CREATIVE MATH. & INF.
15 (2006), 107 - 112

Some applications of statistical information theory

LEA ŠKROVÁNKOVÁ AND LIBUŠA RÉVÉSZOVÁ

ABSTRACT. The purpose of this paper is to present a short survey of applications in game theory. We will show how the optimization modelling can be used to derive in a unified way classical solution concepts to cooperative n-person games as well as to new solution concepts.

1. INTRODUCTION

During the past years many researchers have examined the application of entropy and generalized entropy functionals in different fields. We mention here works on entropy optimization problems over constraints sets via mathematical programming techniques, applications in a diversity of problems such as traffic engineering, information - communication theory and in many economic and finance models.

Statistical information theory has been developed in the early 1950s. In statistical concepts it is known as the Shannon entropy and the associated Kullback-Leibler divergence measure or relative entropy between probability measures. These probability measures are characterized via the maximum entropy principle. Since many distribution functions cannot be derived by maximizing the classical Shannon entropy, several authors proposed more general concepts of relative entropy. One of them due to Czisar is known in the literature as Φ-divergence.

While the concept of entropy in works of these and other authors was originally restricted to information theory and statistics it was later used in optimization modelling for various problems of management science and engineering.

2. FOUNDATIONS OF INFORMATION THEORY AND PRINCIPLES OF MAXIMUM ENTROPY AND MINIMUM RELATIVE ENTROPY

It is generally known that the amount of information and the amount of uncertainty are inversely related. A classical measure of information introduced by Fisher is known as variance-covariance measure and was developed under assumption of normal distribution.

Consider the problem of estimating a probability distribution given only partial information on the distribution. If the partial information is based on a random sample drawn from a population following the unknown distribution, a Bayesian construction might provide the mode of the posterior distribution as a best estimate. However, when the partial information consists instead of deterministic

Received: 16.10.2006. In revised form: 30.10.2006.
2000 *Mathematics Subject Classification.* 91A35, 91A12, 62P20, 62B10.
Key words and phrases. Maximum entropy principle, minimum relative entropy principle, games theory, cooperative game.

constraints on the known distribution it seems that no one best estimate can be selected from the set of distributions consistent with the information.

Based on the classical Shannon entropy in the 1950 is a **principle of maximum entropy** has been introduced. Its main idea is to choose a probability distribution of maximum entropy $H(q)$ among those consistent with the constraints. This choice is justified with a correspondence principle demonstrating that in the generating distribution were the uniform distribution, then the maximum entropy estimate was the most likely empirical distribution among these consistent with the constraints. This principle has been generalized using KullbackLeibler separator $I(q, p)$ between two distributions q and p to the **principle of minimum relative entropy**. This principle reduces to maximum entropy estimate when p is the uniform distribution.

Relative entropy can be used as a procedure for updating the distribution function of one or more random variables. This procedure has its origin in a related inference rule known as the **principle of maximum entropy**. The rule is applied to estimate a distribution function under given constraints on that function. These constraints present the information available to the decision maker. The distribution that is maximally informative but still satisfies the constraints.

Suppose X assumes values in $X^* = (x_0, x_1, \ldots, x_m)$. Let $P(X = x_i) = p_i$ and $I(x_i)$ be a measure of the information contained in a message that $X = x_i$. If for two experiments on X^* represented by X_1 and X_2, it is true that $I(x_i \wedge xj) = I(x_i)+I(x_j)$ when $P(X_1 = x_i, X_2 = x_j) = p_i \cdot p_j$ and $I(x_i) \geq 0$ for all $i = 0, 1, \ldots, m$, then $I(x_i) = -k \cdot \log p_i$, where k is a constant. For our purposes $I(x_i) = -\ln p_i$. For this discrete distribution, a measure of information or uncertainty is the expected information gained when it is known that $X = x_i$ for $i = 0, 1, \ldots, m$, or

$$H(p) = -\sum_{i=0}^{m} p_i \cdot \ln p_i. \tag{1}$$

The expression (1) is known as the **entropy** of $p = (p_0, \ldots, p_m)$. If the inference problem is to estimate a distribution that corresponds to a convex constraint set, say Λ, then the **principle of maximum entropy** prescribes the $H(p)$ in equation (**??**) subject to $p \in \Lambda$ and sum of p_j equal to 1.

The **principle of minimum relative entropy** is a more general inference procedure. Suppose a decision maker initially believes $P(X = x_i) = p_i$, $i = 0, \ldots, m$. Additional information in the form of a convex constraint set Λ becomes available and $p \notin \Lambda$. The principle of minimum relative entropy requires selecting $q = (q_0, \ldots, q_m$, which minimizes

$$I(q, p) = \sum_{i=0}^{m} q_i \cdot \ln \left(\frac{q_i}{p_i} \right) \tag{2}$$

subject to $q \in \Lambda$ and sum of q_i equal to 1. Clearly, the principle of minimum relative entropy and maximum entropy are equivalent when $p = \dfrac{1}{m+1}$. Relative entropy minimization is the more general procedure that admits an initial or prior distribution.

