

TRANSITIVITY AND APPROXIMATE TRANSITIVITY IN PROBLEMS OF OPTIMAL STOPPING

A. IRLE

*Institute of Mathematical Statistics, University of Münster,
Münster, G.F.R.*

In a sequential decision problem it is usually assumed that the available information is represented by an increasing family \mathcal{F} of σ -algebras. Often a reduction, e.g., according to principles of sufficiency or invariance, is performed which yields a smaller family \mathcal{G} . The consequences of such a reduction for problems of optimal stopping are investigated in this paper.

It is shown that \mathcal{G} is transitive for \mathcal{F} (in the Bahadur sense) if and only if for any stochastic process adapted to \mathcal{G} the value (i.e., maximal reward by optimal stopping) under \mathcal{G} and the value under \mathcal{F} are equal.

Then a numerical quantity which describes the possible reduction in value is introduced and some of its properties are investigated. In this context, the relationship between transitivity and invariance is considered, and an example is treated in some detail.

1. Introduction and basic concepts

When we consider a sequential decision problem we usually assume that the amount of available information is increasing with time. We then represent the possible data which we can obtain up to the time $t \in T$ by a σ -algebra \mathcal{F}_t , and thus come to the formal requirement " $\mathcal{F}_s \subset \mathcal{F}_t$ for $s < t$ ". But it is often hard to store and handle all the data as represented by $\mathcal{F} = (\mathcal{F}_t)_{t \in T}$ since their actual amount may be very large, so we want to apply a reduction procedure which leads to a smaller family $\mathcal{G} = (\mathcal{G}_t)_{t \in T}$ of σ -algebras. Such a reduction should of course take into account the underlying statistical structure and may, e.g., be accomplished according to the principles of sufficiency or invariance. The consequences of such reductions for non-sequential decision problems have been thoroughly investigated, starting with the concept of sufficiency already appearing in Fisher (1925) and leading to

the theory of comparison of experiments (see the monograph by LeCam (1974) which also contains historical remarks).

The case of reduction in sequential decision problems, when observing a sequence of random variables, was treated by Bahadur (1954) who introduced the notion of transitivity. The model as considered by Bahadur consists of the following:

- (Ω, \mathcal{A}) — the sample space;
- $T = N$ — the set of time parameters;
- $\mathcal{F} = (\mathcal{F}_n)_{n \in T}$ — an increasing family of sub- σ -algebras of \mathcal{A} where \mathcal{F}_n consists of the events which may be observed up to the time n ;
- $(P_\vartheta)_{\vartheta \in \Theta}$ — a family of probability measures on (Ω, \mathcal{A}) representing the possible states of nature;
- $(D_n, \mathcal{D}_n)_{n \in T}$ — a family of measurable spaces where D_n is the space of terminal decisions at the time n .

A decision rule of the statistician w.r.t. \mathcal{F} consists of a randomized stopping time N w.r.t. \mathcal{F} and a terminal decision function $v = (v_n)_{n \in T}$ w.r.t. \mathcal{F} where each v_n is a transition probability from (Ω, \mathcal{F}_n) to (D_n, \mathcal{D}_n) .

Suppose now that for every $n \in T$ there is given a sufficient σ -algebra $\mathcal{G}_n \subset \mathcal{F}_n$, then — as shown by Bahadur (1954) — for every decision rule (N, v) w.r.t. \mathcal{F} there exists a terminal decision function v' w.r.t. \mathcal{G} such that (N, v) and (N, v') are equivalent, i.e., for all $\vartheta \in \Theta$ and $n \in T$, $C \in \mathcal{D}_n$, $P_\vartheta(N = n, v_n \in C) = P_\vartheta(N = n, v'_n \in C)$; but in general we may not replace N by a randomized stopping time w.r.t. \mathcal{G} . To answer the question when such a replacement is possible, Bahadur introduced the notion of transitivity: \mathcal{G} — which in problems of reduction by sufficiency often does not form an increasing family — is called *transitive* (w.r.t. \mathcal{F} and $(P_\vartheta)_{\vartheta \in \Theta}$) iff for all $n \in T$, $G_{n+1} \in \mathcal{G}_{n+1}$ and $\vartheta \in \Theta$

$$P_\vartheta(G_{n+1} | \mathcal{G}_n) = P_\vartheta(G_{n+1} | \mathcal{F}_n).$$

It is then shown by Bahadur that for every decision rule (N, v) w.r.t. \mathcal{F} there exists an equivalent decision rule (N', v') w.r.t. \mathcal{G} if \mathcal{G} is sufficient and transitive. The case that Θ contains only one element is trivial from the point of view of decision theory, and in this case the family \mathcal{O} where $\mathcal{O}_n = \{\emptyset, \Omega\}$ for all $n \in T$ is obviously sufficient and transitive. But from the point of view of optimal stopping there seems to emerge a problem of some significance.

Let us assume that all the available information is represented by a family $\mathcal{F} = (\mathcal{F}_t)_{t \in T}$ and that a data reduction has led to a smaller family $\mathcal{G} = (\mathcal{G}_t)_{t \in T}$, and furthermore that we are given a probability measure on the underlying sample space and a real-valued stochastic process $X = (X_t)_{t \in T}$ such that each X_t is \mathcal{G}_t -measurable (the set of all processes with this property

is denoted by $M(\mathcal{G})$ from now on). The problem of optimal stopping for X w.r.t. \mathcal{G} , respectively \mathcal{F} , then is to maximize the reward EX_τ in a certain set of "reasonable" stopping times τ w.r.t. \mathcal{G} , respectively \mathcal{F} , i.e., find the supremum $v(X, \mathcal{G})$, respectively $v(X, \mathcal{F})$, of possible rewards and look for optimal stopping times. It will usually be easier to determine $v(X, \mathcal{G})$ since there are less stopping times w.r.t. the smaller family \mathcal{G} , and so the problem arises to characterize those families \mathcal{G}, \mathcal{F} for which $v(X, \mathcal{G})$ and $v(X, \mathcal{F})$ are equal and, more generally, to investigate the possible difference between $v(X, \mathcal{G})$ and $v(X, \mathcal{F})$.

For a formalization of these considerations let us introduce some notions and notations which will be used from now on.

Let (Ω, \mathcal{A}, P) be the basic probability space, $T \neq \emptyset$ an ordered set — the set of time parameters. Furthermore let

$$\begin{aligned} g(T) &= \{\mathcal{G} = (\mathcal{G}_t)_{t \in T} : \mathcal{G}_t \text{ sub-}\sigma\text{-algebra of } \mathcal{A}\}, \\ m(T) &= \{\mathcal{G} = (\mathcal{G}_t)_{t \in T} : \mathcal{G} \in g(T), \mathcal{G}_s \subset \mathcal{G}_t \text{ for } s < t\}. \end{aligned}$$

For $\mathcal{G} \in g(T)$ define $\mathcal{G}^* \in m(T)$ by $\mathcal{G}_t^* = \sigma(\bigcup_{s \leq t} \mathcal{G}_s)$, thus $\mathcal{G}_t \subset \mathcal{G}_t^*$. Let there be given a point ∞ such that $\infty \notin T$ and $\infty > t$ for all $t \in T$ and set $T^* = T \cup \{\infty\}$ and for $\mathcal{G} \in g(T)$ $\mathcal{G}_\infty = \sigma(\bigcup_{t \in T} \mathcal{G}_t)$.

DEFINITION (1.1). A mapping $\tau : \Omega \rightarrow T^*$ is called a *stopping rule* w.r.t. $\mathcal{G} \in g(T)$ iff for all $t \in T$ the following holds:

- (i) There exists $D_t \in \mathcal{G}_t$ such that

$$\{\tau = t\} = \{\tau \geq t\} \cap D_t.$$

- (ii) $\{\tau \geq t\} \in \mathcal{G}_t^*$.

A stopping rule τ is called a *stopping time* iff $P(\{\tau = \infty\}) = 0$ holds.

The set of all stopping rules (stopping times) w.r.t. \mathcal{G} is denoted by $r(\mathcal{G})$ ($s(\mathcal{G})$).

Condition (i) has the following interpretation: Having observed up to the time t the decision to stop the observation at this time is only to be based on events in \mathcal{G}_t .

Remark (1.2). (a) If \mathcal{G} is an increasing family, i.e., $\mathcal{G}_t = \mathcal{G}_t^*$ for all $t \in T$ then the conditions (i) and (ii) are equivalent to " $\{\tau \leq t\} \in \mathcal{G}_t$ and $\{\tau = t\} \in \mathcal{G}_t$ " and thus yield the usual notion.

(b) If T is a totally ordered set, then for $\tau_1, \tau_2 \in r(\mathcal{G})$ also $\min\{\tau_1, \tau_2\} \in r(\mathcal{G})$, since

$$\{\min\{\tau_1, \tau_2\} = t\} = (\{\tau_1 = t\} \cup \{\tau_2 = t\}) \cap \{\min\{\tau_1, \tau_2\} \geq t\},$$

but in general $\max\{\tau_1, \tau_2\}$ does not yield an element of $r(\mathcal{G})$.

We now introduce some more notations:

For $\mathcal{G} \in g(T)$ let $M(\mathcal{G})$ denote the set of all real-valued stochastic processes $X = (X_t)_{t \in T}$ on (Ω, \mathcal{A}, P) such that X_t is \mathcal{G}_t -measurable, and $M_1(\mathcal{G})$ all $X \in M(\mathcal{G})$ with $X_t(\Omega) \subset [0; 1]$ P -a.s. for all $t \in T$. We set

$$M = \bigcup \{M(\mathcal{G}): \mathcal{G} \in g(T)\}, \quad M_1 = \bigcup \{M_1(\mathcal{G}): \mathcal{G} \in g(T)\}$$

and call $X \in M$ *integrable* iff EX_t is finite for all $t \in T$.

For two random variables f, g we write $f \leq g$ ($f = g$) iff $P(\{f > g\}) = 0$ ($P(\{f \neq g\}) = 0$), and for $X, Y \in M$ we write $X \leq Y$ ($X = Y$) iff $X_t \leq Y_t$ ($X_t = Y_t$) for all $t \in T$.

We proceed to give some examples for stopping rules.

EXAMPLE (1.3). (a) Let $T = \mathbb{N}$, $\mathcal{G} \in g(T)$ and $X \in M(\mathcal{G})$. For any family $(E_t)_{t \in T}$ of Borel sets of the real line

$$\tau = \inf \{t \in T: X_t \in E_t\}$$

belongs to $r(\mathcal{G})$ (where by definition $\inf \emptyset = \infty$).

(b) Let $T = [0; \infty)$, $\mathcal{G} \in g(T)$, $X \in M(\mathcal{G})$ and E an analytic subset of the real line. Suppose that for every $t \in T$ the process $(X_s)_{s \in [t; \infty)}$ is progressively measurable w.r.t. $(\sigma(\bigcup_{t \leq r \leq s} \mathcal{G}_r))_{s \in [t; \infty)}$, which is, e.g., fulfilled if X has right-continuous paths, and that $\sigma(\bigcup_{t \leq r \leq s} \mathcal{G}_r)$ is P -complete for every $s \geq t$.

If we set $\mathcal{G}_t^+ = \bigcap_{\varepsilon > 0} \sigma(\bigcup_{t \leq s \leq t+\varepsilon} \mathcal{G}_s)$ then

$$\tau = \inf \{t \in T: X_t \in E\}$$

belongs to $r(\mathcal{G}^+)$.

For $X \in M$, $\mathcal{G} \in g(T)$ and $\tau \in r(\mathcal{G})$ we define a mapping X_τ from Ω to the real line by

$$X_\tau(\omega) = I_{\{\tau < \infty\}}(\omega) X_{\tau(\omega)}(\omega).$$

If τ takes only countably many values then X_τ is obviously measurable.

The problem to maximize EX_τ in a certain set of "reasonable" stopping times w.r.t. \mathcal{G} is called *the problem of optimal stopping for X w.r.t. \mathcal{G}* .

For the following formal definition we denote by $d(\mathcal{G})$ the set of all stopping times w.r.t. \mathcal{G} such that $\tau(\Omega) \cap T$ is order-isomorphic to a subset of the natural numbers, i.e.,

$$d(\mathcal{G}) = \{\tau \in s(\mathcal{G}): \tau(\Omega) \cap T \text{ order-isomorphic to } N' \subset \mathbb{N}\}.$$

DEFINITION (1.4). For $X \in M$, $\mathcal{G} \in g(T)$ and $t \in T^*$ let

$$v(X, \mathcal{G}, t) = \sup \{EX_\tau: \tau \in d(\mathcal{G}), \tau \leq t, EX_\tau^- < \infty\}$$

($\sup \emptyset = -\infty$).

