

Mathematical notes for “When Should We Use Top Coding in Locally Private Estimation?”

Hajime Ono^{a,1}, Kazuhiro Minami², Hideitsu Hino^{2,3}

¹Security Fundamentals Laboratory, Cybersecurity Research Institute, National Institute of Information and Communications Technology, 4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan

²The Institute of Statistical Mathematics, 10-3 Midoricho, Tachikawa, Tokyo 190-0014, Japan

³RIKEN AIP, 1-4-1 Nihonbashi, Tokyo, 103-0027, Japan

Received: date / Accepted: date

Appendix A: Detailed description of Fig. 2

Here, we follow the notation defined in Section 4.

Let $T(P) = \mathbb{E}_{X \sim P}[X]$ and $w \circ \rho = (\cdot - \cdot)^2$. We use *bit flip* [13], a generalization of the randomized response for continual domains, as a randomization mechanism Q . Treating $\perp$ as τ and letting $\mathcal{Z} = \{e^\epsilon \tau / (e^\epsilon - 1), -\tau(e^\epsilon - 1)\}$, for $y \in [0, \tau]$,

$$Q(Z_i = z | Y_i = y) = \begin{cases} \frac{e^\epsilon}{e^\epsilon + 1} \frac{y}{\tau} + \frac{1}{e^\epsilon + 1} \frac{\tau - y}{\tau} & \text{if } z = \frac{e^\epsilon}{e^\epsilon - 1} \tau, \\ \frac{1}{e^\epsilon + 1} \frac{y}{\tau} + \frac{e^\epsilon}{e^\epsilon + 1} \frac{\tau - y}{\tau} & \text{if } z = -\frac{1}{e^\epsilon - 1} \tau, \\ 0 & \text{otherwise.} \end{cases}$$

This Z_i is an unbiased estimator for Y_i , that is, $\mathbb{E}[Z_i | Y_i = y] = y$. Using these Z_i s, let $\hat{T}(Z_{1:n})$ be $\sum_{i=1}^n Z_i / n$.

For the $Z_{1:n}$, T , $\hat{T}$ and $w \circ \rho$, the non-maximum risk is decomposed to the bias and variance terms as follows:

$$\begin{aligned} & \mathbb{E}[w \circ \rho(\hat{T}(Z_{1:n}), T(P))] \\ &= \mathbb{E}[(\hat{T}(Z_{1:n}) - T(P))^2] \\ &= \mathbb{E}\left[\left(\hat{T}(Z_{1:n}) - \mathbb{E}[\hat{T}(Z_{1:n})]\right)^2\right] \\ & \quad + \left(\mathbb{E}[\hat{T}(Z_{1:n})] - T(P)\right)^2, \end{aligned}$$

where P is fixed, unlike the maximum and minimax risks. The first term of the last line is the variance, and the second term is the squared bias.

We analyze the bias and variance terms for the T , $\hat{T}$, and Z_i s. The variance term is evaluated as follows:

$$\begin{aligned} & \mathbb{E}\left[\left(\hat{T}(Z_{1:n}) - \mathbb{E}[\hat{T}(Z_{1:n})]\right)^2\right] \\ &= \mathbb{E}[\hat{T}(Z_{1:n})^2] - \mathbb{E}[\hat{T}(Z_{1:n})]^2 \\ &= \frac{1}{n} \frac{e^\epsilon}{(e^\epsilon - 1)^2} \tau^2 + \frac{1}{n} \mathbb{E}[Y](\tau - \mathbb{E}[Y]). \end{aligned}$$

The bias is

$$\begin{aligned} & \left| \mathbb{E}[\hat{T}(Z_{1:n})] - T(P) \right| \\ &= \left| \frac{1}{n} \sum_{i=1}^n (\mathbb{E}[Y_i] - \mathbb{E}[X_i]) \right| \\ &= |\Pr(X_1 \geq \tau)\tau - \mathbb{E}[X_i | X_i \geq \tau]| \\ &= \mathbb{E}[X_i - \tau | X_i \geq \tau]. \end{aligned}$$

Figure 2 is the visualization of the non-maximum risk and squared-bias and variance terms varying threshold τ , where the non-maximum risk is replaced by the simple word “error”. In the figure, P is fixed to the exponential distribution with inverse scale parameter $\lambda = 1$, and $n = 1000$. The squared-bias term is monotonically decreasing, and the variance term is monotonically increasing. Furthermore, the error is not constant and is minimum around $\tau = 1.5$.

