Principal Investigator: M. Nussbaum, Department of Mathematics, Cornell University, Ithaca, NY

New Horizons in Statistical Decision Theory

Project description

Results from Prior NSF Support

(a) (i) NSF award DMS-0805632, amount $240,000, period of support 7/1/08-6/31/11
(ii) NSF award DMS-1106460, amount $368,972, period of support 7/1/11-6/31/14

(b) Title: (i) Asymptotic Methods in Quantum Statistics, (ii) Asymptotic Inference for Locally Stationary Processes

(c) Summary of results of the completed work

Intellectual merit. A point of departure and a unifying topic for both previous awards has been the concept of asymptotic equivalence of statistical experiments. An experiment $\mathcal{E}$ is a collection of probability measures $(P_\theta, \theta \in \Theta)$ on a (measurable) sample space $(\mathcal{X}, \mathcal{B}_\mathcal{X})$. When studying experiments with the same parameter space $\Theta$, but different sample spaces, one is led to ask the question of how informative different experiments are with respect to the unknown parameter $\theta \in \Theta$. The primary example of an experiment which is equally informative to $\mathcal{E}$ is an $\mathcal{F} = \mathcal{E}_T$ generated by a sufficient statistic $T$ in $\mathcal{E}$. More generally, for $\mathcal{E}, \mathcal{F}$ given on different sample spaces, one defines an informativity distance using Markov kernels $K$ as follows: let $\mathcal{F} = (Q_\theta, \theta \in \Theta)$ define $\delta(\mathcal{E}, \mathcal{F}) = \inf_K \sup_{\psi \in \Theta} \|Q_\theta - K P_\theta\|_1$ where $\|\cdot\|_1$ is total variation distance. Then Le Cam’s $\Delta$-distance between experiments (or deficiency distance) is defined as $\Delta(\mathcal{E}, \mathcal{F}) = \max(\delta(\mathcal{E}, \mathcal{F}), \delta(\mathcal{F}, \mathcal{E}))$. Two experiments $\mathcal{E}, \mathcal{F}$ are equivalent if $\Delta(\mathcal{E}, \mathcal{F}) = 0$; and indeed the $\mathcal{E}_T$ generated from $\mathcal{E}$ by a sufficient statistic $T$ is then equivalent. Two sequences $\mathcal{E}_n, \mathcal{F}_n$ are called asymptotically equivalent if $\Delta(\mathcal{E}_n, \mathcal{F}_n) \to 0$. This notion has been shown by Le Cam [82] to provide the adequate framework for decision theoretic limit theorems. For finite dimensional parameter spaces $\Theta$, sequences of rescaled (localized) experiments around a given $\theta_0 \in \Theta$ converge to a Gaussian limit (local asymptotic normality, LAN). That fundamental concept has permeated many branches of statistics, up to the level of applications such as clinical trials. In conjunction with classical Fisher information, it is widely applied to create rigorous benchmarks for optimal procedures, cf. van der Vaart [132].

An extension to infinite dimensional parameter spaces and to a global approximation (without localization), in the context of i.i.d. experiments has been given in Nussbaum [102]: $n$ i.i.d. observations $y_i$ with density $f$ on the unit interval are asymptotically equivalent to a white noise model

$$dZ_t = f^{1/2}(t)dt + \frac{1}{2}n^{-1/2}dW_t, \ t \in [0, 1]$$

(1)

if $f \in \Sigma_{\alpha,M}$, a Hölder class with smoothness $\alpha > 1/2$. Subsequent developments (Carter [22], [23], [24], Brown et al. [12], [13], Delattre and Hoffmann [34], Wang [133], Rohde [118], Grama and Neumann [50], Dalalyan and Reiss [31], [32], Low and Zhou [86], Reiss [116], [117], Cai and Zhou [19], Brown et al [11], Meister [91], Meister and Reiss [92], Wang [134]), justify a claim that asymptotic equivalence theory is emerging as a recognizable research area in statistics.

[R1] Asymptotic equivalence of spectral density estimation and white noise. Assume a sample $y^{(n)} = (y(1), \ldots, y(n))'$ from a real Gaussian stationary sequence $y(t)$ with zero mean, autocovariance function $\gamma_h = E y(t)y(t + h)$ and spectral density $f$ on $[-\pi, \pi]$. Define a nonparametric
set $\Sigma_{\alpha,M} = H^{\alpha}(M) \cap F_M$, where $H^{\alpha}(M)$ is an ellipsoid in terms of autocovariances $\gamma_h$, $h \in \mathbb{Z}$, with smoothness coefficient $\alpha$ and $F_M$ is the set of even positive functions $f$ on $[-\pi, \pi]$ such that $|\log f| \leq M$. Then it is shown that observations $y_n$ with spectral density $f$ are asymptotically equivalent to a white noise model of type (1) above with $\log f$ in place of $f^{1/2}$, provided $f \in \Sigma_{\alpha,M}$ for some $M > 0$ and $\alpha > 1/2$. This represents the nonparametric asymptotic equivalence version of the classical LAN property for parametric models $\{f_\theta, \theta \in \Theta\}$ of spectral densities, cf. Davies [33], Dzhaparidze [37], Taniguchi and Kakizawa [125].

[R2] Large deviations in testing simple hypotheses for locally stationary Gaussian processes (LSGP). This paper is an outgrowth of the thesis research with Ph. D. student Inder Tecuapetla Gómez. For the basic theory of LSGP cf. Dahlhaus and Polonik [28] and references therein. The study of large deviation techniques for these models, which was originally aimed at acquiring technical tools for the treatment of nonparametric asymptotic equivalence, as a byproduct generated some new results on hypothesis testing. A large deviation result for the log-likelihood ratio for testing simple hypotheses in LSGP is derived. This allows to find explicitly the rates of exponential decay of the error probabilities of type I and type II for Neyman-Pearson tests. Furthermore, as an application the analog of classical results on asymptotic efficiency of tests such as Stein’s lemma and the Chernoff bound is obtained, as well as the more general Hoeffding bound concerning best possible joint exponential rates for the two error probabilities. The thesis also treats the asymptotic equivalence problem for LSGP, but a solution is obtained only for finite dimensional parameter spaces using Le Cam’s connection theorem [81].

[R3] Sharp adaptive nonparametric testing for Sobolev ellipsoids. This result is an outgrowth of the thesis research with Ph. D. student Pengsheng Ji. It concerns a problem of asymptotic inference in the classical Gaussian white noise model, but with potential applications in stationary and locally stationary Gaussian processes. We consider testing for presence of a signal in Gaussian white noise with small intensity depending on $n$, when the alternatives are given by smoothness ellipsoids with an $L^2$-ball of radius $\rho$ removed. It is known that, for a fixed Sobolev type ellipsoid of smoothness $\beta$ and size $M$, a certain sequence of radii depending on $n$ is the critical separation rate, in the sense that the minimax error of second kind over $\alpha$-tests stays asymptotically between 0 and 1 strictly (Ingster [66]). In addition, Ermakov [39] found the sharp asymptotics of the minimax error of second kind at the separation rate. For adaptation over $M$ only, it is established that the sharp risk asymptotics can be replicated in that adaptive setting, if $\rho$ tends to zero slower than the separation rate. The penalty for adaption here turns out to be a sequence tending to infinity arbitrarily slowly. This result is of preliminary nature, since sharp adaptation over both $\beta$ and $M$ appears to be possible (cf. proposed research, part C).

