The Annals of Statistics
2005, Vol. 33, No. 2, 959–976
DOI 10.1214/009053604000001219

ON THE CHERNOFF BOUND FOR EFFICIENCY OF QUANTUM HYPOTHESIS TESTING

BY VLADISLAV KARGIN

Cornerstone Research

The paper estimates the Chernoff rate for the efficiency of quantum hypothesis testing. For both joint and separate measurements, approximate bounds for the rate are given if both states are mixed, and exact expressions are derived if at least one of the states is pure. The efficiencies of tests with separate and joint measurements are compared. The results are illustrated by a test of quantum entanglement.

1. Introduction. In his preface to a book about integral geometry [Santaló (1976)] Mark Kac wrote: "... Probability Theory is measure theory with a 'soul' which [in the case of integral geometry] is provided not by Physics or by games of chance or by Economics but by the most ancient and noble of all mathematical disciplines, namely Geometry." In a sense, then, Quantum Statistics can be called probability theory with a "subconscious." The probability distributions, so important for classical statistics, are no longer the deepest foundation layer but only an outward manifestation of geometry in the Hilbert space of quantum states. This foundational change begs for a new look at the classical statistics results, and this paper contributes by reconsidering the Chernoff–Hoeffding results about hypothesis testing.

Why quantum statistics? Today, quantum states can be manufactured. For example, in one method [Cirac and Zoller (1995)] ions are placed in a trap created by electrostatic potential and radio-frequency oscillations. The ions then are cooled by laser emission and arranged on a line in the trap. After that, each individual ion can be accessed by laser pulses and their joint quantum state can be altered according to the researcher's wishes. This ability to build and manipulate quantum systems is changing our thinking about computation and information transmission. Suddenly, certain classic problems—the factorization of large integers, the search in an unstructured database, secure communication—are not as difficult as they used to be.

This conceptual change also affects statistics.

For example, how can a quantum state manufacturer check if states have been generated faithfully? We can anticipate the statistician's answer: Select a sample

Received September 2003; revised April 2004.
AMS 2000 subject classifications. 62P35, 62G10.
Key words and phrases. Quantum statistics, fidelity, quantum relative entropy, joint measurement, separate measurement, entanglement.

of the states and perform a statistical test. But now, besides designing the test, the statistician must play an additional role, the role of advisor on how to perform measurements of a sample of quantum states. Since in quantum mechanics both the measurement and the state determine the probability distribution of outcomes, the choice of measurement affects the properties of the statistical test.

Not all measurements are readily available. Sometimes it is possible to measure sample states jointly, as one large quantum state, and sometimes the states can only be measured separately and simultaneously. Yet another possibility is that the states must be measured separately and sequentially. Finally, sometimes the sample states can only be measured partially, for example, when each state represents several remote particles that cannot be measured jointly. Clearly, the efficiency of the optimal test will depend on which measurements are available.

For a single state the problem of quantum hypothesis testing was solved by Holevo (1976) and Helstrom (1976). Here I consider a different situation: when the researcher has access to several copies of the same state but may not be able to measure them jointly.

The problem of testing using a sample of states was considered in Helstrom (1976), Ogawa and Nagaoka (2000), Parthasarathy (2001) and Ogawa and Hayashi (2002). These authors considered only joint measurements and only the situation when one of the errors may go to zero arbitrarily slowly. In contrast, I consider a Bayesian version of the problem, in which the researcher aims to minimize a weighted average of both errors, and I consider both joint and separate measurements.

When joint measurements are available, the problem of testing using a sample can be solved by applying the Holevo–Helstrom result to the case of tensor powers of primary states. In this case, my main results provide useful bounds on both the expected error when the sample is finite and the rate of decline in error as the number of sample states grows. The bounds are given in terms of fidelity distance between quantum hypotheses. In addition, when one of the hypotheses specifies a pure state, I derive an explicit expression for the rate of error decline.

For the separate measurements, I concentrate mainly on the asymptotic case. If at least one of the hypotheses is pure, then the optimal separate measurement leads to the same exponential rate of error decline as the optimal joint measurement. The error can, however, be twice as large as the error of the optimal joint measurement. If both hypotheses are mixed, then there is a large window of possibility that a joint measurement can lead to a better exponential rate than the optimal separate measurement. The improvement in the exponential rate, however, cannot be made greater than a factor of 2.

This paper contributes only to the theory of quantum hypothesis testing. I do not touch on another rapidly growing area of research: quantum state estimation. One of the most surprising discoveries in this area is that for mixed states there is a large gap between the rate of convergence of estimates based on the best separate and joint measurements: see Gill and Massar (2000), Gill (2001) and

Ballester (2004). This discovery is in agreement with the possibility of better hypothesis testing by joint measurements suggested by the results of this paper. For additional information about recent progress in the area of quantum estimation see the excellent review article by Barndorff-Nielsen, Gill and Jupp (2003).

The rest of the paper is organized as follows. Section 2 gives some basic information about quantum states and measurements and formulates the problem of quantum hypothesis testing. Section 3 gives a short summary of the Chernoff–Hoeffding results about hypothesis testing. Sections 4 and 5 discuss joint and separate measurements, respectively. Section 6 presents an illustration, and Section 7 concludes.

2. Quantum hypothesis testing. States of quantum-mechanical objects—electrons, photons, atoms, molecules, and so on—are described by density matrices. A density matrix is a self-adjoint, nonnegative operator of a complex Hilbert space with a trace of 1. In this paper we will be concerned only with finite-dimensional Hilbert spaces, so the operator is indeed represented by a finite Hermitian matrix. A particular case is projectors on one-dimensional subspaces. They represent states that are called pure.

States are not directly observable: they can be measured but the outcome of a measurement is a random variable. In the case of a countable number of outcomes, every measurement can be represented by a set of nonnegative operators which are required to add up to the identity operator. Each operator in the set corresponds to a particular outcome of the measurement, and if the state is ρ and outcome i is represented by operator M_i, then the probability of the outcome is $\mathrm{tr}\{M_i\rho\}$. An important subclass is formed by projective measurements, in which the outcomes are represented by orthogonal projectors: $M_iM_j = \delta_{ij}M_i$, where δ_{ij} is the Kronecker delta function. [For a more complete review of the mathematical apparatus of modern quantum mechanics see papers by Gill (2001) and Barndorff-Nielsen, Gill and Jupp (2003), or the book by Peres (1995).]

We consider the following problem: a researcher is given a sample of N identical quantum states, which are either ρ_0 or ρ_1 with the prior probability $1/2$. He aims to minimize the average probability of making an incorrect decision about the state by devising a system of measurements and a decision rule. Can we safely assume that all measurements are available to the researcher? No.