$v(X, \mathcal{G}) = v(X, \mathcal{G}, \infty)$ is called *value of X w.r.t. \mathcal{G}* .

The quantity $v(X, \mathcal{G}, t)$ thus gives the supremum which can be achieved in the problem of optimal stopping for X w.r.t. certain "well-behaved" stopping times in $s(\mathcal{G})$ and obviously depends on \mathcal{G} in general.

Remark (1.5). Suppose $\mathcal{G}^1, \mathcal{G}^2 \in g(T)$ such that for every $t \in T$ $\mathcal{G}_t^1 \subset \mathcal{G}_t^2$ P -a.s. holds, i.e., for any $A^1 \in \mathcal{G}_t^1$ exists $A^2 \in \mathcal{G}_t^2$ with $P(A^1 \triangle A^2) = 0$; from now on this will be denoted by $\mathcal{G}^1 \leq \mathcal{G}^2$.

If τ_1 belongs to $d(\mathcal{G}^1)$ then obviously there exists $\tau_2 \in d(\mathcal{G}^2)$ with $P(\{\tau_1 \neq \tau_2\}) = 0$, i.e., according to the convention following (1.2) $\tau_1 = \tau_2$. This implies for every $X \in M$, $t \in T$,

$$v(X, \mathcal{G}^1, t) \leq v(X, \mathcal{G}^2, t).$$

If we define an equivalence " \sim " on $g(T)$ by

$$"\mathcal{G}^1 \sim \mathcal{G}^2" \quad \text{iff} \quad "\mathcal{G}^1 \leq \mathcal{G}^2 \text{ and } \mathcal{G}^2 \leq \mathcal{G}^1",$$

then $v(X, \cdot, t)$ is constant on each equivalence class.

To obtain a unified definition for arbitrary time sets T we only used stopping times in $d(\mathcal{G})$ for the definition of the value. This is certainly no restriction if already $T \subset N$ holds, but also in other interesting cases $v(X, \mathcal{G}, t)$ is an upper bound of EX_τ for arbitrary $\tau \in s(\mathcal{G})$, $\tau \leq t$. To illustrate this fact we will treat the case $T = [0; \infty)$ – commonly called the *continuous parameter case*. For this the following definitions will be useful:

(1.6) For a measurable process $X \in M$ and $\mathcal{G} \in m(T)$ we call X \mathcal{G} -*bounded from below* iff there exists a random variable h with finite expectation such that for all $\tau \in s(\mathcal{G})$

$$X_\tau \geq E(h|\mathcal{G}_\tau)$$

holds. X is called \mathcal{G} -*bounded* iff X and $-X$ are \mathcal{G} -bounded from below.

X is called \mathcal{G} -*lower semi-continuous from the right* iff for every $\tau \in s(\mathcal{G})$ and every decreasing sequence $(\tau_n)_n$ in $s(\mathcal{G})$ with $\lim_n \tau_n = \tau$

$$X_\tau \leq \liminf_n X_{\tau_n}$$

holds.

THEOREM (1.7). Let $T = [0; \infty)$, $X \in M$ a measurable process.

Suppose that for $\mathcal{G} \in m(T)$ X is \mathcal{G} -bounded from below and \mathcal{G} -lower semi-continuous from the right.

Then for any $\tau \in s(\mathcal{G})$ $EX_\tau^- < \infty$ and for $t \in T^*$ with $\tau \leq t$

$$EX_\tau \leq v(X, \mathcal{G}, t).$$

The proof is omitted; compare Thompson (1971), p. 310.

2. Bahadur transitivity

For the following let (Ω, \mathcal{A}, P) be the basic probability space, $T \neq \emptyset$ an ordered set.

DEFINITION (2.1). For $\mathcal{G}, \mathcal{F} \in g(T)$, $s, t \in T$ with $s < t$ set

$$b_t(\mathcal{G}_s, \mathcal{F}_s) = \sup \{ |E[E(f|\mathcal{G}_s) - E(f|\mathcal{F}_s)]| : f \text{ } \mathcal{G}_t\text{-measurable, } 0 \leq f \leq 1 \}$$

\mathcal{G} is called *Bahadur transitive* for \mathcal{F} iff $\mathcal{G} \leq \mathcal{F}$ and for all $s, t \in T$ with $s < t$

$$b_t(\mathcal{G}_s, \mathcal{F}_s) = 0.$$

In the following we will write *B-transitive* for Bahadur transitive.

Remark (2.2). (a) $\mathcal{G} \leq \mathcal{F}$ is *B-transitive* for \mathcal{F} iff for all $s, t \in T$ with $s < t$ and for all $G_t \in \mathcal{G}_t$, $P(G_t|\mathcal{G}_s) = P(G_t|\mathcal{F}_s)$.

(b) Part (a) is strengthened by the following:

$$b_t(\mathcal{G}_s, \mathcal{F}_s) = \sup \{ |E[P(G_t|\mathcal{G}_s) - P(G_t|\mathcal{F}_s)]| : G_t \in \mathcal{G}_t \}.$$

Denoting the above supremum by γ we have

$$\gamma \leq b_t(\mathcal{G}_s, \mathcal{F}_s).$$

The reverse inequality is proved as in Rogge (1974).

Let us now define for $T = N$ and integrable $X \in M(\mathcal{G})$: For $n \in T$ set $Z_n^n(X, \mathcal{G}) = X_n$ and for $k < n$

$$Z_k^n(X, \mathcal{G}) = \max \{ X_k, E(Z_{k+1}^n(X, \mathcal{G})|\mathcal{G}_k) \},$$

furthermore for $k, m \in T$,

$$X(k, m)_n = \max \{ \min \{ X_n, m \}, -k \}$$

$$Z_n(X, \mathcal{G}) = \lim_{m \rightarrow \infty} \lim_{k \rightarrow \infty} \lim_{j \rightarrow \infty} Z_n^j(X(k, m), \mathcal{G}),$$

see Chow, Robbins and Siegmund (1971), p. 67, 81.

LEMMA (2.3). For $T = N$ and $\mathcal{G}, \mathcal{F} \in g(T)$ we have for every $X \in M_1(\mathcal{G})$:

$$(i) \quad E(Z_1^n(X, \mathcal{F}) - Z_1^n(X, \mathcal{G}))^+ \leq \frac{1}{2} \sum_{j=1}^{n-1} b_{j+1}(\mathcal{G}_j, \mathcal{F}_j),$$

$$E|Z_1^n(X, \mathcal{F}) - Z_1^n(X, \mathcal{G})| \leq \sum_{j=1}^{n-1} b_{j+1}(\mathcal{G}_j, \mathcal{F}_j).$$

$$(ii) \quad E(Z_1(X, \mathcal{F}) - Z_1(X, \mathcal{G}))^+ \leq \frac{1}{2} \sum_{j=1}^{\infty} b_{j+1}(\mathcal{G}_j, \mathcal{F}_j),$$

$$E|Z_1(X, \mathcal{F}) - Z_1(X, \mathcal{G})| \leq \sum_{j=1}^{\infty} b_{j+1}(\mathcal{G}_j, \mathcal{F}_j).$$

Proof. (i) is shown by backward induction; (ii) follows from (i) by passing to the limit. ■

The following result shows the importance of B -transitivity for problems of optimal stopping; see also Irle (1981).

THEOREM (2.4). *Assume $\mathcal{G} \in g(T)$, $\mathcal{F} \in m(T)$ with $\mathcal{G} \leq \mathcal{F}$. Then the following statements are equivalent:*

- (i) \mathcal{G} is B -transitive for \mathcal{F} .
- (ii) For every integrable $X \in M(\mathcal{G})$ we have

$$v(X, \mathcal{G}) = v(X, \mathcal{F}).$$

- (iii) For every $X \in M_1(\mathcal{G})$ we have

$$v(X, \mathcal{G}) = v(X, \mathcal{F}).$$

Proof. (i) \Rightarrow (ii): By the definition of the value it is enough to show that for every $T' \subset T$, T' order-isomorphic to a subset of the natural numbers,

$$\begin{aligned} \sup \{EX_\tau: \tau \in d(\mathcal{G}), \tau(\Omega) \cap T \subset T', EX_\tau^- < \infty\} \\ = \sup \{EX_\tau: \tau \in d(\mathcal{F}), \tau(\Omega) \cap T \subset T', EX_\tau^- < \infty\} \end{aligned}$$

holds for every integrable $X \in M(\mathcal{G})$.

But this follows from (2.3) as in Chow, Robbins and Siegmund (1971), p. 103.

(ii) \Rightarrow (iii) is obviously true.

(iii) \Rightarrow (i): Consider $s, t \in T$ with $s < t$ and $G_t \in \mathcal{G}_t$. We define $X \in M_1(\mathcal{G})$ by

$$X_r = \begin{cases} 0, & r \notin \{s, t\}, \\ I_{G_t}, & r = t, \\ P(G_t | \mathcal{G}_s), & r = s. \end{cases}$$

Obviously $P(G_t) \leq v(X, \mathcal{G})$.

For $\tau \in d(\mathcal{G})$ there exists $D_s \in \mathcal{G}_s$ with $\{\tau = s\} = \{\tau \geq s\} \cap D_s$. This implies $\{\tau = s\} \subset D_s$ and $\{\tau = t\} \subset D_s^c$, thus

$$EX_\tau = \int_{\{\tau=s\}} P(G_t | \mathcal{G}_s) dP + \int_{\{\tau=t\}} I_{G_t} dP \leq \int_{D_s} P(G_t | \mathcal{G}_s) dP + \int_{D_s^c} I_{G_t} dP = P(G_t).$$

From this we have

$$v(X, \mathcal{G}) = P(G_t),$$

and by (iii)

$$P(G_t) = v(X, \mathcal{F}) \geq \sup \{EX_\tau: \tau \in d(\mathcal{F}), \tau(\Omega) \subset \{s, t\}\} \geq P(G_t).$$

This implies

$$P(G_t) = E \max \{P(G_t|\mathcal{G}_s), P(G_t|\mathcal{F}_s)\},$$

thus

$$\max \{P(G_t|\mathcal{G}_s), P(G_t|\mathcal{F}_s)\} = P(G_t|\mathcal{F}_s).$$

But this yields at once

$$P(G_t|\mathcal{G}_s) = P(G_t|\mathcal{F}_s). \quad \blacksquare$$

For increasing families one obtains:

PROPOSITION (2.5). For $\mathcal{G}, \mathcal{F} \in m(T)$ with $\mathcal{G}_t \subset \mathcal{F}_t$ for all $t \in T$ the following statements are equivalent:

- (i) \mathcal{G} is B -transitive for \mathcal{F} .
- (ii) Every supermartingale w.r.t. \mathcal{G} is a supermartingale w.r.t. \mathcal{F} .
- (iii) Every martingale w.r.t. \mathcal{G} is a martingale w.r.t. \mathcal{F} .

The easy proof is omitted.

The definition of transitivity furthermore has the following measure-theoretical consequences.

THEOREM (2.6). Assume $\mathcal{G} \in g(T)$, $\mathcal{F} \in m(T)$ and T linearly ordered. If \mathcal{G} is B -transitive for \mathcal{F} , then \mathcal{G}^* is B -transitive for \mathcal{F} and

$$\mathcal{G}^* \sim (\mathcal{G}_\infty \cap \mathcal{F}_t)_{t \in T}.$$

Proof. It is easily seen that \mathcal{G}^* is B -transitive for \mathcal{F} and thus by (2.5) every martingale w.r.t. \mathcal{G}^* is a martingale w.r.t. \mathcal{F} . This implies, see Sekiguchi (1976), p. 213, that for every $t \in T$, $\mathcal{G}_t^* \supset \mathcal{G}_\infty \cap \mathcal{F}_t$ P -a.s., thus $\mathcal{G}^* \sim (\mathcal{G}_\infty \cap \mathcal{F}_t)_{t \in T}$. \blacksquare

A necessary condition for the B -transitivity of an increasing family \mathcal{G} is thus given by $\mathcal{G} \sim (\mathcal{G}_\infty \cap \mathcal{F}_t)_{t \in T}$; the following example shows that this condition is not sufficient.