Further below in this proposal an introduction to quantum hypothesis testing is given; the framework there is noncommutative probability theory based on the postulates of quantum mechanics. This topic has been of major interest in Quantum Information Theory for some time, beginning with Helstrom [58].

[R4] The Chernoff lower bound for symmetric quantum hypothesis testing. A testing or discrimination problem between states of quantum system is considered, where the hypotheses are density operators on a finite-dimensional complex Hilbert space. A lower bound on the asymptotic rate exponents of Bayesian error probabilities is proved. 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. The classical result is obtained as a special case if the two hypothetic density operators commute. Independently it has been
shown in the literature (cf [3] and [R5] below) that the lower bound is achievable also in the
general quantum (noncommutative) case. This implies that the result is one part of the definitive
quantum Chernoff bound.

[R5] Asymptotic error rates in quantum hypothesis testing. This paper gives a comprehensive
account of the quantum Chernoff bound, both with regard to the lower risk bound (first shown in
[R4] above) and with regard to attainability, first shown in Audenaert et al. [3]. The attainability
part is thus not a result claimed by the proposer under the previous grant, but what is relevant
in [R5] is the Hoeffding extension of the Chernoff bound, well known in the classical case.

[R6] Asymptotically optimal discrimination between multiple pure quantum states. From classi-
cal (non-quantum) statistics it is known that for the problem of discriminating between \( r \geq 2 \)
probability measures, the optimal error decreases exponentially with an exponent which equals
the exponent for testing between the least favorable (closest) pair of measures. This paper proves
the analogous fact for quantum states which are pure (given by density matrices of rank one).
Thus, the decision problem between a finite number of states of a quantum system is considered,
assuming that an arbitrarily large number of copies of the system is available for measurement.
An upper bound is provided on the asymptotic exponential decay of the averaged probability of
rejecting the true state. It represents a generalization of the bound discussed above in [R4]-[R5]
(multiple quantum Chernoff bound, MQCB) for a set of \( r \geq 2 \) states. The main result is that the
bound is sharp in the case of pure states.

the problem of a large deviation type optimality for discriminating between several \( r \geq 2 \)
quantum states is considered, but the solution was found under the restriction that all states are
pure. To solve the problem for general (mixed) states, as a first stage it is of interest to obtain an
exponential rate of decay of the error probability. The second stage would then be to specify
the optimal exponent, i.e. the MQCB. An exponential rate has been found by Parthasarathy
[111]; here an alternative approach is taken. Observations are subdivided into blocks, and pairs
of states are tested within a block using the optimal method for the binary case. In this way, a
sequence of quantum multiple tests in increasing block length is found, with an error exponent
which equals up to a factor the least favorable mean quantum Chernoff distance over all pairs.
The factor can be arbitrarily close to 1, but may also be large depending on the configuration of
the \( r \) basic states.

[R8] An asymptotic error bound for testing multiple quantum hypotheses. A decision is to be
made between \( r \geq 2 \) quantum states, on the basis of measurements of \( n \) of copies of a quantum
system. The corresponding states are described by an \( n \)-fold tensor product of the associated
basic density operators. The optimal test (or quantum detector) minimizing the sum of error
probabilities and thus generalizing the Holevo-Helstrom test is well known [63], [137]. The open
problem is to determine the corresponding optimal error exponent. Here again, the solution in
the classical case is known (discrimination between \( r \geq 2 \) probability measures): according to
[120], the optimal error exponent is the minimal binary Chernoff bound between any two density
operators taken from the finite set. Define analogously the multiple quantum Chernoff bound
(MQCB), considering all possible pairs of states. In [R6] it has been shown that this asymptotic
error is attainable in the case of \( r \) pure states, and that it is unimprovable in general. Here the
attainability result is extended to a larger class of \( r \)-tuples of states which are possibly mixed,
but pairwise linearly independent. A bound valid for all ensembles of states is also obtained, by
constructing a quantum detector which universally attains the MQCB up to a factor 1/3.
[R9] **Attainment of the multiple quantum Chernoff bound for certain ensembles of mixed states.** This result can be considered a supplement to [R8]. It is established that for certain special configurations of mixed quantum states, the asymptotic risk bound found in [R8] can be improved. In these cases the target bound MQCB is exactly attainable, suppressing the factor 1/3. Additional reasoning in connection with this result shows that there is a multitude of special configurations of the set of general (possibly mixed) states where the MQCB is attained, lending further support to the conjecture that it is attainable in general.

**Broader impacts.** Topics [R2] and [R3] were connected to Ph.D. thesis research and contributed to the career launch of two new researchers in Statistics: Pengsheng Ji, now Assistant Professor (tenure track) at the University of Georgia at Athens, and Inder Tecuapetla-Gómez, now Postdoc researcher at the University of Göttingen in Germany.

A significant part of the research effort was directed at furthering the interaction between the fields of Statistics and Quantum Information Science. In this regard, the paper [104] co-authored by the PI, which (in conjunction with Audenaert et al. [3]) solved the long-standing problem of the quantum Chernoff bound, had some impact. Applications, discussion and generalizations can be found in [1], [4], [8], [9], [20], [21], [27], [29], [30], [42], [56], [57], [59], [60], [61], [68], [69], [71], [79], [89], [90], [93], [95], [96], [97], [99], [100], [110], [121], [123], [128], [129], [135] and in the textbook [112]. The result has been described as seminal in the paper [85] (to appear in Ann. Statist; author affiliation: Centre for Quantum Technologies, Singapore). The mapping from a pair of quantum states to a pair of associated probability distributions, introduced in [104] to prove the converse part of the quantum Chernoff bound, has been termed the Nussbaum-Szkoł mapping [29], [30], [128], [129]. The PI was invited speaker at the First International Workshop on Entangled Coherent States and Its Application to Quantum Information Science in Tokyo, 2012.

(d) Publications resulting from the NSF award (full text available under http://www.math.cornell.edu/~nussbaum/papers.html)

**R1:** Golubev, Nussbaum, and Zhou [48]; **R2:** Tecuapetla-Gómez and Nussbaum [127], Tecuapetla-Gómez [126]; **R3:** Ji and Nussbaum [75], Ji [74]; **R4:** Nussbaum and Szkoła [104]; **R5:** Audenaert, Nussbaum, Szkoła and Verstraete [5]; **R6:** Nussbaum and Szkoła [107]; **R7:** Nussbaum and Szkoła [105]; **R8:** Nussbaum and Szkoła [106]; **R9:** Nussbaum [103]

(e) available data, samples etc.: N/A

**Proposed future research**

(A) **Information contained in additional observations** Le Cam [80] considered a regular parametric experiment of n i.i.d. observations $X^{(n)} = (X_1, \ldots, X_n)$ where $X_i \sim P_\theta$, with parameter $\theta \in \Theta \subset \mathbb{R}^k$, and asked the question under what conditions m additional observations $X_{n+1}, \ldots, X_{n+m}$ do not essentially increase the information in the data about $\theta$. He found that if $m = o(n)$, then the experiments given by $X^{(n)}$ and $X^{(n+m)}$ are asymptotically equivalent, provided the family $(P_\theta, \theta \in \Theta)$ is regular in the usual sense. Mammen [87] extended these results by giving explicit equivalence maps connecting observations $X^{(n)}$ and $X^{(n+m)}$.

