Math. Struct. in Comp. Science (2010), *vol.* 20, *pp.* 977–993. © Cambridge University Press 2010
doi:10.1017/S0960129510000289

A separation between divergence and Holevo information for ensembles

RAHUL JAIN[†], ASHWIN NAYAK[‡] and YI SU[§]

[†] *Centre for Quantum Technologies, and Department of Computer Science,*
National University of Singapore, Block S15, 3 Science Drive 2, Singapore 11754
Email: rahul@comp.nus.edu.sg
[‡] *Department of Combinatorics and Optimization, and Institute for Quantum Computing,*
University of Waterloo, and Perimeter Institute for Theoretical Physics.
200 University Ave. W., Waterloo, ON N2L 3G1, Canada
Email: ashwin.nayak@uwaterloo.ca
[§] *Department of Mathematics, University of Michigan*
2074 East Hall, 530 Church Street, Ann Arbor, MI 48109-1043, U.S.A.
Email: yisu@umich.edu

Received 22 May 2008; revised 3 June 2010

The notion of *divergence information* of an ensemble of probability distributions was introduced by Jain, Radhakrishnan and Sen in Jain *et al.* (2002; 2009) in the context of the 'substate theorem'. Since then, divergence has been recognised as a more natural measure of information in several situations in both quantum and classical communication.
We construct ensembles of probability distributions for which divergence information may be significantly smaller than the more standard Holevo information. As a result, we establish that bounds previously shown for Holevo information are weaker than similar ones shown for divergence information.

1. Introduction

In this article, we study the relationship between two different measures of information contained in an ensemble of probability distributions. The first measure, *Holevo information*, is a standard notion from information theory, and is equivalent to the notion of *mutual information* between two random variables. Consider jointly distributed random variables XY, with X taking values in a sample space $\mathcal{X}$. Consider the ensemble of distributions $\mathcal{E} = \{(\lambda_i, Y_i) \; : \; i \in \mathcal{X}\}$, where $\lambda_i = \Pr(X = i)$ and $Y_i = Y|(X = i)$, which is obtained by conditioning on values assumed by X. The Holevo information of the ensemble is given by $\chi(\mathcal{E}) = \mathrm{I}(X : Y) = \mathbb{E}_{i \sim X} \mathrm{S}(Y_i \| Y)$, where $\mathrm{S}(\cdot\|\cdot)$ measures the *relative*

[†] This work was conducted while Rahul Jain was at the Institute for Quantum Computing, University of Waterloo, where he was supported in part by an ARO/NSA USA grant.
[‡] Ashwin Nayak's research work was supported in part by NSERC Canada, CIFAR, MITACS, QuantumWorks, and an Ontario ERA. His research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI.
[§] Yi Su was supported by the Department of Mathematics, University of Michigan, and this research was carried out while he was at the University of Waterloo, and supported in part by an NSERC Canada Undergraduate Research Award.

entropy of a random variable (equivalently, distribution) with respect to another. This notion may be extended to ensembles of quantum states (see, for example, Nielsen and Chuang (2000)), and the term 'Holevo information' is derived from the literature in quantum information theory.

The second measure, the *divergence information*, was originally introduced in Jain *et al.* (2002; 2009) and arises in the study of relative entropy, and its connection with a 'substate property'. The *observational divergence* (or simply *divergence*) of two classical distributions P, Q on the same finite sample space is $\max_E P(E) \log_2(P(E)/Q(E))$, where E ranges over all events. We may view this as a (scaled) measure of the factor by which P may exceed Q for an event of interest. The notion of *divergence information* is derived from this as $\mathrm{D}(\mathscr{E}) = \mathbb{E}_{i \sim X} \mathrm{D}(Y_i \| Y)$, in analogy with Holevo information. A quantum generalisation of this measure may also be defined (Jain *et al.* 2009).

Relative entropy and Holevo (or mutual) information have been studied extensively in communication theory and beyond (see, for example, Cover and Thomas (1991)) since they arise in a variety of applications. Since the discovery of the substate theorem (Jain *et al.* 2002), divergence has been recognised as a more natural measure of information in a growing number of applications (Jain *et al.* 2009, Section 1). The applications include privacy trade-offs in communication protocols for computing relations (Jain *et al.* 2005), message compression (Jain *et al.* 2005), bit-string commitment (Jain 2008) and the communication complexity of remote state preparation (Jain 2006). In particular, divergence captures, up to a constant factor, the substate property for probability distributions. It thus becomes relevant in every application where the substate theorem is used.

We begin by constructing ensembles of probability distributions (equivalently, jointly distributed random variables) for which the Holevo and divergence information are quantitatively different.

Theorem 1.1. For every positive integer N, and real number $k \geqslant 1$ such that $N > 2^{36k^2}$, there is an ensemble $\mathscr{E}$ of distributions over a sample space of size N such that $\mathrm{D}(\mathscr{E}) = k$ and $\chi(\mathscr{E}) \in \Theta(k \log \log N)$.

A more precise statement of this theorem (Theorem 3.1) and related results are given in Section 3.

The ensembles we construct satisfy the property that the ensemble average (that is, the distribution of the random variable Y in the description above) is uniform. We show that the above separation is essentially the best possible whenever the ensemble average is uniform (Theorem 3.5). The result also applies to ensembles of quantum states, where the ensemble average is the completely mixed state (Theorem 3.6). We leave open the possibility of larger separations for classical or quantum ensembles with non-uniform averages.

The difference between the two measures demonstrated by Theorem 1.1 shows that in certain applications, divergence is quantitatively a more relevant measure of information. In Appendix A, we describe three applications where functionally similar bounds have been established in terms of both measures. This article shows that the bounds in terms of divergence information are, in fact, stronger.

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

In earlier work on the subject, Jain *et al.* (2009, Appendix A) compared relative entropy and divergence for both classical and quantum states. For pairs of distributions P, Q over a sample space of size N, they showed that $\mathrm{D}(P \| Q) \leqslant \mathrm{S}(P \| Q) + 1$, and $\mathrm{S}(P \| Q) \leqslant \mathrm{D}(P \| Q) \cdot (N - 1)$. This extends to the corresponding measures of information in an ensemble: $\mathrm{D}(\mathscr{E}) \leqslant \chi(\mathscr{E}) + 1$ and $\chi(\mathscr{E}) \leqslant \mathrm{D}(\mathscr{E}) \cdot (N - 1)$. They also showed qualitatively similar relations for ensembles of quantum states. In addition, they constructed a pair of distributions P, Q such that $\mathrm{S}(P \| Q) = \Omega(\mathrm{D}(P \| Q) \cdot N)$. However, they did not translate their construction to a similar separation for *ensembles* of probability distributions. Our work in this paper fills this gap for ensembles (of classical or quantum states) with a uniform average.

