V. 8, n.º 3 (setembre 1984) pp. 121-126

THE QUADRATIC SCORING RULE AS A BASIS FOR EXPERIMENT COMPARISON

PILAR GARCÍA - CARRASCO

UNIVERSIDAD COMPLUTENSE DE MADRID

In this paper, the role of strictly proper quadratic utility measures in Bayesian inference and experiment comparison is investigated. This utility function, which is not local, provides a good basis for the resolution of many real situations which are not purely inferential. By recognizing the decision problem underlying statistical inference, we obtain the criteria for experiment comparison based on the quadratic scoring rule. Properties and relations with other criteria are studied, particularly with those based on sufficiency and on the maximation of the informational energy gain. Finally some particular cases are commented and new areas of research suggested.

Keywords: UTILITY. COMPARISON OF EXPERIMENTS. INFORMATIONAL ENERGY GAIN. INFORMATION. DECISION THEORY. STATISTICAL INFERENCE.

1. INTRODUCTION.

In the scientific literature many comparison of experiments criteria have been proposed and studied. As basic works in this topic we shall mention those performed by Bohnenblust, Shapley and Sherman (unpublished), Blackwell /4/, /5/, /6/, Lindley /15/, Lehmann /14/, Raiffa and Shlaifer /16/, De Groot /12/,/13/, Fedorov /9/, etc...

The comparison among different criteria is a topic of fundamental interest in which works of Blackwell /4/, /5/ and Lindley /15/, stand out.

García-Carrasco /10/ performs a unified revision of the former criteria, and a study of their main properties. All of them were considered as convenient procedures in different situations and mainly defended because of their good behaviour.

From a Bayesian point of view the only correct way of obtaining the optimal experiment is maximizing the expected utility.

Bernardo /2/ consideres the inference problem as a particular case of decision problem in which the different posterior distributions of the parameter of interest constitute the decision space; he justifies the necesity of working with proper utility functions and he shows that, if these functions are local, then the only adequate criterion for comparison of experiment is that proposed by Lindley, based on the maximization of Shanon information.

The constraint to proper utilities for the treatment of inference problems is unquestionable. It is different from the property of local utility, which only seems convenient in pure inference problems, in which the decisor is only interested in the knowledge of the true value of the parameter. Therefore, there is an open a field of investigation which is going to be the objective of this paper: to explore the consequences of working with a proper but non local utility function.

The problem is wide and in this paper we have only developed some primary ideas. We start with the most simple case of non-local proper utility, the quadratic utility. Concretely, considering the convenience of the invariance under one - to - one utility function trans-

- Pilar García-Carrasco - Universidad Complutense de Madrid - Ciudad Universitaria - Madrid.

formations of the utility function, we define following Bernardo /1/

$$u(p(.),\theta) = A\left\{2\,\frac{p(\theta)}{\pi(\theta)} - \int_\Omega \frac{p(\theta)^2}{\pi(\theta)}\,d\theta - 1\right\}$$

where $\theta \in \Omega$ (parameter space), $p(.) \in D$ (decision set of posterior densities on Ω), $A > 0$ is an arbitrary constant and $\pi(.)$ is a reference density which sets up the origin of the utilities scale since $u(\pi(.),\theta) = 0$.

Applying to this utility function the general criterion of maximization of the expected utility, we obtain in this paper a criterion for comparing experiments which reproduces in the discrete case the criterion of maximization of the informational energy gain one, proposed and studied by Garcia-Carrasco /11/, and, in the continuous case, is tightly connected with it; in fact, it is possible to transfer the results obtained there to the problem we are interested in.

2. QUADRATIC INFORMATION SUPPLIED BY AN EXPERIMENT

Let Ω be a parameter space, let $p(\theta)$, $\theta \in \Omega$ be the prior density and let $X = \{\mathcal{X};\ p(x/\theta),\ \theta \in \Omega\}$ be an experiment.

