The Annals of Statistics
1984, Vol. 12, No. 4, 1310–1323

ON THE STABILITY OF BAYES ESTIMATORS FOR GAUSSIAN PROCESSES[1]

BY IAN W. MCKEAGUE

The Florida State University, Tallahassee

We consider the Bayes estimator δ_0 for a Gaussian signal process observed in the presence of additive Gaussian noise under contamination of the signal or noise by QN-laws, introduced by Gualtierotti (1979). Upper bounds on the increase in the mean square error of δ_0 over the minimum possible mean square error under contaminated noise or contaminated signal are given. It is shown that the performance of δ_0 is relatively close to optimal for small amounts of contamination.

1. Introduction. Consider a Gaussian signal process $X = (X_t)$ observed in the presence of an additive Gaussian noise process $N = (N_t)$ for $t \in [0, T]$. The observed process Y is given by $Y_t = X_t + N_t$, $t \in [0, T]$. The Bayes estimator, $\delta_0(Y) = E(X \mid Y)$, of the signal X can be calculated explicitly in terms of the means and covariances of the prior and noise distributions. See Mandelbaum (1984) for a recent study of this estimator.

In the present paper we study the behavior of the Bayes estimator δ_0 under departures from Gaussian law by the prior and noise distributions. This work is similar in spirit but conceptually distinct from work on the asymptotic robustness of estimators, as in Huber (1981), where an unknown parameter is fixed throughout repeated observations and the sample size goes to infinity. It is well known that no linear estimator, such as a Bayes estimator, can be asymptotically robust with respect to contamination in the noise. However, in the present situation where contamination in the noise is restricted to a single realization of the process it is to be expected that δ_0 is qualitatively robust, that is, insensitive to small deviations from the assumptions of Gaussian noise and Gaussian prior. In the present paper we attempt to assess this robustness of δ_0 in quantitative terms. We propose using a specific contamination model to obtain analytic expressions for the amount of deterioration in the performance of δ_0 under contamination.

An important consideration in the choice of a contamination model is the presence of mutual absolute continuity between the contaminated and uncontaminated Gaussian law. Without absolute continuity it is possible to discriminate with zero probability of error between these two laws. This phenomenon of singular discrimination is not found in practical situations and should not be allowed in our contamination model. Consequently, we are looking for a contaminated Gaussian law which preserves absolute continuity. The requirement of

Received October 1982; revised May 1984.
[1] Research supported by a Florida State University COFRS Grant and US Army Research Office Grant DAAG29-82-K-0168.
AMS (1980) *subject classifications.* Primary 62F35, 62M20; secondary 62F15, 60G35, 60B11.
Key words and phrases. Robustness, Bayes estimators, Gaussian processes, QN-laws.

absolute continuity is a severe restriction since in infinite dimensions two measures can be close in terms of their defining parameters yet orthogonal. For example, let μ_1, μ_2 be the measures on $C[0, T]$ induced by two Wiener processes with covariances $s \wedge t$ and $(1 + \delta)(s \wedge t)$ respectively. Then μ_1 and μ_2 are orthogonal for any $\delta > 0$. This rules out consideration of a mixture model $(1 - \varepsilon)\mu_1 + \varepsilon\mu_2$, where $0 < \varepsilon < 1$, as a realistic contamination model since it can be determined with zero probability of error from which law, μ_1 or μ_2, the sample path originated. This can be done by calculating the quadratic variation of the observed sample path (Y_t), $t \in [0, T]$:

$$\lim_{n\to\infty} \sum_{j=1}^{n} (Y_{t_j} - Y_{t_{j-1}})^2 = \begin{cases} T & \text{a.e. } d\mu_1 \\ (1 + \delta)T & \text{a.e. } d\mu_2, \end{cases}$$

where $t_j = jTn^{-1}$. In practice, where continuous observation of the process is not possible, one can still reduce the probability of error to an arbitrarily small positive quantity by observing the process at time points $t_j = jTn^{-1}$ for n sufficiently large. The mixture model is inappropriate in infinite dimensions where, in order to satisfy the absolute continuity requirement, stringent conditions need to be imposed on the means and covariances of the Gaussian laws involved.

An alternative contamination model, which we use in the present paper, is the QN-law introduced by Gualtierotti (1979). QN-laws are defined by a relation $dQ = q dP$, where P is Gaussian and q is a quadratic form. They have more mass in the tails than the Gaussian laws while being equivalent to them and they are sufficiently tractable to allow calculation of Bayes loss which plays an important role in this work. Although the usual concern with robustness is contaminated tails, which in the case of QN-laws are fairly mild, over-heavy tail behavior is ruled out here by the requirement of absolute continuity.

The following example will be used to illustrate the abstract setting in which the main results of the paper are established.

EXAMPLE 1. Let the signal process (X_t) and the observation process (Y_t) satisfy the stochastic differential equations

$$dX_t = -\beta X_t\, dt + dW_t^1, \quad t \in [0, 1],$$
$$dY_t = X_t\, dt + dW_t^2, \quad t \in [0, 1],$$

where W^1 and W^2 are independent Wiener processes, $\beta > 0$, $Y_0 = 0$, and X_0 is a $N(0, 1/(2\beta))$ random variable which is independent of W^1 and W^2. This is a simple example of a linear dynamical system arising in the Kalman-Bucy filtering theory. Formulae for the Bayes estimator $\delta_0[Y](t) = E(X \mid Y)_t$, $t \in [0, 1]$, have been derived by Liptser and Shiryayev (1978, Chapter 12). Let μ_X and μ_N denote the measures induced on the Borel σ-field of $C[0, 1]$ by the signal process (X_t) and the noise process (W_t^2) respectively. Consider a contaminated noise distribution given by the QN-law $d\nu_N(x) = c_N(1 + \varepsilon \int_0^1 x_t^2\, dt)\, d\mu_N(x)$, where $\varepsilon > 0$ and c_N is a constant. Let $r_\varepsilon(\delta)$ denote the integrated mean square error of an estimator

δ under the contaminated noise ν_N, that is

$$r_\varepsilon(\delta) = \int_0^1 E(X_t - \delta[Y](t))^2\,dt,$$

where the expectation is with respect to the probability measure determined by μ_X and ν_N. Theorem 4.4 gives a bound on the ratio of $r_\varepsilon(\delta_0)$ to the minimum possible integrated mean square error under ν_N,

$$r_\varepsilon(\delta_0)/\inf_\delta r_\varepsilon(\delta) \le 1 + (3/\beta)(1 + o(1))\varepsilon^2, \quad \text{as} \quad \varepsilon \to 0.$$

An analogous result holds for the case of a contaminated prior distribution given by the QN-law $d\nu_X(x) = c_X(1 + \varepsilon \int_0^1 x_t^2\,dt)d\mu_X(x)$, except that the $3/\beta$ can be replaced by $2/\beta$.