While in some situations the researcher can make a joint measurement of the state that represents the total sample, in other situations he can do only separate measurements of each state in the sample. If the separate measurements are done independently of each other, then we will call them separate independent measurements. If the measurements can be done sequentially and the researcher adjusts the current measurement according to the results obtained in the previous measurements, then they are separate adaptive measurements. Sometimes the researcher can return to the states that he has already measured and measure them again using the information that he has already obtained. The class of these

measurements is often called separable measurements. See, for example, Bennett et al. (1999). Sometimes, the researcher is even more restricted. This happens, for example, if a sample quantum state consists of two spatially remote parts and the researcher can only measure them separately. This restriction may further decrease the efficiency of statistical inference.

This paper concentrates on two cases that are in a certain sense extreme. In one of them the researcher can make any joint measurement he wishes; in the other he can do only separate independent measurements. We are interested in knowing how this restriction affects the efficiency of hypothesis testing.

3. Classical Chernoff–Hoeffding bounds. This section reviews results by Chernoff (1952), Sanov (1957) and Hoeffding (1965) about asymptotic error rates in hypothesis testing. For details the reader can also consult the book by Cover and Thomas (1991).

Consider two multinomial distributions, P_1 and P_2, on a finite space $X = \{x_i\}$, $i = 1, \ldots, n$. Suppose a sample X_N of size N is drawn from one of these distributions and provided to a researcher, whose task is to guess the distribution. A nonrandomized decision rule is characterized by a pair of complementary subsets of the outcome space, A_N and A_N^c. If the sample belongs to A_N, hypothesis P_1 is accepted; otherwise, P_2 is accepted. The Bayesian probability of making an error is

$$R_N = \pi_{P_1} \Pr\{X_N \in A_N^c | P_1\} + \pi_{P_2} \Pr\{X_N \in A_N | P_2\}, \tag{3.1}$$

where π_{P_1} and π_{P_2} are prior probabilities of the distributions.

The Chernoff–Hoeffding theorem claims that the probability of error R_N in the optimal test declines exponentially and the best achievable rate of decline is

$$\frac{1}{N} \log R_N \sim D(P_{\lambda^*} || P_1), \tag{3.2}$$

where

$$P_\lambda(x_i) = \frac{P_1^\lambda(x_i) P_2^{1-\lambda}(x_i)}{\sum_{i=1}^n P_1^\lambda(x_i) P_2^{1-\lambda}(x_i)}, \tag{3.3}$$

$D(P_\lambda || P_1)$ is the Kullback–Leibler distance from P_1 to P_λ,

$$D(P_\lambda || P_1) = \sum_{i=1}^n P_\lambda(x_i) \ln \frac{P_1(x_i)}{P_\lambda(x_i)}, \tag{3.4}$$

and λ^* is chosen in such a way that

$$D(P_{\lambda^*} || P_2) = D(P_{\lambda^*} || P_1). \tag{3.5}$$

This rate is sometimes called the Chernoff information distance between distributions P_1 and P_2. We denote it as $D_c(P_1, P_2)$.

It is also possible [see Cover and Thomas (1991)] to derive another expression for the asymptotic probability of error, which is easier to calculate:

$$\frac{1}{N}\ln R_N \sim \min_{0\le\lambda\le 1} \log \sum_{i=1}^{n} P_1^{\lambda}(x_i)\, P_2^{1-\lambda}(x_i). \tag{3.6}$$

In quantum statistics the researcher has the ability to vary distributions over outcomes by choosing the measurement of the given sample of quantum states. In addition, his task is to test hypotheses not about the distributions over outcomes but about the states themselves. How does this affect the classical results?

4. Joint measurements.

4.1. *Generalities.* Joint measurement of all sample states is by definition a measurement of the tensor product of the sample states. Here is an example of a joint measurement [from Keyl and Werner (2001)] that cannot be reduced to separate measurements.

EXAMPLE 1. Let H denote the Hilbert space where the quantum state lives, let d be the dimension of this space, and let N be the size of the sample. Then the group $SU(d)$ acts on $H^{\otimes N}$ by acting naturally on each term in the tensor power. Consequently, $H^{\otimes N}$ can be decomposed as a direct sum of the subspaces Y_i invariant under this action, and projectors on these subspaces, P_{Y_i}, can be taken as elements of a joint measurement. Keyl and Werner use this measurement for estimating the individual state spectrum.

If joint measurements are allowed, then in effect we have the problem of testing two alternative hypotheses about a single—although huge—quantum state, the problem that was solved by Holevo and Helstrom [see, e.g., Holevo (2001)]. In our situation we only need to determine what additional implications follow from the special structure of the state.

Suppose the hypotheses about the quantum state are given by matrices ρ_0 and ρ_1 with prior probability of $1/2$, and the task is to find a measurement and a decision procedure that result in the lowest possible expected probability of error. Then according to the Holevo–Helstrom result, the optimal measurement can be represented by projectors on the eigenvectors of the operator $\rho_0 - \rho_1$. The optimal decision is given by the following rule: If the measurement outcome corresponds to an eigenvector with a positive eigenvalue, then ρ_0 is chosen; otherwise, ρ_1 is chosen. The expected error probability for the optimal measurement and decision rule is

$$R = \tfrac{1}{2}\bigl(1 - \tfrac{1}{2}\|\rho_0 - \rho_1\|_1\bigr), \tag{4.1}$$

where $\|\cdot\|_1$ denotes the sum of the absolute values of eigenvalues.

In our case the hypothetical states are tensor powers of the individual states, $\rho_0^{\otimes N}$ and $\rho_1^{\otimes N}$, where

$$\rho_i^{\otimes N} \equiv \underbrace{\rho_i \otimes \rho_i \otimes \cdots \otimes \rho_i}_{N}. \tag{4.2}$$

The corresponding minimal expected error is

$$R = \tfrac{1}{2}(1 - \tfrac{1}{2}\|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1). \tag{4.3}$$

How does the error decline as N grows to infinity? Can we answer this question by explicitly calculating the distribution of eigenvalues of $\rho_0^{\otimes N} - \rho_1^{\otimes N}$? This matrix has very large dimensionality so directly finding its eigenvectors and eigenvalues is a hard computational problem. One way to circumvent this difficulty is by calculating moments of the eigenvalue distribution.

Initial moments are indeed easy to calculate. Let us introduce notation for the moments:

$$\mu_n =: \int_0^1 t^n \, dF(t) = \frac{1}{d^N} \operatorname{tr}(\rho_0^{\otimes N} - \rho_1^{\otimes N})^n, \tag{4.4}$$

where $F(t)$ is the discrete probability distribution that puts equal probability weight on each eigenvalue. Then the following proposition holds.

