Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

The Chernoff lower bound for symmetric quantum hypothesis testing

(revised version: August 2007)

by

Michael Nussbaum, and Arleta Szkola

Preprint no.: 69 2006

The Chernoff lower bound for symmetric quantum hypothesis testing

Michael Nussbaum[1] and Arleta Szkoła[2]

Cornell University and Max Planck Society

Abstract

We consider symmetric hypothesis testing in quantum statistics, where the hypotheses are density operators on a finite-dimensional complex Hilbert space, representing states of a finite quantum system. We prove a lower bound on the asymptotic rate exponents of Bayesian error probabilities. The bound represents a quantum extension of the Chernoff bound, which gives the best asymptotically achievable error exponent in classical discrimination between two probability measures on a finite set. In our framework the classical result is reproduced if the two hypothetic density operators commute.

Recently it has been shown elsewhere [1] that the lower bound is achievable also in the generic quantum (noncommutative) case. This implies that our result is one part of the definitive quantum Chernoff bound.

1 Introduction

One typical problem in hypothesis testing is to decide between two equiprobable hypotheses, say H_0 and H_1, where H_i assumes that the observed data are generated by an i.i.d. process with law P_i, $i = 0, 1$. In the classical setting P_0, P_1 are probability measures on a measurable space, the sample space. One discriminates between them by means of test functions, which are nonnegative measurable functions on the n-fold product sample space. An error occurs if according to the given decision rule based on the value of the test function, one accepts hypothesis H_0 while the data are generated with law P_1, or vice versa.

If one declares one of the hypotheses to be the null hypothesis and the other one the alternative, then errors occuring while the null hypothesis is true are called of first kind, otherwise of second kind. Due to Stein's lemma there exist test functions maintaining a given

[1] Supported in part by NSF Grant DMS-03-06497

[2] Supported in part by German Research Foundation (DFG) via the project "Entropy, Geometry and Coding of Large Quantum Information Systems"

1991 Mathematics Subject Classification. 62P35, 62G10

Key words and phrases. Quantum statistics, density operators, Bayesian discrimination, exponential error rate, Holevo-Helstrom tests, quantum Chernoff bound.

upper bound α on the error probability of first kind such that the probability of error of the second kind decreases to 0 with the optimal asymptotic rate exponent equal to the Kullback-Leibler distance from the null hypothesis to the alternative. Sanov's theorem extends this result to the case where instead of a single measure P_0, a family Ω of measures is associated with the null hypothesis. Then the negative Kullback-Leibler distance from the set Ω to P_1 gives the minimal asymptotic error exponent, [17], see also [6].

In *symmetric* hypothesis testing one treats the errors of first and second kind in a symmetric way. We will focus here on the *Bayesian error probability*, which is the average of the two kinds of error probabilities. It is minimized by the likelihood ratio test and vanishes exponentially fast as the sample size n tends to infinity. The corresponding optimal asymptotic rate exponent is equal to the *Chernoff bound*

$$\inf_{0\leq s\leq 1} \log \int p_0^{1-s}(\omega) p_1^s(\omega) \mu(d\omega) \tag{1}$$

pertaining to probability measures P_0 and P_1 with respective densities p_0 and p_1 (wrt dominating measure $\mu = P_0 + P_1$). These results go back to papers by Chernoff and Hoeffding, [5, 10]. Chentsov and Morozova [4] present a thorough and illuminating discussion of the Chernoff bound, relating it to the differential geometry of statistical inference.

If the data are obtained from quantum systems then one has to replace probability measures by quantum states, i.e. by normalized positive linear functionals on an appropriate algebra of observables. In the present paper this is assumed to be the algebra of linear operators on a finite-dimensional complex Hilbert space. One discriminates between two states ρ_0 and ρ_1 by means of quantum tests, which are defined as positive operator valued measures on n-fold tensor products of the algebra of observables of a single quantum system. Here we employed the standard language of quantum mechanics; throughout the paper however we will utilize an elementary and accessible mathematical framework based on complex linear algebra only. It will become apparent that quantum tests are analogs of test functions defined on finite sample spaces and their n-fold products.

While the basic problems in nonsymmetric quantum hypothesis testing (pertaining to α-tests) were solved in [9], [15] and [3] by obtaining quantum versions of Stein's lemma and Sanov's theorem, the case of discrimination (or equally weighted hypotheses) has not yet received full teatment. Although quantum tests minimizing the generalized Bayesian error probabilities were constructed about 30 years ago by Helstrom and Holevo in [8, 11], a closed form expression for the optimal asymptotic quantum error exponent similar to the classical Chernoff distance remained an open problem. A reason has been that there is no obvious canonical way to extend (1) to a quantum setting. On the very formal level, due to non-commutativity effects, there are different non equivalent ways of generalizing the distance. In [15] Ogawa and Hayashi list three candidates for the optimal quantum rate exponent, relying on three different extensions of the target function in the variational formula (1). However, two of these candidate expressions are not well defined if the hypotheses are not faithful states, i.e. if the associated density operators do not have full rank.

Recently the problem of symmetric quantum testing was treated by Kargin [12], with partial progress towards the definitive Chernoff bound. Lower and upper bounds on the optimal error exponent in terms of fidelity between the two density operators were given; the lower bound was shown to be sharp in the case that one of the density operators has rank one (i.

e. represents a pure quantum state). We remark that fidelity is a notion of distinguishability between density operators which is frequently used in quantum information theory, see e.g. [13].

Our main result, which we formulate rigorously in Section 2, states that $\inf_{0\le s\le 1} \log \mathrm{Tr}\left[\ \rho_0^{1-s}\rho_1^s\right]$ is a lower bound on the general asymptotic error exponent, ρ_0 and ρ_1 being density operators replacing the probability densities p_0 and p_1 of the classical setting. We remark that our quantum bound coincides with one of the three candidates for a quantum Chernoff bound discussed in [15]. We prove the main theorem in Section 3. Recently, Audenaert et al. have shown in [1] that in accordance with our conjecture stated in a previous version of the present work, [14], the lower bound is indeed achievable. This justifies to refer to it as quantum Chernoff bound.

