Data Privacy for a ρ-Recoverable Function

Ajaykrishnan Nageswaran and Prakash Narayan, *Fellow, IEEE*

***Abstract*— A user's data is represented by a finite-valued random variable. Given a function of the data, a querier is required to recover, with at least a prescribed probability, the value of the function based on a query response provided by the user. The user devises the query response, subject to the recoverability requirement, so as to maximize privacy of the data from the querier. Privacy is measured by the probability of error incurred by the querier in estimating the data from the query response. We analyze single and multiple independent query responses, with each response satisfying the recoverability requirement, which provide maximum privacy to the user. In the former setting, we also consider privacy for a predicate of the user's data. Achievability schemes with explicit randomization mechanisms for query responses are given and their privacy compared with converse upper bounds.**

***Index Terms*— Chernoff radius, function computation, predicate privacy, privacy, recoverability.**

I. Introduction

CONSIDER a (legitimate) user's data that is represented by a finite-valued random variable (rv) with known probability mass function (pmf). A querier wishes to compute a given function of the data from a user-provided query response which is a suitably randomized version of the data. The user devises the query response so as to enable the querier to recover from it the function value with a prescribed accuracy while maximizing privacy of the data, i.e., minimizing the likelihood of the querier learning the data value from it. A generalization entails the user devising multiple independent such query responses with each query response adhering to the prescribed recoverability requirement, while maximizing overall privacy.

We consider a new and rudimentary formulation of this setting in which the user forms a query response from which the querier can recover the function value with probability at least ρ, $0 \leq \rho \leq 1$. Under this requirement, the chosen query response must afford maximum data privacy. Specifically, the query response must inflict – on the querier's best estimate from it of the data value – a maximum probability of error. Beginning with a single query response, we give an explicit characterization of a randomization mechanism that enables ρ-recoverability of the function value and yields the corresponding maximum privacy, termed ρ-privacy. In particular, our query-response scheme is tantamount to an "add-noise" mechanism with the user computing first the function value and then adding to it a suitable value-dependent noise. Our optimal single query response depends on the pmf of the data rv only in a limited way through associated minentropies. Furthermore, when privacy is sought for a predicate of the user data, we obtain a characterization of *predicate ρ-privacy* and an explicit randomization mechanism that attains it. Next, when the querier elicits $n \geq 1$ ρ-recoverable and independent query responses, privacy of user data can degrade while accuracy of function estimation by the querier improves. We provide a converse upper bound for maximum privacy with respect to such responses, i.e., ρ-privacy, for every n. When $0.5 < \rho \leq 1$, this upper bound decays exponentially in n to a limit which is the querier's data-estimation error on the basis of a knowledge of the exact function value (i.e., corresponding to $\rho = 1$). The rate of this decay is shown to be (the Kullback-Leibler divergence) $D\left(\text{Ber}(0.5)||\text{Ber}(\rho)\right)$. We provide an explicit add-noise achievability scheme with privacy that converges to the mentioned limit at the same exponential rate. When $0 \leq \rho \leq 0.5$, we again provide an explicit add-noise achievability scheme. While it remains unknown whether the corresponding privacy is optimal, this scheme is shown to prevent the querier from estimating exactly the function value for *any* n. Neither achievability scheme depends on a knowledge of the pmf of the data rv. Finally, these two achievability schemes are shown to be asymptotically superior in privacy to a scheme made up of (conditionally) independent and identically distributed (i.i.d.) repetitions of our optimal single query response; this is done by means of suitable asymptotic approximations of privacy in terms of Chernoff radii. The superiority of the former schemes is enabled by rendering an estimation by the querier of the exact function value to be more error-prone than by the latter scheme, while conforming to the ρ-recoverability requirement.

An explanation of our approach is in order. In a model for private function computation, the querier can possess initial knowledge or beliefs of the user's data in the form of a family of *prior* pmfs that describe said data. Accordingly, the user must fashion a query response or responses that assure data privacy in the form of an adequate querier's probability of data-estimation error for *every* prior in said family. As indicated in Section VI, the minmax of the probability of data-estimation error (maximum and minimum, respectively, over query responses and prior pmfs) serves as a minimum guarantee of privacy for user data. In this approach, our concept of ρ-privacy developed below plays a basal role whose operational significance is clear also if the

Manuscript received February 21, 2018; revised September 26, 2018; accepted December 27, 2018. Date of publication January 22, 2019; date of current version May 20, 2019. This work was supported by the U.S. National Science Foundation under Grant CCF-1527354. This paper was presented in part at the 2018 International Symposium on Information Theory [2].

The authors are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA, and also with the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: ajayk@umd.edu; prakash@umd.edu).

Communicated by V. M. Prabhakaran, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2019.2894147

querier's uncertainty regarding the user's data were reflected by a (single) known pmf or if the user's data were known to be *generated* by said pmf. It should be added that the maximum probability of error criterion is eminently tractable – as our work shows – compared with more discerning measures, e.g., L_1- or L_2-distances between user data and the querier's estimate of it. The latter measures serve to penalize deviation of the querier's estimate of user data from its true value, a discriminating feature missing in our work (and one which is currently under study).

Our approach is in the spirit of prior works that deal with information leakage of a user's private data with associated nonprivate correlated data. A randomized version of the nonprivate data is released publicly under a constraint on the expected distortion between the nonprivate and public data. For instance, in [8], [28], and [30], leakage as measured by the mutual information between the private and public data is minimized with respect to the "channel" from the former to the latter, while constraining a distortion between the nonprivate and public data. In a more elaborate setting [32], temporally i.i.d. private and nonprivate data that are correlated across multiple users are encoded into a bin index. With this index and additional side-information as inputs, a decoder reconstructs the nonprivate data under a distortion constraint. Privacy is gauged by the conditional entropy rate of the private data given the decoder's inputs, and achievable privacy-distortion pairs are characterized. These works are based on principles of rate distortion theory. A variant model in [27] considers private and public data as the input and output of a channel with a "hard" distortion requirement between them being met with probability 1. Based on a concept of "maximal leakage" introduced in [21], privacy-recoverability tradeoffs are characterized with privacy measured in terms of α-Rényi divergence. In a separate vein, in [20] a possibly randomized function of the private and nonprivate data is released publicly while constraining the expected distortion between the nonprivate and public data. Measuring privacy in terms of a minimal expected loss function of private data and its estimate based on public data, optimal privacy mechanisms are "learned" in binary and Gaussian settings using techniques based on generative adversarial nets [18]. By comparison, for finite-valued data and query responses, upon limiting ourselves to information leakage as a probability of error and recoverability as a (pointwise) conditional probability of error, we obtain exact and approximate utility-privacy tradeoffs for single and multiple query responses, respectively. This is in contrast to a prior approach [3] in which maximum a posteriori (MAP) estimates of private and nonprivate data are formed on basis of a randomized version of the latter which is made public. The private, nonprivate and public data are assumed to form a Markov chain. Under a constraint on the probability of estimating correctly the private data, mechanisms are sought for said randomization so as to maximize correct MAP estimation of the nonprivate data.

An important movement that has received dominant attention in recent years is differential privacy, introduced in [12] and [13] and explored further in [4], [7], [25], and [29], among others. Consider a database that hosts multiple users' data that, in our framework, constitutes a data vector. The notion of differential privacy stipulates that altering a data vector slightly leads only to a near-imperceptible change in the corresponding probability distribution of the output of the privacy mechanism, i.e., query responses that are randomized functions of data vectors. We note that unlike in differential privacy, our work lacks a notion of closeness of datasets. Upon imposing a differential privacy constraint, there exists a large body of work that seeks to maximize function recoverability by minimizing a discrepancy cost between function value and randomized query response; a sampling is mentioned below. *In contrast, our work maximizes privacy under a constraint on recoverability*, and may be viewed as a companion approach. Considering a class of linear functions of data, tradeoffs between recoverability as measured by the worst-case L_2-distance (over user data) between function value and query response, and differential privacy, are examined in [19]. Similar tradeoffs for add-noise differential private mechanisms with an additional restriction are characterized in [17]. Other pertinent works include parameter estimation [33], empirical-frequency-of-data estimation [5] and distribution estimation [11], [22], [37], all under differential privacy constraints. A relaxation of the concept of differential privacy is examined in [4] in the form of distributional differential privacy as part of a larger framework of "coupled-worlds privacy." Distributional differential privacy requires the mentioned indistinguishability to hold for a random data vector over probability distributions in a specified family (to which our allusion above to the querier's initial knowledge of a family of prior pmfs for the user's data is redolent). This is in contrast to a worst-case requirement over the family of all probability distributions of the data vector in a differential privacy context.

Directions other than differential privacy also have been pursued. As mentioned in [36], these include studies based on clustering (e.g., [35]), t-closeness (e.g., [26]), data pertubation (e.g., [15]), etc; see [36] for a comprehensive list. Other methods include (ρ_1, ρ_2)-privacy (e.g., [16]), confidence intervals (e.g., [1]), and cryptographic approaches (e.g., [6]).

Our model for ρ-recoverable function computation with associated privacy is described in Section II. The ρ-privacy and predicate ρ-privacy for a single query response are characterized in Section III, and ρ-privacy is extended to multiple independent query responses in Section IV. The inadequacy of (conditionally) i.i.d. repetitions of the optimum ρ-privacy scheme of Section III in the context of Section IV is brought out in Section V. The concluding Section VI mentions unanswered questions even in our simple setting of multiple independent query responses.

II. Preliminaries

A (legitimate) user's data is represented by a rv X taking values in a finite set $\mathcal{X}$ with $|\mathcal{X}| = r$, say, and of known pmf P_X with $P_X(x) > 0$, $x \in \mathcal{X}$. Throughout, we shall consider a given mapping $f : \mathcal{X} \to \mathcal{Z} = \{0, 1, \ldots, k-1\}$, $2 \le k \le r$. For a realization $X = x$ in $\mathcal{X}$, a querier – who does not know x – wishes to compute $f(x)$ from a *query response* (QR) $F(x)$ provided by the user, where $F(x)$ is a rv with values in $\mathcal{Z}$. A QR must satisfy the following recoverability condition.

Definition 1: Given $0 \leq \rho \leq 1$, a QR $F(X)$ is *ρ-recoverable* if

$$P\big(F(X) = f(x) \big| X = x\big) \geq \rho, \quad x \in \mathcal{X}. \tag{1}$$

Condition (1) can be written equivalently in terms of a stochastic matrix $W : \mathcal{X} \to \mathcal{Z}$ with the requirement

$$W\big(f(x)|x\big) \geq \rho, \quad x \in \mathcal{X}; \tag{2}$$

which, too, will constitute a *ρ-recoverable* QR. Such a ρ-recoverable $F(X)$ or W will be termed ρ-QR. Note that ρ-recoverability in (1), (2) does not depend on P_X.

Definition 2: A ρ-QR $F(X)$ will be called an *add-noise* ρ-QR if it can be expressed as

$$F(X) = f(X) + N \text{ mod } k \tag{3}$$

where N is a $\mathcal{Z}$-valued rv that satisfies

$$N \multimap\!\!- f(X) \multimap\!\!- X \tag{4}$$

and with conditional pmf given by

$$\begin{aligned} P\big(N = i \big| X = x\big) &= P\big(N = i \big| f(X) = f(x), X = x\big) \\ &= P\big(N = i \big| f(X) = f(x)\big) \qquad (5) \\ &= V\big(i + f(x) \text{ mod } k \,\big| f(x)\big) \qquad (6) \end{aligned}$$

for some stochastic matrix $V : \mathcal{Z} \to \mathcal{Z}$ with $V(i|i) \geq \rho$, $i \in \mathcal{Z}$; we shall refer to it also as add-noise ρ-QR V. Thus, an add-noise ρ-QR is obtained by adding to the function value $f(x)$ a noise N whose (conditional) pmf can depend on $f(x)$.

By (3)-(6), an add-noise ρ-QR $F(X)$ with $V : \mathcal{Z} \to \mathcal{Z}$ has the following property:

$$\begin{aligned} &P\big(F(X) = i \big| f(X) = f(x), X = x\big) \\ &\qquad = V(i|f(x)), \quad i \in \mathcal{Z},\ x \in \mathcal{X}. \qquad (7) \end{aligned}$$

Definition 3: Denoting by Z the rv $F(X)$ with values in $\mathcal{Z}$, the *privacy* of a ρ-QR $F(X)$ (or equivalently ρ-QR W) satisfying (1) $\big($respectively (2)$\big)$ is

$$\pi_\rho(F) = \pi_\rho(W) = \min_g P\big(g(Z) \neq X\big) \tag{8}$$

where the minimum is over all estimators $g : \mathcal{Z} \to \mathcal{X}$ of X on the basis of $F(X)$. Clearly, the minimum in (8) is attained by the MAP estimator $g_{MAP} = g_{MAP(W)} : \mathcal{Z} \to \mathcal{X}$ given by

$$g_{MAP(W)}(i) = \arg\max_{x \in \mathcal{X}} P_X(x) W(i|x), \quad i \in \mathcal{Z} \tag{9}$$

so that (8) equals $P\big(g_{MAP(W)}(Z) \neq X\big)$. When $F(X)$ is an add-noise ρ-QR V as in Definition 2, we shall denote $\pi_\rho(F)$ in (8) by $\pi_\rho(V)$. The corresponding minimum in (8) will be denoted by $P\big(g_{MAP(V)}(Z) \neq X\big)$ where

$$g_{MAP(V)}(i) = \arg\max_{x \in \mathcal{X}} P_X(x) V\big(i|f(x)\big), \quad i \in \mathcal{Z}. \tag{10}$$

Ties in (9) and (10) are broken arbitrarily.

Remark: We assume throughout that the querier knows P_X and W for computing the MAP estimate in (9).

Definition 4: For each $0 \leq \rho \leq 1$, the maximum privacy that can be attained by a ρ-QR is termed *ρ-privacy* and denoted by $\pi(\rho)$, i.e.,

$$\pi(\rho) = \max_{\substack{W : W(f(x)|x) \geq \rho \\ x \in \mathcal{X}}} \pi_\rho(W). \tag{11}$$

Remark: That the maximum in (11) exists will be seen below.

The following simple lemma shows when a ρ-QR W is also an add-noise ρ-QR, and will be helpful in our proofs of achievability of privacy by ρ-QRs.

Lemma 1: Given $0 \leq \rho \leq 1$, for a ρ-QR $W : \mathcal{X} \to \mathcal{Z}$ with identical rows for all $x \in f^{-1}(i)$, $i \in \mathcal{Z}$, there exists an add-noise ρ-QR $V = V(W) : \mathcal{Z} \to \mathcal{Z}$ with the same privacy, i.e., with $\pi_\rho(V) = \pi_\rho(W)$. Conversely, for an add-noise ρ-QR $V : \mathcal{Z} \to \mathcal{Z}$, there exists a ρ-QR $W = W(V) : \mathcal{X} \to \mathcal{Z}$ with identical rows as above, and with $\pi_\rho(W) = \pi_\rho(V)$.

Proof: For a stochastic matrix $W : \mathcal{X} \to \mathcal{Z}$ which satisfies (2) and has rows $\{(W(i'|x),\ i' \in \mathcal{Z}), x \in \mathcal{X}\}$ that are identical for all $x \in f^{-1}(i)$, $i \in \mathcal{Z}$, consider a stochastic matrix $V = V(W) : \mathcal{Z} \to \mathcal{Z}$ given by

$$V(i|j) = W(i|x) \quad \text{for every } x \in f^{-1}(j),\ i,\ j \in \mathcal{Z} \tag{12}$$

and an associated add-noise QR $F'(X)$ defined as in (3)-(6) with V as above. Since $V(i|i) \geq \rho$, $i \in \mathcal{Z}$, in (12), $F'(X)$ is an add-noise ρ-QR. To see that $\pi_\rho(V) = \pi_\rho(W)$, we have

$$\begin{aligned} &P\Big(g_{MAP(V)}\big(F'(X)\big) = X\Big) \\ &\qquad = \sum_{i \in \mathcal{Z}} \max_{x \in \mathcal{X}} P\big(X = x, F'(X) = i\big) \qquad (13) \end{aligned}$$

where in the right-side,

$$\begin{aligned} &P\big(X = x, F'(X) = i\big) \\ &= \sum_{j \in \mathcal{Z}} P\big(X = x, f(X) = j, F'(X) = i\big) \\ &= P\big(X = x, f(X) = f(x)\big) P_{F'(X)|f(X)}(i|f(x)), \quad \text{by (7)} \\ &= P_X(x) V\big(i|f(x)\big) \\ &= P_X(x) W\big(i|x\big), \quad \text{by (12)}. \end{aligned}$$

Hence, by (13),

$$\begin{aligned} 1 - \pi_\rho(V) &= P\Big(g_{MAP(V)}\big(F'(X)\big) = X\Big) \\ &= \sum_{i \in \mathcal{Z}} \max_{x \in \mathcal{X}} P_X(x) W\big(i|x\big) \\ &= 1 - \pi_\rho(W). \end{aligned}$$

Conversely, given an add-noise ρ-QR $V : \mathcal{Z} \to \mathcal{Z}$, consider a stochastic matrix $W = W(V) : \mathcal{X} \to \mathcal{Z}$ with identical rows for all $x \in f^{-1}(i)$, $i \in \mathcal{Z}$, defined by (12). By the same steps as above, this W is a ρ-QR and, furthermore, $\pi_\rho(W) = \pi_\rho(V)$. ■

A justification of our model above is warranted. Our choice of the probability of error as a measure of recoverability as well as privacy is driven by considerations of tractability and obtaining exact answers, as indicated in Section I. In particular, it enables us to identify optimal or asymptotically optimal ρ-QRs in our achievability proofs. In symmetry with the pointwise measure of recoverability in $X = x$ in (1) or (2), it would be preferable to consider a redolent measure of privacy in (8), (11) that is pointwise in $Z = i$, viz.

$$\max_{\substack{W: W(f(x)|x) \geq \rho \\ x \in \mathcal{X}}} \max_{i \in \mathcal{Z}} P\big(g_{MAP(W)}(Z) \neq X \mid Z = i\big).$$

However, such conservatism leads to intractability; by contrast, our liberal choice in (8), (11), which is equivalent to

$$\max_{\substack{W: W(f(x)|x) \geq \rho \\ x \in \mathcal{X}}} \max_{i \in \mathcal{Z}} P\big(g_{MAP(W)}(Z) \neq X,\ Z = i\big), \quad (14)$$

makes for comprehensive analysis as will be seen below; the equivalence of (14) is explained in the remark following the proof of Theorem 2 in the next section.