For a discussion of robust Kalman filtering in discrete time we refer to the papers of Ershov and Liptser (1978) and Masreliez and Martin (1977). References to other aspects of Bayesian robustness can be found in Berger (1982). QN-laws have been applied to signal detection and information theory by Gualtierotti (1980, 1982, 1983).

The results of this paper are given in terms of noise distributions on separable Banach spaces. This is general enough to cover the space $C[0, 1]$ arising in Example 1. In fact our methods can be carried over to a large class of locally convex spaces including $\mathbb{R}^{[0,1]}$. Section 2 contains some preliminary material on measures on separable Banach spaces and a derivation of the Bayes estimator δ_0. Section 3 contains a discussion of QN-laws and posterior distributions when either the prior or noise is a QN-law. Upper bounds for the increase in the mean square error of δ_0 over the minimum possible mean square error under a QN-law prior or QN-law noise are given in Section 4.

2. Preliminaries.

Let E denote a separable Banach space with topological dual E'. The Borel σ-field on E is denoted $\mathscr{B}(E)$. Let μ be a probability measure on $\mathscr{B}(E)$ such that $\int_E \langle f, x\rangle^2\,d\mu(x) < \infty$, for all $f \in E'$. Then, by Weron (1976), μ has a mean element m in E and a covariance operator $R\colon E' \to E$, defined by

$$\langle f, m\rangle = \int_E \langle f, x\rangle\,d\mu(x),\ \langle Rf, g\rangle = \int_E \langle f, x - m\rangle\,\langle g, x - m\rangle\,d\mu(x),$$

for all f, g in E'. There exists a separable Hilbert space H and a continuous linear injection $j\colon H \to E$ such that $R = jj^*$; see Schwartz (1964), Baxendale (1976). The Hilbert space H is called the reproducing kernel Hilbert space (RKHS) of R. The identity on H is denoted I. For u, v in E, z in E', $(u \otimes v)(z) = \langle v, z\rangle\,u$. The notation $R = \sum_n u_n \otimes u_n$ for $\{u_n, n \ge 1\} \subset E$ means that $\sum_1^N \langle f, u_n\rangle\,u_n \to Rf$ in the norm topology of E, for all f in E'. If $\{e_n, n \ge 1\}$ is any CONS in H, then $R = \sum je_n \otimes je_n$; Vakhania and Tarieladze (1978). If each f in E' is a Gaussian random variable under μ, then μ is said to be Gaussian and we write $\mu = N(m, R)$.

Let $(S, \mathscr{S})$ and $(T, \mathscr{T})$ be measurable spaces, μ_{XY} a probability measure on

$\mathscr{S} \times \mathscr{T}$, μ_X and μ_Y the projections of μ_{XY}. The conditional distribution $\mu_{X|y}$, if it exists, is defined to be a probability measure on $\mathscr{S}$ for a.e. $d\mu_Y(y)$ such that $\mu_{X|y}(A)$ is measurable as a function of y for each fixed $A \in \mathscr{S}$ and

$$\mu_{XY}(A \times B) = \int_B \mu_{X|y}(A)\, d\mu_Y(y) \quad \text{for all} \quad A \in \mathscr{S} \quad \text{and} \quad B \in \mathscr{T}.$$

The following lemma, which is proved using Fubini's theorem, is a version of the abstract Bayes formula of Kallianpur and Striebel (1968).

LEMMA 2.1. *Suppose that the conditional distribution $\mu_{Y|x}$ exists, $\mu_{Y|x} \ll \mu_Y$* a.e. *$d\mu_X(x)$ and the map $(x, y) \to (d\mu_{Y|x}/d\mu_Y)(x)$ is $\mathscr{S} \times \mathscr{T}$ measurable. Then the conditional distribution $\mu_{X|y}$ exists, $\mu_{X|y} \ll \mu_X$* a.e. *$d\mu_Y(y)$ and*

$$\frac{d\mu_{X|y}}{d\mu_X}(x) = \frac{d\mu_{Y|x}}{d\mu_Y}(y) \quad \text{a.e.} \quad d\mu_X \otimes \mu_Y(x, y).$$

The probability measure μ_{XY} will be defined through a prior distribution μ_X on $\mathscr{S}$ for the signal and a noise distribution μ_N on $\mathscr{T}$. The spaces S and T represent the signal and observation spaces respectively. For a signal $x \in \mathscr{S}$ and noise $y \in T$, the observation is given by $f(x, y)$, where $f\colon S \times T \to T$ is an $\mathscr{S} \times \mathscr{T}/\mathscr{T}$ measurable map. Thus, assuming independent signal and noise, the joint distribution of signal and observation is given by

$$\mu_{XY}(A) = \mu_X \otimes \mu_N\{(x, y)\colon (x, f(x, y)) \in A\}, \quad A \in \mathscr{S} \times \mathscr{T}.$$

It is easily seen that $\mu_{Y|x}$ exists and is equal to $\mu_N \circ f_x^{-1}$, where $f_x\colon T \to T$ is defined by $f_x(y) = f(x, y)$. Under the hypotheses of Lemma 2.1, $\mu_{X|y}$ exists and is called the posterior distribution of the signal.