EXAMPLE (2.7). Consider random variables X_1, X_2, Y_1, Y_2, Y_0 on a probability space (Ω, \mathcal{A}, P) with the following properties.

$$X_1 = Y_1 = 0;$$

$$Y_0: \Omega \rightarrow [0; 1] \text{ is uniformly distributed};$$

$$X_2: \Omega \rightarrow \{-1, 1\} \text{ with } P(X_1 = 1|Y_0) = Y_0 = 1 - P(X_2 = -1|Y_0);$$

$$Y_2: \Omega \rightarrow \{-1, 1\} \text{ with } P(Y_2 = 1) = P(Y_2 = -1) = \frac{1}{2};$$

$$Y_0 \text{ and } Y_2 \text{ are independent};$$

$$\sigma(X_2) \cap \sigma(Y_0) = \{\emptyset, \Omega\}.$$

If we define for $T = \{1, 2\}$

$$\mathcal{G}_1 = \sigma(X_1), \quad \mathcal{G}_2 = \sigma(X_2), \quad \mathcal{H}_1 = \sigma(Y_1), \quad \mathcal{H}_2 = \sigma(Y_2),$$

$$\mathcal{F}_1 = \sigma(Y_0), \quad \mathcal{F}_2 = \mathcal{A},$$

then the following statements are easily verified:

- (i) \mathcal{G} is not B -transitive for \mathcal{F} , but for $i = 1, 2$, $\mathcal{G}_i = \mathcal{G}_\infty \cap \mathcal{F}_i$,
- (ii) \mathcal{H} is B -transitive for \mathcal{F} ,
- (iii) $v(X, \mathcal{F}, 2) = \frac{1}{4} > v(Y, \mathcal{F}, 2) = 0$,
- (iv) $L((X_1, X_2)|P) = L((Y_1, Y_2)|P)$,

where here and in the following $L(\cdot|P)$ stands for "distribution under P ".

This shows that (X_1, X_2) and (Y_1, Y_2) behave differently in optimal stopping problems w.r.t. \mathcal{F} , although they have the same distribution. ■

For the important special case that \mathcal{G} is induced by a stochastic process we introduce the following notations.

(2.8) For $X \in M$ let $\sigma(X) = (\sigma(X_t))_{t \in T} \in g(T)$, then $\sigma(X)^* = (\sigma(X_s; s \leq t))_{t \in T}$ is the corresponding increasing family.

For $\mathcal{F} \in m(T)$ and $X \in M(\mathcal{F})$, $\sigma(X)$ is B -transitive for \mathcal{F} iff X has the Markov property w.r.t. \mathcal{F} , i.e., $P(\{X_t \in B\}|\mathcal{F}_s) = P(\{X_t \in B\}|X_s)$ for every Borel set B and $s, t \in T$ with $s < t$.

We define $U = \{u = (u_t)_{t \in T}: u_t: \mathbf{R} \rightarrow \mathbf{R} \text{ measurable}\}$ and for $u \in U$, $X \in M$ an element $u(X) \in M(\sigma(X))$ by $u(X)_t = u_t \circ X_t$. Let $U_X = \{u \in U: u(X) \text{ integrable}\}$. Then (2.5) implies

COROLLARY (2.9). For $\mathcal{F} \in m(T)$ and $X \in M(\mathcal{F})$ are equivalent:

- (i) X has the Markov property w.r.t. \mathcal{F} .
- (ii) For every $u \in U_X$ we have

$$v(u(X), \sigma(X)) = v(u(X), \mathcal{F}).$$

- (iii) For every $u \in U_X$ with $u(X) \in M_1(\mathcal{F})$ we have

$$v(u(X), \sigma(X)) = v(u(X), \mathcal{F}).$$

Thus (2.9) gives a characterization of Markov processes by considering the behaviour in problems of optimal stopping. We remark that the statement "(i) \Rightarrow (ii)" is well known, see, e.g., Chow, Robbins and Siegmund (1971), p. 103.

The statements in (2.4) and (2.5) – given for arbitrary time sets – where essentially based on a discrete parameter argument (compare also (1.5)). In the following we will treat the continuous parameter case, i.e., $T = [0; \infty)$, and consider arbitrary stopping times for well-measurable processes where in general an approximation by stopping times taking only countably many values does not seem to be possible. The following notions will be useful.

A family $\mathcal{G} \in m(T)$ is called *regular* iff \mathcal{G} is P -complete, \mathcal{G}_0 contains all P -zero sets of \mathcal{G}_∞ and \mathcal{G} is right-continuous. These are the families usually encountered in the "general theory of processes".

For a regular family $\mathcal{G} \in m(T)$ we define (compare Mertens (1972)): An integrable $X \in M(\mathcal{G})$ is called a *strong \mathcal{G} -supermartingale* iff X is \mathcal{G} -well-measurable and for all bounded $\varrho, \tau \in s(\mathcal{G})$ with $\varrho \leq \tau$,

$$EX_\tau > -\infty \quad \text{and} \quad E(X_\tau | \mathcal{G}_\varrho) \leq X_\varrho.$$

X is called a *strong \mathcal{G} -martingale* iff X and $-X$ are strong \mathcal{G} -supermartingales.

A strong \mathcal{G} -supermartingale is called *regular* iff for all $\varrho, \tau \in s(\mathcal{G})$ with $\varrho \leq \tau$

$$EX_\tau \text{ exists} \quad \text{and} \quad E(X_\tau | \mathcal{G}_\varrho) \leq X_\varrho.$$

The continuous parameter version of the optional sampling theorem yields:

(2.10) *Every supermartingale w.r.t. \mathcal{G} with right-continuous paths (P-a.s.) is a strong \mathcal{G} -supermartingale; furthermore it is easily seen, that every strong \mathcal{G} -supermartingale, which is \mathcal{G} -bounded from below, is regular.*

Conversely we have according to Mertens (1972):

(2.11) *Every strong \mathcal{G} -supermartingale has (P-a.s.) paths which are upper semi-continuous from the right; every strong \mathcal{G} -martingale has (P-a.s.) right-continuous paths.*

We can now prove:

THEOREM (2.12). *Assume $T = [0; \infty)$, $\mathcal{G}, \mathcal{F} \in m(T)$, \mathcal{G} and \mathcal{F} regular with $\mathcal{G}_t \subset \mathcal{F}_t$ for all $t \in T$. Then the following statements are equivalent:*

- (i) \mathcal{G} is B -transitive for \mathcal{F} .
- (ii) $\sup \{EX_\tau; \tau \in s(\mathcal{G})\} = \sup \{EX_\tau; \tau \in s(\mathcal{F})\}$ for all \mathcal{G} -well-measurable $X \in M_1(\mathcal{G})$.
- (iii) $\sup \{EX_\tau; \tau \in s(\mathcal{G})\} = \sup \{EX_\tau; \tau \in s(\mathcal{F})\}$ for all \mathcal{G} -well-measurable $X \in M(\mathcal{G})$, which are \mathcal{G} -bounded from below.
- (iv) Every strong \mathcal{G} -supermartingale is a strong \mathcal{F} -supermartingale.
- (v) Every strong \mathcal{G} -martingale is a strong \mathcal{F} -martingale.

Proof. We will prove (i) \Rightarrow (v) \Rightarrow (iv) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i).

(i) \Rightarrow (v): Let Y be a strong \mathcal{G} -martingale; then Y is a martingale w.r.t. \mathcal{G} (in the usual sense), thus by (2.5) Y is a martingale w.r.t. \mathcal{F} . According to (2.11) Y has (P-a.s.) right-continuous paths, thus by (2.10) Y is a strong \mathcal{F} -martingale.

(v) \Rightarrow (iv): Let X be a strong \mathcal{G} -supermartingale, then, for $n \in \mathbb{N}$, $\min \{X, n\}$ defines a strong \mathcal{G} -supermartingale, for which we may apply the Doob-Meyer decomposition theorem (see Mertens (1972)). Thus there is a strong \mathcal{G} -martingale Y^n and a process $A^n \in M(\mathcal{G})$ with increasing paths, such that $P(\bigcup_{t \in T} \{\min \{X_t, n\} \neq Y_t^n - A_t^n\}) = 0$. By (v) Y^n is a strong \mathcal{F} -martingale,

which implies that $\min\{X, n\}$ is a strong \mathcal{F} -supermartingale. Letting n tend to infinity we obtain that X is a strong \mathcal{F} -supermartingale.

(iv) \Rightarrow (iii): We need the following auxiliary statement:

(2.13) *If (iv) holds, then every \mathcal{G} -well-measurable $X \in M(\mathcal{G})$, which is \mathcal{G} -bounded from below, is also \mathcal{F} -bounded from below.*

Proof of (2.13) Consider a \mathcal{G} -well-measurable $X \in M(\mathcal{G})$ and an integrable random variable h such that $X_t \geq E(h|\mathcal{G}_t)$ for all $t \in s(\mathcal{G})$. From the regularity of \mathcal{G} we may choose versions $E(h|\mathcal{G}_t)$, such that $(E(h|\mathcal{G}_t))_{t \in T}$ has right-continuous paths. Now the well-measurability of X implies together with the assumption of \mathcal{G} -boundedness that

$$P\left(\bigcup_{t \in T} \{X_t < E(h|\mathcal{G}_t)\}\right) = 0.$$

We may assume w.l.o.g., that h with this property is \mathcal{G}_∞ -measurable. Now as in (2.5), (iv) implies (v) and — using the regularity of \mathcal{G} — (v) implies (i).

From (i) one concludes by the usual extension procedure, that for every $G \in \mathcal{G}_\infty$ we have $P(G|\mathcal{G}_t) = P(G|\mathcal{F}_t)$ for every $t \in T$, thus also $E(h|\mathcal{G}_t) = E(h|\mathcal{F}_t)$ for every $t \in T$.

Choosing versions $E(h|\mathcal{F}_t)$, such that $(E(h|\mathcal{F}_t))_{t \in T}$ has right-continuous paths, we may conclude

$$P\left(\bigcup_{t \in T} \{E(h|\mathcal{F}_t) \neq E(h|\mathcal{G}_t)\}\right) = 0,$$

and

$$P\left(\bigcup_{t \in T} \{X_t < E(h|\mathcal{F}_t)\}\right) = 0.$$

This yields $X_t \geq E(h|\mathcal{F}_t)$ for every $t \in s(\mathcal{F})$, thus X is \mathcal{F} -bounded from below.

To prove the implication “(iv) \Rightarrow (iii)” consider X as in (2.13), thus $EX_t > -\infty$ for all $t \in s(\mathcal{F}) \supset s(\mathcal{G})$. Since for all $t \in s(\mathcal{F})$ $EX_t = \sup_{n \in \mathbb{N}} E \min\{X_t, n\}$, we may assume w.l.o.g. that $\sup_t EX_t \leq k$ for some $k \in \mathbb{N}$. According to Mertens (1972) there exists the minimal dominating regular strong \mathcal{G} -(resp. \mathcal{F} -)supermartingale $Z(\mathcal{G})$ (resp. $Z(\mathcal{F})$) for X such that

$$\sup \{EX_t: t \in s(\mathcal{G})\} = EZ(\mathcal{G})_0, \quad \sup \{EX_t: t \in s(\mathcal{F})\} = EZ(\mathcal{F})_0,$$

see Mertens (1972), p. 54–55, for details.

(The condition “ $\sup_t EX_t \leq k$ ” here ensures the integrability of $Z(\mathcal{G})$, $Z(\mathcal{F})$, but otherwise does not enter into the argument.)

Since X is \mathcal{F} -bounded from below, $Z(\mathcal{G})$ is also \mathcal{F} -bounded from below and by (iv) $Z(\mathcal{G})$ is a strong \mathcal{F} -supermartingale. Thus by (2.10) $Z(\mathcal{G})$ is

regular. Now the minimality of $Z(\mathcal{F})$ implies

$$Z(\mathcal{F})_0 \leq Z(\mathcal{G})_0,$$

thus

$$\sup \{EX_\tau: \tau \in s(\mathcal{G})\} = EZ(\mathcal{G})_0 \geq EZ(\mathcal{F})_0 = \sup \{EX_\tau: \tau \in s(\mathcal{F})\}.$$

Since the other inequality is obvious this proves (iii).

(iii) \Rightarrow (ii) is obvious.

(ii) \Rightarrow (i) follows as in (2.4). \blacksquare

In the following we want to consider a situation – appearing rather naturally in statistical problems – in which we find B -transitivity.

EXAMPLE (2.14). Let $T = [0; \infty)$ and P, Q Gaussian measures on (Ω, \mathcal{A}) where $\Omega = \mathbf{R}^T$ and \mathcal{A} the product- σ -algebra of the Borel sets of \mathbf{R} .

For $t \in T$ let \mathcal{F}_t be the σ -algebra generated by the coordinate mappings up to the time t . We denote the mean value function of P (resp. Q) by m_P (resp. m_Q); to simplify the exposition we assume $m_P = 0$ and $m_Q(0) = 0$.

