NGR-05-009-079

N69-25589

Quantum Detection and Estimation Theory

Carl W. Helstrom*

Department of Applied Electrophysics
University of California, San Diego
La Jolla, California 92037

ABSTRACT

A review. Quantum detection theory is a reformulation, in quantum-mechanical terms, of statistical decision theory as applied to the detection of signals in random noise. Density operators take the place of the probability density functions of conventional statistics. The optimum procedure for choosing between two hypotheses, and an approximate procedure valid at small signal-to-noise ratios and called threshold detection, are presented. Quantum estimation theory seeks best estimators of parameters of a density operator. A quantum counterpart of the Cramér-Rao inequality of conventional statistics sets a lower bound to the mean-square errors of such estimates. Applications at present are primarily to the detection and estimation of signals of optical frequencies in the presence of thermal radiation.

Keywords:
signal detection, detection theory, parameter estimation, statistical estimation, estimation theory, quantum theory, decision theory, hypothesis testing, statistical decisions.

I. Quantum Statistical Theory

Much of statistical theory can be viewed as the calculation of expected values. Classically, a system characterized by the variables $x_1, x_2, \ldots, x_n$ has associated with it a probability density function (p.d.f.) $p(x_1, x_2, \ldots, x_n)$, and the expectations of certain measurable functions $f(x_1, x_2, \ldots, x_n)$,

$$\underset{\sim}{E}[f(x_1, \ldots, x_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n, \tag{1.1}$$

are required. Quantum-mechanically a system is described by a density operator ρ, which is a function of the dynamical variables of the system, and the expected value of an observable whose quantum-mechanical operator is F is given by the trace[1]

$$\underset{\sim}{E}(F) = \mathrm{Tr}(\rho F). \tag{1.2}$$

The density operator ρ is the quantum counterpart of the p.d.f. $p(x_1, \ldots, x_n)$. When, as in the classical limit, ρ is diagonal in a representation based on the simultaneous eigenstates $|x_1 \ldots x_n\rangle$ of the operators $X_1, \ldots, X_n$ corresponding to the variables $x_1, \ldots, x_n$,

$$\rho = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |x_1 \ldots x_n\rangle\, p(x_1, \ldots, x_n)\, \langle x_1 \ldots x_n|\, dx_1 \ldots dx_n, \tag{1.3}$$

the expectation in Eq. (1.2) reduces to Eq. (1.1), with

$$f(x_1, \ldots, x_n) = \langle x_1 \ldots x_n | F | x_1 \ldots x_n \rangle. \tag{1.4}$$

Quantum statistical theory includes the classical as a special case[2].

Modern statistical theory also has a normative and methodological aspect, which appears in its treatment of hypothesis testing and estimation.

It seeks the best procedures for making statements about the condition of a system under observation, statements that are framed as decisions among hypotheses about the system, or as estimates of numerical parameters characterizing it. The statements are based on observational data subject to unavoidable random error. The best methods are those that minimize the influence of error, and by evaluating their quality it is possible to determine the ultimate limits imposed by statistical uncertainty on the accuracy of decisions and measurements[3].

In classical physics statistical uncertainty is largely due to the presence of random noise, which originates primarily in molecular chaos. Statistical hypothesis-testing or decision theory has been extensively applied to the detection of acoustic and electromagnetic signals in noise and permits defining the weakest signal that can be detected with a specified probability of error, as a function of the strength of the interfering noise[4-9]. Estimation theory has been applied to the measurement of signal parameters such as amplitude, carrier frequency, and time of arrival, which are important in telemetry and radar. The noise sets a limit to the accuracy of such measurements.

The subject of this review is the formulation of statistical decision and estimation theory in quantum-mechanical terms. It involves replacing the probability density functions that appear in the classical theory by quantum-mechanical density operators. Although the context will be the detection of signals at optical frequencies and the estimation of their parameters, the application of these concepts is not limited thereto. The aim of quantum detection and estimation theory is to determine how the reliability of decisions and parameter estimates is affected both by random noise and by quantum-mechanical uncertainty.

Classical Decision and Estimation Theory

Decision theory treats the choice among hypotheses about the system at hand. In the simplest binary decision there are two hypotheses, exemplified by the absence or presence of a signal $s(t)$ of known form in the input $x(t)$ to a receiver during a certain observation interval $(0, T)$. The hypotheses are then

H_0 (null hypothesis): $x(t) = n(t)$,

H_1 (alternative hypothesis): $x(t) = n(t) + s(t)$,

where $n(t)$ is a random process representing noise with certain specified statistical properties. We suppose that the decision is to be based on n samples $x_i = x(t_i)$ of the input $x(t)$ during the interval $(0, T)$, $(i = 1, 2, \ldots, n)$. The p.d.f.'s $p_0(x_1, \ldots, x_n)$ and $p_1(x_1, \ldots, x_n)$ of these data under the two hypotheses are known. The best method of deciding between them is sought.

The adjective "best" is principally defined in two ways. In "Bayesian" decision theory the observer knows the prior probabilities ζ and $(1-\zeta)$ of hypotheses H_0 and H_1, and he also knows the four costs C_{ij} of choosing hypothesis H_i when H_j is true $(i, j = 0, 1)$[10]. The costs are entailed by the actions and circumstances following the decision, which is to be made in such a way that the average cost is minimum. This so-called "Bayes strategy" requires H_1 to be selected whenever[11]

$$\Lambda(x_1, \ldots, x_n) = \frac{p_1(x_1, \ldots, x_n)}{p_0(x_1, \ldots, x_n)} \geq \frac{\zeta(C_{10}-C_{00})}{(1-\zeta)(C_{01}-C_{11})} = \Lambda_0 \; . \tag{1.5}$$

Otherwise H_0 is selected. The function $\Lambda(x_1, \ldots, x_n)$ is called the <u>likelihood ratio</u>.

Decisions among more than two hypotheses can be treated in a similar manner. Often the costs associated with the various errors can be set equal, whereupon it is the average probability of error that is to be minimized. The best strategy is then to choose the hypothesis whose posterior or conditional probability, given the data $(x_1, \ldots, x_n)$, is greatest[12]. The posterior probability can be expressed in terms of likelihood ratios between pairs of p.d.f.'s for the data under the several hypotheses.

The second way of defining a "best" binary decision procedure is provided by the theory of Neyman and Pearson[13, 14]. Two kinds of errors can occur. Choosing H_1 when H_0 is true is called an error of the first kind, or false alarm; its probability under a given decision strategy is denoted by Q_0. Choosing H_0 when H_1 is true is an error of the second kind, or false dismissal; its probability is Q_1. The complement $Q_d = 1 - Q_1$ is often called the probability of detection[15]. That strategy is now considered best that attains the maximum probability Q_d of detection for a set false-alarm probability Q_0. It leads to the same comparison of the likelihood ratio $\Lambda(x_1, \ldots, x_n)$ with a decision level Λ_0 as in Eq. (1.3), but with Λ_0 fixed so that the false-alarm probability equals the pre-assigned value[16]. The Neyman-Pearson criterion dispenses with the prior probabilities and costs needed for the Bayesian approach, but is not easily generalized to decisions among more than two hypotheses.

Estimation theory typically treats data $\underset{\sim}{x} = (x_1, \ldots, x_n)$ whose joint p.d.f. $p(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m) = p(\underset{\sim}{x}; \underset{\sim}{\theta})$ depends on some unknown parameters $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_m)$ that are to be estimated. For instance, the data may be samples $x_j = x(t_j)$ of the input

$$x(t) = s(t; \underset{\sim}{\theta}) + n(t)$$

to a receiver, composed of noise $n(t)$ of known statistical properties and a signal $s(t; \underset{\sim}{\theta})$ depending on parameters $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_m)$ such as amplitude, time of arrival, and carrier frequency. On the basis of the n data $\underset{\sim}{x}$, the values of these parameters are to be estimated as accurately as possible.

Estimation theory sets up a measure of the cost or seriousness of errors in the estimates $\hat{\underset{\sim}{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)$ of the parameters. The most common cost function is a weighted sum of the squared errors,

$$C(\hat{\underset{\sim}{\theta}}, \underset{\sim}{\theta}) = \sum_{k=1}^{m} w_k(\hat{\theta}_k - \theta_k)^2. \tag{1.6}$$

The problem is to find estimates $\hat{\theta}_k = \hat{\theta}_k(x_1, \ldots, x_n)$ as such functions of the data that the average cost is minimum. Of interest also are lower bounds on the sizes of the errors, measured usually by the mean-square deviations $\underset{\sim}{E}(\hat{\theta}_k - \theta_k)^2$, as well as the bias of each estimate, defined as the deviation

$$E(\hat{\theta}_k) - \theta_k$$

of the expected value of the estimate from the true value of the parameter[17].