2. Preliminaries

In this section, we summarise our notation and the information-theoretic concepts we will use later in the paper – see Cover and Thomas (1991) for a deeper treatment of (classical) information theory. While the bulk of this article refers to classical information theory, as mentioned in Section 1, it was motivated by studies in (and has implications for) quantum information – see Nielsen and Chuang (2000) for an introduction to quantum information.

For a positive integer N, let $[N]$ represent the set $\{1, \ldots, N\}$. We view probability distributions over $[N]$ as vectors in $\mathbb{R}^N$. The probability assigned by distribution P to a sample point $i \in [N]$ is denoted by p_i (that is, with the same letter in small case). We use $P^\downarrow$ to denote the distribution obtained from P by composing it with a permutation π on $[N]$ so $p_i^\downarrow = p_{\pi(i)}$ and $p_1^\downarrow \geqslant p_2^\downarrow \geqslant \cdots \geqslant p_N^\downarrow$. For an event $E \subseteq [N]$, we use $P(E) = \sum_{i \in E} p_i$ to denote the probability of that event. We use U_N to denote the uniform distribution over $[N]$. The expected value of a function $f : [N] \to \mathbb{R}$ with respect to the distribution P over $[N]$ is abbreviated to $\mathbb{E}_P f$.

We will appeal to the *majorisation* relation for some of our arguments. The relation tells us which of two given distributions is 'more random'.

Definition 2.1 (Majorisation). Let P, Q be distributions over $[N]$. We say that P *majorises* Q, denoted by $P \succeq Q$, if

$$\sum_{j=1}^{i} p_j^\downarrow \geqslant \sum_{j=1}^{i} q_j^\downarrow$$

for all $i \in [N]$.

The following fact is straightforward.

Fact 2.1. Any probability distribution P on $[N]$ majorises U_N, the uniform distribution over $[N]$.

Throughout this paper, we will use 'log' to denote the logarithm with base 2 and 'ln' to denote the logarithm with base e.

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

Definition 2.2 (Entropy and relative entropy). Let P, Q be probability distributions on $[N]$. The *entropy* of P is defined as $\mathrm{H}(P) \overset{\text{def}}{=} -\sum_{i=1}^{N} p_i \log p_i$. The *relative entropy* between P and Q, denoted $\mathrm{S}(P \| Q)$, is defined by

$$\mathrm{S}(P \| Q) \overset{\text{def}}{=} \sum_{i=1}^{N} p_i \log \frac{p_i}{q_i}.$$

Note that the relative entropy with respect to the uniform distribution is connected to entropy since $\mathrm{S}(P \| \mathrm{U}_N) = \log N - \mathrm{H}(P)$.

We can formalise the connection between majorisation and randomness through the following fact.

Fact 2.2. If P, Q are distributions over $[N]$ such that P majorises Q, that is, $P \succeq Q$, then $\mathrm{H}(P) \leqslant \mathrm{H}(Q)$.

The notion of *observational divergence* was defined in Jain *et al.* (2002) in the context of the 'substate theorem'.

Definition 2.3 (Observational divergence). Let P, Q be probability distributions on $[N]$. Then the *observational divergence* between them, denoted $\mathrm{D}(P \| Q)$, is defined by

$$\mathrm{D}(P \| Q) \overset{\text{def}}{=} \max_{f:[N] \to [0,1]} (\mathbb{E}_P f) \log \frac{\mathbb{E}_P f}{\mathbb{E}_Q f}.$$

Note that we allow the quantity to take the value $+\infty$. Throughout this paper we will simply refer to 'observational divergence' as 'divergence'.

The divergence $\mathrm{D}(P \| Q)$ is always non-negative, and it is finite precisely when the support of P is contained in the support of Q (Jain *et al.* 2002). Due to convexity, the divergence between two distributions is attained by the characteristic function of an event.

Lemma 2.3.

$$\mathrm{D}(P \| Q) = \max_{E \subseteq [N]} P(E) \log \frac{P(E)}{Q(E)}.$$

Proof. Let $\mathscr{F}$ denote the (convex) set of functions from $[N]$ to $[0, 1]$. The extreme points of $\mathscr{F}$ are precisely the characteristic functions of events in $[N]$. For an extreme point, say the characteristic function f_E of the event $E \subseteq [N]$, we have $\mathbb{E}_P f_E = P(E)$.

If the divergence is $+\infty$, there is an event for which the right-hand side also takes the value $+\infty$. So assume that the divergence is finite. In this case, the right-hand side is also finite since the support of P is contained in the support of Q. By restricting $f : [N] \to [0, 1]$ to characteristic functions of events, we see that $\mathrm{D}(P \| Q)$ is at least the expression on the right-hand side above.

For the inequality in the other direction, note that the function

$$g(x) = (ax + b) \log \left(\frac{ax + b}{cx + d} \right)$$

defined on $[0, 1]$ is convex in x, for any $a, b, c, d \in \mathbb{R}$ such that $ax + b \geqslant 0$ and $cx + d > 0$ when $x \in [0, 1]$. Therefore, the function $g(x)$ attains its maximum at either $x = 0$ or $x = 1$.

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

The convexity of $g(x)$ implies that for any $\alpha \in [0,1]$, and functions $f, f' \in \mathscr{F}$, we have

$$
\begin{aligned}
&(\mathbb{E}_P(\alpha f + (1-\alpha) f'))\ \log \frac{\mathbb{E}_P(\alpha f + (1-\alpha) f')}{\mathbb{E}_Q(\alpha f + (1-\alpha) f')} \\
&\quad = (\alpha(\mathbb{E}_P f - \mathbb{E}_P f') + \mathbb{E}_P f')\ \log \frac{\alpha(\mathbb{E}_P f - \mathbb{E}_P f') + \mathbb{E}_P f'}{\alpha(\mathbb{E}_Q f - \mathbb{E}_Q f') + \mathbb{E}_Q f'} \\
&\quad \leqslant \max \left\{ (\mathbb{E}_P f) \log \frac{\mathbb{E}_P f}{\mathbb{E}_Q f}, \quad (\mathbb{E}_P f') \log \frac{\mathbb{E}_P f'}{\mathbb{E}_Q f'} \right\}.
\end{aligned}
$$

So the divergence is attained at an extreme point of $\mathscr{F}$, which proves the claim. □

From now on, we will only use the equivalent definition of divergence given by this lemma.