It appears possible to extend these results to a nonparametric setting, taking the Gaussian white noise model (1) as a point of departure, where $g := f^{1/2} \in \Sigma \subset L_2(0,1)$. Taking a similar route as
in Nussbaum [102], one starts by considering a subexperiment where \( g \) is restricted to a shrinking neighborhood \( \Sigma_n (g_0) \) around some \( g_0 \) which is initially fixed. Then by subtracting the ‘known’ \( g_0(t) dt \) and multiplying by \( n^{1/2} \) the model is equivalent to

\[
dZ_t = n^{1/2} (g(t) - g_0(t)) dt + dW_t, \ t \in [0, 1], \ g \in \Sigma_n (g_0)
\]

(we use \( dZ_t \) generically for the observations in each model). Since \( n \) plays the role of sample size, we have to compare (2) with its version where \( n \) is replaced by \( n + m \). By a well known formula, the Hellinger distance between the two laws for given \( g \) tends to zero if the \( L_2 \)-distance of the two drift functions tends to zero, that is (assuming \( m/n \to 0 \))

\[
(n+m)^{1/2} - n^{1/2} 2 \|g-g_0\|^2 \sim \frac{1}{4n} m^2 \|g-g_0\|^2 = o(1).
\]

We can now link the possible growth of \( m \) with the shrinking rate of the neighborhood \( \Sigma_n (g_0) \). If \( \Sigma_n (g_0) \) is a metric \( L_2 \)-neighborhood shrinking with rate \( \gamma_n \) and the initial parameter space \( \Sigma \) allows an estimation rate of \( g \) as \( \|\hat{g} - g\|_2 = O_P (\gamma_n) \) then asymptotic equivalence results obtained over shrinking neighborhoods \( \Sigma_n (g_0) \) can be ‘pieced together’ to a result valid over the whole of \( \Sigma \), by the sample splitting method of [102]. So if we assume that \( \Sigma \) is a smoothness class with index \( \alpha \) then the estimation rate will be \( \gamma_n = n^{-\alpha/(2\alpha+1)} \), and consequently, if \( \|g-g_0\|_2 = O_P (\gamma_n^2) \) in (3) then

\[
\frac{m}{n^{1/2}} \gamma_n = \frac{m}{n^{1/2}} n^{-\alpha/(2\alpha+1)} = \frac{m}{n^{(2\alpha+1)/(2\alpha+1)}} = o(1)
\]

will be sufficient for asymptotic equivalence of (1) with its counterpart where \( n \) is replaced by \( n + m \), globally over parameter space \( f^{1/2} = g \in \Sigma \).

In (4) as a limit for \( \alpha \to \infty \) we obtain Le Cam’s condition \( m/n = o(1) \) valid for the finite dimensional case. For small \( \alpha \), that is \( \alpha \to 0 \) we would obtain a condition \( m/n^{1/2} = o(1) \) in the limit (cp. also Low and Zhou [86] for Poisson experiments). As a summary, we see that the larger the parameter space \( \Sigma \), the smaller the allowed number \( m \) of supplementary observations in addition to \( n \) which would be asymptotically negligible.

Here we have assumed \( \Sigma \) to be a smoothness class with index \( \alpha \), but in the white noise setting it appears straightforward to describe the attainable rates of convergence over \( \Sigma \) (thus also the shrinking rates \( \gamma_n \) of \( \Sigma_n (g_0) \)) in terms of entropy characteristics. We expect the rate for \( m \) found in (4) to be sharp, with nonequivalence for larger \( m \) to be proved by methods similar to Brown and Zhang [15]. The question of explicit equivalence maps also poses itself.

The result about the information in additional observations will automatically carry over to those settings where equivalence to the nonparametric white noise model (1) has been established, such as the i.i.d. model on a finite interval (Nussbaum [102]) and the stationary Gaussian sequence model under weak dependence (Golubev et al. [48]). In general we expect the reasoning with ‘additional observations’ to be helpful when comparing the informational content of different statistical models; for applications in extreme value theory cf. Marohn [88] and Falk and Marohn [40].

(B) Asymptotic equivalence for estimating a Toeplitz covariance matrix Consider a real-valued stationary Gaussian time series \( \{X(t), t \geq 1\} \) with zero mean and autocovariance function \( \gamma_h = EX(t)X(t+h) \) for \( h \geq 0 \). Let \( X = (X(1), \ldots, X(p)) \) be a corresponding series
of length \( p \) and let \( \mathbf{X}_1, \ldots, \mathbf{X}_n \) be independent copies of \( \mathbf{X} \). Cai, Ren and Zhou [18] considered estimation of the \( p \times p \) Toeplitz covariance matrix \( \Sigma_p := \mathbf{E} \mathbf{X} \mathbf{X}^\top \) from these \( n \) independent observations, in a setting where the dimension \( p \) may tend to infinity along with sample size \( n \). They established minimax rates of convergence, in the spectral norm for matrices defined by \( \|A\|^2 := \lambda_{\max}(A^\top A) \) for two commonly used parameter spaces. These parameter spaces are given in terms of rates of decay for the autocovariance function \( \gamma_h \), thus they are similar to smoothness classes with index \( \alpha \) for the associated spectral density function \( f \). Their basic finding is that the optimal rate of convergence in terms of \( \|\cdot\|^2 \), apart from some log-factor, is given by \((np)^{-2\alpha/(2\alpha+1)}\). Note that the rate for estimating a spectral density in the classical model where only one time series \( \mathbf{X}_1 \) of length \( p \) is observed (thus \( n = 1 \)) is \( p^{-2\alpha/(2\alpha+1)} \), so that the result for \( n \) i.i.d. copies of a time series of length \( p \) is in agreement with the intuition that the 'total number of observations' then is \( np \).

Cai et al. [18] also raise the question of asymptotic equivalence of the corresponding statistical experiment to some Gaussian white noise model. Indeed Golubev, Nussbaum and Zhou [48] have shown that observations given by \( \mathbf{X}_1 \), as \( p \to \infty \) and \( n = 1 \), when the spectral density \( f \) varies in a certain smoothness class, can be approximated by a white noise experiment

\[
dZ_\lambda = \log f(\lambda) \, d\lambda + 2\pi^{1/2} (np)^{-1/2} \, dW_\lambda, \quad \lambda \in [-\pi, \pi]
\]

in the sense of the Le Cam deficiency distance. This suggests an analogous equivalence result for the model of i.i.d. copies \( \mathbf{X}_1, \ldots, \mathbf{X}_n \), where in (5) above we now have \( n \to \infty \) in addition to \( p \to \infty \). The intuition here is that the presence of i.i.d. copies, in addition to the aspect of weakly dependent observations as \( p \to \infty \), should in fact improve the Gaussian approximation to the model.

The nonparametric estimation results by Cai et al. [18] described above were obtained under a restrictive condition which limits the growth of \( n \) relative to \( p \). The condition can be described, apart from a log factor, by \((np)^{1/(2\alpha+1)} \leq p/2\), which essentially means

\[
n \leq O(p^{2\alpha})\,.
\]

That some growth condition on \( n \) relative to \( p \) will occur here is clear from the following heuristic observation. Assume the extreme case that \( p \) is fixed and \( n \to \infty \); we observe i.i.d. zero mean Gaussian vectors of fixed dimension \( p \) with general Toeplitz covariance matrix \( \Sigma \), which depends on the parameter vector \( \gamma_p := (\gamma_0, \ldots, \gamma_{p-1}) \). If \( \gamma_p \) is further restricted by a periodicity condition, then \( \Sigma \) is a circulant matrix and its eigenvectors do not depend on the unknown parameter \( \gamma_p \). In this case (cf. below) observations are equivalent to a regression type model (9), which for fixed \( p \) and \( n \to \infty \) can be reduced to a discrete version of (5). However without a periodicity condition on \( \gamma_p \), already for \( p = 3 \) the eigenvectors of \( \Sigma \) depend on the parameter \( \gamma_p \), so that a reduction to a simple model like (9) or (5) appears not to be possible.