Concerning ρ-recoverability, we add that there is no loss of generality in (1), (2) by not considering an *estimate* of $f(X)$ on the basis of $F(X)$; this is so, because the user can emulate any such estimation strategy of the querier to produce yet another ρ-QR.

Lastly, a seemingly more general setting comprising "private data X, correlated nonprivate data Y and publicly released data Z" is addressed below; see Remark (iii) after the proof of Proposition 3.

III. ρ-PRIVACY FOR A SINGLE QUERY RESPONSE

A characterization of ρ-privacy is provided by obtaining first an upper bound for $\pi(\rho)$ and then identifying explicitly an add-noise ρ-QR whose privacy meets the bound.

Let

$$x^* = \arg\max_{x \in \mathcal{X}} P_X(x), \quad x_i^* = \arg\max_{x \in f^{-1}(i)} P_X(x),\ i \in \mathcal{Z} \quad (15)$$

and suppose that $x^* \in f^{-1}(i^*)$ for some $i^* \in \mathcal{Z}$, where x^*, i^* and x_i^*, $i \in \mathcal{Z}$, need not be unique. Further, set

$$\rho_c = \frac{P_X(x^*)}{\sum_{i \in \mathcal{Z}} P_X(x_i^*)} \quad (16)$$

and observe that $1/k \leq \rho_c < 1$, where the left inequality is by

$$\frac{P_X(x^*)}{\sum_{i \in \mathcal{Z}} P_X(x_i^*)} \geq \frac{P_X(x^*)}{\sum_{i \in \mathcal{Z}} P_X(x^*)} = \frac{1}{k}.$$

The following choice of ρ-QR $W = W_o : \mathcal{X} \to \mathcal{Z}$ will play a material role in the achievability proof of ρ-privacy in Theorem 2 below:

$$W_o(i|x) = \begin{cases} \max\{\rho_c, \rho\}, & i = f(x) \\ \Big(1 - \max\{\rho_c, \rho\}\Big) \frac{P_X(x_i^*)}{\sum_{l \neq f(x)} P_X(x_l^*)}, & i \neq f(x), \\ & x \in \mathcal{X},\ i \in \mathcal{Z}. \end{cases} \quad (17)$$

We note that W_o has the property that for each $i \in \mathcal{Z}$, all rows of W_o corresponding to $x \in f^{-1}(i)$ are identical. By dint of Lemma 1, the associated stochastic matrix $V_o : \mathcal{Z} \to \mathcal{Z}$ given by

$$\begin{aligned} V_o(i|j) &= W_o(i|x) \ \text{ for every } x \in f^{-1}(j) \\ &= \begin{cases} \max\{\rho_c, \rho\}, & i = j \\ \Big(1 - \max\{\rho_c, \rho\}\Big) \frac{P_X(x_i^*)}{\sum_{l \neq j} P_X(x_l^*)}, & i \neq j, i, j \in \mathcal{Z} \end{cases} \end{aligned} \quad (18)$$

will be also of consequence in achieving ρ-privacy.

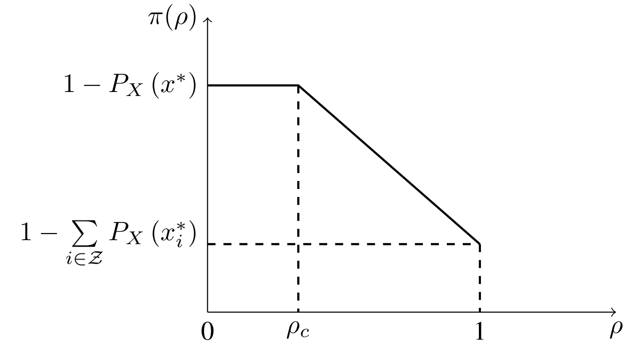

Fig. 1. $\pi(\rho)$ vs. ρ.

An exact characterization of ρ-privacy is provided by

Theorem 2: ρ-privacy equals

$$\begin{aligned} \pi(\rho) &= 1 - \max\left\{P_X(x^*), \rho \sum_{i \in \mathcal{Z}} P_X(x_i^*)\right\} \\ &= 1 - \max\{\rho_c, \rho\} \sum_{i \in \mathcal{Z}} P_X(x_i^*), \quad 0 \leq \rho \leq 1. \end{aligned} \quad (19)$$

Furthermore, ρ-privacy is achieved by the ρ-QR W_o in (17) *and, additionally, by the add-noise ρ-QR V_o in* (18).

Remarks:

(i) The choice of W_o and V_o in (17) and (18), and the value of ρ-privacy in (19), depend on P_X only through $P_X(x_i^*), i \in \mathcal{Z}$, i.e., $P_X\big(f^{-1}(i)\big) 2^{-H_{min}(P_i)}$, $i \in \mathcal{Z}$, where $H_{min}(P_i)$ is the minentropy of the pmf $P_i = \big(P_X(x)/P_X\big(f^{-1}(i)\big),\ x \in f^{-1}(i)\big)$.

(ii) By Theorem 2,

$$\pi(\rho) = \begin{cases} 1 - P_X(x^*), & 0 \leq \rho \leq \rho_c \\ 1 - \rho \sum_{i \in \mathcal{Z}} P_X(x_i^*), & \rho_c \leq \rho \leq 1 \end{cases}$$

and is plotted in Fig. 1. In particular, for $0 \leq \rho \leq \rho_c$, $\pi(\rho) = 1 - P_X(x^*)$ and is the error of a MAP estimator of X without any observation. For $\rho = 1$, $\pi(1) = 1 - \sum_{i \in \mathcal{Z}} P_X(x_i^*)$ is the error of a MAP estimator of X on the basis of $f(X)$.

(iii) The ρ-privacy achieving ρ-QR W_o and the corresponding add-noise ρ-QR V_o in Theorem 2 are not unique. For instance, see Remark (ii) after the proof of Proposition 3, Remark (ii) following Theorem 5 and the first part of the Remark following Theorem 6.

Proof: That the two characterizations of $\pi(\rho)$ in (19) are identical follows by straightforward manipulation. We first show that ρ-privacy cannot exceed the right-side(s) of (19), and then identify a ρ-QR that attains it.

Converse: Clearly

$$P\big(g_{MAP(W)}(Z) = X\big) \geq P_X(x^*)$$

and for every $W : \mathcal{X} \to \mathcal{Z}$ satisfying (2),

$$\begin{aligned} P\big(g_{MAP(W)}(Z) = X\big) &= \sum_{i \in \mathcal{Z}} \max_{x \in \mathcal{X}} P_X(x) W(i|x) \\ &\geq \sum_{i \in \mathcal{Z}} \max_{x \in f^{-1}(i)} P_X(x) W(i|x) \\ &\geq \rho \sum_{i \in \mathcal{Z}} P_X(x_i^*) \end{aligned}$$

leading to

$$P\big(g_{MAP(W)}(Z) = X\big) \geq \max\left\{P_X(x^*), \rho \sum_{i \in \mathcal{Z}} P_X(x_i^*)\right\}. \tag{20}$$

Hence

$$\begin{aligned} \pi_\rho(W) &= P\big(g_{MAP(W)}(Z) \neq X\big) \\ &\leq 1 - \max\left\{P_X(x^*), \rho \sum_{i \in \mathcal{Z}} P_X(x_i^*)\right\}, \quad 0 \leq \rho \leq 1 \end{aligned} \tag{21}$$

so that the same upper bound, valid for all $W : \mathcal{X} \to \mathcal{Z}$ subject to (2), applies to $\pi(\rho)$, too.

Achievability: We show that the choice of the ρ-QR $W_o : \mathcal{X} \to \mathcal{Z}$ in (17) has privacy $\pi_\rho(W_o)$ equal to the right-side(s) of (19). To this end,

$$\begin{aligned} 1 - \pi_\rho(W_o) &= P\big(g_{MAP(W_o)}(Z) = X\big) \\ &= \sum_{i \in \mathcal{Z}} \max_{x \in \mathcal{X}} P_X(x) W_o(i|x) \\ &= \sum_{i \in \mathcal{Z}} \max\left\{\max_{x \in f^{-1}(i)} P_X(x) W_o(i|x), \right. \\ &\qquad \left. \max_{x \notin f^{-1}(i)} P_X(x) W_o(i|x)\right\} \\ &= \sum_{i \in \mathcal{Z}} \max\left\{P_X(x_i^*) \max\{\rho_c, \rho\}, \right. \\ &\qquad \left. \max_{x \notin f^{-1}(i)} P_X(x) W_o(i|x)\right\}, \quad \text{by (17).} \end{aligned} \tag{22}$$

We claim that

$$P_X(x_i^*) \max\{\rho_c, \rho\} \geq \max_{x \notin f^{-1}(i)} P_X(x) W_o(i|x), \quad i \in \mathcal{Z} \tag{23}$$

whereupon (22) becomes

$$1 - \pi_\rho(W_o) = \max\{\rho_c, \rho\} \sum_{i \in \mathcal{Z}} P_X(x_i^*)$$

so that the privacy $\pi_\rho(W_o)$ equals the right-side(s) of (19). It remains to establish (23). Considering first the case $0 \leq \rho \leq \rho_c$, we must show for each $x \notin f^{-1}(i)$ that

$$\begin{aligned} P_X(x_i^*) \rho_c &\geq P_X(x) W_o(i|x) \\ &= P_X(x)(1 - \rho_c) \frac{P_X(x_i^*)}{\sum_{j \neq f(x)} P_X(x_j^*)}, \quad \text{by (17)} \end{aligned}$$

i.e.,

$$\frac{\rho_c}{1 - \rho_c} \geq \frac{P_X(x)}{\sum_{j \neq f(x)} P_X(x_j^*)} \tag{24}$$

which, in turn, would follow if

$$\frac{\rho_c}{1 - \rho_c} \geq \frac{P_X(x^*)}{\sum_{j \neq f(x)} P_X(x_j^*)},$$

which is tantamount to showing that

$$\frac{\sum_{j \neq i^*} P_X(x_j^*)}{\sum_{j \neq f(x)} P_X(x_j^*)} \leq 1. \tag{25}$$

Clearly, (25) holds for each $x \notin f^{-1}(i)$, as the denominator is either larger than or equal to the numerator for all $i \in \mathcal{Z}$. For the case $\rho_c \leq \rho < 1$, we must show (24) with ρ_c replaced by ρ; this follows readily since

$$\frac{\rho}{1 - \rho} \geq \frac{\rho_c}{1 - \rho_c}, \quad \rho_c \leq \rho < 1.$$

For $\rho = 1$, we have by (17) that $W_o(i|x) = \mathbb{1}\big(i = f(x)\big)$, $x \in \mathcal{X}$, $i \in \mathcal{Z}$, whereby (23) holds trivially.

Finally, that the add-noise ρ-QR V_o achieves ρ-privacy follows by Lemma 1. ■

Remark: The equivalence in (14) is justified by the proof of Theorem 2 which shows, in effect, that a common maximizer in

$$\begin{aligned} &\arg \max_{\substack{W : W(f(x)|x) \geq \rho \\ x \in \mathcal{X}}} P\big(g_{MAP(W)}(Z) \neq X\big) \\ &= \arg \max_{\substack{W : W(f(x)|x) \geq \rho \\ x \in \mathcal{X}}} \sum_{i \in \mathcal{Z}} P\big(g_{MAP(W)}(Z) \neq X,\ Z = i\big) \\ &= \arg \max_{\substack{W : W(f(x)|x) \geq \rho \\ x \in \mathcal{X}}} \max_{i \in \mathcal{Z}} P\big(g_{MAP(W)}(Z) \neq X,\ Z = i\big) \end{aligned}$$

is W_o in (17).

We close this section by considering ρ-privacy for a predicate $Y = h(X)$ of X, where $h : \mathcal{X} \to \mathcal{Y} = \{0, 1, \ldots, m-1\}$, $2 \leq m \leq r$, is a given mapping. Analogously as in Definition 3, *predicate privacy*[1] of a ρ-QR $F(X)$ or W in (1), (2) is

$$\pi'_\rho(F) = \pi'_\rho(W) = P\big(g'_{MAP(W)}(Z) \neq Y\big)$$

and *predicate ρ-privacy* is

$$\pi'(\rho) = \max_{\substack{W : W(f(x)|x) \geq \rho \\ x \in \mathcal{X}}} \pi'_\rho(W). \tag{26}$$

Clearly, $\pi'(\rho) \leq \pi(\rho)$, $0 \leq \rho \leq 1$, and when $h : \mathcal{X} \to \mathcal{Y} = \mathcal{X}$ is the identity mapping, predicate ρ-privacy in (26) specializes to ρ-privacy in (11).

Proposition 3 below provides an exact characterization of $\pi'(\rho)$. Its proof builds on that for $\pi(\rho)$ in Theorem 2.

[1]Notation used in the context of ρ-privacy will be primed throughout for predicate ρ-privacy.

The following additional notation is convenient. Define

$$j^* = \arg\max_{j\in\mathcal{Y}} P_X\left(h^{-1}(j)\right),$$
$$P_X(i,j) = P_X\left(f^{-1}(i)\cap h^{-1}(j)\right), \quad i\in\mathcal{Z},\ j\in\mathcal{Y},$$
$$j_i^* = \arg\max_{j\in\mathcal{Y}} P_X(i,j), \quad i\in\mathcal{Z}$$

and

$$\rho_c' = \frac{P_X\left(h^{-1}(j^*)\right)}{\sum_{i\in\mathcal{Z}} P_X\left(i,j_i^*\right)}, \tag{27}$$

where the maxima above need not be attained uniquely. Observe that $\max\left\{\frac{1}{m},\frac{1}{k}\right\} \le \rho_c' \le 1$. The right inequality is by

$$\rho_c' = \frac{\sum_{i\in\mathcal{Z}} P_X\left(i,j^*\right)}{\sum_{i\in\mathcal{Z}} P_X\left(i,j_i^*\right)} \le 1$$

while the left inequality follows from

$$\frac{P_X\left(h^{-1}(j^*)\right)}{\sum_{i\in\mathcal{Z}} P_X\left(i,j_i^*\right)} \ge \frac{P_X\left(h^{-1}(j^*)\right)}{\sum_{i\in\mathcal{Z}} P_X\left(f^{-1}(i)\right)} = P_X\left(h^{-1}(j^*)\right) \ge \frac{1}{m}$$

and

$$\frac{P_X\left(h^{-1}(j^*)\right)}{\sum_{i\in\mathcal{Z}} P_X\left(i,j_i^*\right)} \ge \frac{P_X\left(h^{-1}(j^*)\right)}{\sum_{i\in\mathcal{Z}} P_X\left(h^{-1}(j_i^*)\right)} \ge \frac{P_X\left(h^{-1}(j^*)\right)}{\sum_{i\in\mathcal{Z}} P_X\left(h^{-1}(j^*)\right)} = \frac{1}{k}.$$

Proposition 3: Predicate ρ-privacy equals

$$\pi'(\rho) = 1 - \max\left\{P_X\left(h^{-1}(j^*)\right), \rho\sum_{i\in\mathcal{Z}} P_X\left(i,j_i^*\right)\right\}$$
$$= 1 - \max\left\{\rho_c',\rho\right\}\sum_{i\in\mathcal{Z}} P_X\left(i,j_i^*\right), \quad 0\le\rho\le 1. \tag{28}$$

Proof: Starting with the converse part, we have that

$$P\left(g'_{MAP(W)}(Z)=Y\right) \ge P_X\left(h^{-1}(j^*)\right)$$

and for every ρ-QR $W:\mathcal{X}\to\mathcal{Z}$ in (2),

$$P\left(g'_{MAP(W)}(Z)=Y\right) = \sum_{i\in\mathcal{Z}}\max_{j\in\mathcal{Y}}\sum_{x\in h^{-1}(j)} P_X(x)W(i|x)$$
$$\ge \rho\sum_{i\in\mathcal{Z}}\max_{j\in\mathcal{Y}} P_X(i,j)$$
$$= \rho\sum_{i\in\mathcal{Z}} P_X(i,j_i^*)$$

where the first equality uses

$$P\left(Y=j\middle|X=x,Z=i\right) = \mathbb{1}\left(j=h(x)\right).$$

The converse proof is completed similarly as for Theorem 2.

Turning to the achievability part, consider the ρ-QR $W_o':\mathcal{X}\to\mathcal{Z}$ specified for $x\in\mathcal{X}$, $i\in\mathcal{Z}$, $j\in\mathcal{Y}$, as follows.