The basic framework of our signal + noise model is now described as follows. The observation space T is assumed to be a separable Banach space E. A Gaussian noise distribution $\mu_N = N(0, R_N)$ on $\mathscr{B}(E)$ is specified. The RKHS of the noise is denoted H_N and the corresponding injection denoted $j_N\colon H_N \to E$. The signal space S is taken to be H_N and $f(x, y) = j_N(x) + y$, for $x \in H_N$, $y \in E$. The prior distribution of the signal is assumed to be a Gaussian measure $\mu_X = N(m_X, R_X)$ on $\mathscr{B}(H_N)$. The reason for these assumptions is that $j_N(H_N)$ is the set of admissible translates of μ_N, i.e. translates of μ_N which are mutually absolutely continuous with respect to μ_N, Kuelbs (1970). In Example 1 we have

$$E = C[0, 1], \quad \mu_N = \text{Wiener measure on } \mathscr{B}(E), \quad H_N = L^2[0, 1]$$

$$j_N(x)(t) = \int_0^t x(s)\, ds, \quad 0 \le t \le 1, \quad x \in H_N.$$

$$R_X(x)(t) = \int_0^1 \frac{1}{2\beta} e^{-\beta|t-s|} x(s)\, ds, \quad 0 \le t \le 1, \quad x \in H_N.$$

Let $\mathscr{L}_N$ denote the closure of E' in $L^2(E, \mu_N)$, $U_N\colon \mathscr{L}_N \to H_N$ the unitary operator defined by $U_N f = j_N^* f$ for f in E'. Gaussian covariance operators on

Hilbert space are trace-class, see Kuo (1975), so that R_X has a series representation $R_X = \sum_n \tau_n e_n \otimes e_n$, where $\{e_n, n \geq 1\}$ is a CONS in H_N, $\tau_n \geq 0$ and $\text{tr}(R_X) = \sum_n \tau_n < \infty$. This particular CONS for H_N will be fixed throughout the remainder of the paper. The norm in H_N is denoted $\|\cdot\|$. The following result describes the posterior distribution $\mu_{X|y}$ on $\mathscr{B}(H_N)$.

PROPOSITION 2.2. *The posterior distribution $\mu_{X|y}$ exists as a probability measure on $\mathscr{B}(H_N)$ and is given by $\mu_{X|y} = N(m_{X|y}, R_{X|y})$, where*

$$m_{X|y} = \sum_n \frac{\tau_n}{1+\tau_n}\left\{[U_N^{-1}(e_n)](y) + \frac{\langle e_n, m_X\rangle}{\tau_n}\right\} e_n,$$

$$R_{X|y} = R_X(I + R_X)^{-1}.$$

PROOF. Denote $[U_N^{-1}(e_n)](y)$ by $\alpha_n(y)$. The α_n are i.i.d. $N(0, 1)$ random variables under μ_N so that $m_{X|y} \in H_N$ a.e. $d\mu_N(y)$. But, $\mu_N \circ f_x^{-1} \sim \mu_N$ for each $x \in H_N$ (cf. McKeague, 1982, Theorem 2.1) so that by Baker (1976) $\mu_Y \sim \mu_N$. Thus $m_{X|y} \in H_N$ a.e. $d\mu_Y(y)$ and the pair $(m_{X|y}, R_{X|y})$ defines a Gaussian measure on $\mathscr{B}(H_N)$ a.e. $d\mu_Y(y)$. Now check the conditions of Lemma 2.1. $\mu_{Y|x}$ exists and is equal to $\mu_N \circ f_x^{-1}$. Also $\mu_{Y|x} \sim \mu_N \sim \mu_Y$ for all $x \in H_N$. The map $(x, y) \to d\mu_{Y|x}/d\mu_Y(y)$ is $\mathscr{B}(H_N) \times \mathscr{B}(E)$ measurable since

$$\frac{d\mu_{Y|x}}{d\mu_Y}(y) = \frac{d\mu_N \circ f_x^{-1}}{d\mu_N}(y)\,\frac{d\mu_N}{d\mu_Y}(y)$$

$$= \frac{d\mu_N}{d\mu_Y}(y)\exp\{[U_N^{-1}(x)] - \tfrac{1}{2}\|x\|^2\}$$

$$= \frac{d\mu_N}{d\mu_Y}(y)\exp \sum \{\alpha_n(y)\,\langle e_n, x\rangle - \tfrac{1}{2}\langle e_n, x\rangle^2\},$$

where the Radon-Nikodym derivative $d\mu_N \circ f_x^{-1}/d\mu_N$ is given in McKeague (1982, Theorem 2.1), for instance. Now applying Lemma 2.1, the characteristic functional $\hat{\mu}_{X|y}(u) = \int_{H_N} e^{i\langle u,x\rangle}\, d\mu_{X|y}(x)$, for $u \in H_N$, as a function of u, is proportional to $\int_{H_N} \lim_{k\to\infty} Z_k(x)\, d\mu_X(x)$, where

$$Z_k(x) = \exp \sum_{n=1}^k \{i\langle e_n, u\rangle\langle e_n, x\rangle + \alpha_n(y)\langle e_n, x\rangle - \tfrac{1}{2}\langle e_n, x\rangle^2\}.$$

Provided that $\{Z_k, k \geq 1\}$ is uniformly integrable, the result now follows from routine calculations since the $\langle e_n, x\rangle$, $n \geq 1$ are independent $N(\langle e_n, m_X\rangle, \tau_n)$ random variables under μ_X. But

$$\int_{H_N} |Z_k(x)|^2\, d\mu_X(x) \leq \int_{H_N} \exp\{2 \sum_{n=1}^k \alpha_n(y)\langle e_n, x\rangle\}\, d\mu_X(x)$$

$$= \exp\{2 \sum_{n=1}^k (\alpha_n^2(y)\tau_n + \alpha_n(y)\langle e_n, m_X\rangle)\},$$

which shows that $\{Z_k, k \geq 1\}$ is a.e. $d\mu_Y(y)$ uniformly integrable with respect to μ_X, as required. □

3. QN-laws. In this section we derive the posterior distribution for the signal + noise model when either the prior or noise distribution is allowed to be a QN-law. We start with the definition and formulae for the mean and covariance of QN-laws.