2 Mathematical setting and the main theorem

For an elementary introduction to quantum statistics with physical background, see Gill [7]. We will describe here only the formalism for the simplest possible nonclassical setup of discrimination between two hypotheses. A *density matrix* ρ is a complex, self-adjoint, positive, $d \times d$ matrix satisfying the normalization condition $\mathrm{Tr}\,[\rho] = 1$, where $\mathrm{Tr}\,[\cdot]$ is the trace operation. Here "positive" means nonnegative definite. We identify a density matrix with a state of a quantum system; we also use "matrix" and "operator" interchangeably. The two hypotheses are described by two states $H_0 : \rho = \rho_0$, $H_1 : \rho = \rho_1$. Decisions are made using a *test* r, which is a complex self-adjoint positive $d \times d$ operator satisfying the inequality $r \le \mathbf{1}$. Here $\mathbf{1}$ is the unit matrix and $\le$ is in the sense of matrix order, i.e. $\mathbf{1}-r_n$ is positive (nonnegative definite). In particular, projection operators are tests. Applying the test to a state ρ the experimenter or observer creates a random variable taking values in the spectrum (set of eigenvalues) of r ; the expectation of this random variable is $\mathrm{Tr}[\rho r]$. Thus a test gives a r.v. with at most d possible values in $[0, 1]$. In line with the usual understanding of a randomized test, these values are interpreted as a conditional probability of rejecting the null hypothesis H_0. Then $\mathrm{Tr}[\rho r]$ is the overall probability of rejecting H_0 when ρ is the true state. Accordingly, $\mathrm{Tr}\ [\rho_0 r]$ is the *error probability of first kind* and $\mathrm{Tr}[(\mathbf{1} - r)\rho_1] = 1-\mathrm{Tr}\ [\rho_0 r]$ is the *error probability of second kind*. When both ρ_0, ρ_1 and also r are diagonal matrices then the setup reduces to the classical testing problem for two probability measures on an appropriate index set Ω, $|\Omega| = d$ given by ρ_0, ρ_1 respectively. The same is true when ρ_0, ρ_1 have the same set of eigenvectors; then ρ_0, ρ_1 are said to commute (*commutative case*). In this sense, commuting states describe the classical discrimination problem between two probability measures on a finite sample space Ω, as a special case of the present quantum setting.

A *pure state* is given by a density matrix which has rank 1, which means it is a projection onto a subspace of (complex) dimension one. We will also use the following notation: we set $\mathcal{H} = \mathbb{C}^d$, with the understanding that $\mathcal{H}$ can be any d-dimensional complex Hilbert space, and we write $\mathcal{B}(\mathcal{H})$, $\mathcal{B}(\mathcal{H}^{\otimes n})$ for the set of complex $d \times d$ or $dn \times dn$ matrices, respectively. In the bra-ket notation, $|v\rangle$ and $\langle v|$ denote a vector in $\mathcal{H}$ and its dual vector with respect to the scalar product in $\mathcal{H}$ (essentially a column and a row vector). A one dimensional projection onto a subspace of $\mathcal{H}$ spanned by a unit vector v may be written as $|v\rangle\langle v|$. It is a density

operator of a pure state.

The above describes the basic setup where the finite dimension d is arbitrary. We consider the quantum analog of having n i.i.d. observations. For this, the two hypotheses are assumed to be $\rho_0^{\otimes n}$ and $\rho_1^{\otimes n}$ for two basic d-dimensional states ρ_0, ρ_1, where $\rho^{\otimes n}$ is the n-fold tensor product of ρ with itself. (Recall that the tensor product $a \otimes b$ of two matrices is a matrix which consists of blocks $a_{ij}b$, arranged according to the indices i, j. Thus $\rho_0^{\otimes n}$ is a $dn \times dn$ matrix.) The tests r_n now operate on the states $\rho_0^{\otimes n}$ and $\rho_1^{\otimes n}$, i.e. their dimension is $dn \times dn$, but they need not have tensor product structure. The corresponding *Bayesian error probability* is

$$\begin{aligned} Err(r_n) :=& \frac{1}{2}\mathrm{Tr}\left[\ (r_n\rho_0^{\otimes n} + (\mathbf{1} - r_n)\rho_1^{\otimes n})\right] \\ &= \frac{1}{2}\left(1 - \mathrm{Tr}\left[r_n(\rho_1^{\otimes n} - \rho_0^{\otimes n})\right]\right). \end{aligned}$$

The optimal hypothesis tests minimizing the error probability are known to be the *Holevo-Helstrom hypothesis tests*, [11, 8]. They are given for each $n \in \mathbb{N}$ by the projections

$$\Pi_n^* := \mathrm{supp}\ (\rho_1^{\otimes n} - \rho_0^{\otimes n})_+,$$

where supp a denotes the support projection of a linear operator a and a_+ means the positive part of a self-adjoint operator a. Thus if $a = \sum_i \lambda_i E_i$ is the the spectral decomposition using projections E_i then $a_+ := \sum_{\lambda_i > 0} \lambda_i E_i$ and $\mathrm{supp} a_+ = \sum_{\lambda_i > 0} E_i$. Indeed we have for an arbirtrary test operator in $\mathcal{B}(\mathcal{H}^{\otimes n})$

$$\begin{aligned} Err(r_n) &= \frac{1}{2}\left(1 - \mathrm{Tr}\left[\ r_n(\rho_1^{\otimes n} - \rho_0^{\otimes n})\right]\right) \\ &\geq \frac{1}{2}\left(1 - \sup\{\mathrm{Tr}\left[\ \tilde{r}(\rho_1^{\otimes n} - \rho_0^{\otimes n})\right] :\ \tilde{r} \in \mathcal{B}(\mathcal{H}^{\otimes n})\ \text{test}\}\right) \\ &= \frac{1}{2}\left(1 - \sup\{\mathrm{Tr}\ \left[\Pi(\rho_1^{\otimes n} - \rho_0^{\otimes n})\right] :\ \Pi \in \mathcal{B}(\mathcal{H}^{\otimes n})\ \text{projection}\}\right) \\ &= \frac{1}{2}\left(1 - \mathrm{Tr}\ \left[\Pi_n^*(\rho_1^{\otimes n} - \rho_0^{\otimes n})\right]\right) \\ &= \frac{1}{2}\left(1 - \frac{1}{2}\|\rho_1^{\otimes n} - \rho_0^{\otimes n}\|_1\right) \end{aligned}$$

where $\|a\|_1 = \mathrm{Tr}[a_+] + \mathrm{Tr}[a_+ - a]$ is the generalization of the L_1-norm. Note that the last line above gives an exact closed form expression of the best error probability for every n, but its asymptotics as $n \to \infty$ (rate of exponential decay) is the subject of the present paper.

