PHYSICAL REVIEW LETTERS

VOLUME 76 | 1 JANUARY 1996 | NUMBER 1

Minimum Decision Cost for Quantum Ensembles

Dorje Brody[1,*] and Bernhard Meister[1,2,†]

[1]*Blackett Laboratory, Imperial College, South Kensington, London SW7 2BZ, United Kingdom*

[2]*Isaac Newton Institute For Mathematical Science, 20 Clarkson Road, Cambridge CB3 0EH, United Kingdom*

(Received 18 July 1995)

We calculate the binary decision costs of different strategies for measurement of a given ensemble of N independent and identically prepared polarized spin $1/2$ particles. The results obtained prove that, for arbitrarily given values of the prior probabilities and an arbitrary number of constituent particles, the cost of a combined measurement is the same as that for sequential measurements of the individual particles.

PACS numbers: 02.50.Le, 03.65.Bz, 05.30.Ch, 89.70.+c

The task of the experimentalist in a problem of experimental design is to find an optimal observational strategy. Ordinarily one must choose among different strategies before the data can be obtained, and hence one must perform a *preposterior analysis.* When the experiment involves a decision among different quantum mechanical states, such an analysis is indeed important, since, unlike the classical case, repeated samplings of the same system are not generally permitted.

There are a number of different approaches to the determination of optimal strategies. In the information-theoretic approach, one typically ascertains the strategy that maximizes the mutual information (see, e.g., [1]), but this is generally difficult, owing to the nonlinear nature of the Shannon information. In the minimax approach [2], one finds the strategy that minimizes the maximum cost (or loss) incurred by the decision among different strategies. When certain *a priori* knowledge concerning the nature of the state is available, then one may use a Bayes procedure to seek a strategy that minimizes the expected cost [2,3]. This approach is based upon repeated use of Bayes' theorem in order to replace prior by posterior distributions in accordance with the data obtained from experiments.

In the present Letter, we study the Bayesian approach to a binary decision problem (a decision requiring choices between two different states). First, we briefly introduce the Bayesian approach to quantum hypothesis testing. These notions, developed by Helstrom and others [4–6] and also recently extended by Jones [7], are then applied to obtain the optimal strategy for a Bayes decision between two quantum mechanical pure states, for an ensemble of polarized spin $1/2$ particles. In this example, we first consider the application of (quantum) Bayes sequential analysis to this ensemble. The result is then compared with a combined measurement of the entire ensemble, treated as a single composite system (i.e., simultaneous measurement of all the individual particles). Other strategies consisting of combined measurements of subensembles are also considered. The Bayes solution to the problem demonstrates that the Bayes cost for separate sequential measurements of the individual particles is the same as that of a single combined measurement. This result differs from that predicted by Peres and Wootters [8]. Any other strategy turns out to entail a higher expected cost.

First, consider a decision problem requiring a choice among M hypotheses $H_1, \ldots, H_M$ concerning a quantum system. Hypothesis H_k asserts that the system is in the state having the associated density operator $\hat{\rho}_k$ ($k = 1, \ldots, M$), and the prior probability of the jth state is ξ_j, with

$$\sum_{k=1}^{M} \xi_k = 1 \, . \tag{1}$$

From past experience, one knows that the system is in the jth state with a relative frequency ξ_j. The self-adjoint operators $\hat{\rho}_k$ act on the vectors of a Hilbert space $\mathcal{H}$, are non-negative definite, and have unit trace.

A *quantum decision strategy* is characterized by a *probability operator measure* (POM) [4] on $\mathcal{H}$, i.e., a set of M non-negative definite self-adjoint operators Π_j satisfying

$$\sum_{j=1}^{M} \Pi_j = \mathbf{1} . \tag{2}$$

If this POM is applied to the system when hypothesis H_k is true, then the conditional probability of choosing hypothesis H_j is given by

$$\Pr(X = j \,|\, W = k) = \mathrm{Tr}(\hat{\rho}_k \Pi_j) . \tag{3}$$

Here X denotes the random variable that is to be observed and W, typically being the parameter, is the unknown state of nature.