For $\rho_c'=1$

$$W_o'(i|x) = \mathbb{1}\left(f(x)=i\right) \tag{29}$$

and for $\rho_c'<1$,

$$W_o'(i|x) = \begin{cases} \max\{\rho_c',\rho\} + (1-\max\{\rho_c',\rho\})\,\frac{P_X(i,j_i^*)-P_X(i,j)}{\sum_{l\in\mathcal{Z}} P_X\left(l,j_l^*\right)-P_X\left(h^{-1}(j)\right)}, & i=f(x),\ j=h(x)\\ (1-\max\{\rho_c',\rho\})\,\frac{P_X(i,j_i^*)-P_X(i,j)}{\sum_{l\in\mathcal{Z}} P_X\left(l,j_l^*\right)-P_X\left(h^{-1}(j)\right)}, & i\ne f(x),\ j=h(x). \end{cases} \tag{30}$$

Since $\rho_c'<1$, observe in (30) that

$$\sum_{l\in\mathcal{Z}} P_X\left(l,j_l^*\right) - P_X\left(h^{-1}(j)\right) \ge \sum_{l\in\mathcal{Z}} P_X\left(l,j_l^*\right) - P_X\left(h^{-1}(j^*)\right) > 0.$$

We show in Appendix A that $\pi_\rho'(W_o')$ equals the right-side of (28).

This completes the proof of the proposition. ■

Remarks:

(i) Note that W_o' has the property that for each i in $\mathcal{Z}$, j in $\mathcal{Y}$, all rows corresponding to x in $f^{-1}(i)\cap h^{-1}(j)$ are identical. Pursuant to Lemma 1, W_o' can be interpreted as an add-noise ρ-QR obtained by adding to $f(X)$ a noise N that satisfies the Markov condition $N -\!\circ\!- f(X),h(X) -\!\circ\!- X$.

(ii) When h is the identity mapping, W_o' in (30) *does not* coincide with W_o in (17); in fact, W_o' reduces to

$$W_o''(i|x) = \begin{cases} \max\{\rho_c,\rho\} + (1-\max\{\rho_c,\rho\})\,\frac{P_X(x_i^*)-P_X(x)}{\sum_{l\in\mathcal{Z}} P_X\left(x_l^*\right)-P_X(x)}, & i=f(x),\\ (1-\max\{\rho_c,\rho\})\,\frac{P_X(x_i^*)-P_X(x)}{\sum_{l\in\mathcal{Z}} P_X\left(x_l^*\right)-P_X(x)}, & i\ne f(x). \end{cases}$$

By Proposition 3 and Theorem 2, $\pi_\rho'(W_o'') = \pi(\rho) = \pi_\rho(W_o)$. On the other hand, and unlike W_o, the ρ-QR W_o'' is not of the add-noise type in the sense of Definition 2 and also depends on the entirety of P_X.

(iii) Proposition 3 covers the setting when privacy is sought for a randomized function Y of the data X with the (finite-valued) rvs X,Y having a given joint pmf. Specifically ρ-privacy for Y corresponds to predicate ρ-privacy in Proposition 3 with $\bar{X}$, $h(\bar{X})$, $f(\bar{X})$ and $F(\bar{X})$, where

$$\bar{X}=(X,Y), \quad h(\bar{X})=Y,\ f(\bar{X})=f(X),\ F(\bar{X})=Z,$$

(with an abuse of notation in f).

(iv) In a similar vein, Proposition 3 also covers the setting with "private data X, correlated nonprivate data Y and publicly released data Z" alluded to in Section I. Formally, ρ-privacy for the mentioned setting equals predicate ρ-privacy in Proposition 3 applied to $\tilde{X}$, $h(\tilde{X})$, $f(\tilde{X})$ and $F(\tilde{X})$, where

$$\tilde{X}=(X,Y), \quad h(\tilde{X})=X,\ f(\tilde{X})=Y,\ F(\tilde{X})=Z.$$

IV. Multiple Independent Query Responses

In a general setting, given a mapping $f : \mathcal{X} \to \mathcal{Z}$, a querier wishes to compute $f(x)$, $x \in \mathcal{X}$, from ρ-QRs $\{(F_t(x), x \in \mathcal{X})\}_{t=1}^n$, $n \geq 1$. The rvs $\{F_t(X)\}_{t=1}^n$ are taken to be conditionally mutually independent, conditioned on X, but not necessarily identically distributed, with *each* $F_t(X)$ satisfying the ρ-recoverability condition (1). Correspondingly, consider stochastic matrices $\{W_t : \mathcal{X} \to \mathcal{Z}\}_{t=1}^n$ such that

$$\begin{aligned} P\big(F_1(X) = i_1, \ldots, F_n(X) = i_n \big| X = x\big) &= \prod_{t=1}^n P\big(F_t(X) = i_t | X = x\big) \\ &= \prod_{t=1}^n W_t(i_t|x), \quad x \in \mathcal{X},\ i_1, \ldots, i_n \in \mathcal{Z} \end{aligned} \tag{31}$$

say, with *each* W_t satisfying (2). Similarly, for add-noise ρ-QRs $F_t(X)$ as in Definition 2 with $\{V_t : \mathcal{Z} \to \mathcal{Z}$ where $V_t(i|i) \geq \rho$, $i \in \mathcal{Z}\}_{t=1}^n$,

$$\begin{aligned} P\big(F_1(X) = i_1, \ldots, F_n(X) = i_n \big| X = x\big) = \prod_{t=1}^n V_t\big(i_t|f(x)\big), \quad x \in \mathcal{X},\ i_1, \ldots, i_n \in \mathcal{Z}. \end{aligned} \tag{32}$$

In all contexts, denote $Z_t = F_t(X)$, $t = 1, \ldots, n$.

This formulation is apposite when the main objective of the querier is to improve its estimation accuracy (beyond ρ) of a *given* $f(X)$ by soliciting multiple ρ-QRs whereas the user designs said ρ-QRs so as to maximize the privacy of its data X. A different formulation in which an "adversarial querier" elicits multiple ρ-QRs for various choices of functions in order to destroy data privacy by isolating the value of X, is beyond the scope of this paper.

Remark: In addition to possibly eroding privacy, multiple ρ-QRs enable the querier to estimate $f(X)$ with a probability that can exceed ρ. Precisely, for a MAP estimator h_{MAP} of $f(X)$ on the basis of $\{F_t(X)\}_{t=1}^n$ in (31), we have

$$\begin{aligned} &P\big(h_{\text{MAP}}(F_1(X), \ldots, F_n(X)) = f(X)\big) \\ &\geq \max\left\{\rho, \max_{i \in \mathcal{Z}} P(f(X) = i), P\left(\text{Bin}(n, \rho) \geq \left\lfloor \frac{n}{2} \right\rfloor + 1\right)\right\} \end{aligned} \tag{33}$$

where Bin(n, ρ) is a binomial rv with parameters $n \geq 1$ and $0 \leq \rho \leq 1$. In particular, for $0.5 < \rho \leq 1$, the right-side of (33) tends to 1 as $n \to \infty$. See Appendix C and Lemma 7.

Definition 5: For each $0 \leq \rho \leq 1$ and $n \geq 1$, the ρ-privacy that can be attained by ρ-QRs $\{F_t(X)\}_{t=1}^n$ as in (31) with each $F_t(X)$ satisfying (1) $\big($or equivalently each W_t satisfying (2)$\big)$ is

$$\pi_n(\rho) = \max_{\substack{W_1, \ldots, W_n: \\ W_t(f(x)|x) \geq \rho,\ x \in \mathcal{X}}} \pi_\rho(W_1, \ldots, W_n),$$

where

$$\pi_\rho(W_1, \ldots, W_n) = \min_{g_n} P\big(g_n(Z_1, \ldots, Z_n) \neq X\big),$$

with the minimum being taken over all estimators $g_n : \mathcal{Z}^n \to \mathcal{X}$ on the basis of $\{F_t(X)\}_{t=1}^n$. Thus,

$$\pi_\rho(W_1, \ldots, W_n) = P\big(g_{\text{MAP}\,(W_1, \ldots, W_n)}(Z_1, \ldots, Z_n) \neq X\big) \tag{34}$$

where

$$\begin{aligned} &g_{MAP(W_1, \ldots, W_n)}(i_1, \ldots, i_n) \\ &\quad = \arg\max_{x \in \mathcal{X}} P_X(x) \prod_{t=1}^n W_t(i_t|x), \quad i_1, \ldots, i_n \in \mathcal{Z}. \end{aligned}$$

Similarly, for add-noise ρ-QRs $\{F_t(X)\}_{t=1}^n$ as in (32), we define

$$\pi_\rho(V_1, \ldots, V_n) = P\big(g_{\text{MAP}\,(V_1, \ldots, V_n)}(Z_1, \ldots, Z_n) \neq X\big) \tag{35}$$

with

$$\begin{aligned} &g_{MAP(V_1, \ldots, V_n)}(i_1, \ldots, i_n) \\ &\quad = \arg\max_{x \in \mathcal{X}} P_X(x) \prod_{t=1}^n V_t\big(i_t|f(x)\big), \quad i_1, \ldots, i_n \in \mathcal{Z}. \end{aligned}$$

Of particular interest will be the cases $W_t = W$ or $V_t = V$, $t = 1, \ldots, n$, when we write (34) and (35) as

$$\pi_\rho(W^n) = P\big(g_{\text{MAP}\,(W^n)}(Z_1, \ldots, Z_n) \neq X\big)$$

and

$$\pi_\rho(V^n) = P\big(g_{\text{MAP}\,(V^n)}(Z_1, \ldots, Z_n) \neq X\big).$$

We provide first in Section IV-A an upper bound for ρ-privacy $\pi_n(\rho)$ which is valid for each $0 \leq \rho \leq 1$ and every $n \geq 1$. Next, in Section IV-B, considering the realms $0.5 < \rho \leq 1$ and $0 \leq \rho \leq 0.5$ separately, we show corresponding explicit achievability schemes. However, unlike in Section III for the case $n = 1$, the lower bound for $\pi_n(\rho)$ from the achievability schemes below, that use add-noise ρ-QRs, need not coincide with the upper bound in Theorem 4 for any finite $n \geq 1$. These upper and lower bounds for $\pi_n(\rho)$ are rendered into more convenient, albeit blunter forms in Section IV-C.

A. Converse

We provide next, as a converse result, an upper bound for $\pi_n(\rho)$, $n \geq 1$. For $0 \leq \rho \leq 1$, set

$$\begin{aligned} \Gamma_n(\rho) = \min\left\{1 - \rho_c, \min\left\{1 - \rho, P\left(\text{Bin}(n, \rho) \leq \left\lfloor \frac{n}{2} \right\rfloor\right)\right\}\right\} \\ \times \sum_{i \in \mathcal{Z}} P_X(x_i^*), \quad n \geq 1 \end{aligned} \tag{36}$$

and note that $0 \leq \Gamma_n(\rho) \leq 1$.

Theorem 4: For each $0 \leq \rho \leq 1$ and for every $n \geq 1$,

$$\pi_n(\rho) \leq 1 - \sum_{i \in \mathcal{Z}} P_X(x_i^*) + \Gamma_n(\rho).$$

Remark: For $0 \leq \rho \leq 1$ and $n = 1$, since

$$\Gamma_1(\rho) = (1 - \max\{\rho_c, \rho\}) \sum_{i \in \mathcal{Z}} P_X(x_i^*),$$

we have that the upper bound for $\pi_n(\rho)$ above reduces to that for $\pi(\rho)$ in the right-side of (21).

Proof: For $W_1, \ldots, W_n$ satisfying (2),

$$\begin{aligned} P\big(g_{\mathrm{MAP}\,(W_1,\ldots,W_n)}\,(Z_1,\ldots,Z_n) = X\big) \\ \geq P\big(g_{\mathrm{MAP}\,(W_1)}\,(Z_1) = X\big) \\ \geq \max\left\{P_X\left(x^*\right), \rho \sum_{i\in\mathcal{Z}} P_X\left(x_i^*\right)\right\} \end{aligned} \tag{37}$$

by (20). Also,

$$P\big(g_{\mathrm{MAP}\,(W_1,\ldots,W_n)}\,(Z_1,\ldots,Z_n) = X\big) = \sum_{(i_1,\ldots,i_n)\in\mathcal{Z}^n} \max_{x\in\mathcal{X}} P_X\left(x\right) \prod_{t=1}^n W_t\left(i_t|x\right). \tag{38}$$

For each $i \in \mathcal{Z}$ and for $l = \left\lfloor \frac{n}{2} \right\rfloor + 1, \ldots, n$, set

$$\mathcal{A}_l(i) = \Big\{(i_1,\ldots,i_n)\in\mathcal{Z}^n : i \text{ occurs } l \text{ times in } (i_1,\ldots,i_n)\Big\}. \tag{39}$$

Then, in (38),

$$\begin{aligned} &P\big(g_{\mathrm{MAP}\,(W_1,\ldots,W_n)}\,(Z_1,\ldots,Z_n) = X\big) \\ &\geq \sum_{i\in\mathcal{Z}} \sum_{l=\lfloor\frac{n}{2}\rfloor+1}^{n} \sum_{(i_1,\ldots,i_n)\in\mathcal{A}_l(i)} \max_{x\in\mathcal{X}} P_X\left(x\right)\prod_{t=1}^n W_t\left(i_t|x\right) \\ &\geq \sum_{i\in\mathcal{Z}} P_X\left(x_i^*\right) \sum_{l=\lfloor\frac{n}{2}\rfloor+1}^{n} \sum_{(i_1,\ldots,i_n)\in\mathcal{A}_l(i)} \prod_{t=1}^n W_t\left(i_t|x_i^*\right) \\ &= \sum_{i\in\mathcal{Z}} P_X\left(x_i^*\right) s_i(n) \end{aligned} \tag{40}$$

where

$$s_i(n) = \sum_{l=\lfloor\frac{n}{2}\rfloor+1}^{n} s_i^l(n) \tag{41}$$

with

$$s_i^l(n) = \sum_{(i_1,\ldots,i_n)\in\mathcal{A}_l(i)} \prod_{t=1}^n W_t\left(i_t|x_i^*\right), \quad i \in \mathcal{Z}. \tag{42}$$

To understand the functional dependence of $s_i^l(n)$ on $\big(W_1\left(i|x_i^*\right),\ldots,W_n\left(i|x_i^*\right)\big)$, consider as an instance all $(i_1,\ldots,i_n) \in \mathcal{A}_l(i)$ with $i_1 = \ldots = i_l = i$ and $i_t \neq i$, $t = l+1,\ldots,n$. The corresponding sum for such $(i_1,\ldots,i_n) \in \mathcal{A}_l(i)$ in (42) equals

$$\begin{aligned} &\left(\prod_{t=1}^{l} W_t\left(i|x_i^*\right)\right) \sum_{\substack{(i_{l+1},\ldots,i_n)\in\mathcal{Z}^{n-l} \\ i_t\neq i,\ t=l+1,\ldots,n}} \prod_{t=l+1}^{n} W_t\left(i_t|x_i^*\right) \\ &= \left(\prod_{t=1}^{l} W_t\left(i|x_i^*\right)\right) \prod_{t=l+1}^{n} \left(1 - W_t\left(i|x_i^*\right)\right). \end{aligned}$$

In this manner, we observe that $s_i^l(n)$ reduces to a sum of $\binom{n}{l}$ terms (corresponding to the locations of l is), each of which is a product of $W_t\left(i|x_i^*\right)$-terms for l locations of t in $\{1,\ldots,n\}$ corresponding to occurrences of i, and $\big(1 - W_t\left(i|x_i^*\right)\big)$-terms in the remaining $(n-l)$ locations. Thus, $s_i^l(n)$ is a function of $\big(W_1\left(i|x_i^*\right),\ldots,W_n\left(i|x_i^*\right)\big)$.

We seek a suitable lower bound for $s_i(n)$ in terms of ρ and n, to which end we make the

Claim: For $i \in \mathcal{Z}$, $s_i(n)$ is a nondecreasing function of each $W_t\left(i|x_i^*\right)$, $t = 1,\ldots,n$.

By (41), the claim and the observation following (42), $s_i(n)$ is bounded below in an identical manner for $i = 0, 1, \ldots, k-1$, upon replacing each $W_1\left(i|x_i^*\right),\ldots,W_n\left(i|x_i^*\right)$ by ρ, in accordance with (2). By said observation, we have from (41) for $i = 0, 1, \ldots, k-1$ that

$$\begin{aligned} s_i(n) &\geq \sum_{l=\lfloor\frac{n}{2}\rfloor+1}^{n} \binom{n}{l} \rho^l \left(1-\rho\right)^{n-l} \\ &= P\left(\mathrm{Bin}(n,\rho) \geq \left\lfloor\frac{n}{2}\right\rfloor + 1\right). \end{aligned} \tag{43}$$

Then from (40),

$$\begin{aligned} &P\big(g_{\mathrm{MAP}\,(W_1,\ldots,W_n)}\,(Z_1,\ldots,Z_n) = X\big) \\ &\geq \left(\sum_{i\in\mathcal{Z}} P_X\left(x_i^*\right)\right) P\left(\mathrm{Bin}(n,\rho) \geq \left\lfloor\frac{n}{2}\right\rfloor + 1\right). \end{aligned} \tag{44}$$

Combining (37) and (44), we get

$$\begin{aligned} &P\big(g_{\mathrm{MAP}\,(W_1,\ldots,W_n)}\,(Z_1,\ldots,Z_n) = X\big) \\ &\geq \max\Bigg\{P_X\left(x^*\right), \left(\sum_{i\in\mathcal{Z}} P_X\left(x_i^*\right)\right) \\ &\qquad \times \max\left\{\rho, P\left(\mathrm{Bin}(n,\rho) \geq \left\lfloor\frac{n}{2}\right\rfloor + 1\right)\right\}\Bigg\} \\ &= \max\left\{\rho_c, \max\left\{\rho, P\left(\mathrm{Bin}(n,\rho) \geq \left\lfloor\frac{n}{2}\right\rfloor + 1\right)\right\}\right\} \\ &\qquad \times \sum_{i\in\mathcal{Z}} P_X\left(x_i^*\right) \\ &= \sum_{i\in\mathcal{Z}} P_X\left(x_i^*\right) - \Gamma_n(\rho) \end{aligned}$$

from which the assertion of the theorem follows since $W_1,\ldots,W_n$ were arbitrary subject to (2).