The Holevo-Helstrom tests Π_n^* are non-commutative generalizations of the likelihood ratio tests: if the hypotheses H_0 and H_1 correspond to commuting density operators ρ_0 and ρ_1 then for all $n \in \mathbb{N}$ the Holevo-Helstrom projections Π_n^* commute with $\rho_0^{\otimes n}$ and $\rho_1^{\otimes n}$, too. The density operators ρ_i may be completely specified by their eigenvalues forming discrete probability measures P_i, $i = 0, 1$ on an appropriate index set Ω, $|\Omega| = d$ for the mutually commuting spectral projectors on $\mathcal{H}$. For each $n \in \mathbb{N}$ the set of eigenvalues of the tensor product $\rho_i^{\otimes n}$, $i = 0, 1$, corresponds to the respective product measure $P_i^n := \prod_{j=1}^n P_i$ on the cartesian product $\Omega^n := \times_{i=1}^n \Omega$ while the Holevo-Helstrom projection Π_n^* generalizes the indicator function $\lambda_n^* = \mathbf{1}\{p_1^n - p_0^n > 0\}$ on Ω^n, which is the well known maximum likelihood

decision. Here p_i denote the probability densities of the laws P_i. Define the classical error probability $Err(\lambda)$ of a test function λ ($0 \leq \lambda \leq 1$) by

$$Err(\lambda) = \frac{1}{2}\left(E_{P_0}\lambda + E_{P_1}(1-\lambda)\right). \tag{2}$$

As already mentioned in the Introduction the Bayesian error probability $Err(\lambda_n^*)$ vanishes, as $n \to \infty$, with a minimal asymptotical rate exponent equal to the *Chernoff bound* $\delta(P_0, P_1)$:

$$\lim_{n\to\infty} \frac{1}{n} \log Err(\lambda_n^*) = \delta(P_0, P_1) := \inf_{0\leq s\leq 1} \log \sum_{x\in\Omega} p_0^{1-s}(x) p_1^s(x). \tag{3}$$

We remark that

$$\sum_{x\in\Omega} p_0^{1-s}(x) p_1^s(x) =: A(s), \qquad s \in [0,1], \tag{4}$$

represent the normalization factors of the parametric family of probability measures

$$p_s(x) := \frac{1}{A(s)} p_0^{1-s}(x) p_1^s(x), \qquad x \in \Omega.$$

The family is called a *Hellinger arc* in the literature. It interpolates between p_0 and p_1 if their supports $D_0, D_1 \subseteq \Omega$ coincide. Otherwise p_s, $s \in [0,1]$, is discontinuous (in the Euclidian metric of $\mathbb{R}^{|\Omega|}$) at the endpoints $s = 0, 1$ such that over the open parameter interval $(0,1)$ it represents an interpolation between the densities of the conditional probablities $Q_0 := P_0(\cdot|B)$ and $Q_1 := P_1(\cdot|B)$, where $B := D_0 \cap D_1$.

There is an equivalent expression for the Chernoff bound (3) in terms of the KL-distance (relative entropy):

$$\delta(P_0, P_1) = \inf_{s\in[0,1]} \left(-(1-s)K(Q_s\|Q_0) - sK(Q_s\|Q_1) + \log \pi_0^{1-s}\pi_1^s\right), \tag{5}$$

where Q_s denotes the conditional probability $P_s(\cdot|B)$, for $s \in [0,1]$, and $\pi_i := P_i(B)$, for $i = 0, 1$. Observe that if the supports D_0 and D_1 coincide, i.e. $B = \Omega$, then the target function in (5) -we will refer to it as $H(s)$ in the sequel- becomes simply $-(1-s)K(P_s\|P_0) - sK(P_s\|P_1)$. What is remarkable is that in this case we have

$$\delta(P_0, P_1) = -K(P_\sigma\|P_0) = -K(P_\sigma\|P_1),$$

where the parameter $\sigma \in [0,1]$ is uniquely defined by the second equality above. In the generic case of possibly different supports a modified version of the above formula is valid. One distinguishes two cases: if there exists a $\sigma \in (0,1)$ such that $H'(\sigma) = 0$, which is equivalent to $K(Q_\sigma\|Q_0) - K(Q_\sigma\|Q_1) = \log(\pi_0/\pi_1)$, then

$$\delta(P_0, P_1) = -K(Q_\sigma\|P_0) + \log \pi_0 = -K(Q_\sigma\|P_1) + \log \pi_1.$$

Otherwise, the infimum in (5) is attained either at $s = 0$ or at $s = 1$ and the corresponding values of the Chernoff bound are $\log \pi_0$ and $\log \pi_1$.

The identity (5) and the other claims in the above paragraph follow from (25) in the Appendix and attendant reasoning. To our knowledge, no quantum generalization of (5) has yet been found.

In the following theorem we formulate the classical result (3) for the general case of probability measures P_0, P_1 on an arbitrary measurable space (Ω, Σ), not necessarily finite. Consider the Bayesian error probability of discrimination between P_0, P_1 by means of test functions $0 \leq \lambda \leq 1$:

$$\Delta\left(P_0, P_1\right) := \inf_{\lambda \text{ test function}} Err(\lambda). \tag{6}$$

where $Err(\lambda)$ is given by (2). Let λ^* be the maximum likelihood test function $\lambda^* = \mathbf{1}\{p_1 - p_0 > 0\}$ on Ω in terms of densities p_0, p_1 for some dominating measure μ. It is well known that $\Delta\left(P_0, P_1\right)$ can be expressed as

$$\Delta\left(P_0, P_1\right) = Err(\lambda^*) = \frac{1}{2}\int \min\left(p_0, p_1\right) d\mu. \tag{7}$$

Theorem 2.1 *Let P_0, P_1 be two probability measures on (Ω, Σ). For product measures P_0^n, P_1^n corresponding to n i.i.d. observations $\omega_1, \ldots, \omega_n$ all having law P_0 or P_1, the Bayesian error probability satisfies*

$$\lim_{n\to\infty} n^{-1} \log\ \Delta\left(P_0^n, P_1^n\right) = \inf_{0\leq s\leq 1} \log \int p_1^s p_0^{1-s} d\mu \tag{8}$$

where $p_i = dP_i/d\mu$, $i = 0, 1$, $\mu := P_0 + P_1$.

For strictly positive p_0 and p_1 with $p_0 \neq p_1$ the proof can be found in the literature, cf. e.g. [4], p. 164, or for finite sample space in [6], p. 312. For completeness, we present a proof for the general case of possibly different support of P_0, P_1 in the Appendix. Indeed if P_0, P_1 have the same support then the function $A(s) = \int p_1^s p_0^{1-s} d\mu$ is analytic and strictly convex, hence a minimizer $\sigma \in [0, 1]$ of $A(s)$ exists and the infimum is in fact a minimum. However, if the supports are different then $A(s)$ may be discontinuous at the endpoints of the interval $[0, 1]$. Hence a minimizer need not exist and the r.h.s. in (8) is only an infimum. The proof of our main theorem, Theorem 2.2 below, uses the above classical result for the general case of possibly different support.

We intend to investigate the asymptotic behavior of the Bayesian error probability in the case where the hypotheses are quantum states on $\mathcal{B}(\mathcal{H})$, where $\dim \mathcal{H} = d < \infty$. In order to derive the optimal asymptotic rate exponent we replace the target function in the variational formula (3) or (8), which defines the classical Chernoff bound, by

$$\hat{A}(s) := \mathrm{Tr}\left[\rho_0^{1-s}\rho_1^s\right], \qquad s \in [0, 1].$$

Our main theorem, formulated below, confirms that the logarithm of the infimum of $\hat{A}(s)$ over $[0, 1]$ gives a lower bound on the optimal quantum error exponent.