The divergence of any distribution with respect to the uniform distribution is bounded.

Lemma 2.4. For any probability distribution P on $[N]$, we have $0 \leqslant \mathrm{D}(P \| \mathrm{U}_N) \leqslant \log N$.

Proof. Consider the event E that achieves the divergence between P and U_N. Without loss of generality, the event E is non-empty. Therefore

$$P(E) \geqslant \mathrm{U}_N(E) \geqslant 1/N,$$

and

$$0 \leqslant \mathrm{D}(P \| U_N) \leqslant P(E) \log(P(E)N) \leqslant \log N. \qquad \square$$

Note that we only need to maximise over N events to calculate the divergence with respect to the uniform distribution.

Lemma 2.5. For any probability distribution P on $[N]$ such that $P^{\downarrow} = P$, that is, $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_N$, we have

$$\mathrm{D}(P \| \mathrm{U}_N) = \max_{i \in [N]} P([i]) \log \frac{N \cdot P([i])}{i}.$$

Proof. By the definition of observational divergence, the right-hand side in the above is bounded by $\mathrm{D}(P \| \mathrm{U}_N)$. Note that for the inequality in the other direction, the probability $P(E)$ of any event E with size $n_E = |E|$ is bounded by $P([n_E])$, the probability of the first n_E elements in $[N]$. So we have

$$
\begin{aligned}
\mathrm{D}(P \| Q) &= \max_{E \subseteq [N]} P(E) \log \frac{N \cdot P(E)}{n_E} \\
&\leqslant \max_{E \subseteq [N]} P(E) \log \frac{N \cdot P([n_E])}{n_E} \\
&\leqslant \max_{E \subseteq [N]} P([n_E]) \log \frac{N \cdot P([n_E])}{n_E},
\end{aligned}
$$

since P majorises U_N (Fact 2.1) and

$$P([n_E]) \geqslant \tfrac{n_E}{N}.$$

This is equivalent to the right-hand side in the statement of the lemma. □

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

Definition 2.4 (Ensemble). An *ensemble* is a sequence of pairs $\{(\lambda_j, Q_j) \; : \; j \in [M]\}$ for some positive integer M, where $\Lambda = (\lambda_j) \in \mathbb{R}^M$ is a probability distribution on $[M]$ and Q_j are probability distributions over the same sample space $\mathscr{Y}$.

Definition 2.5 (Holevo information). The *Holevo information* of an ensemble $\mathscr{E} = \{(\lambda_j, Q_j) \; : \; j \in [M]\}$, denoted $\chi(\mathscr{E})$, is defined by

$$\chi(\mathscr{E}) \overset{\text{def}}{=} \sum_{j=1}^{M} \lambda_j \, \mathrm{S}(Q_j \| Q),$$

where $Q = \sum_{j=1}^{M} \lambda_j Q_j$ is the *ensemble average*.

Definition 2.6 (Divergence information). The *divergence information* of an ensemble $\mathscr{E}$ given by $\mathscr{E} = \{(\lambda_j, Q_j) \; : \; j \in [M]\}$, denoted $\mathrm{D}(\mathscr{E})$, is defined by

$$\mathrm{D}(\mathscr{E}) \overset{\text{def}}{=} \sum_{j=1}^{M} \lambda_j \, \mathrm{D}(Q_j \| Q),$$

where $Q = \sum_{j=1}^{M} \lambda_j Q_j$ is the *ensemble average*.

3. Divergence versus relative entropy

In this section we describe the construction of an ensemble for which there is a large separation between the divergence and Holevo information. The ensemble has the property that the ensemble average is uniform. As a by-product of our construction, we will also obtain a bound on the maximum possible separation for ensembles with a uniform average.

We begin with the construction of the ensemble. Let

$$f_{\mathrm{L}}(k, N) = k(\ln \log(kN) - \ln(6k) + 1) - \log(1 + k \ln 2) - 1 - \frac{1}{\ln 2}$$

on points in the positive orthant in $\mathbb{R}^2$ with $Nk > 1$.

Theorem 3.1. For every integer $N > 1$, and every positive real number k with

$$\tfrac{16}{N} \leqslant k < \log N,$$

there is an ensemble

$$\mathscr{E} = \left\{ (\tfrac{1}{N}, Q_i) \; : \; i \in [N] \right\}$$

with

$$\tfrac{1}{N} \sum_i Q_i = \mathrm{U}_N,$$

where U_N is the uniform distribution over $[N]$, and with $\mathrm{D}(\mathscr{E}) \leqslant k$ and

$$\chi(\mathscr{E}) \geqslant f_{\mathrm{L}}(k, N).$$

To construct the ensemble described in the above theorem, we first construct a probability distribution P on $[N]$ with observational divergence $\mathrm{D}(P \| \mathrm{U}_N) \leqslant k$ such that

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

its relative entropy $\mathrm{S}(P\,\|\,\mathrm{U}_N)$ is large in comparison with k. Let $f_{\mathrm{U}} = k(\ln\log(Nk) - \ln k + 1)$ be defined on points in the positive orthant of $\mathbb{R}^2$ with $kN > 1$.

Theorem 3.2. For every integer $N > 1$, and every positive real number k, with

$$\tfrac{16}{N} \leqslant k < \log N,$$

there is a probability distribution P with $\mathrm{D}(P\,\|\,\mathrm{U}_N) = k$, and

$$f_{\mathrm{L}}(k,N) \leqslant \mathrm{S}(P\,\|\,\mathrm{U}_N) \leqslant f_{\mathrm{U}}(k,N).$$

The construction of the ensemble is now immediate.

