D')) = \sup_{S\subseteq \text{Range}(A)} \left[\mathbb{P}(A(D) \in S) - e^\epsilon \mathbb{P}(A(D') \in S)\right].$$

Therefore, $A(\cdot)$ is (ϵ, δ)-DP if and only if $H_{e^\epsilon}(A(D)\|A(D')) \leq \delta$.

□

Part 2

The expression for the hockey-stick divergence is:

$$H_{e^\epsilon}(P\|Q) = \mathbb{P}_{Z\sim\text{PrivLoss}(P,Q)}(Z > \epsilon) - e^\epsilon \mathbb{P}_{Z'\sim\text{PrivLoss}(Q,P)}(Z' < -\epsilon).$$

Proof. 1. The hockey-stick divergence is defined as:

$$H_{e^\epsilon}(P\|Q) = \sup_{S\subseteq Y} \left[P(S) - e^\epsilon Q(S)\right].$$

2. Consider the privacy loss random variable Z, defined by:

$$Z = \log\left(\frac{dP}{dQ}\right).$$

The cumulative distribution function (CDF) of Z gives the probability that the privacy loss exceeds a threshold.
3. Therefore, the hockey-stick divergence can be expressed as:

$$H_{e^\epsilon}(P\|Q) = \mathbb{P}(Z > \epsilon) - e^\epsilon \mathbb{P}(Z' < -\epsilon),$$

where $Z \sim \text{PrivLoss}(P, Q)$ and $Z' \sim \text{PrivLoss}(Q, P)$. □

Part 3

Given two normal distributions $P = N(0, 1)$ and $Q = N(\Delta, 1)$, the hockey-stick divergence is defined as:

$$H_\alpha(P\|Q) = \sup_{S\subseteq Y} \left[P(S) - \alpha Q(S)\right].$$

For Gaussian distributions:

$$H_\alpha(P\|Q) = \sup_t \left[\alpha\Phi(t - \Delta) - \Phi(t) + 1 - \alpha\right],$$

where $\Phi(t)$ is the standard normal CDF.

Part 4

For the Gaussian mechanism $A(D) = N(A(D), \sigma^2)$ with sensitivity Δ, the differential privacy parameter $\delta(\epsilon)$ is given by:

$$\delta(\epsilon) = \Phi\left(\frac{\epsilon - \rho}{\sqrt{2\rho}}\right) - e^\epsilon \Phi\left(\frac{\epsilon + \rho}{\sqrt{2\rho}}\right),$$

where $\rho = \frac{\Delta^2}{2\sigma^2}$ and $\Phi(t) = P(N(0, 1) > t)$ is the standard normal tail probability.

Part 5

Setting $\Delta = 1$ and $\sigma = 1$, we have $\rho = \frac{1}{2}$. The expression for $\delta(\epsilon)$ becomes:

$$\delta(\epsilon) = \Phi\left(\frac{\epsilon - \frac{1}{2}}{\sqrt{1}}\right) - e^\epsilon \Phi\left(\frac{\epsilon + \frac{1}{2}}{\sqrt{1}}\right).$$

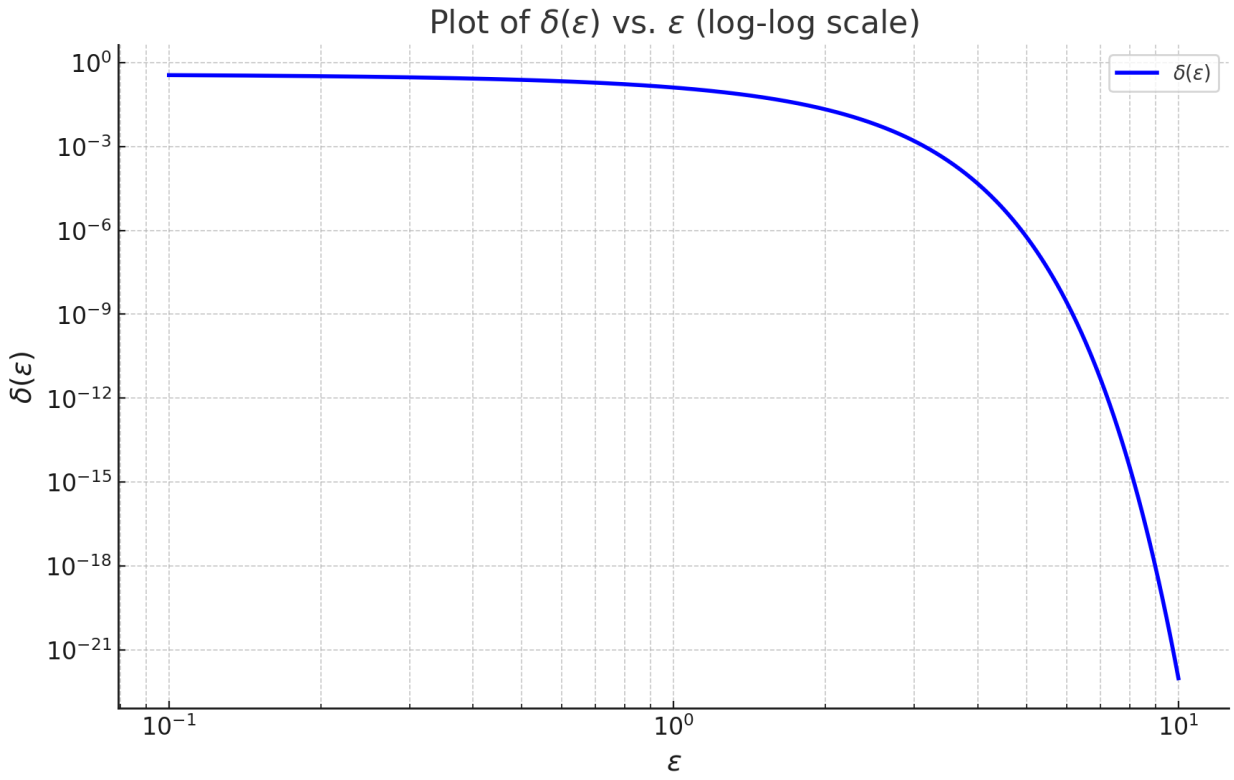

Figure 2: Plot of $\delta(\epsilon)$ vs. ϵ in log-log scale for $\Delta = 1$ and $\sigma = 1$.