Theorem 2.2 *[Quantum Chernoff Lower Bound] Let ρ_0, ρ_1 be two density operators representing quantum states on a finite-dimensional complex Hilbert space $\mathcal{H}$. Then any sequence of test projections $\Pi_n \in \mathcal{B}(\mathcal{H}^{\otimes n})$, $n \in \mathbb{N}$, satisfies*

$$\liminf_{n\to\infty} \frac{1}{n} \log Err(\Pi_n) \geq \inf_{0\leq s\leq 1} \log\ Tr\left[\rho_0^{1-s}\rho_1^s\right]. \tag{9}$$

We point out that indeed $\hat{A}(s)$ represents the proper generalization of (4) in the context of symmetric hypothesis testing. As already noted in the Introduction, and as conjectured in [14], it turns out to be achievable, see [1].

3 Proof of the main theorem

We will prove Theorem 2.2 applying the corresponding classical result, Theorem 2.1, to appropriate probability distributions appearing in the general non-commutative setting. Another ingredient is the following lemma.

Lemma 3.1 *Let x, y be two unit vectors in a finite-dimensional Hilbert space $\mathcal{H}$ and $\lambda, \gamma \geq 0$. Then it holds for all projections $\Pi \in \mathcal{B}(\mathcal{H})$*

$$\lambda|\langle \Pi x|y\rangle|^2 + \gamma|\langle (\mathbf{1}-\Pi)x|y\rangle|^2 \geq \frac{1}{2}|\langle x|y\rangle|^2 \min\{\lambda,\gamma\}.$$

Proof. Let ξ, α be vectors in $\mathbb{R}^2$ identified with the respective complex numbers $\langle x|y\rangle$ and $\langle \Pi x|y\rangle$, where $\Pi \in \mathcal{B}(\mathcal{H})$ is a projection. We intend to prove the stronger claim that for all $\alpha \in \mathbb{R}^2$

$$\lambda\|\alpha\|^2 + \gamma\|\xi-\alpha\|^2 \geq \frac{1}{2}\|\xi\|^2 \min\{\lambda,\gamma\}. \tag{10}$$

while Lemma 3.1 claims (10) only for $\alpha \in \Gamma_\xi = \{\langle \Pi x|y\rangle : \Pi \in \mathcal{B}(\mathcal{H})\}$ again identifying $\mathbb{C}$ with $\mathbb{R}^2$.

Let P_ξ be the projection onto the subspace spanned by $\xi \in \mathbb{R}^2$, then

$$\begin{aligned}\lambda\|\alpha\|^2 + \gamma\|\xi-\alpha\|^2 &= \lambda\|P_\xi\alpha\|^2 + \lambda\|(\mathbf{1}-P_\xi)\alpha\|^2 + \gamma\|\xi-P_\xi\alpha\|^2 + \gamma\|(\mathbf{1}-P_\xi)\alpha\|^2 \\ &\geq \lambda\|P_\xi\alpha\|^2 + \gamma\|\xi-P_\xi\alpha\|^2 \\ &= \lambda a^2\|\xi\|^2 + \gamma(1-a)^2\|\xi\|^2, \end{aligned} \tag{11}$$

where in the last line we set $P_\xi\alpha = a\xi$ for some $a \in \mathbb{R}$.

Assume $\|\xi\| > 0$, otherwise the lemma is trivially true. We calculate the minimum of (11) as a function of a taking the derivative. The solution of $(2\lambda a - 2\gamma(1-a))\|\xi\|^2 = 0$ is $a = \frac{\gamma}{\lambda+\gamma}$, which leads to the value of (11) at the minimum

$$\left(\lambda a^2 + \gamma(1-a)^2\right)\|\xi\|^2 = \frac{\lambda\gamma}{\lambda+\gamma}\|\xi\|^2.$$

Finally, the claim (10) follows from the estimate

$$\frac{\lambda\gamma}{\lambda+\gamma}\|\xi\|^2 \geq \frac{\lambda\gamma}{2\max(\lambda,\gamma)}\|\xi\|^2 = \frac{1}{2}\min(\lambda,\gamma)\|\xi\|^2.$$

■

Proof of Theorem 2.2. We will establish

$$\liminf_{n\to\infty}\frac{1}{n}\log\left(Err(\Pi_n)\right) \geq \inf_{0\leq s\leq 1}\log \mathrm{Tr}\ \rho_0^{1-s}\rho_1^s,$$

for any sequence of projections $\Pi_n \in \mathcal{B}(\mathcal{H}^{\otimes n}),\ n \in \mathbb{N}$.

We consider two arbitrary density operators ρ_0, ρ_1 on a finite-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^d$ with spectral representations

$$\rho_0 = \sum_{i=1}^{d} \lambda_i \, |x_i\rangle\langle x_i|, \quad \rho_1 = \sum_{i=1}^{d} \gamma_i \, |y_i\rangle\langle y_i|,$$

i.e. $|x_i\rangle$, $i = 1, \ldots, d$, and $|y_i\rangle$, $i = 1, \ldots, d$ are two orthonormal bases (ONB) of eigenvectors in $\mathbb{C}^d$, and $\lambda_i, \gamma_i \in [0, 1]$ are the respective eigenvalues of ρ_0 and ρ_1.

Let Π be a projection onto a subspace of $\mathbb{C}^d$, then

$$\begin{aligned}
\mathrm{Tr}\,[\Pi\rho] &= \mathrm{Tr}\left[\Pi\left(\sum_{i=1}^{d} \lambda_i \, |x_i\rangle\langle x_i|\right)\right] \\
&= \sum_{i=1}^{d} \lambda_i \, \langle x_i|\Pi x_i\rangle \\
&= \sum_{i=1}^{d} \lambda_i \, \|\Pi x_i\|^2 \\
&= \sum_{i=1}^{d} \lambda_i \sum_{j=1}^{d} |\langle \Pi x_i|y_j\rangle|^2,
\end{aligned}$$

where the third identity is true since Π is a projection and the last one is by Parseval's identity for the ONB $|y_j\rangle$, $j = 1, \ldots, d$. In the same way we obtain

$$\mathrm{Tr}\ \,[(\mathbf{1} - \Pi)\rho_1] = \sum_{j=1}^{d} \gamma_j \sum_{i=1}^{d} |\langle (\mathbf{1} - \Pi)y_j|x_i\rangle|^2.$$