Proof of Theorem 3.1. Let $Q_j = P \circ \pi_j$, where π_j is the cyclic permutation of $[N]$ by $j-1$ places. We endow the set of the N cyclic permutations $\{Q_j \,:\, j \in [N]\}$ of P with the uniform distribution. By construction, the ensemble average is U_N. Since both observational divergence and relative entropy with respect to the uniform distribution are invariant under permutations of the sample space, $\mathrm{D}(\mathscr{E}) = \mathrm{D}(P\,\|\,\mathrm{U}_N) \leqslant k$, and $\chi(\mathscr{E}) = \mathrm{S}(P\,\|\,\mathrm{U}_N) \geqslant f_{\mathrm{L}}(k,N)$. □

We turn to the construction of the distribution P. Our construction is such that $P^{\downarrow} = P$, that is, $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_N$. Lemma 2.5 tells us that we only need to ensure that

$$P([i])\log\frac{N\cdot P([i])}{i} \leqslant k, \quad \forall\, i \in [N], \tag{1}$$

to guarantee $\mathrm{D}(P\,\|\,Q) \leqslant k$. Since $\mathrm{S}(P\,\|\,\mathrm{U}_N) = \log N - \mathrm{H}(P)$, we wish to minimise the entropy of P subject to the constraints in Equation (1). This is equivalent to successively maximising $p_1, p_2, \ldots$, and motivates the following definitions.

Define the function $g(y,x) = y\log(Ny/x) - k$ on the positive orthant of $\mathbb{R}^2$. Consider the function $h : \mathbb{R}^+ \to \mathbb{R}^+$ defined implicitly by the equation $g(h(x),x) = 0$.

Lemma 3.3. The function $h \,:\, \mathbb{R}^+ \to \mathbb{R}^+$ is well defined, strictly increasing and concave.

Proof. Fix an $x \in \mathbb{R}^+$, and consider the function $g_x(y) = g(y,x)$. This function is continuous on $\mathbb{R}^+$, tends to $-k < 0$ as $y \to 0^+$, and tends to ∞ as $y \to \infty$. By the Intermediate Value Theorem, for some $y > 0$, we have $g_x(y) = 0$. Moreover, $g_x(y) < -k$ for $0 < y \leqslant x/N$, and is strictly increasing for $y > x/N\mathrm{e}$ (its derivative is $g'_x(y) = \log(\mathrm{e}Ny/x)$). Therefore, there is a unique y such that $g_x(y) = 0$ and $h(x)$ is well defined.

The function h satisfies the equation $h\log(Nh/x) = k$, and thus the identity

$$x = Nh\exp\left(-\tfrac{k\ln 2}{h}\right).$$

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

Differentiating with respect to h, we see that

$$\frac{dx}{dh} = N\left(1 + \frac{k\ln 2}{h}\right)\exp\left(-\tfrac{k\ln 2}{h}\right)$$
$$\frac{d^2x}{dh^2} = \frac{N(k\ln 2)^2}{h^3}\ \exp\left(-\tfrac{k\ln 2}{h}\right).$$

So $dh/dx > 0$ for all $x > 0$, and h is a strictly increasing function. Note that $d^2x/dh^2 > 0$ for all $h > 0$, so x is a convex function of h. Since h is an increasing function, convexity of $x(h)$ implies concavity of $h(x)$. □

Let $v_0 = 0$. For $i \in [N]$, let $v_i = h(i)$, that is, $v_i \log(Nv_i/i) = k$. Let $s_i \overset{\text{def}}{=} \min\{1, v_i\}$ for $i \in [N]$. Let $p_1 = s_1$, and $p_i = s_i - s_{i-1}$ for all $2 \leqslant i \leqslant N$. Lemma 3.3 guarantees that these numbers are well defined. We claim the following lemma.

Lemma 3.4. The vector $P = (p_i) \in \mathbb{R}^N$ defined above is a probability distribution, and $P^{\downarrow} = P$, that is, $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_N$.

Proof. By definition, we have $v_i > 0$ for all $i \in [N]$. Therefore $s_1 = \min\{1, v_1\} > 0$. Since $h(x)$ is an increasing function in x, the sequence (v_i) is also increasing, so (s_i) is non-decreasing. Therefore $p_i = s_i - s_{i-1} \geqslant 0$ for $i > 1$.

Now $v_N \log v_N = k > 0$. Since $x \log x \leqslant 0$ for $x \in (0, 1]$, we have $v_N > 1$. So $s_N = \min\{1, v_N\} = 1$. Therefore $\sum_{i=1}^N p_i = s_N = 1$. So P is a probability distribution on $[N]$.

Note that

$$(v_2/2)\log(Nv_2/2) = k/2 < k,$$

so $v_1 > v_2/2$. So $s_1 \geqslant s_2/2$, that is, $p_1 \geqslant p_2$. For $i \geqslant 2$, we have

$$p_i - p_{i+1} = (s_i - s_{i-1}) - (s_{i+1} - s_i) = 2s_i - s_{i-1} - s_{i+1}.$$

Since $h(x)$ is concave, the function $\min\{1, h(x)\}$ is also concave. Therefore, $s_i \geqslant (s_{i-1} + s_{i+1})/2$, and the sequence (p_i) is non-decreasing. □

The vector $S = (s_i) \in \mathbb{R}^N$, and thus represents the (cumulative) distribution function corresponding to P.

Proof of Theorem 3.2. We claim that the probability distribution P constructed above satisfies the properties stated in the theorem.

Since $P^{\downarrow} = P$, by Lemma 2.5, we only need to verify that $s_i \log(Ns_i/i) \leqslant k$ for $i \in [N]$. If $s_i = v_i$, the condition is satisfied with the equality. (Note that since $k < \log N$, we have $s_1 = v_1 < 1$.) Otherwise, $s_i = 1 < v_i$, so $s_i \log(Ns_i/i) < v_i \log(Nv_i/i) = k$.

We now bound the relative entropy $\mathrm{S}(P \,\|\, \mathrm{U}_N)$ from below. Let n be the smallest positive integer such that $v_{n-1} \leqslant 1$ and $v_n > 1$. Note that $n > 1$. We also have $n \leqslant N$ since $v_N > 1$ (as $v_N \log v_N = k > 0$). Therefore, we have $s_i = v_i$ (equivalently, $Ns_i = i2^{k/s_i}$) for $i \in [n-1]$

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

and $s_n = 1 < v_n$. Thus, for $1 < i < n$,

$$\begin{aligned} Np_i &= i2^{\frac{k}{s_i}} - (i-1)2^{\frac{k}{s_{i-1}}} \\ &= 2^{\frac{k}{s_i}} + (i-1)(2^{\frac{k}{s_i}} - 2^{\frac{k}{s_{i-1}}}) \\ &= 2^{\frac{k}{s_i}} + (i-1)2^{\frac{k}{s_{i-1}}}(2^{\frac{k}{s_i}-\frac{k}{s_{i-1}}} - 1) \\ &= 2^{\frac{k}{s_i}} + Ns_{i-1}(2^{\frac{k}{s_i}-\frac{k}{s_{i-1}}} - 1) \\ &\geqslant 2^{\frac{k}{s_i}} + Ns_{i-1}\left(\frac{k}{s_i} - \frac{k}{s_{i-1}}\right)\ln 2 \\ &= 2^{\frac{k}{s_i}} - \frac{Np_ik}{s_i}\ln 2. \end{aligned}$$

The penultimate line follows from the inequality $2^x \geqslant 1 + x\ln 2$ for all $x \in \mathbb{R}$. Thus we have

$$Np_i \geqslant \frac{2^{\frac{k}{s_i}}}{1 + \frac{k}{s_i}\ln 2}. \tag{2}$$

Since $Np_1 = Ns_1 = 2^{(k/s_1)}$, this also holds for $i = 1$.