If there are certain characteristics of the distributions which are presumed known, then these can be incorporated as constraints into the analysis. In such a case as it was shown in [2] and [3] the minimization of $I(q,p)$ subject to these constraints results in a convex programming problem with a related duality theory. This can simplify computations and provide interesting interpretations. The constrained minimum $I^*(q,p)$ is called **minimum discrimination information** (MDI) statistic. It gives the expected amount of information that an observation X yields in favor of the distribution p as opposed to the distribution q. Minimizing $I(q,p)$ for discrimination between probability distributions p and q subject to any constraints results in estimates (say p^*). In many cases these estimates are also maximum likelihood estimates. Moreover, as the sample size increases the asymptotic distribution theory for the MDI value $I^*(q,p)$ leads to a chi squared test of the null hypothesis that p and q are identical.

Thus estimation and hypothesis can be achieved simultaneously. If we suppose a discrete probability distribution $p = (p_1, \ldots, p_m)$, the constrained minimum $I^*(q,p)$, (MDI) can be written as the following convex programming problem:

$$\text{Minimize} \quad I(q,p) = \sum_{i=1}^{m} p_i \cdot \ln\left(\frac{q_i}{p_i}\right) \tag{3}$$

subject to

$$\sum_{i-1}^{m} a_{ij} \cdot p_i = \theta_j, \quad j = 1, \ldots, n,$$

$$\sum_{i=1}^{m} p_i = 1, \quad p_i \geq 0, \quad i = 1, \ldots, m,$$

where probabilities q_i and constraint values θ are constants. A generalization of the relative entropy concept called Φ-divergence has been introduced by Czisar. Now we summarize some of its basic properties. Let be a given differentiable convex function defined on an interval $I \subset \mathbb{R}$, containing the interval $[0,1]$ and such that $\Phi(0) = \Phi(1) = 0$. Let

$$P = \left\{ p \in \mathbb{R}^n : \sum_{i=1}^{m} p_i = 1, \quad p_i \geq 0 \right\}$$

be the set of finite discrete probability measures associated with discrete random variables taking a finite number of values. For $p \in P$, Φ-entropy functional is defined as

$$H_\Phi(p) = -\sum_{i=1}^{m} \Phi(p_i). \tag{4}$$

(4) It can be easily verified that $H_\Phi(p)$ is a continuous function of its n variables and is invariant under permutations. Besides
1. $H_\Phi(p) \geq 0$ with equality if p is degenerate,
2. $H_\Phi(p) \leq -m\,\Phi\left(\frac{1}{m}\right)$ with equality if $p_i = \frac{1}{m}$, $i = 1, \ldots, m$.

Associated with the Φ-entropy H_Φ is the Φ-divergence functional

$$I_\Phi(x,y) = \sum_{i=1}^{m} y_i \, \Phi\left(\frac{x_i}{y_i}\right) \tag{5}$$

where $x, y \in P$ and dom $\Phi = \mathbb{R}^+$.

The concept of Φ-relative entropy has been developed as a generalization of various entropy type functionals widely used in probability theory and statistics. If, for example, we replace kernel $\Phi(\cdot)$ in (5) by $t \log t$ or $(1-t)^2$, we recover Kullback-Leibler discrimination measure or chi squared-divergence. The choice of $I_\Phi(x,y)$ as a measure of distance between two probability distributions is supported by the fact that (5) is well defined and nonnegative, it is equal to zero if and only if $x_i = y_i$ for all i. Moreover I_Φ is convex in each of its arguments. Note, however, that I_Φ is not a distance in the usual sense. Now we turn our attention to some applications of these concepts in game theory.

3. Applications to Solution Concepts of Cooperative Games

First we briefly describe basic notations used in n-person cooperative games. Let (N, v) be a game in characteristic function form, where $N = \{1, 2, \ldots, n\}$ is the set of players and v is normalized (strictly positive) characteristic function given by the payoff vector (imputation) $x = (x_1, \ldots, x_n)$ and $v(S) > 0$ is its value for the coalition S (any subset of N). Let $x(S)$ be the sum of the payoffs x_i to players i in coalition S, $w_i > 0$ be the weight associated with player i, $v_i = v(\{i\})$ and $x(N) = v(N)$ be the sum of payoffs x_i to all players $i \in N$, the grand coalition value.

Solutions of n-person cooperative games are in general given as a vector of payoffs to the individual players. As is known several solution concepts to cooperative games have been proposed in literature. There are also various ways of determination of such solutions. Besides other methods some of these solution concepts can be determined by solving a corresponding mathematical programming problem. For example, to determine the nucleolus of a characteristic function game Schmeidle used a finite sequence minimization problems. Charnes showed that both core and the Shapley value as solution concepts of cooperative games are special convex nucleus solutions.

Later the authors of [1] for n-person game solution studied a special class of problems in which the functional to be minimized is based on Φ-divergence. Such an approach allows to derive in a unified way classical solution concepts to cooperative games as well as to generate new solution concepts.

