The Annals of Statistics
1990, Vol. 18, No. 4, 1501–1555

A SURVEY OF PRODUCT-INTEGRATION WITH A VIEW TOWARD APPLICATION IN SURVIVAL ANALYSIS

BY RICHARD D. GILL AND SØREN JOHANSEN

University of Utrecht and University of Copenhagen

The correspondence between a survival function and its hazard or failure-rate is a central idea in survival analysis and in the theory of counting processes. This correspondence is shown to be a special case of a more general correspondence between multiplicative and additive matrix-valued measures on the real line. Additive integration of the survival function produces the hazard, while the multiplicative integral, or so-called *product-integral*, of the hazard yields the survival function. The easy generalization to the *matrix* case (noncommutative multiplication) allows an elegant and completely parallel treatment of intensity measures of Markov processes, with many possible applications in multistate survival models. However, the difficulties and multiplicity of theories of product-integration in multivariate time explain why so many different multivariate product-limit estimators exist. We give a complete and elementary treatment of the basic theory of the product-integral $\prod(1 + dX)$ together with a discussion of some of its applications. New results are given on the compact differentiability of the product-integral, to be used along with the functional δ-method for getting large-sample results for product-limit estimators.

1. History. Product-integration has a long history in pure and applied mathematics. At the same time it has many applications in statistics and probability. However, the product-integral (also known as the multiplicative integral) is almost unknown among statisticians and probabilists, and its properties are continually being rediscovered. We shall try to remedy this situation by collecting together the key facts on matrix product-integration over intervals of the real line, giving self-contained and elementary proofs. We discuss the connection with the not so trivial theory of exponential semimartingales. We include new results, in particular, on functional differentiability of the product-integral. We give applications of product-integration in survival analysis (in a wide sense): to the product-limit estimator, in the probabilistic and statistical theory of Markov processes (the correspondence between transition probabilities and cumulative intensities, and their estimation), in the estimation of branching process models, and in likelihood expressions for counting process experiments. We also discuss product-integration with multidimensional time, with applications to multivariate product-limit estimators. The differentiability of the product-integral allows a rigorous

Received June 1987; revised April 1990.

AMS 1980 *subject classifications*. Primary 60J27, 62G05; secondary 60H20, 45D05.

Key words and phrases. Multiplicative integral, Volterra integral equation, Kolmogorov forward and backward equations, Duhamel equation, Doléans-Dade exponential semimartingale, compact (Hadamard) differentiability, Markov process, intensity measure, branching processes, product-limit (Kaplan–Meier) estimator.

derivation of the large sample properties of product-limit estimators under varied sampling schemes by the functional version of the δ-method. Efficiency and bootstrap results also follow immediately with this approach.

The present paper does not cover all possible applications of product-integration in statistics, by any means. An important omission is the use of product-integration in the prediction theory of multivariate time series; see Masani (1966, 1979) for introductions to this area. Indeed the kind of application which motivates us has also strongly coloured our own approach to the theory of product-integration.

We propose here the new notation $\pi(1 + dX)$ for the product-integral. This seems to be a natural way to complete the two-by-two table

$$\begin{array}{cc} \Sigma & \prod \\ \int & \pi \end{array}$$

though many other notations have been used in the past.

The product-integral was introduced by Volterra (1887) as the solution of a certain basic integral equation. The notation $\widehat{\int}(1 + dX)$ is due to Schlesinger; Volterra confusingly just wrote $\int dx$. The notion was further exploited and developed, especially by Schlesinger (1931, 1932), Rasch (1930, 1934), Birkhoff (1938) and Masani (1947). In particular, Rasch (1930) introduced the notation $\prod(1 + dX)$ for the product-integral that is now favoured by many authors, though Arley's $\mathscr{P}(1 + dX)$ [Arley (1943)] is also nice. All these papers are concerned with the (absolutely) continuous case. More recently, starting with a paper by Wall (1953), a highly abstract theory of product-integration was established by Mac Nerney (1963), B. W. Helton (1966) and J. C. Helton (1975a, 1975b) among others (here we have just indicated the contributions of those authors which are most relevant for our purposes). Especially important for us is the fact that this theory allows discrete as well as continuous integrating measures. However, most statisticians will find the setting and notation in these papers very unfamiliar.

The major textbook on product-integration, by Dollard and Friedman (1979), is a mine of information but has for our purposes a serious defect. The last-mentioned school of product-integration theory is only summarily covered (though extensive references are given), while instead the late and only chapter on Stieltjes product-integration—that is, with respect to possibly discrete measures—treats a completely different and, as far as our statistical applications are concerned, completely inappropriate product-integral, $\pi\, e^{dX}$. Moreover, the theory of this "exponential integral" contains in the discrete case many unpleasant complications which simply do not arise with $\pi(1 + dX)$. As we shall see, the theory of the ordinary product-integral which we present here allows an effortless unification of the continuous and discrete, in both the scalar and matrix cases. This is vital for statistical applications in which underlying measures may be continuous, but natural estimators are discrete.

Product-integration makes a very natural appearance in the theory of Markov processes, as was already pointed out in the book of Volterra and

Hostinsky (1938). A notable early work building on this is by Arley (1943); see also Arley and Borschenius (1945). From here it found its way into survival analysis in an informal way via Cox (1972) and Kalbfleisch and Prentice (1980).

A very important early paper on Markov processes is by Dobrushin (1953), who earlier than Wall's followers established the correspondence between certain sum- and product-integrals and removed the continuity restriction in the right way for our purposes. This paper has never appeared in English translation and very few authors have followed up his results. Another connection in probability and statistics is that Doléans-Dade's exponential semimartingale [Doléans-Dade (1970)], which plays such an important role in stochastic analysis, is in a sense just a product-integral, as we shall see later. The same holds for Jacod's formula [Jacod (1975)] for the likelihood ratio (Radon–Nikodym derivative) for a counting process experiment. Less surprisingly, so is the product-limit estimator from survival analysis of Kaplan and Meier (1958).