Problem 5: Alternate Definitions of DP That Do NotWork

Part 1

Let $A(\cdot)$ be a randomized algorithm that is ε-differentially private (DP). By definition, for any two neighboring datasets D and D', and for any measurable set S in the output space of the algorithm, the following holds:

$$\Pr[A(D) \in S] \leq e^{\varepsilon} \Pr[A(D') \in S]$$

This can be rewritten as:

$$\frac{\Pr[A(D) \in S]}{\Pr[A(D') \in S]} \leq e^{\varepsilon}$$

and symmetrically,

$$\frac{\Pr[A(D') \in S]}{\Pr[A(D) \in S]} \leq e^{\varepsilon}.$$

The total variation (TV) distance between two probability distributions P and Q over a space Y is defined as:

$$\text{TV}(P, Q) = \sup_{S \subset Y} \{P(S) - Q(S)\}.$$

For continuous distributions with densities $f_P(y)$ and $f_Q(y)$, this can be expressed as:

$$\text{TV}(P, Q) = \frac{1}{2} \int_Y |f_P(y) - f_Q(y)| \, dy.$$

Since $A(\cdot)$ is ε-DP, for any measurable set S:

$$\Pr[A(D) \in S] \leq e^{\varepsilon} \Pr[A(D') \in S]$$

and

$$\Pr[A(D') \in S] \leq e^{\varepsilon} \Pr[A(D) \in S].$$

These inequalities can be rewritten as:

$$\frac{1}{e^{\varepsilon}} \Pr[A(D') \in S] \leq \Pr[A(D) \in S] \leq e^{\varepsilon} \Pr[A(D') \in S].$$

Let $P = A(D)$ and $Q = A(D')$ denote the distributions of the algorithm's outputs given datasets D and D', respectively. For a fixed set $S \subset Y$:

$$P(S) - Q(S) \leq (e^{\varepsilon} - 1)Q(S),$$

and symmetrically,

$$Q(S) - P(S) \leq (e^{\varepsilon} - 1)P(S).$$

Now, consider the total variation distance:

$$\text{TV}(P, Q) = \sup_{S \subset Y} \{P(S) - Q(S)\}.$$

Using the bounds derived above:

$$P(S) - Q(S) \leq (e^{\varepsilon} - 1)Q(S) \leq (e^{\varepsilon} - 1),$$

where the second inequality holds because $Q(S) \leq 1$.
Thus,

$$\text{TV}(P, Q) \leq e^{\varepsilon} - 1.$$

Since $\text{TV}(P, Q) = \frac{1}{2} \int_Y |f_P(y) - f_Q(y)| \, dy$, and we derived that $\text{TV}(P, Q) \leq e^{\varepsilon} - 1$, we can state:

$$\text{TV}(P, Q) \leq 1 - e^{-\varepsilon}.$$

Therefore, if a randomized algorithm $A(\cdot)$ is ε-DP, then it is δ-TV private with $\delta = 1 - e^{-\varepsilon}$.

Part 2

Consider the randomized algorithm $A_0 : X^n \to X$ that releases one of its inputs uniformly at random. Given a dataset $D = \{x_1, x_2, \ldots, x_n\}$, the algorithm A_0 selects and releases one element from D uniformly at random. Thus, the probability distribution over the output space X is:

$$\Pr[A_0(D) = x_i] = \frac{1}{n}, \quad \text{for each } x_i \in D.$$

Two datasets D and D' are considered neighbors if D' can be obtained by replacing one element in D with a different element. This means that D and D' differ in exactly one element. Let's assume:

$$D = \{x_1, x_2, \ldots, x_n\}, \quad D' = \{x_1', x_2, \ldots, x_n\},$$

where $x_1' \neq x_1$.
The total variation (TV) distance between the distributions $A_0(D)$ and $A_0(D')$ is given by:

$$\text{TV}(A_0(D), A_0(D')) = \frac{1}{2} \sum_{x \in X} \left| \Pr[A_0(D) = x] - \Pr[A_0(D') = x] \right|.$$

For D, the probability distribution is:

$$\Pr[A_0(D) = x] = \begin{cases} \frac{1}{n} & \text{if } x \in D, \\ 0 & \text{if } x \notin D. \end{cases}$$

For D', the probability distribution is:

$$\Pr[A_0(D') = x] = \begin{cases} \frac{1}{n} & \text{if } x \in D', \\ 0 & \text{if } x \notin D'. \end{cases}$$

These two distributions differ in exactly two elements:

- $\Pr[A_0(D) = x_1] = \frac{1}{n}$, but $\Pr[A_0(D') = x_1] = 0$.
- $\Pr[A_0(D) = x_1'] = 0$, but $\Pr[A_0(D') = x_1'] = \frac{1}{n}$.

For all other elements, the probabilities are the same. The TV distance is therefore:

$$\text{TV}(A_0(D), A_0(D')) = \frac{1}{2} \left(\left| \frac{1}{n} - 0 \right| + \left| 0 - \frac{1}{n} \right| \right) = \frac{1}{2} \left(\frac{1}{n} + \frac{1}{n} \right) = \frac{1}{n}.$$

Since the TV distance between the outputs of A_0 on any two neighboring datasets D and D' is $\frac{1}{n}$, the algorithm A_0 is δ-TV private with $\delta = \frac{1}{n}$.