The Generalization to Quantum Theory

Central to classical decision and estimation theory are the p.d.f.'s $p_0(\underset{\sim}{x})$, $p_1(\underset{\sim}{x})$, and $p(\underset{\sim}{x}; \underset{\sim}{\theta})$ of the outcomes of observations of the system. It is natural to consider analogous theories based instead on quantum-mechanical density operators ρ_0, ρ_1, $\rho(\underset{\sim}{\theta})$ of the system, a generalization that leads to quantum decision and estimation theory[18-20].

The system under observation might, for instance, be a lossless cavity that functions as an ideal receiver of electromagnetic radiation. The cavity is initially empty. In one wall is an aperture that faces the source of the signal, and during an interval (0, T) when the signal, if present in the external field, is expected to arrive, the aperture is open. At time T the aperture is closed, and thereafter the cavity contains background radiation and, possibly, a field due to the signal. The density operator of the field will be ρ_0 when only background radiation is present (hypothesis H_0) and ρ_1 when a signal of the specified type has arrived (hypothesis H_1). Detection involves a choice between these hypotheses. In particular, one would like to know the weakest signal that can be detected with a certain probability Q_d as a function of the false-alarm probability Q_0 and the nature of the background radiation.

If, on the other hand, the signal field is known to be present, it may be necessary to measure certain of its parameters, such as its amplitude or carrier frequency. These can be regarded as parameters of the density operator $\rho(\underset{\sim}{\theta}) = \rho(\theta_1, \ldots, \theta_m)$ of the net field in the cavity. One would like to know the minimum mean-square errors with which the field parameters can be estimated, as functions of the characteristics of the signal and background fields.

Crucial in quantum decision and estimation theory is the question of which dynamical variables of the system shall be measured. In the classical theory it is possible in principle to measure all the variables and to conceive of their having the joint probability density functions $p_0(\underset{\sim}{x})$, $p_1(\underset{\sim}{x})$, and $p(\underset{\sim}{x}; \underset{\sim}{\theta})$ required for setting up the optimum procedures. Quantum-mechanically only observables -- dynamical variables represented by Hermitian operators -- can be measured, and since they are to be measured simultaneously on the same system, their operators must commute. Different sets of commuting observables may yield different costs in a Bayes decision or estimation strategy, and the problem remains of finding the set that entails the lowest cost of all.

If there exists a representation in which all the density operators involved are simultaneously diagonal, they all commute, and by working in this representation, the decision or estimation problem can be reduced to one that can be handled by the classical theory. Quantum-mechanical decision and estimation theory is presently formulated entirely within the framework of the conventional interpretation of quantum mechanics, and questions of the simultaneous measurability of variables whose operators do not commute have not been treated.

II. Binary Decisions

The Detection Operator

A choice is to be made between two hypotheses about a system, (H_0) that its density operator is ρ_0 and (H_1) that its density operator is ρ_1. The prior probability of H_0 is ζ and of H_1 $(1-\zeta)$, and the cost attendant upon choosing H_i when H_j is true ($i, j = 0, 1$) is C_{ij}. Suppose that some set of commuting observables $X_1, X_2, \ldots,$ has been measured, with outcomes $x_1, x_2, \ldots.$ The decision will be based on the value of some function $f(x_1, x_2, \ldots)$ of the outcomes. Equivalently, it could be based on the outcome of a measurement of the operator $f(X_1, X_2, \ldots)$. What operator should this be?

All that we really require is that the outcome be one of two numbers, 0 and 1, and we choose H_0 if it is 0, H_1 if it is 1. The operator $f(X_1, X_2, \ldots)$ should therefore be one whose only eigenvalues are 0 and 1, and such an operator is a projection operator. We denote it by Π and call it the <u>detection operator.</u>

Which of all the projection operators Π for the system is best? To determine it we put down an expression for the average cost and minimize it over the set of all Π's. The average cost depends on the probabilities Q_0 and Q_1 of errors of the first and second kinds. The former is the probability under hypothesis H_0 that H_1 is chosen, that is, that measurement of Π yields the value 1,

$$Q_0 = \Pr\{\Pi \to 1 \mid H_0\} = \underset{\sim}{E}(\Pi \mid H_0) = \mathrm{Tr}\,\rho_0 \Pi. \tag{2.1}$$

Similarly

$$Q_1 = \mathrm{Tr}[\rho_1(\underset{\sim}{1} - \Pi)] = 1 - \mathrm{Tr}\,\rho_1 \Pi, \tag{2.2}$$

and the average cost is

$$\overline{C} = \zeta[C_{00}(1-Q_0) + C_{10}Q_0] + (1-\zeta)[C_{01}Q_1 + C_{11}(1-Q_1)] =$$

$$\zeta C_{00} + (1-\zeta)C_{01} - (1-\zeta)(C_{01} - C_{11})\,\mathrm{Tr}(\rho_1 - \lambda\rho_0)\Pi, \tag{2.3}$$

where

$$\lambda = \frac{\zeta(C_{10} - C_{00})}{(1-\zeta)(C_{01} - C_{11})}. \tag{2.4}$$

Since $C_{01} > C_{11}$, $\overline{C}$ will be minimum if $\mathrm{Tr}(\rho_1 - \lambda\rho_0)\,\Pi$ is maximum.

Choose a representation in terms of the eigenstates $|\eta_k\rangle$ of the operator $\rho_1 - \lambda\rho_0$, whose eigenvalues we suppose discrete,

$$(\rho_1 - \lambda\rho_0)\,|\eta_k\rangle = \eta_k|\eta_k\rangle. \tag{2.5}$$

It is then necessary to maximize

$$\mathrm{Tr}(\rho_1 - \lambda\rho_0)\Pi = \sum_k \eta_k \langle\eta_k|\Pi|\eta_k\rangle, \tag{2.6}$$

and this will be accomplished if

$$\langle\eta_k|\Pi|\eta_k\rangle = 1, \quad \eta_k \geq 0,$$

$$\langle\eta_k|\Pi|\eta_k\rangle = 0, \quad \eta_k < 0.$$

Hence the best projection operator to measure in order to choose between H_0 and H_1 is

$$\Pi = \sum_{\substack{k: \\ \eta_k \geq 0}} |\eta_k\rangle\langle\eta_k| . \tag{2.7}$$

Equivalently, $\rho_1 - \lambda\rho_0$ is measured, and H_1 is chosen if the outcome is positive[19, 21].

The probabilities of error are

$$Q_0 = \sum_{\substack{k: \\ \eta_k \geq 0}} \langle\eta_k|\rho_0|\eta_k\rangle ,$$

$$Q_1 = 1 - \sum_{\substack{k: \\ \eta_k \geq 0}} \langle\eta_k|\rho_1|\eta_k\rangle , \tag{2.8}$$

and the minimum average cost is

$$\overline{C}_{min} = \zeta\, C_{00} + (1 - \zeta)C_{01} - (1-\zeta)(C_{01} - C_{11}) \sum_{\substack{k: \\ \eta_k > 0}} \eta_k . \tag{2.9}$$

Denote the eigenvalues of the density operators ρ_0 and ρ_1 by P_{0k} and P_{1k}, respectively, numbering them in descending order. If the operators are completely continuous, these eigenvalues form discrete spectra. A theorem in analysis then assures us that the eigenvalues η_m of $\rho_1 - \lambda\rho_0$ are also discrete, that its k-th positive eigenvalue is less than or equal to P_{1k}, and that its k-th negative eigenvalue is greater than or equal to $-\lambda P_{0k}$. Here the positive eigenvalues are counted by beginning with the largest, the negative ones by beginning with the most negative[22].

If the density operators ρ_0 and ρ_1 commute, the eigenvalues of $\rho_1 - \lambda\rho_o$ are $P_{1k} - \lambda P_{0k}$, and these are positive when

$$P_{1k}/P_{0k} > \lambda .$$

The best procedure is then to measure either ρ_0, ρ_1, or a suitable operator commuting with both. When the system is found in the kth common eigenstate, choose H_1 if $P_{1k}/P_{0k} \geq \lambda$, H_0 if $P_{1k}/P_{0k} < \lambda$. This is just the likelihood-ratio test of classical decision theory.

Let the system be a simple harmonic oscillator, such as a single mode of the field in our ideal receiver, and assume it to be in thermal equilibrium with an average number of photons equal to N_0 (hypothesis H_0) or to N_1 (hypothesis H_1). The density operators are[23]

$$\rho_k = \sum_{m=0}^{\infty} |m\rangle P_{km} \langle m| , \qquad (2.10)$$

$$P_{km} = (1 - v_k) v_k^m , \quad v_k = N_k/(N_k + 1), \ k = 0, 1,$$

in terms of the eigenstates $|m\rangle$ of the number operator n. It then suffices to measure n itself and to choose hypothesis H_1 when

$$\frac{(1 - v_1)}{(1 - v_0)} (v_1/v_0)^m > \lambda$$

where m is the outcome of the measurement.