The expected utility of X when the prior density is $p(\theta)$ may generally be defined, following of principle maximization of the expected utility as:

$$u(X\ ;\ p(\theta)) = \int_\mathcal{X} p(x) \sup_{d \in D} \int_\Omega u(d\ ,\theta)\, p(\theta/x)\, d\theta\, dx - \sup_{d \in D} \int_\Omega u(d\ ,\theta)\, p(\theta)\, d\theta$$

In our case, since D is the set of densities on Ω and u a proper utility function,

$$u(X\ ;\ p(\theta)) = \int_\mathcal{X} p(x) \int_\Omega u(p(./x),\theta)\, p(\theta/x)\, d\theta\, dx - \int_\Omega u(p(.),\theta)\, p(\theta)\, d\theta$$

and, in particular, for the invariant quadratic utility,

$$u(X;\ p(\theta)) = \int_\mathcal{X} p(x) \int_\Omega A\left(2\,\frac{p(\theta/x)}{\pi(\theta)} - \int_\Omega \frac{p(\theta/x)^2}{\pi(\theta)}\,d\theta - 1\right) p(\theta/x)\, d\theta\, dx - \int_\Omega A\left(2\,\frac{p(\theta)}{\pi(\theta)} - \int_\Omega \frac{p(\theta)^2}{\pi(\theta)}\,d\theta - 1\right) p(\theta)\, d\theta =$$

$$= A\left(\int_\mathcal{X}\int_\Omega p(x)\,\frac{p(\theta/x)^2}{\pi(\theta)}\,d\theta\, dx - \int_\Omega \frac{p(\theta)^2}{\pi(\theta)}\,d\theta\right) =$$

$$= A\, G_\pi(X\ ;\ p(\theta)) \qquad (2.3)$$

where, for posterior simplification, G is defined as the expression in brackets.

Properties

1. If $w = g(\theta)$ is a differentiable one - to - one transformation, $W = g(\Omega)$, and $q(.)$ and $\pi^*(.)$ are the densities on W implied by $p(.)$ and $\pi(.)$ respectively, it is verified as expected since u was defined to be invariant, that

$$G_{\pi^*}(X;\ q(w)) = \int_\mathcal{X}\int_W p(x)\,\frac{q(w/x)^2}{\pi^*(w)}\,dw\,dx - \int_W \frac{q(w)^2}{\pi^*(w)}\,dw = \int_\mathcal{X}\int_\Omega p(x)\,\frac{p(\theta/x)^2}{\pi(\theta)}\,d\theta\,dx - \int_\Omega \frac{p(\theta)^2}{\pi(\theta)}\,d\theta = G_\pi(X;\ p(\theta)).$$

2. In particular, considering any reference $\pi(\theta)$, such that $\pi(\theta) > 0 \quad \forall\ \theta \in \Omega$, and denoting $g = F_\pi$ the correspondent distribution function, we have the biyection $w = g(\theta)$ for which $\pi(\theta)$ implies a distribution $\pi^*(w)$: $U(0, 1)$, and it is verified that:

$$G_\pi(X;\ p(\theta)) = G_{U(0,1)}(X;\ q(w)) =$$

$$\int_X \int_W p(x)q(w/x)^2\, dxdw -$$

$$- \int_W q(w)^2\, dw = GAN\ (X;\ q(w)) \qquad (2.4)$$

where GAN (X; q(w)) is the informational energy gain mentioned in the introduction and defined exactly as appearing in the last expression.

3. COMPARISON OF EXPERIMENTS CRITERION. DEFINITION AND PROPERTIES.

Let us consider an inference problem on $\theta \in \Omega$ with a prior distribution $p(\theta)$, and let us suppose that given the concrete conditions of the problem, we are interested in a quadratic (proper, non-local and invariant) utility function with reference density $\pi(\theta)$.