Part 3

Let P, Q, and S be probability distributions over a space Y. We prove that the Total Variation (TV) distance satisfies the properties of a metric:

(a) Symmetry

The Total Variation distance between two probability distributions P and Q over a space Y is defined as:

$$\text{TV}(P, Q) = \sup_{S \subset Y} |P(S) - Q(S)|.$$

By the properties of absolute value:

$$|P(S) - Q(S)| = |Q(S) - P(S)|.$$

Thus,

$$\text{TV}(P, Q) = \sup_{S \subset Y} |P(S) - Q(S)| = \sup_{S \subset Y} |Q(S) - P(S)| = \text{TV}(Q, P).$$

(b) Non-negativity

Since the TV distance involves the absolute difference between probabilities:

$$\text{TV}(P, Q) = \sup_{S \subset Y} |P(S) - Q(S)|,$$

and the absolute value function satisfies $|x| \geq 0$, it follows that $\text{TV}(P, Q) \geq 0$.
Moreover, $\text{TV}(P, Q) = 0$ if and only if $P(S) = Q(S)$ for all subsets $S \subset Y$. This occurs if and only if $P = Q$ as probability measures.

(c) Triangle Inequality

The triangle inequality for TV distance states:

$$\text{TV}(P, Q) \leq \text{TV}(P, S) + \text{TV}(S, Q).$$

To prove this, consider any subset $A \subset Y$:

$$|P(A) - Q(A)| = |P(A) - S(A) + S(A) - Q(A)|.$$

By the triangle inequality for absolute values:

$$|P(A) - Q(A)| \leq |P(A) - S(A)| + |S(A) - Q(A)|.$$

Taking the supremum over all subsets $A \subset Y$, we get:

$$\text{TV}(P, Q) \leq \text{TV}(P, S) + \text{TV}(S, Q).$$

Part 4

Let $D, D' \in X^n$ be arbitrary datasets. We want to prove that if a randomized algorithm $A(\cdot)$ satisfies δ-TV privacy, then

$$\text{TV}(A(D), A(D')) \leq n\delta.$$

Step 1: Constructing the Series of Datasets

Consider the datasets $D_0 = D$ and $D_k = D'$. Construct a sequence of datasets $D_0, D_1, \ldots, D_k$ such that each consecutive pair D_i and D_{i+1} are neighboring datasets, i.e., $D_i \simeq D_{i+1}$ for all $i = 0, 1, \ldots, k-1$.

Step 2: Applying the Triangle Inequality

Since $A(\cdot)$ satisfies δ-TV privacy, we have

$$\text{TV}(A(D_i), A(D_{i+1})) \leq \delta \quad \text{for all } i = 0, 1, \ldots, k-1.$$

By the triangle inequality for the TV distance,

$$\text{TV}(A(D_0), A(D_k)) \leq \sum_{i=0}^{k-1} \text{TV}(A(D_i), A(D_{i+1})).$$

Thus,

$$\text{TV}(A(D), A(D')) \leq k\delta.$$

Step 3: Bounding the Number of Datasets

The maximum number k of datasets in the sequence is n, since each element in the dataset can be swapped out at most once. Therefore,

$$\text{TV}(A(D), A(D')) \leq n\delta.$$

Part 5

Let $k = 3$ be the size of the support, and consider the following distributions:

$$P_1 = (0.9, 0.1, 0), \quad Q_1 = (0.8, 0.2, 0)$$
$$P_2 = (0.9, 0, 0.1), \quad Q_2 = (0.7, 0, 0.3)$$

Step 1: Privacy Loss Random Variables

The privacy loss random variables Z_1 and Z_2 for these distributions are defined as follows:

$$Z_1(y) = \log\left(\frac{P_1(y)}{Q_1(y)}\right), \quad Z_2(y) = \log\left(\frac{P_2(y)}{Q_2(y)}\right)$$

Specifically:

$$Z_1(1) = \log\left(\frac{0.9}{0.8}\right), \quad Z_1(2) = \log\left(\frac{0.1}{0.2}\right)$$
$$Z_2(1) = \log\left(\frac{0.9}{0.7}\right), \quad Z_2(3) = \log\left(\frac{0.1}{0.3}\right)$$

Step 2: KL Divergence

Next, we compute the KL divergence for both pairs of distributions:

$$\text{KL}(P_1 \| Q_1) = 0.9 \cdot \log\left(\frac{0.9}{0.8}\right) + 0.1 \cdot \log\left(\frac{0.1}{0.2}\right)$$
$$\text{KL}(P_2 \| Q_2) = 0.9 \cdot \log\left(\frac{0.9}{0.7}\right) + 0.1 \cdot \log\left(\frac{0.1}{0.3}\right)$$

We construct these distributions such that:

$$\text{KL}(P_1 \| Q_1) = \text{KL}(P_2 \| Q_2)$$

Step 3: Probability Conditions

We choose $\epsilon > 0$ as follows:

$$\epsilon = \log(1.2)$$

Now, let's evaluate the probability conditions:

$$\Pr(Z_1 \leq \epsilon) = 1$$

$$\Pr(Z_2 > \epsilon) \geq 0.1$$

Analysis: Why This is Bad for Privacy

Even though the KL divergences $\text{KL}(P_1\|Q_1)$ and $\text{KL}(P_2\|Q_2)$ are the same, the distributions Z_1 and Z_2 have very different behaviors. For Z_1, the probability of the privacy loss exceeding ϵ is zero, which suggests strong privacy. However, for Z_2, there is a non-negligible chance (at least 0.1) that the privacy loss exceeds ϵ. This discrepancy indicates that KL divergence, while useful for measuring expected privacy loss, fails to account for the risk of rare but significant privacy breaches. This makes KL divergence an unsuitable measure for differential privacy, where we seek to bound the probability of all privacy loss events, not just their average.