Now in view of the identity $|\langle (\mathbf{1} - \Pi)y_j|x_i\rangle|^2 = |\langle (\mathbf{1} - \Pi)x_i|y_j\rangle|^2$ we have

$$\begin{aligned}
Err(\Pi) &= \frac{1}{2}\left(\mathrm{Tr}\ \,[\rho_0\Pi] + \mathrm{Tr}\ \,[\rho_1(\mathbf{1} - \Pi)]\right) \\
&= \frac{1}{2} \sum_{i,j=1}^{d} \left(\lambda_i |\langle \Pi x_i|y_j\rangle|^2 + \gamma_j |\langle (\mathbf{1} - \Pi)x_i|y_j\rangle|^2\right).
\end{aligned}$$

We introduce the abbreviation

$$Err_{i,j}(\Pi) := \frac{1}{2}\left(\lambda_i |\langle \Pi x_i|y_j\rangle|^2 + \gamma_j |\langle (\mathbf{1} - \Pi)x_i|y_j\rangle|^2\right).$$

It holds

$$\begin{aligned}
Err(\Pi) &= \inf_{\Pi \text{ projection}} \sum_{i,j=1}^{d} Err_{i,j}(\Pi) \geq \sum_{i,j=1}^{d} \inf_{\Pi \text{ projection}} Err_{i,j}(\Pi) \\
&\geq \sum_{i,j=1}^{d} \frac{1}{4} \min\{\lambda_i, \gamma_j\} |\langle x_i|y_j\rangle|^2 \qquad (12)
\end{aligned}$$

where the first inequality is obvious and the second is an application of Lemma 3.1. Note that

$$p_{i,j} := \lambda_i |\langle x_i | y_j \rangle|^2, \quad q_{i,j} := \gamma_j |\langle x_i | y_j \rangle|^2, \qquad i,j = 1, \ldots, d, \tag{13}$$

define probability measures P and Q on d^2 elements, respectively. Indeed

$$\sum_{i,j=1}^{d} p_{i,j} = \sum_{i,j=1}^{d} \lambda_i |\langle x_i | y_j \rangle|^2 = \sum_{i=1}^{d} \lambda_i \|x_i\|^2 = \sum_{i=1}^{d} \lambda_i = 1$$

and similarly for $(q_{i,j})$. Now, inequality (12) may be written

$$Err(\Pi) \geq \frac{1}{4} \sum_{i,j=1}^{d} \min\{p_{i,j}, q_{i.j}\}. \tag{14}$$

Observe according to (6) and (7), the r.h.s. above is up to the factor $1/2$ equal to the classical minimal Bayesian error probability $\Delta(P,Q)$ of discrimination between probability measures P and Q :

$$\frac{1}{2} \sum_{i,j=1}^{k} \min\left\{p_{i,j}, q_{i,j}\right\} = \Delta\left(P, Q\right). \tag{15}$$

Next we consider the case where the quantum hypotheses are $\rho_0^{\otimes n}$ and $\rho_1^{\otimes n}$. Then the corresponding classical probability measures according to (13) are product measures P^n and Q^n, for P, Q corresponding to ρ_0, ρ_1, respectively. Applying inequality (14), (15) and subsequently combining it with the classical result on the Chernoff bound for $\Delta\left(P^n, Q^n\right)$, Theorem 2.1, we obtain for any sequence of projections $\Pi_n \in \mathcal{B}(\mathcal{H}^{\otimes n})$, $n \in \mathbb{N}$,

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log Err(\Pi_n) &\geq \lim_{n\to\infty} \frac{1}{n} \log \left(\frac{1}{2} \Delta(P^n, Q^n) \right) \\ &= \log \left(\inf_{0 \leq s \leq 1} \sum_{i,j=1}^{d} p_{i,j}^{1-s} q_{i,j}^{s} \right). \end{aligned}$$

We finish the proof by verifying

$$\begin{aligned} \sum_{i,j=1}^{d} p_{i,j}^{1-s} q_{i,j}^{s} &= \sum_{i,j=1}^{d} \lambda_i^{1-s} \gamma_j^{s} |\langle x_i | y_j \rangle|^2 = \sum_{i,j=1}^{d} \lambda_i^{1-s} \langle x_i | y_j \rangle \gamma_j^{s} \langle y_j | x_i \rangle \\ &= \mathrm{Tr} \left[\sum_{i,j=1}^{d} \lambda_i^{1-s} |x_i\rangle\langle x_i| \ \gamma_j^{s} |y_j\rangle\langle y_j| \right] \\ &= \mathrm{Tr} \left[\rho_0^{1-s} \rho_1^{s} \right]. \end{aligned}$$

■

4 Appendix

As announced in Section 2 we give a proof for Theorem 2.1 for the general case where the two probability measures involved are allowed to have different supports. As far as possible we follow the proof in the case of same support by Chentsov and Morozova [4].

Proof of Theorem 2.1. *1. Preliminary observations:* Assume that two probability measures P_0, P_1 on a measurable space (Ω, Σ) have support $D_i = \mathrm{supp}(P_i)$, $i = 0, 1$. Denote $B = D_1 \cap D_2$ and for $i = 0, 1$

$$S_i = D_i \setminus B. \tag{16}$$

We introduce the measure $\mu = P_0 + P_1$ and define the densities $p_i = dP_i/d\mu$, $i = 0, 1$. Then clearly $p_1 + p_2 = 1$. We assume the densities and the sets D_i are chosen such that

$$D_i = \{\omega : p_i(\omega) > 0\}, \qquad i = 0, 1,$$

hence

$$B = \{\omega : p_0(\omega) > 0,\ p_1(\omega) > 0\}.$$

Recall the definition of the Hellinger arc of densities for parameter $s \in [0, 1]$:

$$p_s(\omega) = p_1^s(\omega) p_0^{1-s}(\omega) A^{-1}(s),$$

where

$$A(s) = \int p_1^s(\omega) p_0^{1-s}(\omega) \mu(d\omega)$$

is a normalizing factor. Note that for $s = 0$ and $s = 1$ we obtain the initial densities p_0, p_1 respectively, so that $A(0) = A(1) = 1$. However the function $A(s)$ is not continuous in general at the endpoints $0, 1$. Indeed, the integral is over the set B,

$$A(s) = \int_B p_1^s(\omega) p_0^{1-s}(\omega) \mu(d\omega)$$

and by dominated convergence it follows that

$$A_+(0) := \lim_{s \searrow 0} A(s) = \int_B p_0(\omega) \mu(d\omega) = P_0(B),$$
$$A_-(1) := \lim_{s \nearrow 1} A(s) = \int_B p_1(\omega) \mu(d\omega) = P_1(B).$$

Furthermore, observe that for $s \in (0, 1)$ the densities p_s have support B, with limits at the endpoints

$$p_{0+}(\omega) = p_0(\omega)/P_0(B), \quad p_{1-}(\omega) = p_1(\omega)/P_1(B).$$