It remains to establish the claim, and it suffices to do so with $i = 0$, $t = 1$, i.e., we show that $s_0(n)$ is nondecreasing in $W_1\left(0|x_0^*\right)$. From the observation following (42), $s_0^l(n)$ is a sum of $\binom{n}{l}$ terms, each of which is a product of $W_t\left(0|x_0^*\right)$-terms for l locations of t in $\{1,\ldots,n\}$ where 0s occur and $\big(1 - W_t\left(0|x_0^*\right)\big)$-terms for the remaining $(n-l)$ locations. Thus, each of these $\binom{n}{l}$ terms will have either $W_1\left(0|x_0^*\right)$ or $1 - W_1\left(0|x_0^*\right)$ in it (depending on whether or not $i_1 = 0$). The latter possibility yields a term with $-W_1\left(0|x_0^*\right)$ which is seen to be canceled by a suitable term from $s_0^{l+1}(n)$. Also, $s_0^n(n) = W_1\left(0|x_0^*\right) \prod_{t=2}^{n} W_t\left(0|x_0^*\right)$. Thus, $s_0(n)$ consists of terms with $+W_1\left(0|x_0^*\right)$ or with no $W_1\left(0|x_0^*\right)$, and thereby is linear and nondecreasing in $W_1\left(0|x_0^*\right)$. This proves the claim. ∎

B. Achievability

Throughout our achievability proofs, for the sake of convenience and without loss of essential generality, we assume

$$
\left[\begin{array}{cc|cc|c|cc}
\rho & 1-\rho & 0 & 0 & \cdots & & \\
1-\rho & \rho & 0 & 0 & \cdots & & \\
\hline
0 & 0 & \rho & 1-\rho & \cdots & & \\
0 & 0 & 1-\rho & \rho & \cdots & & \\
\hline
 & & & & \ddots & & \\
\hline
 & & \cdots & & & \rho & 1-\rho \\
 & & \cdots & & & 1-\rho & \rho
\end{array}\right]
$$

(a)

$$
\left[\begin{array}{cc|cc|c|cc|c}
\rho & 1-\rho & 0 & 0 & \cdots & & & \\
1-\rho & \rho & 0 & 0 & \cdots & & & \\
\hline
0 & 0 & \rho & 1-\rho & \cdots & & & \\
0 & 0 & 1-\rho & \rho & \cdots & & & \\
\hline
 & & & & \ddots & & & \\
 & & \cdots & & & \rho & 1-\rho & 0 \\
0 & 0 & \cdots & & & 1-\rho & \rho & 0 \\
\hline
1-\rho & 0 & \cdots & & & & & \rho
\end{array}\right]
$$

(b)

Fig. 2. Add-noise ρ-QR V_1. (a) k even. (b) k odd.

that

$$P_X(x_i^*) \geq P_X(x_{i+1}^*), \quad i = 0, 1, \ldots, k-2. \tag{45}$$

1) Realm $0.5 < \rho \leq 1$: Our achievability scheme uses the following stochastic matrix $V_1 : \mathcal{Z} \to \mathcal{Z}$, not depending on P_X, given by

$$V_1(i|j) = \begin{cases} \rho, & i = j \\ 1-\rho, & j \text{ even, } i = j+1 \bmod k \text{ or } j \text{ odd, } i = j-1 \\ 0, & \text{otherwise,} \end{cases} \tag{46}$$

for $i, j \in \mathcal{Z}$. Thus, for k even, the $k \times k$-matrix V_1 is block-diagonal with exactly $k/2$ blocks of 2×2-matrices

$$\begin{bmatrix} \rho & 1-\rho \\ 1-\rho & \rho \end{bmatrix}.$$

For k odd, the upper-left $(k-1) \times (k-1)$-submatrix of V_1 is similarly structured with $(k-1)/2$ such blocks, and with the kth row being $V_1(0|k-1) = 1-\rho$ and $V_1(k-1|k-1) = \rho$. See Fig. 2. Corresponding to $V_1 : \mathcal{Z} \to \mathcal{Z}$ in (46), consider the conditionally i.i.d. ρ-QRs $\{Z_t = F_t(X)\}_{t=1}^n$ given by (32) as

$$P\big(F_1(X) = i_1, \ldots, F_n(X) = i_n \big| X = x\big) = \prod_{t=1}^n V_1(i_t|f(x)). \tag{47}$$

For $0 \leq \rho \leq 1$, set

$$\Lambda_n(\rho) = P\left(\mathrm{Bin}(n, \rho) \leq \left\lfloor \frac{n}{2} \right\rfloor\right) \left(\sum_{i \in \mathcal{Z}:\ i \text{ odd}} P_X(x_i^*)\right), \quad n \geq 1 \tag{48}$$

and note that $0 \leq \Lambda_n(\rho) \leq 1$.

Theorem 5: Let $0.5 < \rho \leq 1$. For every $n \geq 1$, the add-noise ρ-QRs $\{Z_t = F_t(X)\}_{t=1}^n$ in (47) with $V_1 : \mathcal{Z} \to \mathcal{Z}$ in (46) yield privacy

$$\pi_\rho(V_1^n) \geq 1 - \sum_{i \in \mathcal{Z}} P_X(x_i^*) + \Lambda_n(\rho). \tag{49}$$

Remarks:

(i) The choice of $V_1 : \mathcal{Z} \to \mathcal{Z}$ takes its cue from the proof of Theorem 4. The first lower bound in (40) results upon discarding those $(i_1, \ldots, i_n)$ in $\mathcal{Z}^n$ in which the most frequent symbol from $\mathcal{Z}$ occurs no more than $\lfloor \frac{n}{2} \rfloor$ times. The specific choice of V_1 in (46) ensures that the number of such occurrences is at least $\lfloor \frac{n}{2} \rfloor + 1$.

(ii) Observe that when P_X is the uniform pmf on $\mathcal{X}$, for $n = 1$, $\pi_\rho(V_1) = 1 - k\rho/r = \pi(\rho)$, the latter by (19). On the other hand, $\pi_\rho(V_1)$ can be strictly smaller than $\pi(\rho)$; for instance for $\mathcal{X} = \mathcal{Z} = \{0, 1, 2\}$, $P_X = (0.5, 0.3, 0.2)$, $f(x) = x$, and $\rho = 0.6$, it is straightforward to show that $\pi(\rho) = 0.4$ whereas $\pi_\rho(V_1) = 0.38$.

Proof: We have

$$P\big(g_{\mathrm{MAP}(V_1^n)}(Z_1, \ldots, Z_n) = X\big) = \sum_{(i_1, \ldots, i_n) \in \mathcal{Z}^n} \max_{x \in \mathcal{X}} P_X(x) \prod_{t=1}^n V_1\big(i_t | f(x)\big). \tag{50}$$

When $\rho = 1$, $V_1 : \mathcal{Z} \to \mathcal{Z}$ in (46) has 1s along its diagonal and 0s elsewhere. Hence, the right-side of (50) equals $\sum_{i \in \mathcal{Z}} P_X(x_i^*)$. Since $\Lambda_n(1) = 0$, (49) holds (with equality).

Hereafter we take $0.5 < \rho < 1$. By the form of V_1 in (46), for each $x \in \mathcal{X}$ only those $(i_1, \ldots, i_n) \in \mathcal{Z}^n$ yield nonzero contributions in (50) when consisting of $i_t = f(x)$; and $i_t = f(x) + 1 \bmod k$ for $f(x)$ even or $i_t = f(x) - 1$ for $f(x)$ odd. Accordingly, we distinguish between the cases when k is even or it is odd.

(i) *k even:* For $i = 0, 2, \ldots, k-2$, set

$$\mathcal{B}_n(i) = \big\{(i_1, \ldots, i_n) \in \mathcal{Z}^n : i_t = i \text{ or } i_t = i+1\big\}. \tag{51}$$

Then in (50),

$$P\big(g_{\mathrm{MAP}(V_1^n)}(Z_1, \ldots, Z_n) = X\big) = \sum_{i=0,2,\ldots,k-2} \sum_{(i_1, \ldots, i_n) \in \mathcal{B}_n(i)} \max_{x \in f^{-1}(i) \cup f^{-1}(i+1)} P_X(x) \prod_{t=1}^n V_1\big(i_t | f(x)\big) \tag{52}$$

where for each $i = 0, 2, \ldots, k-2$ and for each $(i_1, \ldots, i_n) \in \mathcal{B}_n(i)$,

$$\max_{x \in f^{-1}(i) \cup f^{-1}(i+1)} P_X(x) \prod_{t=1}^n V_1\big(i_t | f(x)\big) = \max\Big\{P_X(x_i^*) \rho^{l_i(i_1, \ldots, i_n)} (1-\rho)^{n - l_i(i_1, \ldots, i_n)}, P_X(x_{i+1}^*) (1-\rho)^{l_i(i_1, \ldots, i_n)} \rho^{n - l_i(i_1, \ldots, i_n)}\Big\} \tag{53}$$

with $l_i(i_1,\ldots,i_n)$ being the number of is in $(i_1,\ldots,i_n)$. The first term in $\{\cdot,\cdot\}$ above is no larger than the second if

$$l_i(i_1,\ldots,i_n) \le \tau_n(i,\rho) \triangleq \left\lfloor \frac{1}{2}\left(n - \frac{\log \frac{P(x_i^*)}{P(x_{i+1}^*)}}{\log \frac{\rho}{1-\rho}}\right)\right\rfloor.$$

Since $0.5<\rho<1$, we observe by the assumption in (45) that $\tau_n(i,\rho) \le \lfloor \frac{n}{2} \rfloor$; and $\tau_n(i,\rho) \le \frac{n}{2}-1$ for even[2] n.

Then for $i=0,2,\ldots,k-2$ and when $\tau_n(i,\rho)\ge 0$, by (53) we get in (52) that

$$\begin{aligned}
&\sum_{(i_1,\ldots,i_n)\in\mathcal{B}_n(i)} \max_{x\in f^{-1}(i)\cup f^{-1}(i+1)} P_X(x)\prod_{t=1}^n V_1\big(i_t|f(x)\big)\\
&= P_X(x_{i+1}^*)\sum_{l=0}^{\tau_n(i,\rho)}\binom{n}{l}(1-\rho)^l\rho^{n-l}\\
&\quad + P_X(x_i^*)\sum_{l=\tau_n(i,\rho)+1}^{n}\binom{n}{l}\rho^l(1-\rho)^{n-l}\\
&= P_X(x_{i+1}^*)\sum_{l=n-\tau_n(i,\rho)}^{n}\binom{n}{l}\rho^l(1-\rho)^{n-l}\\
&\quad + P_X(x_i^*)\sum_{l=\tau_n(i,\rho)+1}^{n}\binom{n}{l}\rho^l(1-\rho)^{n-l}\\
&\le P_X(x_{i+1}^*)\sum_{l=\lfloor \frac{n}{2}\rfloor+1}^{n}\binom{n}{l}\rho^l(1-\rho)^{n-l}\\
&\quad + P_X(x_i^*)\sum_{l=0}^{n}\binom{n}{l}\rho^l(1-\rho)^{n-l} \qquad (54)
\end{aligned}$$

where the first term in the previous inequality readily follows from the observation above, since

$$n-\tau_n(i,\rho) \ge \begin{cases} n-\lfloor\frac{n}{2}\rfloor \ge \lfloor\frac{n}{2}\rfloor+1 & \text{for odd } n\\ \frac{n}{2}+1 & \text{for even } n.\end{cases}$$

Note that when $\tau_n(i,\rho)<0$, this upper bound in (54) remains valid. By (52) and (54),

$$\begin{aligned}
&P\big(g_{\mathrm{MAP}(V_1^n)}(Z_1,\ldots,Z_n)=X\big)\\
&\le \sum_{i=0,2,\ldots,k-2}\Big[P_X(x_{i+1}^*)\,P\Big(\mathrm{Bin}(n,\rho)\ge\Big\lfloor\frac{n}{2}\Big\rfloor+1\Big)\\
&\quad + P_X(x_i^*)\,P\Big(\mathrm{Bin}(n,\rho)\ge\Big\lfloor\frac{n}{2}\Big\rfloor+1\Big)\\
&\quad + P_X(x_i^*)\,P\Big(\mathrm{Bin}(n,\rho)\le\Big\lfloor\frac{n}{2}\Big\rfloor\Big)\Big]\\
&= \Big(\sum_{i\in\mathcal{Z}}P_X(x_i^*)\Big)P\Big(\mathrm{Bin}(n,\rho)\ge\Big\lfloor\frac{n}{2}\Big\rfloor+1\Big)\\
&\quad + \Big(\sum_{i=0,2,\ldots,k-2}P_X(x_i^*)\Big)P\Big(\mathrm{Bin}(n,\rho)\le\Big\lfloor\frac{n}{2}\Big\rfloor\Big) \qquad (55)\\
&= \sum_{i\in\mathcal{Z}}P_X(x_i^*)-\Lambda_n(\rho). \qquad (56)
\end{aligned}$$

[2]When $P(x_i^*)=P(x_{i+1}^*)$, we get $\tau_n(i,\rho)=n/2$ for even n. In this case, replacing $\tau_n(i,\rho)=n/2$ by $\tau_n(i,\rho)=n/2-1$ does not alter subsequent calculations.

(ii) *k odd:* For $i=0,2,\ldots,k-3$, set $\mathcal{B}_n(i)$ as in (51), and

$$\mathcal{B}_n(k-1)=\big\{(i_1,\ldots,i_n)\in\mathcal{Z}^n : i_t=0 \text{ or } i_t=k-1\big\}.$$

Then

$$\begin{aligned}
&P\big(g_{\mathrm{MAP}(V_1^n)}(Z_1,\ldots,Z_n)=X\big)\\
&\le \sum_{i=0,2,\ldots,k-3}\ \sum_{(i_1,\ldots,i_n)\in\mathcal{B}_n(i)}\\
&\qquad \max_{x\in f^{-1}(i)\cup f^{-1}(i+1)} P_X(x)\prod_{t=1}^n V_1\big(i_t|f(x)\big) \qquad (57)\\
&\quad + \sum_{(i_1,\ldots,i_n)\in\mathcal{B}_n(k-1)}\ \max_{x\in f^{-1}(k-1)} P_X(x)\prod_{t=1}^n V_1\big(i_t|f(x)\big)\\
&\le \Big(\sum_{i=0}^{k-2}P_X(x_i^*)\Big)P\Big(\mathrm{Bin}(n,\rho)\ge\Big\lfloor\frac{n}{2}\Big\rfloor+1\Big)\\
&\quad + \sum_{i=0,2,\ldots,k-3}P_X(x_i^*)\,P\Big(\mathrm{Bin}(n,\rho)\le\Big\lfloor\frac{n}{2}\Big\rfloor\Big)+P_X(x_{k-1}^*)\\
&\quad \times\Big(P\Big(\mathrm{Bin}(n,\rho)\ge\Big\lfloor\frac{n}{2}\Big\rfloor+1\Big)+P\Big(\mathrm{Bin}(n,\rho)\le\Big\lfloor\frac{n}{2}\Big\rfloor\Big)\Big)\\
&= \Big(\sum_{i\in\mathcal{Z}}P_X(x_i^*)\Big)P\Big(\mathrm{Bin}(n,\rho)\ge\Big\lfloor\frac{n}{2}\Big\rfloor+1\Big)\\
&\quad + \Big(\sum_{i=0,2,\ldots,k-1}P_X(x_i^*)\Big)P\Big(\mathrm{Bin}(n,\rho)\le\Big\lfloor\frac{n}{2}\Big\rfloor\Big)\\
&= \sum_{i\in\mathcal{Z}}P_X(x_i^*)-\Lambda_n(\rho) \qquad (58)
\end{aligned}$$

where in the inequality above, the first two terms on the right-side obtain *a la* (55). When $(i_1,\ldots,i_n)=(0,\ldots,0)$ (the all-zero sequence), the maximum in (57) is over x in $f^{-1}(0)\cup f^{-1}(1)\cup f^{-1}(k-1)$. The preceding calculations are, in effect over x in $f^{-1}(0)$, and are justified since $P_X(x_0^*)\rho^n \ge P_X(x_1^*)(1-\rho)^n \ge P_X(x_{k-1}^*)(1-\rho)^n$ for $0.5\le\rho\le 1$.