In a setting where both \( n, p \to \infty \), but under a more restrictive growth condition on \( n \) than (6):

\[
n = o\left(p^{\alpha-1/2}\right),
\]

preliminary investigations by the PI and Tecuapetla-Gomez have shown that asymptotic equivalence to (5) holds. However we conjecture that (roughly) the square of the r.h.s. above is the sharp rate for asymptotic equivalence, namely

\[
n = o\left(p^{2\alpha-1}\right).
\]
To gain insight, one may refer to the periodic submodel of the Gaussian stationary series, which has been crucially used to approximate the general stationary model in Golubev et al. [48]. Here for even \( q := p \) the autocovariances are restricted by \( \gamma_{q/2+h} = \gamma_{q/2-(h-1)}, h = 1, \ldots, q/2 \), which implies that the resulting Toeplitz covariance matrices \( \Sigma_q \) are circulants. In this case they can all be simultaneously diagonalized by the same discrete Fourier transform; the eigenvalues \( \lambda_j \) fulfill

\[
\lambda_j = 2\pi \hat{f}_p(\omega_j), \quad |j| \leq q/2
\]

where \( \omega_j = 2\pi j/p, |j| \leq q/2 \) and the function \( \hat{f}_p \) is a truncated Fourier series approximation to the spectral density \( f \). As a result, by an orthogonal transformation, the model of observed i.i.d. time series \( X_1, \ldots, X_n \) is equivalent to observing independent r.v.’s

\[
y_{ij} \sim N(0, \lambda_j), \quad |j| \leq q/2, \; i = 1, \ldots, n. \tag{9}
\]

This model, for \( n = 1 \), is a ‘Gaussian variance’ nonparametric regression model where the (approximate) values of the spectral density \( f(\omega_j) \) enter through the variance of Gaussians. Moreover, since \( i = 1, \ldots, n \), one observes \( n \) i.i.d. replications of such a model.

To further simplify the setup, for heuristic purposes, consider the more standard nonparametric Gaussian regression of location type

\[
y_{ij} = f(x_j) + \xi_{ij}, \quad j = 1, \ldots, p, \; i = 1, \ldots, n \tag{10}
\]

where points \( x_j, j = 1, \ldots, p \) are equidistant on \([0, 1] \), the \( \xi_{ij} \) are independent standard Gaussian and the parameter \( f \) varies in a function class \( \mathcal{F} \). Here also if \( i = 1, \ldots, n \), one observes i.i.d. replications of a basic model where \( n = 1 \). Rohde [118] for the case \( n = 1 \) considered an extension of the basic Brown and Low [14] result about the continuous approximation of the discrete Gaussian regression model, allowing smoothness classes of higher degree \( \alpha \) and finding associated rates of convergence for the deficiency distance. We note the following straightforward extension of [118] to the case of \( n > 1 \) replications. Consider a continuous white noise model

\[
dZ_t = f(t)dt + (np)^{-1/2}dW_t, \; t \in [0, 1]; \tag{11}
\]

then (11) and (10) are asymptotically equivalent if the condition

\[
np \sup_{f \in \mathcal{F}} \| f - \hat{f}_p \|_2^2 \to 0 \tag{12}
\]

is fulfilled, where \( \hat{f}_p \) is the truncated Fourier series approximation to \( f \) in a special ONB (the standard Fourier system is used in [118]).

**Case a).** Assume that \( \mathcal{F} \) is a Sobolev type ellipsoid given by a restriction \( \sum_{j=1}^{\infty} a_j f_j^2 \leq M \) with \( a_j = j^{2\alpha} \). Then to achieve (12) one notes

\[
np \sum_{j=p+1}^{\infty} f_j^2 \leq np (p+1)^{-2\alpha} M \leq Mnp^{1-2\alpha}
\]

so that a condition (8) is sufficient for the equivalence of (11) and (10).

**Case b).** Assume that \( \mathcal{F} \) is a function set as above with \( a_j = \exp(\mu j) \), for some \( \mu > 0 \); such ellipsoids are associated to analytic functions. In this case we obtain that \( n \) may grow exponentially relative to \( p \), namely \( n = o \left( p^{-1} \exp(\mu p) \right) \).
Case c). Assume that \( \mathcal{F} \) is a regular parametric class of functions \( \mathcal{F} = \{f_\theta, \vartheta \in \Theta\} \) where \( \Theta \subset \mathbb{R}^k \), \( k \) fixed. Assume initially that \( p \) is fixed also; then by taking averages over \( i = 1, \ldots, n \) in (10) one obtains a parametric regression model

\[
\bar{y}_j = f_\theta (x_j) + n^{-1/2} \xi_j, \quad j = 1, \ldots, p
\]

where the Fisher information is (for dimension \( k = 1 \)) \( J_\theta = n \sum_{j=1}^p (f_\theta'(x_j))^2 \). Now let \( p \to \infty \); if we assume smoothness properties in \( x \) for \( f_\theta \) and \( f_\theta' \) then the asymptotic Fisher information will be \( J_\theta \sim np \int_0^1 (f_\theta'(x))^2 \, dx \), that is, it will be the same as in the model (11) if \( f \) is replaced by \( f_\theta \). This leads to the conjecture that if \( \mathcal{F} \) is a regular parametric class \( \mathcal{F} = \{f_\theta, \vartheta \in \Theta\} \) then 

\[
n \text{may grow arbitrarily quickly compared to } p \to \infty \text{ for the equivalence of (10) and (11) to hold. Carried over to the model of i.i.d. copies } X_1, \ldots, X_n \text{ of a stationary series } X \text{ of length } p, \text{ this would mean that in the parametric case } \mathcal{F} = \{f_\theta, \vartheta \in \Theta\}, \text{ the approximation by the continuous model (5) would always hold as soon as } p \to \infty.\]

The summary of considerations in cases a)-c) above can be formulated as the following conjecture: if we observe \( n \) independent replications of a time series of length \( p \), then the larger the parameter space \( \mathcal{F} \), the smaller the allowed number \( n \) turns out to be, such that asymptotic equivalence to a white noise model (5) holds, and thus also to a full time series of length \( np \).

(C) Sharp Adaptive Nonparametric Testing for Sobolev Ellipsoids This subproject is based on the results described in [R3] above. Consider the Gaussian white noise model (11) for \( p = 1 \) with unknown, nonrandom signal \( f \), given in terms of its Fourier coefficients as \( f = (f_j)_{j=1}^\infty \).