The Choice between Pure States

There are few pairs of noncommuting operators ρ_0 and ρ_1 for which the eigenvalue problem in Eq. (2.5) has been solved. One general case of interest is that in which the system is in a pure state under each hypothesis[24, 25],

$$\rho_0 = |\psi_0\rangle\langle\psi_0|, \quad \rho_1 = |\psi_1\rangle\langle\psi_1|. \tag{2.11}$$

There are then just two states $|\eta_0\rangle$, $|\eta_1\rangle$ satisfying Eq.(2.5) with non-zero eigenvalues, and they are linear combinations of $|\psi_0\rangle$ and $|\psi_1\rangle$,

$$|\eta_k\rangle = z_{k0}|\psi_0\rangle + z_{k1}|\psi_1\rangle, \quad k = 0, 1. \tag{2.12}$$

By substituting Eqs. (2.11) and (2.12) into Eq. (2.5), a set of linear homogeneous equations for z_{k0}, z_{k1} is obtained. A solution exists only when the determinant of their coefficients vanishes, which yields a quadratic equation for the eigenvalues η_0 and η_1. The solution is

$$\begin{aligned} \eta_k &= \tfrac{1}{2}(1-\lambda) - (-1)^k R, \quad k = 0, 1, \\ R &= \left\{\left[\tfrac{1}{2}(1-\lambda)\right]^2 + \lambda q\right\}^{\frac{1}{2}}, \\ q &= 1 - |\langle\psi_1|\psi_0\rangle|^2. \end{aligned} \tag{2.13}$$

The detection operator to be measured is $\Pi = |\eta_1\rangle\langle\eta_1|$, the false-alarm and detection probabilities are

$$\begin{aligned} Q_0 &= |\langle\eta_1|\psi_0\rangle|^2 = (\eta_1 - q)/2R, \\ Q_d &= 1 - Q_1 = |\langle\eta_1|\psi_1\rangle|^2 = (\eta_1 + \lambda q)/2R, \end{aligned} \tag{2.14}$$

and the minimum average cost $\overline{C}$ can be calculated by eq. (2.9), in which the sum now has a single term η_1.

In the choice between two coherent states $|\mu_0\rangle$ and $|\mu_1\rangle$ of a harmonic oscillator such as the field in a single mode of our ideal receiver, now devoid of background radiation, the parameter q entering Eq. (2.13) is[26]

$$q = 1 - |\langle\mu_1|\mu_0\rangle|^2 = 1 - \exp(-|\mu_1 - \mu_0|)^2.$$

If, for instance, $\mu_0 = 0$, the choice is between the presence and the absence of a coherent signal in the mode, and the probabilities of error depend, through $q = 1 - \exp(-N_s)$, only on the mean number $N_s = |\mu_1|^2$ of signal photons, as in Eq. (2.14).

The Coherent Signal in Thermal Radiation

Let hypothesis H_1 assert the presence, H_0 the absence, of a coherent signal of complex amplitude μ in a single mode of a cavity in thermal equilibrium at absolute temperature $\mathcal{T}$. If the thermal radiation were gone, the oscillator representing the mode would be in a coherent state $|\mu\rangle$. The density operators are, in the P-representation[26],

$$\begin{aligned} \rho_0 &= (\pi N)^{-1}\int \exp(-|\alpha|^2/N)\,|\alpha\rangle\langle\alpha|\,d^2\alpha, \\ \rho_1 &= (\pi N)^{-1}\int \exp(-|\alpha-\mu|^2/N)\,|\alpha\rangle\langle\alpha|\,d^2\alpha, \\ N &= \left[\exp(\hbar\Omega/|K\mathcal{T})-1\right]^{-1}, \end{aligned} \tag{2.15}$$

where $\hbar$ = Planck's constant $h/2\pi$, Ω = the angular frequency of the mode, and K = Boltzmann's constant. The diagonalization of $\rho_1 - \lambda\rho_0$, as in Eq. (2.5), with these density operators remains an outstanding unsolved problem of quantum detection theory. By taking μ as real, which voids no generality, and using the co-ordinate (q-) representation, Eq. (2.5) can be expressed as a homogeneous integral equation, whose kernel is a linear combination of Gaussian functions[27]. Evaluation of the probability of detection would permit specifying the minimum detectable coherent signal of known phase in the presence of thermal background radiation.

When, as is most reasonable at optical frequencies, the absolute phase of the complex signal amplitude is unknown and is assigned a uniform prior distribution over $(0, 2\pi)$, both ρ_0 and ρ_1 are diagonal in the number representation, and the best detector simply measures the energy in the mode[18].

If a coherent signal of random phase is present in a number of modes of a receiver cavity in thermal equilibrium, a linear transformation of the mode amplitudes permits approximate reduction of the problem to the detection of a signal in a single harmonic oscillator[19]. For this it is required that the signal occupy a frequency band so narrow that the average number of thermal photons is the same for all the modes that it excites. In effect, the optimum processing of the field creates a single mode "matched" to the signal, and it is the energy or the excitation level of this composite mode that is to be measured.

The receiver decides that no signal is present whenever the number of photons counted in the matched mode is less than an integer M. The false-alarm probability is then

$$Q_0 = \left[N/(N+1) \right]^M ,$$

and the detection probability is

$$Q_d = 1 - (N+1)^{-1} \exp(-N_s/(N+1)) \times \sum_{m=0}^{M-1} \left[N/(N+1)\right]^m L_m(-N_s/[N(N+1)]), \qquad (2.16)$$

where $L_m(x)$ is the mth Laguerre polynomial[19, 28].

If this receiver is designed to meet the Neyman-Pearson criterion, randomization will in general be necessary in order to attain the pre-assigned false-alarm probability. There will then be a certain photon count M' for which hypothesis H_1 (signal present) is chosen with probability f, H_0 with probability 1-f. For counts less than M', H_0 is always chosen, for counts greater than M', H_1. The required value of f is easily calculated. Graphs of detection probability versus signal strength for such a receiver have been published[29].

III. Threshold Detection

The Classical Threshold Receiver

It would be useful if a receiver set to incur a fixed false-alarm probability attained maximum detection probability for all expected amplitudes of the signal. This is seldom the case with a receiver based on the classical likelihood-ratio test. Only in particularly simple instances, as when the signal is completely known except for amplitude and phase and is received in Gaussian noise, is the likelihood-ratio test uniformly most powerful with respect to signal amplitude. It is usually necessary to set it up for a "standard" signal of specific amplitude and to accept less than maximum probability of detecting signals of other amplitudes. Furthermore, the likelihood ratio is often difficult to generate from a receiver input.

In a compromise that is often expedient, the likelihood ratio is replaced by the so-called threshold statistic

$$U = \frac{\partial}{\partial A} \ln \Lambda(x_1, \ldots, x_n; A) \Big|_{A=0}, \tag{3.1}$$

where

$$\Lambda(x_1, \ldots, x_n; A) = \frac{p_1(x_1, \ldots, x_n; A)}{p_0(x_1, \ldots, x_n)} \tag{3.2}$$

is the likelihood ratio, with $p_1(x_1, \ldots, x_n; A)$ the p.d.f. of the data when a signal of strength A is present; $p_0(x_1, \ldots, x_n) = p_1(x_1, \ldots, x_n; 0)$. The threshold statistic is the logarithmic derivative, with respect to A, of the likelihood ratio for detecting a signal of strength A, evaluated in the limit of vanishing amplitude. It is compared with a decision level U_0, and hypothesis H_1, "signal present", is selected when $U > U_0$. The measure A of signal strength is so chosen that the derivative in Eq. (3.1) does not vanish; it is usually proportional to the energy of the signal[30].

This threshold statistic U is most nearly optimum when the decision is based on data collected in a large number M of independent trials. Compared with the decision level U_0 then is the sum $U_1 + U_2 + \ldots + U_M$ of the threshold statistics calculated from the data obtained in each trial. The sum has nearly a Gaussian distribution, by virtue of the central limit theorem, and the false-alarm and detection probabilities are approximately

$$Q_0 = \operatorname{erfc} x = (2\pi)^{-\frac{1}{2}} \int_x^\infty \exp(-t^2/2)\, dt, \qquad (3.3)$$
$$Q_d = \operatorname{erfc}\,(x - M^{\frac{1}{2}} D),$$

where D is an equivalent signal-to-noise ratio defined by

$$D^2 = \left[\underset{\sim}{E}(U \mid H_1) - \underset{\sim}{E}(U \mid H_0) \right]^2 / \operatorname{Var}_0 U \qquad (3.4)$$

with Var_0 U the variance of the statistic U in the absence of the signal. In Eq. (3.3), x is related to the decision level U_0 on the sum of the threshold statistics.

The false-alarm and detection probabilities will be given approximately as in Eq. (3.3) for any statistic $U(x_1, \ldots, x_n)$ when the decision is based on the sum of such statistics for a large number M of independent trials. For a fixed pair of probabilities Q_0 and Q_d and for $M >> 1$, that detector is best for which the equivalent signal-to-noise ratio D is largest, for such a detector will require the least number of M of independent trials. The threshold detector as defined in Eqs. (3.1) and (3.2) is best in this sense[31].