We have seen that in these conditions the utility of an experiment X is $u(X;\ p(\theta)) = AG_\pi(X;\ p(\theta))$. Therefore, if we have at our disposal two experiments with similar cost, $X = \{\mathcal{X};\ p(x/\theta),\ \theta \in \Omega\}$ and $Y = \{\mathcal{Y};\ p(y/\theta),\ \theta \in \Omega\}$ we will prefer X to Y ($X \overset{\pi}{\gtrsim} Y$) if and only if

$$u(X;\ p(\theta)) \geqslant u(Y;\ p(\theta)) \iff G_\pi(X;\ p(\theta)) \geqslant$$

$$\geqslant G_\pi(Y;\ p(\theta)) \iff GAN\ (X;\ q(w)) \geqslant GAN\ (Y;\ q(w))$$

using the notation given in (2.4). Moreover, X is defined to be equivalent to Y ($X \overset{\pi}{\sim} Y$) if and only if $X \overset{\pi}{\gtrsim} Y$ and $Y \overset{\pi}{\gtrsim} X$.

Owing to this close relation with the maximization of informational energy gain criterion all the properties and relationships with other criteria studied by Garcia-Carrasco /11/ may be stated here. Their proof simply consists of using the similar property already proved by the GAN, and stating it in terms of quadratic utilities making the transformation $q = F_\pi$.

Moreover, taking notice that the expressions (2.1), (2.2) and (2.3) are no more than a particular case of the expected value of the sample information (EVSI) which was defined by Raiffa and Schlafer /16/(4,5), all the properties and relationships studied for that expected sample information by García-Carrasco /10/, can also be stated for the criterion here considered, ($X \overset{\pi}{\gtrsim} Y$).

We now formulate the main results obtained from both points of view. We assume unless specifically otherwise stated that a prior density $p(\theta)$, and a reference distribution $\pi(\theta)$ have been chosen beforehand.

Properties

1.-The relationship $\overset{\pi}{\gtrsim}$ is a complete preorder.

2.-For all experiment X, $G_\pi(X;\ p(\theta)) \geq 0$, with equality if and only if $p(x/\theta)$ is independent of θ a.e. (almost everywhere).

3.-$X \overset{\pi}{\gtrsim} N$, where X is any experiment and $N = \{\mathcal{Y};\ B;\ Q\}$ is the null experiment.

4.-For all compound experiment $(X_1,\ X_2)$, and being X_1 the correspondent marginal experiment, it is verified that $(X_1, X_2) \overset{\pi}{\gtrsim} X_1$ verifying $\overset{\pi}{\sim}$ if and only if $p(x_2/x_1,\ \theta)$ is independent of θ a.e.

5.-Let $X^{(n)}$ be the experiment resulting from a sampling on X with size n, then $X^{(n+1)} \overset{\pi}{\gtrsim} X^{(n)}$ $\forall\ n \geq 1$.

6.-Let X_1, X_2 and X_3 be three experiments on Ω, and let X_3 be independent of both, X_1 and X_2; then if $X_1 \overset{\pi}{\gtrsim} X_2$ for all prior distributions, then $(X_1,\ X_3) \overset{\pi}{\gtrsim} (X_2,\ X_3)$ for all prior distribution.

7.-Let X_1, X_2, X_3 and X_4 be four experiments on , such that $X_1 \overset{\pi}{\gtrsim} X_2$ and $X_3 \overset{\pi}{\gtrsim} X_4$ for all prior distributions; moreover, assume that X_1 is independent of X_3 and that X_2 is independent of X_4. Then $(X_1,\ X_3) \overset{\pi}{\gtrsim} (X_2,\ X_4)$ for all prior distributions.