Furthermore we assume that P and Q have a common covariance kernel K , K continuous and nonsingular, and that, for every $t \in T$, $m_Q|_{[0; t]}$ belongs to the reproducing kernel Hilbert space of $K|_{[0; t] \times [0; t]}$.

Then, for every $t \in T$, $P|_{\mathcal{F}_t}$ and $Q|_{\mathcal{F}_t}$ are equivalent with density

$$\frac{dQ}{dP} \Big|_{\mathcal{F}_t} = \exp \left(Y_t - \frac{1}{2} r(t) \right),$$

where $r: T \rightarrow T$ is increasing, and $(Y_t)_{t \in T}$ is w.r.t. P a Gaussian process with mean value function 0 and covariance kernel $r(\min \{s, t\})$; see Bhattacharya and Smith (1972); now $m_Q(0) = 0$ implies $r(0) = 0$ and $Y_0 = 0$ P -a.s.. Assume now that r is continuous, strictly increasing and unbounded. According to Irle (1980) any separable version of $(Y_{r^{-1}(t)})_{t \in T}$ then is a Wiener process w.r.t. the increasing right-continuous family $(\mathcal{F}_{r^{-1}(t)}^+)_{t \in T}$ (under P). This implies, that for all $s, t \in T$ with $s < t$ and every Borel set B

$$P(\{Y_t \in B\} | \mathcal{F}_s^+) = P(\{Y_t \in B\} | Y_s).$$

Setting $\mathcal{G}_t = \sigma(Y_t)$ it follows that \mathcal{G} is B -transitive for \mathcal{F}^+ .

We remark that this can be used to obtain locally best tests for Gaussian processes; see Irle (1980).

3. ε -transitivity

If all the available information is represented by the family $\mathcal{F} \in g(T)$ and a data-reduction has led to a family $\mathcal{G} \in g(T)$, $\mathcal{G} \leq \mathcal{F}$, then in general \mathcal{G} will

not be transitive for \mathcal{F} . According to (2.4) there exists $X \in M_1(\mathcal{G})$ with

$$v(X, \mathcal{F}) - v(X, \mathcal{G}) > 0.$$

If one wants to define a numerical quantity to characterize the possible reduction of value by looking at stopping times in $s(\mathcal{G})$ instead of $s(\mathcal{F})$, it seems natural to use the supremum of $v(X, \mathcal{F}) - v(X, \mathcal{G})$, $X \in M_1(\mathcal{G})$. It would not make sense to take the supremum for $X \in M(\mathcal{G})$ since for $X \in M_1(\mathcal{G})$ with $v(X, \mathcal{F}) - v(X, \mathcal{G}) = \delta > 0$ the quantity $v(kX, \mathcal{F}) - v(kX, \mathcal{G}) = k\delta$ tends to infinity with k ; thus we would obtain ∞ as supremum.

This leads to the following definition:

DEFINITION (3.1). For $\mathcal{G}, \mathcal{F} \in g(T)$, $t \in T^*$ set

$$\delta_t(\mathcal{G}, \mathcal{F}) = \sup \{v(X, \mathcal{F}, t) - v(X, \mathcal{G}, t) : X \in M_1(\mathcal{G})\},$$

$$\Delta_t(\mathcal{G}, \mathcal{F}) = \max \{\delta_t(\mathcal{G}, \mathcal{F}), \delta_t(\mathcal{F}, \mathcal{G})\}.$$

If $\delta(\mathcal{G}, \mathcal{F}) = \delta_\infty(\mathcal{G}, \mathcal{F}) = \varepsilon$, we call \mathcal{G} ε -transitive for \mathcal{F} .

Thus $\delta(\mathcal{G}, \mathcal{F})$ is a numerical quantity describing the behaviour of \mathcal{F} in problems of optimal stopping for \mathcal{G} . It will often be the case that \mathcal{F} represents all the available information and \mathcal{G} is obtained by data reduction, thus $\mathcal{G} \leq \mathcal{F}$, $\mathcal{F} \in m(T)$. In this situation theorem (2.4) implies:

Remark (3.2). Assume $\mathcal{G} \in g(T)$, $\mathcal{F} \in m(T)$, $\mathcal{G} \leq \mathcal{F}$. Then \mathcal{G} is B -transitive for \mathcal{F} if and only if \mathcal{G} is 0-transitive for \mathcal{F} .

Thus the notion of ε -transitivity yields a generalization of the notion of B -transitivity.

We now list some simple properties of $\delta_t(\mathcal{G}, \mathcal{F})$.

PROPOSITION (3.3). For $\mathcal{G}, \mathcal{F}, \mathcal{H} \in g(T)$, $t \in T^*$, we have

- (i) $0 \leq \delta_t(\mathcal{G}, \mathcal{F}) \leq 1$;
- (ii) $\delta_t(\mathcal{F}, \mathcal{G}) = 0$ if $\mathcal{G} \leq \mathcal{F}$;
- (iii) $\Delta_t(\mathcal{G}, \mathcal{F}) = \delta_t(\mathcal{G}, \mathcal{F})$ if $\mathcal{G} \leq \mathcal{F}$;
- (iv) $\delta_t(\mathcal{G}, \mathcal{F}) \leq \delta_t(\mathcal{G}, \mathcal{H}) + \delta_t(\mathcal{H}, \mathcal{F})$ if $\mathcal{G} \leq \mathcal{H}$;
- (v) $\Delta_t(\mathcal{G}, \mathcal{F}) \leq \Delta_t(\mathcal{G}, \mathcal{H}) + \Delta_t(\mathcal{H}, \mathcal{F})$ if $\mathcal{G} \leq \mathcal{H}$, $\mathcal{F} \leq \mathcal{H}$.

If there are some $\mathcal{G}, \mathcal{F} \in g(T)$ with $\delta_t(\mathcal{G}, \mathcal{F}) > 0$ then δ_t and Δ_t do not fulfil the triangle inequality; namely for \mathcal{O} defined by $\mathcal{O}_t = \{\emptyset, \Omega\}$ we have for $X \in M(\mathcal{O})$

$$v(X, \mathcal{O}, t) = \sup_{s \leq t} X_s = v(X, \mathcal{H}, t) \quad \text{for all } \mathcal{H} \in g(T),$$

thus $\delta_t(\mathcal{G}, \mathcal{F}) > 0 = \Delta_t(\mathcal{G}, \mathcal{O}) + \Delta_t(\mathcal{O}, \mathcal{F})$.

The definition of δ_t has furthermore the following easy consequences:

PROPOSITION (3.4). For $\mathcal{G}, \mathcal{F} \in g(T)$ we have

- (i) $\delta_s(\mathcal{G}, \mathcal{F}) \leq \delta_t(\mathcal{G}, \mathcal{F})$ for $s < t$,
- (ii) $\delta(\mathcal{G}, \mathcal{F}) = \sup_{t \in T} \delta_t(\mathcal{G}, \mathcal{F})$.

Although it seems natural to consider only processes between 0 and 1 in the definition of ε -transitivity, often more general processes occur in problems of optimal stopping. To allow such cases and also cases in which only a subset of the processes between 0 and 1 is of interest, we introduce the following concept; for this we denote the power set of $M(\mathcal{G})$ by $p(M(\mathcal{G}))$.

Consider a subset \mathcal{L}_φ of $g(T)$ and a mapping φ ,

$$\varphi: \mathcal{L}_\varphi \rightarrow \bigcup \{p(M(\mathcal{G})): \mathcal{G} \in \mathcal{L}_\varphi\}$$

such that $\varphi(\mathcal{G}) \in p(M(\mathcal{G}))$ for all $\mathcal{G} \in \mathcal{L}_\varphi$; thus for every $\mathcal{G} \in \mathcal{L}_\varphi$ there is given a subset $\varphi(\mathcal{G})$ of $M(\mathcal{G})$.

DEFINITION (3.5). For $\mathcal{G} \in \mathcal{L}_\varphi$, $\mathcal{F} \in g(T)$, $t \in T^*$ set

$$\delta_t(\mathcal{G}, \mathcal{F}; \varphi) = \sup \{v(X, \mathcal{F}, t) - v(X, \mathcal{G}, t): X \in \varphi(\mathcal{G})\},$$

$$\Delta_t(\mathcal{G}, \mathcal{F}; \varphi) = \max \{\delta_t(\mathcal{G}, \mathcal{F}; \varphi), \delta_t(\mathcal{F}, \mathcal{G}; \varphi)\} \quad \text{if } \mathcal{F} \in \mathcal{L}_\varphi.$$

If $\delta(\mathcal{G}, \mathcal{F}; \varphi) = \delta_\infty(\mathcal{G}, \mathcal{F}; \varphi) = \varepsilon$ then \mathcal{G} is called (ε, φ) -transitive for \mathcal{F} .

Remark (3.6). For $\mathcal{G} \in \mathcal{L}_\varphi$, $\mathcal{F} \in m(T)$ with $\mathcal{G} \leq \mathcal{F}$ the B -transitivity of \mathcal{G} for \mathcal{F} implies the $(0, \varphi)$ -transitivity of \mathcal{G} for \mathcal{F} .

As a simple example consider $a, b \in \mathbf{R}$, $a < b$, and define $\mathcal{L}_{\varphi(a,b)} = g(T)$ and $\varphi(a, b)(\mathcal{G}) = \{X \in M(\mathcal{G}): a \leq X \leq b\}$; thus $\delta(\mathcal{G}, \mathcal{F}; \varphi(0, 1)) = \delta(\mathcal{G}, \mathcal{F})$.

As we want to use $\delta(\mathcal{G}, \mathcal{F})$ for the investigation of problems of optimal stopping, the following two problems arise:

(i) Find φ as simple as possible with $\delta = \delta(\cdot; \varphi)$.

(ii) Use δ to describe the behaviour of $\delta(\cdot; \varphi)$ for large classes of φ which are of interest for applications.

We begin with a treatment of problem (i).

Define $\varphi(m)$ by $\mathcal{L}_{\varphi(m)} = m(T)$ and

$$\varphi(m)(\mathcal{G}) = \{X \in M_1(\mathcal{G}): X = X(f) = (E(f|\mathcal{G}_t))_{t \in T},$$

$$0 \leq f \leq 1, f \text{ } \mathcal{G}_x\text{-measurable}\},$$

thus $\varphi(m)(\mathcal{G}) \subset M_1(\mathcal{G})$ and $\delta_t(\cdot; \varphi) \leq \delta_t$.

Furthermore according to the optional sampling theorem for $\mathcal{G} \in m(T)$ and $X(f) \in \varphi(m)(\mathcal{G})$

$$v(X(f), \mathcal{G}, t) = Ef \quad \text{for all } t \in T^*.$$

$\varphi(m)(\mathcal{G})$ may be a strict subset of the set of all martingales in $M_1(\mathcal{G})$ (if T is not linearly ordered). One obtains the simple representation:

$$\delta_t(\mathcal{G}, \mathcal{F}; \varphi(m)) = \sup \{v(X(f), \mathcal{F}, t) - Ef: X(f) \in \varphi(m)(\mathcal{G})\}.$$

THEOREM (3.7). For any $\mathcal{G} \in m(T)$, $\mathcal{F} \in g(T)$ and $t \in T^*$

$$\delta_t(\mathcal{G}, \mathcal{F}) = \delta_t(\mathcal{G}, \mathcal{F}; \varphi(m)).$$

Two lemmas are used in the proof.

LEMMA (3.8). Let $\mathcal{C}_1, \mathcal{C}_2$ be sub- σ -algebras of \mathcal{A} , $\mathcal{C}_1 \subset \mathcal{C}_2$, and X_i \mathcal{C}_i -measurable, $0 \leq X_i \leq 1$ ($i = 1, 2$) such that $X_1 \geq E(X_2|\mathcal{C}_1)$.

Then there exists a \mathcal{C}_2 -measurable random variable X_0 with the following properties:

- (i) $0 \leq X_0 \leq 1$,
- (ii) $X_0 \geq X_2$,
- (iii) $E(X_0|\mathcal{C}_1) = X_1$.

The easy proof is omitted.

Now (3.8) implies:

LEMMA (3.9). Suppose that, for $\mathcal{C} \in m(N)$, $X \in M_1(\mathcal{C})$ is a supermartingale w.r.t. \mathcal{C} .

Then there exists a martingale $Y \in M(\mathcal{C})$ with the following properties:

- (i) $EY_1 = EX_1$,
- (ii) $X_n \leq Y_n \leq 1$ for all n .

