# 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


# The Chernoff lower bound for symmetric quantum hypothesis testing 

Michael Nussbaum ${ }^{1}$ and Arleta Szkola ${ }^{2}$<br>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

[^0]upper bound $\alpha$ 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 KullbackLeibler 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 $\Omega$ of measures is associated with the null hypothesis. Then the negative Kullback-Leibler distance from the set $\Omega$ 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

$$
\begin{equation*}
\inf _{0 \leq s \leq 1} \log \int p_{0}^{1-s}(\omega) p_{1}^{s}(\omega) \mu(d \omega) \tag{1}
\end{equation*}
$$

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 $\rho_{0}$ and $\rho_{1}$ 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 $\alpha$ 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 noncommutativity 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 $\inf _{0 \leq s \leq 1} \log \operatorname{Tr}\left[\rho_{0}^{1-s} \rho_{1}^{s}\right]$ is a lower bound on the general asymptotic error exponent, $\rho_{0}$ and $\rho_{1}$ 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 $\rho$ is a complex, self-adjoint, positive, $d \times d$ matrix satisfying the normalization condition $\operatorname{Tr}[\rho]=1$, where $\operatorname{Tr}[\cdot]$ 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 $r \leq \mathbf{1}$. Here $\mathbf{1}$ is the unit matrix and $\leq$ 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 $\rho$ 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 $\operatorname{Tr}[\rho r]$. 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 $\operatorname{Tr}[\rho r]$ is the overall probability of rejecting $H_{0}$ when $\rho$ is the true state. Accordingly, $\operatorname{Tr}\left[\rho_{0} r\right]$ is the error probability of first kind and $\operatorname{Tr}\left[(\mathbf{1}-r) \rho_{1}\right]=1-\operatorname{Tr}\left[\rho_{0} r\right]$ is the error probability of second kind. When both $\rho_{0}, \rho_{1}$ and also $r$ are diagonal matrices then the setup reduces to the classical testing problem for two probability measures on an appropriate index set $\Omega,|\Omega|=d$ given by $\rho_{0}, \rho_{1}$ respectively. The same is true when $\rho_{0}$, $\rho_{1}$ have the same set of eigenvectors; then $\rho_{0}, \rho_{1}$ 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 $\Omega$, 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 $\mathcal{B}(\mathcal{H}), \mathcal{B}\left(\mathcal{H}^{\otimes n}\right)$ for the set of complex $d \times d$ or $d n \times d n$ 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 $\rho_{0}, \rho_{1}$, where $\rho^{\otimes n}$ is the $n$-fold tensor product of $\rho$ with itself. (Recall that the tensor product $a \otimes b$ of two matrices is a matrix which consists of blocks $a_{i j} b$, arranged according to the indices $i, j$. Thus $\rho_{0}^{\otimes n}$ is a $d n \times d n$ matrix.) The tests $r_{n}$ now operate on the states $\rho_{0}^{\otimes n}$ and $\rho_{1}^{\otimes n}$, i.e. their dimension is $d n \times d n$, but they need not have tensor product structure. The corresponding Bayesian error probability is

$$
\begin{aligned}
\operatorname{Err}\left(r_{n}\right):= & \frac{1}{2} \operatorname{Tr}\left[\left(r_{n} \rho_{0}^{\otimes n}+\left(\mathbf{1}-r_{n}\right) \rho_{1}^{\otimes n}\right)\right] \\
& =\frac{1}{2}\left(1-\operatorname{Tr}\left[r_{n}\left(\rho_{1}^{\otimes n}-\rho_{0}^{\otimes n}\right)\right]\right)
\end{aligned}
$$

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

$$
\Pi_{n}^{*}:=\operatorname{supp}\left(\rho_{1}^{\otimes n}-\rho_{0}^{\otimes n}\right)_{+},
$$

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 suppa$a_{+}=\sum_{\lambda_{i}>0} E_{i}$. Indeed we have for an arbirtrary test operator in $\mathcal{B}\left(\mathcal{H}^{\otimes n}\right)$