For some \( \beta, M, \rho > 0 \), define sets of functions

\[
\Sigma(\beta, M) = \{f = (f_j)_{j=1}^\infty : \sum_{j=1}^\infty j^{2\beta} f_j^2 \leq M\}; \quad B_\rho = \{f \in l_2 : \|f\|_2^2 \geq \rho\}.
\]

We intend to test the null hypothesis of “no signal” \( H_0 : f = 0 \) against nonparametric alternatives \( H_a : f \in \Sigma(\beta, M) \cap B_\rho \). Assuming that \( n \to \infty \), one expects that for a fixed radius \( \rho \), consistent \( \alpha \)-testing (power tending to 1) in that setting is possible. Assume now \( \rho \to 0 \); if that happens too quickly compared to \( n \), all \( \alpha \)-tests will have the trivial asymptotic (worst case) power \( \alpha \). According to a fundamental result of Ingster [66] there is a critical rate for \( \rho_n \), the so-called separation rate

\[
\rho_n \propto n^{-4\beta/(4\beta+1)}
\]

at which the transition in the power behaviour occurs. A precise formulation in terms of \( \alpha \)-tests (cf. Ingster and Suslina [67]) identifies (13) as the minimax rate in nonparametric testing, over \( f \in \Sigma(\beta, M) \cap B_\rho \).

These minimax rates have been extended in two ways. In the first of these, Ermakov [39] found the exact asymptotics of the minimax type II error (equivalently, of the maximin power) at the separation rate. The shape of that result and its derivation from an underlying Bayes-minimax theorem on ellipsoids exhibit an analogy to the Pinsker [113] constant in nonparametric estimation. In another direction, Spokoiny [124] considered the adaptive version of the minimax nonparametric testing problem, where both \( \beta \) and \( M \) are unknown, and showed that the rate at which \( \rho_n \to 0 \) has to be slowed down by a \( \log \log n \)-factor if nontrivial asymptotic power is to be achieved. Thus an “adaptive minimax rate” was specified, analogous to Ingster’s nonadaptive.
separation rate (13), where the additional log log \( n \)-factor is interpreted as a penalty for adaptation. However this result did not involve a sharp asymptotics of type II error in the sense of Ermakov [39]. In a regression setting, analogous adaptive rate results have been obtained by Gayraud and Pouet [41].

It is noteworthy that in nonparametric estimation over \( f \in \Sigma(\beta, M) \) with \( l_2 \)-loss (as opposed to testing), where the risk asymptotics is given by the Pinsker constant, there is a series of results showing that adaptation is possible with neither a penalty in the rate nor in the constant, cf. Efromovich and Pinsker [38], Golubev [46], [47], Tsybakov [131]. It can now be asked whether the sharp risk asymptotics for testing in the sense of Ermakov [39] can be reproduced in an adaptive setting, in the context of a possible rate penalty for adaptation.

In Ji and Nussbaum [75], cf. [R3], a partial solution has been obtained, in the sense that for a fixed \( \beta \) a sharp risk asymptotics for adaptation over \( \beta \) was found. The penalty for adaptation in this context was the fact that the critical rate (13) has to be slowed down by a factor \( c_n \to \infty \) (arbitrarily slowly), for the Ermakov [39] sharp risk bound to hold in terms of the log-asymptotics of the risk (moderate deviation approach). We conjecture that this result can be reproduced also when both \( \beta \) and \( M \) are unknown. For this, the log log \( n \)-penalty factor on the critical radius, known from Spokoiny [124], has to be further amended by a factor \( \delta_n \to \infty \) (arbitrarily slowly), and risk has again to be evaluated in the sense of a log-asymptotics for the tail probability.

To formulate conjectures, define a critical radius \( \hat{\rho}_{n,\beta,M} \) depending on \( M \) and \( \beta \) by

\[
\left( \hat{\rho}_{n,\beta,M} \right)^{(4\beta+1)/4\beta} = n^{-1} \left( A(\beta) M \right)^{1/4\beta} (2 \log \log n)^{1/2} d_n
\]

where \( d_n \to \infty \) and \( A(\beta) = (4\beta + 1)^{(2\beta+1)/(2\beta + 1)} \) is a constant found by Ermakov [39]. Assume the smoothness parameter \( (\beta, M) \) varies over a range \( S := [\beta_1, \beta_2] \times [M_1, M_2] \), where \( \beta_1, M_1 > 0 \). Define for a test \( \phi \) and for given \( \beta, M > 0 \),

\[
\Psi(\phi, \rho, \beta, M) := \sup_{f \in \Sigma(\beta,M) \cap B_\rho} (1 - E_{n,f} \phi).
\]

**Conjecture (i)** For any asymptotic \( \alpha \)-test \( \phi_n \) we have

\[
\frac{1}{2d_n^2 \log \log n} \log \sup_{(\beta,M) \in S} \Psi(\phi_n, \hat{\rho}_{n,\beta,M}, \beta, M) \geq -\frac{1}{2} + o(1). \tag{14}
\]

(ii) If \( d_n = o(\log \log n) \) then there is a sequence of asymptotic \( \alpha \)-tests \( \phi_n^* \) which fulfills the converse inequality to (14).

The sharp asymptotics is contained here in the radius \( \hat{\rho}_{n,\beta,M} \) over which the upper bound on the (worst case) error of second kind is attained, in the sense of the log-asymptotics for a tail probability. This conjectured result combines preliminary considerations on sharp asymptotics for unknown \( \beta \), known \( M \) in section 7.1.3 of Ingster and Suslina [67] with the methods developed for the case of only \( M \) unknown in [R3]. The result, if confirmed, would also be an analog, for nonparametric testing, of the sharp risk asymptotics for estimation in the presence of a penalty for adaptation such as Tsybakov [130]. An analogous result for the sup-norm can also be envisaged, building on the non-adaptive sharp asymptotics for testing of Lepski and Tsybakov [84], the adaptive rate results for testing (unknown \( M \) only) of Dümbgen and Spokoiny [36], Rohde [119], the non-adaptive sharp asymptotics for estimation of Korostelev [78] and the adaptive estimation results of Golubev, Lepski and Levit [49].

9
Asymptotic inference in quantum statistical experiments. For a brief introduction to the following subprojects, we describe the simplest possible setup of discrimination between several quantum hypotheses. For an introduction to quantum statistical inference with physical background, see Gill [44], [43], Holevo [64], [65] or Barndorff-Nielsen, Gill and Jupp [6].

A density matrix \( \rho \) is a complex, self-adjoint, positive, \( d \times d \) matrix satisfying the normalization condition \( \text{Tr}[\rho] = 1 \), where \( \text{Tr}[\cdot] \) is the trace operation. Here "positive" means nonnegative definite. We identify a density matrix with a state of a quantum system; the hypotheses are described by the states \( \rho_i, i = 1, \ldots, r \). A multiple test, or detector is given by an \( r \)-tuple of Hermitian positive \( d \times d \) matrices \( \tau = \{ \tau_i, i = 1, \ldots, r \} \) fulfilling \( \sum_i \tau_i = I \) (unit matrix). The total error probability of the detector \( \tau \) is

\[
\text{Err}(\tau) := 1 - \frac{1}{r} \sum_{i=1}^{r} \text{Tr}[\rho_i \tau_i]
\]

where \( \text{Tr}[\rho_i \tau_i] \) can be interpreted as the probability of correct decision when \( \rho_i \) is the true state. If all \( \rho_i \) and also all \( \tau_i \) are real diagonal matrices then the setup reduces to the classical testing problem between probability measures \( P_i, i = 1, \ldots, r \) on an appropriate index set \( \Omega, |\Omega| = d \). In this case each \( \tau_i \), by its set of eigenvalues, represents a function \( t_i \) on \( \Omega \) with values in \([0, 1]\) such that \( \sum_{i=1}^{r} t_i(\omega) = 1 \) for all \( \omega \), so \( t = \{t_i, i = 1, \ldots, r\} \) represents a classical (randomized) decision function, or multiple test. The same is true when all \( \rho_i \) have the same set of eigenvectors; then \( \rho_i \) are said to commute (commutative case). In this sense, commuting states describe the classical discrimination problem between several probability measures on a finite sample space \( \Omega \), as a special case of the present quantum setting.