Let us remark, that the martingale, which appears in the Doob decomposition of X , does not satisfy the second inequality of (ii) in general.

Proof of (3.7). Let $\varepsilon > 0$ and $X \in M_1(\mathcal{G})$ such that

$$v(X, \mathcal{F}, t) - v(X, \mathcal{G}, t) \geq \delta_t(\mathcal{G}, \mathcal{F}) - \varepsilon.$$

By the definition of the value there exists $T' \subset \{s \in T: s \leq t\}$, T' order-isomorphic to a subset of the natural numbers with

$$a = \sup \{EX_\tau: \tau \in s(\mathcal{F}), \tau(\Omega) \cap T \subset T'\} \geq v(X, \mathcal{F}, t) - \varepsilon.$$

Let furthermore

$$b = \sup \{EX_\tau: \tau \in s(\mathcal{G}), \tau(\Omega) \cap T \subset T'\}.$$

We suppose for the following that T' is infinite; the finite case is similar. Thus we have $T' = \{t_n: n \in N\}$ with $t_n < t_{n+1}$.

Define $\mathcal{G}' \in m(N)$, $\mathcal{F}' \in g(N)$, $X' \in M(\mathcal{G}')$ by

$$\mathcal{G}'_n = \mathcal{G}_{t_n}, \quad \mathcal{F}'_n = \mathcal{F}_{t_n}, \quad X'_n = X_{t_n}.$$

Now according to Chow, Robbins and Siegmund (1971), p. 75, p. 81, (compare (2.3)) $Z' = (Z'_n(X', \mathcal{G}'))_{n \in N}$ is a supermartingale w.r.t. \mathcal{G}' , and we have the following properties: $X'_n \leq Z'_n \leq 1$ for all n and $EZ'_1 = b$. By (3.9) there exists a martingale Y' w.r.t. \mathcal{G}' such that $EY'_1 = EZ'_1$ and $Z'_n \leq Y'_n \leq 1$ for all n . This implies

$$\begin{aligned} v(Y', \mathcal{F}') - v(Y', \mathcal{G}') &\geq v(X', \mathcal{F}') - EY'_1 = a - b \\ &\geq v(X, \mathcal{F}, t) - v(X, \mathcal{G}, t) - \varepsilon \geq \delta_t(\mathcal{G}, \mathcal{F}) - 2\varepsilon. \end{aligned}$$

Since Y' is a bounded martingale, there exists a random variable f with $Y'_n = E(f|\mathcal{G}_n)$ for all n , thus $Ef = b$. We may assume that f is \mathcal{G}_r -measurable and $0 \leq f \leq 1$.

Consider $X(f) \in \varphi(m)(\mathcal{G})$, i.e., $X(f)_s = E(f|\mathcal{G}_s)$ for $s \in T$. Obviously $\sup \{EX(f)_\tau : \tau \in s(\mathcal{F}), \tau(\Omega) \cap T \subset T'\} = v(Y', \mathcal{F}')$, and so

$$\begin{aligned} v(X(f), \mathcal{F}, t) - v(X(f), \mathcal{G}, t) &= v(X(f), \mathcal{F}, t) - b \\ &\geq v(Y', \mathcal{F}') - b \geq \delta_t(\mathcal{G}, \mathcal{F}) - 2\varepsilon. \end{aligned}$$

This implies $\delta_t(\mathcal{G}, \mathcal{F}; \varphi(m)) \geq \delta_t(\mathcal{G}, \mathcal{F}) - 2\varepsilon$ from which the assertion follows. ■

In the following example we will give a detailed calculation of $\delta(\mathcal{G}, \mathcal{F})$, using (3.7) as a starting point.

EXAMPLE (3.10). Let $\Omega = [0; 1]$, \mathcal{A} the σ -algebra of Borel sets of Ω and P Lebesgue measure. For any k let \mathcal{G}_k be the smallest σ -algebra containing $[0; 2^{-k})$, $[2^{-k}; 2 \cdot 2^{-k})$, ..., $[(2^k - 1)2^{-k}; 1]$ and $\mathcal{F}_k = \mathcal{G}_{k+1}$. Thus $\mathcal{G} \leq \mathcal{F}$ and the additional information in \mathcal{F} consists of "looking one step into the future".

We will use the following notations:

For $A \in \mathcal{G}_{k+1}$ let $g_k(A)$ be the largest set in \mathcal{G}_k with $g_k(A) \subset A$, $c_k(A) = A \setminus g_k(A)$, and $d_k(A)$ the smallest set $d_k(A) \in \mathcal{G}_{k+1}$ with the properties $d_k(A) \cap c_k(A) = \emptyset$, $d_k(A) \cup c_k(A) \in \mathcal{G}_k$. Obviously $P(d_k(A)) = P(c_k(A))$ and

$$P(A|\mathcal{G}_k) = I_{g_k(A)} + \frac{1}{2}(I_{c_k(A)} + I_{d_k(A)}).$$

For $X(f) \in \varphi(m)(\mathcal{G})$ we have for arbitrary $\tau \in s(\mathcal{F})$

$$\begin{aligned} EX(f)_\tau &= \sum_{k=1}^{\infty} \int P(\{\tau = k\}|\mathcal{G}_k) f dP \\ &= \sum_{k=1}^{\infty} \left(\int_{g_k(\{\tau=k\})} f dP + \frac{1}{2} \int_{d_k(\{\tau=k\}) \cup c_k(\{\tau=k\})} f dP \right). \end{aligned}$$

This implies

$$EX(f)_\tau - Ef = \frac{1}{2} \left(\sum_{k=1}^{\infty} \int_{d_k(\{\tau=k\})} f dP - \sum_{k=1}^{\infty} \int_{c_k(\{\tau=k\})} f dP \right) = \frac{1}{2} \gamma, \quad \text{say.}$$

It is now straightforward to obtain $\gamma \leq 1/2$.

This implies by (3.7)

$$\delta(\mathcal{G}, \mathcal{F}) \leq \frac{1}{4}.$$

Setting $X_2 = I_{[0; 1/4)} + I_{[1/2; 3/4)}$,

$$X_1 = E(X_2|\mathcal{G}_1) = \frac{1}{2} I_{[0; 1]},$$

we obtain

$$\begin{aligned} \sup \{EX_\tau: \tau \in s(\mathcal{F}), \tau \leq 2\} - EX_1 \\ = E \max \{X_1, E(X_2 | \mathcal{F}_1)\} - \frac{1}{2} = E \max \{X_1, X_2\} - \frac{1}{2} = \frac{1}{4}. \end{aligned}$$

This implies $\delta(\mathcal{G}, \mathcal{F}) = 1/4$. ■

So for increasing families $\mathcal{G} \in m(T)$, $\varphi(m)(\mathcal{G})$ gives a fairly simple class of processes such that $\delta_t(\mathcal{G}, \cdot) = \delta_t(\mathcal{G}, \cdot; \varphi(m))$ holds.

Such a simple characterization does not seem to be possible in general for families which do not increase, since there does not seem to exist an analogue of the optional sampling theorem for such families.

We now turn to the second problem which we mentioned in connection with (3.6), i.e., find large classes of φ such that δ describes the behaviour of $\delta(\cdot; \varphi)$.

We introduce the following notion:

DEFINITION (3.11). φ is called *compatible* iff for every $\varepsilon > 0$ there exists $\eta > 0$ such that for all $\mathcal{G}, \mathcal{F} \in \mathcal{L}_\varphi$

$$\delta(\mathcal{G}, \mathcal{F}) \leq \eta \quad \text{implies} \quad \delta(\mathcal{G}, \mathcal{F}; \varphi) \leq \varepsilon.$$

Consider the simple case $\varphi = \varphi(a, b)$.

PROPOSITION (3.12). For $a, b \in \mathbf{R}$, $a < b$ we have $\delta(\cdot; \varphi(a, b)) = (b - a)\delta$, and $\varphi(a, b)$ is compatible.

As remarked in the beginning of this chapter we cannot find a meaningful bound for arbitrary classes of unbounded processes. In the following we will treat the important special case that the considered processes are composed of a bounded part and a decreasing part, where the decreasing part may represent the costs of observations. We assume $T = N$ and consider first the set C of all deterministic cost functions, i.e.,

$$C = \{\xi: T \rightarrow [0; \infty): \xi(n) \leq \xi(n+1) \text{ for all } n, \xi(1) = 0\}.$$

The condition $\xi(1) = 0$ is no real restriction since it could always be achieved by passing from ξ to $\xi - \xi(1)$. We define $\varphi(C)$ by $\mathcal{L}_{\varphi(C)} = g(T)$ and for $\mathcal{G} \in g(T)$

$$\varphi(C)(\mathcal{G}) = \{Y = X - \xi: X \in M_1(\mathcal{G}), \xi \in C\}.$$

Thus $\varphi(C)$ describes processes which are of interest in statistical decision theory, and the condition $X \in M_1(\mathcal{G})$ only gives a suitable normalization and could be replaced by a different boundedness condition.

THEOREM (3.13). Assume $T = N$ and $\mathcal{G}, \mathcal{F} \in g(T)$.

(i) For every $k \in T$

$$\delta_k(\mathcal{G}, \mathcal{F}; \varphi(C)) \leq k\delta_k(\mathcal{G}, \mathcal{F}).$$

(ii) For every $k \in T$

$$\delta(\mathcal{G}, \mathcal{F}; \varphi(C)) \leq \frac{1}{k} + (k+2)\delta(\mathcal{G}, \mathcal{F}).$$

(iii) $\varphi(C)$ is compatible.

Proof. For $\xi \in C$ define $S(\xi) \in T^*$ by

$$S(\xi) = \inf \{n \in T: \xi(n+1) - \xi(n) \geq 1\}.$$

Then it follows for $Y = X - \xi$ with $X \in M_1(\mathcal{G})$ and for $\tau \in s(\mathcal{G}) \cup s(\mathcal{F})$

$$\begin{aligned} EY_{\min\{\tau, S(\xi)\}} - EY_\tau &= \int_{\{\tau > S(\xi)\}} (X_{S(\xi)} - \xi(S(\xi)) - X_\tau + \xi(\tau)) dP \\ &\geq \int_{\{\tau > S(\xi)\}} (\xi(\tau) - \xi(S(\xi)) - 1) dP \geq 0 \end{aligned}$$

by definition of $S(\xi)$, thus

$$EY_{\min\{\tau, S(\xi)\}} \geq EY_\tau.$$

Define $X'_n = X_n$ and $\xi'(n) = \xi(n)$ for $n \leq S(\xi)$, $X'_n = 0$ and $\xi'(n) = n - S(\xi) + \xi(S(\xi))$ for $n > S(\xi)$. It follows

$$E(X'_{\min\{\tau, S(\xi)\}} - \xi'(\min\{\tau, S(\xi)\})) \geq EY_\tau \quad \text{for all } \tau \in s(\mathcal{G}) \cup s(\mathcal{F}),$$

and

$$E(X'_\tau - \xi'(\tau)) = EY_\tau \quad \text{for } \tau \leq S(\xi),$$

which implies for $Y' = X' - \xi'$

$$v(Y', \mathcal{F}, k) = v(Y, \mathcal{F}, k), \quad v(Y', \mathcal{G}, k) = v(Y, \mathcal{G}, k) \quad \text{for all } k \in T^*.$$

By definition $\xi'(j) \leq j - 1$ for all $j \in T$, thus

$$-(k-1) \leq Y'_j \leq 1 \quad \text{for } k \in T \text{ and } j = 1, \dots, k.$$

It follows by (3.12)

$$v(Y, \mathcal{F}, k) - v(Y, \mathcal{G}, k) = v(Y', \mathcal{F}, k) - v(Y', \mathcal{G}, k) \leq k\delta_k(\mathcal{G}, \mathcal{F}),$$

which yields assertion (i).