The Quantum Threshold Receiver

The quantum counterpart of the likelihood-ratio receiver is one in which the optimum detection operator Π is measured. It has been found uniformly most powerful with respect to signal amplitude only for detecting a known signal of random phase in the presence of thermal noise, a detection problem in which, as we have seen, the density operators commute and the classical likelihood-ratio test is optimum.

Furthermore, the mathematical problem of determining the optimum projection operator Π presents great difficulty in most cases of practical interest. For these reasons, a quantum-mechanical counterpart to the classical threshold statistic is of interest.

The quantum threshold statistic Π_θ is defined as [18]

$$\Pi_\theta = \partial \Pi_a / \partial A \big|_{A=0} , \tag{3.5}$$

where Π_a is that operator for which the equivalent signal-to-noise ratio D given by

$$D^2 = \frac{\left[\mathrm{Tr}\rho_1(A)\,\Pi_a - \mathrm{Tr}\,\rho_0\Pi_a\right]^2}{\mathrm{Tr}\,\rho_0\Pi_a^2 - (\mathrm{Tr}\,\rho_0\Pi_a)^2} \tag{3.6}$$

is maximum. This equivalent signal-to-noise ratio is the quantum-mechanical form of the one defined in Eq. (3.4); $\rho_1(A)$ is the density operator of the observed system when a signal of strength A is present, and $\rho_0 = \rho_1(0)$. There is no loss of generality if Π_a is so defined that

$$\mathrm{Tr}\,\rho_0\Pi_a = 0, \tag{3.7}$$

since an arbitrary multiple of the identity operator $\underset{\sim}{1}$ can be subtracted from Π_a without changing D^2.

We define the Hermitian operator Θ (A) as the solution of the equation

$$\rho_1(A) - \rho_0 = \tfrac{1}{2}(\rho_0\Theta + \Theta\rho_0), \tag{3.8}$$

and we show that $\Pi_a = \Theta$. First of all,

$$\mathrm{Tr}(\rho_1 - \rho_0) = 0 = \tfrac{1}{2}\,\mathrm{Tr}(\rho_0\Theta + \Theta\rho_0) = \mathrm{Tr}\,\rho_0\Theta,$$

so that Eq. (3.7) is satisfied. We now show that $\Pi_a = \Theta$ maximizes

$$D^2 = \left[\mathrm{Tr}(\rho_1 - \rho_0)\,\Pi_a\right]^2 / \mathrm{Tr}(\rho_0 \Pi_a^2). \tag{3.9}$$

Substituting from Eq. (3.8), we find

$$\left[\mathrm{Tr}(\rho_1 - \rho_0)\,\Pi_a\right]^2 = \left[\tfrac{1}{2}(\mathrm{Tr}\rho_0\Theta\Pi_a + \mathrm{Tr}\Theta\rho_0\Pi_a)\right]^2 =$$

$$\left[\mathrm{R}\ell\ \mathrm{Tr}\rho_0\Theta\Pi_a\right]^2 \le \left|\mathrm{Tr}\rho_0\Theta\Pi_a\right|^2 =$$

$$\left|\mathrm{Tr}\rho_0^{\frac{1}{2}}\Theta\Pi_a\rho_0^{\frac{1}{2}}\right|^2 \le \mathrm{Tr}(\rho_0^{\frac{1}{2}}\Theta^2\rho_0^{\frac{1}{2}})\ \mathrm{Tr}(\Pi_a\rho_0\Pi_a)$$

$$= \mathrm{Tr}(\rho_0\Theta^2)\ \mathrm{Tr}(\rho_0\Pi_a^2)$$

by the Schwarz inequality for traces. Hence

$$D^2 \le \mathrm{Tr}(\rho_0\Theta^2)$$

with equality when $\Pi_a = \Theta$.

The threshold operator is thus[32]

$$\Pi_\theta = \partial\Theta(A)/\partial A\big|_{A=0}.$$

As the solution of the operator equation

$$\partial\rho_1(A)/\partial A\big|_{A=0} = \tfrac{1}{2}(\rho_0\Pi_\theta + \Pi_\theta\rho_0), \tag{3.10}$$

it can be regarded as the symmetrized logarithmic derivative (s.l.d.) of $\rho_1(A)$, evaluated at $A = 0$.

In the quantum threshold receiver the operator Π_θ is measured and the outcome compared with a decision level π_θ set to yield a preassigned false-alarm probability. The operator Π_θ is not a projection operator; the equivalent projection operator is

$$\int_{\pi_\theta}^{\infty} |\theta\rangle\langle\theta|\, d\theta$$

with $|\theta\rangle$ the eigenstate of Π_θ with eigenvalue θ, assumed here part of a continuous spectrum.

Threshold Detection of a Coherent Signal

In the cavity that furnishes our model of a quantum receiver the electric field at time t at point $\underset{\sim}{r}$ is represented by a quantum-mechanical operator $\underset{\sim}{\varepsilon}(\underset{\sim}{r}, t)$, which is conveniently decomposed into its positive- and negative-frequency parts,

$$\underset{\sim}{\varepsilon}(\underset{\sim}{r}, t) = \underset{\sim}{\varepsilon}^{(+)}(\underset{\sim}{r}, t) + \underset{\sim}{\varepsilon}^{(-)}(\underset{\sim}{r}, t),$$

$$\underset{\sim}{\varepsilon}^{(-)}(\underset{\sim}{r}, t) = [\underset{\sim}{\varepsilon}^{(+)}(\underset{\sim}{r}, t)]^{+},$$

the one being the Hermitian conjugate of the other. In terms of the mode eigenfunctions $\underset{\sim}{u}_{\underset{\sim}{m}}(\underset{\sim}{r})$, which are solutions of the Helmholtz equation with suitable boundary conditions at the walls of the cavity, the positive-frequency part of the electric-field operator is written as [33]

$$\underset{\sim}{\varepsilon}^{(+)}(\underset{\sim}{r}, t) = i \sum_{\underset{\sim}{m}} (\hbar\omega_{\underset{\sim}{m}}/2)^{\frac{1}{2}} \, a_{\underset{\sim}{m}} \underset{\sim}{u}_{\underset{\sim}{m}}(\underset{\sim}{r}) \exp(-i\omega_{\underset{\sim}{m}} t), \qquad (3.11)$$

where $\omega_{\underset{\sim}{m}}$ is the angular frequency of mode $\underset{\sim}{m}$. The mode index $\underset{\sim}{m}$ accounts for both the spatial configuration and the polarization of the mode.

The operator $a_{\underset{\sim}{m}}$ and its Hermitian conjugate $a_{\underset{\sim}{m}}^{+}$ are the annihilation and creation operators for photons in mode $\underset{\sim}{m}$ and obey the usual commutation rules,

$$\begin{aligned} a_{\underset{\sim}{m}} a_{\underset{\sim}{n}}^{+} - a_{\underset{\sim}{n}}^{+} a_{\underset{\sim}{m}} &= [a_{\underset{\sim}{m}}, a_{\underset{\sim}{n}}^{+}] = \delta_{\underset{\sim}{mn}}, \\ [a_{\underset{\sim}{m}}, a_{\underset{\sim}{n}}] &= [a_{\underset{\sim}{m}}^{+}, a_{\underset{\sim}{n}}^{+}] = 0. \end{aligned} \qquad (3.12)$$

The number operator for mode $\underset{\sim}{m}$ is $n_{\underset{\sim}{m}} = a_{\underset{\sim}{m}}^{+} a_{\underset{\sim}{m}}$.

Suppose that under hypothesis H_0 the cavity is filled with random Gaussian radiation characterized by the mode correlation matrix $\underset{\sim}{\varphi}$, whose elements are

$$\varphi_{\underset{\sim}{k}\underset{\sim}{m}} = \mathrm{Tr}(\rho_0 a_{\underset{\sim}{m}}^{+} a_{\underset{\sim}{k}}) \tag{3.13}$$

The density operator ρ_0 for L modes of the field is then, in the P-representation[26],

$$\rho_0 = \pi^{-L} |\det \underset{\sim}{\varphi}|^{-1} \int \cdots \int \exp(-\underset{\sim}{\alpha}^{+} \underset{\sim}{\varphi}^{-1} \underset{\sim}{\alpha}) \times |\underset{\sim}{\alpha}\rangle\langle\underset{\sim}{\alpha}| d^{2L}\underset{\sim}{\alpha}, \tag{3.14}$$

where $\underset{\sim}{\alpha}$ is a column vector of complex mode variables

$$\alpha_{\underset{\sim}{m}} = \alpha_{\underset{\sim}{m}x} + i\alpha_{\underset{\sim}{m}y},$$