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. In the quantum "large sample" setting the hypotheses are assumed to be the \( n \)-fold tensor product of \( \rho_i \) with itself: \( \rho_i^{\otimes n}, i = 1, \ldots, r \). The detectors \( \tau = \{ \tau_i, i = 1, \ldots, r \} \) now operate on the states \( \rho_i^{\otimes n} \) which are \( d^n \times d^n \) matrices, but the \( \tau_i \) need not have tensor product structure. The corresponding total error probability \( \text{Err}_n(\tau) \) of a detector \( \tau \) is given by (15) where \( \rho_i \) are replaced by \( \rho_i^{\otimes n} \). The object of study is the asymptotics as \( n \to \infty \) of the optimal error probability

\[
\Delta_n(\Sigma) := \inf_{\tau \text{ detector}} \text{Err}_n(\tau) \text{ where } \Sigma = \{ \rho_1, \ldots, \rho_r \}.
\]

For \( r = 2 \), the optimal hypothesis tests minimizing the error probability have long been known to be the Holevo-Helstrom hypothesis tests [58], [63]; they are given by certain projectors derived from the two states \( \rho_1^{\otimes n}, \rho_2^{\otimes n} \), generalizing the Neyman-Pearson test. But the log-asymptotics of \( \Delta_n(\rho_1, \rho_2) \) has only recently been found by Nussbaum and Szkola [104], cf. [R4] above, and by Audenaert et al [3], as

\[
\lim_{n \to \infty} \frac{1}{n} \log \Delta_n(\rho_1, \rho_2) = -C(\rho_1, \rho_2) \text{ where } C(\rho_1, \rho_2) = -\log \inf_{0 \leq s \leq 1} \text{Tr}[\rho_1^{1-s}\rho_2^s] \geq 0
\]

generalizing the classical Chernoff [26] bound for discriminating between two probability measures. For several quantum hypotheses \( \Sigma = \{ \rho_1, \ldots, \rho_r \} \), the detectors minimizing \( \Delta_n(\Sigma) \) in (16) have been implicitly described in [62] and [137]. As to the asymptotics of the error probability \( \Delta_n(\Sigma) \),
there exists the conjecture (the multiple quantum Chernoff bound, MQCB)

$$\lim_{n \to \infty} \frac{1}{n} \log \Delta_n (\Sigma) = - \min_{i \neq j} C (\rho_i, \rho_j),$$

(18)
in view of the known generalization of the Chernoff bound to the case of several probability measures \( (r \geq 2) \) in [120]. The conjecture has been confirmed in a number of special cases, cf. [R6]-[R9] above. The following subprojects concern particular topics in quantum discrimination, and more generally in quantum statistical inference.

**(D) Realizable receivers for quantum discrimination**  A pure state, in the finite dimensional setting described above, is given by a density matrix of rank one, i.e. \( \rho = |x \rangle \langle x| \) where \( |x\rangle \) is a column unit vector in \( \mathbb{C}^d \) and \( \langle x| \) its transpose (Dirac notation). In quantum optics, one works with Gaussian states, described by density operators in Hilbert space \( L_2 (\mathbb{R}) \). If the density operator is described in terms of the ONB consisting of the normed Hermite polynomials, then the setting is said to be that of the harmonic oscillator. In Nair et al. [99] a scheme for discriminating an \( r \)-ensemble of certain pure states of the harmonic oscillator is developed, the so-called coherent states which play a central role in quantum optics. The quantum decision functions between states (multiple tests, or detectors) are also called receivers in this context. The starting point for the method of [99] has been the detector found in Nussbaum and Szkola [107], cf. also [R6], for the problem of discriminating between \( r \) pure states in the case of quantum-i.i.d. observations. In this case the states are of form \( |x_j\rangle \otimes^{n} \langle x_j| \otimes^{n} \), \( j = 1, \ldots, r \) for some unit vectors \( x_j \in \mathbb{C}^d \); the method proposed in [107] to achieve the optimal bound (18) is based on a Gram-Schmidt orthonormalization of the "large" unit vectors \( x_j \otimes^{n} \).

For quantum optics, where each of the \( n \) observed quantum system represents a photon, such a scheme is considered non-realizable, since it is not possible in practice to store and process a large number \( n \) of photons. This served as a motivation in [99] to develop the SWN receiver (for sequential waveform nulling receiver), which achieves the multiple quantum Chernoff bound (18) asymptotically while using only sequential operations, that is methods where the decision among indices \( \{1, \ldots, r\} \) is updated after each observation step \( i \), where \( i = 1, \ldots, n \). In the physics literature, such methods are said to follow the principle of local operations and classical communication (LOCC); in statistics the prime example for this idea would be the method of sequentially updating Bayesian posteriors.

The method of [99] is developed in the context of coherent states [83]. These are pure Gaussian states which can be described, for a complex number \( \alpha \), as a pure state \( |\alpha\rangle \langle \alpha| \) where

$$|\alpha\rangle = \exp \left( -\frac{1}{2} |\alpha|^2 \right) \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle$$

and where \( |n\rangle, n = 0, 1, \ldots \) is the ONB of the Hilbert space \( L_2 (\mathbb{R}) \) associated with the harmonic oscillator. The states are Gaussian because the basic (canonical) observables \( P, Q \) are normally distributed. At the same time other observables have a discrete distribution; for instance the observed number of photons in that state is Poisson \( Po \left( |\alpha|^2 \right) \). The latter fact is in agreement with the interpretation of \( |\alpha|^2 \) as the energy of the state. Suppose there are given complex numbers \( \alpha_j, j = 1, \ldots, r \) and some \( N > 0 \); the states of light to be discriminated are \( \rho_{j,N} := N |\alpha_j\rangle \langle \alpha_j| \),
It is shown that as $N \to \infty$, the SWN receiver attains an optimal discrimination bound equivalent to (18), called the Helstrom limit in this case; this takes the form

$$
\lim_{N \to \infty} \frac{1}{N} \log \inf_{\tau \text{ receiver}} \text{Err}_N(\tau) = -\min_{j \neq k} |\alpha_j - \alpha_k|^2
$$

as $N \to \infty$, where $N$ is an energy parameter for the ensemble. Here it is notable that the asymptotics for energy $N \to \infty$ replaces the standard scheme of a large number of identical copies of basic states. Based on physical intuition, in [99] also a variant of the method for i.i.d. copies of basic states is presented, which achieves the bound (18) along with the orthogonalization detector of [R6], but is sequential, i.e. follows the LOCC principle.