$$
\begin{aligned}
\operatorname{Err}\left(r_{n}\right) & =\frac{1}{2}\left(1-\operatorname{Tr}\left[r_{n}\left(\rho_{1}^{\otimes n}-\rho_{0}^{\otimes n}\right)\right]\right) \\
& \geq \frac{1}{2}\left(1-\sup \left\{\operatorname{Tr}\left[\tilde{r}\left(\rho_{1}^{\otimes n}-\rho_{0}^{\otimes n}\right)\right]: \tilde{r} \in \mathcal{B}\left(\mathcal{H}^{\otimes n}\right) \text { test }\right\}\right) \\
& =\frac{1}{2}\left(1-\sup \left\{\operatorname{Tr}\left[\Pi\left(\rho_{1}^{\otimes n}-\rho_{0}^{\otimes n}\right)\right]: \Pi \in \mathcal{B}\left(\mathcal{H}^{\otimes n}\right) \text { projection }\right\}\right) \\
& =\frac{1}{2}\left(1-\operatorname{Tr}\left[\Pi_{n}^{*}\left(\rho_{1}^{\otimes n}-\rho_{0}^{\otimes n}\right)\right]\right) \\
& =\frac{1}{2}\left(1-\frac{1}{2}\left\|\rho_{1}^{\otimes n}-\rho_{0}^{\otimes n}\right\|_{1}\right)
\end{aligned}
$$

where $\|a\|_{1}=\operatorname{Tr}\left[a_{+}\right]+\operatorname{Tr}\left[a_{+}-a\right]$ 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 $n \rightarrow \infty$ (rate of exponential decay) is the subject of the present paper.

The Holevo-Helstrom tests $\Pi_{n}^{*}$ are non-commutative generalizations of the likelihood ratio tests: if the hypotheses $H_{0}$ and $H_{1}$ correspond to commuting density operators $\rho_{0}$ and $\rho_{1}$ then for all $n \in \mathbb{N}$ the Holevo-Helstrom projections $\Pi_{n}^{*}$ commute with $\rho_{0}^{\otimes n}$ and $\rho_{1}^{\otimes n}$, too. The density operators $\rho_{i}$ may be completely specified by their eigenvalues forming discrete probability measures $P_{i}, i=0,1$ on an appropriate index set $\Omega,|\Omega|=d$ 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 $\Pi_{n}^{*}$ generalizes the indicator function $\lambda_{n}^{*}=1\left\{p_{1}^{n}-p_{0}^{n}>0\right\}$ on $\Omega^{n}$, 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 $\operatorname{Err}(\lambda)$ of a test function $\lambda(0 \leq \lambda \leq 1)$ by

$$
\begin{equation*}
\operatorname{Err}(\lambda)=\frac{1}{2}\left(E_{P_{0}} \lambda+E_{P_{1}}(1-\lambda)\right) . \tag{2}
\end{equation*}
$$

As already mentioned in the Introduction the Bayesian error probability $\operatorname{Err}\left(\lambda_{n}^{*}\right)$ vanishes, as $n \rightarrow \infty$, with a minimal asymptotical rate exponent equal to the Chernoff bound $\delta\left(P_{0}, P_{1}\right)$ :

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \frac{1}{n} \log \operatorname{Err}\left(\lambda_{n}^{*}\right)=\delta\left(P_{0}, P_{1}\right):=\inf _{0 \leq s \leq 1} \log \sum_{x \in \Omega} p_{0}^{1-s}(x) p_{1}^{s}(x) \tag{3}
\end{equation*}
$$

We remark that

$$
\begin{equation*}
\sum_{x \in \Omega} p_{0}^{1-s}(x) p_{1}^{s}(x)=: A(s), \quad s \in[0,1] \tag{4}
\end{equation*}
$$

represent the normalization factors of the parametric family of probability measures

$$
p_{s}(x):=\frac{1}{A(s)} p_{0}^{1-s}(x) p_{1}^{s}(x), \quad x \in \Omega
$$

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 $\left.\mathbb{R}^{|\Omega|}\right)$ 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 $Q_{0}:=P_{0}(\cdot \mid B)$ and $Q_{1}:=P_{1}(\cdot \mid B)$, 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):

$$
\begin{equation*}
\delta\left(P_{0}, P_{1}\right)=\inf _{s \in[0,1]}\left(-(1-s) K\left(Q_{s} \| Q_{0}\right)-s K\left(Q_{s} \| Q_{1}\right)+\log \pi_{0}^{1-s} \pi_{1}^{s}\right) \tag{5}
\end{equation*}
$$

where $Q_{s}$ denotes the conditional probability $P_{s}(\cdot \mid B)$, 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 $-(1-s) K\left(P_{s} \| P_{0}\right)-s K\left(P_{s} \| P_{1}\right)$. What is remarkable is that in this case we have

$$
\delta\left(P_{0}, P_{1}\right)=-K\left(P_{\sigma} \| P_{0}\right)=-K\left(P_{\sigma} \| P_{1}\right)
$$