$\underset{\sim}{\alpha}^{+}$ is the Hermitian conjugate row vector, $\underset{\sim}{\alpha}^{+} = \{\ldots \alpha_{\underset{\sim}{m}}^{*} \ldots\}$, and $d^{2L}\underset{\sim}{\alpha} = \prod_{\underset{\sim}{m}} d\alpha_{\underset{\sim}{m}x}\, d\alpha_{\underset{\sim}{m}y}$ is the element of integration in the space of the $\alpha_{\underset{\sim}{m}}$'s. Here

$$|\underset{\sim}{\alpha}\rangle = \prod_{\underset{\sim}{m}} |\alpha_{\underset{\sim}{m}}\rangle$$

is the Glauber coherent state for a field with complex amplitude $\alpha_{\underset{\sim}{m}}$ in mode $\underset{\sim}{m}$. In thermal equilibrium at absolute temperature $\mathcal{T}$,

$$\varphi_{\underset{\sim}{k}\underset{\sim}{m}} = N_{\underset{\sim}{k}}\, \delta_{\underset{\sim}{k}\underset{\sim}{m}},$$

$$N_{\underset{\sim}{k}} = [\exp(\hbar\omega_{\underset{\sim}{k}}/K\mathcal{T}) - 1]^{-1}. \tag{3.15}$$

Were a coherent signal of amplitude A and known phase present in the absence of the random radiation, the field would be in a coherent state $|A\underset{\sim}{\mu}\rangle$, in which the complex amplitude in mode $\underset{\sim}{m}$ is $A\mu_{\underset{\sim}{m}}$. If this coherent signal is superimposed on the random radiation described by ρ_0 of Eq. (3.14), the density operator for the field is

$$\rho_1(A) = \pi^{-L} |\det \underset{\sim}{\varphi}|^{-1} \int \cdots \int \exp [-(\underset{\sim}{\alpha}^+ - A\underset{\sim}{\mu}^+)\underset{\sim}{\varphi}^{-1}(\underset{\sim}{\alpha} - A\underset{\sim}{\mu})] \times |\underset{\sim}{\alpha}\rangle\langle\underset{\sim}{\alpha}| d^{2L}\underset{\sim}{\alpha}, \tag{3.16}$$

which can also be written as [34]

$$\rho_1(A) = V^+(A)\rho_0 V(A), \tag{3.17}$$

$$V(A) = \exp [\tfrac{1}{2}A\Pi_\theta - A^2 \underset{\sim}{\mu}^+ (\underset{\sim}{I} + 2\underset{\sim}{\varphi})^{-1} \underset{\sim}{\mu}], \tag{3.18}$$

where

$$\Pi_\theta = 2[\underset{\sim}{\mu}^+ (\underset{\sim}{I} + 2\underset{\sim}{\varphi})^{-1} \underset{\sim}{a} + \underset{\sim}{a}^+ (\underset{\sim}{I} + 2\underset{\sim}{\varphi})^{-1} \underset{\sim}{\mu}] \tag{3.19}$$

with $\underset{\sim}{a}$ the column vector of annihilation operators $a_{\underset{\sim}{m}}$ and $\underset{\sim}{a}^+ = (\ldots, a_{\underset{\sim}{m}}^+, \ldots)$ the row vector of the creation operators for the modes. $\underset{\sim}{I}$ is the identity matrix.

The threshold operator for deciding whether the coherent field with amplitudes $A\mu_{\underset{\sim}{m}}$ is present is the operator Π_θ given by Eq. (3.18), as can be verified by differentiating $\rho_1(A)$ with respect to A, setting A = 0, and comparing with Eq. (3.10). The outcomes of measurements of the operator Π_θ have a Gaussian distribution under each hypothesis, and the false-alarm and detection probabilities are given exactly by Eq. (3.3), with x related to the decision level π_θ with which the outcomes are compared. The equivalent signal-to-noise ratio D is given by

$$D^2 = \mathrm{Tr}(\rho_0 \Pi_\theta^2) = 4A^2 \underset{\sim}{\mu}^+ (\underset{\sim}{I} + 2\underset{\sim}{\varphi})^{-1} \underset{\sim}{\mu}. \tag{3.20}$$

For detection in thermal radiation this signal-to-noise ratio reduces to

$$D^2 = 4N_s/(2N + 1), \tag{3.21}$$

where $N_s = E_s/\hbar\Omega$ is the average number of photons in the field of the coherent signal and N is given by Eq. (2.15). For Eq. (3.21) to hold it is necessary that the average numbers $N_{\underset{\sim}{m}}$ of thermal photons in all modes excited by the signal be nearly equal to N, as will be the case when, as usually, the signal occupies only a narrow band of frequencies about Ω.

In the classical limit the term $2\underset{\sim}{\varphi}$ dominates in the factor $(\underset{\sim}{I} + 2\underset{\sim}{\varphi})^{-1}$ in Eq. (3.19). and the threshold operator becomes, except for an additive constant, proportional to the logarithm of the classical likelihood ratio for choosing between hypotheses H_0 and H_1. The false-alarm and detection probabilities for the classically optimum detector are given by Eq. (3.3) with $D^2 = 2A^2 \underset{\sim}{\mu}^+ \underset{\sim}{\varphi}^{-1} \underset{\sim}{\mu}$, or for thermal equilibrium, $D^2 = 2E_s/K\mathcal{T}$. Thus in this case the quantum threshold operator becomes equivalent in the classical limit to the optimum likelihood-ratio statistic[35].

Detection of Gaussian Radiation

If the signal field itself has the character of random Gaussian radiation, the density operator ρ_1 has the same form as ρ_0 in Eq. (3.14). We suppose that under hypothesis H_0 the mode correlation matrix, defined by Eq. (3.13), is $\underset{\sim}{\varphi}_0$; under hypothesis H_1 it is

$$\underset{\sim}{\varphi}_1 = \underset{\sim}{\varphi}_0 + A\underset{\sim}{\varphi}_s, \tag{3.22}$$

where $A\underset{\sim}{\varphi}_s$ is the mode correlation matrix of the random signal components of the field. This is the quantum-mechanical counterpart of what is sometimes called the "noise-in-noise" detection problem, and it corresponds to the detection of light from an incoherent source.

The optimum detection operator Π for deciding between hypotheses H_0 and H_1 remains undiscovered. The threshold operator, however, can be calculated[36]. It is an Hermitian quadratic form in the annihilation and creation operators of the modes,

$$\Pi_\theta = \sum_{k,m} a_k^+ q_{km} a_m + b\underset{\sim}{1} = \underset{\sim}{a}^+ \underset{\sim}{Q}\underset{\sim}{a} + b\underset{\sim}{1},$$

$$b = - \mathrm{Tr}\left[(\underset{\sim}{I} + \underset{\sim}{\varphi}_0)^{-1}\underset{\sim}{\varphi}_s\right] = - \mathrm{Tr}(\underset{\sim}{Q}\underset{\sim}{\varphi}_0), \tag{3.23}$$

where the matrix $\underset{\sim}{Q}$ is the solution of the equation

$$2\underset{\sim}{\varphi}_s = \underset{\sim}{\varphi}_0\underset{\sim}{Q}(\underset{\sim}{I} + \underset{\sim}{\varphi}_0) + (\underset{\sim}{I} + \underset{\sim}{\varphi}_0)\underset{\sim}{Q}\underset{\sim}{\varphi}_0. \tag{3.24}$$

The constant b serves to make $\mathrm{Tr}\,(\rho_0 \Pi_\theta)$ vanish.

The p.d.f.'s, under the two hypotheses, of the outcomes of measurements of Π_θ are difficult to calculate, and only approximate forms of the false-alarm and detection probabilities are accessible in the general case. The moment-generating functions of the observable $Q' = \underset{\sim}{a}^+\underset{\sim}{Q}\underset{\sim}{a}$ are given by[36]

$$h_i(z) = \mathrm{Tr}\,\rho_i e^{zQ'} = \exp\left[-\mathrm{Tr}\,\ln\,(\underset{\sim}{I} - \underset{\sim}{\varphi}_i\underset{\sim}{P})\right], \tag{3.25}$$

$$\underset{\sim}{P} = \exp\,(z\underset{\sim}{Q}) - \underset{\sim}{I}, \quad i = 0,\ 1.$$

The p.d.f of the outcome of a measurement of Q' is the inverse Laplace transform of $h_i(-s)$, and approximation methods, such as the method of steepest descents, are available for calculating the false-alarm and detection probabilities[37].

Reception at an Aperture

An unsatisfactory aspect of quantum detection theory is its formulation in terms of simultaneous measurements of the electromagnetic field in a closed volume. An optical instrument, such as a telescope, is more appropriately considered as processing the field at its aperture throughout a finite observation interval. An advantage of the threshold operator for detecting a Gaussian random field is that it can be translated into a form involving only the field operators at the aperture of the receiver, and it can thus be applied to the detection of light from an incoherent source[38]. This translation is possible because the classical mode amplitudes for the cavity receiver after its aperture is closed are linearly related to the field at the aperture itself during the observation interval (0, T).