Hence the corresponding limiting measures are the conditional probability measures

$$P_{0+}(\cdot) = P_0(\cdot|B), \quad P_{1-}(\cdot) = P_1(\cdot|B).$$

If the sample space is restricted to B, the densities p_s, $s \in (0, 1)$, can be written in exponential family form

$$p_s(\omega) = \exp\left(s \log \frac{p_1(\omega)}{p_0(\omega)}\right) p_0(\omega) A^{-1}(s), \qquad \omega \in B. \tag{17}$$

and for $s = 0, 1$ the above holds if $B = D_s$. Also, for $s = 0, 1$, if $B \neq D_s$ then the restriction $p_s|B$ is not a probability density. We denote

$$H(s) = \log A(s), \quad H_+(0) = \log P_0(B), \quad H_-(1) = \log P_1(B).$$

2. Bayesian error probabilities $Err(\lambda_n^*)$ by change of measure to P_s: Recall the form of the optimal test λ_n^* on Ω^n for equiprobable hypothetic densities p_0 and p_1 on Ω:

$$\lambda_n^* = \mathbf{1}\left\{\prod_{j=1}^n p_1(\omega_j) > \prod_{j=1}^n p_0(\omega_j)\right\},$$

where $\omega_1, \ldots, \omega_n$ are n i.i.d. observations. (One may also take "$\geq$" or decide arbitrarily on the "=" set). We partition the set Ω^n into disjoint subsets $S_{0,n}$, $S_{1,n}$ and B_n:

$$S_{0,n} := \{\text{there is } j \in \{1, \ldots, n\} \text{ such that } \omega_j \in S_0\},$$
$$S_{1,n} := \{\text{there is } j \in \{1, \ldots, n\} \text{ such that } \omega_j \in S_1\},$$

where S_i, $i = 0, 1$ were defined in (16). The remaining case is the event

$$B_n := \{\omega^n \in \Omega : \ \omega_j \in B \text{ for } j = 1, \ldots, n\}.$$

Denote $\omega^n = (\omega_1, \ldots, \omega_n) \in \Omega^n$. We have $\lambda_n^*(\omega^n) = 1$ (decision in favor of P_1) if $\omega^n \in S_{1,n}$, i.e. an event happens which excludes P_0. Similarly we have $\lambda_n^*(\omega^n) = 0$ for $\omega^n \in S_{0,n}$. For $\omega^n \in B_n$ define the (normed) log-likelihood ratio by

$$L_n(\omega^n) := n^{-1} \sum_{i=1}^n \log \frac{p_1}{p_0}(\omega_i).$$

Then we can describe the test λ_n^*

$$\lambda_n^*(\omega^n) = \mathbf{1}\{L_n(\omega^n) > 0, \ \omega^n \in B_n\} + \mathbf{1}\{\omega^n \in S_{1,n}\}. \tag{18}$$

Further we define for $i = 0, 1$ functions

$$G_{s,n}^{(i)}(\omega^n) = \mathbf{1}\{\omega^n \in B_n\} \ n^{-1} \sum_{j=1}^n \log \frac{p_i}{p_s}(\omega_j).$$

We note the following relations, for $\omega \in B$:

$$\log \frac{p_0}{p_s}(\omega) = -s \log \frac{p_1}{p_0}(\omega) + H(s), \tag{19}$$
$$\log \frac{p_1}{p_s}(\omega) = (1 - s) \log \frac{p_1}{p_0}(\omega) + H(s). \tag{20}$$

To prove (20), observe that

$$\log \frac{p_1}{p_s} = \log \frac{p_1 A(s)}{\exp\left(s \log \frac{p_1}{p_0}\right) p_0} = \log \frac{p_1}{p_0} - s \log \frac{p_1}{p_0} + H(s) = (1 - s) \log \frac{p_1}{p_0} + H(s).$$

Furthermore it holds

$$\log \frac{p_0}{p_s} = \log \frac{p_0 A(s)}{\exp\left(s \log \frac{p_1}{p_0}\right) p_0} = -s \log \frac{p_1}{p_0} + H(s),$$

which implies (19). As a consequence of (19) and (20) we have for $\omega^n \in B_n$

$$G^{(0)}_{s,n}(\omega^n) = -sL_n(\omega^n) + H(s), \tag{21}$$

$$G^{(1)}_{s,n}(\omega^n) = (1-s)L_n(\omega^n) + H(s). \tag{22}$$

In the sequel we write E_s for expectation under the density p_s and denote by E^n_s the expectation under the product density for the respective basic density p_s. Notice that the test λ^*_n necessarily decides correctly if $\omega^n \in B^c_n = S_{0,n} \cup S_{1,n}$. Thus the minimal Bayesian error probabilities can be expressed for any $s \in (0,1)$ as

$$\begin{aligned} Err(\lambda^*_n) &= E^n_0 \lambda^*_n + E^n_1(1-\lambda^*_n) = E^n_0 \mathbf{1}_{B_n} \lambda^*_n + E^n_1 \mathbf{1}_{B_n}(1-\lambda^*_n) \\ &= E^n_s \lambda^*_n \exp\left(nG^{(0)}_{s,n}\right) + E^n_s(1-\lambda^*_n)\exp\left(nG^{(1)}_{s,n}\right) \qquad (23) \\ &= E^n_s \lambda^*_n \exp\left(-nsL_n + nH(s)\right) + E^n_s(1-\lambda^*_n)\exp\left(n(1-s)L_n + nH(s)\right) \\ &= \exp\left(nH(s)\right)\left\{E^n_s\left(\lambda^*_n \exp\left(-nsL_n\right) + (1-\lambda^*_n)\exp\left(n(1-s)L_n\right)\right)\right\}. \qquad (24) \end{aligned}$$

3. Upper risk bound:

From the expression (18) for λ^*_n we see that for all $\omega^n \in B_n$

$$\lambda^*_n \exp\left(-nsL_n\right) + (1-\lambda^*_n)\exp\left(n(1-s)L_n\right) \le 1$$

so that (24) implies for all $n \in \mathbb{N}$

$$Err(\lambda^*_n) \le \exp\left(nH(s)\right)$$

and hence

$$\frac{1}{n} \log Err(\lambda^*_n) \le H(s).$$

Since $s \in (0,1)$ was arbitrary, and since the bounds $H(0) = H(1) = 0$ are trivial, we obtain

$$\frac{1}{n} \log Err(\lambda^*_n) \le \inf_{0 \le s \le 1} H(s).$$

4. Convexity of $H(s)$ on $(0,1)$: Using the exponential family expression (17) for densities p_s the function $H(s)$ may be written for $s \in (0,1)$

$$H(s) = \log \int_B \exp\left(s \log \frac{p_1(\omega)}{p_0(\omega)}\right) p_0(\omega) d\mu(\omega). \tag{25}$$