The assertion of the theorem holds by (56) and (58). ■

2) Realm $0\le\rho\le 0.5$: Our achievability scheme uses ρ-QRs as in (47) with V_1 replaced by $V_2:\mathcal{Z}\to\mathcal{Z}$, not depending on P_X, which is: for $0\le\rho\le 1/k$,

$$V_2(i|j)=\frac{1}{k},\quad i,j\in\mathcal{Z} \qquad (59)$$

and for $1/k<\rho\le 0.5$,

$$V_2(i|j) = \begin{cases}
\frac{1}{\lfloor\frac{1}{\rho}\rfloor}, & j=0,\ldots,\Big\lfloor\frac{k}{\lfloor\frac{1}{\rho}\rfloor}\Big\rfloor\Big\lfloor\frac{1}{\rho}\Big\rfloor-1,\\
& i=\Big\lfloor\frac{j}{\lfloor\frac{1}{\rho}\rfloor}\Big\rfloor\Big\lfloor\frac{1}{\rho}\Big\rfloor,\ldots,\Big(\Big\lfloor\frac{j}{\lfloor\frac{1}{\rho}\rfloor}\Big\rfloor+1\Big)\Big\lfloor\frac{1}{\rho}\Big\rfloor-1;\\
\frac{1}{k \bmod \lfloor\frac{1}{\rho}\rfloor}, & \\
& j=\Big\lfloor\frac{k}{\lfloor\frac{1}{\rho}\rfloor}\Big\rfloor\Big\lfloor\frac{1}{\rho}\Big\rfloor,\ldots,k-1,\\
& i=\Big\lfloor\frac{j}{\lfloor\frac{1}{\rho}\rfloor}\Big\rfloor\Big\lfloor\frac{1}{\rho}\Big\rfloor,\ldots,k-1;\\
0, & \text{otherwise.}
\end{cases} \qquad (60)$$

$$\left[\begin{array}{ccc:ccc:cc} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{array}\right]$$

Fig. 3. Add-noise ρ-QR V_2 for $\rho = 1/3$ and $k = 8$.

In particular, for $1/k < \rho \leq 0.5$, the $k \times k$-matrix V_2 consists of $\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor$ diagonal blocks of $\left\lfloor \frac{1}{\rho} \right\rfloor \times \left\lfloor \frac{1}{\rho} \right\rfloor$-matrices, each with identical elements equal to $1/\left\lfloor \frac{1}{\rho} \right\rfloor$; and a single "filler" block of size $k \bmod \left\lfloor \frac{1}{\rho} \right\rfloor \times k \bmod \left\lfloor \frac{1}{\rho} \right\rfloor$ with identical elements equal to $1/\left(k \bmod \left\lfloor \frac{1}{\rho} \right\rfloor\right)$. The latter is vacuous if $\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor \left\lfloor \frac{1}{\rho} \right\rfloor = k$, i.e., $k \bmod \left\lfloor \frac{1}{\rho} \right\rfloor = 0$. See Fig. 3.

Theorem 6: Let $0 \leq \rho \leq 0.5$. *For every* $n \geq 1$, *the add-noise* ρ*-QRs* $\{Z_t = F_t(X)\}_{t=1}^n$ *in* (47) *with* V_1 *replaced by* $V_2 : \mathcal{Z} \to \mathcal{Z}$ *in* (59), (60) *yield privacy*

$$\pi_\rho\left(V_2^n\right) = \begin{cases} 1 - P_X\left(x^*\right), & 0 \leq \rho \leq \frac{1}{k} \\ 1 - \displaystyle\sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor} P_X\left(x^*_{i\lfloor \frac{1}{\rho} \rfloor}\right), & k \bmod \left\lfloor \frac{1}{\rho} \right\rfloor \neq 0, \ \frac{1}{k} < \rho \leq 0.5 \\ 1 - \displaystyle\sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor - 1} P_X\left(x^*_{i\lfloor \frac{1}{\rho} \rfloor}\right), & k \bmod \left\lfloor \frac{1}{\rho} \right\rfloor = 0, \ \frac{1}{k} < \rho \leq 0.5. \end{cases} \tag{61}$$

Remark: For $0 \leq \rho \leq 0.5$, the privacy $\pi_\rho\left(V_2^n\right)$ above lacks dependence on n. However, for $0 \leq \rho \leq 1/k$,

$$\pi_\rho\left(V_2^n\right) = \pi_\rho\left(V_2\right) = 1 - P_X\left(x^*\right) = \pi(\rho)$$

where the last identity is by (19). Thus, for $n = 1$, the add-noise ρ-QR with V_2 too achieves ρ-privacy, as did V_o in Theorem 2.

On the other hand, for $1/k < \rho \leq 0.5$, V_2 can be strictly inferior to V_o for $n = 1$; for instance, with P_X being the uniform pmf on $\mathcal{X}$, by Theorem 6 with $k \bmod \left\lfloor \frac{1}{\rho} \right\rfloor \neq 0$,

$$\begin{aligned} \pi_\rho\left(V_2\right) &= 1 - \left(\left\lfloor \frac{k}{\left\lfloor \frac{1}{\rho} \right\rfloor} \right\rfloor + 1\right) \frac{1}{r} \\ &< 1 - \frac{k}{\left\lfloor \frac{1}{\rho} \right\rfloor} \frac{1}{r} \leq 1 - \frac{\rho k}{r} = \pi(\rho) = \pi(V_o) \end{aligned}$$

where the last two identities are by Theorem 2.

Proof: We have

$$P\big(g_{\mathrm{MAP}(V_2^n)}\left(Z_1, \ldots, Z_n\right) = X\big) = \sum_{(i_1,\ldots,i_n) \in \mathcal{Z}^n} \max_{x \in \mathcal{X}} P_X(x) \prod_{t=1}^n V_2\big(i_t | f(x)\big). \tag{62}$$

When $0 \leq \rho \leq 1/k$, we get from (59) that

$$P\big(g_{\mathrm{MAP}(V_2^n)}\left(Z_1, \ldots, Z_n\right) = X\big) = P_X\left(x^*\right)$$

so that $\pi_\rho\left(V_2^n\right) = 1 - P_X\left(x^*\right)$. Considering next $1/k < \rho \leq 0.5$, by the form of V_2 in (60), for each

$$x \in \mathcal{X} \setminus \left\{ f^{-1}\left(\left\lfloor \frac{k}{\left\lfloor \frac{1}{\rho} \right\rfloor} \right\rfloor \left\lfloor \frac{1}{\rho} \right\rfloor\right) \cup \ldots \cup f^{-1}(k-1) \right\}$$

only those $(i_1, \ldots, i_n) \in \mathcal{Z}^n$ yield nonzero contributions in (62) when

$$i_t \in \left\{ \left\lfloor \frac{f(x)}{\left\lfloor \frac{1}{\rho} \right\rfloor} \right\rfloor \left\lfloor \frac{1}{\rho} \right\rfloor, \ldots, \left(\left\lfloor \frac{f(x)}{\left\lfloor \frac{1}{\rho} \right\rfloor} \right\rfloor + 1\right) \left\lfloor \frac{1}{\rho} \right\rfloor - 1 \right\}, \quad t = 1, \ldots, n,$$

and for each

$$x \in \left\{ f^{-1}\left(\left\lfloor \frac{k}{\left\lfloor \frac{1}{\rho} \right\rfloor} \right\rfloor \left\lfloor \frac{1}{\rho} \right\rfloor\right) \cup \ldots \cup f^{-1}(k-1) \right\}$$

only those $(i_1, \ldots, i_n) \in \mathcal{Z}^n$ yield nonzero contributions in (62) when

$$i_t \in \left\{ \left\lfloor \frac{f(x)}{\left\lfloor \frac{1}{\rho} \right\rfloor} \right\rfloor \left\lfloor \frac{1}{\rho} \right\rfloor, \ldots, k-1 \right\}, \quad t = 1, \ldots, n.$$

For $i = 0, \ldots, \left(\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor - 1\right)$, set

$$\mathcal{C}_n(i) = \left\{ (i_1, \ldots, i_n) \in \mathcal{Z}^n : i_t \in \left\{ i \left\lfloor \frac{1}{\rho} \right\rfloor, \ldots, (i+1) \left\lfloor \frac{1}{\rho} \right\rfloor - 1 \right\} \right\}$$

and, when the filler block above exists,

$$\mathcal{C}_n\left(\left\lfloor \frac{k}{\left\lfloor \frac{1}{\rho} \right\rfloor} \right\rfloor\right) = \left\{ (i_1, \ldots, i_n) \in \mathcal{Z}^n : i_t \in \left\{ \left\lfloor \frac{k}{\left\lfloor \frac{1}{\rho} \right\rfloor} \right\rfloor \left\lfloor \frac{1}{\rho} \right\rfloor, \ldots, k-1 \right\} \right\}.$$

Then in (62), with the filler block existing

$$\begin{aligned} &P\left(g_{\mathrm{MAP}(V_2^n)}\left(Z_1, \ldots, Z_n\right) = X\right) \\ &= \sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor - 1} \sum_{(i_1,\ldots,i_n) \in \mathcal{C}_n(i)} \max_{x \in f^{-1}\left(i \lfloor \frac{1}{\rho} \rfloor\right) \cup \ldots \cup f^{-1}\left((i+1)\lfloor \frac{1}{\rho} \rfloor - 1\right)} P_X(x) \prod_{t=1}^n V_2\big(i_t | f(x)\big) \\ &+ \sum_{(i_1,\ldots,i_n) \in \mathcal{C}_n\left(\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor\right)} \max_{x \in f^{-1}\left(\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor \lfloor \frac{1}{\rho} \rfloor\right) \cup \ldots \cup f^{-1}(k-1)} P_X(x) \prod_{t=1}^n V_2\big(i_t | f(x)\big) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor - 1} \sum_{(i_1,\ldots,i_n)\in\mathcal{C}_n(i)} \max_{x\in f^{-1}\left(i\lfloor\frac{1}{\rho}\rfloor\right)\cup\ldots\cup f^{-1}\left((i+1)\lfloor\frac{1}{\rho}\rfloor - 1\right)} P_X(x)\left(\frac{1}{\lfloor\frac{1}{\rho}\rfloor}\right)^n \\
&\quad + \sum_{(i_1,\ldots,i_n)\in\mathcal{C}_n\left(\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor}\right\rfloor\right)} \max_{x\in f^{-1}\left(\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor}\right\rfloor \lfloor\frac{1}{\rho}\rfloor\right)\cup\ldots\cup f^{-1}(k-1)} P_X(x)\left(\frac{1}{k \bmod \lfloor\frac{1}{\rho}\rfloor}\right)^n \\
&= \sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor - 1} \left(\frac{1}{\lfloor\frac{1}{\rho}\rfloor}\right)^n \sum_{(i_1,\ldots,i_n)\in\mathcal{C}_n(i)} \max_{x\in f^{-1}\left(i\lfloor\frac{1}{\rho}\rfloor\right)\cup\ldots\cup f^{-1}\left((i+1)\lfloor\frac{1}{\rho}\rfloor - 1\right)} P_X(x) \\
&\quad + \left(\frac{1}{k \bmod \lfloor\frac{1}{\rho}\rfloor}\right)^n \sum_{(i_1,\ldots,i_n)\in\mathcal{C}_n\left(\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor}\right\rfloor\right)} \max_{x\in f^{-1}\left(\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor}\right\rfloor \lfloor\frac{1}{\rho}\rfloor\right)\cup\ldots\cup f^{-1}(k-1)} P_X(x) \\
&= \sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor - 1} \left(\frac{1}{\lfloor\frac{1}{\rho}\rfloor}\right)^n \sum_{(i_1,\ldots,i_n)\in\mathcal{C}_n(i)} P_X\left(x^*_{i\lfloor\frac{1}{\rho}\rfloor}\right) \\
&\quad + \left(\frac{1}{k \bmod \lfloor\frac{1}{\rho}\rfloor}\right)^n \sum_{(i_1,\ldots,i_n)\in\mathcal{C}_n\left(\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor}\right\rfloor\right)} P_X\left(x^*_{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor}\right\rfloor \lfloor\frac{1}{\rho}\rfloor}\right) \\
&= \sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor} P_X\left(x^*_{i\lfloor\frac{1}{\rho}\rfloor}\right), \qquad (63)
\end{aligned}
$$

where (63) uses (45) and $|\mathcal{C}_n(i)| = \left(\lfloor\frac{1}{\rho}\rfloor\right)^n$, $\left|\mathcal{C}_n\left(\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor}\right\rfloor\right)\right| = \left(k \bmod \lfloor\frac{1}{\rho}\rfloor\right)^n$. In the absence of the filler block, clearly

$$
P\left(g_{\mathrm{MAP}(V_2^n)}(Z_1,\ldots,Z_n) = X\right) = \sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor - 1} P_X\left(x^*_{i\lfloor\frac{1}{\rho}\rfloor}\right).
$$

The assertion of the theorem follows. ■

C. Useful Bounds for $\pi_n(\rho)$

Theorems 4 and 5 yield effective upper and lower bounds for $\pi_n(\rho)$. Upon rewriting these bounds with a slight weakening, useful information can be extracted concerning the limiting behaviour of $\pi_n(\rho)$ as $n \to \infty$. Specifically by Theorem 4, for each $0 \le \rho \le 1$ and for every $n \ge 1$,

$$
\pi_n(\rho) \le 1 - \sum_{i\in\mathcal{Z}} P_X(x_i^*) + \Gamma_n(\rho) \qquad (64)
$$

and by Theorem 5, for $0.5 < \rho \le 1$ and for every $n \ge 1$,

$$
\pi_n(\rho) \ge \pi_\rho(V_1^n) \ge 1 - \sum_{i\in\mathcal{Z}} P_X(x_i^*) + \Lambda_n(\rho). \qquad (65)
$$

Estimates of $P\left(\mathrm{Bin}(n,\rho) \le \lfloor\frac{n}{2}\rfloor\right)$ appearing in $\Gamma_n(\rho)$ and $\Lambda_n(\rho)$ (cf. (36) and (48)) lead to useful bounds for $\pi_n(\rho)$ in (64) and (65). Let $\mathrm{Ber}(\alpha)$ denote a Bernoulli rv with the probability of "1" being α, $0 \le \alpha \le 1$. Hereafter, all logarithms and exponentials are with respect to the base 2.

Lemma 7:

(i) For each $0.5 \le \rho \le 1$ *and every* $n \ge 1$,

$$
\begin{aligned}
&\frac{1}{n+1}\exp\left[-nD\left(\mathrm{Ber}\left(\frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor\right)\Big|\Big|\mathrm{Ber}(\rho)\right)\right] \\
&\le P\left(\mathrm{Bin}(n,\rho) \le \left\lfloor\frac{n}{2}\right\rfloor\right) \\
&\le \left(\left\lfloor\frac{n}{2}\right\rfloor + 1\right)\exp\left[-nD\left(\mathrm{Ber}\left(\frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor\right)\Big|\Big|\mathrm{Ber}(\rho)\right)\right].
\end{aligned}
$$

(ii) For each $0 \le \rho \le 0.5$ *and for every* $n \ge 1$,

$$
P\left(\mathrm{Bin}(n,\rho) \le \left\lfloor\frac{n}{2}\right\rfloor\right) \ge 1 - \rho.
$$

Proof: See Appendix B. ■

Lemma 7(i) leads to the following useful bounds for $\pi_n(\rho)$.

Proposition 8: For each $0.5 < \rho \le 1$,

(i)

$$
\begin{aligned}
\pi_n(\rho) \le 1 - \sum_{i\in\mathcal{Z}} P_X(x_i^*) &+ \left(\left\lfloor\frac{n}{2}\right\rfloor + 1\right) \\
&\times \exp\left[-nD\left(\mathrm{Ber}\left(\frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor\right)\Big|\Big|\mathrm{Ber}(\rho)\right)\right] \\
&\times \sum_{i\in\mathcal{Z}} P_X(x_i^*)
\end{aligned}
$$

for all n *such that*

$$
\left(\left\lfloor\frac{n}{2}\right\rfloor + 1\right)\exp\left[-nD\left(\mathrm{Ber}\left(\frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor\right)\Big|\Big|\mathrm{Ber}(\rho)\right)\right] \le 1 - \min\{\rho,\rho_c\}.
$$

(ii) for every $n \ge 1$,

$$
\begin{aligned}
\pi_n(\rho) \ge \pi_\rho(V_1^n) \ge 1 - \sum_{i\in\mathcal{Z}} P_X(x_i^*) &+ \frac{1}{n+1} \\
&\times \exp\left[-nD\left(\mathrm{Ber}\left(\frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor\right)\Big|\Big|\mathrm{Ber}(\rho)\right)\right] \\
&\times \left(\sum_{i\in\mathcal{Z}:\, i \text{ odd}} P_X(x_i^*)\right).
\end{aligned}
$$

Proof: The assertions follow directly by applying the upper and lower bounds in Lemma 7(i) to the right-sides of (64) and (65), respectively, and recalling (36) and (48). ■

D. Asymptotic Implications

We close this section with useful asymptotic implications of Theorem 4, 5, 6 and Proposition 8. Considering first the (more interesting) realm $0.5 < \rho \leq 1$, the upper bounds for $\pi_n(\rho)$ in Theorem 4 and Proposition 8(i), as also the lower bounds in Theorem 5 and Proposition 8(ii), converge according to

$$\lim_n \pi_n(\rho) = 1 - \sum_{i \in \mathcal{Z}} P_X\left(x_i^*\right) = \pi(1), \quad 0.5 < \rho \leq 1 \quad (66)$$

(see Remark (ii) after Theorem 2), i.e., the error probability of a MAP estimator of X on the basis of a knowledge of $f(X)$. Furthermore, both the sets of bounds converge at the same exponential rate in n with the (n-dependent) exponent itself tending to $D\big(\text{Ber}(0.5)||\text{Ber}(\rho)\big) > 0$. Thus, in the realm $0.5 < \rho \leq 1$, the asymptotic privacy in (66) is that which is afforded when the querier forms an accurate MAP estimate of $f(X)$ w.p. 1 from ρ-QRs $\{F_t(X)\}_{t=1}^n$, followed by a MAP estimate of X that is compatible with the estimated $f(X)$.