Now let C_{ij} be the cost of choosing hypothesis H_i when H_j is true. Then the expected cost of the observational strategy specified by the POM $\{\Pi_j\}$ is [4]

$$\overline{C} = \sum_{i=1}^{M} \sum_{j=1}^{M} \xi_j C_{ij} \mathrm{Tr}(\hat{\rho}_j \Pi_i) \equiv \mathrm{Tr} \sum_{i}^{M} R_i \Pi_i , \tag{4}$$

where the Hermitian *risk operators* R_i are defined by

$$R_i = \sum_{j=1}^{M} \xi_j C_{ij} \hat{\rho}_j . \tag{5}$$

A set $\{\Pi_j\}$ of POM that minimizes the cost (4) under the constraints (2) is defined as optimal, and the cost is Bayes, i.e., $\overline{C} = \overline{C}^*$ (the superscript $*$ here corresponds to the optimal strategy). Necessary and sufficient conditions for the optimality of a POM are known to be [5,6] the self-adjointness of the operator

$$\Upsilon = \sum_{j=1}^{M} R_j \Pi_j = \sum_{j=1}^{M} \Pi_j R_j \tag{6}$$

and the non-negative definiteness of the operator $R_j - \Upsilon$ for all $j = 1, \ldots, M$. The minimum expected Bayes cost is thus

$$\overline{C}^*(\xi, \{\Pi_j^*\}) = \mathrm{Tr}\Upsilon . \tag{7}$$

In a simple case where $M = 2$, i.e., for binary decisions, one can easily verify [4] that the optimal POM is projection valued, and the Bayes cost becomes

$$\overline{C}^*(\xi, \{\Pi_j^*\}) = \xi_1 C_{11} + \xi_2 C_{12} - \xi_2 (C_{12} - C_{22}) \sum_{\eta_i > 0} \eta_i , \tag{8}$$

where η_i are the eigenvalues of the operator $\hat{\rho}_2 - \gamma \hat{\rho}_1$, with

$$\gamma = \frac{\xi_1 (C_{21} - C_{11})}{\xi_2 (C_{12} - C_{22})} = \frac{\xi}{1 - \xi} . \tag{9}$$

Here and in the sequel, we choose a 0-1 cost structure $C_{ij} = 1 - \delta_{ij}$, i.e., assign cost 1 to an incorrect decision and 0 to a correct decision. Also the prior probability for state 1 is given by $\xi_1 = \xi$, and hence $\xi_2 = 1 - \xi$.

Now consider an experiment where a physicist must estimate (decide) the direction of polarization of a given ensemble of N spin $1/2$ particles, using a Stern-Gerlach (SG) device. The physicist knows that the particles have been filtered through another SG device with a magnetic field in the x-y plane at a constant angle θ_1 or θ_2 from the x axis, and in either case the spin-up state has been selected. The physicist can select the orientation angle ϕ of the detector relative to the x axis. When the particle passes through the field of the detector magnet, the physicist observes either the spin-up (head) or spin-down (tail) state, whereupon he must decide between the alternatives θ_1 (i.e., the polarization direction $\theta = \theta_1$) and θ_2. The values of the angles $\{\theta_k\}$ are not specified, but the difference between the two angles is given by $|\theta_2 - \theta_1| = 2\delta$.

First, consider the case where the physicist performs sequential observations of the individual spin $1/2$ particles. Suppose, for simplicity, that $N = 1$. The physicist must decide, either before or after the observation, whether the particle is polarized in the θ_1 or θ_2 direction. If a decision were to be chosen without any observation, then a Bayes decision against the prior distribution $\xi(W)$ of W (in this case, $W = 1$ or 2) would be optimal. Suppose that X (spin "up" or "down") is observed before a decision is chosen. Then the physicist follows the same decision procedure as in the previous case. However, the difference here is that the distribution of W has changed from the prior to the posterior distribution. Hence a Bayes decision against the posterior distribution of W is now optimal.

When the state of the system is $\hat{\rho}_k$, the conditional probability for observing the spin-up (+1) state is given by

$$b_k(\phi) \equiv \Pr(X = +1 \,|\, W = \theta_k) = \cos^2\!\left(\frac{\theta_k - \phi}{2}\right). \tag{10}$$

If one fixes the angle ϕ, then the experiment is entirely analogous to the classical coin tossing problem [9] for coins with bias given by the above b_k. However, having the freedom to choose the angle ϕ for each value of the prior ξ, the physicist must choose an optimal direction given by [10]

$$\phi_{\mathrm{opt}}(\xi) = \tan^{-1}\!\left(\frac{\xi \sin\theta_1 - (1 - \xi)\sin\theta_2}{\xi \cos\theta_1 - (1 - \xi)\cos\theta_2}\right), \tag{11}$$

which can be calculated by finding the optimal POM $\{\Pi_j^*\}$ [9]. Hence we have a problem of tossing quantum coins with bias depending upon the prior probability ξ.