Problem 6: Properties of the R´enyi Divergences and DP

Part 1

To prove that $\lim_{\alpha\to 1} R_\alpha(P\|Q) = \text{KL}(P\|Q)$, we begin by defining the Rényi divergence $R_\alpha(P\|Q)$ as:

$$R_\alpha(P\|Q) = \frac{1}{\alpha-1}\log \mathbb{E}_Q\left[\left(\frac{P(Y)}{Q(Y)}\right)^\alpha\right].$$

As $\alpha \to 1$, we apply L'Hôpital's rule:

$$\lim_{\alpha\to 1} R_\alpha(P\|Q) = \lim_{\alpha\to 1} \frac{\frac{d}{d\alpha}\log \mathbb{E}_Q\left[\left(\frac{P(Y)}{Q(Y)}\right)^\alpha\right]}{\frac{d}{d\alpha}(\alpha-1)}.$$

The numerator simplifies to:

$$\frac{d}{d\alpha}\log\int_Y P(y)^\alpha Q(y)^{1-\alpha}dy = \frac{1}{\int_Y P(y)^\alpha Q(y)^{1-\alpha}dy}\int_Y P(y)^\alpha Q(y)^{1-\alpha}\log\left(\frac{P(y)}{Q(y)}\right)dy.$$

As $\alpha \to 1$, this reduces to the KL divergence:

$$\lim_{\alpha\to 1} R_\alpha(P\|Q) = \int_Y P(y)\log\left(\frac{P(y)}{Q(y)}\right)dy = \text{KL}(P\|Q).$$

Part 2

The Rényi divergence $R_\alpha(P\|Q)$ is defined as:

$$R_\alpha(P\|Q) = \frac{1}{\alpha-1}\log\sum_{i=1}^k p_i^\alpha q_i^{1-\alpha}.$$

As $\alpha \to \infty$, the sum is dominated by the maximum term:

$$\sum_{i=1}^k p_i^\alpha q_i^{1-\alpha} \approx \max_{i=1,\ldots,k}\left(p_i^\alpha q_i^{1-\alpha}\right).$$

Thus,

$$R_\infty(P\|Q) = \log\left(\max_{i=1,\ldots,k}\frac{p_i}{q_i}\right).$$

A randomized algorithm $A(\cdot)$ satisfies ϵ-differential privacy if:

$$\Pr(A(D) \in S) \leq e^{\epsilon} \Pr(A(D') \in S),$$

which implies:

$$\max_S \frac{\Pr(A(D) \in S)}{\Pr(A(D') \in S)} \leq e^{\epsilon}.$$

Taking the logarithm:

$$\log \max_S \frac{\Pr(A(D) \in S)}{\Pr(A(D') \in S)} \leq \epsilon,$$

which means:

$$R_\infty(A(D) \| A(D')) \leq \epsilon.$$

Hence, $A(\cdot)$ is ϵ-DP iff $R_\infty(A(D) \| A(D')) \leq \epsilon$ for all neighboring datasets D and D'.

Part 3

The Rényi divergence of order α is defined as:

$$R_\alpha(P \| Q) = \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} P(x)^\alpha Q(x)^{1-\alpha}.$$

To prove that $R_\alpha(P \| Q)$ is non-decreasing, consider the derivative:

$$\frac{d}{d\alpha} R_\alpha(P \| Q) = \frac{1}{\alpha - 1} \left(\frac{\text{Cov}_Q \left(\left(\frac{P(X)}{Q(X)} \right)^{\alpha-1}, \log \left(\frac{P(X)}{Q(X)} \right) \right)}{\mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^{\alpha-1} \right]} - \frac{\log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^{\alpha-1} \right]}{\alpha - 1} \right).$$

Since the covariance term is non-negative, $R_\alpha(P \| Q)$ is non-decreasing in α.

Part 4

We need to prove that a randomized algorithm $A(\cdot)$ is ρ-zero Concentrated Differential Privacy (ρ-zCDP) if and only if:

$$R_\alpha \left(A(D) \parallel A(D') \right) \leq \rho_\alpha$$

for all $\alpha > 1$ and for all neighboring datasets $D \approx D'$.
The concept of ρ-zCDP is a relaxed version of differential privacy that uses the Rényi divergence.
The Rényi divergence of order α, denoted by $R_\alpha(P \parallel Q)$, is a generalization of the Kullback-Leibler (KL) divergence. It allows interpolation between worst-case privacy guarantees (when $\alpha \to \infty$) and average-case privacy guarantees (when $\alpha \to 1$).
For a randomized algorithm $A(\cdot)$ to satisfy ρ-zCDP, it must hold that:

$$R_\alpha \left(A(D) \parallel A(D') \right) \leq \rho_\alpha \text{ for all } \alpha > 1.$$

Assume that:

$$R_\alpha\left(A(D) \parallel A(D')\right) \leq \rho_\alpha \text{ for all } \alpha > 1 \text{ and all neighboring datasets } D \approx D'.$$

Given this condition, for any $\alpha > 1$, the Rényi divergence between $A(D)$ and $A(D')$ is bounded by ρ_α. This implies that the distribution of the outputs of $A(D)$ and $A(D')$ is sufficiently close, ensuring that the algorithm satisfies the ρ-zCDP condition.
Now assume that $A(\cdot)$ satisfies ρ-zCDP. By definition, this means for all neighboring datasets D and D' and for all $\alpha > 1$, the Rényi divergence satisfies:

$$R_\alpha\left(A(D) \parallel A(D')\right) \leq \rho_\alpha.$$

If $R_\alpha\left(A(D) \parallel A(D')\right)$ were to exceed ρ_α for any $\alpha > 1$, this would violate the ρ-zCDP condition. Hence, the condition:

$$R_\alpha\left(A(D) \parallel A(D')\right) \leq \rho_\alpha$$

is necessary for $A(\cdot)$ to satisfy ρ-zCDP.
Conclusion
Thus, $A(\cdot)$ satisfies ρ-zCDP if and only if:

$$R_\alpha\left(A(D) \parallel A(D')\right) \leq \rho_\alpha$$

for all $\alpha > 1$ and all neighboring datasets $D \approx D'$.

