ON EXTREMAL MARTINGALE DISTRIBUTIONS

LESTER E. DUBINS and GIDEON SCHWARZ
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

The question

Q(1)   *How does one construct all martingales X_t?*

is one that we have found quite stimulating. Though we have had little success in answering this somewhat vague query, it does not seem inappropriate to call attention to it and to several related questions.

Since, for each subset T of the real line, the set $\mathfrak{M}_T$ of all distributions of martingales X_t for $t \in T$ is convex, the following related questions suggest themselves:

Q(2)   *How does one characterize the extreme points of $\mathfrak{M}_T$?*

Q(3)   *How does one construct all extreme points of $\mathfrak{M}_T$?*

Q(4)   *Is every element of $\mathfrak{M}_T$ a mixture of the extreme ones?*

However, though every mixture of a finite number of elements of $\mathfrak{M}_T$ is an element of $\mathfrak{M}_T$, such is not the case for more general mixing. This is so simply because there are sequences $\theta_1, \theta_2, \cdots$ of probability measures of mean 0 on the real line and positive numbers $\alpha_1, \alpha_2, \cdots$ whose sum is 1 such that $\sum \alpha_i \theta_i$ has no mean. It seems, therefore, that in studying Q(4), and possibly some of the other queries, it may be desirable to shift attention from the set of martingales to a closely related set.

The difficulty in answering the queries above varies considerably with the parameter set T. If T is finite or has the order structure of the positive integers, the queries are easy to answer. The martingales $X_1, X_2, \cdots$ whose distributions are extremal are merely those that possess these two properties:

   (i) X_1 is a constant;
and
   (ii) the conditional distribution of each X_n given the past up to time $n - 1$ is almost surely a two-valued distribution.

Moreover, as is easily verified, every martingale distribution parametrized by the positive integers is a mixture of the extremal ones. Thus, if T is the set of

positive integers, there seem to be satisfactory answers to the four queries above. Moreover, the situation seems to be fairly similar if T is any well-ordered subset of the real line.

However, if T is not well-ordered, our knowledge is far less complete. To take the simplest not well-ordered set, the negative integers, though an answer to Q(2) can be given, and though it may not be difficult to settle Q(4), we are completely in the dark with regard to Q(1) and Q(3).

The answer to Q(2) is here formulated as a formal proposition which will be used and proved later.

PROPOSITION 1. *The distribution of a martingale* X_n, *as* n *ranges over the negative integers, or all the integers, is extremal among all such distributions if and only if* (ii) *obtains, and every event in the tail* σ-*field has probability* 0 *or* 1.

(An event E is in the *tail* σ-field if and only if, for every integer k, E is in the σ-field generated by the X_n for $n \leq k$.)

When T is the nonnegative real line we do not know even how to characterize the extremal martingale distributions. The situation is perhaps more tractable if attention is restricted to martingales with continuous paths, for, in addition to other advantages, the result in [1] and [2], (and in unpublished work of Itô and Watanabe) would seem to be applicable.

2. Symmetric martingales. A counterexample

If the conditional distribution of each increment $X_{n+1} - X_n$ given the past is symmetric about the origin, then the martingale X_n is *symmetric*. This definition is applicable whether n ranges over the positive, the negative, or all, integers.

The position of the martingales with continuous paths among all real-parameter martingales seems similar to that of the symmetric martingales among all martingales with discrete time-parameter. But the only point in introducing the symmetric martingales here is to point out, by means of an example, a distinction between symmetric martingales based on the positive integers and those based on the negative integers. For the purposes of this paper, the requirement that the increments $X_{n+1} - X_n$ have a mean and the requirement that the increments be summable in n are irrelevant and will therefore be dropped. Therefore, of interest here are the *symmetric processes* $\{D_n\}$, that is, the processes such that for each n, the conditional distribution of D_n given the past is symmetric.

Plainly, a distribution of a real-valued random variable D is symmetric if and only if it is the distribution of a product of a nonnegative random variable s with an independent random variable b that assumes the values 1 and -1 with probability $\frac{1}{2}$ each. (For s one can always choose $|D|$; unless D is 0 with positive probability, $D/|D|$ can be chosen for b.) Likewise, as is analogous to the theorem in [2], any symmetric process can be similarly factored into two processes, one of which is nonnegative real-valued, and the other is a fair-coin process. For simplicity, assume henceforth, for all processes $\{D_n\}$ to be considered, that

$|D_n| > 0$ almost surely for all n. To $\{D_n\}$ associate the process $b(\{D_n\}) = \{b_n\}$ where $b_n = 1$ or -1, according as $D_n > 0$ or $D_n < 0$. Plainly, the convex set $\sum$ of distributions of symmetric processes is a subset of the convex set $\mathfrak{B}$ of all distributions of stochastic processes $\{D_n\}$ such that $\{b_n\}$ is the fair-coin distribution (that is, the b_n are independent and $b_n = \pm 1$ with probability $\frac{1}{2}$ each).