Appendix B: Proof of Theorem 1

We prove the Theorem 1 with the following two steps:

1. Lower bound the minimax risk with the error probability of a hypothesis test.

^ae-mail: hono@nict.go.jp

- Further lower bound the error probability with an information-theoretic distance between two distributions.

First, we derive a lower bound of the risk, which is characterized by the probability that $\rho(\hat{T}(X_{1:n}), T(P)) \geq \delta$. For any weighting function w and any $\delta > 0$ such that $w(\delta) > 0$, we obtain

$$\begin{aligned} & \mathbb{E}_{X_{1:n} \sim P^n} \left[w \circ \rho(\hat{T}(X_{1:n}), T(P)) \right] \\ & \geq w(\delta) P(\rho(\hat{T}(X_{1:n}), T(P)) \geq \delta). \end{aligned}$$

We then lower bound $P(\rho(\hat{T}(X_{1:n}), T(P)) \geq \delta)$ using the error probability of a finite hypothesis test. Obviously, we have

$$\begin{aligned} & \inf_{\hat{T}} \sup_{P \in \mathcal{P}} P(\rho(\hat{T}(X_{1:n}), T(P)) \geq \delta) \\ & \geq \inf_{\hat{T}} \max_{P \in \{P_0, P_1\}} P(\rho(\hat{T}(X_{1:n}), T(P)) \geq \delta) \end{aligned} \quad (\text{B.1})$$

for any subset $\{P_0, P_1\}$ of $\mathcal{P}$. Here, we refer to P_0 and P_1 as *hypotheses*, and we refer to any $\mathcal{A}$ -measurable function $\psi : \mathcal{X}^n \rightarrow \{0, 1\}$ as a *test*. We select 2δ -separated hypotheses such that

$$\rho(T(P_0), T(P_1)) \geq 2\delta.$$

Note that δ is not required to be a universal constant. Then, for any estimator $\hat{T}$ and $j = 0, 1$, we obtain

$$P_j(\rho(\hat{T}(X_{1:n}), T(P_j)) \geq \delta) \geq P_j(\psi^*(X_{1:n}; \hat{T}) \neq j), \quad (\text{B.2})$$

where $\psi^* : \mathcal{X}^n \rightarrow \{0, 1\}$ is the *minimum distance test* defined by $\psi^*(X_{1:n}; \hat{T}) = \arg \min_{j \in \{0, 1\}} \rho(\hat{T}(X_{1:n}), T(P_j))$. In this context, (B.2) is obtained by the triangle inequality and the property of ψ^* . Combining (B.1) and (B.2), we obtain the following:

$$\begin{aligned} & \inf_{\hat{T}} \sup_{P \in \mathcal{P}} P(\rho(\hat{T}(X_{1:n}), T(P)) \geq \delta) \\ & \geq \inf_{\hat{T}} \max_{P \in \{P_0, P_1\}} P(\rho(\hat{T}(X_{1:n}), T(P)) \geq \delta) \geq p_e, \end{aligned}$$

where p_e is the error probability of the best test:

$$p_e \equiv \inf_{\psi} \max_{j \in \{0, 1\}} P_j(\psi(X_{1:n}) \neq j).$$

The infimum is taken over all tests $\{\psi : \mathcal{X}^n \rightarrow \{0, 1\}\}$.

The following lemma is introduced to lower bound the error probability.

Lemma 1 (Theorem 2.2 of [16]) *Let P_0 and P_1 be two probability measures on $(\mathcal{X}, \mathcal{A})$. If there exists a real value α such that:*

$$D_{\text{TV}}(P_0^n, P_1^n) \equiv \sup_{S \in \mathcal{A}^n} |P_0^n(S) - P_1^n(S)| \leq \alpha < 1,$$

then $p_e \geq (1 - \alpha)/2$.

Lemma 1 implies that we obtain a lower bound for the error probability if we have an upper bound for the total variation between the two hypotheses.

Combining inequalities (B.1), (B.2) and Lemma 1, we obtain Theorem 1.

Appendix C: Derivation of (14)

The basic proof idea is identical to that of Le Cam's method. We reduce the estimation problem to a binary hypothesis test and evaluate the error probability of the test. However, unlike in the standard estimation problems, considered in Theorem 1, an estimator $\hat{T}$ is a function of top-coded and randomized records $Z_{1:n}$. Due to this difference, we must modify the proof.