A further confusing factor in the history of product-integration is the fact that the theory can be built up from some very different starting points. We therefore conclude this first part of the paper by summarizing three *equivalent* definitions of the same product-integral, and stating its most important property of multiplicativity. We also give a fourth equivalent definition for the scalar (commutative) case. In Section 2 we start with a version of the first of these definitions (the product-limit definition) and prove its equivalence with the others and derive other basic properties of product-integrals. In Section 3 further properties of continuity and differentiability are derived and we discuss the stochastic product-integral of Doléans-Dade and product-integration over general spaces (multivariate time). Finally, in Section 4 we sketch a number of statistical applications.

We define the product-integral for finite real matrix-valued measures defined on the Borel subsets $\mathscr{B}$ of the interval $]0, \tau]$, say. Let X be such a measure. Thus each component X_{ij} of X, $1 \le i, j \le p$, is a finite real (signed) measure on $]0, \tau]$. We can represent X by its distribution function, which we shall denote by the same symbol; thus $X(t) = X(]0, t])$ is a $p \times p$ matrix. Let $\mathbf{1}$ be the identity matrix and $\mathbf{0}$ the matrix of zeros. The product-integral of X, written $Y = \prod(\mathbf{1} + dX)$, will initially be defined as a matrix-valued function on $]0, \tau]$. Both X and Y are cadlag (right continuous with left-hand limits). We will show in the next section that Y can also be considered as a *multiplicative interval function*, just as we can consider X as a measure, thus as an *additive interval function*.

DEFINITION 1 (The product-limit).

$$Y(t) = \prod_{s \in]0, t]} (\mathbf{1} + X(ds)) = \lim_{\max|t_i - t_{i-1}| \to 0} \prod_i (\mathbf{1} + X(]t_{i-1}, t_i])),$$

where $0 = t_0 < t_1 \cdots < t_n = t$ is a partition of $]0, t]$.

The terms in the product here are to be taken in their natural order.

DEFINITION 2 (The Volterra integral equation [Volterra (1887)]). Y is the unique solution of the equation

$$Y(t) = \mathbf{1} + \int_{s \in]0,t]} Y(s-)X(ds).$$

DEFINITION 3 (The Péano series [Péano (1888)]).

$$Y(t) = \mathbf{1} + \sum_{n=1}^{\infty} \int \cdots \int_{0 < t_1 < \cdots < t_n \le t} X(dt_1) \cdots X(dt_n).$$

Each of these definitions has its own merits and when starting from each there is an existence problem to be solved first. The product-limit definition, which we shall take as primary [following Wall (1953) and especially Mac Nerney (1963)], motivates the notation for the product-integral and suggests many of its properties (we will show that one actually has uniform convergence over points of the partitions). The Volterra integral equation is the historical definition and provides also a vital property of the product-integral; moreover, it seems to be the best starting point when defining the product-integral for semimartingales. It turns out to be equivalent to the Kolmogorov forward equation from Markov processes. Finally, the Péano series, while perhaps intuitively unappealing, is a useful technical tool and helps one to derive efficiently the required results. In particular, its existence problem is very easily solved. This starting point was taken by Johansen (1986) and (for $\prod e^{dX}$) by Schmidt (1971). By the way, the discrete approximation to the product-integral given in the product-limit definition is simply the result of applying the first-order Euler scheme for the approximate numerical solution of the integral equation. (Numerical mathematics contains a large literature under the name of product integration, meaning, however, the integration of the product of two functions.)

In the case $p = 1$, when the matrix-valued measure X becomes an ordinary signed measure, a fourth definition is possible. Denote by X^c the continuous part of X, thus $X^c(t) = X(t) - \sum_{s \le t} X(\{t\})$. This is more generally available in the commutative case, that is, when the matrices $X(B)$, $B \in \mathscr{B}$, all commute.

DEFINITION 4 (Commutative case only).

$$Y(t) = \prod_{s \in]0,t]} (\mathbf{1} + X(\{s\})) \exp(X^c(t)).$$

In particular, when $X(s) = sA$ for all s and for a fixed matrix A, $\prod_{]0,t]}(\mathbf{1} + dX) = \exp(tA)$.

So far we have only defined the product-integral of X over Borel sets of the form $]0,t]$ for some t. We can define it over any $B \in \mathscr{B}$ in the following way. Letting χ_B denote the indicator function of the set B, denote by X_B the measure $dX_B = \chi_B \, dX$. Let Y_B be the product-integral of X_B. Finally, we define

$$Y(B) = \prod_{s \in B} (\mathbf{1} + X(ds)) = Y_B(\tau).$$

(We can similarly define the product-integral of a measurable matrix-valued function H with respect to X via the product-integral of X_H defined by $dX_H = H\,dX$ provided H is integrable with respect to X in the ordinary way.) We now formulate the most important *multiplicative* property of the product-integral:

PROPERTY 1 (Multiplicativity). For any $s < u < t$ we have

$$Y(]s,t]) = Y(]s,u])Y(]u,t]).$$

In the commutative case the multiplicativity property extends to the product-integral over any two disjoint sets.

As we shall see in Section 2, one can construct a one-to-one correspondence between additive and multiplicative interval functions. Thus, we shall take Definition 1 and Property 1 as the basis of our treatment of product-integration. The integrals in Definitions 2 and 3 are Lebesgue–Stieltjes integrals. However, Definition 1 represents the product-integral by means of Riemann–Stieltjes approximating finite products and correspondingly an abstract theory is possible [Mac Nerny (1963)] in which all of the ordinary integrals are also of Riemann–Stieltjes type.