PROPOSITION 1.

$$\mu_n = \frac{1}{d^N} \sum_{\{k_1,\ldots,k_n\}} (-1)^{\sum k_i} \big(\operatorname{tr}(\rho_{k_1} \cdots \rho_{k_n})\big)^N, \tag{4.5}$$

where $\{k_1, \ldots, k_n\}$ *run over the set of all* n*-sequences of* 0 *and* 1.

PROOF. The proposition follows from the noncommutative binomial expansion of $(\rho_0^{\otimes N} - \rho_1^{\otimes N})^n$ and the fact that $\operatorname{tr}(\rho_{k_1}^{\otimes N} \cdots \rho_{k_n}^{\otimes N}) = (\operatorname{tr}(\rho_{k_1} \cdots \rho_{k_n}))^N$. □

The advantage of this formula is that for a fixed n, the calculation is as easy for large as for small values of the sample size N. The difficulty is that the number of terms in this formula grows exponentially with moment size n. Therefore the standard map from the set of moment sequences to the set of distributions is impractical.

In the next sections we will pursue other approaches to estimating $\|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1$ based on consideration of important special cases and on construction of measurements that approximate the optimal measurement.

4.2. *Special cases.* To get more insight about the behavior of $\|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1$, it is useful to consider two special cases: 1. When both states are pure. 2. When the density operators commute. In the first case let $\rho_0 = |\psi_0\rangle\langle\psi_0|$ and $\rho_1 = |\psi_1\rangle\langle\psi_1|$. (For convenience, we use the Dirac ket-bra notation: the elements of the Hilbert space are denoted as $|\psi\rangle$, and the linear functionals on the Hilbert space are denoted as $\langle\psi|$. In particular, $|\psi\rangle\langle\psi|$ is the orthogonal projector on $|\psi\rangle$.) Then we have the following result:

THEOREM 1. *If both states are pure, then the expected error probability is*

$$R = \tfrac{1}{2}\bigl(1 - \sqrt{1 - |\langle\psi_0|\psi_1\rangle|^{2N}}\bigr). \tag{4.6}$$

Asymptotically,

$$\frac{1}{N}\log R \sim 2\log|\langle\psi_0|\psi_1\rangle| \qquad \text{as } N \to \infty. \tag{4.7}$$

PROOF. Because of (4.1), we need only to prove that for pure states

$$\|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1 = 2\sqrt{1 - |\langle\psi_0|\psi_1\rangle|^{2N}}. \tag{4.8}$$

We can write

$$\|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1 = \bigl\||\psi_0^{\otimes N}\rangle\langle\psi_0^{\otimes N}| - |\psi_1^{\otimes N}\rangle\langle\psi_1^{\otimes N}|\bigr\|_1. \tag{4.9}$$

The operator $|\psi_0^{\otimes N}\rangle\langle\psi_0^{\otimes N}| - |\psi_1^{\otimes N}\rangle\langle\psi_1^{\otimes N}|$ acts nontrivially only in a two-dimensional space spanned by $\psi_0^{\otimes N}$ and $\psi_1^{\otimes N}$, and it is easy to compute the operator eigenvalues in this space. They are

$$\pm\sqrt{1 - |\langle\psi_0^{\otimes N}|\psi_1^{\otimes N}\rangle|^2} = \pm\sqrt{1 - |\langle\psi_0|\psi_1\rangle|^{2N}}. \tag{4.10}$$

From this and the fact that all other eigenvalues are zero, the first equality of the theorem follows. The asymptotic expression follows from the Taylor series for the square root. □

Now consider the other simple case, that of commuting ρ_0 and ρ_1, and let the distributions of eigenvalues be P for ρ_0 and Q for ρ_1. This is essentially the classical case, so the error rate is obviously classical. For completeness we state it as a theorem:

THEOREM 2. *If states commute, then asymptotically*

$$\frac{1}{N}\log R \sim -D_c(P, Q). \tag{4.11}$$

PROOF. The conclusion follows by diagonalizing simultaneously ρ_0 and ρ_1 and applying (4.1) and definitions. □

4.3. *Bounds.* Let us now derive some simple bounds on the error probability that follow from known inequalities. These bounds are useful because they are rather narrow and easy to compute. The first set of bounds follows from inequalities between quantum fidelity and probability of error.

Recall that *fidelity* between two states is defined as

$$F(\rho_0, \rho_1) = \operatorname{tr} \sqrt{\sqrt{\rho_0}\rho_1\sqrt{\rho_0}}, \tag{4.12}$$

where $\sqrt{X}$ is the unique nonnegative definite, Hermitian matrix Y such that $Y^2 = X$.

THEOREM 3. *Probability of error for the optimal test with joint measurement satisfies the bounds*

$$\tfrac{1}{2}\bigl(1 - \sqrt{1 - [F(\rho_0, \rho_1)]^{2N}}\bigr) \le R \le \tfrac{1}{2}[F(\rho_0, \rho_1)]^N. \tag{4.13}$$

Asymptotically,

$$2\log F(\rho_0, \rho_1) \lesssim \frac{1}{N}\log R \lesssim \log F(\rho_0, \rho_1) \qquad \text{as } N \to \infty. \tag{4.14}$$

If ρ_0 is pure, $\rho_0 = |\psi_0\rangle\langle\psi_0|$, the probability of error satisfies a tighter upper bound,

$$R \le \tfrac{1}{2}[F(\rho_0, \rho_1)]^{2N} = \tfrac{1}{2}\langle\psi_0|\rho_1|\psi_0\rangle^N. \tag{4.15}$$

Asymptotically,

$$\frac{1}{N}\log R \sim \log\langle\psi_0|\rho_1|\psi_0\rangle \qquad \text{as } N \to \infty. \tag{4.16}$$

PROOF. The first result follows from inequalities (44) in Fuchs and van de Graaf (1999) applied to the case of the sample of N independent states, and from the fact that $F(\rho_0^{\otimes N}, \rho_1^{\otimes N}) = [F(\rho_0, \rho_1)]^N$. The result about the case where ρ_0 is pure follows from Exercise 9.21 in Nielsen and Chuang (2000). For the reader's convenience, I include below short proofs of these results.