In order for the threshold operator for detecting incoherent light in the presence of thermal background radiation to take a simple form when expressed in terms of the aperture fields, it is necessary that the duration T of the observation interval be much longer than the reciprocal of the bandwidth W of the light to be detected ($WT >> 1$), and that the diameter of the aperture $\mathcal{A}$ be much greater than the correlation length $\hbar c/K\mathcal{T}$ of the thermal radiation. Both these conditions are normally met. The threshold operator is then proportional to

$$\int_{\mathcal{A}} \int_{\mathcal{A}} d^2 \underset{\sim}{r}_1 \, d^2 \underset{\sim}{r}_2 \int_0^T \int_0^T dt_1 dt_2 \qquad (3.26)$$
$$\times \, \psi^{(-)}(\underset{\sim}{r}_1 t_1) \, G(\underset{\sim}{r}_1, t_1; \underset{\sim}{r}_2, t_2) \, \psi^{(+)}(\underset{\sim}{r}_2, t_2),$$

in which for simplicity a scalar field

$$\psi(\underset{\sim}{r}, t) = \psi^{(+)}(\underset{\sim}{r}, t) + \psi^{(-)}(\underset{\sim}{r}, t)$$

has been assumed. Here

$$G(\underset{\sim}{r}_1, \underset{\sim}{t}_1; \underset{\sim}{r}_2, \underset{\sim}{t}_2) = \mathrm{Tr}\left[\rho_s \psi^{(-)}(\underset{\sim}{r}_2, t_2)\, \psi^{(+)}(\underset{\sim}{r}_1, t_1)\right]$$

is the mutual coherence function of the signal field, where ρ_s is its density operator in the absence of the thermal background. A similar receiver has been derived by Kuriksha on the basis of the classical likelihood ratio[39, 40].

The moment-generating function of the threshold statistic can also be expressed in terms of the mutual coherence function of the signal field at the aperture by similarly translating the form given in Eq. (3.25), and from this the false-alarm and detection probabilities can be approximated. Details are presented elsewhere[38].

IV. Choices Among Many Hypotheses.

The choice among M hypotheses, of which the kth asserts, "The system has the density operator ρ_k", $k = 1, 2, \ldots, M$, can be based on the outcome of a measurement of M commuting projection operators $\Pi_1, \Pi_2, \ldots, \Pi_m$ forming a resolution of the identity operator $\underset{\sim}{1}$,

$$\Pi_1 + \Pi_2 + \ldots + \Pi_m = \underset{\sim}{1}. \qquad (4.1)$$

Quantum logic was formulated in terms of projection operators by von Neumann[41]. Our problem is to pick such a set of operators Π_k that the decision among the M hypotheses can be made with minimum average cost. It will arise, for instance, in designing and evaluating the best receiver for a communication system in which messages are coded into an alphabet of more than two symbols, a different signal being transmitted for each.

Let ζ_k be the prior probability of hypothesis H_k and C_{ij} be the cost incurred upon choosing H_i when H_j is true. The average cost per decision is

$$\bar{C} = \sum_{i=1}^{M} \sum_{j=1}^{M} \zeta_j C_{ij} \, \mathrm{Tr}\,(\rho_j \Pi_i), \tag{4.2}$$

which is to be minimized by a set of commuting projection operators Π_k that satisfy Eq. (4.1). If in particular $C_{ii} = 0$, $C_{ij} = 1$, $i \neq j$, $\bar{C}$ equals the average probability of error. This problem of minimizing $\bar{C}$ remains unsolved for $M > 2$, except when the M operators ρ_j commute, whereupon it reduces to a standard problem in classical decision theory.

If under each hypothesis the system is in a pure state, $\rho_k = |\psi_k\rangle\langle\psi_k|$, the projection operators will have the form

$$\Pi_j = |\eta_j\rangle\langle\eta_j| \ ,$$

where the $|\eta_j\rangle$ are linear combinations of the states $|\psi_k\rangle$. To find what linear combination minimizes $\bar{C}$ is also, for $M > 2$, an unsolved problem, although one that appears simpler than the general problem.

V. Estimation Theory

Bayesian estimation theory determines a strategy $\hat{\theta}(\underset{\sim}{x}) = \hat{\theta}(x_1, x_2, \ldots, x_n)$ for estimating a parameter θ of the p. d. f. $p(\underset{\sim}{x};\theta) = p(x_1, x_2, \ldots, x_n; \theta)$

of the data $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ by minimizing the average cost

$$\bar{C} = \int d^n \underset{\sim}{x} \int d\theta \, z(\theta) \quad C(\hat{\theta}(\underset{\sim}{x}), \theta) \, p(\underset{\sim}{x}; \theta), \tag{5.1}$$

where $z(\theta)$ is the prior probability density function of the parameter θ and $C(\hat{\theta}, \theta)$ is the cost associated with a discrepancy between the estimate $\hat{\theta}$ and the true value of the parameter[42].

Quantum-mechanically the parameter θ of a density operator $\rho(\theta)$ is estimated by means of a resolution of the identity[41]

$$\int dE(\theta') = \underset{\sim}{1} \quad , \tag{5.2}$$

where $dE(\theta')$ is a projection operator corresponding to the statement, "The value of the parameter θ lies between θ' and $\theta' + d\theta'$."

Equivalently we can define such an operator

$$\hat{\theta} = \int \theta' \, dE(\theta') \tag{5.3}$$

that the outcome of a measurement of $\hat{\theta}$ yields the value of the estimate of the parameter θ. Corresponding to Eq. (5.1), the average cost associated with the estimate is

$$\bar{C} = \int \int z(\theta) \, C(\theta', \theta) \, \mathrm{Tr}\left[\rho(\theta) \, dE(\theta')\right] d\theta. \tag{5.4}$$

The best estimator of the parameter θ is that resolution of the identity $dE(\theta')$, or the associated operator $\hat{\theta}$, for which the average cost $\bar{C}$ is minimum. How to find it remains an unsolved problem. If estimation is viewed as a choice among a continuum of hypotheses about the system, Eq. (5.4) is the counterpart of Eq. (4.2). If there is a representation in which the density operators $\rho(\theta)$ are simultaneously diagonal, they all commute, and the problem reduces to the classical one of minimizing $\overline{C}$ of Eq. (5.1).

Even in classical statistical estimation the full apparatus of the Bayesian theory is seldom called upon, for prior probability density functions of the parameters are usually unknown. Instead, estimators are sought that have small or zero bias and at the same time incur a small mean-square error over a broad range of true values of the parameters. In quantum-mechanical terms the bias of an estimate $\hat{\theta}$ of a parameter θ of the density operator $\rho(\theta)$ is defined by

$$b(\theta) = \underset{\sim}{E}(\hat{\theta} - \theta) = \mathrm{Tr}\left[\hat{\theta}\,\rho(\theta)\right] - \theta, \tag{5.5}$$

where $\hat{\theta}$ is the operator whose measurement yields the value of the estimate of θ. (Parameters are c-numbers.) The mean-square error is

$$\mathcal{E} = \underset{\sim}{E}(\hat{\theta} - \theta)^2 = \mathrm{Tr}\left[\rho(\theta)\ (\hat{\theta} - \theta\underset{\sim}{1})^2\right]. \tag{5.6}$$

An estimate that has zero bias and attains the minimum value of $\mathcal{E}$ for all values of the parameter θ is said to have uniformly minimum variance.

The Cramér-Rao Inequality

In classical statistics an inequality due to Cramér[43] and Rao[44] sets a lower bound to the mean-square error attainable by any estimator of a parameter θ of a p.d.f. $p(\underset{\sim}{x}; \theta)$,

$$E(\hat{\theta} - \theta)^2 \geq \left[1 + b'(\theta)\right]^2 \left\{E\left[\frac{\partial}{\partial\theta}\ln p(\underset{\sim}{x}; \theta)\right]^2\right\}^{-1}, \tag{5.7}$$

where $b'(\theta) = db(\theta)/d\theta$ and $b(\theta)$ is the bias. For unbiased estimates $b'(\theta) = 0$..

Furthermore, equality is achieved in Eq. (5.7) by an estimator $\hat{\theta}(\underset{\sim}{x})$ satisfying the equation

$$\frac{\partial}{\partial\theta} \ln p(\underset{\sim}{x};\ \theta) = k(\theta)\left[\hat{\theta}(\underset{\sim}{x}) - \theta\right], \tag{5.8}$$

with $k(\theta)$ independent of the data $\underset{\sim}{x}$, provided that such an estimator exists. If it exists, it is unbiased and a sufficient statistic, and it is called an efficient estimator.

In order for a function $\hat{\theta}(\underset{\sim}{x})$ to be a sufficient statistic for estimating θ, it must be possible to factor the density function $p(\underset{\sim}{x};\ \theta)$ into a part depending on the data $\underset{\sim}{x}$ only through $\hat{\theta}(\underset{\sim}{x})$ and a remainder that is independent of the parameter θ,

$$p(\underset{\sim}{x};\ \theta) = g(\hat{\theta}(\underset{\sim}{x});\ \theta)\ r(\underset{\sim}{x}).$$

Such a factorization is seldom possible.