It follows

$$H'(s) = \frac{A'(s)}{A(s)} = \frac{\int_B \log \frac{p_1(\omega)}{p_0(\omega)} \exp\left(s \log \frac{p_1(\omega)}{p_0(\omega)}\right) p_0(\omega) d\mu(\omega)}{A(s)},$$

where the fact that $A(s)$ can be differentiated under the integral sign, and the integral is finite for all $s \in (0,1)$ is from the basic theory of exponential families. In the sequel we identify expectation under p_s and its restriction $p_s|B$ for $s \in (0,1)$. We can thus write (for a random variable ω taking values in B)

$$H'(s) = E_s \log \frac{p_1(\omega)}{p_0(\omega)} = E_s \log \frac{p_s(\omega)}{p_0(\omega)} - E_s \log \frac{p_s(\omega)}{p_1(\omega)}. \tag{26}$$

For the second derivative we obtain

$$\begin{aligned} H''(s) &= \frac{A''(s)A(s) - (A'(s))^2}{A^2(s)} \\ &= \frac{\int \left(\log \frac{p_1(\omega)}{p_0(\omega)}\right)^2 \exp\left(s \log \frac{p_1(\omega)}{p_0(\omega)}\right) p_0(\omega) d\mu(\omega)}{A(s)} - \left(H'(s)\right)^2 \\ &= E_s \left(\log \frac{p_1(\omega)}{p_0(\omega)}\right)^2 - \left(E_s \log \frac{p_1(\omega)}{p_0(\omega)}\right)^2 \geq 0, \end{aligned}$$

since the last expression is the variance of the random variable $\log\left(p_1/p_0\right)(\omega)$ under p_s. Thus $H(s)$ is convex on $(0,1)$. There are two cases.

Case 1: There is some $s \in (0,1)$ such that $H''(s) = 0$. Then $\log\left(p_1/p_0\right)(\omega)$ is constant P_s-almost surely. Since all P_s, $s \in (0,1)$, dominate each other, $\left(p_1/p_0\right)(\omega)$ is also constant P_s-almost surely, for all $s \in (0,1)$ and $H''(s) = 0$ for all these s. Hence $H(s)$ is linear on $(0,1)$. Furthermore, each P_s, $s \in (0,1)$, dominates μ on B (i.e. dominates $\mu|B$). It follows

$$\frac{p_1}{p_0}(\omega) = c, \qquad \mu\text{-a.s. on } B,$$

for some constant $c > 0$. In that case

$$P_1(B) = \int_B c dP_0 = cP_0(B)$$

and

$$c = \frac{P_1(B)}{P_0(B)}.$$

This implies

$$\begin{aligned} P_0(\cdot|B) = P_1(\cdot|B) = P_s, &\qquad s \in (0,1), \\ A(s) = \left(P_0(B)\right)^{1-s} \left(P_1(B)\right)^s, &\qquad s \in (0,1). \end{aligned} \tag{27}$$

Case 2: For all $s \in (0,1)$ we have $H''(s) > 0$. Then $H(s)$ is strictly convex on $(0,1)$.

5. Lower risk bound: Since, according to (26), for arbitrary $s \in (0,1)$

$$H'(s) = E_s \log \frac{p_1}{p_0}(\omega)$$

we have in view of (21) and (22) for each $n \in \mathbb{N}$:

$$E_s^n G_{s,n}^{(0)} = -sH'(s) + H(s) =: \gamma_0(s),$$
$$E_s^n G_{s,n}^{(1)} = (1-s)H'(s) + H(s) =: \gamma_1(s).$$

Since $G_{s,n}^{(i)}$ is an i.i.d. average, we have by the Law of Large Numbers as n tends to infinity

$$G_{s,n}^{(0)}(\omega^n) \to \gamma_0(s), \quad G_{s,n}^{(1)}(\omega^n) \to \gamma_1(s),$$

almost surely under P_s. Let $\delta, \eta > 0$ be arbitrary and consider the subsets

$$U_n := \left\{ \omega^n : G_{s,n}^{(i)}(\omega^n) - \gamma_i(s) \geq -\eta,\, i = 0,1 \right\}, \qquad n \in \mathbb{N}.$$

Then, again by the Law of Large Numbers, there is an $n_\delta \in \mathbb{N}$ such that

$$P_s^n(U_n) \geq 1 - \delta, \qquad \text{for all } n \geq n_\delta.$$

Starting with identity (23) we estimate the minimal error probability for $n \geq n_\delta$:

$$\begin{aligned} Err(\lambda_n^*) &= E_s^n \lambda_n^* \exp\left(nG_{s,n}^{(0)}\right) + E_s^n (1-\lambda_n^*) \exp\left(nG_{s,n}^{(1)}\right) \\ &\geq E_s^n \mathbf{1}\{U_n\} \left(\lambda_n^* \exp\left(n\gamma_0(s) - n\eta\right) + (1-\lambda_n^*) \exp\left(n\gamma_1(s) - n\eta\right)\right) \\ &\geq E_s^n \mathbf{1}\{U_n\} \exp\left(n \min\left(\gamma_0(s), \gamma_1(s)\right) - n\eta\right) \\ &\geq (1-\delta) \exp\left(n \min\left(\gamma_0(s), \gamma_1(s)\right) - n\eta\right). \end{aligned}$$

Consequently we have for any sequence of test functions λ_n, $n \in \mathbb{N}$,

$$\liminf_{n\to\infty} n^{-1} \log Err(\lambda_n) \geq \min\left(\gamma_0(s), \gamma_1(s)\right) - \eta.$$

Since η was arbitrary, we obtain for any $s \in (0,1)$

$$\liminf_{n\to\infty} n^{-1} \log Err(\lambda_n) \geq \min\left(\gamma_0(s), \gamma_1(s)\right)$$

and hence

$$\liminf_{n\to\infty} n^{-1} \log Err(\lambda_n) \geq \sup_{0<s<1} \min\left(\gamma_0(s), \gamma_1(s)\right).$$

It remains to show that

$$\sup_{0<s<1} \min\left(\gamma_0(s), \gamma_1(s)\right) \geq \inf_{0\leq s\leq 1} H(s). \tag{28}$$

Recall that the values $H'(s)$ are well defined for $s \in (0,1)$ and that $H(s)$ is convex in that domain. Hence there exist limits

$$H'_+(0) = \lim_{s \searrow 0} H'(s), \quad H'_-(1) = \lim_{s \nearrow 1} H'(s).$$

Observe that the limits are possibly infinite. However due to convexity, only $H'_+(0) = -\infty$ or $H'_+(1) = \infty$ may occur.

Again, in view of the convexity of $H(s)$ on $(0,1)$, the following cases may occur.

a) $H'_+(0) < 0$, $H'_-(1) > 0$

b) $H'_+(0) < 0$, $H'_-(1) \leq 0$

c) $H'_+(0) \geq 0$, $H'_-(1) > 0$

d) $H'_+(0) \geq 0$, $H'_-(1) \leq 0$.