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 $H^{\prime}(\sigma)=0$, which is equivalent to $K\left(Q_{\sigma} \| Q_{0}\right)-K\left(Q_{\sigma} \| Q_{1}\right)=\log \left(\pi_{0} / \pi_{1}\right)$, then

$$
\delta\left(P_{0}, P_{1}\right)=-K\left(Q_{\sigma} \| P_{0}\right)+\log \pi_{0}=-K\left(Q_{\sigma} \| P_{1}\right)+\log \pi_{1} .
$$

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 $(\Omega, \Sigma)$, not necessarily finite. Consider the Bayesian error probability of discrimination between $P_{0}, P_{1}$ by means of test functions $0 \leq \lambda \leq 1$ :

$$
\begin{equation*}
\Delta\left(P_{0}, P_{1}\right):=\inf _{\lambda \text { test function }} \operatorname{Err}(\lambda) \tag{6}
\end{equation*}
$$

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

$$
\begin{equation*}
\Delta\left(P_{0}, P_{1}\right)=\operatorname{Err}\left(\lambda^{*}\right)=\frac{1}{2} \int \min \left(p_{0}, p_{1}\right) d \mu \tag{7}
\end{equation*}
$$

Theorem 2.1 Let $P_{0}, P_{1}$ be two probability measures on $(\Omega, \Sigma)$. 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

$$
\begin{equation*}
\lim _{n \rightarrow \infty} n^{-1} \log \Delta\left(P_{0}^{n}, P_{1}^{n}\right)=\inf _{0 \leq s \leq 1} \log \int p_{1}^{s} p_{0}^{1-s} d \mu \tag{8}
\end{equation*}
$$

where $p_{i}=d P_{i} / d \mu, i=0,1, \mu:=P_{0}+P_{1}$.

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 $\operatorname{dim} \mathcal{H}=d<\infty$. 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

$$
\hat{A}(s):=\operatorname{Tr}\left[\rho_{0}^{1-s} \rho_{1}^{s}\right], \quad s \in[0,1]
$$

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 $\rho_{0}, \rho_{1}$ be two density operators representing quantum states on a finite-dimensional complex Hilbert space $\mathcal{H}$. Then any sequence of test projections $\Pi_{n} \in \mathcal{B}\left(\mathcal{H}^{\otimes n}\right)$, $n \in \mathbb{N}$, satisfies

$$
\begin{equation*}
\liminf _{n \rightarrow \infty} \frac{1}{n} \log \operatorname{Err}\left(\Pi_{n}\right) \geq \inf _{0 \leq s \leq 1} \log \operatorname{Tr}\left[\rho_{0}^{1-s} \rho_{1}^{s}\right] \tag{9}
\end{equation*}
$$

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 $\Pi \in \mathcal{B}(\mathcal{H})$

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

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

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

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

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

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

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

$$
\liminf _{n \rightarrow \infty} \frac{1}{n} \log \left(\operatorname{Err}\left(\Pi_{n}\right)\right) \geq \inf _{0 \leq s \leq 1} \log \operatorname{Tr} \rho_{0}^{1-s} \rho_{1}^{s}
$$

for any sequence of projections $\Pi_{n} \in \mathcal{B}\left(\mathcal{H}^{\otimes n}\right), n \in \mathbb{N}$.
We consider two arbitrary density operators $\rho_{0}, \rho_{1}$ on a finite-dimensional Hilbert space $\mathcal{H}=$ $\mathbb{C}^{d}$ with spectral representations

$$
\rho_{0}=\sum_{i=1}^{d} \lambda_{i}\left|x_{i}\right\rangle\left\langle x_{i}\right|, \quad \rho_{1}=\sum_{i=1}^{d} \gamma_{i}\left|y_{i}\right\rangle\left\langle y_{i}\right|,
$$

i.e. $\left|x_{i}\right\rangle, i=1, \ldots, d$, and $\left|y_{i}\right\rangle, i=1, \ldots, d$ are two orthonormal bases (ONB) of eigenvectors in $\mathbb{C}^{d}$, and $\lambda_{i}, \gamma_{i} \in[0,1]$ are the respective eigenvalues of $\rho_{0}$ and $\rho_{1}$.