An analogous lower bound exists in quantum estimation theory[45]. Let $\hat{\theta}$ be an operator, the outcome of a measurement of which provides an estimate of a parameter θ of the density operator $\rho(\theta)$. Then the mean-square error is bounded below by

$$\underset{\sim}{E}(\hat{\theta} - \theta)^2 = \mathrm{Tr}\left[\rho(\theta)(\hat{\theta} - \theta\underset{\sim}{1})^2\right] \geq$$

$$\left[1 + b'(\theta)\right]^2 (\mathrm{Tr}\,\rho L^2)^{-1} = \left[1 + b'(\theta)\right]^2 (\mathrm{Tr}\ \frac{\partial\rho}{\partial\theta} L)^{-1}, \tag{5.9}$$

where L is the symmetrized logarithmic derivative (s. l. d.) of $\rho(\theta)$ with respect to θ, defined by

$$\rho L + L\rho = 2\,\partial\rho/\partial\theta. \tag{5.10}$$

The inequality becomes an equality if

$$L = k(\theta)(\hat{\theta} - \theta\underset{\sim}{1}), \tag{5.11}$$

with $k(\theta)$ a numerical function of the true value θ only. This requires the density operator $\rho(\theta)$ to have the form

$$\rho(\theta) = V^{+}(\hat{\theta}; \theta)\, \rho_b V(\hat{\theta}; \theta), \tag{5.12}$$

where ρ_b is independent of the parameter θ, and $V(\hat{\theta}; \theta)$ is an operator satisfying the equation

$$\partial V/\partial\theta = \tfrac{1}{2} VL = \tfrac{1}{2} k(\theta)\, V(\hat{\theta} - \theta \underset{\sim}{1}) \tag{5.13}$$

and depending on the dynamical variables of the system only through the operator $\hat{\theta}$, of which it is a function. If such an estimator $\hat{\theta}$ exists, it is unbiased, attains the minimum variance $\left[k(\theta)\right]^{-1}$, and is termed an efficient estimator.[46]

An example is the estimation of the amplitude A of a coherent field in the presence of incoherent Gaussian radiation. The density operator $\rho(A)$ is then given by $\rho_1(A)$ of Eqs. (3.16) and (3.17), with ρ_0 of Eq. (3.14) taking the place of ρ_b. Comparing Eqs. (3.18) and (5.13), we find the s.l.d.

$$L = \Pi_\theta - 4A \underset{\sim}{\mu}^{+} (I + 2\underset{\sim}{\varphi})^{-1} \underset{\sim}{\mu},$$

where Π_θ is the threshold operator for detecting the field with mode amplitudes $A\mu_{\underset{\sim}{m}}$ in the presence of the same type of background radiation. This threshold operator is given in Eq. (3.19).

An efficient estimator of the amplitude A of the field is, by virtue of Eq. (5.11), the operator

$$\hat{A} = \left[4 \underset{\sim}{\mu}^{+} (\underset{\sim}{I} + 2\underset{\sim}{\varphi})^{-1} \underset{\sim}{\mu}\right]^{-1} \Pi_\theta, \tag{5.14}$$

and it attains the minimum variance

$$\underset{\sim}{E}(\hat{A} - A)^2 = [4 \underset{\sim}{\mu}^{+} (\underset{\sim}{I} + 2\underset{\sim}{\varphi})^{-1} \underset{\sim}{\mu}]^{-1}. \tag{5.15}$$

For background radiation of the thermal variety and a narrow-band signal field, this estimator provides a relative variance

$$\underset{\sim}{E}(\hat{A} - A)^2 / A^2 = (2N + 1)/4N_s, \tag{5.16}$$

where N_s is the mean number of photons in the signal field and N is the mean number of thermal photons per mode. In the classical limit this minimum relative variance becomes equal to $(2E_s/K\mathcal{T})^{-1}$, which is the same as for a classical efficient estimator of the amplitude of a coherent signal of energy E_s and known phase in thermal noise of absolute temperature $\mathcal{T}$.

Efficient estimators can be expected to be at least as rare in quantum estimation theory as in the classical theory, and no general method has been found for producing estimators that come close to the lower bound set by the quantum counterpart, Eq. (5.9), of the Cramér-Rao inequality.

Sufficient Statistics

The density operator $\rho(\theta)$ can sometimes be factored as in Eq. (5.12) into two parts, $V(\hat{\theta}; \theta)$ and its Hermitian conjugate, that depend on the dynamical variables of the system only through the operator $\hat{\theta}$, and a third part ρ_b independent of the unknown parameter θ. The operator $\hat{\theta}$ might then, in analogy with the classical terminology, be called a sufficient estimator, or a sufficient statistic for estimation. The operator $\hat{A}$ in Eq. (5.14) is sufficent for estimating the amplitude of the signal field.

In classical detection theory the sufficient statistic for estimating the amplitude of a coherent signal in Gaussian noise is also sufficent for detecting the signal; that is, the likelihood ratio for detection depends on the input to the receiver only through that statistic, and the optimum decision about the presence of the signal can be based upon it. In the corresponding quantum-detection problem, the amplitude estimator $\hat{A}$ does not provide the optimum detection operator, as is evident from the treatment of detection in the absence of thermal noise, given in Section II. For a coherent signal in Gaussian noise the efficient estimator of signal amplitude is related rather to the threshold statistic for detection. The concept of a sufficient statistic does not, therefore, seem to have the range in quantum-mechanical decision and estimation theory that it possesses in the classical theory.[46]

Multiple Estimation

Thus far we have treated only the estimation of a single parameter of the density operator of the system. In the classical theory the Cramér-Rao inequality has been generalized to cover the simultaneous estimation of several unknown parameters[43, 44], and a corresponding generalization is possible in quantum estimation theory as well[47]. In discussing it we restrict ourselves for simplicity to unbiased estimates.

Let there be m parameters $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_m)$ of the density operator $\rho(\underset{\sim}{\theta})$ to be estimated, and let $\hat{\theta}_j$ be the operator whose measurement yields a number that is taken as the estimate of the parameter θ_j. Since the estimates are assumed to be unbiased,

$$\underset{\sim}{E}(\hat{\theta}_j) = \mathrm{Tr}\left[\hat{\theta}_j \, \rho(\underset{\sim}{\theta})\right] = \theta_j ;$$

we define

$$\delta\hat{\theta}_j = \hat{\theta}_j - \theta_j \mathbf{1}$$

as the operator providing the error in the estimate of θ_j. The co-variance of simultaneous estimates of the parameters θ_i and θ_j is then

$$B_{ij} = \left\{\hat{\theta}_i, \hat{\theta}_j\right\} = \tfrac{1}{2}\,\mathrm{Tr}\left[\rho(\delta\hat{\theta}_i \delta\hat{\theta}_j + \delta\hat{\theta}_j \delta\hat{\theta}_i)\right] \tag{5.17}$$

These covariances form an m x m matrix $\underset{\sim}{B}$, whose diagonal elements are the variances of the errors in the estimates. If the operators $\hat{\theta}_i$ are to be measured on the same system, they must commute in order for the covariances B_{ij} to have a clearly defined physical meaning[48].

The sizes of the errors and their correlations are conveniently visualized in terms of the concentration ellipsoid in an m-dimensional space with Cartesian co-ordinates $\tilde{\underset{\sim}{Z}} = (z_1, z_2, \ldots, z_m)$; its equation is[49]

$$\tilde{\underset{\sim}{Z}}\underset{\sim}{B}^{-1}\underset{\sim}{Z} = m + 2, \tag{5.18}$$

where $\underset{\sim}{Z}$ is a column vector, $\tilde{\underset{\sim}{Z}}$ its transposed row vector. The larger this ellipsoid, the greater the mean-square errors, and an elongation of the ellipsoid in a direction aslant to the co-ordinate axes indicates a correlation among the estimates.

The generalized Cramér-Rao inequality for multiple estimation places this concentration ellipsoid outside the ellipsoid

$$\tilde{\underset{\sim}{Z}}\underset{\sim}{A}\underset{\sim}{Z} = m + 2, \tag{5.19}$$

where $\underset{\sim}{A} = \|A_{ij}\|$,

$$A_{ij} = \tfrac{1}{2}\mathrm{Tr}\ \rho(L_iL_j + L_jL_i) = \mathrm{Tr}(\partial\rho/\partial\theta_i)L_j \tag{5.20}$$

with L_i the s.l.d. of $\rho(\underset{\sim}{\theta})$ with respect to θ_i, defined as in Eq. (5.10). That is, for any column vector $\underset{\sim}{Z}$ of m real elements,

$$\tilde{\underset{\sim}{Z}}\underset{\sim}{B}^{-1}\underset{\sim}{Z} \leq \tilde{\underset{\sim}{Z}}\underset{\sim}{A}\underset{\sim}{Z}.$$

Alternatively, for any column vector $\underset{\sim}{Y}$ of real elements,

$$\tilde{\underset{\sim}{Y}}\underset{\sim}{B}\underset{\sim}{Y} \geq \tilde{\underset{\sim}{Y}}\underset{\sim}{A}^{-1}\underset{\sim}{Y}; \tag{5.21}$$

by picking appropriate values of $\tilde{\underset{\sim}{Y}} = (y_1, \ldots, y_m)$, one can set lower bounds to variances and covariances of unbiased estimates of the unknown parameters. In particular,

$$B_{ii} = \mathrm{Var}\,\hat{\theta}_i = \mathrm{Tr}\,\rho(\hat{\theta}_i - \theta_i\underset{\sim}{1})^2 \geq (\underset{\sim}{A}^{-1})_{ii}, \tag{5.22}$$

which is the i-th diagonal element of the inverse matrix $\underset{\sim}{A}^{-1}$.