In the realm $0 \leq \rho \leq 0.5$, the upper bound for $\pi_n(\rho)$ in Theorem 4, by Lemma 7(ii), equals

$$1 - \max\{\rho_c, \rho\} \sum_{i \in \mathcal{Z}} P_X\left(x_i^*\right) \quad (67)$$

for *all* $n \geq 1$, which is the ρ-privacy for $n = 1$ in Theorem 2. As remarked after Theorem 6, this upper bound is unattainable, in general, by add-noise ρ-QRs $\{F_t(X)\}_{t=1}^n$ with $V_2 : \mathcal{Z} \rightarrow \mathcal{Z}$ in (59), (60). Hence, an interpretation as above in the complementary realm is lacking as is the answer to the putative tightness (or not) of the mentioned bound. However, since

$$\pi_n(\rho) \geq \pi_\rho\left(V_2^n\right) > 1 - \sum_{i \in \mathcal{Z}} P_X\left(x_i^*\right) = \pi(1),$$

where the strict inequality is evident from Theorem 6 (by comparing the expressions in (61) with $1 - \sum_{i \in \mathcal{Z}} P_X\left(x_i^*\right)$), we can conclude that no accurate estimate of $f(X)$ w.p. 1 is possible from ρ-QRs $\{F_t(X)\}_{t=1}^n$ for any n, unlike for $0.5 < \rho \leq 1$.

V. Inadequacy of Conditionally i.i.d W_o for Multiple Query Responses

Theorem 2 establishes the optimality of the add-noise ρ-QR $W_o : \mathcal{X} \rightarrow \mathcal{Z}$, or equivalently $V_o : \mathcal{Z} \rightarrow \mathcal{Z}$, in achieving ρ-privacy $\pi(\rho)$, $0 \leq \rho \leq 1$, for $n = 1$. Upon choosing $W_t = W_o$ or $V_t = V_o$, $t = 1, \ldots, n$, $n \geq 2$, in (31) or (32), respectively, how does the corresponding privacy $\pi_\rho\left(W_o^n\right)$ or $\pi_\rho\left(V_o^n\right)$ compare with the achievable privacy in Theorems 5 and 6? In the regime of all suitably large n, we show below that the former does not exceed the latter and, in fact, can be strictly smaller.

To this end, the concept of Chernoff information [9] plays a material role. Given a stochastic matrix $V : \mathcal{Z} \rightarrow \mathcal{Z}$, define its *Chernoff radius*, denoted $C(V)$, as the minimum of pairwise Chernoff information quantities:

$$\begin{aligned} C(V) &= \min_{\substack{j \neq j' \\ j, j' \in \mathcal{Z}}} C(j, j') \\ &= \min_{\substack{j \neq j' \\ j, j' \in \mathcal{Z}}} \left[- \min_{0 \leq \lambda \leq 1} \log \left(\sum_{i \in \mathcal{Z}} V(i|j)^\lambda V\left(i|j'\right)^{1-\lambda} \right) \right], \end{aligned} \quad (68)$$

noting that $C(V) \geq 0$ with $C(V) > 0$ iff all the rows of V are distinct.

Also useful will be the next two technical lemmas. Let $\tilde{f}(X)$ be a $\mathcal{Z}$-valued rv with pmf

$$P\left(\tilde{f}(X) = i\right) = \frac{P_X(x_i^*)}{\sum_{l \in \mathcal{Z}} P_X\left(x_l^*\right)}, \quad i \in \mathcal{Z}$$

with x_i^*, $i \in \mathcal{Z}$, as in (15). Let $\tilde{Z}_t$, $t = 1, \ldots, n$, be conditionally mutually independent $\mathcal{Z}$-valued rvs conditioned on $\tilde{f}(X)$, with

$$P_{\tilde{Z}_t | \tilde{f}(X)} = V, \quad t = 1, \ldots, n.$$

We use the notation $A \doteq \exp(-nB)$ to mean $\lim_n -\frac{1}{n} \log A = B$ (cf. e.g., [24]).

Lemma 9: For $0 \leq \rho \leq 1$, consider add-noise ρ-QRs $\{F_t(X)\}_{t=1}^\infty$ with (31) *holding for every $n \geq 1$, where $W_t = W$, $t \geq 1$, and $W : \mathcal{X} \rightarrow \mathcal{Z}$ has identical rows for all $x \in f^{-1}(i)$, $i \in \mathcal{Z}$, and has associated $V : \mathcal{Z} \rightarrow \mathcal{Z}$ in* (12).

(i) The corresponding privacy for every $n \geq 1$ is

$$\begin{aligned} \pi_\rho\left(W^n\right) = \pi_\rho\left(V^n\right) = 1 - &\left(\sum_{i \in \mathcal{Z}} P_X\left(x_i^*\right) \right) \\ &\times P\left(g_{\text{MAP }(V^n)}\left(\tilde{Z}_1, \ldots, \tilde{Z}_n\right) = \tilde{f}(X)\right). \end{aligned} \quad (69)$$

(ii) Furthermore,

$$\pi_\rho\left(V^n\right) - \left(1 - \sum_{i \in \mathcal{Z}} P_X\left(x_i^*\right)\right) \doteq \exp\left[-nC(V)\right]. \quad (70)$$

Proof:

(i)

$$\begin{aligned} &P\left(g_{\text{MAP}(W^n)}\left(Z_1, \ldots, Z_n\right) = X\right) \\ &= \sum_{(i_1, \ldots, i_n) \in \mathcal{Z}^n} \max_{x \in \mathcal{X}} P_X(x) \prod_{t=1}^n W\left(i_t | x\right) \\ &= \sum_{(i_1, \ldots, i_n) \in \mathcal{Z}^n} \max_{\substack{x \in \cup f^{-1}(j) \\ j \in \mathcal{Z}}} P_X(x) \prod_{t=1}^n W\left(i_t | x\right) \\ &= \sum_{(i_1, \ldots, i_n) \in \mathcal{Z}^n} \max_{j \in \mathcal{Z}} P_X\left(x_j^*\right) \prod_{t=1}^n W\left(i_t | x_j^*\right) \\ &= \sum_{(i_1, \ldots, i_n) \in \mathcal{Z}^n} \max_{j \in \mathcal{Z}} P_X\left(x_j^*\right) \prod_{t=1}^n V\left(i_t | f\left(x_j^*\right)\right), \end{aligned}$$

by (12)

$$= \sum_{(i_1,\ldots,i_n)\in\mathcal{Z}^n} \max_{j\in\mathcal{Z}} P_X(x_j^*) \prod_{t=1}^{n} V(i_t|j) \tag{71}$$

$$= \left(\sum_{i\in\mathcal{Z}} P_X(x_i^*)\right) \times \sum_{(i_1,\ldots,i_n)\in\mathcal{Z}^n} \max_{j\in\mathcal{Z}} \frac{P_X(x_j^*)}{\sum_{i\in\mathcal{Z}} P_X(x_i^*)} \prod_{t=1}^{n} V(i_t|j)$$

$$= \left(\sum_{i\in\mathcal{Z}} P_X(x_i^*)\right) \times P\left(g_{\mathrm{MAP}(V^n)}\left(\tilde{Z}_1,\ldots,\tilde{Z}_n\right) = \tilde{f}(X)\right) \tag{72}$$

where the third equality above is by the assumed form of W.

The first assertion in (69) follows from (71) and the second from (72).

(ii) By [24, Th. 2],

$$P\left(g_{\mathrm{MAP}(V^n)}\left(\tilde{Z}_1,\ldots,\tilde{Z}_n\right) \neq \tilde{f}(X)\right) \doteq \exp\left[-nC(V)\right],$$

which, applied to (69), yields (70). ■

Remark: Observe that a direct application of [24, Th. 2] to

$$\pi_\rho(W^n) = P\left(g_{\mathrm{MAP}(W^n)}(Z_1,\ldots,Z_n) \neq X\right)$$

is not useful as it yields

$$\pi_\rho(W^n) \doteq \exp\left[-nC(W)\right]$$

where the Chernoff radius of $W : \mathcal{X} \to \mathcal{Z}$ is $C(W) = 0$ owing to the presence of identical rows when $k \leq r - 1$.

Lemma 10: For V_o in (18) *and V_1 in* (46)*, we have*

$$C(V_1) = D\big(\mathrm{Ber}(0.5)||\mathrm{Ber}(\rho)\big) = -\log 2\sqrt{\rho(1-\rho)}, \quad 0 \leq \rho \leq 1 \tag{73}$$

and for $0.5 < \rho < 1$

$$C(V_o) = -\log 2\sqrt{\max\{\rho_c,\rho\}\left(1-\max\{\rho_c,\rho\}\right)}, \quad k = 2 \tag{74}$$

$$C(V_o) > -\log 2\sqrt{\max\{\rho_c,\rho\}\left(1-\max\{\rho_c,\rho\}\right)}, \quad k \geq 3. \tag{75}$$

Proof: First observe that for $0 < \rho < 1$,

$$C(V_1) = \sup_{0<\lambda<1} \log -\left(\rho^\lambda (1-\rho)^{1-\lambda} + \rho^{1-\lambda}(1-\rho)^\lambda\right) = \log \frac{1}{\inf_{0<\lambda<1} \rho^\lambda (1-\rho)^{1-\lambda} + \rho^{1-\lambda}(1-\rho)^\lambda} = \log \frac{1}{2\sqrt{\rho(1-\rho)}} = D\big(\mathrm{Ber}(0.5)||\mathrm{Ber}(\rho)\big) \tag{76}$$

where the infimum is attained as a minimum at $\lambda = 0.5$; and $C(V_1) = \infty$ for $\rho = 0$ and $\rho = 1$. The last equality above is by simple calculation.

Turning to (74), for $k = 2$,

$$C(V_o) = \sup_{0<\lambda<1} \log -\Big(\left(\max\{\rho_c,\rho\}\right)^\lambda \left(1-\max\{\rho_c,\rho\}\right)^{1-\lambda} + \left(\max\{\rho_c,\rho\}\right)^{1-\lambda}\left(1-\max\{\rho_c,\rho\}\right)^\lambda\Big) = \log \frac{1}{2\sqrt{\max\{\rho_c,\rho\}\left(1-\max\{\rho_c,\rho\}\right)}},$$

in the manner of (76).

To show (75), for $j \neq j'$ in $\mathcal{Z}$,

$$C(j,j') = \sup_{0<\lambda<1} \log \frac{1}{\sum_{i\in\mathcal{Z}} V_o(i|j)^\lambda V_o(i|j')^{1-\lambda}} = \sup_{0<\lambda<1} (1-\lambda)\, D_\lambda\big(V_o(.|j)||V_o(.|j')\big) \tag{77}$$

where D_λ is the Rényi divergence of order λ [31]. For each $\lambda \in (0,1)$, since D_λ satisfies the data processing theorem [14, Theorem 1], we get

$$D_\lambda\big(V_o(.|j)||V_o(.|j')\big) \geq D_\lambda\left(\mathrm{Ber}\left(V_o(j'|j)\right)||\mathrm{Ber}\left(V_o(j'|j')\right)\right) = D_\lambda\left(\mathrm{Ber}\left(\frac{P_X(x_{j'}^*)}{\sum_{l\neq j} P_X(x_l^*)}\left(1-\max\{\rho_c,\rho\}\right)\right)\Big|\Big| \mathrm{Ber}\big(\max\{\rho_c,\rho\}\big)\right). \tag{78}$$

Claim: For $k \geq 3$ and $0.5 < \rho < 1$, the right-side of (78) is strictly larger than

$$D_\lambda\left(\mathrm{Ber}\left(1-\max\{\rho_c,\rho\}\right)\Big|\Big|\mathrm{Ber}\big(\max\{\rho_c,\rho\}\big)\right).$$

Then applying the claim to (77), for all $j \neq j'$ in $\mathcal{Z}$,

$$C(j,j') > \sup_{0<\lambda<1} (1-\lambda) D_\lambda\Big(\mathrm{Ber}\left(1-\max\{\rho_c,\rho\}\right)\Big|\Big|\mathrm{Ber}\big(\max\{\rho_c,\rho\}\big)\Big) \geq 0.5\, D_{0.5}\left(\mathrm{Ber}\left(1-\max\{\rho_c,\rho\}\right)\Big|\Big|\mathrm{Ber}\big(\max\{\rho_c,\rho\}\big)\right) = -\log 2\sqrt{\max\{\rho_c,\rho\}(1-\max\{\rho_c,\rho\})}$$

which yields (75).

It remains to prove the claim. Note that for $k \geq 3$, $P_X\left(x_{j'}^*\right)\Big/\sum_{l\neq j} P_X(x_l^*) < 1$ and so

$$\frac{P_X\left(x_{j'}^*\right)}{\sum_{l\neq j} P_X(x_l^*)}\left(1-\max\{\rho_c,\rho\}\right) < 1-\max\{\rho_c,\rho\} < \max\{\rho_c,\rho\}$$

since $\max\{\rho_c,\rho\} > 0.5$. Then, it suffices to show that $D_\lambda\big(\mathrm{Ber}(\alpha)||\mathrm{Ber}(\beta)\big)$ is (strictly) decreasing in α for $0 \leq \alpha < \beta$. We have

$$\frac{d}{d\alpha} D_\lambda\big(\mathrm{Ber}(\alpha)||\mathrm{Ber}(\beta)\big) = \frac{1}{\lambda-1} \frac{\lambda\alpha^{\lambda-1}\beta^{1-\lambda} - \lambda(1-\alpha)^{\lambda-1}(1-\beta)^{1-\lambda}}{\alpha^\lambda\beta^{1-\lambda} + (1-\alpha)^\lambda(1-\beta)^{1-\lambda}}. \tag{79}$$

Since $\lambda \in (0,1)$, the right-side of (79) is negative iff

$$\alpha^{\lambda-1}\beta^{1-\lambda} > (1-\alpha)^{\lambda-1}(1-\beta)^{1-\lambda},$$

i.e.,

$$\left(\frac{1-\alpha}{\alpha}\right)^{1-\lambda} > \left(\frac{1-\beta}{\beta}\right)^{1-\lambda}$$

which holds since $\alpha < \beta$. ■

Finally, we show that the privacy of add-noise ρ-QRs $\{F_t(X)\}_{t=1}^{\infty}$ under (32) for every $n \geq 1$ with $V_t = V_o$, $t \geq 1$, is no better than with $V_t = V_1$ or V_2 accordingly as $0.5 < \rho \leq 1$ or $0 \leq \rho \leq 0.5$; and, in fact, the former can be strictly smaller than the latter.

Proposition 11: For all n suitably large (depending on case below):

(i) $0 \leq \rho \leq 0.5$*:*

$$\pi_\rho\left(V_2^n\right) \geq \pi_\rho\left(V_o^n\right); \tag{80}$$

(ii) $0.5 < \rho < 1$*:*
$k = 2$ –

$$\pi_\rho\left(V_1^n\right) > \pi_\rho\left(V_o^n\right), \quad \rho < \rho_c \tag{81}$$

$$\pi_\rho\left(V_1^n\right) = \pi_\rho\left(V_o^n\right), \quad \rho \geq \rho_c; \tag{82}$$

$k \geq 3$ –

$$\pi_\rho\left(V_1^n\right) > \pi_\rho\left(V_o^n\right). \tag{83}$$

Proof:

(i) See Appendix D.