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

$$
\begin{aligned}
\operatorname{Tr}[\Pi \rho] & =\operatorname{Tr}\left[\Pi\left(\sum_{i=1}^{d} \lambda_{i}\left|x_{i}\right\rangle\left\langle x_{i}\right|\right)\right] \\
& =\sum_{i=1}^{d} \lambda_{i}\left\langle x_{i} \mid \Pi x_{i}\right\rangle \\
& =\sum_{i=1}^{d} \lambda_{i}\left\|\Pi x_{i}\right\|^{2} \\
& =\sum_{i=1}^{d} \lambda_{i} \sum_{j=1}^{d}\left|\left\langle\Pi x_{i} \mid y_{j}\right\rangle\right|^{2}
\end{aligned}
$$

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

$$
\operatorname{Tr}\left[(\mathbf{1}-\Pi) \rho_{1}\right]=\sum_{j=1}^{d} \gamma_{j} \sum_{i=1}^{d}\left|\left\langle(\mathbf{1}-\Pi) y_{j} \mid x_{i}\right\rangle\right|^{2} .
$$

Now in view of the identity $\left|\left\langle(\mathbf{1}-\Pi) y_{j} \mid x_{i}\right\rangle\right|^{2}=\left|\left\langle(\mathbf{1}-\Pi) x_{i} \mid y_{j}\right\rangle\right|^{2}$ we have

$$
\begin{aligned}
\operatorname{Err}(\Pi) & =\frac{1}{2}\left(\operatorname{Tr}\left[\rho_{0} \Pi\right]+\operatorname{Tr}\left[\rho_{1}(\mathbf{1}-\Pi)\right]\right) \\
& =\frac{1}{2} \sum_{i, j=1}^{d}\left(\lambda_{i}\left|\left\langle\Pi x_{i} \mid y_{j}\right\rangle\right|^{2}+\gamma_{j}\left|\left\langle(\mathbf{1}-\Pi) x_{i} \mid y_{j}\right\rangle\right|^{2}\right) .
\end{aligned}
$$

We introduce the abbreviation

$$
\operatorname{Err}_{i, j}(\Pi):=\frac{1}{2}\left(\lambda_{i}\left|\left\langle\Pi x_{i} \mid y_{j}\right\rangle\right|^{2}+\gamma_{j}\left|\left\langle(\mathbf{1}-\Pi) x_{i} \mid y_{j}\right\rangle\right|^{2}\right) .
$$

It holds

$$
\begin{align*}
\operatorname{Err}(\Pi) & =\inf _{\Pi \text { projection }} \sum_{i, j=1}^{d} \operatorname{Err}_{i, j}(\Pi) \geq \sum_{i, j=1}^{d} \inf _{\Pi \text { projection }} \operatorname{Err}_{i, j}(\Pi) \\
& \geq \sum_{i, j=1}^{d} \frac{1}{4} \min \left\{\lambda_{i}, \gamma_{j}\right\}\left|\left\langle x_{i} \mid y_{j}\right\rangle\right|^{2} \tag{12}
\end{align*}
$$

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

$$
\begin{equation*}
p_{i, j}:=\lambda_{i}\left|\left\langle x_{i} \mid y_{j}\right\rangle\right|^{2}, \quad q_{i, j}:=\gamma_{j}\left|\left\langle x_{i} \mid y_{j}\right\rangle\right|^{2}, \quad i, j=1, \ldots, d, \tag{13}
\end{equation*}
$$

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

$$
\sum_{i, j=1}^{d} p_{i, j}=\sum_{i, j=1}^{d} \lambda_{i}\left|\left\langle x_{i} \mid y_{j}\right\rangle\right|^{2}=\sum_{i=1}^{d} \lambda_{i}\left\|x_{i}\right\|^{2}=\sum_{i=1}^{d} \lambda_{i}=1
$$

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

$$
\begin{equation*}
\operatorname{Err}(\Pi) \geq \frac{1}{4} \sum_{i, j=1}^{d} \min \left\{p_{i, j}, q_{i \cdot j}\right\} . \tag{14}
\end{equation*}
$$

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 $Q$ :

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

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 $\rho_{0}, \rho_{1}$, 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 $\Pi_{n} \in \mathcal{B}\left(\mathcal{H}^{\otimes n}\right), n \in \mathbb{N}$,

$$
\begin{aligned}
\liminf _{n \rightarrow \infty} \frac{1}{n} \log \operatorname{Err}\left(\Pi_{n}\right) & \geq \lim _{n \rightarrow \infty} \frac{1}{n} \log \left(\frac{1}{2} \Delta\left(P^{n}, Q^{n}\right)\right) \\
& =\log \left(\inf _{0 \leq s \leq 1} \sum_{i, j=1}^{d} p_{i, j}^{1-s} q_{i, j}^{s}\right)
\end{aligned}
$$