The reader familiar with semimartingale theory will be aware that the equivalence between Definitions 1–4 is well known to hold for (matrix-valued) semimartingales, where the notation $\mathscr{E}(X)$ instead of $\prod(\mathbf{1} + dX)$ is usual; see the beautiful papers of Doléans-Dade (1970) and Emery (1978) in particular. (The integrals are now stochastic integrals and the product-limit result holds "in probability." Definition 4 needs to be modified.) The results are proved using sophisticated functional analysis and appropriate topologies on the space of semimartingales. As far as we know, it is not possible to derive these results in the matrix semimartingale case by a similar elementary approach to the one we shall use. However, in the commutative case one can get some of the way, as we shall see in Section 3.3. Further references are also given there.

2. Basic theory. This section develops the basic theory of the product-integral from the point of view of the duality between multiplicative integration of (real matrix-valued) additive interval functions and additive integration of multiplicative interval functions: Definition 1 (the product-limit) and Property 1 (multiplicativity) are taken as primary. To emphasize this point of view, we change from the notation of Section 1 and use α for an additive and μ for a multiplicative interval function; they play the role of X and Y in the previous part, respectively, via the equivalences $\alpha(s,t) = X(]s,t])$ and $\mu(s,t) = Y(]s,t])$. (We also work on the whole line $]0,\infty]$ instead of just a subinterval $]0,\tau]$.)

We start in Section 2.1 by stating some simple but important algebraic identities for matrices and by defining additive and multiplicative interval functions and their Riemann approximating products and sums over a partition of an interval of the real line. The aim is to show that these approximations converge to limits (called the product-integral and the sum-integral, respectively) when the mesh of the partition (i.e., the length of the largest subinterval) goes to zero. Also, we want to show that these operations are one another's inverses. We shall treat the nonnegative scalar case ($p = 1$, $\alpha = \alpha_0 \geq 0$) in Section 2.2 and show in Section 2.3 that the general matrix case (including the signed scalar case!) follows directly from this using the matrix identities. The treatment closely follows Mac Nerney (1963) except for one point. He proves that the integrals are limits of Riemann sums and products where the limit is taken along refinements of the corrèsponding partitions. We shall also first obtain this result but, following Dobrushin (1953), start with right-continuous functions and thereby connect up with the usual interpretation of additive interval functions as measures and establish the stronger result that the Riemann sums and products converge (uniformly over bounded intervals) if just the mesh of the partition converges to zero. This connection also gives a measure-theoretic interpretation to the further results of Section 2.4, where we treat the equivalence with the alternative definitions of Section 1.

2.1. *The key identities.* The identities presented in the following lemma will reappear in different guises throughout the paper. In fact, (1) and (2) are discrete versions of Kolmogorov's forward and backward equation, respectively, and are special cases of (4), a version of Duhamel's equation, the basis of our differentiability result for the product-integral.

LEMMA 1. *Let $A_1, \dots, A_n$ and $B_1, \dots, B_n$ be $p \times p$ matrices. Then*

$$\prod_{1 \leq i \leq n} (\mathbf{1} + A_i) - \mathbf{1} = \sum_{1 \leq i \leq n} (\mathbf{1} + A_1) \cdots (\mathbf{1} + A_{i-1}) A_i, \tag{1}$$

$$\prod_{1 \leq i \leq n} (\mathbf{1} + A_i) - \mathbf{1} = \sum_{1 \leq i \leq n} A_i (\mathbf{1} + A_{i+1}) \cdots (\mathbf{1} + A_n), \tag{2}$$

$$\begin{aligned} &\prod_{1 \leq i \leq n} (\mathbf{1} + A_i) - \mathbf{1} - \sum_{1 \leq i \leq n} A_i \\ &\quad = \sum_{1 \leq i < j \leq n} A_i (\mathbf{1} + A_{i+1}) \cdots (\mathbf{1} + A_{j-1}) A_j, \end{aligned} \tag{3}$$

$$\begin{aligned} &\prod_{1 \leq i \leq n} (\mathbf{1} + A_i) - \prod_{1 \leq i \leq n} (\mathbf{1} + B_i) \\ &\quad = \sum_{1 \leq i \leq n} (\mathbf{1} + A_1) \cdots (\mathbf{1} + A_{i-1})(A_i - B_i)(\mathbf{1} + B_{i+1}) \cdots (\mathbf{1} + B_n). \end{aligned} \tag{4}$$

PROOF. The relations (1), (2) and (4) can be written as telescoping sums. To prove (3), note that from (1) we get

$$\prod_{1 \leq i \leq n} (\mathbf{1} + A_i) - \mathbf{1} = \sum_{1 \leq j \leq n} \prod_{1 \leq i < j} (\mathbf{1} + A_i) A_j, \tag{5}$$

but (2) implies that

$$\prod_{1\le i<j}(\mathbf{1}+A_i)=\mathbf{1}+\sum_{1\le i<j}A_i(\mathbf{1}+A_{i+1})\cdots(\mathbf{1}+A_{j-1}).$$

Inserting this into (5) gives (3). □

We next define an *additive interval function* $\alpha(s,t)$, $0\le s\le t<\infty$, with values in the $p\times p$ matrices. Such a function has the properties

$$\alpha(s,t)=\alpha(s,u)+\alpha(u,t)\quad\text{for all } s\le u\le t, \tag{6}$$

$$\alpha(s,s)=\mathbf{0}\quad\text{for all } s, \tag{7}$$

$$\alpha(s,t)\to\mathbf{0}\quad\text{as } t\downarrow s \text{ for all } s. \tag{8}$$