Here, we consider a testing problem described as follows: First, $j = 0$ or $j = 1$ is selected uniformly at random. Then, $X_{1:n}$ are sampled from P_j . To satisfy ϵ -LDP, $X_{1:n}$ are top-coded and randomized, which are denoted by $Z_{1:n}$. Finally, a test decides $j = 0$ or $j = 1$ using $Z_{1:n}$. Figure 11 describes the difference between the tests used in here and Appendix B, respectively.

We modify the inequalities that appeared in Appendix B for the modified test. First, we consider the counterpart of (B.1).

$$\begin{aligned} & \inf_{\hat{T}} \sup_{P \in \mathcal{P}} \Pr(\rho(\hat{T}(Z_{1:n}), T(P_\theta)) \geq \delta | X \sim P) \\ & \geq \inf_{\hat{T}} \max_{j \in \{0, 1\}} P_{j,Z}(\rho(\hat{T}(Z_{1:n}), T(P_j)) \geq \delta). \end{aligned}$$

This inequality is derived using the basic property of the max operator and the relationship $\{P_0, P_1\} \subset \mathcal{P}$, independent of top coding and LDP.

Next, we consider the counterpart of (B.2). For distribution pair (P_0, P_1) such that $\rho(T(P_0), T(P_1)) \geq 2\delta$, we have

$$P_{j,Z}(\rho(\hat{T}(Z_{1:n}), T(P_j)) \geq \delta) \geq P_{j,Z}(\psi^*(Z_{1:n}; \hat{T}) \neq j), \quad (\text{C.3})$$

where $\psi^* : \mathcal{Z}^n \rightarrow \{0, 1\}$ is the minimum distance test defined by $\psi^*(Z_{1:n}; \hat{T}) = \arg \min_{j \in \{0, 1\}} \rho(\hat{T}(Z_{1:n}), T(P_j))$. This inequality is derived by using the property of the minimum distance test, which is independent of top coding and LDP.

As Lemma 1, the error probability of the test function ψ^* is lower bounded using the total variation between $P_{0,Z}^n$ and $P_{1,Z}^n$ as follows:

$$\begin{aligned} \max_{j \in \{0, 1\}} P_{j,Z}(\psi^*(Z_{1:n})) & \geq \inf_{\psi} \max_{j \in \{0, 1\}} P_{j,Z}(\psi(Z_{1:n}) \neq j) \\ & \geq \frac{1}{2}(1 - D_{\text{TV}}(P_{0,Z}^n, P_{1,Z}^n)). \end{aligned}$$

Combining inequalities that appeared in this section, we obtain (14).

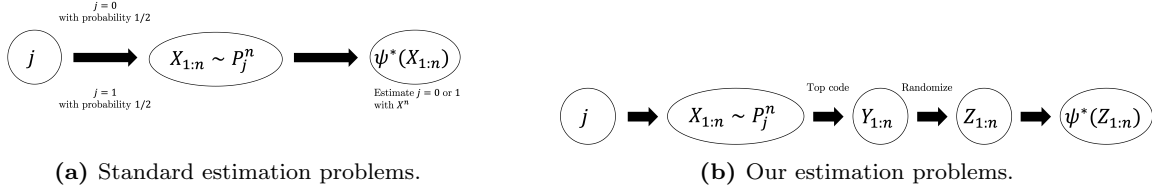


Fig. 11: Binary hypothesis tests used in the minimax risk analysis of the standard and our estimations. Test ψ^* takes raw records $X_{1:n}$ in (a); otherwise, it takes top-coded and randomized records $Z_{1:n}$ in (b).

Appendix D: Derivation of (21)

The KL divergence between $P_{\lambda,Y}$ and $P_{\lambda',Y}$ is written as follows:

$$D_{\text{KL}}(P_{\lambda,Y} \| P_{\lambda',Y}) = \int_0^\tau p_\lambda(x) \log \frac{p_\lambda(x)}{p_{\lambda'}(x)} dx + P_\lambda(X > \tau) \log \frac{P_\lambda(X > \tau)}{P_{\lambda'}(X > \tau)}. \quad (\text{D.4})$$