Let E_1 and E_2 be separable Banach spaces. Suppose that $\mu = N(m, R)$ on $\mathscr{B}(E_1)$ with RKHS denoted H and injection $j:H \to E_1$; also let $A:E_2 \to E_2'$ be a symmetric nonnegative operator, $a \in E_2$ and $J:E_1 \to E_2$ be a bounded linear operator. Denote $c^{-1} = \int_{E_1} (1 + \langle A(J(x)- a), J(x)- a\rangle)\, d\mu(x)$. Note that $c^{-1} < \infty$ since $\int_{E_1} \| x \|_{E_1}^2\, d\mu(x) < \infty$ by Fernique (1970). Define a probability measure ν on $\mathscr{B}(E_1)$ by the relation

$$(d\nu/d\mu)(x) = c(1 + \langle A(J(x)- a), J(x)- a\rangle).$$

The measure ν, written $\nu = \mathrm{QN}((J, a, A), \mu)$, is called a QN-law and was introduced by Gualtierotti (1979). When $E_1 = E_2$ and J is the identity map write $\nu = \mathrm{QN}((a, A), \mu)$.

The statistical significance of the parameters J, a, A can be described as follows. The operator A controls the amount and direction of the non-Gaussian contribution to ν. For instance, if A is a projection onto the span of an element b then ν can be non-Gaussian in the direction of b and Gaussian in directions orthogonal to b. The element a controls the origin of the non-Gaussian contribution. The need for two spaces E_1, E_2 and the operator J transferring the effect of A and a to the E_1 space arises intrinsically in defining the posterior distribution when the noise is a QN-law, as will be seen in Proposition 3.2.

In Example 1 we have $\nu_N = \mathrm{QN}((0, \varepsilon A), \mu_N)$ on $\mathscr{B}(E)$, where $E = C[0, 1]$ and $A:E \to E'$ is defined by

$$\langle Ax, y\rangle = \int_0^1 x(t)y(t)\, dt, \quad x, y \in E.$$

The contaminated prior distribution $\nu_X = \mathrm{QN}((0, \varepsilon A), \mu_X)$ on $\mathscr{B}(H_N)$, where A is the identity operator on H_N.

Gualtierotti (1980) calculated the mean and covariance of QN-laws on separable Hilbert space. It is possible to extend this result to separable Banach spaces as follows.

LEMMA 3.1. (i) *j^*J^*AJj is a trace-class operator on H and* $c^{-1} = 1 + \mathrm{tr}(j^*J^*AJj) + \langle A(J(m)- a), J(m)- a\rangle$.

(ii) *The mean m^Q and covariance operator R^Q of $\nu = \mathrm{QN}((J, a, A), \mu)$ are given by*

$$m^Q = m + u, \quad R^Q = R + 2cRJ^*AJR - u \otimes u,$$

where $u = 2cRJ^*A(J(m)- a)$.

PROOF. (Sketch) Assume that $m = 0$ and consider just the evaluation of R^Q. The operator $J^*AJ:E_1 \to E_1'$, is nonnegative and symmetric so that, by Schwartz (1964), there exists a separable Hilbert space H_1 and a continuous linear injection $i:H_1 \to E_1'$, such that $J^*AJ = ii^*$. Let $\{u_n, n \geq 1\}$ be a CONS in H_1 and let $g_n =$

$i(u_n)$, $n \geq 1$. Then it is easily seen that $J^*AJ = \sum_n g_n \otimes g_n$, where $g_n \in E_1'$. Thus, for $f \in E_1'$

$$\int_{E_1} \langle f, x\rangle^2 \langle AJ(x), J(x)\rangle \, d\mu(x) = \sum_n \int_{E_1} \langle f, x\rangle^2 \langle g_n, x\rangle^2 \, d\mu(x),$$

so that we can reduce to evaluating integrals of the form $\int_{E_1} \langle f, x\rangle^2 \langle g, x\rangle^2 \, d\mu(x)$. Choose $h_n \in E_1'$ such that $j^*(h_n)$, $n \geq 1$ is a CONS for H. Define

$$\pi_k x = \sum_{n=1}^k \langle h_n, x\rangle Rh_n, \quad x \in E_1.$$

Then, by Tien (1978, Lemma 2), $\pi_k x$ converges a.s. $[\mu]$ to x. But

$$\begin{aligned} &\int_{E_1} \langle f, \pi_k x\rangle^4 \langle g, \pi_k x\rangle^4 \, d\mu(x) \\ &\quad \leq \left\{ \int_{E_1} \langle f, \pi_k x\rangle^8 \, d\mu(x)\right\}^{1/2} \left\{\int_{E_1} \langle g, \pi_k x\rangle^8 \, d\mu(x)\right\}^{1/2} \\ &\quad \leq 105 \, \langle Rf, f\rangle^2 \langle Rg, g\rangle^2, \end{aligned}$$

since $\langle f, \pi_k x\rangle$ is a $N(0, \sum_{n=1}^k \langle Rh_n, f\rangle^2)$ random variable and $\sum_{n=1}^k \langle Rh_n, f\rangle^2 \leq \langle Rf, f\rangle$. It follows that $\{\langle f, \pi_k x\rangle^2 \langle g, \pi_k x\rangle^2, k \geq 1\}$ is uniformly integrable and the Lebesgue convergence theorem can be applied. The integral $\int_{E_1} \langle f, \pi_k x\rangle^2 \langle g, \pi_k x\rangle^2 \, d\mu(x)$ can be calculated using the fact that $\langle h_n, x\rangle$, $n \geq 1$ is an iid $N(0, 1)$ sequence of random variables with respect to μ. □

The next proposition shows that the posterior is a QN-law if either the prior is Gaussian and the noise is a QN-law or the prior is a QN-law and the noise is Gaussian. Let $\mu_N = N(0, R_N)$, $\mu_X = N(m_X, R_X)$ as in Section 2 and let $\mu_{X|y}$ denote the corresponding posterior distribution given in Proposition 2.2.