For $k \in T$ define $k^* \geq 2$ by

$$k^* = \inf \{j \in T: \xi'(j) \geq k\}$$

and set $Y''_n = Y'_n$ for $n < k^*$, $Y''_n = -\xi'(k^*)$ for $n \geq k^*$; thus $Y'' \in M(\mathcal{G})$ and by definition of k^* and ξ'

$$-(k+1) \leq -\xi'(k^*) \leq Y'' \leq 1.$$

For every $\tau \in s(\mathcal{G})$

$$EY''_\tau = EY''_{\min\{\tau, k^*\}} \leq EY'_{\min\{\tau, k^*\}},$$

since $Y_k'' \leq Y_k'$. It follows

$$v(Y'', \mathcal{F}) - v(Y', \mathcal{G}) \leq v(Y'', \mathcal{F}) - v(Y'', \mathcal{G}) \leq (k+2)\delta(\mathcal{G}, \mathcal{F}).$$

Let us now give a bound for $v(Y', \mathcal{F}) - v(Y'', \mathcal{F})$:

For every $\tau \in s(\mathcal{F})$

$$EY'_\tau - EY''_\tau = \int_{\{\tau \geq k^*\}} (Y'_\tau + \xi'(k^*)) dP \leq P(\{\tau \geq k^*\}) \leq P(\{\xi'(\tau) \geq k\}),$$

since $\sup_{n \geq k^*} Y'_n + \xi'(k^*) \leq 1$ and $\{\tau \geq k^*\} \subset \{\xi'(\tau) \geq k\}$. If furthermore $EY'_\tau \geq 0$, then

$$1 - E\xi'(\tau) \geq EY'_\tau \geq 0,$$

thus $E\xi'(\tau) \leq 1$ and

$$P(\{\xi'(\tau) \geq k\}) \leq E\xi'(\tau)/k \leq 1/k.$$

This implies

$$v(Y', \mathcal{F}) - v(Y'', \mathcal{F}) \leq \sup \{EY'_\tau - EY''_\tau : \tau \in s(\mathcal{F}), EY'_\tau \geq 0\} \leq 1/k.$$

Altogether we obtain

$$\begin{aligned} v(Y, \mathcal{F}) - v(Y, \mathcal{G}) &= v(Y', \mathcal{F}) - v(Y', \mathcal{G}) \\ &= v(Y', \mathcal{F}) - v(Y'', \mathcal{F}) + v(Y'', \mathcal{F}) - v(Y', \mathcal{G}) \\ &\leq \frac{1}{k} + (k+2)\delta(\mathcal{G}, \mathcal{F}). \end{aligned}$$

Assertion (iii) follows immediately from (ii). ■

For deterministic cost functions we have thus proved compatibility.

We continue to assume $T = N$ and define for $\mathcal{G} \in g(T)$ the set $RC(\mathcal{G})$ of random cost functions in $M(\mathcal{G})$, i.e.,

$$RC(\mathcal{G}) = \{R \in M(\mathcal{G}) : R_n \leq R_{n+1} \text{ for all } n, R_1 = 0\}.$$

$\varphi = \varphi(RC)$ is defined by $\mathcal{L}_\varphi = g(T)$ and

$$\varphi(\mathcal{G}) = \{Y = X - R : X \in M_1(\mathcal{G}), R \in RC(\mathcal{G})\}.$$

Obviously $C \subset RC(\mathcal{G})$, thus $\varphi(C)(\mathcal{G}) \subset \varphi(RC)(\mathcal{G})$.

The following example shows that in general $\varphi(RC)$ is not compatible and that we cannot expect a result similar to (3.13) (i).

EXAMPLE (3.14). For $T = N$ we will define a sequence $(\mathcal{G}^n)_n$ in $m(T)$ and $\mathcal{F} \in m(T)$ with $\mathcal{G}^n \leq \mathcal{F}$ such that the following holds:

- (i) $\lim_n \delta(\mathcal{G}^n, \mathcal{F}) = 0$,
- (ii) $\liminf_n \delta_2(\mathcal{G}^n, \mathcal{F}; \varphi(RC)) \geq \frac{1}{2}$.

Set $\Omega = \{0, 1\}^T$, \mathcal{A} the product- σ -algebra of the power set of $\{0, 1\}$, i.e., $\mathcal{A} = p(\{0, 1\})^T$, and $P = \bigotimes_{i \in T} P_i$ where $P_i(\{0\}) = 1/i = 1 - P_i(\{1\})$. Furthermore we define $\mathcal{F}_i = \mathcal{A}$ for all $i \in T$, $\mathcal{G}_i^n = \mathcal{A}$ for all $i \geq 2$ and

$$\mathcal{G}_1^n = \bigotimes_{i \leq n-1} p(\{0, 1\}) \otimes \{\emptyset, \{0, 1\}\} \otimes \bigotimes_{i \geq n+1} p(\{0, 1\}).$$

Obviously

$$\sup_{F \in \mathcal{F}_1} \inf_{G \in \mathcal{G}_1^n} P(F \triangle G) \xrightarrow{n} 0.$$

Then, for $X \in M_1(\mathcal{G}^n)$ and $k \in T$,

$$\begin{aligned} v(X, \mathcal{F}, k) - v(X, \mathcal{G}^n, k) &= E \max \{X_1, \dots, X_k\} - \\ &\quad - E \max \{X_1, E(\max \{X_2, \dots, X_k\} | \mathcal{G}_1^n)\} \\ &\leq E(\max \{X_2, \dots, X_k\} - E(\max \{X_2, \dots, X_k\} | \mathcal{G}_1^n))^+ \\ &= \frac{1}{2} E |\max \{X_2, \dots, X_k\} - E(\max \{X_2, \dots, X_k\} | \mathcal{G}_1^n)| \\ &\leq \sup_{F \in \mathcal{F}_1} \inf_{G \in \mathcal{G}_1^n} P(F \triangle G), \end{aligned}$$

see Rogge (1974) for the last inequality.

So we have $\delta(\mathcal{G}^n, \mathcal{F}) \xrightarrow{n} 0$.

Let us now define

$$X_1^n = \frac{1}{2}, \quad X_i^n = I_{\{\pi_n = 1\}} \quad \text{for } i \geq 2,$$

where $\pi_n: \Omega \rightarrow \{0, 1\}$ denotes the n th coordinate mapping, and $R^n \in RC(\mathcal{G}^n)$ by

$$R_1^n = 0, \quad R_i^n = n I_{\{\pi_n = 0\}} \quad \text{for } i \geq 2.$$

Set $Y^n = X^n - R^n \in \varphi(RC)(\mathcal{G}^n)$. Obviously

$$\begin{aligned} v(X^n, \mathcal{F}, 2) &= E \max \{X_1, X_2\} = (1 - 1/n) + 1/(2n), \\ v(Y^n, \mathcal{F}, 2) &= v(X^n, \mathcal{F}, 2). \end{aligned}$$

We have for $\tau \in s(\mathcal{G}^n)$, $\tau \leq 2$,

$$\{\tau = 1\} = A \times \{0, 1\}, \quad \text{where} \quad A \in \bigotimes_{i \neq n} p(\{0, 1\}),$$

$$\{\tau = 2\} = \{\tau = 1\}^c = A^c \times \{0, 1\}.$$

Setting $P' = \bigotimes_{i \neq n} P_i$ this implies

$$\begin{aligned} E Y_1^n &= \frac{1}{2} P'(A) + P'(A^c) P_n(\{1\}) - P'(A^c) P_n(\{0\}) n \\ &= \frac{1}{2} P'(A) + P'(A^c) \left(1 - \frac{1}{n}\right) - P'(A^c) \leq \frac{1}{2}. \end{aligned}$$

This yields

$$v(Y^n, \mathcal{F}, 2) - v(Y^n, \mathcal{G}^n, 2) \geq \frac{1}{2n} + \left(1 - \frac{1}{n}\right) - \frac{1}{2},$$

thus

$$\liminf \delta_2(\mathcal{G}^n, \mathcal{F}; \varphi(RC)) \geq 1/2. \quad \blacksquare$$

4. ε -transitivity and invariance

In this chapter we consider the model of invariance for sequential experiments, as stated, e.g., in Hall, Wijsman and Ghosh (1965). Sampling from an infinite sequence of random variables X_n with values in a measurable space (D, \mathcal{D}) is represented by the product space $(\Omega, \mathcal{A}) = (D^{\mathbb{N}}, \mathcal{D}^{\mathbb{N}})$ where X_n is the n th coordinate mapping. An increasing sequence is defined by $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$.

Assume that there is given a family $\mathcal{P} = (P_{\vartheta})_{\vartheta \in \Theta}$ of probability measures on (Ω, \mathcal{A}) and a group G of measurable transformations on (Ω, \mathcal{A}) which leaves \mathcal{P} invariant:

Let again $T = \mathbb{N}$ and, for $n \in T$, $(\Omega_n, \mathcal{A}_n) = (D^n, \mathcal{D}^n)$ and $\omega_n = (x_1, \dots, x_n)$ for $\omega = (x_1, x_2, \dots) \in \Omega$. $P_{\vartheta, n}$ denotes the measure on $(\Omega_n, \mathcal{A}_n)$ which is induced by P_{ϑ} , and in the following we will also denote $P_{\vartheta}|_{\mathcal{F}_n}$ by $P_{\vartheta, n}$. Suppose now that for every $n \in T$ the following holds:

Every $g \in G$ induces a measurable transformation g_n on $(\Omega_n, \mathcal{A}_n)$ such that, for all $\omega \in \Omega$, $(g(\omega))_n = g_n(\omega_n)$. This yields a group G_n of measurable transformations on $(\Omega_n, \mathcal{A}_n)$, and we let \mathcal{I}_n denote the σ -algebra of invariant sets w.r.t. G_n . Furthermore there is given a sub- σ -algebra \mathcal{S}_n of \mathcal{A}_n , \mathcal{S}_n sufficient for $\{P_{\vartheta, n}; \vartheta \in \Theta\}$.

A result due to Stein says that under certain regularity conditions $\mathcal{S}_n \cap \mathcal{I}_n$ is sufficient for $\{P_{\vartheta, n}|_{\mathcal{I}_n}; \vartheta \in \Theta\}$. Next consider \mathcal{I}_n and \mathcal{S}_n as σ -algebras on Ω and assume furthermore that, for every $\vartheta \in \Theta$, $\mathcal{S} = (\mathcal{S}_n)_{n \in T}$ is B -transitive for \mathcal{F} w.r.t. P_{ϑ} . By a result of Hall, Wijsman and Ghosh (1965) $\mathcal{S} \cap \mathcal{I} = (\mathcal{S}_n \cap \mathcal{I}_n)_{n \in T}$ is B -transitive for $(\mathcal{I}_n)_{n \in T}$ w.r.t. P_{ϑ} for every ϑ (under certain regularity conditions).

In a problem of optimal stopping for a process in $M(\mathcal{S} \cap \mathcal{I})$, respectively, $M(\mathcal{I})$ there now arises the following question: How much can we gain if we use stopping times in $s(\mathcal{F})$ instead of using only stopping times in $s(\mathcal{S} \cap \mathcal{I})$, respectively $s(\mathcal{I})$. Denoting by δ_{ϑ} , respectively $\delta_{\vartheta, k}$, the quantities δ , respectively δ_k , defined w.r.t. P_{ϑ} we have

$$\delta_{\vartheta}(\mathcal{S}, \mathcal{F}) = 0, \quad \delta_{\vartheta}(\mathcal{S} \cap \mathcal{I}, \mathcal{I}) = 0,$$

and from (3.3)

$$\delta_{\vartheta}(\mathcal{S} \cap \mathcal{I}, \mathcal{F}) \leq \delta_{\vartheta}(\mathcal{S} \cap \mathcal{I}, \mathcal{I}) + \delta_{\vartheta}(\mathcal{S}, \mathcal{F}) = \delta_{\vartheta}(\mathcal{S} \cap \mathcal{I}, \mathcal{I}) \leq \delta_{\vartheta}(\mathcal{S} \cap \mathcal{I}, \mathcal{F}),$$

$$\delta_{\vartheta}(\mathcal{S} \cap \mathcal{I}, \mathcal{F}) \leq \delta_{\vartheta}(\mathcal{S} \cap \mathcal{I}, \mathcal{I}) + \delta_{\vartheta}(\mathcal{I}, \mathcal{F}) = \delta_{\vartheta}(\mathcal{I}, \mathcal{F}).$$

This implies

$$\delta_{\mathfrak{g}}(\mathcal{I} \cap \mathcal{J}, \mathcal{F}) = \delta_{\mathfrak{g}}(\mathcal{I} \cap \mathcal{J}, \mathcal{I}) \leq \delta_{\mathfrak{g}}(\mathcal{I}, \mathcal{F}).$$

Now this suggests the use of $\delta_{\mathfrak{g}}(\mathcal{I}, \mathcal{F})$ ($\delta_{\mathfrak{g},k}(\mathcal{I}, \mathcal{F})$) to describe the consequences which a reduction by invariance has for problems of optimal stopping. In the remaining part of this chapter we will consider a special invariance structure to illustrate these ideas:

Let $(\Omega, \mathcal{A}) = (\mathbf{R}^T, \mathcal{B}^T)$ where \mathcal{B} denotes the Borel sets of the real line, $\Theta = \mathbf{R}$, and for $\mathfrak{g} \in \Theta P_{\mathfrak{g}} = N(\mathfrak{g})^T$ where $N(\mathfrak{g})$ denotes the normal distribution with mean \mathfrak{g} and variance 1; thus $P_{\mathfrak{g},n} = N(\mathfrak{g})^n$. Furthermore let $G = \{g^+, g^-\}$ where $g^+(\omega) = \omega$ and $g^-(\omega) = -\omega$, and $G_n = \{g_n^+, g_n^-\}$ with $g_n^+(\omega_n) = \omega_n$ and $g_n^-(\omega_n) = -\omega_n$. This yields for \mathcal{I}_n -considered as a sub- σ -algebra of \mathcal{A} (where for easier notation we set $T(n) = T \setminus \{1, \dots, n\}$)

$$\mathcal{I}_n = \{A \times \mathbf{R}^{T(n)} : A \in \mathcal{B}^n, A = -A\}.$$

In this situation, a problem of optimal stopping for a process in $M(\mathcal{I})$ and stopping times in $s(\mathcal{I})$ is considered, e.g., by Lai (1973).