If $\mu \in \mathfrak{B}$ and, under μ, the conditional distribution of the sequence $\cdots, D_n, \cdots$ given the sequence $\cdots, b_n, \cdots$ is degenerate or, equivalently, if each D_n is a function of the sequence $\cdots, b_n, \cdots$, then μ is *pure*. As is not difficult to verify, μ is an extreme point of $\mathfrak{B}$ if and only if μ is pure. Plainly then, if μ is pure and $\mu \in \Sigma$, then μ is an extreme point of Σ. As is also not difficult to verify, if the time-parameter set T is the set of positive integers, then every symmetric process is a mixture of the pure ones; so a process is extremal if and only if it is pure.

An example will now be given which shows that *when T is the set of negative integers, there is a process that is extremal among the symmetric ones and yet is not pure.*

Let a and b be two distinct positive numbers and define the process thus. Let D_{-1} be a, $-a$, b, and $-b$ with probabilities $\frac{1}{4}$ each. Given $D_{-1}, D_{-2}, \cdots, D_n$, the conditional distribution of D_{n-1} puts weight $\frac{1}{2}$ on each of the values a and $-b$ when $|D_n| = a$, and weight $\frac{1}{2}$ on each of the values $-a$ and b when $|D_n| = b$.

Thus defined, $\{D_n, n = \cdots, -2, -1\}$ is a stationary Markov process, and its forward transition probabilities (which are easily obtained from the stationary distribution and the given backward transition probabilities) are: the conditional distribution of D_{n+1} given $\cdots, D_{n-1}, D_n$ puts weight $\frac{1}{2}$ on each of $\pm a$ when D_n is a or $-b$, and weight $\frac{1}{2}$ on each of $\pm b$ when D_n is $-a$ or b. Clearly, the process is symmetric, and has property (ii). Furthermore, since the variables D_m preceding any D_n are independent—not just conditionally independent—of the D_j following D_n, all events in the tail field have probability 0 or 1. As will be evident from the proof of proposition 1, the two conditions given there also characterize the extreme points of Σ. Therefore, the distribution of $\{D_n\}$ is an extreme point of Σ. However, $\{D_n\}$ is not pure: the distribution of $\{|D_n|\}$ given $\{b_n\}$ gives weight $\frac{1}{2}$ each to two sequences, one of which ends with $|D_{-1}| = a$, and the other with $|D_{-1}| = b$.

3. Proof of proposition 1

Throughout this section, the parameter space is the set of all integers, or of the negative integers.

(a) Let α be the distribution of a martingale $\{X_n\}$. Let A and A^c be two complementary events in the tail field, such that $\alpha(A) > 0$ and $\alpha(A^c) > 0$. The conditional distributions of the process, given A and given A^c, are easily seen to be martingales as well, and their average, weighted by $\alpha(A)$ and $\alpha(A^c)$, is α.

(b) Let β be the distribution of a martingale such that for some m, the event $B = \{$The conditional distribution of X_m given the past is not two-valued$\}$ has positive probability. As is well known, a distribution with finite mean that is not

two-valued, can be expressed as the average of two distinct distributions with the same mean, and as is not difficult to verify, the two can be chosen to depend measurably on the given distribution. Clearly, such a decomposition yields a decomposition of β as an average of two martingale distributions, each defined by adopting β off B, while on B, the joint distribution of $\{X_n, n < m\}$ as well as the conditional distribution of $\{X_n, n > m\}$ given $\{X_n, n \leq m\}$ are as under β, and the conditional distribution of X_m given $\{X_n, n < m\}$ is one of the two components of the decomposition.

The "only if" part of proposition 1 follows from (a) and (b).

(c) Let γ_1 and γ_2 be two distributions of $\{X_n\}$, such that for some event C in the tail field, $\gamma_1(C) \neq \gamma_2(C)$. Then, for $\gamma = \frac{1}{2}(\gamma_1 + \gamma_2)$, $0 \neq \gamma(C) \neq 1$.

(d) Let δ_1 and δ_2 be two martingale distributions of $\{X_n\}$, and put $\delta = \frac{1}{2}(\delta_1 + \delta_2)$. If, for every $n \in T$, the conditional δ-distribution of X_n, given the past, is two-valued almost surely, then the conditional distributions of X_n under δ_1 and δ_2 are supported by the same two values, and since, furthermore, the mean of both conditional distributions of X_n is X_{n-1}, they are equal to each other.