8.-Let be $X = \{\mathcal{X};\ A;\ p(x/\theta),\ \theta \in \Omega\}$ and let $(E_i)_{i \in N}$ be a partition of $\mathcal{X}$ form elements of the σ-algebra A. Let us consider the new experiment $Y = \{\mathcal{Y};\beta;\ Q_\theta,\ \theta \in \Omega\}$, being

β the σ-algebra generated by $(E_i)_{i\in N}$ and Q_θ such that $Q_\theta(E_i) = \int_{E_i} p(x/\theta)dx$. Then $X \overset{\pi}{\gtrsim} Y$ with equivalence if, and only if, $p(\theta/x) = p(\theta/E_i)$ a.e. $\forall i \in N$ and $\forall x \in E_i$.

9.- For all statistic $T = T(X^{(n)})$ based on the sample of size n, $X^{(n)} \overset{\pi}{\gtrsim} T$ verifying with equality if, and only if, T is sufficient statistic.

10.-Let S and T be two sufficient statistics of $X^{(n)}$ and $Y^{(m)}$ respectively; then $X^{(m)} \overset{\pi}{\gtrsim} Y^{(n)}$ if, and only if, $S \overset{\pi}{\gtrsim} T$.

11.-Let be $X = \{\mathcal{X} = R, \beta, p(x/\theta), \theta \in \Omega\}$ and $t = t(x)$ a $R \to R$ strictly monotone and differentiable function. Then $X \overset{\pi}{\sim} T$.

12.-Let $X_1,\ldots,X_n$ be experiments on (R, β) and $y_1 = y_1(x_1,\ldots,x_n),\ldots,y_m = y_m(x_1,\ldots,x_n)$ a $R^n \to R^m$ transformation such that the application $(x_{n-m+1},\ldots,x_n) \to (y_1,\ldots,y_m)$ is one -to- one and with continuous first partial derivatives $\forall x_1,\ldots,x_{n-m}$. Then, the compound experiments $X = (X_1,\ldots,X_n)$ and $Y = (X_1,\ldots,X_{n-m}, Y_1,\ldots,Y_m)$ are such that $X \overset{\pi}{\sim} Y$.

13.-Let X and Y be two experiments on Ω, such that X is preferred to Y according to Lehmann's criterion $(X \overset{Lh}{\gtrsim} Y)$. Then, for any prior distribution $p(\theta)$ and for any reference distribution $\pi(\theta)$, $X \overset{\pi}{\gtrsim} Y$.

14.-Let $X = \{\mathcal{X}; A; P_\theta, \theta \in \Omega\}$ and $Y = \{\mathcal{Y}; \beta, Q_\theta, \theta \in \Omega\}$ be two experiments on Ω, such that $\{P_\theta, \theta \in \Omega\}$ is a complete family of distributions. Let X be preferred to Y according to the sufficiency criterion $(X \overset{S}{\gtrsim} Y)$. Then, for any prior distribution $p(\theta)$ and for any reference distribution $\pi(\theta)$, $X \overset{\pi}{\gtrsim} Y$.

15.-Let X and Y be two experiments on $\Omega = \{\theta_1,\ldots, \theta_N\}$. Then if X is preferred to Y according to Blackwells' criterion $(X \overset{B}{\gtrsim} Y)$, then $X \overset{\pi}{\gtrsim} Y$ for any prior distribution $p(\theta)$ and for any reference distribution $\pi(\theta)$.

16.-Let X and Y be two experiments on $\Omega = \{\theta_1,\ldots,\theta_N\}$. Then if X is preferred to Y according to the sufficiency criterion $(X \overset{S}{\gtrsim} Y)$, then $X \overset{\pi}{\gtrsim} Y$ for any prior distribution $p(\theta)$ and for any reference distribution $\pi(\theta)$.

Observation. The definitions of null experiment, compound experiment and marginal experiment, as well as those of the sufficiency, Lehmann's and Blackwell's criteria can be founded in the references cited in the introduction.