The Fuchs–Graaf result states that for every pair of quantum states, ρ_0 and ρ_1, it is true that

$$1 - F(\rho_0, \rho_1) \le \tfrac{1}{2}\|\rho_0 - \rho_1\|_1 \le \sqrt{1 - F(\rho_0, \rho_1)^2}. \tag{4.17}$$

These inequalities hold because of the following results:

1. $F(\rho_0, \rho_1) = \min_{P,Q} F(P, Q)$, where distributions P and Q arise from a measurement of states ρ_0 and ρ_1, and where $F(P, Q) =: \sum_i \sqrt{p_i q_i}$.
2. $\|\rho_0 - \rho_1\|_1 = \max_{P,Q} \|P - Q\|_1$, where P and Q come from a measurement, and where $\|P - Q\|_1 =: \sum_i |p_i - q_i|$.
3. The corresponding inequality holds for probability distributions

$$1 - F(P, Q) \le \tfrac{1}{2}\|P - Q\|_1 \le \sqrt{1 - F(P, Q)^2}. \tag{4.18}$$

Indeed, given results 1–3, the left-hand side inequality in (4.17) follows because

$$1 - F(\rho_0, \rho_1) \underset{(1)}{=} 1 - F(P, Q) \qquad \text{(for certain } P \text{ and } Q\text{)} \tag{4.19}$$

$$\underset{(3)}{\leq} \tfrac{1}{2}\|P - Q\|_1 \underset{(2)}{\leq} \tfrac{1}{2}\|\rho_0 - \rho_1\|_1. \tag{4.20}$$

The right-hand side inequality follows similarly.

Result 1 is from Fuchs and Caves (1995). Result 2 is a restatement of the Holevo–Helstrom result (4.1). The left-hand side inequality in result 3 holds because

$$\sum_i |p_i - q_i| \geq \sum_i \left(\sqrt{p_i} - \sqrt{q_i}\right)^2 \geq 2\left(1 - \sum_i \sqrt{p_i q_i}\right). \tag{4.21}$$

The right-hand side inequality in result 3 holds because

$$\sum_i |p_i - q_i| = \sum_i |\sqrt{p_i} - \sqrt{q_i}\,||\sqrt{p_i} + \sqrt{q_i}\,| \tag{4.22}$$

$$\leq \sqrt{\sum_i |\sqrt{p_i} - \sqrt{q_i}|^2 \sum_i |\sqrt{p_i} + \sqrt{q_i}|^2} \tag{4.23}$$

$$= 2\sqrt{1 - \left(\sum_i \sqrt{p_i q_i}\right)^2}. \tag{4.24}$$

To prove the second part of the theorem, we need to prove that if $\rho_0 = |\psi_0\rangle\langle\psi_0|$, then there are such a measurement and a decision rule such that

$$R \leq \tfrac{1}{2}\langle\psi_0|\rho_1|\psi_0\rangle. \tag{4.25}$$

Take measurement $\{P_{\psi_0}, I - P_{\psi_0}\}$, where P_{ψ_0} is the projector on vector ψ_0. Then the probabilities of the first and second outcomes are, respectively, 1 and 0 if the state is ψ_0, and $\langle\psi_0|\rho_1|\psi_0\rangle$ and $1 - \langle\psi_0|\rho_1|\psi_0\rangle$ if the state is ρ_1. Define the decision rule as follows: state ψ_0 is accepted if and only if the first outcome occurs. The expected error of this rule is $\frac{1}{2}\langle\psi_0|\rho_1|\psi_0\rangle$.

In the case of tensor powers (4.25) becomes

$$R \leq \tfrac{1}{2}\langle\psi_0|\rho_1|\psi_0\rangle^N. \tag{4.26}$$

□

REMARK. The lower bound of inequality (4.13) binds for pure states, because for pure states $F(\rho_0, \rho_1) = |\langle\psi_0|\psi_1\rangle|$. The upper bound of inequality (4.13) binds for certain commuting operators.

Another pair of bounds follows from results by Ogawa and Nagaoka (2000) and Ogawa and Hayashi (2002). Define quantum relative entropy:

$$D(\rho_0||\rho_1) = -\operatorname{tr}[\rho_0(\log\rho_0 - \log\rho_1)]. \tag{4.27}$$

Let also

$$\Psi(\rho_0||\rho_1) = \min_{0\le s\le 1} \log \mathrm{tr}[\rho_0 \rho_1^{s/2} \rho_0^{-s} \rho_1^{s/2}]. \tag{4.28}$$

THEOREM 4.

$$\max\{D(\rho_0||\rho_1), D(\rho_1||\rho_0)\} \lesssim \frac{1}{N} \log R \lesssim \min\{\Psi(\rho_0||\rho_1), \Psi(\rho_1||\rho_0)\}. \tag{4.29}$$

PROOF. The lower bound is proved as follows. The error R is the average of error probabilities of two types, $R = \frac{1}{2}R_1 + \frac{1}{2}R_2$. For the optimal test R_1 and R_2 must have the same rate of decline. Therefore, we have two possibilities. If both $(1/N)\log R_1$ and $(1/N)\log R_2$ satisfy the inequality, then $(1/N)\log R$ also does. If both violate the inequality, then results of Ogawa and Nagaoka imply that one of the error probabilities must approach 1 as the sample size grows. Therefore, $(1/N)\log R$ approaches 0 and the inequality must hold. The upper bound is a consequence of the results in Theorem 1 of Ogawa and Hayashi (2002). □

EXAMPLE 2. Figures 1–4 illustrate the bounds. The states here are linear combinations of the Pauli matrices

$$\rho_0 = \tfrac{1}{2}(I + a\sigma_1), \tag{4.30}$$

$$\rho_1 = \tfrac{1}{2}\big(I + (b\cos\theta)\sigma_1 + (b\sin\theta)\sigma_2\big), \tag{4.31}$$

where

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}. \tag{4.32}$$

Figures 1 and 2 are for a small sample. Figure 1 shows that the test with fidelity-optimal measurements may underperform the optimal test in small samples. However, Figure 2 shows that the performance of the fidelity-optimal measurement in small samples is very close to that of the optimal measurement except for the situations when the states are close to being pure states.

Figures 3 and 4 are for a large sample. They show that the lower bounds in Theorems 3 and 4 are not achieved by the fidelity-optimal measurement, and that the lower bound from Theorem 4 (based on quantum relative entropy) is not as good an estimate of the error as the lower bound from Theorem 3 (based on fidelity). They also suggest that the upper bound from Theorem 4 is very close to the upper bound from Theorem 3.

Let us briefly summarize our findings for the case of mixed states. We found several inequalities on the optimal asymptotic rate and a good measurement that guarantees that we can realize the upper bounds. However, we can neither determine if the lower bounds in the inequalities are achievable for a given pair of states, nor check easily if a given measurement has an asymptotically best rate. These problems remain open.