Let $g_{\mathfrak{g},n}(\omega_n) = (2\pi)^{-n/2} \exp(-\frac{1}{2} \sum_{j=1}^n (x_j - \mathfrak{g})^2)$ denote the density of $P_{\mathfrak{g},n}$

with respect to n -dimensional Lebesgue measure; for easier notation we will not distinguish in the following between \mathcal{F}_n -measurable mappings defined on Ω and \mathcal{A}_n -measurable mappings defined on Ω_n .

(4.1) For a bounded \mathcal{F}_n -measurable random variable f_n a version $\tilde{f}_n^{\mathfrak{g}}$ of $E_{\mathfrak{g}}(f_n | \mathcal{I}_n)$ is given by

$$\tilde{f}_n^{\mathfrak{g}}(\omega) = (f_n(\omega) g_{\mathfrak{g},n}(\omega_n) + f_n(-\omega) g_{\mathfrak{g},n}(-\omega_n)) / (g_{\mathfrak{g},n}(\omega_n) + g_{\mathfrak{g},n}(-\omega_n)).$$

From this it is easily seen that, for $\mathfrak{g} \neq 0$, \mathcal{I} is not B -transitive for \mathcal{F} w.r.t. $P_{\mathfrak{g}}$.

(4.2) Set $f_2 = (X_1 + X_2)^2$; then we have for $f_1^{\mathfrak{g}} = E_{\mathfrak{g}}(f_2 | \mathcal{F}_1)$ $f_1^{\mathfrak{g}}(x_1) = E_{\mathfrak{g}}(x_1 + X_2)^2 = 1 + (x_1 + \mathfrak{g})^2$, so that $f_1^{\mathfrak{g}}(x_1) \neq f_1^{\mathfrak{g}}(-x_1)$, for $x_1 \neq 0$, $\mathfrak{g} \neq 0$ and $\tilde{f}_1^{\mathfrak{g}}(x_1) \neq f_1^{\mathfrak{g}}(x_1)$, since by (5.1) $\tilde{f}_1^{\mathfrak{g}}(x_1)$ is a nontrivial convex combination of $f_1^{\mathfrak{g}}(x_1)$ and $f_1^{\mathfrak{g}}(-x_1)$. It is easy to give a direct proof that B -transitivity holds for $\mathfrak{g} = 0$; but this will also be a consequence of the following more general considerations.

Let $n \in T$, $n \geq 2$, and $\mathfrak{g} \in \Theta$. Since \mathcal{I} is an increasing family we have according to (3.7)

$$\begin{aligned} \delta_{\mathfrak{g},n}(\mathcal{I}, \mathcal{F}) &= \delta_{\mathfrak{g},n}(\mathcal{I}, \mathcal{F}; \varphi(m)) \\ &= \sup \{v_{\mathfrak{g}}(X(f), \mathcal{F}, n) - E_{\mathfrak{g}}f : X(f) = (E_{\mathfrak{g}}(f | \mathcal{I}_k))_{k \in T}, \\ &\quad 0 \leq f \leq 1, f \text{ } \mathcal{I}_n\text{-measurable}\}. \end{aligned}$$

Let f be \mathcal{I}_n -measurable and bounded. For $k = 1, \dots, n$ set

$$f_k = E_{\mathfrak{g}}(f | \mathcal{F}_k), \tilde{f}_k = E_{\mathfrak{g}}(f | \mathcal{I}_k), \text{ thus } f = f_n = \tilde{f}_n, \tilde{f}_k = E_{\mathfrak{g}}(f_k | \mathcal{I}_k).$$

(Here and in the following we suppress ϑ in some notations.)

(4.3) For $\tau \in s(\mathcal{F})$ with $\tau \leq n$

$$\begin{aligned} E_{\vartheta} \tilde{f}_{\tau} - E_{\vartheta} f \\ = \sum_{k=1}^{n-1} \int_{\{\tau=k\}} (f_k(-\omega) - f_k(\omega)) \exp(-\tfrac{1}{2} \vartheta^2 k) / (\exp(\vartheta S_k) + \exp(-\vartheta S_k)) dP_0, \end{aligned}$$

where $S_k(\omega) = x_1 + \dots + x_k$.

For $k < n$

$$f_k(\omega_k) = (2\pi)^{-(n-k)/2} \int_{\mathbb{R}^{n-k}} f(\omega_k, x_{k+1}, \dots, x_n) \exp(-\tfrac{1}{2} \sum_{j=k+1}^n (x_j - \vartheta)^2) d\lambda^{n-k},$$

thus

$$\begin{aligned} f_k(-\omega_k) \\ = (2\pi)^{-(n-k)/2} \int_{\mathbb{R}^{n-k}} f(-\omega_k, x_{k+1}, \dots, x_n) \exp(-\tfrac{1}{2} \sum_{j=k+1}^n (x_j - \vartheta)^2) d\lambda^{n-k} \\ = (2\pi)^{-(n-k)/2} \int_{\mathbb{R}^{n-k}} f(-\omega_k, -x_{k+1}, \dots, -x_n) \exp(-\tfrac{1}{2} \sum_{j=k+1}^n (x_j + \vartheta)^2) d\lambda^{n-k} \\ = E_{-\vartheta}(f | \hat{\mathcal{F}}_k), \end{aligned}$$

since $f(\omega_n) = f(-\omega_n)$. This implies

$$\begin{aligned} (4.4) \quad f_k(-\omega_k) - f_k(\omega_k) \\ = \int_{\mathbb{R}^{n-k}} f(\omega_k, x_{k+1}, \dots, x_n) (g_{-\vartheta, n-k}(x_{k+1}, \dots, x_n) - g_{\vartheta, n-k}(x_{k+1}, \dots, x_n)) d\lambda^{n-k}. \end{aligned}$$

Using these computations we can now prove for $n \in T$ and $\vartheta \in \Theta$

THEOREM (4.5).

$$\delta_{\vartheta, n}(\mathcal{J}, \mathcal{F}) \leq \min \{ \exp(-\tfrac{1}{2} \vartheta^2), \exp(-\tfrac{1}{2} \vartheta^2) (2 - 2 \exp(-\tfrac{1}{2} \vartheta^2 (n-1)))^{1/2} \}.$$

Proof. Let f be \mathcal{J}_n -measurable, $0 \leq f \leq 1$. Since

$$|f_k(\omega) - f_k(-\omega)| \leq 1 \quad \text{and} \quad \exp(\vartheta S_k) + \exp(-\vartheta S_k) \geq 1$$

(4.3) implies that for every $\tau \in s(\mathcal{F})$, $\tau \leq n$

$$|E_{\vartheta} \tilde{f}_{\tau} - E_{\vartheta} f| \leq \sum_{k=1}^{n-1} \exp(-\tfrac{1}{2} \vartheta^2 k) P_0(\{\tau = k\}) \leq \exp(-\tfrac{1}{2} \vartheta^2);$$

furthermore from (4.3) and (4.4)

$$|E_{\vartheta} \tilde{f}_{\tau} - E_{\vartheta} f| \leq \sum_{k=1}^{n-1} \exp(-\tfrac{1}{2} \vartheta^2 k) P_0(\{\tau = k\}) \int |g_{\vartheta, n-k} - g_{-\vartheta, n-k}| d\lambda^{n-k}.$$

To obtain an easy bound for $\int |g_{\vartheta, n-k} - g_{-\vartheta, n-k}| d\lambda^{n-k}$ we use the Hellinger distance H which for two probability measures Q_1, Q_2 with densities q_1, q_2 w.r.t. a σ -finite measure μ is defined as

$$H(Q_1, Q_2) = (1 - \int \sqrt{q_1 q_2} d\mu)^{1/2}.$$

It is well known (see, e.g., LeCam (1974), p. 58) that

$$H^2(Q_1, Q_2) \leq \int |q_1 - q_2| d\mu \leq 2^{1/2} H(Q_1, Q_2).$$

From the definition of H one obtains for the product measures

$$H(Q_1^n, Q_2^n) = (1 - (\int \sqrt{q_1 q_2} d\mu)^n)^{1/2}.$$

Using these facts and

$$\int \sqrt{g_{\vartheta, 1} g_{-\vartheta, 1}} d\lambda^1 = \exp(-\tfrac{1}{2} \vartheta^2)$$

we conclude

$$\int |g_{\vartheta, k} - g_{-\vartheta, k}| d\lambda^k \leq (2 - 2 \exp(-\tfrac{1}{2} \vartheta^2 k))^{1/2}.$$

This implies

$$\begin{aligned} |E_{\vartheta} \tilde{f}_\tau - E_{\vartheta} f| &\leq \sum_{k=1}^{n-1} \exp(-\tfrac{1}{2} \vartheta^2 k) (2 - 2 \exp(-\tfrac{1}{2} \vartheta^2 (n-k)))^{1/2} P_0(\{\tau = k\}) \\ &\leq \exp(-\tfrac{1}{2} \vartheta^2) (2 - 2 \exp(-\tfrac{1}{2} \vartheta^2 (n-1)))^{1/2}. \end{aligned}$$

The assertion now follows from (3.7). ■

COROLLARY (4.6). (i) $\delta_{\vartheta}(\mathcal{I}, \mathcal{F}) = 0$ for $\vartheta = 0$.

(ii) $\delta_{\vartheta}(\mathcal{I}, \mathcal{F}) \leq \exp(-\tfrac{1}{2} \vartheta^2)$ and $\lim_{\vartheta \rightarrow \infty} \delta_{\vartheta}(\mathcal{I}, \mathcal{F}) = 0$.

(iii) For all $n \in T$

$$\lim_{\vartheta \rightarrow 0} \delta_{\vartheta, n}(\mathcal{I}, \mathcal{F}) = 0.$$

The proof is immediate from (4.5).

For stopping problems in the presence of cost functions we obtain at once

$$\delta_{\vartheta}(\mathcal{I}, \mathcal{F}; \varphi(C)) = 0 \quad \text{for } \vartheta = 0$$

and for $\vartheta \neq 0$:

COROLLARY (4.7). (i) For all $\vartheta \neq 0$

$$\delta_{\vartheta}(\mathcal{I}, \mathcal{F}; \varphi(C)) \leq (\max\{1, \exp(\tfrac{1}{4} \vartheta^2) - 1\})^{-1} + \exp(-\tfrac{1}{4} \vartheta^2) + 2 \exp(-\tfrac{1}{2} \vartheta^2).$$

(ii) $\lim_{\vartheta \rightarrow \infty} \delta_{\vartheta}(\mathcal{I}, \mathcal{F}; \varphi(C)) = 0$.

Proof. It follows from (3.13) and (4.6) that

$$\delta_{\vartheta}(\mathcal{I}, \mathcal{F}; \varphi(C)) \leq \inf_{k \in T} \{(1/k) + (k+2) \exp(-\frac{1}{2} \vartheta^2)\}.$$

The mapping $(1/x) + (x+2) \exp(-\vartheta^2/2)$ attains its minimum at $\exp(\vartheta^2/4)$. Setting $k = [\exp(\vartheta^2/4)] \in T$ we have

$$\max\{1, \exp(\vartheta^2/4) - 1\} \leq k \leq \exp(\vartheta^2/4),$$

which implies (i).