4. SOME OBSERVATIONS AND PARTICULAR CASES.

4.1. DIFFERENCES WITH THE LOGARITHMIC UTILITY.

Bernardo /2/ proved that the unique utility function which is proper, local, regular and invariant under one - to - one transformations for inference problems is the logarithmic one which implies a criterion for comparison of experiments based on the maximization of Shannon's information as proposed by Lindley /15/.

There are two fundamental differences between this case and there are stated here:

a) The quadratic utility is not the unique proper, non local and invariant with respect to one - to - one transformations, but it is perhaps the simplest among them (Bernardo /1/).

b) The comparison of experiments based on the logarithmic utility does not depend on the reference distribution $\pi(\theta)$, what implies important simplifications. This is not the case with quadratic utility, where the ordering among experiments depends on the reference distribution. $\pi(\theta)$.

4.2. THE PRIOR AS REFERENCE DISTRIBUTION.

Stating the reference distribution $\pi(\theta)$ is equivalent to defining an origin on the scale of the utilities; the election of $\pi(\theta) = p(\theta)$, the prior distribution, seems reasonable and usefull. Indeed, the condition $u(p(.),\theta) = 0$

means that, if we perform an experiment X and we obtain a result x which does not modify our knowledge about θ i.e. such that $p(\theta/x) = p(\theta)$, then the utility of this result is zero. To put it in other words, the value of the information supplied by irrelevant data is zero.

If $\pi(\theta) = p(\theta)$, then the utility of an experiment X when the prior distribution is $p(\theta)$ reduces to $u(X;\ p(\theta)) = AG_p(X;\ p(\theta)) =$

$$= A\left(\int_X \int_\Omega p(\theta/x)\, p(x/\theta)\, d\theta dx - 1\right).$$

And hence $X \overset{p}{\succsim} Y \iff$

$$\int_X \int_\Omega p(\theta/x) p(x/\theta) d\theta dx \geq \int_Y \int_\Omega p(\theta/y) p(y/\theta) d\theta dy.$$

Even in this case and even with the usual conjugate families, the obtention of $u(x;\ p(\theta))$ is extremely complicated. Even in the normal model, where mathematics are simple, the expression obtained for $u(X;\ p(\theta))$ is too complicated, to be used as an example for intuitive considerations.

4.3. FINITE PARAMETER SPACES

It is particularly interesting to consider Ω finite, so that $\Omega = \{\theta_1, \ldots, \theta_N\}$. In this case the proper local and invariant quadratic utility is

$$u(p(.), \theta_k) = 2p(\theta_k) - \sum_{j=1}^{N} p(\theta_j)^2$$

and, given the experiment $X = \{X = \{x_1, \ldots, x_m\};\ p(x_i/\theta_k),\ \theta_k \in \Omega\}$,

$$u(X;\ p(\theta)) = \sum_{i=1}^{m} p(x_i) \left(\sum_{k=1}^{N} p(\theta_k/x_i)^2 \right) - \sum_{k=1}^{N} p(\theta_k)^2 = GAN\ (X\ ;\ p(\theta))$$

which leads directly to the criterion of maximization of the informational energy gain.

4.4. CONTINUOUS PARAMETER SPACES WITH NON-INVARIANT QUADRATIC UTILITY.

Considering for the continuous case the proper but not invariant quadratic utility function

$$u(p(.), \theta) = A\left(2p(\theta) - \int_\Omega p(\theta)^2\, d\theta\right) + B(\theta),$$

where $A > 0$ is a constant and $B(\theta)$ an arbitrary function of θ, one has

$$u(X;\ p(\theta)) =$$

$$= A\left(\int_X P(x) \int_\Omega p(\theta/x)^2\, d\theta dx - \int_\Omega p(\theta)^2\, d\theta\right) =$$

$$= A\ .\ GAN\ (X\ ;\ p(\theta))$$

which again leads directly to the criterion of maximization of the informational energy gain. This points out an important limitation of such criterion so far as invariance is considered to be an important requirement.