(e) As an application of martingale convergence shows, two distributions of $\{X_n\}$ that have the same projections on the tail field and the same conditional distributions of X_n given the past are identical.

The "if" part of proposition 1 follows from (c), (d), and (e).

4. Two consequences of the counterexample

(a) *Continuous martingales.* Let the parameter space T be the set of non-negative reals. As [1] and [2] establish, every continuous martingale $\{X_t\}$ can be transformed into standard Brownian motion by a path-dependent transformation of the time scale. If this transformation almost surely maps distinct paths of $\{X_t\}$ into distinct paths of the Brownian motion, the martingale is *pure*, and its distribution is an extreme point of the set of all distributions of continuous martingales. *There is, however, a martingale which is not pure, and its distribution is extremal among all distributions of continuous martingales.* The following is an example.

Let $B(t, \omega)$ be a standard Brownian motion process and let $\cdots$, $D_{-2}(\omega)$, $D_{-1}(\omega)$ have the distribution of the counterexample in section 2. Furthermore, the following two additional requirements on the joint distribution of $\{D_n\}$ and $\{B(t)\}$ are compatible with their distributions, and suffice to specify their joint distribution: (i) for $n = \cdots$, -2, -1, D_n and the Brownian motion increment $B(|n|^{-1}) - B((|n| + 1)^{-1})$ have the same sign; (ii) $|D_{-1}|$ is independent of $\{B(t)\}$. Now let $\tau(\cdot, \omega)$ be the unique continuous function whose value at 0 is 0, whose derivative in the open interval $((|n| + 1)^{-1}, |n|^{-1})$ is $|D_n|$, and whose derivative in $(1, \infty)$ is 1. Let $\theta(\cdot, \omega)$ be the inverse function of $\tau(\cdot, \omega)$, and define $Y(t, \omega) = B(\theta(t, \omega), \omega)$. As is now easily verified, $\{Y(t, \cdot)\}$ is a continuous martingale whose distribution is extremal, yet $\{Y(t, \cdot)\}$ is not pure.

(b) *Randomized strategies.* In the theory of games, where moves are made

only at time $t = 1, 2, \cdots$, randomization can be achieved in one of two ways: (i) by letting moves at certain times depend not only on the information available at that time to the player, but also on a random device; (ii) by "mixing pure strategies," that is, choosing at random a pure strategy for the entire game. When the available information does not decrease in time, the two methods of randomization are equivalent in the sense that they induce the same set of distributions of the history of the game [3], [4], [5]. (Incidentally, [5] contains the essential ideas for answering questions (1) through (4) for discrete well-ordered time.) But, as shown in the next paragraph, this equivalence does not generalize to games with continuous time.

Consider a player in a game. The information available to him at time $t > 0$ consists of his past moves and the sequence of values of $\{b_i, i < -t^{-1}\}$, where $\{b_i\}$ is a fair-coin tossing process. His moves occur at times $t_n = -n^{-1}, n = \cdots, -2, -1$, and consist of choosing one of the positive numbers a or b. Let d_n be the number chosen at time t_n. One way of choosing is to let the products $\{b_i\, d_i\}$ be distributed as $\{D_i\}$ in the counterexample. This is a randomized strategy, in that, given the b_i for $i < -t^{-1}$, the totality of moves made before time t is independent of the future information, that is, of the b_i for $i \geq -t^{-1}$. However, this distribution of $\{D_n\}$ cannot be achieved by mixing "pure" strategies, if the latter are defined as strategies that determine the moves of the player in the time interval $[0, t]$ as a function of his information history during that time interval.

REFERENCES

[1] K. E. DAMBIS, "On decomposition of continuous sub-martingales," *Teor. Verojatnost. i Primenen.*, Vol. 10 (1965), pp. 438–448. (In Russian; English summary.)

[2] L. E. DUBINS and G. SCHWARZ, "On continuous martingales," *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 53 (1965), pp. 913–916.

[3] H. W. KUHN, "Extensive games," *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 36 (1950), pp. 570–576.

[4] ———, "Extensive games and the problem of information," *Contributions to the Theory of Games*, Vol. 2, pp. 193–216, *Annals of Mathematical Studies*, No. 28, Princeton, Princeton University Press, 1953.

[5] A. WALD and J. WOLFOWITZ, "Two methods of randomization in statistics and the theory of games," *Ann. of Math.*, Vol. 53 (1951), pp. 581–586.