(ii) For $0 \leq \rho < 1$, we have by Lemma 9(ii),

$$\pi_\rho\left(V_o^n\right) - \left(1 - \sum_{i\in\mathcal{Z}} P_X\left(x_i^*\right)\right) \doteq \exp\left[-nC\left(V_o\right)\right], \tag{84}$$

and by Theorem 5 for $0.5 < \rho \leq 1$,

$$\begin{aligned}\pi_\rho\left(V_1^n\right) &- \left(1 - \sum_{i\in\mathcal{Z}} P_X\left(x_i^*\right)\right) \\ &\geq \Lambda_n(\rho), \quad n \geq 1 \\ &\doteq \exp\big[-nD\big(\mathrm{Ber}(0.5)||\mathrm{Ber}(\rho)\big)\big], \\ &\qquad \text{by (48) and Lemma 7(i)} \\ &= \exp\big[-nC(V_1)\big], \quad \text{by (73).}\end{aligned} \tag{85}$$

For $k = 2$ and $0.5 < \rho < \rho_c$, by (73) and (74),

$$\begin{aligned}C(V_1) &= -\log 2\sqrt{\rho(1-\rho)} \\ &< -\log 2\sqrt{\rho_c(1-\rho_c)} = C(V_o)\end{aligned}$$

so that (81) holds by (84) and (85). For $k = 2$ and $\rho \geq \rho_c$, observe in (46) and (18) that $V_1 = V_o$ whereby (82) holds. For $k \geq 3$ and $0.5 < \rho < 1$, by (73) and (75),

$$\begin{aligned}C\left(V_1\right) &= -\log 2\sqrt{\rho(1-\rho)} \\ &\leq -\log 2\sqrt{\max\{\rho_c,\rho\}\left(1-\max\{\rho_c,\rho\}\right)} \\ &< C(V_o)\end{aligned}$$

and so (83) holds by (84) and (85). ■

VI. Discussion

The choice of $W_o : \mathcal{X} \to \mathcal{Z}$ or $V_o : \mathcal{Z} \to \mathcal{Z}$ in (17), (18), depending on P_X through $P_X\left(x_i^*\right)$, $i \in \mathcal{Z}$, yields maximal privacy for a single ρ-QR for all $0 \leq \rho \leq 1$. However, for the case of multiple conditionally independent ρ-QRs, our achievability schemes in Section IV, that are "universal" in the sense of not depending on P_X, perform variously according to the value of ρ. In particular, for $0.5 < \rho \leq 1$, conditionally i.i.d. add-noise ρ-QRs $\{F_t(X)\}_{t=1}^{\infty}$ with $V_1 : \mathcal{Z} \to \mathcal{Z}$ in (46) are asymptotically optimal with privacy $\pi_\rho\left(V_1^n\right)$ converging to the limit of the upper bounds for ρ-privacy $\pi_n(\rho)$, $n \geq 1$, in Theorem 4. However, when $0 \leq \rho \leq 0.5$, our add-noise ρ-QRs with $V_2 : \mathcal{Z} \to \mathcal{Z}$ in (59), (60) yield privacy $\pi_\rho\left(V_2^n\right)$ not depending on n, which, in general, does not meet the corresponding upper bound in (67). Thus, it remains open whether conditionally independent ρ-QRs $\{F_t(X)\}_{t=1}^{\infty}$, that depend on P_X or are not necessarily of the add-noise variety, can outperform $\pi_\rho\left(V_1^n\right)$ or $\pi_\rho\left(V_2^n\right)$. Indeed, the goodness of our upper bound for $\pi_n(\rho)$ in (67), $0 \leq \rho \leq 0.5$ (that does not depend on n), is unresolved. These observations are analogous – in our setting – to the "composition" results for differential privacy (cf. e.g., [23]).

We conclude with a simple observation in explication of our approach mentioned in Section I. Suppose that the querier's family of priors $\mathcal{P}$ consists of a specified set of pmfs P on $\mathcal{X}$ with $P_X(x) > 0$, $x \in \mathcal{X}$. For a single ρ-QR, the ρ-privacy $\pi(\rho) = \pi(\rho; P)$ for any P in $\mathcal{P}$ is attained by $W_o = W_o(P)$ or $V_o = V_o(P)$ as remarked after Theorem 2. With

$$P_* = P_*(\rho) = \arg\min_{P\in\mathcal{P}} \ \pi(\rho; P), \quad 0 \leq \rho \leq 1$$

a ρ-QR $W_o(P_*)$ or $V_o(P_*)$ will yield privacy $\pi(\rho; P_*)$ in (19) that serves as a guaranteed lower bound for ρ-privacy computed according to *any* prior pmf P in $\mathcal{P}$. In the same vein, for $n \geq 1$ conditionally independent query responses, the minima with respect to P in $\mathcal{P}$ of the lower bound for $\pi_\rho\left(V_1^n\right)$ in (49) or of $\pi_\rho\left(V_2^n\right)$ in Theorem 6, respectively, serve as privacy guarantees in the realms $0.5 < \rho \leq 1$ or $0 \leq \rho \leq 0.5$, computed for any P in $\mathcal{P}$.

Appendix A
Proof of Achievability in Proposition 3

We have that $1 - \pi'_\rho\left(W'_o\right)$ equals

$$\begin{aligned}P\big(g'_{MAP(W'_o)}(Z) = Y\big) &= \sum_{i\in\mathcal{Z}} \max_{j\in\mathcal{Y}} P\left(Z = i, Y = j\right) \\ &= \sum_{i\in\mathcal{Z}} \max_{j\in\mathcal{Y}} \sum_{x\in h^{-1}(j)} P_X(x) W'_o(i|x).\end{aligned} \tag{86}$$

When $\rho'_c = 1$, we get in (86), upon using (29), that

$$\begin{aligned}&P\big(g'_{MAP(W'_o)}(Z) = Y\big) \\ &\quad = \sum_{i\in\mathcal{Z}} \max_{j\in\mathcal{Y}} \sum_{x\in h^{-1}(j)\cap f^{-1}(i)} P_X(x) = \sum_{i\in\mathcal{Z}} P_X\left(i, j_i^*\right).\end{aligned} \tag{87}$$

When $\rho'_c < 1$, for each $i \in \mathcal{Z}$ the summand in (86), upon using (30), is

$$\max_{j\in\mathcal{Y}} \Bigg[\sum_{x\in h^{-1}(j)\cap f^{-1}(i)} P_X(x)W'_o(i|x) + \sum_{x\in h^{-1}(j)\setminus f^{-1}(i)} P_X(x)W'_o(i|x) \Bigg]$$
$$= \max_{j\in\mathcal{Y}} \Bigg[\sum_{x\in h^{-1}(j)\cap f^{-1}(i)} P_X(x)\Big\{ \max\{\rho'_c,\rho\} + (1-\max\{\rho'_c,\rho\}) \frac{P_X(i,j_i^*) - P_X(i,j)}{\sum_{l\in\mathcal{Z}} P_X(l,j_l^*) - P_X(h^{-1}(j))} \Big\}$$
$$+ \sum_{x\in h^{-1}(j)\setminus f^{-1}(i)} P_X(x)\Big\{ (1-\max\{\rho'_c,\rho\}) \times \frac{P_X(i,j_i^*) - P_X(i,j)}{\sum_{l\in\mathcal{Z}} P_X(l,j_l^*) - P_X(h^{-1}(j))} \Big\} \Bigg]$$
$$= \max_{j\in\mathcal{Y}} \Bigg[\max\{\rho'_c,\rho\}P_X(i,j) + (1-\max\{\rho'_c,\rho\}) \times \frac{P_X(i,j_i^*) - P_X(i,j)}{\sum_{l\in\mathcal{Z}} P_X(l,j_l^*) - P_X(h^{-1}(j))} P_X\left(h^{-1}(j)\right) \Bigg]. \tag{88}$$

It suffices to show that the right-side of (88) is bounded above by $\max\{\rho'_c,\rho\}P_X(i,j_i^*)$ for each $i \in \mathcal{Z}$; this is done below. Then, in fact, the right-side of (88) equals $\max\{\rho'_c,\rho\}P_X(i,j_i^*)$ as seen by setting $j = j_i^*$ in the term within $[\cdots]$.

First consider the case when $\max\{\rho'_c,\rho\} < 1$. It is seen from (27) that for each $j \in \mathcal{Y}$,

$$\frac{\max\{\rho'_c,\rho\}}{1-\max\{\rho'_c,\rho\}} \ge \frac{\rho'_c}{1-\rho'_c} = \frac{P_X\left(h^{-1}(j^*)\right)}{\sum_{l\in\mathcal{Z}} P_X(l,j_l^*) - P_X(h^{-1}(j^*))} \ge \frac{P_X\left(h^{-1}(j)\right)}{\sum_{l\in\mathcal{Z}} P_X(l,j_l^*) - P_X(h^{-1}(j))}. \tag{89}$$

Using (89) in (88), and since $P_X\left(h^{-1}(j)\right) > 0$, $j \in \mathcal{Y}$, we get that the right-side of (88) is bounded above by $\max\{\rho'_c,\rho\}P_X(i,j_i^*)$. Also, this is true trivially when $\max\{\rho' c,\rho\} = 1$. Hence, we get that for each $i \in \mathcal{Z}$, the right-side of (88) equals $\max\{\rho'_c,\rho\}P_X(i,j_i^*)$. This, combined with (86)-(88), yields

$$\pi'_\rho(W'_o) = 1 - \max\{\rho'_c,\rho\}\sum_{i\in\mathcal{Z}} P_X(i,j_i^*). \qquad \blacksquare$$

Appendix B
Proof of Lemma 7

(i) For each $0 \le \rho \le 1$,

$$P\left(\text{Bin}(n,\rho) \le \left\lfloor\frac{n}{2}\right\rfloor\right) = \sum_{t=0}^{\lfloor \frac{n}{2}\rfloor} P\left(T_{\text{Ber}\left(\frac{t}{n}\right)}\right)$$

where $T_{\text{Ber}\left(\frac{t}{n}\right)}$ denotes the set of all n-length binary sequences of "type" $\text{Ber}\left(\frac{t}{n}\right)$, i.e., with t 1s (and $(n-t)$ 0s), so that

$$\max_{0\le t\le \lfloor\frac{n}{2}\rfloor} P\left(T_{\text{Ber}\left(\frac{t}{n}\right)}\right) \le P\left(\text{Bin}(n,\rho) \le \left\lfloor\frac{n}{2}\right\rfloor\right) \le \left(\left\lfloor\frac{n}{2}\right\rfloor + 1\right) \max_{0\le t\le \lfloor\frac{n}{2}\rfloor} P\left(T_{\text{Ber}\left(\frac{t}{n}\right)}\right). \tag{90}$$

Using well-known bounds for the probability of all n-length sequences of a given type (cf. [10, Lemma 2.6]), for each $0 \le \rho \le 1$, and noting that the number of types for binary sequences of length n equals $n+1$,

$$\frac{1}{n+1}\exp\left[-nD\left(\text{Ber}\left(\frac{t}{n}\right)\Big|\Big|\text{Ber}(\rho)\right)\right] \le P\left(T_{\text{Ber}\left(\frac{t}{n}\right)}\right) \le \exp\left[-nD\left(\text{Ber}\left(\frac{t}{n}\right)\Big|\Big|\text{Ber}(\rho)\right)\right] \tag{91}$$

and noting that for $0.5 \le \rho \le 1$,

$$\min_{0\le t\le \lfloor\frac{n}{2}\rfloor} D\left(\text{Ber}\left(\frac{t}{n}\right)\Big|\Big|\text{Ber}(\rho)\right) = D\left(\text{Ber}\left(\frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor\right)\Big|\Big|\text{Ber}(\rho)\right) \tag{92}$$

we have, by (91) and (92), from (90) that

$$\frac{1}{n+1}\exp\left[-nD\left(\text{Ber}\left(\frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor\right)\Big|\Big|\text{Ber}(\rho)\right)\right] \le P\left(\text{Bin}(n,\rho) \le \left\lfloor\frac{n}{2}\right\rfloor\right) \le \left(\left\lfloor\frac{n}{2}\right\rfloor + 1\right)\exp\left[-nD\left(\text{Ber}\left(\frac{1}{n}\left\lfloor\frac{n}{2}\right\rfloor\right)\Big|\Big|\text{Ber}(\rho)\right)\right].$$

(ii) We have that

$$P\left(\text{Bin}(n,\rho) \ge \left\lfloor\frac{n}{2}\right\rfloor + 1\right) = \sum_{t=\lfloor\frac{n}{2}\rfloor+1}^{n} \binom{n}{t}\rho^t(1-\rho)^{n-t}$$
$$= (1-\rho)^n \sum_{t=\lfloor\frac{n}{2}\rfloor+1}^{n} \binom{n}{t}\left(\frac{\rho}{1-\rho}\right)^t$$
$$\le (1-\rho)^n \left(\frac{\rho}{1-\rho}\right)^{\lfloor\frac{n}{2}\rfloor+1} \sum_{t=\lfloor\frac{n}{2}\rfloor+1}^{n} \binom{n}{t}, \quad \text{since } 0 \le \rho \le 0.5$$
$$\le (1-\rho)^n \left(\frac{\rho}{1-\rho}\right)^{\lfloor\frac{n}{2}\rfloor+1} 2^{n-1}$$
$$\le (1-\rho)^n \left(\frac{\rho}{1-\rho}\right)^{\frac{n-1}{2}+1} 2^{n-1}$$
$$= \rho\left(2\sqrt{\rho(1-\rho)}\right)^{n-1}$$
$$\le \rho, \quad \text{since } 2\sqrt{\rho(1-\rho)} \le 1 \text{ for } 0 \le \rho \le 1.$$

The assertion follows. $\blacksquare$

Appendix C
Proof of (33)

Since ρ and $\max_{i\in\mathcal{Z}}\ P\left(f(X)=i\right)$ are obvious lower bounds for the left-side of (33), it suffices to show that

$$P\Big(h_{\text{MAP}}\left(F_1(X),\ldots,F_n(X)\right)=f(X)\Big) \geq P\left(\text{Bin}(n,\rho)\geq\left\lfloor\frac{n}{2}\right\rfloor+1\right). \tag{93}$$

The proof bears a resemblance to that of Theorem 4 above and so we shall refer to pertinent details therein. We have

$$\begin{aligned}P\Big(h_{\text{MAP}}\left(F_1(X),\ldots,F_n(X)\right)=f(X)\Big) &= \sum_{(i_1,\ldots,i_n)\in\mathcal{Z}^n}\max_{j\in\mathcal{Z}}\ P\left(f(X)=j\right)\\ &\quad\times P\left(F_1(X)=i_1,\ldots,F_n(X)=i_n|f(X)=j\right).\end{aligned} \tag{94}$$

Since

$$\begin{aligned}&P\left(F_1(X)=i_1,\ldots,F_n(X)=i_n|f(X)=j\right)\\ &=\sum_{x\in\mathcal{X}}P\left(F_1(X)=i_1,\ldots,F_n(X)=i_n|f(X)=j,X=x\right)\\ &\quad\times P\left(X=x|f(X)=j\right)\\ &=\sum_{x\in f^{-1}(j)}\prod_{t=1}^{n}W_t\left(i_t|x\right)\frac{P_X(x)}{P\left(f(X)=j\right)},\end{aligned}$$

we get in (94) with $\mathcal{A}_l(i)$ in (39) that

$$\begin{aligned}&P\Big(h_{\text{MAP}}\left(F_1(X),\ldots,F_n(X)\right)=f(X)\Big)\\ &=\sum_{(i_1,\ldots,i_n)\in\mathcal{Z}^n}\max_{j\in\mathcal{Z}}\left(\sum_{x\in f^{-1}(j)}P_X(x)\prod_{t=1}^{n}W_t\left(i_t|x\right)\right)\\ &\geq\sum_{i\in\mathcal{Z}}\sum_{l=\left\lfloor\frac{n}{2}\right\rfloor+1}^{n}\sum_{(i_1,\ldots,i_n)\in\mathcal{A}_l(i)}\\ &\quad\max_{j\in\mathcal{Z}}\left(\sum_{x\in f^{-1}(j)}P_X(x)\prod_{t=1}^{n}W_t\left(i_t|x\right)\right)\\ &\geq\sum_{i\in\mathcal{Z}}\sum_{x\in f^{-1}(i)}P_X(x)\\ &\quad\left(\sum_{l=\left\lfloor\frac{n}{2}\right\rfloor+1}^{n}\sum_{(i_1,\ldots,i_n)\in\mathcal{A}_l(i)}\prod_{t=1}^{n}W_t\left(i_t|x\right)\right).\end{aligned}$$

Mimicking (40)-(43), observe that the sum above within $(\cdot)$ is bounded below by $P\left(\text{Bin}(n,\rho)\geq\left\lfloor\frac{n}{2}\right\rfloor+1\right)$. Clearly, (93) follows. ■

Appendix D
Proof of Proposition 11(i)

The following two lemmas are pertinent. Recall from (15) that $x^*=\arg\max_{x\in\mathcal{X}}\ P_X\left(x\right)$ is in $f^{-1}\left(i^*\right)$ for some (fixed) $i^*\in\mathcal{Z}$.

Lemma 12: For $V_o:\mathcal{Z}\to\mathcal{Z}$ in (18)*,*

(i) when $\rho_c<\rho\leq 1$, no two rows can be identical;

(ii) when $0\leq\rho\leq\rho_c$, if the rows $V_o(\cdot|j)$ and $V_o(\cdot|j')$, $j\neq j'$, are identical, then each coincides with the row $V_o(\cdot|i^*)$, in which case $P_X\left(x_j^*\right)=P_X\left(x_{j'}^*\right)=P_X\left(x^*\right)$. Furthermore, the number of identical rows of V_o cannot exceed $\left\lfloor\frac{1}{\rho_c}\right\rfloor$.

Proof: With $0\leq\rho\leq 1$, if the rows of $V_o:\mathcal{Z}\to\mathcal{Z}$ corresponding to $j\neq j'$ in $\mathcal{Z}$ are identical, then

$$V(i|j)=V(i|j'),\quad i\in\mathcal{Z}\setminus\{j,j'\}$$

i.e.,

$$\begin{aligned}&(1-\max\{\rho_c,\rho\})\frac{P_X\left(x_i^*\right)}{\sum_{l\neq j}P_X\left(x_l^*\right)}\\ &=(1-\max\{\rho_c,\rho\})\frac{P_X\left(x_i^*\right)}{\sum_{l'\neq j'}P_X\left(x_{l'}^*\right)},\quad i\in\mathcal{Z}\setminus\{j,j'\}\end{aligned}$$

whence

$$P_X\left(x_j^*\right)=P_X\left(x_{j'}^*\right); \tag{95}$$

and furthermore

$$V(i|j)=V(i|j'),\quad i\in\{j,j'\}$$

which, using (95), gives straightforwardly that

$$\max\{\rho_c,\rho\}=\frac{P_X\left(x_i^*\right)}{\sum_{l\in\mathcal{Z}}P_X\left(x_l^*\right)},\quad i\in\{j,j'\}. \tag{96}$$

(i) When $\rho>\rho_c$, recalling (16)

$$\frac{P_X\left(x_i^*\right)}{\sum_{l\in\mathcal{Z}}P_X\left(x_l^*\right)}\leq\frac{P_X\left(x^*\right)}{\sum_{l\in\mathcal{Z}}P_X\left(x_l^*\right)}=\rho_c<\max\{\rho_c,\rho\}$$

which violates (96) for $i\in\{j,j'\}$, so that no two rows of $V_o:\mathcal{Z}\to\mathcal{Z}$ can be identical.