5. DISCUSSION.

In this paper, inference problems are stated as a particular case of decision problems, where the action space consists of probability distributions on the parameter space and explore the use of non-local utilities.

We admit, with the arguments stated by Bernardo /1/, that the utility function must be proper and invariant, and choose among these the simplest which is non-local, i.e. the quadratic utility function. When we apply the general criterion for the maximization of the expected utility we arrive to a criterion for comparison of experiments which in the finite case reproduces the informational energy gain criterion and in the continuous case is tightly connected with this.

A large number of problems remain open. We are presently interested in the asintotic behaviour of the quadratic information supplied by an experiment; in the definition of $G_\pi(\psi/X,\ p(\theta))$ where $\psi = \psi(\theta)$ is in general non-one - to one function of θ in whose knowledge we are interested; in the prior minimal quadratic information distribution in the sense stated by Bernardo /3/ (resulting in the finite case in those which minimize the informational energy); in the application of our criterion to the discrimination among regression models, in the sense stated by Fedorov /9/, Box and Hill /8/ and del Rio /17/; in the sequential design of experiments based on the quadratic information following to De Groot /13/. All those problems become

interesting, once justified from the decision-theoretical point of view that the quadratic information is a sensible utility measure.

6. REFERENCES.

/1/ BERNARDO, J.M.: "The use of Information in the Design and Analysis of Scientific Experimentation". Ph. D. Thesis. University of London. (1975).

/2/ BERNARDO, J.M.: "Expected information as expected utility". Ann. Statist. 7, 686-690. (1979a).

/3/ BERNARDO, J.M.: "Reference postaior distribution for Bayesian inference". J. Roy. Statis. Soc. B., 41, 113-147. (With discussion). (1979b).

/4/ BLACKWELL, D.: "Comparison of experiments". Proc. 2nd Berkeley Symp. Berkeley: University of California Press, 93-102. (1951).

/5/ BLACKWELL, D.: "Equivalent comparisons of experiments". Ann. Math. Statist. 24, 265-272. (1953).

/6/ BLACKWELL, D. y GIRSHICK, M.A.: "Theory of Games and Statistical Decisions". New York, Wiley. (1954).

/7/ BOHNENBLUST, H.F., SHAPLEY, L.S. AND SHERMAN, S.: "Reconnaissance in Game Theory". Research Memorandum RM 208. Santa Mónica, California, Rand Corporation.

/8/ BOX, G.E.P. AND HILL, W.J.: "Discrimination among mechanistic models" .Technometrics, 9, 57-71. (1967).

/9/ FEDOROV, V.V.: "Design of experiments for linear optimality criteria". Theory Prob. Application, 16, 189-195. (1971).

/10/ GARCIA-CARRASCO, M.P.: "Criterios para la comparación de experimentos". Trab. Estadística, 29, 28-51. (1978).

/11/ GARCIA-CARRASCO, M.P.: "Criterio bayesiano para la comparación de experimentos basado en la maximización de la ganancia de energía informacional". Actas XIII Reunión Nacional de Estadística. 2, 65-72. (1982).

/12/ DE GROOT, M.H.: "Optimal Allocation of observations". Ann. Inst. Math. 18. 13-28. (1966).

/13/ DE GROOT, M. H.: "Optimal Statistical Decisions". New York: Mc. Graw-Hill. (1970).

/14/ LEHMANN, E.L.: "Testing Statistical Hypotheses". New York: Wiley.(1959).

/15/ LINDLEY, D.V.: "On a measure of the information provided by an experiment". Ann. Math. Statist., 27, 986-1005. (1956).

/16/ RAIFFA, H. AND SCHLAIFER, R.: "Applied Statistical Decision Theory". Cambridge, Mass: The M.I.T. Press. (1961).

/17/ RIO BUENO, M.: "Algunas cuestiones sobre discriminación en modelos de regresión". Tesis Doctoral. Universidad de Madrid. (1977).