Similarly, a *multiplicative interval function* $\mu(s,t)$ is defined by

$$\mu(s,t)=\mu(s,u)\mu(u,t)\quad\text{for all } s\le u\le t, \tag{9}$$

$$\mu(s,s)=\mathbf{1}\quad\text{for all } s, \tag{10}$$

$$\mu(s,t)\to\mathbf{1}\quad\text{as } t\downarrow s \text{ for all } s. \tag{11}$$

Now let $\mathscr{T}=\{t_i,\ i=0,\dots,n\}$ be a partition of $]s,t]$, that is, $s=t_0<t_1<\cdots<t_n=t$. We define $D_i=]t_{i-1},t_i]$, $i=1,\dots,n$, $D=]s,t]$ and $|\mathscr{T}|=\max(t_i-t_{i-1})$, the mesh of the partition. We shall use the notation $\mu(D_i)=\mu(t_{i-1},t_i)$ and $\alpha(D_i)=\alpha(t_{i-1},t_i)$. Define the Riemann sum,

$$\sum_{\mathscr{T}}\Delta(\mu-\mathbf{1})=\sum_{1\le i\le n}(\mu(D_i)-\mathbf{1}), \tag{12}$$

and the Riemann product,

$$\prod_{\mathscr{T}}(\mathbf{1}+\Delta\alpha)=\prod_{1\le i\le n}(\mathbf{1}+\alpha(D_i)). \tag{13}$$

We shall study the limits of these quantities as $|\mathscr{T}|\to 0$ and define thereby additive and multiplicative integrals and show that these are inverse operations. Note that it follows from (6) to (8) that $\alpha(s,t)$ is right continuous in t for fixed s and in s for fixed t. The same is true for $\mu(s,t)$.

2.2. *The nonnegative scalar case.* In this section we take $p=1$ and assume throughout that $\alpha_0(s,t)$ is an additive and nonnegative interval function and that $\mu_0(s,t)$ is a multiplicative interval function which is greater than or equal to 1. The reason for this approach is that the existence and properties of the product-integral are very easy to establish in this case, while the general matrix case, including the signed scalar case, then follows by building on the nonnegative scalar case via a concept of domination. We temporarily define the product-integral of α_0 by

$$\prod_{]s,t]}(1+d\alpha_0)=\sup_{\mathscr{T}}\prod_{\mathscr{T}}(1+\Delta\alpha_0) \tag{14}$$

and the additive integral of μ_0 by

$$\int_{]s,t]}d(\mu_0-1)=\inf_{\mathscr{T}}\sum_{\mathscr{T}}\Delta(\mu_0-1), \tag{15}$$

where $\mathscr{T}$ is an arbitrary partition of $]s,t]$. The next proposition shows that these are also (unique) limits under refinements.

PROPOSITION 1. *If* α_0 *is an additive nonnegative interval function, then* $(s,t) \to \prod_{]s,t]}(1 + d\alpha_0)$ *is a multiplicative function bounded below by* $1 + \alpha_0$ *and above by* $\exp(\alpha_0)$, *and if* μ_0 *is a multiplicative function greater than or equal to* 1, *then* $(s,t) \to \int_{]s,t]} d(\mu_0 - 1)$ *is an additive function bounded above by* $\mu_0 - 1$ *and below by* $\log(\mu_0)$. *Furthermore*,

$$\prod_{]s,t]}(1 + d\alpha_0) = \lim_{\mathscr{T}} \prod_{\mathscr{T}} (1 + \Delta\alpha_0) \tag{16}$$

and

$$\int_{]s,t]} d(\mu_0 - 1) = \lim_{\mathscr{T}} \sum_{\mathscr{T}} \Delta(\mu_0 - 1), \tag{17}$$

where the limits are taken over refinements of $\mathscr{T}$.

PROOF. From the inequalities

$$1 + a + b \le (1 + a)(1 + b) \le \exp(a + b), \qquad a \ge 0, b \ge 0, \tag{18}$$

it follows that $\prod_{\mathscr{T}}(1 + \Delta\alpha_0)$ is increasing over refinements of $\mathscr{T}$ and bounded above by $\exp(\alpha_0)$. Hence the product-integral $\prod_{]s,t]}(1 + d\alpha_0)$ is also bounded above by $\exp(\alpha_0)$. Now for any $\varepsilon > 0$ we can find $\mathscr{T}_\varepsilon$ such that

$$\prod_{\mathscr{T}_\varepsilon}(1 + \Delta\alpha_0) \ge \prod_{]s,t]}(1 + d\alpha_0) - \varepsilon.$$

Since a refinement of $\mathscr{T}_\varepsilon$ increases the left-hand side, but remains below $\prod_{]s,t]}(1 + d\alpha_0)$, we have proved that $\prod_{]s,t]}(1 + d\alpha_0)$ is a limit along refinements of $\prod_{\mathscr{T}}(1 + \Delta\alpha_0)$. Since any two partitions have a common refinement, this limit must be unique. It also follows that $\prod(1 + d\alpha_0)$ is multiplicative. Since the Riemann products are bounded between $1 + \alpha_0(s,t)$ and $\exp(\alpha_0(s,t))$, it follows that the product-integral is bounded between the same quantities and hence satisfies (11).