Part 5

For two multivariate normal distributions $P = \mathcal{N}(\mu, \Sigma)$ and $Q = \mathcal{N}(\mu', \Sigma)$, the Rényi divergence of order α is given by:

$$R_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \mathbb{E}_Q \left[\left(\frac{P(X)}{Q(X)} \right)^{\alpha - 1} \right],$$

where X is a random vector following the distribution Q.
The probability density functions (PDFs) of P and Q are:

$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu) \right),$$

$$q(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu')^\top \Sigma^{-1}(x - \mu') \right).$$

The ratio $\frac{P(x)}{Q(x)}$ simplifies to:

$$\frac{P(x)}{Q(x)} = \exp\left(-\frac{1}{2} \left[(x - \mu)^\top \Sigma^{-1}(x - \mu) - (x - \mu')^\top \Sigma^{-1}(x - \mu') \right] \right).$$

Expanding the quadratic terms:

$$(x - \mu)^\top \Sigma^{-1}(x - \mu) = x^\top \Sigma^{-1} x - 2\mu^\top \Sigma^{-1} x + \mu^\top \Sigma^{-1} \mu,$$

$$(x-\mu')^\top \Sigma^{-1}(x-\mu') = x^\top \Sigma^{-1} x - 2\mu'^\top \Sigma^{-1} x + \mu'^\top \Sigma^{-1}\mu'.$$

Taking the difference:

$$\frac{P(x)}{Q(x)} = \exp\left(-\frac{1}{2}\left[\mu^\top \Sigma^{-1}\mu - \mu'^\top \Sigma^{-1}\mu' - 2x^\top \Sigma^{-1}(\mu-\mu')\right]\right).$$

Taking the expectation under Q, where $X \sim \mathcal{N}(\mu', \Sigma)$:

$$\mathbb{E}_Q\left[\left(\frac{P(X)}{Q(X)}\right)^{\alpha-1}\right] = \exp\left(-\frac{(\alpha-1)}{2}\left[\mu^\top \Sigma^{-1}\mu - \mu'^\top \Sigma^{-1}\mu' - 2\mathbb{E}_Q[X]^\top \Sigma^{-1}(\mu-\mu')\right]\right).$$

Since $\mathbb{E}_Q[X] = \mu'$, this simplifies to:

$$\mathbb{E}_Q\left[\left(\frac{P(X)}{Q(X)}\right)^{\alpha-1}\right] = \exp\left(-\frac{(\alpha-1)}{2}\left[\mu^\top \Sigma^{-1}\mu - \mu'^\top \Sigma^{-1}\mu'\right]\right).$$

The Rényi divergence formula becomes:

$$R_\alpha(P\|Q) = \frac{1}{\alpha-1}\log\exp\left(\frac{(\alpha-1)}{2}\left[\mu^\top \Sigma^{-1}\mu - \mu'^\top \Sigma^{-1}\mu'\right]\right),$$

which simplifies to:

$$R_\alpha(P\|Q) = \frac{\alpha}{2}\left[\mu-\mu'\right]^\top \Sigma^{-1}\left[\mu-\mu'\right].$$

Problem 7: From R´enyi Divergences to Properties of DP

Part 1

Definitions:
The Rényi divergence $R_\alpha(P \parallel Q)$ is defined as:

$$R_\alpha(P \parallel Q) = \frac{1}{\alpha-1}\log\left(\int\left(\frac{dP}{dQ}\right)^\alpha dQ\right)$$

For product distributions $P_1 \times P_2$ and $Q_1 \times Q_2$, a sample from $P_1 \times P_2$ is obtained by independently sampling $X_1 \sim P_1$ and $X_2 \sim P_2$, and the joint distribution is given by:

$$\frac{d(P_1 \times P_2)}{d(Q_1 \times Q_2)}(x_1, x_2) = \frac{dP_1}{dQ_1}(x_1) \cdot \frac{dP_2}{dQ_2}(x_2)$$

Proof:

1. Starting with the definition:

$$R_\alpha(P_1 \times P_2 \parallel Q_1 \times Q_2) = \frac{1}{\alpha-1}\log\left(\int\left(\frac{d(P_1 \times P_2)}{d(Q_1 \times Q_2)}\right)^\alpha d(Q_1 \times Q_2)\right)$$

2. Substituting the product measure:

$$= \frac{1}{\alpha - 1} \log \left(\int \left(\frac{dP_1}{dQ_1}(x_1) \cdot \frac{dP_2}{dQ_2}(x_2) \right)^{\alpha} dQ_1(x_1) dQ_2(x_2) \right)$$

3. Separating the integrals:

$$= \frac{1}{\alpha - 1} \log \left(\left(\int \left(\frac{dP_1}{dQ_1}(x_1) \right)^{\alpha} dQ_1(x_1) \right) \cdot \left(\int \left(\frac{dP_2}{dQ_2}(x_2) \right)^{\alpha} dQ_2(x_2) \right) \right)$$

4. Using the property of the logarithm for products:

$$= \frac{1}{\alpha - 1} \log \left(\int \left(\frac{dP_1}{dQ_1}(x_1) \right)^{\alpha} dQ_1(x_1) \right) + \frac{1}{\alpha - 1} \log \left(\int \left(\frac{dP_2}{dQ_2}(x_2) \right)^{\alpha} dQ_2(x_2) \right)$$

5. Recognizing each term as the Rényi divergence of the respective distributions:

$$= R_\alpha(P_1 \parallel Q_1) + R_\alpha(P_2 \parallel Q_2)$$

Thus, we have proved that:

$$R_\alpha(P_1 \times P_2 \parallel Q_1 \times Q_2) = R_\alpha(P_1 \parallel Q_1) + R_\alpha(P_2 \parallel Q_2)$$