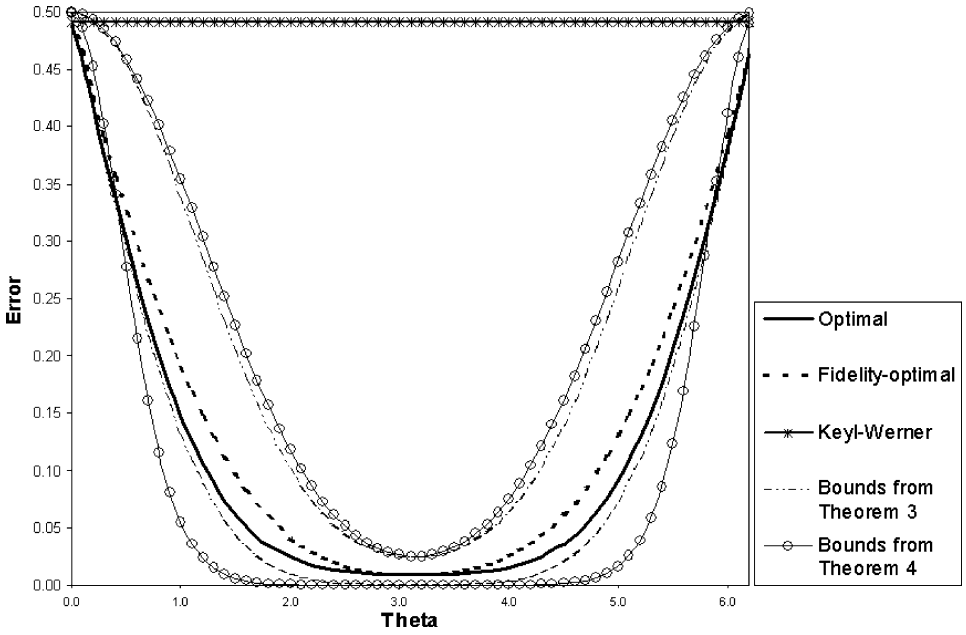

FIG. 1. *Expected test errors in a small sample. The number of sample states* $N = 3$. *The hypotheses are* $\rho_0 = \frac{1}{2}(I + \frac{8}{9}\sigma_1)$ *and* $\rho_1 = \frac{1}{2}(I + (\frac{7}{8}\cos\theta)\sigma_1 + (\frac{7}{8}\sin\theta)\sigma_2)$. *The horizontal axis shows* θ.

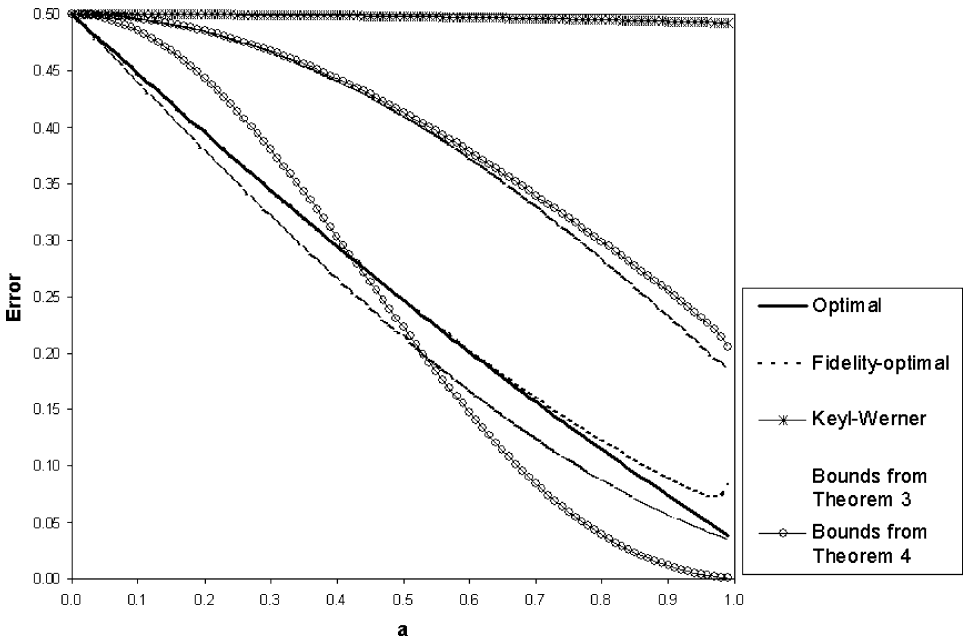

FIG. 2. *Expected test errors in a small sample. The number of sample states* $N = 3$. *The hypotheses are* $\rho_0 = \frac{1}{2}(I + a\sigma_1)$ *and* $\rho_1 = \frac{1}{2}(I + \frac{63}{64}a\sigma_2)$. *The horizontal axis shows* a.

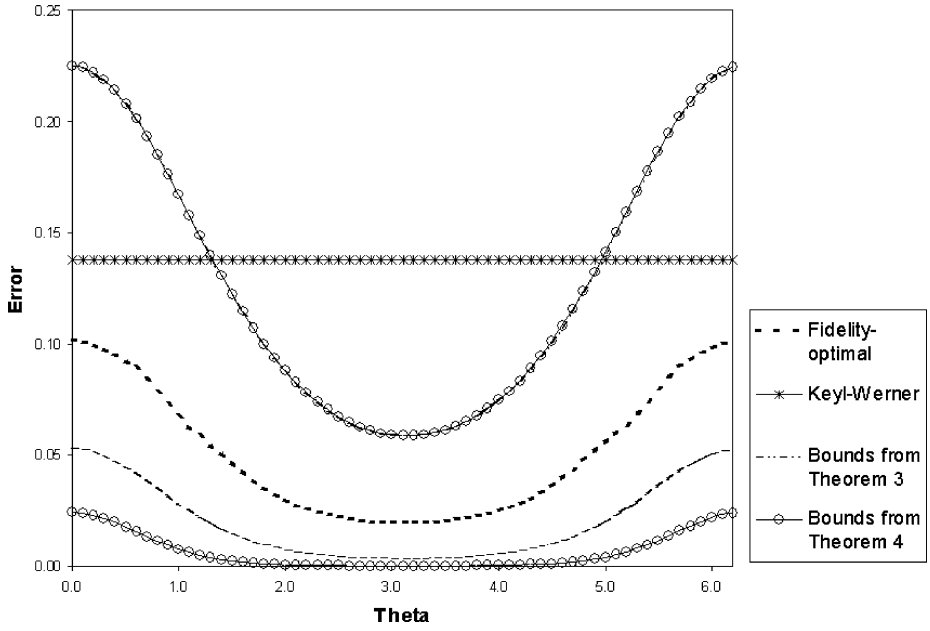

FIG. 3. *Expected test errors in a large sample. The number of sample states* $N = 40$. *The hypotheses about states are* $\rho_0 = \frac{1}{2}(I + \frac{1}{2}\sigma_1)$ *and* $\rho_1 = \frac{1}{2}(I + (\frac{1}{8}\cos\theta)\sigma_1 + (\frac{1}{8}\sin\theta)\sigma_2)$. *The horizontal axis shows* θ.