PROPOSITION 3.2. (i) *If the prior is* $\mu_X = N(m_X, R_X)$ *and the noise is* $\nu_N = \mathrm{QN}((a, A), \mu_N)$ *then the posterior is* $\nu_{X|y} = \mathrm{QN}((j_N, y - a, A), \mu_{X|y})$.

(ii) *If the prior is* $\nu_X = \mathrm{QN}((a, A), \mu_X)$ *and the noise is* $\mu_N = N(0, R_N)$ *then the posterior is* $\nu_{X|y} = \mathrm{QN}((a, A), \mu_{X|y})$.

The proof of this proposition uses the following consequence of Lemma 2.1.

LEMMA 3.3. *Let* μ_{XY} *and* ν_{XY} *be probability measures on* $\mathscr{S} \times \mathscr{T}$ *such that*

(a) $\mu_X \sim \nu_X$ *and* $\mu_Y \sim \nu_Y$;

(b) $\mu_{Y|x}$ *and* $\nu_{Y|x}$ *exist and* $\mu_{Y|x} \sim \nu_{Y|x}$ *a.e.* $d\mu_X(x)$;

(c) $\mu_{Y|x} \ll \mu_Y$ *a.e.* $d\mu_X(x)$;

(d) *the maps* $(x, y) \mapsto d\nu_{Y|x}/d\mu_{Y|x}(y)$, $(x, y) \mapsto d\mu_{Y|x}/d\mu_Y(y)$ *are* $\mathscr{S} \times \mathscr{T}$ *measurable. Then* $\nu_{X|y}$ *exists,* $\nu_{X|y} \sim \mu_{X|y}$ *a.e.* $d\mu_Y(y)$ *and*

$$\frac{d\nu_{X|y}}{d\mu_{X|y}}(x) = \frac{d\mu_Y}{d\nu_Y}(y) \, \frac{d\nu_{Y|x}}{d\mu_{Y|x}}(y) \, \frac{d\nu_X}{d\mu_X}(x) \quad a.e. \quad d\mu_X \otimes \mu_Y(x, y).$$

Proof. From (a) – (c) it follows that $\nu_{Y|x} \ll \nu_Y$ a.e. $d\nu_X(x)$ and

$$\frac{d\nu_{Y|x}}{d\nu_Y}(y) = \frac{d\nu_{Y|x}}{d\mu_{Y|x}}(y)\,\frac{d\mu_{Y|x}}{d\mu_Y}(y)\,\frac{d\mu_Y}{d\nu_Y}(y) \quad \text{a.e.} \quad d\mu_X \otimes \mu_Y(x, y)$$

so that, by (d), the function $(x, y) \mapsto d\nu_{Y|x}/d\nu_Y(y)$ is $\mathscr{S} \times \mathscr{T}$ measurable and $\nu_{X|y}$ exists by Lemma 2.1. The proof is completed by applying Bayes formula.

Proof of Proposition 3.2. (i) We check the conditions of Lemma 3.3. $\mu_Y \sim \nu_Y$ since $\mu_{Y|x} \sim \nu_{Y|x}$ for all x in H_N. $\mu_{Y|x} \ll \mu_Y$ a.e. $d\mu_X(x)$ by the proof of Proposition 2.2.

$$\frac{d\nu_{Y|x}}{d\mu_{Y|x}}(y) = c_N(1 + \langle A(y - a - j_N x), y - a - j_N x\rangle),$$

so that the map $(x, y) \mapsto d\nu_{Y|x}/d\mu_{Y|x}(y)$ is $\mathscr{B}(H_N) \times \mathscr{B}(E)$ measurable. The map $(x, y) \mapsto d\mu_{Y|x}/d\mu_Y(y)$ is $\mathscr{B}(H_N) \times \mathscr{B}(E)$ measurable from the proof of Proposition 2.2. Thus, by Lemma 3.3, $\nu_{X|y}$ exists and

$$\frac{d\nu_{X|y}}{d\mu_{X|y}}(x) = \frac{d\mu_Y}{d\nu_Y}(y)c_N(1 + \langle A(j_N x - (y - a)), j_N x - (y - a)\rangle),$$

which shows that $\nu_{X|y} = \mathrm{QN}((j_N, y - a, A), \mu_{X|y})$. The proof of (ii) is similar. □

4. The performance of δ_0 under QN-law contamination. Let δ denote a decision rule for estimating the true signal. δ is a measurable function from the observation space E into the signal space H_N. For prior ν_X and noise ν_N the mean square error of δ is given by

$$r(\nu_x, \nu_N, \delta) = \int_{H_N \times E} \| x - \delta(y) \|^2 \, d\nu_{XY}(x, y),$$

where $\| \cdot \|$ is the norm in the signal space H_N. In Example 1, where $H_N = L^2[0, 1]$, we have

$$r(\nu_X, \nu_N, \delta) = E \int_0^1 (X_t - \delta[Y](t))^2 \, dt,$$

which is a reasonable measure of the closeness of the estimate $\delta[Y]$ to the signal $X = (X_t)$.

The following quantities are natural measures of the performance of an estimator δ_0: the increase in the mean square error of δ_0 over the Bayes loss (the minimum possible mean square error),

$$\Delta(\nu_X, \nu_N, \delta_0) = r(\nu_X, \nu_N, \delta_0) - \inf_\delta r(\nu_X, \nu_N, \delta),$$

and the ratio of the mean square error of δ_0 to the Bayes loss,

$$\Phi(\nu_X, \nu_N, \delta_0) = \frac{r(\nu_X, \nu_N, \delta_0)}{\inf_\delta r(\nu_X, \nu_N, \delta)}.$$

If $\Phi(\nu_X, \nu_N, \delta_0)$ is close to 1 we would be satisfied that the performance of δ_0 is close to optimal under ν_X and ν_N.