In just the same way, from

$$0 \le \log(ab) \le a - 1 + b - 1 \le ab - 1, \qquad a \ge 1, b \ge 1, \tag{19}$$

it is seen that $\sum_{\mathscr{T}}\Delta(\mu_0 - 1)$ is decreasing in $\mathscr{T}$, that (17) holds and that $(s,t) \to \int_{]s,t]} d(\mu_0 - 1)$ is a nonnegative additive interval function bounded above by $\mu_0(s,t) - 1$ and below by $\log(\mu_0(s,t))$. □

From (1), (3), (16) and the inequalities

$$e^x - \sum_{k=0}^{n-1} \frac{x^k}{k!} \le e^x \frac{x^n}{n!},$$

$$-\frac{1}{2}(x - 1)^2 \le \log x - x + 1, \qquad x > 1,$$

one deduces immediately that the product and additive integrals satisfy the following inequalities [we already have (20) and (23) from Proposition 1]:

$$\alpha_0(s,t) + 1 \le \prod_{]s,t]} (1 + d\alpha_0) \le \exp(\alpha_0(s,t)); \tag{20}$$

$$\alpha_0(s,t) \le \prod_{]s,t]} (1 + d\alpha_0) - 1 \le \alpha_0(s,t)\exp(\alpha_0(s,t)); \tag{21}$$

$$0 \le \prod_{]s,t]} (1 + d\alpha_0) - 1 - \alpha_0(s,t) \le \tfrac{1}{2}(\alpha_0(s,t))^2 \exp(\alpha_0(s,t)); \tag{22}$$

$$\log(\mu_0(s,t)) \le \int_{]s,t]} d(\mu_0 - 1) \le \mu_0(s,t) - 1; \tag{23}$$

$$-\tfrac{1}{2}(\mu_0(s,t) - 1)^2 \le \int_{]s,t]} d(\mu_0 - 1) - \mu_0(s,t) + 1 \le 0. \tag{24}$$

PROPOSITION 2. *If*

$$\mu_0(s,t) = \prod_{]s,t]} (1 + d\alpha_0), \tag{25}$$

then

$$\alpha_0(s,t) = \int_{]s,t]} d(\mu_0 - 1) \tag{26}$$

and vice versa. Moreover, we then have for any partition $\mathscr{T}$,

$$\begin{aligned} 0 &\le \sum_{\mathscr{T}} \Delta(\mu_0 - 1) - \alpha_0(s,t) \\ &\le \mu_0(s,t) - \prod_{\mathscr{T}} (1 + \Delta\alpha_0) \\ &\le \mu_0(s,t)\Big(\sum_{\mathscr{T}} \Delta(\mu_0 - 1) - \alpha_0(s,t)\Big). \end{aligned} \tag{27}$$

PROOF. Let α_0 be additive and nonnegative and define μ_0 by (25). Then evaluate, by relation (45) of Lemma 1 and the left-hand side of (22),

$$\begin{aligned} 0 &\le \sum_{\mathscr{T}} \Delta(\mu_0 - 1) - \alpha_0(s,t) \\ &= \sum_{1\le i\le n} (\mu_0(D_i) - 1 - \alpha_0(D_i)) \\ &\le \sum_{1\le i\le n} (1 + \alpha_0(D_1)) \cdots (1 + \alpha_0(D_{i-1}))(\mu_0(D_i) - 1 - \alpha_0(D_i)) \\ &\qquad\qquad \times \mu_0(D_{i+1}) \cdots \mu_0(D_n) \\ &= \prod_{1\le i\le n} \mu_0(D_i) - \prod_{1\le i\le n} (1 + \alpha_0(D_i)) \\ &= \mu_0(s,t) - \prod_{\mathscr{T}} (1 + \Delta\alpha_0). \end{aligned}$$

This gives the first part of (27) and taking limits over refinements of $\mathscr{T}$ gives (26). The converse and the rest of (27) is proved similarly: let μ_0 be multiplicative and greater than or equal to 1, and define α_0 by (26). Then we evaluate, by (4) and (23),

$$\begin{aligned}
0 &\le \mu_0(s,t) - \prod_{\mathscr{T}} (1 + \Delta\alpha_0) \\
&= \prod_{1\le i\le n} \mu_0(D_i) - \prod_{1\le i\le n} (1 + \alpha_0(D_i)) \\
&= \sum_{1\le i\le n} (1 + \alpha_0(D_1)) \cdots (1 + \alpha_0(D_{i-1}))(\mu_0(D_i) - 1 - \alpha_0(D_i)) \\
&\qquad\quad \times \mu_0(D_{i+1}) \cdots \mu_0(D_n) \\
&\le \sum_{1\le i\le n} \mu_0(D_1) \cdots \mu_0(D_{i-1})(\mu_0(D_i) - 1 - \alpha_0(D_i))\mu_0(D_{i+1}) \cdots \mu_0(D_n) \\
&\le \mu_0(s,t)\Big(\sum_{\mathscr{T}} \Delta(\mu_0 - 1) - \alpha_0(s,t)\Big).
\end{aligned}$$

Hence, letting $\mathscr{T}$ converge through refinements, we obtain (25). □

We proved in Proposition 1 that the product-integral and the additive integral were limits under refinements. We shall now show that we have an even stronger approximation result, namely, that (16) and (17) hold with $\lim_{\mathscr{T}}$ replaced by $\lim_{|\mathscr{T}|\to 0}$, and moreover, that this holds uniformly in bounded s and t.

First note that an additive nonnegative interval function, thanks to the continuity assumption (8), determines by the identification $\alpha_0(]s,t]) = \alpha_0(s,t)$ a σ-additive measure which is finite on bounded intervals. In the following we shall think of α_0 as a measure. The next lemma has been adapted from Dobrushin [(1953), Lemma 1].