We finish the proof by verifying

$$
\begin{aligned}
\sum_{i, j=1}^{d} p_{i, j}^{1-s} q_{i, j}^{s} & =\sum_{i, j=1}^{d} \lambda_{i}^{1-s} \gamma_{j}^{s}\left|\left\langle x_{i} \mid y_{j}\right\rangle\right|^{2}=\sum_{i, j=1}^{d} \lambda_{i}^{1-s}\left\langle x_{i} \mid y_{j}\right\rangle \gamma_{j}^{s}\left\langle y_{j} \mid x_{i}\right\rangle \\
& =\operatorname{Tr}\left[\sum_{i, j=1}^{d} \lambda_{i}^{1-s}\left|x_{i}\right\rangle\left\langle x_{i}\right| \gamma_{j}^{s}\left|y_{j}\right\rangle\left\langle y_{j}\right|\right] \\
& =\operatorname{Tr}\left[\rho_{0}^{1-s} \rho_{1}^{s}\right] .
\end{aligned}
$$

## 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 $(\Omega, \Sigma)$ have support $D_{i}=\operatorname{supp}\left(P_{i}\right), i=0,1$. Denote $B=$ $D_{1} \cap D_{2}$ and for $i=0,1$

$$
\begin{equation*}
S_{i}=D_{i} \backslash B . \tag{16}
\end{equation*}
$$

We introduce the measure $\mu=P_{0}+P_{1}$ and define the densities $p_{i}=d P_{i} / d \mu, i=0,1$. Then clearly $p_{1}+p_{2}=1$. We assume the densities and the sets $D_{i}$ are chosen such that

$$
D_{i}=\left\{\omega: p_{i}(\omega)>0\right\}, \quad i=0,1,
$$

hence

$$
B=\left\{\omega: p_{0}(\omega)>0, p_{1}(\omega)>0\right\} .
$$

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

$$
\begin{aligned}
& A_{+}(0):=\lim _{s \backslash 0} A(s)=\int_{B} p_{0}(\omega) \mu(d \omega)=P_{0}(B) \\
& A_{-}(1):=\lim _{s \nearrow 1} A(s)=\int_{B} p_{1}(\omega) \mu(d \omega)=P_{1}(B)
\end{aligned}
$$

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

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

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

$$
\begin{equation*}
p_{s}(\omega)=\exp \left(s \log \frac{p_{1}(\omega)}{p_{0}(\omega)}\right) p_{0}(\omega) A^{-1}(s), \quad \omega \in B \tag{17}
\end{equation*}
$$

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 $p_{s} \mid B$ 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 $\operatorname{Err}\left(\lambda_{n}^{*}\right)$ by change of measure to $P_{s}$ : Recall the form of the optimal test $\lambda_{n}^{*}$ on $\Omega^{n}$ for equiprobable hypothetic densities $p_{0}$ and $p_{1}$ on $\Omega$ :

$$
\lambda_{n}^{*}=\mathbf{1}\left\{\prod_{j=1}^{n} p_{1}\left(\omega_{j}\right)>\prod_{j=1}^{n} p_{0}\left(\omega_{j}\right)\right\}
$$

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 $\Omega^{n}$ into disjoint subsets $S_{0, n}, S_{1, n}$ and $B_{n}$ :

$$
\begin{aligned}
& S_{0, n}:=\left\{\text { there is } j \in\{1, \ldots, n\} \text { such that } \omega_{j} \in S_{0}\right\}, \\
& S_{1, n}:=\left\{\text { there is } j \in\{1, \ldots, n\} \text { such that } \omega_{j} \in S_{1}\right\},
\end{aligned}
$$

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

$$
B_{n}:=\left\{\omega^{n} \in \Omega: \omega_{j} \in B \text { for } j=1, \ldots, n\right\}
$$

Denote $\omega^{n}=\left(\omega_{1}, \ldots, \omega_{n}\right) \in \Omega^{n}$. We have $\lambda_{n}^{*}\left(\omega^{n}\right)=1$ (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 $\lambda_{n}^{*}\left(\omega^{n}\right)=0$ for $\omega^{n} \in S_{0, n}$. For $\omega^{n} \in B_{n}$ define the (normed) log-likelihood ratio by

$$
L_{n}\left(\omega^{n}\right):=n^{-1} \sum_{i=1}^{n} \log \frac{p_{1}}{p_{0}}\left(\omega_{i}\right) .
$$