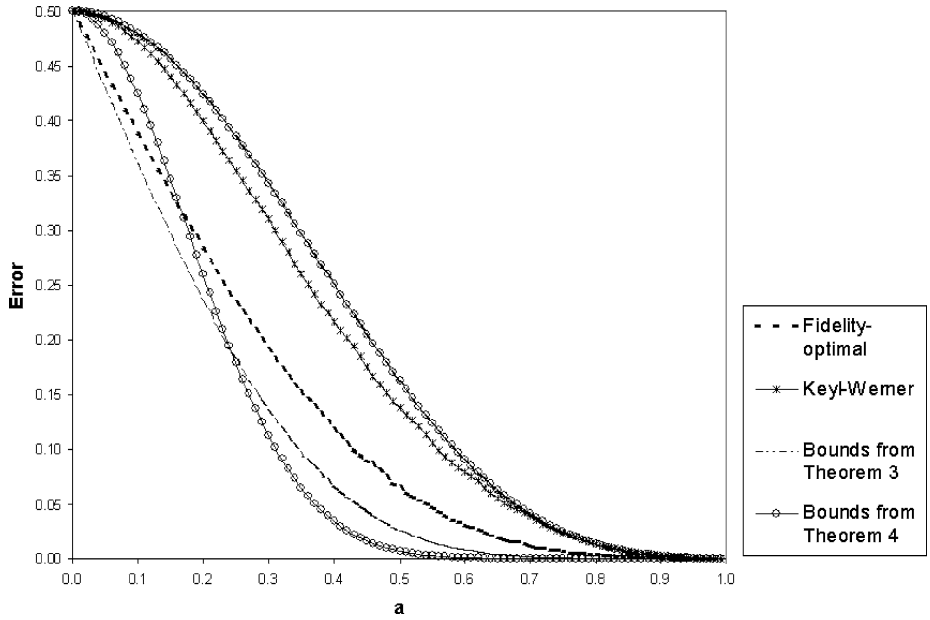

FIG. 4. *Expected test errors in a large sample. The number of sample states* $N = 40$. *The hypotheses are* $\rho_0 = \frac{1}{2}(I + a\sigma_1)$ *and* $\rho_1 = \frac{1}{2}(I + \frac{1}{8}a\sigma_1 + \frac{\sqrt{3}}{8}a\sigma_2)$. *The horizontal axis shows* a.

5. Separate measurements. In the previous section we have seen that it is difficult to compute the optimal joint measurement because of the high dimensionality of the problem involved. Besides, even if the optimal joint measurement is found, it can be hard to realize it in the laboratory. For example, proposed designs of the quantum computer typically use circuits built from standard quantum states (called qubits) and small quantum gates. In this situation, it appears desirable to avoid joint measurements of a large number of qubits. In this section we turn our attention to separate independent measurements. The goal is to compare the efficiency of optimal separate and joint measurements.

It is not difficult to find examples where joint measurements offer a performance advantage over separate measurements. The main question is how this advantage behaves when the sample size grows. Is it washed out? Does it grow? Does it stay relatively constant? Figure 5 gives some numerical evidence that the advantage grows. But before we address this question in detail, let us look more closely at the optimal separate measurements.

What is the structure of the optimal separate measurements? This is a difficult question but we can easily demonstrate the existence of measurements with a simple structure and the optimal exponential error rate.

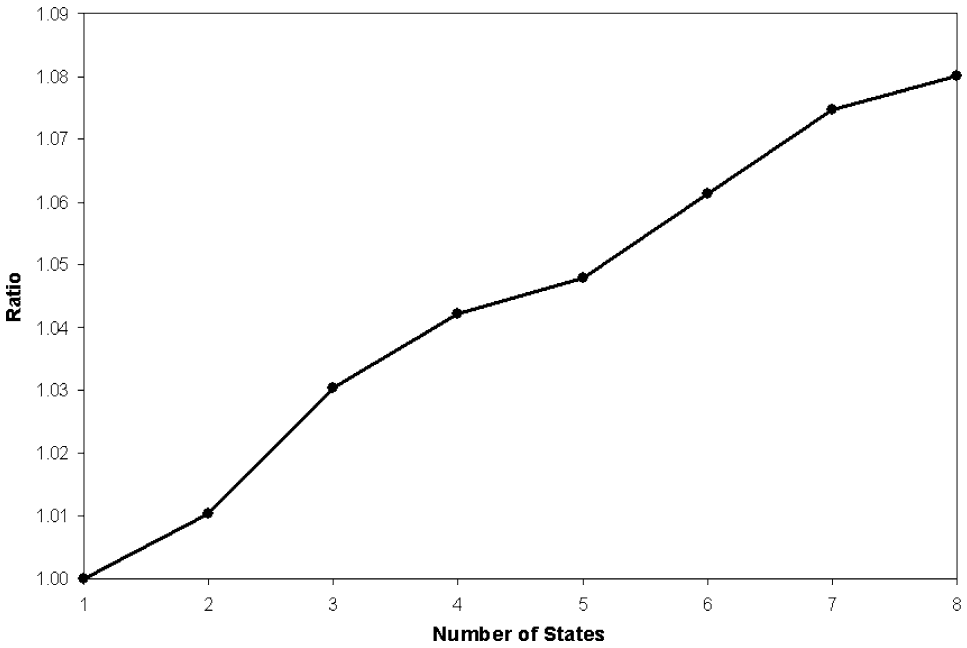

FIG. 5. *Ratio of expected errors in tests with joint and separate measurements. The hypotheses are* $\rho_0 = \frac{1}{2}(I + \frac{5}{6}\sigma_1)$ *and* $\rho_1 = \frac{1}{2}(I + \frac{3\sqrt{3}}{8}\sigma_1 + \frac{3}{8}\sigma_2)$. *The horizontal axis shows the number of states in the sample. The vertical axis shows the ratio of the expected errors in the optimal separate and joint tests.*

THEOREM 5. *There is a separate measurement that is represented by operators proportional to one-dimensional projectors and that has the same asymptotic rate of error as the optimal separate measurement.*