LEMMA 2. *Let θ be a positive measure on $\mathbb{R}_+$ which is finite on bounded intervals, and let $\mathscr{T}$ be a partition of $]s,t]$. If s_i denotes the position of the largest atom in D_i, then*

$$\lim_{|\mathscr{T}|\to 0} \max_{1\le i\le n} \theta(D_i \setminus \{s_i\}) = 0$$

uniformly in $s, t \le u$ for each fixed $u < \infty$.

PROOF. Let $a_1 \ge a_2 \ge \cdots$ be the sizes of the atoms of the measure θ in the interval $]0,u]$, and let $b_1, b_2, \ldots$ be the positions of these atoms. For any $\varepsilon > 0$ we take $n(\varepsilon)$ such that $\Sigma_{n\ge n(\varepsilon)} a_n \le \varepsilon/2$. Now decompose θ into the continuous part θ^c and the discrete part θ^d. Then $\theta^c(]0,v])$ is uniformly continuous on the interval $[0,u]$, and we can hence choose a $\delta_1(\varepsilon)$ such that any interval of length less than or equal to $\delta_1(\varepsilon)$ has θ^c measure less than or equal to $\varepsilon/2$. If the interval is also chosen of length less than $\delta_2(\varepsilon) = \min|b_i - b_j|$ for i and $j \ge n(\varepsilon)$, then the interval can contain at most one of the large atoms $a_1, \ldots, a_{n(\varepsilon)}$. Since s_i is the position of the largest atom in D_i the total mass of the remaining atoms in D_i must be less than or equal to $\varepsilon/2$, by the choice of $n(\varepsilon)$. Thus, for any partition $\mathscr{T}$ of any $]s,t] \subseteq]0,u]$ with $|\mathscr{T}|$ less than $\min(\delta_1(\varepsilon), \delta_2(\varepsilon))$, we have that $\max_{1 \le i \le n} \theta(D_i \setminus \{s_i\}) \le \varepsilon$, which completes the proof. □

LEMMA 3. *Let θ be a nonnegative additive interval function and let $D =]s,t]$. Then for any $u \in D$ we have*

$$\prod_{D}(1 + d\theta) - 1 - \theta(D) \le \theta(D \setminus \{u\})\theta(D)\exp(\theta(D)).$$

PROOF. Let $\prod_{D\setminus\{u\}}(1 + d\theta) = \prod_{]s,u]}(1 + d\theta)\prod_{]u,t]}(1 + d\theta)$. Then

$$\begin{aligned}
0 &\le \prod_{D}(1 + d\theta) - 1 - \theta(D)\\
&= (1 + \theta(\{u\}))\prod_{D\setminus\{u\}}(1 + d\theta) - 1 - \theta(D \setminus \{u\}) - \theta(\{u\}))\\
&= \prod_{D\setminus\{u\}}(1 + d\theta) - 1 - \theta(D \setminus \{u\}) + \theta(\{u\})\Big(\prod_{D\setminus\{u\}}(1 + d\theta) - 1\Big)\\
&\le \tfrac{1}{2}\theta(D \setminus \{u\})^2 \exp(\theta(D \setminus \{u\})) + \theta(\{u\})\theta(D \setminus \{u\})\exp(\theta(D \setminus \{u\}))\\
&\le \theta(D \setminus \{u\})\theta(D)\exp(\theta(D)).
\end{aligned}$$

□

PROPOSITION 3. *If α_0 is an additive nonnegative interval function, then (uniformly over bounded intervals)*

$$\prod_{]s,t]}(1 + d\alpha_0) = \lim_{|\mathscr{T}|\to 0} \prod_{\mathscr{T}}(1 + \Delta\alpha_0). \tag{28}$$

If μ_0 is a multiplicative interval function and $\mu_0 \ge 1$, then (uniformly over bounded intervals)

$$\int_{]s,t]} d(\mu_0 - 1) = \lim_{|\mathscr{T}|\to 0} \sum_{\mathscr{T}} \Delta(\mu_0 - 1). \tag{29}$$

PROOF. Let $M_i = \prod_{D_i}(1 + d\alpha_0)$ and $N_i = 1 + \alpha_0(D_i)$, then M_i and N_i are bounded by $\exp(\alpha_0(D_i))$ and by Lemma 3, $|M_i - N_i| \le \alpha_0(D_i \setminus \{s_i\})\alpha_0(D_i)\exp(\alpha_0(D_i))$. We then get from (4) that

$$\begin{aligned}
0 &\le \prod_{]s,t]}(1 + d\alpha_0) - \prod_{\mathscr{T}}(1 + \Delta\alpha_0)\\
&= \prod_{1\le i\le n} M_i - \prod_{1\le i\le n} N_i = \sum_{1\le i\le n} M_1 \cdots M_{i-1}(M_i - N_i)N_{i+1} \cdots N_n \qquad (30)\\
&\le \max_{1\le i\le n} \alpha_0(D_i \setminus \{s_i\})\alpha_0(D)(\exp(\alpha_0(D)))^2,
\end{aligned}$$

where s_i is the position of the largest atom in D_i. Now, as $|\mathscr{T}| \to 0$, this converges by Lemma 2 uniformly to zero, which proves (28). It follows from (27) that (29) holds. □

2.3. *The general matrix case.* Here we show that the results of the previous section carry over directly from the real, nonnegative case to the general case (including the signed scalar case) by using a concept of domination. Theorem 1 establishes the existence of the product-integral as formulated in Definition 1 of Section 1 and its multiplicative property.