(ii) When $0\leq\rho\leq\rho_c$, suppose that the rows $V_o(\cdot|j)$ and $V_o(\cdot|j')$ are identical for some $j\neq j'$. Then (96) holds which, upon recalling (16), is tantamount to

$$P_X\left(x_j^*\right)=P_X\left(x_{j'}^*\right)=P_X(x^*)=P_X(x_{i^*}^*). \tag{97}$$

To show for $j\neq i^*$ that $V_o(i|j)=V_o(i|i^*)$, $i\in\mathcal{Z}$, consider first $i\in\{j,i^*\}$. Then, using (97),

$$\begin{aligned}V_o(j|j)&=\rho_c,\\ V_o(i^*|j)&=(1-\rho_c)\frac{P_X\left(x_{i^*}^*\right)}{\sum_{l\neq j}P_X\left(x_l^*\right)}\\ &=(1-\rho_c)\frac{P_X\left(x^*\right)}{\sum_{l\in\mathcal{Z}}P_X\left(x_l^*\right)-P_X\left(x_j^*\right)}=\rho_c,\end{aligned}$$

and similarly,

$$\begin{aligned}V_o(j|i^*)&=(1-\rho_c)\frac{P_X\left(x_j^*\right)}{\sum_{l\in\mathcal{Z}}P_X\left(x_l^*\right)-P_X\left(x_{i^*}^*\right)}=\rho_c,\\ V_o(i^*|i^*)&=\rho_c.\end{aligned}$$

And for $i\in\mathcal{Z}\setminus\{j,i^*\}$,

$$\begin{aligned}V_o(i|j)&=(1-\rho_c)\frac{P_X\left(x_i^*\right)}{\sum_{l\in\mathcal{Z}}P_X\left(x_l^*\right)-P_X\left(x_j^*\right)}\\ &=\frac{P_X\left(x_i^*\right)}{\sum_{l\in\mathcal{Z}}P_X\left(x_l^*\right)}=V_o(i|i^*).\end{aligned}$$

Lastly, if the number of identical rows of $V_o : \mathcal{Z} \to \mathcal{Z}$ is α, then $\alpha P_X(x^*) \leq \sum_{l \in \mathcal{Z}} P_X(x_l^*)$, whence $\alpha \leq \left\lfloor \frac{1}{\rho_c} \right\rfloor$. ∎

For $S \subseteq \mathcal{Z}$, let

$$j_S = \arg\max_{l \in S} P_X(x_l^*), \quad R_S = (\mathcal{Z} \setminus S) \cup \{j_S\}, \quad (98)$$

where j_S and R_S need not be unique. Let $\tilde{f}_{R_S}(X)$ be a R_S-valued rv with pmf

$$P\left(\tilde{f}_{R_S}(X) = i\right) = \frac{P_X(x_i^*)}{\sum_{l \in R_S} P_X(x_l^*)}, \quad i \in R_S. \quad (99)$$

Consider the stochastic matrix $V_{R_S} : R_S \to \mathcal{Z}$ given by

$$V_{R_S} = \{V(i|j), i \in \mathcal{Z}, j \in R_S\} \quad (100)$$

and let $\left\{\tilde{Z}_t^{R_S}\right\}_{t=1}^n$ be conditionally mutually independent $\mathcal{Z}$-valued rvs conditioned on $\tilde{f}_{R_S}(X)$, with

$$P_{\tilde{Z}_t^{R_S} | \tilde{f}_{R_S}(X)} = V_{R_S}, \quad t = 1, \ldots, n. \quad (101)$$

Let $C(V_{R_S})$ be the Chernoff radius restricted to R_S, i.e., with the minimum in (68) being instead over all $j \neq j'$ in R_S.

Lemma 13: For $0 \leq \rho \leq 1$, consider add-noise ρ-QRs $\{F_t(X)\}_{t=1}^n$ with (32) *holding for every $n \geq 1$, where $V_t = V$, $t \geq 1$. If $V : \mathcal{Z} \to \mathcal{Z}$ has identical rows $\{V(\cdot|j), j \in S\}$, then*

$$\pi_\rho(V^n) - \left(1 - \sum_{i \in R_S} P_X(x_i^*)\right) \doteq \exp\left[-nC(V_{R_S})\right]$$

for R_S and V_{R_S} in (98) *and* (100), *respectively.*

Remark: If the rows of $V_{R_S} : R_S \to \mathcal{Z}$ are distinct in Lemma 13, then $C(V_{R_S}) > 0$. If the rows of $V : \mathcal{Z} \to \mathcal{Z}$ are distinct, then $S = \phi$, $R_S = \mathcal{Z}$ and $V_{R_S} = V$.

Proof:

$$\begin{aligned}
&P\left(g_{\mathrm{MAP}(V^n)}(Z_1, \ldots, Z_n) = X\right) \\
&= \sum_{(i_1,\ldots,i_n) \in \mathcal{Z}^n} \max_{x \in \mathcal{X}} P_X(x) \prod_{t=1}^n V(i_t | f(x)) \\
&= \sum_{(i_1,\ldots,i_n) \in \mathcal{Z}^n} \max_{\substack{x \in \cup f^{-1}(j) \\ j \in \mathcal{Z}}} P_X(x) \prod_{t=1}^n V(i_t | f(x)) \\
&= \sum_{(i_1,\ldots,i_n) \in \mathcal{Z}^n} \max_{j \in \mathcal{Z}} P_X\left(x_j^*\right) \prod_{t=1}^n V(i_t | j) \\
&= \sum_{(i_1,\ldots,i_n) \in \mathcal{Z}^n} \max_{j \in (\mathcal{Z} \setminus S) \cup S} P_X\left(x_j^*\right) \prod_{t=1}^n V(i_t | j) \\
&= \sum_{(i_1,\ldots,i_n) \in \mathcal{Z}^n} \max_{j \in R_S} P_X(x_j^*) \prod_{t=1}^n V(i_t | j) \qquad (102) \\
&= \left(\sum_{i \in R_S} P_X(x_i^*)\right) \\
&\quad \times \sum_{(i_1,\ldots,i_n) \in \mathcal{Z}^n} \max_{j \in R_S} \frac{P_X\left(x_j^*\right)}{\sum_{i \in R_S} P_X(x_i^*)} \prod_{t=1}^n V(i_t | j) \\
&= \left(\sum_{i \in R_S} P_X(x_i^*)\right) \\
&\quad \times P\left(g_{\mathrm{MAP}\left(V_{R_S}^n\right)}\left(\tilde{Z}_1^{R_S}, \ldots, \tilde{Z}_n^{R_S}\right) = \tilde{f}_{R_S}(X)\right) \qquad (103)
\end{aligned}$$

where (102) is by the identicality of the rows $\{V(\cdot|j), j \in S\}$, and $\tilde{f}_{R_S}(X)$ and $\{\tilde{Z}_t^{R_S}\}_{t=1}^n$ are as in (99) and (101), respectively. The assertion follows by applying [24, Th. 2] to (103). ∎

Turning to the proof of Proposition 11(i), first observe by Theorem 6 that for $0 \leq \rho \leq 0.5$ and *every* $n \geq 1$,

$$\begin{aligned}
&\pi_\rho(V_2^n) - \left(1 - \sum_{i \in \mathcal{Z}} P_X(x_i^*)\right) \\
&\geq \sum_{i \in \mathcal{Z}} P_X(x_i^*) \\
&\quad - \max\left\{P_X(x^*), \sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho} \rfloor} \right\rfloor - \mathbb{1}\left(k \bmod \lfloor \frac{1}{\rho} \rfloor = 0\right)} P_X\left(x^*_{i\lfloor \frac{1}{\rho} \rfloor}\right)\right\} \\
&> 0. \qquad (104)
\end{aligned}$$

We consider two cases: $\rho > \rho_c$ and $0 \leq \rho \leq \rho_c$.

When $\rho > \rho_c$, by Lemma 12(i), all the rows of $V_o : \mathcal{Z} \to \mathcal{Z}$ are distinct so that $C(V_o) > 0$. Then, by (84),

$$\lim_n \pi_\rho(V_o^n) - \left(1 - \sum_{i \in \mathcal{Z}} P_X(x_i^*)\right) = 0$$

which upon comparison with (104), yields (80) in this case.

In the case $0 \leq \rho \leq \rho_c$, $V_o : \mathcal{Z} \to \mathcal{Z}$ can contain identical rows. By Lemma 12(ii) and upon invoking assumption (45) without loss of generality, the identical rows must be those corresponding to $\{0, 1, \ldots, a-1\}$ (with the remaining rows being all distinct), where $a \leq \left\lfloor \frac{1}{\rho_c} \right\rfloor$ is the number of identical rows. By applying Lemma 13, with $S = \{0, 1, \ldots, a-1\}$, $j_S = 0$ and observing that $C\left((V_o)_{R_S}\right) > 0$, we get

$$\begin{aligned}
&\lim_n \pi_\rho(V_o^n) \\
&= 1 - \sum_{i \in R_S} P_X(x_i^*) \\
&= \begin{cases} 1 - P_X(x^*), & a = k \Leftrightarrow \left\lfloor \frac{1}{\rho_c} \right\rfloor = k \\ 1 - \sum_{i \in \{0, a, a+1, \ldots, k-1\}} P_X(x_i^*), & a < k \Leftrightarrow \left\lfloor \frac{1}{\rho_c} \right\rfloor < k \end{cases} \\
&\leq \begin{cases} 1 - P_X(x^*), & \left\lfloor \frac{1}{\rho_c} \right\rfloor = k \\ 1 - \sum_{i \in \left\{0, \lfloor \frac{1}{\rho_c} \rfloor, \lfloor \frac{1}{\rho_c} \rfloor + 1, \ldots, k-1\right\}} P_X(x_i^*), & \left\lfloor \frac{1}{\rho_c} \right\rfloor < k \end{cases} \\
&\leq \begin{cases} 1 - P_X(x^*), & \rho_c = \frac{1}{k} \\ 1 - \sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho_c} \rfloor} \right\rfloor} P_X\left(x^*_{i\lfloor \frac{1}{\rho_c} \rfloor}\right), & k \bmod \left\lfloor \frac{1}{\rho_c} \right\rfloor \neq 0, \ \frac{1}{k} < \rho_c \leq 0.5 \\ 1 - \sum_{i=0}^{\left\lfloor \frac{k}{\lfloor \frac{1}{\rho_c} \rfloor} \right\rfloor - 1} P_X\left(x^*_{i\lfloor \frac{1}{\rho_c} \rfloor}\right), & k \bmod \left\lfloor \frac{1}{\rho_c} \right\rfloor = 0, \ \frac{1}{k} < \rho_c \leq 0.5 \end{cases} \\
&= \pi_{\rho_c}(V_2^n) \\
&\leq \pi_\rho(V_2^n), \quad 0 \leq \rho \leq \rho_c,
\end{aligned}$$

which, upon recalling by Theorem 6 that $\pi_\rho\left(V_2^n\right)$ is the same for all n, establishes (80). ■

Acknowledgments

The authors are indebted to: Himanshu Tyagi and Shun Watanabe for educating us in private function computation, helping formulate the models in this paper, and for numerous beneficial discussions that informed our approach; Adam Smith for raising the question of predicate privacy [34]; and all three anonymous reviewers for examining our work under a microscope and suggesting several improvements.

References

[1] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, May 2000, pp. 439–450.
[2] A. Nageswaran and P. Narayan, "Data privacy for a ρ-recoverable function," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Feb. 2018, pp. 1046–1050.
[3] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder. (2018) "Estimation efficiency under privacy constraints." [Online]. Available: https://arxiv.org/abs/1707.02409
[4] R. Bassily, A. Groce, J. Katz, and A. Smith, "Coupled-worlds privacy: Exploiting adversarial uncertainty in statistical data privacy," in *Proc. IEEE 54th Annu. Symp. Found. Comput. Sci.*, Oct. 2013, pp. 439–448.
[5] R. Bassily and A. Smith, "Local, private, efficient protocols for succinct histograms," in *Proc. 47th Annu. ACM Symp. Theory Comput.*, Jun. 2015, pp. 127–135.
[6] A. Blum, C. Dwork, F. McSherry, and K. Nissim, "Practical privacy: The SuLQ framework," in *Proc. 24th ACM SIGMOD-SIGACT-SIGART Symp. Princ. Database Syst.*, Jun. 2005, pp. 128–138.
[7] A. Blum, K. Ligett, and A. Roth, "A learning theory approach to non-interactive database privacy," in *Proc. 40th ACM Symp. Theory Comput.*, May 2008, pp. 609–618.
[8] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2012, pp. 1401–1408.
[9] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, vol. 23, no. 4, pp. 493–507, 1952.
[10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[11] J. C. Duchi, M. J. Wainwright, and M. Jordan. (2016). "Minimax optimal procedures for locally private estimation." [Online]. Available: https://arxiv.org/abs/1604.02390
[12] C. Dwork, "Differential privacy," in *International Colloquium on Automata, Languages, and Programming*, Springer, 2006, pp. 1–12.
[13] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory Cryptogr. Conf.* Springer, 2006, pp. 265–284.
[14] T. van Erven and P. Harremös, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.
[15] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving mining of association rules," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 217–228, Jul. 2002.
[16] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," in *Proc. 22nd ACM SIGMOD-SIGACT-SIGART Symp. Princ. Database Syst.*, Jun. 2003, pp. 211–222.
[17] Q. Geng and P. Viswanath, "The optimal noise-adding mechanism in differential privacy," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 925–951, Feb. 2016.
[18] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[19] M. Hardt and K. Talwar, "On the geometry of differential privacy," in *Proc. 42nd ACM Symp. Theory Comput.*, Jun. 2010, pp. 705–714.
[20] C. Huang, P. Kairouz, X. Chen, L. Sankar, and R. Rajagopal, "Context-aware generative adversarial privacy," *Entropy*, vol. 19, no. 12, p. 565, Dec. 2017.
[21] I. Issa, S. Kamath, and A. B. Wagner, "An operational measure of information leakage," in *Proc. Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2016, pp. 520–524.
[22] P. Kairouz, K. Bonawitz, and D. Ramage, "Discrete distribution estimation under local privacy," in *Proc. 33rd Int. Conf. Mach. Learn.*, vol. 48, Jun. 2016, pp. 2436–2444.
[23] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 4037–4049, Jun. 2017.
[24] F. Kanaya and T. S. Han, "The asymptotics of posterior entropy and error probability for Bayesian estimation," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1988–1992, Nov. 1995.
[25] S. P. Kasiviswanathan and A. Smith, "On the 'Semantics' of differential privacy: A Bayesian formulation," *J. Privacy Confidentiality*, vol. 6, no. 1, pp. 1–16, Aug. 2014.
[26] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *Proc. IEEE 23rd Int. Conf. Data Eng.*, Apr. 2007, pp. 106–115.
[27] J. Liao, O. Kosut, L. Sankar, and F. P. Calmon. (2018). "Privacy under hard distortion constraints." [Online]. Available: https://arxiv.org/abs/1806.00063
[28] A. Makhdoumi and N. Fawaz, "Privacy-utility tradeoff under statistical uncertainty," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2013, pp. 1627–1634.
[29] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proc. 48th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Oct. 2007, pp. 94–103.
[30] D. Rebollo-Monedero, J. Forné and J. Domingo-Ferrer, "From t-closeness-like privacy to postrandomization via information theory," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 11, pp. 1623–1636, Nov. 2010.
[31] A. Rényi, "On measures of entropy and information," in *Proc. Berkeley Symp. Math. Statist. Probab.*, vol. 1, Jan. 1961, pp. 547–561.
[32] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 838–852, Jun. 2013.
[33] A. Smith, "Privacy-preserving statistical estimation with optimal convergence rates," in *Proc. 43rd Annu. ACM Symp. Theory Comput.*, Jun. 2011, pp. X813–822.
[34] A. Smith, personal communication, Mar. 2018.
[35] L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 10, no. 5, pp. 557–570, Oct. 2002.
[36] L. Wasserman and S. Zhou, "A statistical framework for differential privacy," *J. Amer. Statistical Assoc.*, vol. 105, no. 489, pp. 375–389, Jan. 2010.
[37] M. Ye and A. Barg. (2017). "Optimal schemes for discrete distribution estimation under locally differential privacy." [Online]. Available: https://arxiv.org/abs/1702.00610

Ajaykrishnan Nageswaran received the Bachelor of Technology degree in Electronics and Communication Engineering from the National Institute of Technology Karnataka, India, in 2015. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering at the University of Maryland, College Park.

Prakash Narayan received the Bachelor of Technology degree in Electrical Engineering from the Indian Institute of Technology, Madras in 1976. He received the M.S. degree in Systems Science and Mathematics in 1978 and the D.Sc. degree in Electrical Engineering in 1981, both from Washington University, St. Louis, MO.

He is Professor of Electrical and Computer Engineering at the University of Maryland, College Park, with a joint appointment at the Institute for Systems Research. His research interests are in multiuser information theory, communication theory, communication networks, cryptography, and applied probability and statistics.