Now fix δ_0 as the Bayes estimator (i.e. the optimal estimator in the mean square sense) for Gaussian prior $\mu_X = N(m_X, R_X)$ and Gaussian noise $\mu_N = N(0, R_N)$. δ_0 is given by the posterior mean calculated in Proposition 2.2, $\delta_0(y) = m_{X|y}$. The results of this section give some upper bounds on $\Delta(\nu_X, \nu_N, \delta_0)$ and $\Phi(\nu_X, \nu_N, \delta_0)$ for ν_X and ν_N as QN-law contaminations of μ_X and μ_N respectively. First we evaluate the mean square error of δ_0 under contaminated prior or contaminated noise. Denote $R_1 = R_{X|y} = R_X(I + R_X)^{-1}$.

LEMMA 4.1. (i) *Let* $\nu_X = \mathrm{QN}((a, A), \mu_X)$. *Then*

$$r(\nu_X, \mu_N, \delta_0) = \mathrm{tr}(R_1) + 2c_X\mathrm{tr}(AR_1^2),$$

where $c_X^{-1} = 1 + \mathrm{tr}\, AR_X + \langle A(m_X - a), m_X - a\rangle$.

(ii) *Let* $\nu_N = \mathrm{QN}((a, A), \mu_N)$. *Then*

$$r(\mu_X, \nu_N, \delta_0) = \mathrm{tr}(R_1) + 2c_N\mathrm{tr}(A_N R_1^2),$$

where $A_N = j_N^* A j_N$ *and* $c_N^{-1} = 1 + \mathrm{tr}(A_N) + \langle Aa, a\rangle$.

PROOF. (i) $r(\nu_X, \nu_N, \delta_0) = \int_{H_N}\int_E \| m_{X|y} - x \|^2 \, d\mu_{Y|x}(y)\, d\nu_X(x)$. But

$$m_{X|y} - x = \sum_{n\ge1} \frac{\tau}{1+\tau_n}\left\{[U_N^{-1}(e_n)](y) - \langle x, e_n\rangle - \frac{\langle x - m_X, e_n\rangle}{\tau_n}\right\}e_n,$$

so that

$$\int_E \| m_{X|y} - x \|^2 \, d\mu_{Y|x}(y)$$

$$= \sum_{n\ge1}\left(\frac{\tau_n}{1+\tau_n}\right)^2 \int_E \left\{[U_N^{-1}(e_n)](y) - \langle x, e_n\rangle - \frac{\langle x - m_X, e_n\rangle}{\tau_n}\right\}^2 d\mu_{Y|x}(y)$$

$$= \sum_{n\ge1}\left(\frac{\tau_n}{1+\tau_n}\right)^2\left(1 + \frac{\langle x - m_X, e_n\rangle^2}{\tau_n^2}\right),$$

since $[U_N^{-1}(e_n)](y) - \langle e_n, x\rangle$ is a $N(0, 1)$ random variable under $\mu_{Y|x}$. By Lemma 3.1

$$\int_{H_N} \langle e_n, x - m_X\rangle^2 \, d\nu_X(x) = \tau_n + 2c_X\tau_n^2\langle Ae_n, e_n\rangle,$$

so that

$$r(\nu_X, \mu_N, \delta_0) = \sum_{n\ge1}\left(\frac{\tau_n}{1+\tau_n}\right)^2\left(1 + \frac{1}{\tau_n} + 2c_X\langle Ae_n, e_n\rangle\right)$$

$$= \mathrm{tr}(R_X(I + R_X)^{-1}) + 2c_X\mathrm{tr}(AR_X^2(I + R_X)^{-2}).$$

(ii) is proved in a similar way. □

The following theorem gives an upper bound on the increase in the mean square error of δ_0 over the minimum possible mean square error under a contaminated prior distribution. If V is a bounded linear operator on H_N then $\| V \|$ denotes its operator norm.

THEOREM 4.2. *Let $\nu_X = \mathrm{QN}((a, A), \mu_X)$. Then*

$$\Delta(\nu_X, \mu_N, \delta_0)$$
$$\leq 4c_1^2 \| R_1 A \|^2 [\mathrm{tr}\, R_X R_1 + 2c_X \mathrm{tr}\, AR_1^2 + (1 + 4c_X \| AR_X R_1 \|) \| m_X - a \|^2],$$

where $c_1^{-1} = 1 + \mathrm{tr}(AR_1)$.

PROOF. It is easily checked that $\Delta(\nu_X, \mu_N, \delta_0) = \int_E \| m_{X|y} - m^Q_{X|y} \|^2 \, d\nu_Y(y)$. By Proposition 3.2 and Lemma 3.1, $m^Q_{X|y} = m_{X|y} + 2c_{X|y} R_{X|y} A(m_{X|y} - a)$, so that

$$\Delta(\nu_X, \mu_N, \delta_0) \leq 4c_1^2 \| R_1 A \|^2 \int_E \| m_{X|y} - a \|^2 \, d\nu_Y(y).$$

Now consider

$$\int_E \| m_{X|y} - a \|^2 \, d\nu_Y(y) = \int_{H_N} \int_E \| m_{X|y} - a \|^2 \, d\mu_{Y|x}(y) \, d\nu_X(x).$$

$$\int_E \| m_{X|y} = a \|^2 \, d\mu_{Y|x}(y)$$
$$= \sum_{n\geq 1} \left(\frac{\tau_n}{1 + \tau_n}\right)^2$$
$$\int_E \left\{[U_N^{-1}(e_N)](y) - \langle e_n, x\rangle + \langle e_n, x - a\rangle + \frac{\langle e_n, m_X - a\rangle}{\tau_n}\right\}^2 d\mu_{Y|x}(y)$$
$$= \sum_{n\geq 1} \left(\frac{\tau_n}{1 + \tau_n}\right)^2 \left(1 + \left\{\langle e_n, x - a\rangle + \frac{\langle e_n, m_X - a\rangle}{\tau_n}\right\}^2\right).$$