For a $p \times p$ matrix A, we define the norm $|A| = \max_i \sum_j |a_{ij}|$. Note that $|AB| \leq |A|\,|B|$ and that $|A + B| \leq |A| + |B|$. We say that an interval function β with values in the $p \times p$ matrices has bounded variation on $]s,t]$ if

$$|\beta|(s,t) = \sup_{\mathscr{T}} \sum_{1 \leq i \leq n} |\beta(D_i)| \leq c < \infty.$$

We shall say that β *has bounded variation* if $|\beta|(0,t) < \infty$ for each finite t. [Even if the interval function β has bounded variation, the function $t \to \beta(s,t)$ need not be of bounded variation in the usual sense; it will be if β is additive.] We say that β is *dominated* by a real interval function β_0 if $|\beta(s,t)| \leq \beta_0(s,t)$ for all s and t.

Lemma 4. *An additive interval function α is of bounded variation if and only if it is dominated by an additive nonnegative real interval function α_0. A multiplicative interval function μ is such that $\mu - \mathbf{1}$ is of bounded variation if and only if there is a real multiplicative interval function $\mu_0 \geq 1$ such that $\mu - \mathbf{1}$ is dominated by $\mu_0 - 1$.*

Proof. The "if" parts of the lemma are easy. For the "only if," let α be additive and of bounded variation. We then define

$$\alpha_0(s,t) = |\alpha|(s,t) = \sup_{\mathscr{T}} \sum_{1 \leq i \leq n} |\alpha(D_i)|$$

and we are done.

If μ is multiplicative then the function

$$(s,t) \to |\mu - \mathbf{1}|(s,t) = \sup_{\mathscr{T}} \sum_{1 \leq i \leq n} |\mu(D_i) - \mathbf{1}|$$

is not additive, but only super additive [see (19)], that is,

$$|\mu - \mathbf{1}|(s,t) \geq |\mu - \mathbf{1}|(s,u) + |\mu - \mathbf{1}|(u,t).$$

But we can define $\alpha_0(u,t) = |\mu - \mathbf{1}|(0,t) - |\mu - \mathbf{1}|(0,u) \geq |\mu - \mathbf{1}|(u,t)$ and $\mu_0(s,t) = \prod_{]s,t]}(1 + d\alpha_0)$, and then we have, as required,

$$|\mu(s,t) - \mathbf{1}| \leq |\mu - \mathbf{1}|(s,t) \leq \alpha_0(s,t) \leq \mu_0(s,t) - 1. \qquad \square$$

Theorem 1. *Let α be additive and dominated by the additive function α_0; let $\mu_0 = \prod(1 + d\alpha_0)$. Then μ defined by*

$$\mu(s,t) = \prod_{]s,t]} (\mathbf{1} + d\alpha) = \lim_{|\mathscr{T}| \to 0} \prod_{\mathscr{T}} (\mathbf{1} + \Delta\alpha)$$

exists and the limit is uniform in $0 \leq s \leq t \leq u$ for any fixed $u < \infty$. The *interval function μ is multiplicative, $\mu - \mathbf{1}$ is dominated by $\mu_0 - 1$ and $\mu - \mathbf{1} - \alpha$ is dominated by $\mu_0 - 1 - \alpha_0$.*

PROOF. From (3) of Lemma 1 we find

$$\prod_{\mathscr{T}} (\mathbf{1} + \Delta\alpha) - \mathbf{1} - \alpha(D) = \sum_{1 \le i < j \le n} \alpha(D_i)(\mathbf{1} + \alpha(D_{i+1})) \cdots (\mathbf{1} + \alpha(D_{j-1}))\alpha(D_j).$$

Hence,

$$\left| \prod_{\mathscr{T}} (\mathbf{1} + \Delta\alpha) - \mathbf{1} - \alpha(D) \right| \le \prod_{\mathscr{T}} (1 + \Delta\alpha_0) - 1 - \alpha_0(D) \tag{31}$$

since the same identity holds for α_0. Now let $\mathscr{S}$ be any refinement of $\mathscr{T}$ and let $\mathscr{D}_i$ be the corresponding partition of D_i. Then

$$\prod_{\mathscr{S}} (\mathbf{1} + \Delta\alpha) - \prod_{\mathscr{T}} (\mathbf{1} + \Delta\alpha) = \prod_{1 \le i \le n} \left(\prod_{\mathscr{D}_i} (\mathbf{1} + \Delta\alpha) \right) - \prod_{1 \le i \le n} (\mathbf{1} + \alpha(D_i)).$$

Now apply (4) of Lemma 1 and the inequality (31). We find

$$\left| \prod_{\mathscr{S}} (\mathbf{1} + \Delta\alpha) - \prod_{\mathscr{T}} (\mathbf{1} + \Delta\alpha) \right| \le \prod_{\mathscr{S}} (1 + \Delta\alpha_0) - \prod_{\mathscr{T}} (1 + \Delta\alpha_0). \tag{32}$$

But the product-integral μ_0 of α_0 exists and equals the limit as $|\mathscr{T}| \to 0$ of the Riemann products; hence, the same result holds for α. Taking the limit as $|\mathscr{T}| \to 0$ of (31) shows that $\mu - \mathbf{1} - \alpha$ is dominated by $\mu_0 - 1 - \alpha_0$. In the same way, by using (1) of Lemma 1 it is proved that $\mu - \mathbf{1}$ is dominated by $\mu_0 - 1$. Combining (30) and (32), we get

$$\begin{aligned} 0 &\le \left| \prod_{]s,t]} (\mathbf{1} + d\alpha) - \prod_{\mathscr{T}} (\mathbf{1} + \Delta\alpha) \right| \\ &\le \prod_{]s,t]} (1 + d\alpha_0) - \prod_{\mathscr{T}} (1 + \Delta\alpha_0) \\ &\le \max_i \alpha_0(D_i \setminus \{s_i\}) \alpha_0(D) (\exp(\alpha_0(D)))^2. \end{aligned}$$

Thus, by Lemma 2, $\prod_{\mathscr{T}}(\mathbf{1} + \Delta\alpha) \to \prod_{]s,t]}(\mathbf{1} + d\alpha)$ as $|\mathscr{T}| \to 0$ uniformly in $0 \le s \le t \le u$ for any fixed u. □

We continue by proving in Theorem 2 the converse result for the additive integral of a multiplicative function (minus **1**) and, finally, by showing in Theorem 3 that the two operations are inverses.