Part 2

Statement:
For $1 < \alpha < \alpha' < \infty$ and distributions P, Q, S with $R_\infty(S \parallel Q) < \infty$, the inequality is:

$$R_\alpha(P \parallel Q) \leq \frac{\alpha'}{\alpha' - 1} R_\beta(P \parallel S) + R_{\alpha'}(S \parallel Q)$$

where $\beta = \frac{\alpha(\alpha'-1)}{\alpha'-\alpha}$.
Proof:
Start by using Hölder's inequality for random variables X and Y with $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$:

$$\mathbb{E}[XY] \leq (\mathbb{E}[X^p])^{1/p} (\mathbb{E}[Y^q])^{1/q}$$

Apply Hölder's inequality to the integral in the definition of $R_\alpha(P \parallel Q)$:

$$\int \left(\frac{dP}{dQ} \right)^{\alpha} dQ \leq \left(\int \left(\frac{dP}{dS} \right)^{\alpha p} dS \right)^{1/p} \left(\int \left(\frac{dS}{dQ} \right)^{\alpha q} dQ \right)^{1/q}$$

Express the resulting terms in terms of Rényi divergences:

$$\exp\left((\alpha - 1) R_\alpha(P \parallel Q)\right) \leq \exp\left((\alpha p - 1)\, R_{\alpha p}(P \parallel S)\right) \exp\left(\left((\alpha - 1)q + 1\right) R_{(\alpha-1)q+1}(S \parallel Q)\right)$$

Substitute $p = \frac{\alpha'}{\alpha}$, $q = \frac{\alpha'-\alpha}{\alpha'}$, and $\alpha p = \beta$, $((\alpha - 1)q + 1) = \alpha'$ into the inequality:

$$R_\alpha(P \parallel Q) \leq \frac{\alpha'}{\alpha' - 1} R_\beta(P \parallel S) + R_{\alpha'}(S \parallel Q)$$

This completes the proof of the triangle-like inequality for Rényi divergences.

Part 3

Given:
$R_\alpha(P \parallel S) \leq \rho_1 \alpha$ for all $\alpha > 1$, and $R_\alpha(S \parallel Q) \leq \rho_2 \alpha$ for all $\alpha > 1$.
We need to prove that:

$$R_\alpha(P \parallel Q) \leq \left(\sqrt{\rho_1} + \sqrt{\rho_2}\right)^2 \alpha \quad \text{for all } \alpha > 1$$

Proof:

1. Start with the definition of Rényi divergence:

$$R_\alpha(P \parallel Q) = \frac{1}{\alpha - 1} \log \left(\int \left(\frac{dP}{dQ} \right)^\alpha dQ \right)$$

2. Use the chain rule for Rényi divergence:

$$R_\alpha(P \parallel Q) \leq R_\alpha(P \parallel S) + R_\alpha(S \parallel Q)$$

3. Substitute the bounds on the Rényi divergences:

$$R_\alpha(P \parallel Q) \leq \rho_1 \alpha + \rho_2 \alpha$$

4. Consider the quadratic form of the sum:

$$\sqrt{\rho_1} + \sqrt{\rho_2} = \sqrt{(\sqrt{\rho_1} + \sqrt{\rho_2})^2} = \sqrt{\rho_1 + \rho_2 + 2\sqrt{\rho_1 \rho_2}}$$

5. Hence:

$$R_\alpha(P \parallel Q) \leq \left(\sqrt{\rho_1} + \sqrt{\rho_2}\right)^2 \alpha$$

Thus, the triangle inequality for zCDP is established:

$$R_\alpha(P \parallel Q) \leq \left(\sqrt{\rho_1} + \sqrt{\rho_2}\right)^2 \alpha \quad \text{for all } \alpha > 1$$

Problem 8: Concentration of Measure: Simulation

Part 1

We want to compute the upper bound $t_n(\nu)$ implied by Chebyshev's inequality such that:

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}[X] \right| > t_n(\nu) \right) \leq \nu$$

For a Bernoulli random variable X with $p = 0.7$, the variance is:

$$\text{Var}(X) = p(1-p) = 0.7 \times 0.3 = 0.21$$

Chebyshev's inequality states:

$$\mathbb{P}\left(|Y - \mu| > k\sigma\right) \leq \frac{1}{k^2}$$

where $Y = \frac{1}{n} \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[X] = 0.7$, and $\sigma^2 = \frac{\text{Var}(X)}{n} = \frac{0.21}{n}$.
Setting $t_n(\nu)$ using Chebyshev's inequality:

$$t_n(\nu) = \sqrt{\frac{0.21}{n\nu}}$$

Part 2

We want to compute the upper bound $t_n(\nu)$ implied by Hoeffding's inequality.
Hoeffding's inequality states:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| > t_n(\nu)\right) \leq 2\exp\left(-2nt_n(\nu)^2\right)$$

Solve for $t_n(\nu)$ by:

$$t_n(\nu) = \sqrt{\frac{\ln\left(\frac{2}{\nu}\right)}{2n}}$$

Part 3

We want to compute the upper bound $g_n(\nu)$ using the Gaussian approximation from the Central Limit Theorem.
According to CLT, for large n, the deviation $\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]$ is approximately normally distributed with mean 0 and variance $\frac{0.21}{n}$.
The Gaussian approximation gives:

$$g_n(\nu) = C_\nu\sqrt{\frac{0.21}{n}}$$

where C_ν is the $\frac{\nu}{2}$-quantile of the standard normal distribution.