PROOF. If an optimal measurement includes an outcome represented by a matrix, M_0, which is not proportional to a one-dimensional projector, then this matrix can be represented as a sum of one-dimensional projectors with nonnegative coefficients,

$$M_0 = \sum_{i=1}^{n} \alpha_i M_i. \tag{5.1}$$

Therefore

$$p_0 = \sum_{i=1}^{n} \alpha_i p_i \quad \text{and} \quad q_0 = \sum_{i=1}^{n} \alpha_i q_i, \tag{5.2}$$

where $p_i = \operatorname{tr}(M_i \rho_0)$ and $q_i = \operatorname{tr}(M_i \rho_1)$. Since the function $x^\lambda y^{1-\lambda}$ is concave and homogeneous, we have the inequality

$$p_0^\lambda q_0^{1-\lambda} \geq \sum_{i=1}^{n} (\alpha_i p_i)^\lambda (\alpha_i q_i)^{1-\lambda}. \tag{5.3}$$

Because of (3.6), this inequality implies that using the set of outcomes represented by matrices $\{\alpha_i M_i\}$ instead of the outcome represented by M_0 cannot increase the asymptotic rate of error. Consequently, by continuing in this fashion we can refine the optimal measurement to a measurement with components proportional to projectors and having the same asymptotic rate of error. □

If one of the states is pure, $\rho_0 = |\psi_0\rangle\langle\psi_0|$, then we can explicitly write down a test that has the same exponential error rate as the optimal separate measurement:

THEOREM 6. *When one of the states is pure, there is a separate test with the expected error probability that satisfies the bound*

$$R \lesssim \tfrac{1}{2} \langle \psi_0 | \rho_1 | \psi_0 \rangle^N \qquad \text{as } N \to \infty. \tag{5.4}$$

PROOF. Take measurement $\{P_{\psi_0}, I - P_{\psi_0}\}$, where P_{ψ_0} is the projector on vector ψ_0. Then the probabilities of the first and second outcomes are, respectively, 1 and 0 if the state is ρ_0, and $\langle \psi_0 | \rho_1 | \psi_0 \rangle$ and $1 - \langle \psi_0 | \rho_1 | \psi_0 \rangle$ if the state is ρ_1. Define the decision rule as follows: state ρ_0 is accepted if and only if the second outcome never occurs. This rule leads to an error if and only if the true state is ρ_1 and the second outcome never occurs. Thus the average probability of error for this decision rule is

$$R = \tfrac{1}{2} \langle \psi_0 | \rho_1 | \psi_0 \rangle^N. \tag{5.5}$$

□

Noticing that the optimal separate test cannot have a better exponential rate than the optimal joint test, and comparing the rate in (5.4) with the rate of the optimal joint test in (4.16), we conclude that the test devised in Theorem 6 has the same exponential error rate as both the optimal separate and optimal joint tests.

What should be emphasized, however, is that while the exponential rates for optimal separate and joint tests are the same, the expected error may still be significantly larger for the separate test. The next example shows that the ratio of expected errors in optimal separate and joint tests can be as large as 2 to 1.

EXAMPLE 3. Let the dimension of the Hilbert space be $d = 2$, and let both states, ψ_0 and ψ_1, be pure. Then by a simple optimization, we can find that the optimal separate projective measurement has two components that project, on one of the states and on its orthogonal complement. The expected error of this test is

$$R_{\text{sep}} = \tfrac{1}{2}\left|\langle\psi_0|\psi_1\rangle\right|^{2N}. \tag{5.6}$$

At the same time, we know from Theorem 1 that the expected error of the optimal joint test is

$$R_{\text{joint}} = \tfrac{1}{2}\left(1 - \sqrt{1 - \left|\langle\psi_0|\psi_1\rangle\right|^{2N}}\right) \sim \tfrac{1}{4}\left|\langle\psi_0|\psi_1\rangle\right|^{2N}. \tag{5.7}$$

Consequently,

$$\frac{R_{\text{sep}}}{R_{\text{joint}}} \to 2 \qquad \text{as } N \to \infty. \tag{5.8}$$

If both states are mixed, little is known about the optimal separate test and its asymptotic rate. As an approximation, we can use a measurement that maximizes fidelity (also known as Hellinger) distance between distributions of outcomes. In other words, the measurement is chosen in such a way that it minimizes

$$F(P, Q) = \sum \sqrt{p_i q_i}. \tag{5.9}$$

We will call this measurement fidelity-optimal. The advantage of this method is that the fidelity-optimal measurement is easy to compute. It is simply a measurement with outcomes that are orthogonal projectors on the eigenvectors of the operator

$$M = \rho_1^{-1/2}\sqrt{\rho_1^{1/2}\rho_0\rho_1^{1/2}}\,\rho_1^{-1/2}. \tag{5.10}$$

[See Fuchs and Caves (1995) for an explanation why this M is fidelity-optimal.]

THEOREM 7. *The expected error of the test based on the fidelity-optimal measurement satisfies the asymptotic bound*

$$\frac{1}{N}\log R \lesssim \log F(\rho_0, \rho_1). \tag{5.11}$$

PROOF.

$$\frac{1}{N}\ln R = \min_{0\le\lambda\le 1} \log \sum_{i=1}^{N} p_i^{\lambda} q_i^{1-\lambda} \le \log \sum_{i=1}^{N} \sqrt{p_i q_i} \le \log F(\rho_0, \rho_1). \tag{5.12}$$

The equality holds because of (3.6), and the second inequality is inequality (44) in Fuchs and van de Graaf (1999). □

This is the same upper bound that we derived for joint asymptotic measurement in Theorem 3. Comparing (5.11) with (4.14) shows, however, that there is a possible difference between exponential rates of separate and joint measurements, although the maximal size of the ratio of the rates cannot be greater than 2 to 1.

6. Illustration. This section illustrates the concepts developed above with an example of testing for the presence of entanglement. Entanglement is one of the properties of quantum systems that clearly separates them from classical systems. It is a co-dependence of two remote parts of a quantum system that cannot be created (though it can easily be destroyed) by local operations on the parts. Entanglement has recently been recognized as an important part of many quantum technologies including quantum teleportation and quantum cryptography.

There are several methods to produce entanglement. One of them involves a random decay of a laser-pumped atom. The entanglement is then shared between two photons issued by the atom. Another method, proposed by Turchette et al. (1998), creates entanglement by placing two interacting ions in a trap and illuminating them equally by a laser beam. In this illustration we are interested in tests of whether or not the entanglement has been produced in a given sample.

An example of an entangled quantum state is a pure state of the system of two particles that corresponds to the projector on the vector

$$\psi_0 = \tfrac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \tag{6.1}$$

where $|00\rangle$ and $|11\rangle$ denote $|0\rangle \otimes |0\rangle$ and $|1\rangle \otimes |1\rangle$, and $|0\rangle$ and $|1\rangle$ form an orthonormal basis in the Hilbert space corresponding to one of the particles.