We can now bound the relative entropy using Equation (2).

$$\begin{aligned} \mathrm{S}(P\,\|\,\mathrm{U}_N) &= \sum_{i=1}^{N} p_i \log Np_i \\ &= \sum_{i=1}^{n} p_i \log Np_i \\ &\geqslant \sum_{i=1}^{n-1} p_i \log \frac{2^{\frac{k}{s_i}}}{1 + \frac{k}{s_i}\ln 2} + p_n \log Np_n \\ &\geqslant \sum_{i=1}^{n-1} \frac{p_ik}{s_i} - \sum_{i=1}^{n-1} p_i \log\left(1 + \frac{k\ln 2}{s_i}\right) + p_n \log Np_n. \end{aligned} \tag{3}$$

We bound each of the three terms in the right-hand side of Equation (3) separately.

We start with

$$\sum_{i=1}^{n-1} \frac{p_ik}{s_i}.$$

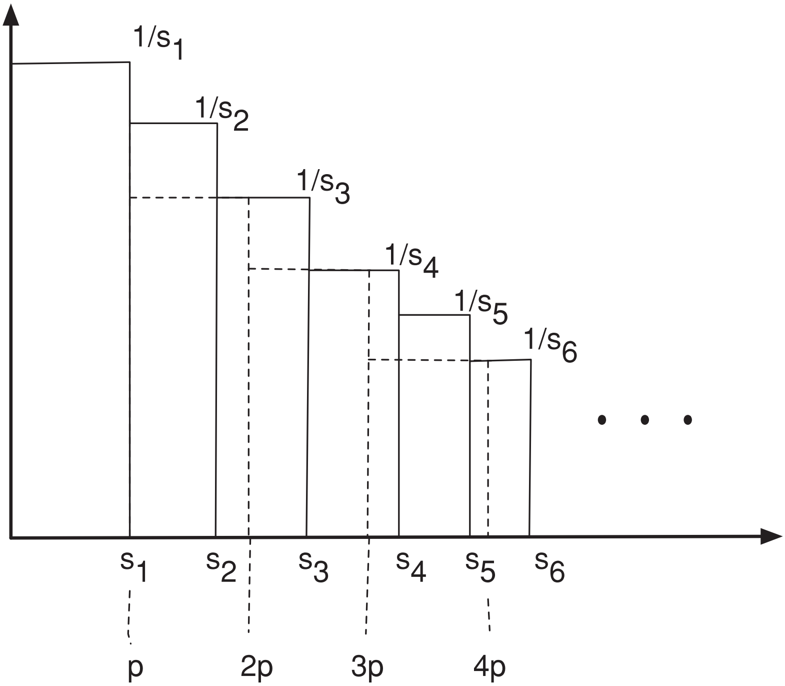
Let $p = p_1$, and let $m = \lfloor 1/p \rfloor$. For every $j \in [m]$, there is an $i \in [n]$, say $i = i_j$, such that $jp \leqslant s_{i_j} \leqslant (j+1)p$. (Otherwise, for some $i > 1$, the probability $p_i = s_i - s_{i-1}$ is strictly larger than p, an impossibility.)

We interpret the sum

$$\sum_{i=2}^{n-1} \frac{p_i}{s_i} = \sum_{i=2}^{n-1} \frac{s_i - s_{i-1}}{s_i}$$

as a Riemann sum approximating the area under the curve $1/x$ between s_1 and s_{n-1} by the area under the solid lines in Figure 3. This area is bounded from below by the area under the dashed lines, which corresponds to the area of rectangles of uniform width p

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press



and height $1/s_{j+1}$ for the jth interval. Thus,

$$\begin{aligned}\sum_{i=1}^{n-1} \frac{p_i k}{s_i} &\geqslant k + k \sum_{j=1}^{m} p \cdot \frac{1}{s_{i_{j+1}}} \\ &\geqslant k + k \sum_{j=1}^{m} p \cdot \frac{1}{(j+2)p} \\ &= k + k \sum_{j=1}^{m} \frac{1}{j+2} \\ &\geqslant k + k \int_{3}^{m+3} \frac{1}{x} dx \\ &= k + k \ln \frac{m+3}{3}. \end{aligned} \tag{4}$$

We lower bound $m = \lfloor (1/p) \rfloor$ next. Recall that $g_1(y) = y \log(Ny) - k$ is an increasing function for $y > 1/\mathrm{e}N$, and $p = p_1 \geqslant 1/N$. Consider the value of $g_1(y)$ at the point $q = 2k/\log kN$:

$$g_1(q) = \frac{2k}{\log kN} \log \frac{2Nk}{\log kN} - k > 2k \left(1 - \frac{\log \log kN}{\log kN}\right) - k \geqslant 0,$$

since $kN \geqslant 16$. As $g_1(q) > g_1(p) > 0$, we have $q > p$. Therefore,

$$m \geqslant \tfrac{1}{p} - 1 \geqslant \tfrac{\log kN}{2k} - 1.$$

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

Taken together with Equation (4), we get

$$\sum_{i=1}^{n-1} \frac{p_i k}{s_i} \geqslant k(\ln \log kN - \ln 6k + 1). \tag{5}$$

We next derive a lower bound for the second term in Equation (3).

$$\begin{aligned} -\sum_{i=1}^{n-1} p_i \log\left(1 + \frac{k \ln 2}{s_i}\right) &= -\sum_{i=1}^{n-1} p_i \log(s_i + k \ln 2) + \sum_{i=1}^{n-1} p_i \log s_i \\ &\geqslant -\log(1 + k \ln 2) + \sum_{i=1}^{n-1} p_i \log s_i. \end{aligned} \tag{6}$$