These results raise the following question: for discriminating between states $\rho_j^\otimes n$, $j = 1, \ldots, r$, where the basic states $\rho_j$ are arbitrary, are there sequential (updating) methods achieving the quantum Cherno bound (18)? For $r = 2$ this question has been answered in the negative, cf. references 15, 17 cited in [99] and [20], [90], [89], [7], [100]. However, it might be interesting to find a subclass of basic states $\rho_j$ such that the quantum Cherno bound is achievable sequentially. A candidate is the class of linearly independent states $\rho_j$, which are close to pure states some sense, and for which it has been shown in [R8] that (18) is achievable for any $r \geq 2$. Moreover, given the fact that the Chernoff bound concerns a Bayes risk for uniform prior over two states, it is of interest to identify those risk bounds which are attainable by sequentially updating posterior distributions.

(E) The Chernoff bound for quantum Markov chains  The extension of the classical Chernoff bound from i.i.d. observations to Markov chains is well known [16]. Consider observations of a Markov chain $X_1, \ldots, X_n$ with finite state space $\{1, \ldots, d\}$, on which two transition probability matrices $Q_i$, $i = 1, 2$ are given. Let $\Delta_n (Q_1, Q_2)$ be the minimal Bayesian error probability for discriminating between the two laws of the Markov chain, assuming they are in the stationary regime. Assume the $d \times d$ stochastic matrices $Q_i$ are given by elements $Q_{i,jk} = P(X_1 = k|X_0 = j)$, $i = 1, 2$, and for $0 \leq s \leq 1$, let $M_{s,Q_1,Q_2}$ be the real positive matrix with elements $M_{jk} = Q_{1,jk}^s Q_{2,jk}^{1-s}$. Then

$$
\lim_{n \to \infty} \frac{1}{n} \log \Delta_n (Q_1, Q_2) = -C_M (Q_1, Q_2) \quad \text{where} \quad C_M (Q_1, Q_2) = \min_{0 \leq s \leq 1} \lambda_{\max} M_{s,Q_1,Q_2} \geq 0
$$

(19)

where $\lambda_{\max} (\cdot)$ is the maximal eigenvalue. This result and the quantum-i.i.d. case (17) have the classical i.i.d. case in common: if the transition probabilities $Q_{i,jk} =: Q_{i,k}$ do not depend on $j$, for $i = 1, 2$, then the matrix $M_{s,Q_1,Q_2}$ has all rows equal, hence the vector 1 can be taken as an eigenvector with eigenvalue $\lambda_{\max} M_{s,Q_1,Q_2} = \sum_{k=1}^d Q_{1,k}^s Q_{2,k}^{1-s}$. The latter expression can also be written $\text{Tr} \left[ \rho_1^{1-s} \rho_2^s \right]$ for two diagonal density matrices $\rho_j$, the diagonal elements of which form the probability distribution $\{Q_{i,k}, k = 1, \ldots, d\}$. This is the setting in which the classical Chernoff bound from i.i.d. observations is obtained from the quantum case (17).

Quantum Markov chains are used in quantum optics to model atom masers, cf. [52], [25]. Consider an initial $d \times d$ density matrix $\rho_0$; then the general Markov transition operation to a density matrix $\rho_1$ is can be described as follows: there is a $k \times k$ density matrix $\phi$ and a unitary $kd \times kd$ matrix $U$ such that

$$
\rho_1 = \text{Tr}_\phi [U (\rho_0 \otimes \phi) U^*]
$$

(20)

where $\text{Tr}_\phi$ is the partial trace operation over the second component, yielding a $d \times d$ matrix (and fulfilling $\text{Tr}_\phi (\rho_0 \otimes \phi) = \rho_0 \text{Tr} [\phi]$). Here the joint state after one Markov operation is $U (\rho_0 \otimes \phi) U^*$,
of which ρₜ is the marginal state. In a quantum Markov process with n steps, the marginal state after n steps is the n-fold iteration of the Markov map (20), but in analogy to the case of classical observations of a Markov chain X₁, ..., Xₙ, one considers the joint state (dimension kdⁿ × kdⁿ) obtained from initial state ρ₀. In [52] both ρ₀ are φ pure states, which implies that the joint state of n observations is also pure, a fact which facilitates asymptotic inference on the basis of local asymptotic normality and Fisher information.

The problem of discriminating between two quantum Markov chains can now be formulated as follows. Given two Markov transition operations of type (20), that is Tᵢ = (Uᵢ, φᵢ), i = 1, 2, with pertaining stationary initial states ρ₀,i; find the analog of the log-asymptotics of (17) and (19), for the Bayesian discrimination error Δₙ(T₁, T₂). The pure state case of [52] should be treated first, since in that case Δₙ(T₁, T₂) = |⟨τ₁,n|τ₂,n⟩|^², where τᵢ,n is the joint pure state vector after n steps (cf. Kargin [77]), and it remains to find the asymptotics as a function of the Markov operators Tᵢ. For an overview of quantum Markov chains cf. [109], sec. III.11 and references therein.

(F) Local asymptotic normality for stationary Gaussian quantum experiments Classical local asymptotic normality (LAN, Le Cam [82]) is a fundamental property of a sequence of statistical experiments, which essentially reduces inference for large sample size to the case of a normal location model. Let {Pₙ,θ, θ ∈ Θ} be a sequence of families of p.m.'s on (Ωₙ, Aₙ) where Θ ⊂ ℜᵏ; the sequence is LAN at θ₀ ∈ int (Θ) if there exists a positive k × k matrix J and random k-vectors Δₙ on Ωₙ such that Δₙ → N(0, J) and

$$
\log \frac{dP_{n,θ_0+h/\sqrt{π}}}{dP_{n,θ_0}} = h^\top Δₙ - \frac{1}{2} h^\top J h + o_P(n,0) \quad \text{as} \quad n \to \infty.
$$

(21)

The underlying idea here is that the log-likelihood ratio asymptotically takes the form pertaining to a Gaussian shift experiment \{Nₖ(h, J⁻¹), h ∈ ℜᵏ\}.

Let now {Pₙ,θ, θ ∈ Θ} be a sequence of quantum statistical experiments, that is (in the simplest case) a family of density matrices ρₙ,θ, of dimension dₙ × dₙ indexed by θ ∈ Θ ⊂ ℜᵏ. Every notion of a quantum analog of LAN (q-LAN in the sequel) must involve the quantum Gaussian distributions; in the most basic description, [65], these are certain families of self-adjoint trace one operators in Hilbert space. Considerations naturally involve the quantum central limit theorem [70].

At present, there are two different approaches to q-LAN. In the first of these (Guta and Kahn [55], [76]), for the case of density matrices ρₙ,θ = ρ₀⊗ₙ being tensor powers of a basic ρ₀ (quantum-i.i.d. case), a direct approach is taken, by showing the strong convergence of the localized experiments to q-Gaussian shifts. Here strong convergence means the existence of quantum channels (the analog of Markov kernels) which allow to approximately transform the prelimiting experiment into the limiting one, and vice versa. This method utilizes the concept of sufficiency for quantum experiments via the quantum analog of Markov kernels, developed in [72], [73]. In the classical case, the strong convergence of localized experiments is known to be related to LAN (Shiryaev and Spokoiny [122]). In the second approach, by Guta and Jenčova [54], the starting point is a possible definition of a quantum analog of the likelihood ratio between two probability measures, the so-called Connes cocycles, which are families of operators with certain algebraic properties. For these objects, a weak convergence property is defined, which can be considered a variant of q-LAN since in the classical case, weak convergence implies LAN. Again for the quantum-i.i.d.
case $\rho_{n,0} = \rho_0^{\otimes n}$, the weak convergence to quantum Gaussian shift experiments is established. In the third approach, by Yamagata et al. [136], a direct generalization of the expansion (21) is proposed. First, for two $d_n \times d_n$ density matrices $\rho$ and $\sigma$, a Hermitian operator $L = L(\sigma|\rho)$ is called the log-likelihood ratio if $\sigma = \exp\left(\frac{1}{2}L\right)\rho\exp\left(\frac{1}{2}L\right)$ where $\exp(\cdot)$ is the matrix exponential. This is shown to exist if $\sigma, \rho$ are of both of full rank or both pure. Suppose that $k = 1$ so that the parameter $\vartheta$ is real; then the q-LAN property of a family $\{\rho_{n,\vartheta}, \vartheta \in \Theta\}$ at $\vartheta_0$ is defined as follows: there exists a sequence of observables $\Delta_n$ satisfying the quantum central limit theorem such that the matrix-valued log-likelihood ratio $L\left(\rho_{n,\vartheta_0+h/\sqrt{n}}|\rho_{n,\vartheta_0}\right)$ can be expanded in a way similar to (21).