The density matrix for this system is

$$\rho_0 = |\psi_0\rangle\langle\psi_0| = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}. \tag{6.2}$$

The alternative hypothesis is that the state is a mix of two nonentangled states given by projectors on vectors $|00\rangle$ and $|11\rangle$, respectively. The density matrix for this hypothesis is

$$\rho_1 = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}. \tag{6.3}$$

This state can be easily produced by local operations but it is useless for technologies that require entanglement.

In application we can expect that ρ_0 and ρ_1 are contaminated by noise. A simple model of noisy hypotheses is represented by the mixed density matrices

$$\widetilde{\rho}_i = \alpha\rho_i + (1-\alpha)\tfrac{1}{4}I, \tag{6.4}$$

where $0 \le \alpha \le 1$, and I is the unit density matrix that represents noise. The parameter α measures the degree of contamination by the noise.

It is easy to calculate fidelity:

$$F(\widetilde{\rho}_0, \widetilde{\rho}_1) = \tfrac{1}{2}\bigl(1-\alpha+\tfrac{1}{2}\sqrt{1-\alpha^2}+\tfrac{1}{2}\sqrt{1+4\alpha+3\alpha^2}\bigr). \tag{6.5}$$

Applying Theorem 3, we have

$$\tfrac{1}{2}\bigl(1-\sqrt{1-[F(\widetilde{\rho}_0, \widetilde{\rho}_1)]^{2N}}\bigr) \le R \le \tfrac{1}{2}[F(\widetilde{\rho}_0, \widetilde{\rho}_1)]^N. \tag{6.6}$$

Consider, for example, the case with a relatively high level of noise: $\alpha = 20\%$. Then fidelity is $F = 0.9913$, and for a sample with 100 quantum states the lower and upper bounds on the expected error are 0.046 and 0.21, respectively. For 300 states the upper bound is 0.037 and we can be sure that the expected error of the test is below 5% threshold.

7. Conclusion. We have bounded the Chernoff efficiency, from above and below, for cases of joint and separate measurements and also calculated it exactly if at least one of the states is pure. In the latter case, the optimal separate measurement results in the same asymptotic error rate as the optimal joint measurement. We have shown by example, however, that the ratio of the error of optimal separate measurement to that of optimal joint measurement can be as large as a factor of 2.

If both states are mixed, there is a distinct possibility that the optimal test with joint measurements can have a better exponential rate than the optimal test with separate measurements. Still, the bounds that we have obtained show that this improvement in the rate cannot be greater than a factor of 2.

Several questions remain open. Most notably, it is not known whether joint measurement can ever have a better exponential error rate than optimal separate measurement. Second, the characteristic properties of optimal separate measurement are not known. In particular, it is not known whether optimal separate measurement of an individual state consists of orthogonal projectors.

Another open direction for research is the small deviations approach to quantum hypothesis testing. In this approach hypotheses about quantum states approach each other when the sample size grows. Asymptotics in this case are likely to be related to the quantum information matrix introduced in the context of quantum estimation theory.

REFERENCES

BALLESTER, M. A. (2004). Estimation of unitary quantum operations. *Phys. Rev. A* **69** 022303.

BARNDORFF-NIELSEN, O. E., GILL, R. D. and JUPP, P. E. (2003). On quantum statistical inference (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 775–816.

BENNETT, C. H., DIVINCENZO, D. P., FUCHS, C. A., MOR, T., RAINS, E., SHOR, P. W., SMOLIN, J. A. and WOOTTERS, W. K. (1999). Quantum nonlocality without entanglement. *Phys. Rev. A* **59** 1070–1091.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

CIRAC, J. I. and ZOLLER, P. (1995). Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74** 4091–4094.

COVER, T. M. and THOMAS, J. A. (1991). *Elements of Information Theory*. Wiley, New York.

FUCHS, C. A. and CAVES, C. M. (1995). Mathematical techniques for quantum communication theory. *Open Systems and Information Dynamics* **3** 345–356. Available at http://arxiv.org/abs/quant-ph/9604001.

FUCHS, C. A. and VAN DE GRAAF, J. (1999). Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inform. Theory* **45** 1216–1227.

GILL, R. D. (2001). Asymptotics in quantum statistics. In *State of the Art in Probability and Statistics* 255–285. IMS, Beachwood, OH.

GILL, R. D. and MASSAR, S. (2000). State estimation for large ensembles. *Phys. Rev. A* **61** 042312.

HELSTROM, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic Press, New York.

HOEFFDING, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–401.

HOLEVO, A. S. (1976). Investigation of a general theory of statistical decisions. *Trudy Mat. Inst. Steklov.* **124**. [English translation in *Proc. Steklov Inst. Math.* **3** (1978) Amer. Math. Soc., Providence, RI.]

HOLEVO, A. S. (2001). *Statistical Structure of Quantum Theory*. Springer, Berlin.

KEYL, M. and WERNER, R. F. (2001). Estimating the spectrum of a density operator. *Phys. Rev. A* **64** 052311.

NIELSEN, M. A. and CHUANG, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge Univ. Press.

OGAWA, T. and HAYASHI, M. (2002). On error exponents in quantum hypothesis testing. Available at http://arxiv.org/abs/quant-ph/0206151.

OGAWA, T. and NAGAOKA, H. (2000). Strong converse and Stein's lemma in quantum hypothesis testing. *IEEE Trans. Inform. Theory* **46** 2428–2433.

PARTHASARATHY, K. R. (2001). On consistency of the maximum likelihood method in testing multiple quantum hypotheses. In *Stochastics in Finite and Infinite Dimensions* (T. Hida et al., eds.) 361–377. Birkhäuser, Boston.

PERES, A. (1995). *Quantum Theory*: *Concepts and Methods*. Kluwer, Dordrecht.

SANOV, I. N. (1957). On the probability of large deviations of random variables. *Mat. Sb. N. S.* **42** 11–44.

SANTALÓ, L. A. (1976). *Integral Geometry and Geometric Probability*. Addison–Wesley, Reading, MA.

TURCHETTE, Q. A., WOOD, C. S., KING, B. E., MYATT, C. J., LEIBFRIED, D., ITANO, W. M., MONROE, C. and WINELAND, D. J. (1998). Deterministic entanglement of two trapped ions. *Phys. Rev. Lett.* **81** 3631–3634.

CORNERSTONE RESEARCH
599 LEXINGTON AVENUE
NEW YORK, NEW YORK 10022
USA
E-MAIL: skarguine@cornerstone.com
URL: www.io.com/˜slava