Case a). In this case H cannot be linear, so that due to the above discussion in 4. (involving Case 1, Case 2) it is strictly convex in $(0,1)$. Hence there is a unique minimum of H on $[0,1]$ at some $\sigma \in (0,1)$ with $H'(\sigma) = 0$. We have

$$\gamma_0(\sigma) = \gamma_1(\sigma) = H(\sigma)$$

hence

$$\sup_{0<s<1} \min\left(\gamma_0(s), \gamma_1(s)\right) \geq H(\sigma) = \inf_{0\leq s\leq 1} H(s).$$

Case b). Again due to convexity, the infimum of H on $[0,1]$ is attained (uniquely) at $s \nearrow 1$:

$$\inf_{0\leq s\leq 1} H(s) = \lim_{s \nearrow 1} H(s) = H_-(1).$$

Now for $s \in (0,1)$ we have $H'(s) \leq 0$ and hence

$$\gamma_0(s) = -sH'(s) + H(s) \geq H(s) \geq (1-s)H'(s) + H(s) = \gamma_1(s)$$

which implies

$$\sup_{0<s<1} \min\left(\gamma_0(s), \gamma_1(s)\right) \geq \sup_{0<s<1} \gamma_1(s) \geq \limsup_{s\nearrow 1} \gamma_1(s) \geq H_-(1) = \inf_{0\leq s\leq 1} H(s).$$

Case c). This is symmetric to case b). We obtain

$$\inf_{0\leq s\leq 1} H(s) = H_+(0)$$

and

$$\sup_{0<s<1} \min\left(\gamma_0(s), \gamma_1(s)\right) \geq H_+(0) = \inf_{0\leq s\leq 1} H(s).$$

Now for $s \in (0,1)$ we have $H'(s) \geq 0$ and hence

$$\gamma_1(s) = (1-s)\, H'(s) + H(s) \geq H(s) \geq -sH'(s) + H(s) = \gamma_0(s)$$

which implies

$$\sup_{0<s<1} \min\left(\gamma_0(s), \gamma_1(s)\right) \geq \sup_{0<s<1} \gamma_0(s) \geq \limsup_{s\searrow 0} \gamma_0(s) \geq H_+(0) = \inf_{0\leq s\leq 1} H(s).$$

Case d). Due to convexity we must have $H'_+(0) = H'_-(1) = 0$; then $H(s)$ is constant on $(0,1)$. By (27) we then have $P_0(B) = P_1(B)$ and

$$H(s) = \log P_0(B) = \log P_1(B),\ s \in (0,1).$$

Consequently

$$\gamma_0(s) = \gamma_1(s) = H(s) = \inf_{0\le s\le 1} H(s)$$

and we obtain trivially

$$\sup_{0<s<1} \min\left(\gamma_0(s), \gamma_1(s)\right) \ge \inf_{0\le s\le 1} H(s).$$

We have verified inequality (28) in all cases a)-d). Hence for any sequence of test functions λ_n on Ω^n, $n \in \mathbb{N}$, we have

$$\liminf_{n\to\infty} n^{-1} \log Err(\lambda_n) \ge \liminf_{n\to\infty} n^{-1} \log Err(\lambda_n^*) \ge \inf_{0\le s\le 1} H(s)$$

The upper and lower bounds together complete the proof. ■

Acknowledgments. The first author wishes to thank Ruedi Seiler and the Mathematical Physics group of Technical University Berlin for the opportunity to spend a research semester there.
The second author is grateful to Ruedi Seiler and her colleagues Nihat Ay, Rainer Siegmund-Schultze, Markus Müller and Tyll Krüger for any kind of support. Further, she wants to thank Prof. I. Csiszár and Dénes Petz. The idea to elaborate the topic of the present paper goes back to stimulating discussions in Information Theory Seminar at Rényi Institute Budapest.

References

[1] Audenaert, K.M.R., Calsamiglia, J., Masanes, Ll., Munoz-Tapia, R., Acin, A., Bagan, E., Verstraete, F. The quantum Chernoff bound, *arXiv: quant-ph/0610027*

[2] Bhatia, R. (1997). *Matrix Analysis*. Springer-Verlag

[3] Bjelakovic, I., Deuschel, J.-D., Krüger, T., R. Seiler, R., Siegmund-Schultze, R. and Szkoła, A. (2005). A quantum version of Sanov's Theorem, *Commun. Math. Phys.* **260** 659-671

[4] Chentsov, N. N. and Morozova, E. A. (1991). Natural geometry of families of probability laws. *Results in Science and Technology, Ser. Modern Problems of Mathematics. Fundamental trends*, **83** 133-265, Moscow (in Russian)

[5] Chernoff, H. (1952). A measure of asymptotic efficiency of tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493-507

[6] Cover, T. M., and Thomas, J.A. (1991). *Elements of Information Theory*. John Wiley and Sons.

[7] Gill, R. (2001). Asymptotics in quantum statistics. In: *State of the Art in Probability and Statistics* (A.W. van der Vaart, M. de Gunst, C.A.J. Klaassen, Eds.), IMS Lecture Notes - Monograph Series 36, 255-285. Also at arXiv:math/0405571v1

[8] Helstrom, C.W. (1976). *Quantum Detection and Estimation Theory*. Academic Press, New York

[9] Hiai, F., Petz, D. (1991). The proper formula for relative entropy and its asymptotics in quantum probability. *Comm. Math. Phys.* **143** 99-114

[10] Hoeffding, W. (1965) Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369-401

[11] Holevo, A. S. (1978). Investigation of a general theory of statistical decisions, *Proceedings of Steklov Institute of Mathematics,* vol. 3

[12] Kargin, V. (2005). On the Chernoff bound for efficiency of quantum hypothesis testing, *Ann. Statist.* **33** 959-976

[13] Nielsen, M. and Chuang, I. (2000). *Quantum Computation and Quantum Information,* Cambridge University Press

[14] Nussbaum, M. and Szkoła, A. (2006). A lower bound of Chernoff type for symmetric quantum hypothesis testing. *arXiv: quant-ph/0607216*

[15] Ogawa, T. and Hayashi, M. (2004). On error exponents in quantum hypothesis Testing, *IEEE Trans. Inform. Theory* **50** (6) 1368-1372

[16] Ogawa, T. and Nagaoka, H., (2000). Strong converse and Stein's lemma in quantum hypothesis testing. *IEEE Trans. Inform. Theory* **46** 2428-2433

[17] Sanov, I. N. (1957). On the probability of large deviations of random variables, *Mat. Sbornik* **42** 11-44

Department of Mathematics
Malott Hall
Cornell University
Ithaca NY 14853
e-mail: nussbaum@math.cornell.edu

Max Planck Institute for Mathematics
in the Sciences
Inselstrasse 22
04103 Leipzig, Germany
e-mail: szkola@mis.mpg.de