Then we can describe the test $\lambda_{n}^{*}$

$$
\begin{equation*}
\lambda_{n}^{*}\left(\omega^{n}\right)=\mathbf{1}\left\{L_{n}\left(\omega^{n}\right)>0, \omega^{n} \in B_{n}\right\}+\mathbf{1}\left\{\omega^{n} \in S_{1, n}\right\} \tag{18}
\end{equation*}
$$

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

$$
G_{s, n}^{(i)}\left(\omega^{n}\right)=\mathbf{1}\left\{\omega^{n} \in B_{n}\right\} n^{-1} \sum_{j=1}^{n} \log \frac{p_{i}}{p_{s}}\left(\omega_{j}\right) .
$$

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

$$
\begin{align*}
& \log \frac{p_{0}}{p_{s}}(\omega)=-s \log \frac{p_{1}}{p_{0}}(\omega)+H(s)  \tag{19}\\
& \log \frac{p_{1}}{p_{s}}(\omega)=(1-s) \log \frac{p_{1}}{p_{0}}(\omega)+H(s) \tag{20}
\end{align*}
$$

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 $\omega^{n} \in B_{n}$

$$
\begin{align*}
& G_{s, n}^{(0)}\left(\omega^{n}\right)=-s L_{n}\left(\omega^{n}\right)+H(s),  \tag{21}\\
& G_{s, n}^{(1)}\left(\omega^{n}\right)=(1-s) L_{n}\left(\omega^{n}\right)+H(s) . \tag{22}
\end{align*}
$$

In the sequel we write $E_{s}$ for expectation under the density $p_{s}$ and denote by $E_{s}^{n}$ the expectation under the product density for the respective basic density $p_{s}$. Notice that the test $\lambda_{n}^{*}$ necessarily decides correctly if $\omega^{n} \in B_{n}^{c}=S_{0, n} \cup S_{1, n}$. Thus the minimal Bayesian error probabilities can be expressed for any $s \in(0,1)$ as

$$
\begin{align*}
\operatorname{Err}\left(\lambda_{n}^{*}\right) & =E_{0}^{n} \lambda_{n}^{*}+E_{1}^{n}\left(1-\lambda_{n}^{*}\right)=E_{0}^{n} \mathbf{1}_{B_{n}} \lambda_{n}^{*}+E_{1}^{n} \mathbf{1}_{B_{n}}\left(1-\lambda_{n}^{*}\right) \\
& =E_{s}^{n} \lambda_{n}^{*} \exp \left(n G_{s, n}^{(0)}\right)+E_{s}^{n}\left(1-\lambda_{n}^{*}\right) \exp \left(n G_{s, n}^{(1)}\right)  \tag{23}\\
& =E_{s}^{n} \lambda_{n}^{*} \exp \left(-n s L_{n}+n H(s)\right)+E_{s}^{n}\left(1-\lambda_{n}^{*}\right) \exp \left(n(1-s) L_{n}+n H(s)\right) \\
& =\exp (n H(s))\left\{E_{s}^{n}\left(\lambda_{n}^{*} \exp \left(-n s L_{n}\right)+\left(1-\lambda_{n}^{*}\right) \exp \left(n(1-s) L_{n}\right)\right)\right\} . \tag{24}
\end{align*}
$$

## 3. Upper risk bound:

From the expression (18) for $\lambda_{n}^{*}$ we see that for all $\omega^{n} \in B_{n}$

$$
\lambda_{n}^{*} \exp \left(-n s L_{n}\right)+\left(1-\lambda_{n}^{*}\right) \exp \left(n(1-s) L_{n}\right) \leq 1
$$

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

$$
\operatorname{Err}\left(\lambda_{n}^{*}\right) \leq \exp (n H(s))
$$

and hence

$$
\frac{1}{n} \log \operatorname{Err}\left(\lambda_{n}^{*}\right) \leq H(s) .
$$

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

$$
\frac{1}{n} \log \operatorname{Err}\left(\lambda_{n}^{*}\right) \leq \inf _{0 \leq s \leq 1} H(s) .
$$

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 $s \in(0,1)$

$$
\begin{equation*}
H(s)=\log \int_{B} \exp \left(s \log \frac{p_{1}(\omega)}{p_{0}(\omega)}\right) p_{0}(\omega) d \mu(\omega) . \tag{25}
\end{equation*}
$$