Part 4

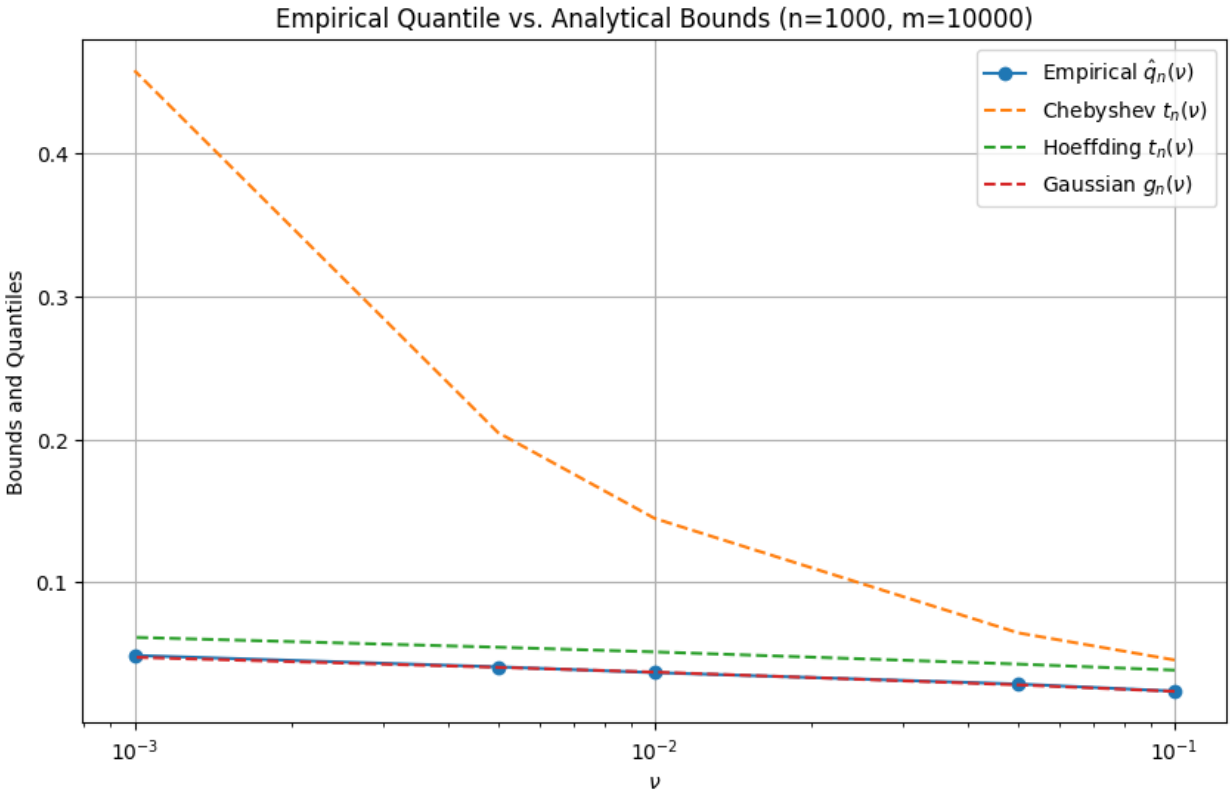

Figure 3: Plot of Empirical Quantile vs. Analytical Bounds for n = 1000 and m = 10000.

Part 5

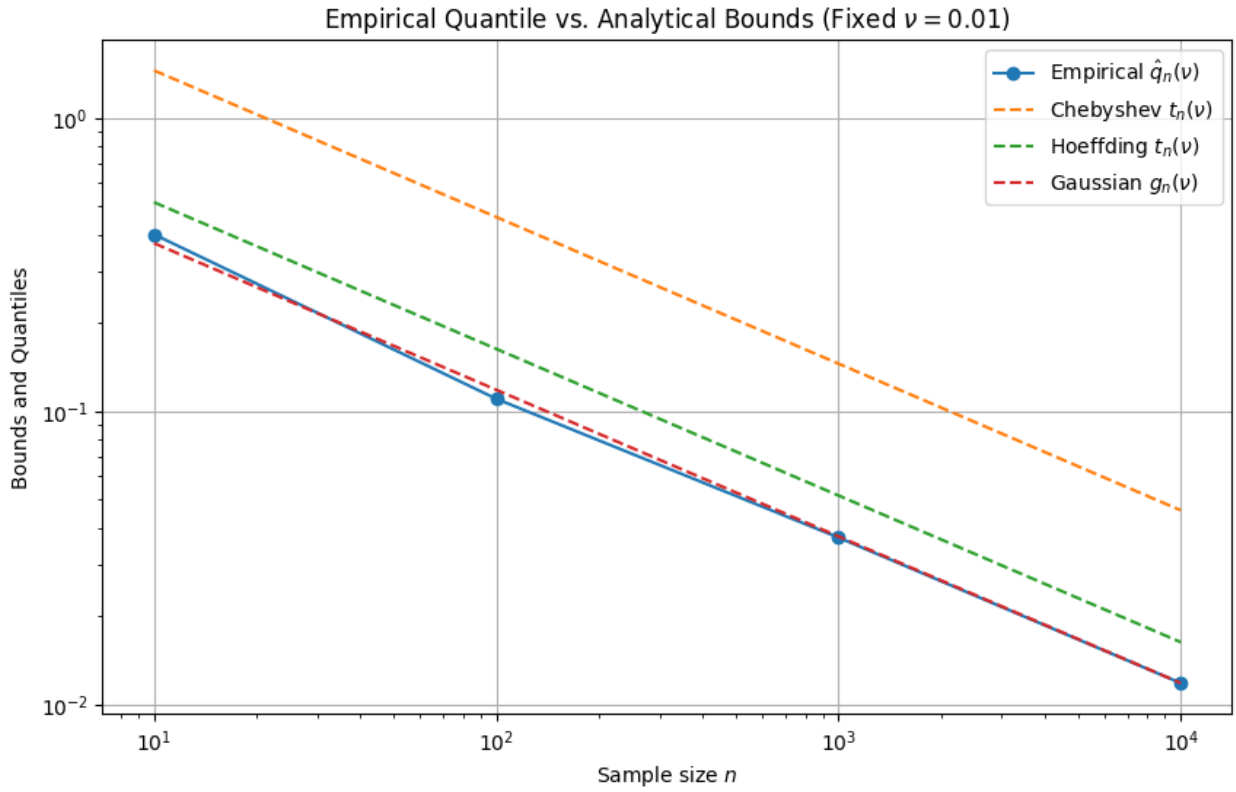

Figure 4: Plot of Empirical Quantile vs. Analytical Bounds for $\nu = 0.01$.