Viewing the second term above as a Riemann sum, we get

$$\begin{aligned} \sum_{i=1}^{n-1} p_i \log s_i &\geqslant \int_0^{s_{n-1}} \log x \, dx \\ &\geqslant \int_0^1 \log x \, dx \\ &= -\frac{1}{\ln 2}. \end{aligned} \tag{7}$$

Combining Equations (6) and (7), we get

$$-\sum_{i=1}^{n-1} p_i \log\left(1 + \frac{k \ln 2}{s_i}\right) \geqslant -\log(1 + k \ln 2) - \frac{1}{\ln 2}. \tag{8}$$

We bound the third term in Equation (3) crudely as $p_n \log N p_n \geqslant -1$. Along with the bounds for the previous two terms, Equations (5) and (8), this shows that

$$\mathrm{S}(P \,\|\, \mathrm{U}_N) \geqslant f_{\mathrm{L}}(k, N) \stackrel{\mathrm{def}}{=} k(\ln \log kN - \ln 6k + 1) - \log(1 + k \ln 2) - 1 - \frac{1}{\ln 2}. \tag{9}$$

This proves the lower bound on the relative entropy.

Moving to an upper bound, we have for $i \geqslant 2$,

$$\begin{aligned} N p_i &= i 2^{\frac{k}{s_i}} - (i-1) 2^{\frac{k}{s_{i-1}}} \\ &= 2^{\frac{k}{s_i}} + (i-1)\left(2^{\frac{k}{s_i}} - 2^{\frac{k}{s_{i-1}}}\right) \\ &\leqslant 2^{\frac{k}{s_i}}, \end{aligned}$$

since the second term is negative. This also holds for $i = 1$, since $p_1 = s_1$ and $s_1 \log N s_1 = k$.

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

Therefore,

$$\begin{aligned}
\mathrm{S}(P\|\mathrm{U}_N) &= \sum_{i=1}^{n} p_i \log N p_i \\
&\leqslant \sum_{i=1}^{n} \frac{k p_i}{s_i} \\
&\leqslant k + k\int_{s_1}^{1} \frac{1}{s} ds \\
&= k - k \ln s_1 \\
&\leqslant k + k \ln\left(\frac{\log Nk}{k}\right) \\
&= k(1 - \ln k + \ln(\log Nk)).
\end{aligned}$$

In the final inequality we have used the lower bound $s_1 \geqslant k/\log Nk$. □

The upper and lower bounds on the relative entropy of P with respect to the uniform distribution both behave as $k \log\log Nk$ up to constant factors.

Proof of Theorem 1.1. The dominating term in both the lower and upper bounds on the relative entropy $\mathrm{S}(P\|\mathrm{U}_N)$, with P as in Theorem 3.2, is $k \ln \log Nk$ when N is large compared with k. Specifically, when $N > 2^{36k^2}$, we have

$$\frac{1}{2} k \log\log Nk \leqslant \mathrm{S}(P\|\mathrm{U}_N) \leqslant 2k \log\log Nk.$$

By hypothesis, $1 \leqslant k$, and by Lemma 2.4, we have $k \leqslant \log N$. Thus,

$$\mathrm{S}(P\|\mathrm{U}_N) \in \Theta(\mathrm{D}(P\|\mathrm{U}_N) \log\log N).$$

The same holds for the ensembles constructed in Theorem 3.1. □

The separation we have demonstrated above is the best possible for ensembles of distributions that have a uniform average distribution.

Theorem 3.5. For any positive integer N, and any ensemble $\mathscr{E} = \{(\lambda_j, Q_j) : j \in [M]\}$ of distributions over $[N]$ such that $\sum_{j=1}^{M} \lambda_j Q_j = \mathrm{U}_N$, we have

$$\chi(\mathscr{E}) \leqslant K(2 \ln \log N - \ln K + 1) + 16,$$

where $K = \mathrm{D}(\mathscr{E})$.

Proof. Let $\mathrm{D}(Q_j\|\mathrm{U}_N) = k_j$. We will show that

$$\mathrm{S}(Q_j\|\mathrm{U}_N) \leqslant k_j(2 \ln \log N - \ln k_j + 1)$$

when $k_j \geqslant 16/N$. When $k_j < 16/N$, we have $\mathrm{S}(Q_j\|\mathrm{U}_N) < 16$. Since $k(2 \ln \log N - \ln k + 1)$ is a concave function in k, averaging over j with respect to the distribution $\Lambda = (\lambda_j)$ gives the claimed bound.

Fix a j such that $k_j > 16/N$. Let $R = Q_j^{\downarrow}$. Note that $\mathrm{D}(R\|\mathrm{U}_N) = k_j$ and $\mathrm{S}(R\|\mathrm{U}_N) = \mathrm{S}(Q_j\|\mathrm{U}_N)$. Consider the distribution P constructed as in Theorem 3.2 with $k = k_j$. Using

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

notation as in this construction, we have $s_i \log(Ns_i/i) = k_j$ for all $i < n$, and $s_n = 1$. Let $t_i = \sum_{l=1}^{i} r_l$, where $r_l \stackrel{\text{def}}{=} \Pr(R = l)$. By definition, we have

$$t_i \log(Nt_i/i) \leqslant k_j = s_i \log(Ns_i/i).$$

Since the function $g_i(y) = y \log(Ny/i)$ is strictly increasing for $y \geqslant i/N\mathrm{e}$, and $t_i \geqslant i/N$ (Fact 2.1), we have $t_i \leqslant s_i$ for $i < n$. Since $s_i = 1$ for $i \geqslant n$, we have $t_i \leqslant s_i$ for these i as well. In other words, $P \succeq R$. By Fact 2.2, we have $\mathrm{H}(P) \leqslant \mathrm{H}(R)$. This is equivalent to $\mathrm{S}(R\|\mathrm{U}_N) \leqslant \mathrm{S}(P\|\mathrm{U}_N)$. By Theorem 3.2,

$$\mathrm{S}(P\|\mathrm{U}_N) \leqslant k_j(\ln \log(Nk_j) - \ln k_j + 1).$$

Since $k_j \leqslant \log N$, this is at most $k_j(2 \ln \log N - \ln k_j + 1)$. □

Finally, note that this is also the best separation possible for an ensemble of quantum states with a completely mixed ensemble average.

Theorem 3.6. For any positive integer N, and any ensemble

$$\mathscr{E} = \{(\lambda_j, \rho_j) \;:\; j \in [M]\}$$

of quantum states ρ_j over a Hilbert space of dimension N such that

$$\sum_{j=1}^{M} \lambda_j \rho_j = \tfrac{\mathrm{I}}{N},$$

the completely mixed state of dimension N, we have

$$\chi(\mathscr{E}) \leqslant K(2 \ln \log N - \ln K + 1) + 16,$$

where $K = \mathrm{D}(\mathscr{E})$.