It follows

$$
H^{\prime}(s)=\frac{A^{\prime}(s)}{A(s)}=\frac{\int_{B} \log \frac{p_{1}(\omega)}{p_{0}(\omega)} \exp \left(s \log \frac{p_{1}(\omega)}{p_{0}(\omega)}\right) p_{0}(\omega) d \mu(\omega)}{A(s)}
$$

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 $p_{s} \mid B$ for $s \in(0,1)$. We can thus write (for a random variable $\omega$ taking values in $B$ )

$$
\begin{equation*}
H^{\prime}(s)=E_{s} \log \frac{p_{1}(\omega)}{p_{0}(\omega)}=E_{s} \log \frac{p_{s}(\omega)}{p_{0}(\omega)}-E_{s} \log \frac{p_{s}(\omega)}{p_{1}(\omega)} . \tag{26}
\end{equation*}
$$

For the second derivative we obtain

$$
\begin{aligned}
H^{\prime \prime}(s) & =\frac{A^{\prime \prime}(s) A(s)-\left(A^{\prime}(s)\right)^{2}}{A^{2}(s)} \\
& =\frac{\int\left(\log \frac{p_{1}(\omega)}{p_{0}(\omega)}\right)^{2} \exp \left(s \log \frac{p_{1}(\omega)}{p_{0}(\omega)}\right) p_{0}(\omega) d \mu(\omega)}{A(s)}-\left(H^{\prime}(s)\right)^{2} \\
& =E_{s}\left(\log \frac{p_{1}(\omega)}{p_{0}(\omega)}\right)^{2}-\left(E_{s} \log \frac{p_{1}(\omega)}{p_{0}(\omega)}\right)^{2} \geq 0,
\end{aligned}
$$

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 $H^{\prime \prime}(s)=0$. 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 $H^{\prime \prime}(s)=0$ for all these $s$. Hence $H(s)$ is linear on $(0,1)$. Furthermore, each $P_{s}, s \in(0,1)$, dominates $\mu$ on $B$ (i.e. dominates $\left.\mu \mid B\right)$. It follows

$$
\frac{p_{1}}{p_{0}}(\omega)=c, \quad \mu \text {-a.s. on } B,
$$

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

$$
P_{1}(B)=\int_{B} c d P_{0}=c P_{0}(B)
$$

and

$$
c=\frac{P_{1}(B)}{P_{0}(B)}
$$

This implies

$$
\begin{align*}
P_{0}(\cdot \mid B) & =P_{1}(\cdot \mid B)=P_{s}, \quad s \in(0,1), \\
A(s) & =\left(P_{0}(B)\right)^{1-s}\left(P_{1}(B)\right)^{s}, \quad s \in(0,1) . \tag{27}
\end{align*}
$$

Case 2: For all $s \in(0,1)$ we have $H^{\prime \prime}(s)>0$. Then $H(s)$ is strictly convex on $(0,1)$.
5. Lower risk bound: Since, according to (26), for arbitrary $s \in(0,1)$

$$
H^{\prime}(s)=E_{s} \log \frac{p_{1}}{p_{0}}(\omega)
$$

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

$$
\begin{aligned}
& E_{s}^{n} G_{s, n}^{(0)}=-s H^{\prime}(s)+H(s)=: \gamma_{0}(s) \\
& E_{s}^{n} G_{s, n}^{(1)}=(1-s) H^{\prime}(s)+H(s)=: \gamma_{1}(s)
\end{aligned}
$$

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

$$
G_{s, n}^{(0)}\left(\omega^{n}\right) \rightarrow \gamma_{0}(s), \quad G_{s, n}^{(1)}\left(\omega^{n}\right) \rightarrow \gamma_{1}(s),
$$

almost surely under $P_{s}$. Let $\delta, \eta>0$ be arbitrary and consider the subsets

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

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

$$
P_{s}^{n}\left(U_{n}\right) \geq 1-\delta, \quad \text { for all } n \geq n_{\delta} .
$$

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

$$
\begin{aligned}
\operatorname{Err}\left(\lambda_{n}^{*}\right) & =E_{s}^{n} \lambda_{n}^{*} \exp \left(n G_{s, n}^{(0)}\right)+E_{s}^{n}\left(1-\lambda_{n}^{*}\right) \exp \left(n G_{s, n}^{(1)}\right) \\
& \geq E_{s}^{n} 1\left\{U_{n}\right\}\left(\lambda_{n}^{*} \exp \left(n \gamma_{0}(s)-n \eta\right)+\left(1-\lambda_{n}^{*}\right) \exp \left(n \gamma_{1}(s)-n \eta\right)\right) \\
& \geq E_{s}^{n} 1\left\{U_{n}\right\} \exp \left(n \min \left(\gamma_{0}(s), \gamma_{1}(s)\right)-n \eta\right) \\
& \geq(1-\delta) \exp \left(n \min \left(\gamma_{0}(s), \gamma_{1}(s)\right)-n \eta\right)
\end{aligned}
$$