Problem 9: [Bonus] Chebyshev's Inequality is Tight in the Worst Case

Part 1

To show that the PDF $f_X(x)$ is well-defined, we need to verify that:

$$\int_2^\infty f_X(x)\,dx = 1.$$

Given:

$$f_X(x) = C\frac{2\log(x)+3}{x^3\log^4 x},$$

we need to check:

$$\int_2^\infty \frac{2\log(x)+3}{x^3\log^4 x}\,dx.$$

Break this into two integrals:

$$\int_2^\infty \frac{2\log(x)}{x^3\log^4 x}\,dx + \int_2^\infty \frac{3}{x^3\log^4 x}\,dx.$$

For the first integral:

$$\int_2^\infty \frac{2\log(x)}{x^3\log^4 x}\,dx = \int_2^\infty \frac{2}{x^3\log^3 x}\,dx.$$

Change variables $u = \log(x)$, $du = \frac{1}{x}dx$, and $dx = e^u du$:

$$\int_2^\infty \frac{2}{x^3\log^3 x}\,dx = \int_{\log(2)}^\infty \frac{2e^{-3u}}{u^3}\,du.$$

For the second integral:

$$\int_2^\infty \frac{3}{x^3\log^4 x}\,dx = \int_{\log(2)}^\infty \frac{3e^{-3u}}{u^4}\,du.$$

Both integrals converge, so:

$$\int_2^\infty f_X(x)\,dx = 1.$$

Thus, $f_X(x)$ is a well-defined PDF.

Part 2

To show that $E[X] < \infty$, we need:

$$E[X] = \int_2^\infty x f_X(x)\,dx.$$

Substitute $f_X(x)$:

$$E[X] = \int_2^\infty x\frac{C(2\log(x)+3)}{x^3\log^4 x}\,dx = C\int_2^\infty \frac{2\log(x)+3}{x^2\log^4 x}\,dx.$$

Break this into two integrals:

$$C\left(\int_2^\infty \frac{2\log(x)}{x^2\log^4 x}\,dx + \int_2^\infty \frac{3}{x^2\log^4 x}\,dx\right).$$

For the first integral:

$$\int_2^\infty \frac{2\log(x)}{x^2\log^4 x}\,dx = \int_2^\infty \frac{2}{x^2\log^3 x}\,dx.$$

Change variables $u = \log(x)$:

$$\int_2^\infty \frac{2}{x^2\log^3 x}\,dx = \int_{\log(2)}^\infty \frac{2e^{-2u}}{u^3}\,du.$$

For the second integral:

$$\int_2^\infty \frac{3}{x^2 \log^4 x}\, dx = \int_{\log(2)}^\infty \frac{3e^{-2u}}{u^4}\, du.$$

Both integrals converge, so $E[X] < \infty$.

Part 3

To show that $E[X^2] < \infty$, we need:

$$E[X^2] = \int_2^\infty x^2 f_X(x)\, dx.$$

Substitute $f_X(x)$:

$$E[X^2] = \int_2^\infty x^2 \frac{C(2\log(x)+3)}{x^3 \log^4 x}\, dx = C\int_2^\infty \frac{2\log(x)+3}{x\log^4 x}\, dx.$$

Break this into two integrals:

$$C\left(\int_2^\infty \frac{2\log(x)}{x\log^4 x}\, dx + \int_2^\infty \frac{3}{x\log^4 x}\, dx\right).$$

For the first integral:

$$\int_2^\infty \frac{2\log(x)}{x\log^4 x}\, dx = \int_2^\infty \frac{2}{x\log^3 x}\, dx.$$

Change variables $u = \log(x)$:

$$\int_2^\infty \frac{2}{x\log^3 x}\, dx = \int_{\log(2)}^\infty \frac{2e^{-u}}{u^3}\, du.$$

For the second integral:

$$\int_2^\infty \frac{3}{x\log^4 x}\, dx = \int_{\log(2)}^\infty \frac{3e^{-u}}{u^4}\, du.$$

Both integrals converge, so $E[X^2] < \infty$.

Part 4

To show $E[X^3] = \infty$, we need:

$$E[X^3] = \int_2^\infty x^3 f_X(x)\, dx.$$

Substitute $f_X(x)$:

$$E[X^3] = \int_2^\infty x^3 \frac{C(2\log(x)+3)}{x^3 \log^4 x}\, dx = C\int_2^\infty \frac{2\log(x)+3}{\log^4 x}\, dx.$$

Break this into two integrals:

$$C\left(\int_2^\infty \frac{2\log(x)}{\log^4 x}\,dx + \int_2^\infty \frac{3}{\log^4 x}\,dx\right).$$

For the first integral:

$$\int_2^\infty \frac{2\log(x)}{\log^4 x}\,dx = \int_{\log(2)}^\infty \frac{2u}{u^4}\,du = \int_{\log(2)}^\infty \frac{2}{u^3}\,du,$$

which diverges.
Thus, $E[X^3] = \infty$.

Part 5

To show:

$$P(X > t) \geq c\frac{1}{t^2\log^3(t)} \text{ for all } t > 4,$$

consider:

$$P(X > t) = \int_t^\infty f_X(x)\,dx = C\int_t^\infty \frac{2\log(x)+3}{x^3\log^4 x}\,dx.$$

Break it into two integrals:

$$C\left(\int_t^\infty \frac{2\log(x)}{x^3\log^4 x}\,dx + \int_t^\infty \frac{3}{x^3\log^4 x}\,dx\right).$$

For large t, the dominant term is:

$$\int_t^\infty \frac{2\log(x)}{x^3\log^4 x}\,dx \approx \frac{2}{t^2\log^3 t},$$

since:

$$\int_t^\infty \frac{\log(x)}{x^3\log^4 x}\,dx \approx \frac{1}{t^2\log^3 t}.$$

Thus:

$$P(X > t) \geq c\frac{1}{t^2\log^3 t}.$$