The symmetrized logarithmic derivatives needed in the error bounds on both single and multiple estimates can be worked out for parameters of coherent fields and random Gaussian fields observed in the presence of random Gaussian background fields. The density operators have then the forms given by Eqs. (3.14) and (3.16). Details have been presented elsewhere[47].

VI. Conclusion

We have omitted from this review the analysis of actual receivers in which quantum effects are significant and the extension of information theory to channels embodying such receivers. Optical heterodyne receivers and optical detectors of incoherent light have been extensively studied, the types of noise encountered in them have been classified and measured, and methods and data required for their design have been compiled. To simplified models, such as the photon-counting receivers, classical detection and estimation theory have been applied[50-59]. Capacities and information rates of communication channels embodying such receivers have been calculated in order to extend into the quantum domain the results of classical information theory[60-70].

A review of quantum detection and estimation theory itself can at the present time be little more than a recital of unsolved problems. Indeed, a collection of ideas in which such fundamental matters as optimum Bayes estimation and optimum multiple-hypothesis testing remain unresolved can hardly be called a theory at all. Nevertheless, it is eminently reasonable that such a theory should exist. If it can be elaborated sufficiently, it will permit us to specify the ultimate limits that the thermal and quantum properties of nature set to the reliable detection of signals and to accurate measurement of parameters of physical systems.

Footnotes

*This paper was prepared under grant NGR-05-009-079 from the National Aeronautics and Space Administration.

1. For a review of the theory of density operators, see

   U. Fano, Revs. Mod. Phys. **29**, 74 (1957),

   D. Ter Haar, Reports on Progress in Physics (Institute of Physics and the Physical Society, London) **24**, 304 (1961).

2. The word "classical" is used here to mean "nonquantum-mechanical". It does not indicate venerability, most of what is here called "classical statistical theory" being younger than quantum mechanics itself.

3. E. Lehmann, Testing of Statistical Hypotheses (John Wiley & Sons, Inc., New York, 1959).

4. D. Middleton, An Introduction to Statistical Communication Theory (McGraw-Hill Book Company, Inc., New York, 1960), chs. 18 - 23.

5. I. Selin, Detection Theory (Princeton University Press, Princeton, N. J., 1965).

6. J. C. Hancock and P. A. Wintz, Signal Detection Theory (McGraw-Hill Book Company, Inc., New York, 1966)

7. A. V. Balakrishnan (ed.), Communication Theory (McGraw Hill Book Company, Inc., New York, 1968), chs. 3, 4.

8. C. W. Helstrom, Statistical Theory of Signal Detection (Pergamon Press, Ltd., Oxford, England, 1968), 2nd ed.

9. H. Van Trees, Detection, Estimation, and Modulation Theory (John Wiley & Sons, Inc., New York, 1968), vol. 1.

10. The Bayes whose name figures so largely here was the Rev. Thomas Bayes (1702 - 1761), whose paper, "An Essay Toward Solving a Problem in the Doctrine of Chances," Phil. Trans. Roy. Soc. (London) **53**, 370 (1763) proposed basing decisions on posterior or conditional probabilities. It has been reprinted in Biometrika **45**, 293 (1958).

11. Ref. 8, pp. 82, 91; ref. 9, p. 26.

12. Ref. 8, p. 194; ref. 9, p. 46.

13. J. Neyman and E. Pearson, Proc. Camb. Phil. Soc. 29, 492 (1933).

14. J. Neyman and E. Pearson, Phil. Trans. Roy. Soc. (London) A231, 289 (1933).

15. The notations $Q_0 = \alpha$ and $Q_d = \beta$ are common in the statistical literature, where these probabilities are termed the size and the power of the test, respectively.

16. Ref. 8, pp. 87, 93; ref. 9, p. 33.

17. Ref. 8, ch. VIII, p. 249; ref. 9, §2.4, p. 52.

18. C. W. Helstrom, Information and Control 10, 254 (1967)

19. C. W. Helstrom, Int. J. Theor. Phys. 1, 37 (1968).

20. C. W. Helstrom, Information and Control 13, 156 (1968).

21. A derivation taking into account the possibility that randomization may be necessary is given in ref. 18.

22. F. Riesz and B. Sz.-Nagy, Functional Analysis (F. Ungar Publishing Co., New York, 1955), p. 238. I am indebted to Professor E. Masry for pointing out this theorem.

23. W. H. Louisell, Radiation and Noise in Quantum Electronics (McGraw-Hill Book Company, Inc., New York, 1964), p. 242.

24. Ref. 20, §8.

25. P. A. Bakut and S. S. Shchurov, Problemy Peredachi Informatsii 4, no. 1, 77 (1968).

26. R. J. Glauber, Phys. Rev. 131, 2766 (1966)

27. Ref. 19, p. 45.

28. G. Lachs, Phys. Rev. 138, B1012 (1965).

29. C. W. Helstrom, Trans. IEEE AES-5, (1969).

30. D. Middleton, Trans. IEEE IT-12, 230 (1966).

31. Ref. 18, p. 273.

32. Ref. 18, p. 275.

33. Ref. 23, p. 153.

34. Ref. 20, p. 165f.

35. Ref. 8, Ch. IV.

36. Ref. 18, p. 284ff.

37. G. Doetsch, Handbuch der Laplace-Transformation (Birkhäuser Verlag, Basel and Stuttgart, 1955) vol. 2, ch. 3.

38. C. W. Helstrom, submitted to J. Opt. Soc. Am.

39. A. A. Kuriksha, Problemy Peredachi Informatsii 3, no. 1, 42 (1967).

40. A. A. Kuriksha, Radiotekhnika i Elektronika 13, 1790 (1968)

41. J. von Neumann, Mathematische Grundlagen der Quantenmechanik (Springer Verlag, Berlin, 1932) p. 130.

42. Ref. 8, ch. VIII; ref. 9, §2.4, p. 52.

43. H. Cramér, Mathematical Methods of Statistics (Princeton University Press, Princeton, N. J., 1946), p. 473ff.

44. C. R. Rao, Bull. Calcutta Math. Soc. 37, 81 (1945).

45. C. W. Helstrom, Phys. Letters 25A, 101 (1967).

46. Ref. 20, pp. 164ff.

47. C. W. Helstrom, Trans. IEEE IT-14, 234 (1968).

48. H. Margenau and R. N. Hill, Progr. Theoret. Phys. (Kyoto) 26, 722 (1961).

49. L. Schmetterer, Mathematische Statistik (Springer Verlag, New York, 1966), p. 63.

50. B. Reiffen and H. Sherman, Proc. IEEE 51, 1316 (1963).

51. C. W. Helstrom, Trans. IEEE IT-10, 274 (1964).

52. J. W. Goodman, IEEE J. Quant. Elec. QE-1, 180 (1965).

53. J. W. Goodman, Trans. IEEE AES-2, 526 (1966).

54. P. A. Bakut, V. G. Vygon, A. A. Kuriksha, V. G. Repin, G. P. Tartakovskii, Problemy Peredachi Informatsii 2, no. 4, 39 (1966).

55. V. L. Stefanyuk, ibid. 2, no. 1, 58 (1966).

56. P. A. Bakut, Radio Engrg. & Electr. Phys. 11, 551 (1966).

57. S. A. Ginzburg, ibid. 11, 1972 (1966).

58. P. A. Bakut, ibid. 12, 1 (1967)

59. I. Bar-David, Trans. IEEE IT-15, 31 (1969).

60. T. E. Stern, Trans. IRE IT-6, 435 (1960)

61. J. P. Gordon, Proc. IRE 50, 1898 (1962)

62. B. E. Goodwin and L. P. Bolgiano, Jr. Proc. IEEE 53, 1745 (1965)

63. L. F. Jelsma and L. P. Bolgiano, IEEE Ann. Commun. Conv. Record, 635 (1965).

64. L. B. Levitin, Problemy Peredachi Informatsii 1, no. 3, 71 (1965).

65. H. Takahasi, Advances in Communication Systems (A. V. Balakrishnan, ed., Academic Press, New York) 1, 227 (1965).

66. D. S. Lebedev and L. B. Levitin, Information and Control 9, 1 (1966).

67. V. V. Mityugov, Problemy Peredachi Informatsii 2, no. 3, 48 (1966).

68. R. L. Stratonovich, ibid. 2, no. 1, 45 (1966).

69. S. I. Borovitskii, V. V. Mityugov, ibid. 3, no. 1, 35 (1967).

70. G. Lachs and G. Fillmore, Trans. IEEE IT-15, (1969).