Having chosen the optimal angle $\phi_{\rm opt}$, the Bayes decision rule specifies that θ_1 is to be chosen if the spin-up state is observed, and θ_2 otherwise. The Bayes cost against the prior ξ, when $N = 1$, can easily be obtained by calculating the eigenvalues of $\hat{\rho}_2 - \gamma\hat{\rho}_1$, with the result [10]

$$\overline{C}^*(\xi, 1) = \frac{1}{2}\left(1 - \sqrt{1 - 2\xi(1 - \xi)(\cos 2\delta + 1)}\right). \tag{12}$$

Now suppose that $N = 2$ and that the result of measurement of the first particle has been obtained. As mentioned above, the physicist must follow the same procedures as in the case $N = 1$, with the prior ξ replaced by the posterior distribution $\xi(\pm)$. From Bayes' theorem, the posterior probability that $\theta = \theta_1$ is given by

$$\xi(+) = \frac{b_1(\phi)\xi}{b_1(\phi)\xi + b_2(\phi)(1 - \xi)} \tag{13}$$

or

$$\xi(-) = \frac{[1 - b_1(\phi)]\xi}{[1 - b_1(\phi)]\xi + [1 - b_2(\phi)](1 - \xi)}, \tag{14}$$

according to the outcome (+ or –) of the first measurement. The optimal orientation angle, before performing the second measurement, is now given by $\phi_{\rm opt}(\xi(+))$ or $\phi_{\rm opt}(\xi(-))$, respectively. The Bayes cost for this case ($N = 2$) is given by the weighted average, i.e.,

$$\overline{C}^* = b_1\xi\overline{C}^*(\xi(+), 1) + b_2(1 - \xi)\overline{C}^*(\xi(-), 1).$$

Next consider an arbitrary number N of particles. Again the procedures are the same as above, except that the prior is now replaced by one of the 2^{N-1} posteriors $\{\xi(+ + \cdots +), \ldots\}$, after observations of $N - 1$ particles. In a classical Bayes decision procedure [2,3], it is difficult to obtain the Bayes cost as a closed function of N. The reason is that, first, one must study the *tree* [11] of the posterior distributions, with branches proliferating as $\sim 2^N$. To each branch (i.e., posterior) of the tree, one associates the cost $\overline{C}^*(\cdot, 1)$ and then calculates the weight (probability) for the sequence of outcomes associated with that branch. After these considerations, one can, in principle, obtain the weighted average of the cost, which involves 2^{N-1} terms. (Note that, for classical procedures, the branches of the posterior tree do recombine and hence proliferate as $\sim N$. However, the weights associated with the branches do not recombine, and therefore one cannot avoid the consideration of 2^{N-1} terms.)

In the case of our "quantum coins," the situation appears even worse, since, after each observation, the physicist must turn the device in accordance with formula (11). This results in changing the bias $b_k(\phi)$ of the "coins" at each stage, and hence one must also incorporate the bias tree (which proliferates $\sim 2^N$). However, it turns out that this optimal orientation forces the posterior tree to recombine into two branches, i.e.,

$$\xi(n, \pm) = \frac{1}{2}\left(1 \pm \sqrt{1 - 4\xi(1 - \xi)\cos^{2(n-1)}\delta}\right), \tag{15}$$

where $\pm$ corresponds to the outcome of the last $[(n - 1)$th$]$ trial being spin up (+) or down (–). This result can be proven by induction as follows. First, for $n = 1$, it is easily verified that $\xi(1, \pm) = \xi(\pm)$ as given in (13) and (14). Next assume that the last (nth) outcome of the trial is (–) and that the posterior is given by the above $\xi(n, -)$. Then, if the next trial outcome is (+), it follows from Bayes' theorem that the posterior distribution, after $n + 1$ observations, is given by

$$\xi(\cdots - +) = \frac{b_1(\phi)\xi(n, -)}{b_1(\phi)\xi(n, -) + b_2(\phi)[1 - \xi(n, -)]},$$

with $\phi = \phi_{\rm opt}(\xi(n, -))$. After some algebra, one can show that the above $\xi(\cdots - +) = \xi(n + 1, +)$. The other three cases $[\xi(\cdots - -)$, etc.$]$ can also be treated in the same manner.

The weights for different branches do not recombine in the quantum case either, however, since $\overline{C}^*(\xi(n, +), 1) = \overline{C}^*(\xi(n, -), 1)$; the final average cost is just $\overline{C}^*(\xi(n, \pm), 1)$ times the sum of all the different weights (which is just 1), and hence we finally deduce that the Bayes cost for sequential observations is

$$\begin{aligned}\overline{C}^*(\xi, N) &= \overline{C}^*(\xi(N, \pm), 1) \\ &= \frac{1}{2}\left(1 - \sqrt{1 - 4\xi(1 - \xi)\cos^{2N}\delta}\right),\end{aligned} \tag{16}$$

for either value of the $(N - 1)$th outcome (+ or –).