Use Lemma 3.1 to get

$$\int_{H_N} \left\{\langle e_n, x - a\rangle + \frac{\langle e_n, m_X - a\rangle}{\tau_n}\right\}^2 d\nu_x(x)$$
$$= \tau_n + 2c_X\langle R_X A R_X e_n, e_n\rangle$$
$$+ 4c_X\left(\frac{1 + \tau_n}{\tau_n}\right)\langle e_n, R_X A(m_X - a)\rangle\langle e_n, m_X - a\rangle + \left(\frac{1 + \tau}{\tau_n}\right)^2 \langle e_n, m_X - a\rangle^2.$$

This yields

$$\int_E \| m_{X|y} - a \|^2 \, d\nu_Y(y)$$

$$= \sum_{n\geq 1} \left\{ \tau_n^2(1+\tau_n)^{-1} + 2c_X\left(\frac{\tau_n}{1+\tau_n}\right)^2 \langle R_X A R_X e_n, e_n\rangle \right.$$

$$\left. + 4c_X \langle A R_X^2 (I+R_X)^{-1} e_n, m_X - a\rangle \langle e_n, m_X - a\rangle + \langle e_n, m_X - a\rangle^2 \right\}$$

$$\leq \operatorname{tr} R_X^2 (I+R_X)^{-1} + 2c_X \operatorname{tr} A R_X^4 (I+R_X)^{-2}$$

$$+ 4c_X \| A R_X^2 (I+R_X)^{-1} \| \, \| m_X - a \|^2 + \| m_X - a \|^2,$$

and the result follows. □

COROLLARY 4.3. *Let* $\nu_X = \mathrm{QN}((a, \varepsilon A), \mu_X)$, *where* $\varepsilon > 0$. *Then*

$$\Phi(\nu_X, \mu_N, \delta_0) \leq 1 + \frac{4\| R_1 A \|^2 [\operatorname{tr}(R_X R_1) + \| m_X - a \|^2]}{\operatorname{tr}(R_1)} (1 + o(1))\varepsilon^2,$$

$$as \quad \varepsilon \to 0.$$

In particular, $\Phi(\nu_X, \mu_N, \delta_0) = 1 + O(\varepsilon^2)$, $\varepsilon \to 0$.

PROOF. The result follows from Proposition 4.1, Theorem 4.2 and the identity

$$\Phi(\nu_X, \mu_N, \delta_0) = 1 + \frac{\Delta(\nu_X, \mu_N, \delta_0)}{r(\nu_X, \mu_N, \delta_0) - \Delta(\nu_X, \mu_N, \delta_0)}.$$

If $\nu_X = \mathrm{QN}((0, \varepsilon I), \mu_X)$ where I is the identity on H_N it follows from Corollary 4.3 and the inequality $\operatorname{tr}(R_X R_1) \leq \| R_X \| \operatorname{tr}(R_1)$ that

$$\Phi(\nu_X, \mu_N, \delta_0) \leq 1 + 4\| R_X \| (1 + o(1))\varepsilon^2, \quad \text{as} \quad \varepsilon \to 0.$$

Thus, in Example 1 where $\| R_X \| < 1/2\beta$ we have

$$\Phi(\nu_X, \mu_N, \delta_0) \leq 1 + (2/\beta)(1 + o(1))\varepsilon^2,$$

as stated in Section 1. The next theorem gives an upper bound on the increase in the mean square error of δ_0 over the minimum possible mean square error under a contaminated noise distribution.

THEOREM 4.4. *Let* $\nu_N = \mathrm{QN}((a, A), \mu_N)$. *Then*

$$\Delta(\mu_X, \nu_N, \delta_0) \leq 8c_2^2 \{ \| R_1 A_N \|^2 [\operatorname{tr} R_X R_1 + 2c_N \operatorname{tr} A_n R_1^2]$$

$$+ \operatorname{tr} R_1^2 (A_N R_X A_N + A_N^2 + 2c_N A_N^3) + (1 + 4c_N \| A_N \|) \langle A R_N A a, a\rangle \},$$

where $A j_N = j_N^* A j_N$ *and* $c_2^{-1} = 1 + \operatorname{tr}(A_N R_1)$.

PROOF. By Proposition 3.2, $\nu_{X|y} = \mathrm{QN}((j_N, y - a, A), \mu_{X|y})$, and by Lemma 3.1, $m^Q_{X|y} = m_{X|y} + 2c_{X|y}R_1 j_N^* A(j_N m_{X|y} - y + a)$. Thus

$$\begin{aligned}\Delta(\mu_X, \nu_N, \delta_0) &= \int_E \| m_{X|y} - m^Q_{X|y} \|^2 \, d\nu_Y(y)\\ &\le 4c_2^2 \int_E \| R_1 j_N^* A(j_N m_{X|y} - y + a) \|^2 \, d\nu_Y(y)\\ &\le 8c_2^2 \Big[\| R_1 A_N \|^2 \int_E \| m_{X|y} - m_X \|^2 \, d\nu_Y(y)\\ &\qquad + \int_E \| R_1 j_N^* A(j_N m_X - y + a) \|^2 \, d\nu_Y(y) \Big].\end{aligned}$$

It is easily checked that

$$\int_E \| m_{X|y} - m_X \|^2 \, d\nu_Y(y) = \mathrm{tr}(R_X R_1) + 2c_N \mathrm{tr}(A_N R_1^2).$$