If we assume that the used Φ-functionals have properties described above, the problem of determining solution concepts to a cooperative game is to find an x solving

$$\min_{x \in I} = \sum_{i \in N} w_i \, v_i \, \Phi\left(\frac{x_i}{v_i}\right) \tag{6}$$

subject to

$$A\,x = b,$$
$$x(N) = v(N),$$

where $I = \operatorname{dom}\Phi$, A is an $m \times n$ matrix and equations $A\,x = b$ represent conditions on x corresponding to the particular solution concept. Some solution concepts can be directly derived by setting a concrete form of the kernel function Φ in (6). For example, if we set $\Phi(t) = (1-t)^2$, $w_i = q_i$ for all $i \in N$ and assume no specific constraints on x (i. e. $A = 0$, $b = 0$), then the problem becomes

$$\min\left\{\sum (q_i - x_i)^2 : \ x(N) = v(N)\right\}.$$

In this case $q(S)$ is the **homomollifier** of the characteristic function $v(S)$ and $q_i = q(\{i\})$.

Some solution concepts to cooperative games are given by extremal principles that apply to the $(2^n - 1)$-dimensional space of the characteristic function vector $v(S)$. A number of formulations given by problem (6) can be applied in such instances. One such a modified formulation can be written as

$$\min = \sum_{S \subset N} w(S)\, v(S)\, \Phi\left(\frac{x(S)}{v(S)}\right) \tag{7}$$

subject to $A\,x = 0$, $B\,x \geq b$, $x(N) = v(N)$.

Here A is an $(m-n) \times m$ matrix and $A\,x = 0$ defines the relations

$$x(S) = \sum_{i \in S} x_i,$$

where $m = 2^n - 1$ is the number of coalitions in an m-person game. Choosing, for example, $\Phi(t) = t\,\ln\frac{t}{e}$, $B = 0$, $b = 0$ in problem (7) the **weighted entropic solution** can be obtained (see [5], [6], [7], [8], [9]).

Finally we consider the Shapley value as a solution concept of cooperative games. Besides the classical formula introduced by Shapley (see [4]) this value can be determined as the solution of the following convex programming problem

$$\min = \left\{\sum_{S \subset N} [v(S) - x(S)]^2\, w(S) : \ x(N) = v(N)\right\} \tag{8}$$

where $w(S) = \binom{n-2}{s-1}^{-1}$ for all $S \subset N$, $S \neq N$ and s denotes the number of players in coalition S.

Turning to the Φ-divergence, it is easy to show that the problem (8) is a special case of the general formulation (7), if we set $\Phi(t) = (1-t)^2$, $B = 0$, $b = 0$ and

$$w(S) = \binom{n-2}{s-1}^{-1} v(S), \quad \text{for all } S \subset N,$$

$$w(S) = 1, \ \text{ if } S = N.$$

It is clear that one can derive other solution concepts to n-person cooperative games by picking different kernel functions Φ.

REFERENCES

[1] Charnes A., Golany B., *On N-person game solutions and convex programs with essentially unconstrained duals*, The UT, Austin, 1989

[2] Charnes A., Cooper W., Seiford L., *Extremal principles and optimization dualities for Khinchin-Kullback-Leibler estimation*, The UT, Austin, 1976

[3] Charnes A., Rousseau J., Seiford L., *Complements, mollifiers and propensity to disrupt*, International Journal of Game Theory 7 (1976)

[4] Chobot M., Turnovec F., Ulašin V., *The game theory*, Alfa, Bratislava, 1991

[5] Chobot M., Škrovánková L., *Use of information theory in marketing models*, Proceedings of The 6th International conference Quantitative methods in bussiness and management, University of Economics, Bratislava, 1999

[6] Cover T., Thomas J. A., *Elements of information theory*, John Wiley and sons, New York, 1999

[7] Peller F., Škrovánková P., *General theory of measurement in marketing research*, Proceedings of The 11^{th} International Scientific Symposium Strategic management and decision support systems in Strategic management, Ekonomski fakultet, Subotica, 2006

[8] Šoltés V., Šoltés M., *Analysis of duration and convexity of coupon obligation*, Creative Math. 14 (2005), 83-86

[9] Šoltés V., Šoltés M., *Analysis of two-asset portfolio*, E+M Economics and Management, Technical University of Liberec, Special Issue, 63-65, 2003

LEA ŠKROVÁNKOVÁ
FACULTY OF ECONOMIC INFORMATICS
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ECONOMICS
DOLNOZEMSKÁ CESTA 1
852 35 BRATISLAVA, SLOVAKIA
E-mail address: `leaskrov@post.sk`

LIBUŠA RÉVÉSZOVÁ
FACULTY OF ECONOMICS
DEPARTMENT OF APPLIED MATHEMATICS AND BUSINESS INFORMATICS
TECHNICAL UNIVERSITY OF KOŠICE
LETNÁ 9
042 00 KOŠICE, SLOVAKIA
E-mail address: `libusa.reveszova@tuke.sk`