A useful point of departure for further clarifying the connection between these concepts of q-LAN would be a quantum analog of the result of Davies [33] on LAN for parametric models of Gaussian stationary sequences. To define the quantum version of such observations, consider first the quantum analog of a centered multivariate normal law (following [135]). This is a Gaussian state $\rho$ of an $n$-mode bosonic system, given by a self-adjoint, positive, trace one operator $\Omega = \sum_{i=1}^{2n} \omega_i \mathbb{I}$, where $\omega_i$ is a positive definite real symmetric matrix such that $\Omega = \mathbb{I} \otimes \omega$. Let $X^\top := (Q_1, P_1, \ldots, Q_n, P_n)$ the vector of canonical observables and for $\xi \in \mathbb{R}^{2n}$, define the Weyl unitary operators as $D(\xi) := D(\mathbf{i} X^\top \xi)$. The characteristic function of a state $\rho$ is defined as

$$\chi(\xi) := \text{Tr} [\mathbf{i} \rho D(\xi)].$$

A state $\rho$ is centered Gaussian if the characteristic function is of form

$$\chi(\xi) = \exp\left(-\frac{1}{2} \xi^\top V \xi\right)$$

where $V$ is a positive definite real symmetric matrix such that $V + \mathbf{i} \Omega \geq 0$, where $\Omega$ is the $2n \times 2n$ real matrix

$$\Omega = \mathbb{I}_n \otimes \omega, \quad \omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$  

This form of $\Omega$ is determined by the canonical commutation relations (22). The matrix $V$ is the analog of the covariance matrix, in the following sense. Every linear combination $L_{\xi} := X^\top \xi$ of the canonical observables is itself an observable, which gives a normal random variable in the usual sense with law $N(0, \xi^\top V \xi)$. However, two such observables $L_{\xi_1}, L_{\xi_2}$ do not have a joint distribution in general; this is only the case if they commute, that is $[L_{\xi_1}, L_{\xi_2}] = 0$. In particular $Q_j, P_j$ are not jointly observable and thus do not have a joint distribution, but the $n$ observables $(Q_1, \ldots, Q_n)$ have a joint normal distribution, and likewise for $(P_1, \ldots, P_n)$. More generally, any set of linearly independent $\xi_1, \ldots, \xi_n$ such that all pairs $L_{\xi_j}, L_{\xi_k}$ commute defines an $n$-variate normal law with covariance matrix $\Sigma = (\xi_j^\top V \xi_k)_{j,k=1}^n$, and in that sense the $n$-mode Gaussian state $\rho$ gives rise to a whole family of $n$-variate centered normal distributions.

Shift invariant Gaussian states were considered in [96]. For these states the joint normal law of $(P_1, Q_2)$ is the same as the joint law of $(P_{1+r}, Q_{2+r})$ for $r \leq n-1$, etc. More generally, the joint
law of \((\gamma_j P_j + \lambda_j Q_j)_{j=1}^k\) is the same as the joint law of \((\gamma_j P_{j+r} + \lambda_j Q_{j+r})_{j=1}^k\) for any \(k < n\) and \(1 \leq r \leq n - k\). In that case the matrix \(V\) must be block Toeplitz, that is \(V = (\Gamma_{j-k})_{j,k=1}^n\) for certain (non-Toeplitz) \(2 \times 2\) matrices \(\Gamma_k\), \(k = 0, \pm 1, \ldots\). Block Toeplitz matrices occur in the analysis of multivariate stationary time series (cf. Brockwell and Davis [10], Pourahmadi [115]); the \(\Gamma_k\) can be understood as the autocovariance function. As in time series analysis, one may postulate the existence of a matrix valued spectral density function \(f(\lambda), \lambda \in [-\pi, \pi]\) such that

\[
\Gamma_k = \int_{-\pi}^{\pi} \exp(i\lambda t) f(\lambda) d\lambda, k = 0, \pm 1, \ldots
\]

In analogy to Davies [33] or Taniguchi and Kakizawa [125], the sequence of quantum statistical experiments \(\{\rho_{n,\vartheta}, \vartheta \in \Theta\}\) would then be determined given by a parametric family of matrix spectral densities \(\{f_{\vartheta}, \vartheta \in \Theta\}\) and pertaining \(n\)-mode Gaussian states with covariance matrix \(V\). To establish q-LAN, it will be helpful that some of the results connecting eigenvalues of Toeplitz matrices and the spectral density (Grenander and Szegö [51]) are also available for the block Toeplitz case [94], [98]. It remains a challenging problem to apply either of the existing q-LAN concepts to the Gaussian stationary case, thereby also clarifying the q-LAN concept itself.

**Broader impacts of the proposed research.** The workshop "New Horizons in Statistical Decision Theory", to take place in September 2014 at the Mathematical Research Institute Oberwolfach, initiated by the proposer and organized jointly with R. Gill (Leiden) and M. Guta (Nottingham), is closely related to the research program outlined here. Currently a new technological era is opening where quantum mechanics is used not only to predict physical behaviour but increasingly to exploit quantum resources in applications such as Quantum Communication, Cryptography, Computation and Metrology [101]. The last decades have witnessed a revolution in the experimental realization of quantum systems, and the ability to control and accurately measure such systems [35]. Due to the probabilistic nature of quantum measurements, statistical inference based on measurement data plays a key role in analyzing and validating the results of quantum experiments. The analysis of quantum statistical models, and of the “classical” models associated to different measurement outcomes, points towards an underlying non-commutative statistical decision theory with deep connections to Operator Algebra, Quantum Information and Quantum Probability. The first steps in uncovering this theory have been made recently, with the development of the theory of quantum sufficiency [72], the quantum Stein Lemma [108], the quantum Chernoff bound [5], local asymptotic normality [76], [54], [136] and the asymptotic theory of state estimation [45].

The purpose of the Oberwolfach meeting is to bring together mathematical statisticians and the leading experts from the different subfields contributing to Quantum Statistics, from experimentalists to theoretical physicists. The timeliness and importance of exchanges between the communities has become evident by the increasing number of instances where researchers in Quantum Information Theory publish in leading statistics journals [2], [104], [17], [53] [106], [136], [85], a trend to which the PI has contributed as author and reviewer and which the current project aims to advance.

The current proposal has a major educational component, as it is intended to accompany the collaboration between the PI and a Ph.D student.