Next consider the case where the physicist treats the entire ensemble as a single composite system. The total spin of a system with N particles is just $N/2$, and the density operator for a spin $N/2$ particle polarized in the direction $\mathbf{n} = (\cos\theta, \sin\theta, 0)$ is given by

$$[\hat{\rho}(\theta)]_{mn} = 2^{-N}\sqrt{{}_NC_m\,{}_NC_n}\,e^{-i(m-n)\theta}, \tag{17}$$

where $(n, m) = 0, \ldots, N$, and ${}_NC_m = N!/m!(N - m)!$. According to the result in (8), one must find the eigenvalues of the matrix $\hat{\rho}_2 - \gamma\hat{\rho}_1$ in order to obtain the Bayes cost. We first calculate the eigenvalues as follows. Define two vectors $\mathbf{u} = \{u_n\}$ and $\mathbf{v} = \{v_n\}$ by

$$u_n \equiv 2^{-N/2}\sqrt{{}_NC_n}\,e^{in\theta_1} \tag{18}$$

and

$$v_n \equiv 2^{-N/2}\sqrt{{}_N C_n}\, e^{in\theta_2}. \tag{19}$$

Then $(\hat{\rho}_1)_{mn} = u_m^* u_n$ and $(\hat{\rho}_2)_{mn} = v_m^* v_n$. Since the inner product $\mathbf{u} \cdot \mathbf{u}^* = \mathbf{v} \cdot \mathbf{v}^* = 1$, one obtains

$$\hat{\rho}_1 \mathbf{u}^* = \sum_n (\hat{\rho}_1)_{mn} u_n^* = \mathbf{u}^*$$

and, similarly, $\hat{\rho}_2 \mathbf{v}^* = \mathbf{v}^*$. Now let $\mathbf{w}$ and λ be, respectively, an eigenvector and the corresponding eigenvalue of the matrix $\hat{\rho}_2 - \gamma \hat{\rho}_1$, i.e.,

$$(\hat{\rho}_2 - \gamma \hat{\rho}_1)\mathbf{w} = \lambda \mathbf{w}\,. \tag{20}$$

We may expand the eigenvector $\mathbf{w}$ in terms of a basis that contains either $\mathbf{u}^*$ or $\mathbf{v}^*$, i.e., $\mathbf{w} = c_1 \mathbf{u}^* + \mathbf{u}_\perp^*$ or $\mathbf{w} = c_2 \mathbf{v}^* + \mathbf{v}_\perp^*$. Here $\mathbf{u}_\perp^*$ denotes some vector orthogonal to $\mathbf{u}^*$, and similarly for $\mathbf{v}_\perp^*$. However, since $\hat{\rho}_1 \mathbf{u}_\perp^* = \hat{\rho}_2 \mathbf{v}_\perp^* = 0$, we have

$$\lambda \mathbf{w} = c_1 \mathbf{u}^* - \gamma c_2 \mathbf{v}^*. \tag{21}$$

On the other hand, if we form the inner product of the two vectors $\mathbf{w} = c_1 \mathbf{u}^* + \mathbf{u}_\perp^*$ and $\mathbf{u}$, we obtain

$$\mathbf{w} \cdot \mathbf{u} = c_1 = \frac{c_1}{\lambda} - \frac{\gamma}{\lambda} c_2 (\mathbf{v}^* \cdot \mathbf{u})\,, \tag{22}$$

and similarly

$$\mathbf{w} \cdot \mathbf{v} = c_2 = \frac{c_1}{\lambda} (\mathbf{u}^* \cdot \mathbf{v}) - \frac{\gamma}{\lambda} c_2\,. \tag{23}$$

Without any loss of generality, we may now set $c_1 = 1$, and then by eliminating c_2 from the above equations, we obtain the eigenvalues of the matrix $\hat{\rho}_2 - \gamma \hat{\rho}_1$, i.e.,

$$\lambda_\pm = \frac{1}{2}\left\{(1 - \gamma) \pm \sqrt{(1 - \gamma)^2 - 4\gamma(\Delta^2 - 1)}\right\}, \tag{24}$$

where

$$\Delta^2 = (\mathbf{v}^* \cdot \mathbf{u})(\mathbf{u}^* \cdot \mathbf{v}) = \cos^{2N}(\delta)\,. \tag{25}$$