We analyze each term of (D.4). The first term is

$$\begin{aligned} & \int_0^\tau p_\lambda(x) \log \frac{p_\lambda(x)}{p_{\lambda'}(x)} dx \\ &= \int_0^\tau \lambda e^{-\lambda x} \log \frac{\lambda e^{-\lambda x}}{\lambda' e^{-\lambda' x}} dx \\ &= \log \frac{\lambda}{\lambda'} \int_0^\tau \lambda e^{-\lambda x} dx \\ & \quad + (-\lambda) \int_0^\tau x \lambda e^{-\lambda x} dx - (-\lambda') \int_0^\tau x \lambda e^{-\lambda x} dx \\ &= (1 - e^{-\lambda \tau}) \log \frac{\lambda}{\lambda'} \\ & \quad + (-\lambda) \frac{1 - e^{-\lambda \tau}(\lambda \tau + 1)}{\lambda} + \lambda' \frac{1 - e^{-\lambda \tau}(\lambda \tau + 1)}{\lambda} \\ &= (1 - e^{-\lambda \tau}) \log \frac{\lambda}{\lambda'} \\ & \quad + (\lambda \tau e^{-\lambda \tau} - 1 + e^{-\lambda \tau}) + \left(-\lambda' \tau e^{-\lambda \tau} + \frac{\lambda'}{\lambda} (1 - e^{-\lambda \tau}) \right). \end{aligned}$$

Since $P_\lambda(X > \tau) = 1 - P_\lambda(X \leq \tau) = 1 - (1 - e^{-\lambda \tau}) = e^{-\lambda \tau}$, the second term is

$$P_\lambda(X > \tau) \log \frac{P_\lambda(X > \tau)}{P_{\lambda'}(X > \tau)} = e^{-\lambda \tau} (-\lambda \tau + \lambda' \tau).$$

Thus,

$$D_{\text{KL}}(P_{\lambda,Y} \| P_{\lambda',Y}) = \left(\log \frac{\lambda}{\lambda'} + \frac{\lambda'}{\lambda} - 1 \right) (1 - e^{-\lambda \tau}).$$

Appendix E: Proof of Theorem 7

Proof Let $\mathcal{Z} = \{-1, +1\}$ and

$$Q(Z = z | Y = y) = \begin{cases} \frac{e^\epsilon}{e^\epsilon + 1} \frac{y}{\tau} + \frac{1}{e^\epsilon + 1} \frac{\tau - y}{\tau} & \text{if } z = +1, y \neq \infty, \\ \frac{1}{e^\epsilon + 1} \frac{y}{\tau} + \frac{e^\epsilon}{e^\epsilon + 1} \frac{\tau - y}{\tau} & \text{if } z = -1, y \neq \infty, \\ \frac{e^\epsilon}{e^\epsilon + 1} & \text{if } z = +1, y = \infty, \\ \frac{1}{e^\epsilon + 1} & \text{if } z = -1, y = \infty. \end{cases}$$

We use the MLE defined as follows:

$$T(Z_{1:n}) = \hat{\lambda} = \max_{\lambda > 0} \log p_{\lambda,Z}(Z_{1:n}).$$

Then, the likelihood is written as follows:

$$\begin{aligned} P_{\lambda,Z}(Z = +1) &= \frac{e^\epsilon - 1}{\tau(e^\epsilon + 1)} \int_0^\lambda x p_\lambda(x) dx \\ & \quad + \frac{1}{e^\epsilon + 1} P_\lambda(X \leq \tau) + \frac{e^\epsilon}{e^\epsilon + 1} P_\lambda(X > \tau) \\ &= \frac{e^\epsilon - 1}{\tau(e^\epsilon + 1)} \frac{1 - e^{-\lambda \tau}(\lambda \tau + 1)}{\lambda} \\ & \quad + \frac{1}{e^\epsilon + 1} (1 - e^{-\lambda \tau}) + \frac{e^\epsilon}{e^\epsilon + 1} e^{-\lambda \tau} \\ &= \frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{1 - e^{-\lambda \tau}}{\lambda \tau} + \frac{1}{e^\epsilon + 1} \end{aligned}$$

and

$$\begin{aligned} P_{\lambda,Z}(Z = -1) &= -\frac{e^\epsilon - 1}{\tau(e^\epsilon + 1)} \int_0^\lambda x p_\lambda(x) dx \\ & \quad + \frac{e^\epsilon}{e^\epsilon + 1} P_\lambda(X \leq \tau) + \frac{1}{e^\epsilon + 1} P_\lambda(X > \tau) \\ &= -\frac{e^\epsilon - 1}{\tau(e^\epsilon + 1)} \frac{1 - e^{-\lambda \tau}(\lambda \tau + 1)}{\lambda} \\ & \quad + \frac{e^\epsilon}{e^\epsilon + 1} (1 - e^{-\lambda \tau}) + \frac{1}{e^\epsilon + 1} e^{-\lambda \tau} \\ &= -\frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{1 - e^{-\lambda \tau}}{\lambda \tau} + \frac{e^\epsilon}{e^\epsilon + 1}. \end{aligned}$$