Proof. Let Q_j be the probability distribution on $[N]$ corresponding to the eigenvalues of ρ_j. By the definition of observational divergence for quantum states,

$$\mathrm{D}(Q_j\|\mathrm{U}_N) \leqslant \mathrm{D}(\rho_j\|\tfrac{\mathrm{I}}{N}).$$

Furthermore, we have

$$\mathrm{S}(\rho_j\|\tfrac{\mathrm{I}}{N}) = \mathrm{S}(Q_j\|\mathrm{U}_N).$$

We now apply the same reasoning as in the proof of Theorem 3.5, and then note that the divergence of the ensemble $\{(\lambda_j, Q_j) \;:\; j \in [M]\}$ is bounded by $\mathrm{D}(\mathscr{E})$ and that the right-hand side in the statement is a non-decreasing function of K. This gives us the stated bound. (Note that we do not need $\sum_{j=1}^{M} \lambda_j Q_j = \mathrm{U}_N$ to use the reasoning in Theorem 3.5.) □

Appendix A. Implications for quantum protocols

A.1. *Quantum string commitment*

A *string commitment* scheme is an extension of the well-studied and powerful cryptographic primitive of *bit commitment*. In such schemes, one party, Alice, wishes to commit an entire

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

string $x \in \{0,1\}^n$ to another party, Bob. The protocol is required to be such that Bob cannot identify the string until it is revealed by Alice. In turn, Alice should not be able to renege on her commitment at the time of revelation. Formally, quantum string commitment protocols are defined as follows (Buhrman *et al.* 2006; Jain 2008).

Definition A.1 (Quantum string commitment (QSC)). Let $P = \{p_x : x \in \{0,1\}^n\}$ be a probability distribution and B be a measure of information contained in an ensemble of quantum states. An (n, a, b)-B-**QSC** protocol for P is a quantum communication protocol between two parties, Alice and Bob. Alice gets an input $x \in \{0,1\}^n$ chosen according to the distribution P. The starting joint state of the qubits of Alice and Bob is some pure state independent of x. The protocol runs in two phases: the commit phase, followed by the reveal phase. There are no intermediate measurements during the protocol. At the end of the reveal phase, Bob measures his qubits according to a POVM $\{M_y : y \in \{0,1\}^n\} \cup \{I - \sum_y M_y\}$ to determine the value of the committed string by Alice or to detect cheating. The protocol satisfies the following properties.

1 **(Correctness)** Suppose Alice and Bob act honestly. Let ρ_x be the state of Bob's qubits at the end of the reveal phase of the protocol, when Alice gets input x. Then $(\forall x, y)$ Tr $M_y \rho_x = 1$ if and only if $y = x$, and 0 otherwise.

2 **(Concealing property)** Suppose Alice acts honestly, and Bob possibly cheats, that is, deviates from the protocol in his local operations. Let σ_x be the state of Bob's qubits after the commit phase when Alice gets input x. Then the B-information $B(\mathscr{E})$ of the ensemble $\mathscr{E} = \{p_x, \sigma_x\}$ is at most b. In particular, this also holds when both Alice and Bob follow the protocol honestly.

3 **(Binding property)** Suppose Bob acts honestly and Alice possibly cheats. Let $c \in \{0,1\}^n$ be a string in a special cheating register C that Alice keeps independent of the rest of the registers until the end of the commit phase. Let τ_c be the state of Bob's qubits at the end of the reveal phase when Alice has c in the cheating register. Let $q_c \stackrel{\text{def}}{=} \text{Tr } M_c \tau_c$. Then

$$\sum_{c \in \{0,1\}^n} p_c q_c \leqslant 2^{a-n}.$$

The idea behind the above definition is as follows. At the end of the reveal phase of an honest run of the protocol, Bob identifies x from ρ_x by performing the POVM measurement $\{M_y\}_y \cup \{I - \sum_y M_y\}$. He accepts the committed string to be x if and only if the observed outcome $y = x$; this happens with probability Tr $M_x \rho_x$. He declares that Alice is cheating if outcome $I - \sum_x M_x$ is observed. Thus, at the end of an honest run of the protocol, with probability 1, Bob accepts the committed string as being exactly Alice's input string. The concealing property ensures that the amount of B-information about x that a possibly cheating Bob gets is bounded by b. In *bit*-commitment protocols, the concealing property is quantified in terms of the probability with which Bob can guess Alice's bit. Here instead we use different notions of information contained in the corresponding ensemble. The binding property ensures that when a cheating Alice wishes to postpone committing to a string until after the commit phase, she succeeds in forcing an honest Bob to accept her choice with bounded probability (in expectation).

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

Strong string commitment, in which both parameters a, b above are required to be 0, is impossible for the same reason that *strong* bit-commitment protocols are impossible (Mayers 1997; Lo and Chau 1997). Weaker versions are nonetheless possible, and exhibit a trade-off between the concealing and binding properties. The trade-off between the parameters a and b has been studied by several researchers (Kent 2003; Buhrman *et al.* 2006; Jain 2008). Buhrman *et al.* (2006) studied this trade-off in both the scenario of a single execution of the protocol and in the asymptotic regime, with an unbounded number of parallel executions of the protocol. In the asymptotic scenario, they showed the following result in terms of Holevo information (which is denoted by χ).

Theorem A.1 (Buhrman *et al.* 2006). Let Π be an (n, a_1, b)-χ-**QSC** scheme. Let Π_m represent m independent, parallel executions of Π (so $\Pi_1 = \Pi$). Let a_m represent the binding parameter of Π_m and let $a \overset{\text{def}}{=} \lim_{m\to\infty} a_m/m$. Then, $a + b \geqslant n$.

Jain (2008) shows a similar trade-off result regarding **QSC**s, but in terms of the divergence information of an ensemble (denoted by D).

Theorem A.2 (Jain 2008). For a single execution of the protocol of an (n, a, b)-D-**QSC** scheme,

$$a + b + 8\sqrt{b+1} + 16 \geqslant n.$$

As we mentioned earlier, for any ensemble $\mathscr{E}$, the divergence information is bounded by the Holevo χ-information $\mathrm{D}(\mathscr{E}) \leqslant \chi(\mathscr{E}) + 1$. This immediately implies the following theorem.