Note that $m^Q_Y = j_N m_X + u$ and $R^Q_Y = j_N R_X j_N^* + R_N + 2c_N R_N A R_N - u \otimes u$, where $u = -2c_N R_N A(a)$. Hence

$$\begin{aligned}&\int_E \| R_1 j_N^* A(j_n m_X - y + a) \|^2 \, d\nu_Y(y)\\ &\quad = \mathrm{tr}(R_1^2 j_N^* A R^Q_Y A j_N) + \| R_1 j_N^* A(a - u) \|^2\\ &\quad = \mathrm{tr}(R_1^2(A_N R_X A_N + A_N^2 + 2c_N A_N^3)) - \| R_1 j_N^* A(u) \|^2 + \| R_1 j_N^* A(a - u) \|^2\\ &\quad = \mathrm{tr}(R_1^2(A_N R_X A_N + A_N^2 + 2c_N A_N^3)) + \| R_1 j_N^* A(a) \|^2\\ &\qquad + 4c_N \langle R_1 j_N^* A(a), R_1 A_N j_N^* A(a) \rangle\\ &\quad \le \mathrm{tr}(R_1^2(A_N R_X A_N + A_N^2 + 2c_N A_N^3)) + (1 + 4c_N \| A_N \|) \langle A R_N A a, a \rangle.\end{aligned}$$

The result follows immediately. □

COROLLARY 4.5. *Let* $\nu_N = \mathrm{QN}((a, \varepsilon A), \mu_N)$, *where* $\varepsilon > 0$. *Then*

$$\Phi(\mu_X, \nu_N, \delta_0) \le 1 + \frac{8[\| R_1 A_N \|^2 \mathrm{tr} R_X R_1 + \mathrm{tr}\, R_1^2(A_N R_X A_N + A_N^2) + \langle A R_N A a, a \rangle]}{\mathrm{tr}(R_1)} \cdot (1 + o(1))\varepsilon^2,$$

as $\varepsilon \to 0$. *In particular*, $\Phi(\mu_X, \nu_N, \delta_0) = 1 + O(\varepsilon^2)$, $\varepsilon \to 0$.

If $\nu_N = \mathrm{QN}((0, \varepsilon A), \mu_N)$ it follows from Corollary 4.5 that

$$\Phi(\mu_X, \nu_N, \delta_0) \le 1 + 24 \| R_X \| \, \| A_N \|^2 (1 + o(1))\varepsilon^2, \quad \text{as} \quad \varepsilon \to 0.$$

In example 1, where $H_N = L^2[0, 1]$, it is easily checked that the operator A_N is given by

$$A_N(x)(t) = \int_0^1 (s \wedge t)x(s)\, ds, \quad 0 \le t \le 1, \quad x \in H_N.$$

Since $\| A_N \| < 1/2$ and $\| R_X \| < 1/(2\beta)$ it follows that

$$\Phi(\mu_X, \nu_N, \delta_0) \le 1 + (3/\beta)(1 + o(1))\varepsilon^2,$$

as $\varepsilon \rightarrow 0$, as stated in Section 1.

The upper bounds obtained for Φ under contaminated signal and contaminated noise differ by a factor of $2/3$. Since these bounds are fairly tight, it seems reasonable to conclude that contamination in the noise is only slightly, if at all, more serious than contamination in the signal.

REFERENCES

BAKER, C. R. (1976). Absolute continuity and applications to information theory. Probability in Banach Spaces, Oberwolfach 1975. *Lecture Notes in Math.* **526** 1–11. Springer-Verlag, Berlin.

BAXENDALE, P. (1976). Gaussian measures on function spaces. *Amer. J. Math.* **98** 891–952.

BERGER, J. O. (1982). The robust Bayesian viewpoint. Technical Report #82-9, Department of Statistics, Purdue University.

ERSHOV, A. A. and LIPTSER, R. S. (1978). Robust Kalman filter in discrete time. *Automat. Remote Control* **39** 359–367.

FERNIQUE, X. (1970). Integrabilite des vecteurs gaussiens. *Comptes Rendus Acad. Sci. Paris* **A270** 1698–1699.

GUALTIEROTTI, A. F. (1979). Sur la détection des signaux, l'information de Shannon, et la capacité de retransmission, quand les lois du modèle ne sont pas exactement normales. *Comptes Rendus Acad. Sci. Paris* **A288** 69–71.

GUALTIEROTTI, A. F. (1980). On average mutual information and capacity for a channel without feedback and contaminated Gaussian noise. *Inform. Control* **46** 46–70.

GUALTIEROTTI, A. F. (1982). A class of stochastic processes with a law generalizing a nondecomposable law on the real line. *Stochastic Process. Appl.* **13** 87–117.

GUALTIEROTTI, A. F. (1983). On the stablity of signal detection. *IEEE Trans. Inform. Theory*, to appear.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

KALLIANPUR, G. and STRIEBEL, C. (1968). Estimation of stochastic processes with additive white noise observation errors. *Ann. Math. Statist.* **39** 785–801.

KUELBS, J. (1970). Gaussian measures on a Banach space. *J. Functional Anal.* **5** 354–367.

KUO, H.-H. (1975). Guassian measures in Banach spaces. *Lecture Notes in Math.* **463**. Springer-Verlag, Berlin.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes*, Volume II. Springer-Verlag, New York.

MANDELBAUM, A. (1984). Linear estimators and measurable linear transformations on a Hilbert space. *Z. Wahrsch. verw. Gebiete* **65** 385–397.

MASRELIEZ, C. G. and MARTIN, R. D. (1977). Robust Bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Trans. Automat. Contr.* **AC-22** 361–371.

MCKEAGUE, I. W. (1982). On the capacity of channels with Gaussian and non-Gaussian noise. *Inform. Control.* **51** 153–173.

SCHWARTZ, L. (1964). Sous-espaces Hilbertiens d'espaces vectoriels topologiques et noyaux associés. *J. Analyse Math.* **13** 115–256.

TIEN, N. Z. (1979). The structure of measurable linear functionals on Banach spaces with Gaussian measures. *Theory Probab. Appl.* **24** 165–168.

VAKHANIA, N. N. and TARIELADZE, V. I. (1978). Covariance operators of probability measures in locally convex spaces. *Theory Probab. Appl.* **23** 1–21.

WERON, A. (1976). On weak second order and Gaussian random elements. *Lecture Notes in Math.* **526**. Springer-Verlag, Berlin.

DEPARTMENT OF STATISTICS
THE FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306