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

$$
\liminf _{n \rightarrow \infty} n^{-1} \log \operatorname{Err}\left(\lambda_{n}\right) \geq \min \left(\gamma_{0}(s), \gamma_{1}(s)\right)-\eta
$$

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

$$
\liminf _{n \rightarrow \infty} n^{-1} \log \operatorname{Err}\left(\lambda_{n}\right) \geq \min \left(\gamma_{0}(s), \gamma_{1}(s)\right)
$$

and hence

$$
\liminf _{n \rightarrow \infty} n^{-1} \log \operatorname{Err}\left(\lambda_{n}\right) \geq \sup _{0<s<1} \min \left(\gamma_{0}(s), \gamma_{1}(s)\right)
$$

It remains to show that

$$
\begin{equation*}
\sup _{0<s<1} \min \left(\gamma_{0}(s), \gamma_{1}(s)\right) \geq \inf _{0 \leq s \leq 1} H(s) . \tag{28}
\end{equation*}
$$

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

$$
H_{+}^{\prime}(0)=\lim _{s \searrow 0} H^{\prime}(s), \quad H_{-}^{\prime}(1)=\lim _{s \nearrow 1} H^{\prime}(s) .
$$

Observe that the limits are possibly infinite. However due to convexity, only $H_{+}^{\prime}(0)=-\infty$ or $H_{+}^{\prime}(1)=\infty$ may occur.

Again, in view of the convexity of $H(s)$ on $(0,1)$, the following cases may occur.
a) $H_{+}^{\prime}(0)<0, H_{-}^{\prime}(1)>0$
b) $H_{+}^{\prime}(0)<0, H_{-}^{\prime}(1) \leq 0$
c) $H_{+}^{\prime}(0) \geq 0, H_{-}^{\prime}(1)>0$
d) $H_{+}^{\prime}(0) \geq 0, H_{-}^{\prime}(1) \leq 0$.

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 $H^{\prime}(\sigma)=0$. 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 $H^{\prime}(s) \leq 0$ and hence

$$
\gamma_{0}(s)=-s H^{\prime}(s)+H(s) \geq H(s) \geq(1-s) H^{\prime}(s)+H(s)=\gamma_{1}(s)
$$

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 $H^{\prime}(s) \geq 0$ and hence

$$
\gamma_{1}(s)=(1-s) H^{\prime}(s)+H(s) \geq H(s) \geq-s H^{\prime}(s)+H(s)=\gamma_{0}(s)
$$

which implies

$$
\sup _{0<s<1} \min \left(\gamma_{0}(s), \gamma_{1}(s)\right) \geq \sup _{0<s<1} \gamma_{0}(s) \geq \limsup _{s \backslash 0} \gamma_{0}(s) \geq H_{+}(0)=\inf _{0 \leq s \leq 1} H(s)
$$

Case d). Due to convexity we must have $H_{+}^{\prime}(0)=H_{-}^{\prime}(1)=0$; 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

$$
\gamma_{0}(s)=\gamma_{1}(s)=H(s)=\inf _{0 \leq s \leq 1} H(s)
$$

and we obtain trivially

$$
\sup _{0<s<1} \min \left(\gamma_{0}(s), \gamma_{1}(s)\right) \geq \inf _{0 \leq s \leq 1} H(s) .
$$

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

$$
\liminf _{n \rightarrow \infty} n^{-1} \log \operatorname{Err}\left(\lambda_{n}\right) \geq \liminf _{n \rightarrow \infty} n^{-1} \log \operatorname{Err}\left(\lambda_{n}^{*}\right) \geq \inf _{0 \leq s \leq 1} H(s)
$$

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 SiegmundSchultze, 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: quantph/0610027
[2] Bhatia, R. (1997). Matrix Analysis. Springer-Verlag
[3] Bjelakovic, I., Deuschel, J.-D., Krüger, T., R. Seiler, R., SiegmundSchultze, R. and Szkola, 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 Szkola, 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


[^0]:    ${ }^{1}$ 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.