Therefore, the binary Bayes decision cost for a spin $N/2$ particle is

$$\overline{C}^*(\xi, N) = \frac{1}{2}\left(1 - \sqrt{1 - 4\xi(1 - \xi)\cos^{2N}\delta}\right). \tag{26}$$

One immediately observes that the above cost (26) is the same as that obtained from sequential analysis, given by (16). Hence the Bayes solution to our optimization problem shows that a combined measurement is as advantageous as sequential measurements. These two strategies, however, are not the only ones, and many other partially combined measurement procedures are possible. However, in the present formalism of sequential analysis, the only effect of any intermediate measurements, either partially combined or not, consists in updating the posterior distributions. That is, the results of any intermediate measurements are not directly involved in the final decision. Since the Bayes cost is a monotonically decreasing function of the number of updating steps, which can be seen from Eqs. (15) and (16), this implies that any partially combined measurements will increase the cost. Therefore, we may now conclude that the optimal measurement strategy consists in either performing a combined measurement of the entire ensemble or performing sequential measurements of the individual particles. Any other strategies will result in higher costs (unless we assign different weights to partially combined intermediate measurements).

This result is quite different from that expected by Peres and Wootters, who conjectured that sequential measurements can never be as efficient as a combined measurement [8]. However, it is important to note that their conjecture is based upon an information-theoretic approach, and the solution of an optimization problem using a Bayesian approach can yield a different result. Massar and Popescu [12], on the other hand, have proved the above-mentioned conjecture explicitly for the case $N = 2$. The method used therein is effectively similar to a Bayesian approach, without the use of the prior distributions, but with a nondiagonal cost function and an infinite number of hypotheses. However, when a prior distribution is available, the Bayes solution is known to be optimal in general [2]. In our case, the prior is given to the observer as a part of the experimental setup; hence the criticism sometimes directed against the subjective nature of the Bayesian approach does not apply. If prior knowledge is not available, one can still employ the Bayesian approach, using a noninformative prior. However, the analysis of such cases is beyond the scope of the present Letter.

Throughout the present Letter, we have considered only the cost associated with making decisions. In any practical situation, on the other hand, one must take into consideration other costs (e.g., the observational cost, the cost of analyzing the results, etc.). In our example of sequential analysis, for example, at each stage before performing an observation, the physicist must analyze the previous results in order to determine the optimal turning angle. Assuming the linearity of the utility function (e.g., that the total cost is just the sum of the decision cost and the observational costs), one can easily find the optimal strategy in a more general situation. These other costs, however, depend upon the specific applications (experiments) concerned and therefore cannot be treated in a systematic manner.

In connection with the decision problem for classical coins which was briefly mentioned above, it is interesting to note that the present quantum binary decision problem corresponds to a classical problem where the number of elements in the decision space is infinite. More details of this, as well as a treatment including the observational costs, may be found in [9].

Although we have considered only the binary decision problem ($M = 2$), our result can be extended to the case $M > 2$ if the number of hypotheses M is finite. This can be seen by introducing Helstrom's cost reduction algorithm [13], where multiple decisions can be reformulated in terms of binary decisions. However, the validity of our result for an infinite number of hypotheses remains to be proved.

The authors acknowledge their gratitude to J. D. Malley for useful discussions of the foregoing topics.

*Electronic address: d.brody@tp.ph.ic.ac.uk
†Electronic address: bm1@ic.ac.uk

[1] E. B. Davies, IEEE Trans. Inf. Theory **24**, 596 (1978).
[2] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Spinger-Verlag, New York, 1985).
[3] M. H. DeGroot, *Optimal Statistical Decisions* (McGraw-Hill, New York, 1970).
[4] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[5] A. S. Holevo, J. Multivar. Anal. **3**, 337 (1973).
[6] H. P. Yuen, R. S. Kennedy, and M. Lax, IEEE Trans. Inf. Theory **21**, 125 (1975).
[7] K. R. W. Jones, Phys. Rev. A **50**, 3682 (1994).
[8] A. Peres and W. K. Wootters, Phys. Rev. Lett. **66**, 1119 (1991).
[9] D. Brody and B. Meister, "Bayesian Inference in Quantum Systems" (to be published).
[10] J. D. Malley and J. Hornstein, Statist. Sci. **8**, 433 (1993).
[11] D. V. Lindley, *Making Decisions* (Wiley, New York, 1971).
[12] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995).
[13] C. W. Helstrom, IEEE Trans. Inf. Theory **28**, 359 (1982).