(ii) is immediate from (i). ■

The bound in (4.5) does not yield that $\lim_{\vartheta \rightarrow 0} \delta_{\vartheta}(\mathcal{I}, \mathcal{F}) = 0$ holds. We will show in the following that this is not due to the method of proof, but that actually

$$\liminf_{\vartheta \rightarrow 0} \delta_{\vartheta}(\mathcal{I}, \mathcal{F}) > 0.$$

THEOREM (4.8). *In the above situation we have*

$$\liminf_{\vartheta \rightarrow 0} \delta_{\vartheta}(\mathcal{I}, \mathcal{F}) > 0.$$

Proof. Let $\vartheta \in (0; 1)$, $c > 0$ and $n \in T$, $n \geq 2$. An \mathcal{I}_n -measurable random variable f , $0 \leq f \leq 1$, is defined by

$$f = I_{\{|S_n| \leq c\}}.$$

Then, for $k = 1, \dots, n-1$,

$$\begin{aligned} f_k &= E_{\vartheta}(f | \mathcal{F}_k) = P_{\vartheta}(\{|S_n| \leq c\} | S_k) \\ &= P_{\vartheta}(\{S_n - S_k \in [-c - S_k; c - S_k]\} | S_k) \\ &= L(S_{n-k} | P_{\vartheta})([-c - S_k; c - S_k]) \\ &= L(S_{n-k} | P_0)([-c - S_k - \vartheta(n-k); c - S_k - \vartheta(n-k)]) \\ &= N(0)([(-c - S_k - \vartheta(n-k))/\sqrt{n-k}; (c - S_k - \vartheta(n-k))/\sqrt{n-k}]). \end{aligned}$$

$E_{-\vartheta}(f | \mathcal{F}_k)$ is obtained by replacing ϑ by $-\vartheta$ in the above formula.

We now choose $n = n(\vartheta) = [\vartheta^{-2}] \in T$ and $c = 1/\vartheta$. Setting

$$\tau(\vartheta) = \min\{\inf\{k \in T: S_k \geq 1/\vartheta\}, n(\vartheta)\}, \quad U(\vartheta) = (S_{\tau(\vartheta)} - 1/\vartheta)^+,$$

we obtain for $f = f(\vartheta)$

$$\begin{aligned} &\sum_{k=1}^{n(\vartheta)} I_{\{\tau(\vartheta)=k\}} (E_{-\vartheta}(f | \mathcal{F}_k) - E_{\vartheta}(f | \mathcal{F}_k)) \\ &= \left\{ N(0) \left(\left[\frac{-2 - \vartheta U(\vartheta)}{\sqrt{\vartheta^2 n(\vartheta) - \vartheta^2 \tau(\vartheta)}} + \sqrt{\vartheta^2 n(\vartheta) - \vartheta^2 \tau(\vartheta)}; \frac{-\vartheta U(\vartheta)}{\sqrt{\vartheta^2 n(\vartheta) - \vartheta^2 \tau(\vartheta)}} + \right. \right. \end{aligned}$$

$$+ \sqrt{\vartheta^2 n(\vartheta) - \vartheta^2 \tau(\vartheta)} \Bigg] - N(0) \left(\left[\frac{-2 - \vartheta U(\vartheta)}{\sqrt{\vartheta^2 n(\vartheta) - \vartheta^2 \tau(\vartheta)}} - \sqrt{\vartheta^2 n(\vartheta) - \vartheta^2 \tau(\vartheta)}; \frac{-\vartheta U(\vartheta)}{\sqrt{\vartheta^2 n(\vartheta) - \vartheta^2 \tau(\vartheta)}} - \sqrt{\vartheta^2 n(\vartheta) - \vartheta^2 \tau(\vartheta)} \right] \right) \Bigg\} I_{\{\tau(\vartheta) < n(\vartheta)\}}$$

We define $g: [0; \infty) \times [0; 1] \rightarrow [0; 1]$ by

$$g(u, v) = N(0) \left(\left[-\frac{2+u}{\sqrt{v}} + \sqrt{v}; -\frac{u}{\sqrt{v}} + \sqrt{v} \right] \right) - N(0) \left(\left[-\frac{2+u}{\sqrt{v}} - \sqrt{v}; -\frac{u}{\sqrt{v}} - \sqrt{v} \right] \right)$$

for $v > 0$, and $g(u, v) = 0$ for $v = 0$.

Obviously g is continuous on $[0; \infty) \times [0; 1]$ with

$$g(0, v) > 0 \quad \text{for } v \in (0; 1].$$

Next let Q be Wiener measure on the Borel sets of the space $\mathcal{C}[0; 1]$ of continuous functions on $[0; 1]$, i.e. the coordinate mappings $(B_t)_{t \in [0; 1]}$ are a Brownian motion under Q .

Set $\tau_1 = \min \{ \inf \{ t: B_t \geq 1 \}, 1 \}$. It is well known that τ_1 is a measurable mapping from $\mathcal{C}[0; 1]$ to $[0; 1]$ and that the set of discontinuities of τ_1 has measure 0 under Q . We will show in the following

$$(4.9) \quad L(\vartheta^2 n(\vartheta) - \vartheta^2 \tau(\vartheta) | P_0) \xrightarrow{\vartheta \downarrow 0} L(1 - \tau_1 | Q).$$

$$(4.10) \quad \vartheta U(\vartheta) \xrightarrow{\vartheta \downarrow 0} 0 \quad \text{in } P_0\text{-probability.}$$

Assuming the validity of (4.9) and (4.10) it follows that

$$L(\vartheta U(\vartheta), \vartheta^2 n(\vartheta) - \vartheta^2 \tau(\vartheta) | P_0) \xrightarrow{\vartheta \downarrow 0} L(0, 1 - \tau_1 | Q),$$

see, e.g., Billingsley (1968, p. 27), so that we obtain for $f(\vartheta) = I_{\{|S_{n(\vartheta)}| \leq 1/\vartheta\}}$

$$E_{\vartheta} \tilde{f}(\vartheta)_{\tau(\vartheta)} - E_{\vartheta} f(\vartheta) \xrightarrow{\vartheta \downarrow 0} \int_{\{\tau_1 < 1\}} \frac{\exp(-\tau_1/2)}{\exp(1) + \exp(-1)} g(0, 1 - \tau_1) dQ > 0,$$

since $Q(\{\tau_1 < 1\}) > 0$ and $g(0, v) > 0$ for $v > 0$. This implies

$$\liminf_{\vartheta \downarrow 0} \delta_{\vartheta}(\mathcal{J}, \mathcal{F}) \geq \lim_{\vartheta \downarrow 0} (E_{\vartheta} \tilde{f}(\vartheta)_{\tau(\vartheta)} - E_{\vartheta} f(\vartheta)) > 0.$$

Similarly one obtains $\liminf_{\vartheta \downarrow 0} \delta_{\vartheta}(\mathcal{J}, \mathcal{F}) > 0$.

Proof of (4.9). Since

$$\lim_{\vartheta \downarrow 0} \vartheta^2 n(\vartheta) = 1$$

it is sufficient to prove $\lim_{g \downarrow 0} L(g^2 \tau(g) | P_0) = L(\tau_1 | Q)$. Set $\varepsilon(g) = n(g)^{-1/2}$. Then for $\varepsilon = \varepsilon(g)$

$$k g^2 \leq k \varepsilon^2 \leq (k+1) g^2 \quad \text{for } k = 1, \dots, n(g).$$

Next if $(s_n)_{n \geq 0}$ is a sequence of real numbers, we define elements $s_{1,g}$, $s_{2,g}$ and $s_{3,g}$ of $\mathcal{C}[0; 1]$ by

$$\begin{aligned} s_{1,g}(i g^2) &= g s_i & \text{for } i = 0, \dots, n(g), \quad s_{1,g}(1) &= g s_{n(g)}, \\ s_{2,g}(i \varepsilon^2) &= g s_i & \text{for } i = 0, \dots, n(g), \\ s_{3,g}(i \varepsilon^2) &= \varepsilon s_i & \text{for } i = 0, \dots, n(g), \end{aligned}$$

and linear between these points. For $S_0 = 0$, $S_1 = X_1, \dots, S_n = X_1 + \dots + X_n$ we thus obtain random elements $S_{1,g}$, $S_{2,g}$ and $S_{3,g}$ with values in $\mathcal{C}[0; 1]$.

According to Donsker's invariance principle (see. e.g., Billingsley (1968), p. 68) we have

$$L(S_{3,g} | P_0) \xrightarrow{g \downarrow 0} Q.$$

Easy geometrical arguments yield

$$\begin{aligned} \|s_{1,g} - s_{2,g}\|_s &\leq g \max_{i \leq n(g)} |s_i - s_{i-1}|, \\ \|s_{2,g} - s_{3,g}\|_s &\leq g, \end{aligned}$$

which implies

$$\|S_{1,g} - S_{2,g}\|_s \leq g \max_{i \leq n(g)} |X_i| \leq n(g)^{-1/2} \max_{i \leq n(g)} |X_i| \xrightarrow{g \downarrow 0} 0 \quad P_0\text{-a.s.}$$

(see, e.g., Barndorff-Nielsen (1963)), thus

$$\|S_{1,g} - S_{3,g}\|_s \xrightarrow{g \downarrow 0} 0 \quad P_0\text{-a.s.}$$

This yields

$$L(S_{1,g} | P_0) \xrightarrow{g \downarrow 0} Q,$$

and so

$$L(\tau_1 \circ S_{1,g} | P_0) \xrightarrow{g \downarrow 0} L(\tau_1 | Q).$$

From

$$\tau_1 \circ S_{1,g} = \min \{ \inf \{ t: S_{1,g}(t) \geq 1 \}, 1 \}$$

it is easily seen that

$$g^2 \tau(g) - g^2 \leq \tau_1 \circ S_{1,g} \leq g^2 \tau(g),$$

which implies

$$L(\eta^2 \tau(\eta)|P_0) \xrightarrow{\eta \downarrow 0} L(\tau_1|Q). \quad \blacksquare$$

Proof of (4.10). We have $\sup_{\eta > 0} E_0 U(\eta) < \infty$, see, e.g., Lorden (1970), thus for every $\eta > 0$

$$P_0(\{\eta U(\eta) > \eta\}) \leq \frac{\eta}{\eta} \sup_{\eta > 0} E_0 U(\eta) \xrightarrow{\eta \downarrow 0} 0. \quad \blacksquare$$

References

- [1] R. R. Bahadur (1954), *Sufficiency and statistical decision functions*, Ann. Math. Statist. **25**, 423–462.
- [2] O. Barndorff-Nielsen (1963), *On the limit behaviour of extreme order statistics*, Ann. Math. Statist. **34**, 994–1002.
- [3] P. K. Bhattacharya and R. P. Smith (1972), *Sequential probability ratio test for the mean value function of a Gaussian process*, Ann. Math. Statist. **43**, 1861–1873.
- [4] P. Billingsley (1968), *Convergence of Probability Measures*, Wiley, New York.
- [5] Y. S. Chow, H. Robbins, and D. Siegmund (1971), *Great Expectations: The Theory of Optimal Stopping*, Houghton-Mifflin, Boston.
- [6] R. A. Fisher (1925), *Theory of statistical estimation*, Proc. Cambridge Philos. Soc. **22**, 700–725.
- [7] W. J. Hall, R. W. Wijsman, and B. K. Ghosh (1965), *The relationship between sufficiency and invariance with applications in sequential analysis*, Ann. Math. Statist. **36**, 575–614.
- [8] A. Irle (1980), *Locally best tests for Gaussian processes*, Metrika **27**, 15–28.
- [9] – (1981), *Transitivity in problems of optimal stopping*, Ann. Probab. **9**, 642–647.
- [10] T. L. Lai (1973), *Optimal stopping and sequential tests which minimize the maximum expected sample size*, Ann. Statist. **1**, 659–673.
- [11] L. LeCam (1974), *Notes on Asymptotic Methods in Statistical Decision Theory*, Centre de Recherches Mathématiques, Montreal.
- [12] G. Lorden (1970), *On excess over the boundary*, Ann. Math. Statist. **41**, 520–527.
- [13] J. F. Mertens (1972), *Théorie des processus stochastiques généraux, applications aux surmartingales*, Z. Wahrsch. Verw. Gebiete **22**, 45–69.
- [14] L. Rogge (1974), *Uniform inequalities for conditional expectations*, Ann. Probab. **2**, 486–489.
- [15] T. Sekiguchi (1976), *On the Krickeberg decomposition for continuous martingales*, Séminaire de Probabilités X, Université de Strasbourg 209–215.
- [16] M. E. Thompson (1971), *Continuous parameter optimal stopping problems*, Z. Wahrsch. Verw. Gebiete **19**, 302–318.

Presented to the semester
Sequential Methods in Statistics
September 7–December 11, 1981