We note $0 < (1 - e^{-\lambda \tau})/\lambda \tau < 1$ for any $\lambda \tau > 0$.

We next analyze the mapping $\lambda \mapsto P_{\lambda,Z}$.

$$\begin{aligned} \frac{d}{d\lambda} P_{\lambda,Z}(Z = +1) &= \frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{(\lambda\tau)\tau e^{-\lambda\tau} - \tau(1 - e^{-\lambda\tau})}{(\lambda\tau)^2} \\ &= \frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{(\lambda\tau e^{-\lambda\tau} - 1 + e^{-\lambda\tau})}{\lambda^2\tau} \\ &= \frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{(\lambda\tau + 1 - e^{\lambda\tau})}{\lambda^2\tau e^{\lambda\tau}} < 0 \end{aligned}$$

for $\lambda > 0$. The last inequality holds by (31). The probability is monotone decreasing with respect to λ , and this monotonicity implies that $\lambda \neq \lambda' \implies P_{\lambda,Z} \neq P_{\lambda',Z}$. Since $P_{\lambda,Z}(Z = -1) = 1 - P_{\lambda,Z}(Z = +1)$,

$$\frac{d}{d\lambda} P_{\lambda,Z}(Z = +1) = -\frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{(\lambda\tau + 1 - e^{\lambda\tau})}{\lambda^2\tau e^{\lambda\tau}} > 0$$

for all $\lambda > 0$. The Fisher information is

$$\begin{aligned} &P_{\lambda,Z}((\log P_{\lambda,Z})^2) \\ &= \left(\frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{\lambda\tau + 1 - e^{\lambda\tau}}{\lambda^2\tau e^{\lambda\tau}} \right)^2 \left(\frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \frac{1}{e^\epsilon + 1} \right)^{-1} \\ &+ \left(-\frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{\lambda\tau + 1 - e^{\lambda\tau}}{\lambda^2\tau e^{\lambda\tau}} \right)^2 \left(-\frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \frac{e^\epsilon}{e^\epsilon + 1} \right)^{-1}. \end{aligned}$$

For $\epsilon \downarrow 0$,

$$\frac{e^\epsilon - 1}{e^\epsilon + 1} \approx \frac{\epsilon}{2} \rightarrow 0 \quad \text{and} \quad \frac{1}{e^\epsilon + 1} \approx \frac{e^\epsilon}{e^\epsilon + 1} \approx \frac{1}{2}.$$

For any $\lambda\tau > 0$,

$$0 < \frac{1 - e^{-\lambda\tau}}{\lambda\tau} < 1.$$

Thus, we have

$$\frac{e^{\lambda\tau} - (1 + \lambda\tau)}{\lambda^2\tau e^{\lambda\tau}} \approx \frac{(\lambda\tau)^2/2}{\lambda^2\tau} = \frac{\tau}{2}$$

for $\lambda\tau \ll 1$. We also have

$$\begin{aligned} &\left(\frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \frac{1}{e^\epsilon + 1} \right)^{-1} \\ &+ \left(-\frac{e^\epsilon - 1}{e^\epsilon + 1} \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \frac{e^\epsilon}{e^\epsilon + 1} \right)^{-1} \approx 4. \end{aligned}$$

The Fisher information is

$$P_{\lambda,Z} \left(\left(\frac{\partial}{\partial \lambda} \log P_{\lambda,Z} \right)^2 \right) \approx \left(\frac{\epsilon}{2} \frac{\tau}{2} \right)^2 4 = \frac{\epsilon^2 \tau^2}{4}.$$

Thus, the distribution of the MLE weakly converges to the normal distribution with variance $\frac{4}{n\epsilon^2\tau^2}$. $\square$