Theorem A.3 (Jain 2008). For a single execution of the protocol of an (n, a, b)-χ-**QSC** scheme,

$$a + b + 8\sqrt{b+2} + 17 \geqslant n.$$

As Jain shows, this implies the asymptotic result in Buhrman *et al.* (2006 Theorem A.1).

The separation that we demonstrate between divergence and Holevo information (Theorem 1.1) shows that for some ensembles over n qubits, $\mathrm{D}(\mathscr{E})$ may be a $\log n$ factor larger than $\chi(\mathscr{E})$. For such ensembles, the binding–concealing trade-off of Theorem A.2 is stronger than that of Theorem A.1.

A.2. *Privacy trade-off for two-party protocols for relations*

Let us consider two-party protocols between Alice and Bob for computing a relation $f \subseteq \mathscr{X} \times \mathscr{Y} \times \mathscr{Z}$. The goal here is to find a $z \in \mathscr{Z}$ such that $(x, y, z) \in f$, when Alice and Bob are given $x \in \mathscr{X}$ and $y \in \mathscr{Y}$, respectively. Jain *et al.* (2002) studied the extent to which the two parties may solve f while keeping their respective inputs hidden from the other party. They showed the following result.

Result A.4 (Jain *et al.* 2005, informal statement). Let μ be a product distribution on $\mathscr{X} \times \mathscr{Y}$. Let $Q_{1/3}^{\mu, A\to B}(f)$ represent the one-way distributional complexity of f for a single communication from Alice to Bob and distributional error under μ at most $1/3$.

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

Let X and Y represent the random variables corresponding to the inputs to Alice and Bob, respectively. If there is a quantum communication protocol for f in which Bob *leaks* divergence information at most b about his input Y, then Alice leaks divergence information at least $\Omega(Q^{\mu,A\to B}_{1/3}(f)/2^{O(b)})$ about her input X. A similar statement also holds with the roles of Alice and Bob interchanged.

From the upper bound on the divergence information in terms of Holevo information, this immediately implies the following result.

Result A.5 (Jain *et al.* 2005, informal statement). Let μ be a product distribution on $\mathcal{X} \times \mathcal{Y}$. Let $Q^{\mu,A\to B}_{1/3}(f)$ represent the one-way distributional complexity of f for a single communication from Alice to Bob and distributional error under μ at most $1/3$. Let X and Y represent the random variables corresponding to the inputs to Alice and Bob, respectively. If there is a quantum communication protocol for f where Bob *leaks* Holevo information at most b about his input Y, then Alice leaks Holevo information at least $\Omega(Q^{\mu,A\to B}_{1/3}(f)/2^{O(b)})$ about her input X. A similar statement also holds with the roles of Alice and Bob interchanged.

It follows from Theorem 1.1 that Result A.4 is much stronger than Result A.5 when the ensemble arising in the protocol between Alice and Bob has divergence information much smaller than its Holevo information.

A.3. *Message compression*

Jain *et al.* (2005) showed the following message compression result.

Result A.6 (Jain *et al.* 2005, informal statement). Let $\mathcal{E} \stackrel{\text{def}}{=} \{p_i, \rho_i\}$ be an ensemble. Alice on getting i, with probability p_i, intends to transmit state ρ_i to Bob. They are willing to tolerate a small constant ε loss in fidelity during transmission. There is a one-way protocol $\mathcal{P}$ that uses prior entanglement between Alice and Bob, and compresses Alice's state ρ_i to a classical message with expected length of the order of $\mathrm{D}(\mathcal{E})$ bits long. Using this classical message and the shared entanglement, Bob can reconstruct a quantum state whose fidelity with ρ_i is $1-\varepsilon$.

This immediately gives us the following result in terms of Holevo information of $\mathcal{E}$ (using $\mathrm{D}(\mathcal{E}) \leqslant \chi(\mathcal{E}) + 1$).

Result A.7 (Jain *et al.* 2005, informal statement). Let $\mathcal{E} \stackrel{\text{def}}{=} \{p_i, \rho_i\}$ be an ensemble. Alice on getting i, with probability p_i, intends to transmit state ρ_i to Bob. They are willing to tolerate a small constant ε loss in fidelity during transmission. There is a one-way protocol $\mathcal{P}$ that uses prior entanglement between Alice and Bob, and compresses Alice's state ρ_i to a classical message with expected length of the order of $\chi(\mathcal{E})$ bits long. Using this classical message and the shared entanglement, Bob can reconstruct a quantum state whose fidelity with ρ_i is $1-\varepsilon$.

It follows from Theorem 1.1 that Result A.6 is much stronger than Result A.7 when the ensemble $\mathcal{E}$ has divergence information much smaller than its Holevo information.

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press

References

Buhrman, H., Christandl, M., Hayden, P., Lo, H.-K. and Wehner, S. (2006) Security of quantum bit string commitment depends on the information measure. *Phys. Rev. Lett.* **97** (25) (article 250501).

Cover, T. M. and Thomas, J. A. (1991) *Elements of Information Theory*, Wiley Series in Telecommunications, John Wiley & Sons.

Jain, R. (2006) Communication complexity of remote state preparation with entanglement. *Quantum Inf. Comput.* **6** (4-5) 461–464.

Jain, R. (2008) New binding-concealing trade-offs for quantum string commitment. *J. Cryptol.* **24** (4) 579–592.

Jain, R., Radhakrishnan, J. and Sen, P. (2002) Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In: *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, IEEE Computer Society Press 429–438.

Jain, R., Radhakrishnan, J. and Sen, P. (2005) Prior entanglement, message compression and privacy in quantum communication. In: *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, IEEE Computer Society Press 285–296.

Jain, R., Radhakrishnan, J. and Sen, P. (2009) A new information-theoretic property about quantum states with an application to privacy in quantum communication. *J. ACM* **56** (6) (article 33).

Kent, A. (2003) Quantum bit string commitment. *Phys. Rev. Lett.* **90** (article 237901).

Lo, H.-K. and Chau, H.-F. (1997) Is quantum bit commitment really possible? *Phys. Rev. Lett.* **78** 3410–3413.

Mayers, D. (1997) Unconditionally secure quantum bit commitment is impossible. *Phys. Rev. Lett.* **78** (17) 3414–3417.

Nielsen, M. A. and Chuang, I. L. (2000) *Quantum Computation and Quantum Information*, Cambridge University Press.

https://doi.org/10.1017/S0960129510000289 Published online by Cambridge University Press