THEOREM 2. *Let μ and μ_0 be multiplicative and $\mu - \mathbf{1}$ be dominated by $\mu_0 - 1$; let $\alpha_0 = \int d(\mu_0 - 1)$. Then α defined by*

$$\alpha(s,t) = \int_{]s,t]} d(\mu - \mathbf{1}) = \lim_{|\mathscr{T}| \to 0} \sum_{\mathscr{T}} \Delta(\mu - \mathbf{1})$$

exists and the limit is uniform in $0 \le s \le t \le u$ for any fixed $u < \infty$. The *interval function α is additive, α is dominated by α_0 and $\mu - \mathbf{1} - \alpha$ is dominated by $\mu_0 - 1 - \alpha_0$.*

PROOF. Let us evaluate

$$\mu(D) - \mathbf{1} - \sum_{\mathscr{T}} \Delta(\mu - \mathbf{1}) = \prod_{1 \le i \le n} \mu(D_i) - \mathbf{1} - \sum_{1 \le i \le n} (\mu(D_i) - \mathbf{1}).$$

By Lemma 1 we get for $A_i = \mu(D_i) - \mathbf{1}$ that this equals

$$\sum_{1 \le i < j \le n} (\mu(D_i) - \mathbf{1})\mu(D_{i+1}) \cdots \mu(D_{j-1})(\mu(D_j) - \mathbf{1})$$

$$= \sum_{1 \le i < j \le n} (\mu(D_i) - \mathbf{1})\mu(D_{i+1} \cup \cdots \cup D_{j-1})(\mu(D_j) - \mathbf{1}),$$

which is bounded in norm by the same expression with μ replaced by μ_0. Hence,

$$\left| \mu(D) - \mathbf{1} - \sum_{\mathscr{T}} \Delta(\mu - \mathbf{1}) \right| \le \mu_0(D) - 1 - \sum_{\mathscr{T}} \Delta(\mu_0 - 1). \tag{33}$$

Now let $\mathscr{S}$ be a refinement of $\mathscr{T}$ and let $\mathscr{D}_i$ be the corresponding partition of D_i. Then

$$\sum_{\mathscr{T}} \Delta(\mu - \mathbf{1}) - \sum_{\mathscr{S}} \Delta(\mu - \mathbf{1}) = \sum_{1 \le i \le n} \left(\mu(D_i) - \mathbf{1} - \sum_{\mathscr{D}_i} \Delta(\mu - \mathbf{1}) \right),$$

hence

$$\left| \sum_{\mathscr{T}} \Delta(\mu - \mathbf{1}) - \sum_{\mathscr{S}} \Delta(\mu - \mathbf{1}) \right| \le \sum_{\mathscr{T}} \Delta(\mu_0 - 1) - \sum_{\mathscr{S}} \Delta(\mu_0 - 1).$$

Now $\alpha_0 = \int d(\mu_0 - 1)$ exists and, moreover, equals the limit as $|\mathscr{T}| \to 0$ of the Riemann sums, which shows that the same results hold for μ. The domination of α by α_0 now follows trivially from the definitions. Taking the limit as $|\mathscr{T}| \to 0$ in (33) gives the domination of $\mu - \mathbf{1} - \alpha$ by $\mu_0 - 1 - \alpha_0$. The uniformity follows from (27) and Theorem 1. □

THEOREM 3. *If α is additive and of bounded variation and μ is defined by*

$$\mu(s,t) = \prod_{]s,t]} (\mathbf{1} + d\alpha), \tag{34}$$

then

$$\alpha(s,t) = \int_{]s,t]} d(\mu - \mathbf{1}). \tag{35}$$

Similarly, if μ is multiplicative and $\mu - \mathbf{1}$ of bounded variation, and if α is defined by (35), *then* (34) *holds.*

PROOF. Assume α to be additive and of bounded variation and define μ by (34). Let α_0 dominate α and define μ_0 to be the product integral of α_0. Then by Theorem 1, $|\mu - \mathbf{1}| \le \mu_0 - 1$, which shows that $\mu - \mathbf{1}$ is of bounded variation, and that $\int d(\mu - \mathbf{1})$ exists. We must show that it coincides with α. Now evaluate

$$\sum_{\mathscr{T}} \Delta(\mu - \mathbf{1}) - \alpha(s,t) = \sum_{1 \le i \le n} (\mu(D_i) - \mathbf{1} - \alpha(D_i)),$$

which by Theorem 1 is dominated by

$$\sum_{1 \le i \le n} (\mu_0(D_i) - 1 - \alpha_0(D_i)) = \sum_{\mathscr{T}} \Delta(\mu_0 - 1) - \alpha_0(s,t),$$

which goes to zero by Proposition 2.

Similarly, if μ is multiplicative and $\mu - \mathbf{1}$ dominated by $\mu_0 - 1$, then by Theorem 2, $\alpha = \int d(\mu - \mathbf{1})$ is dominated by $\alpha_0 = \int d(\mu_0 - 1)$ and hence

$$\begin{aligned}\mu(D) - \prod_{\mathscr{T}} (\mathbf{1} + \Delta\alpha) &= \prod_{1 \le i \le n} (\mu(D_i)) - \prod_{1 \le i \le n} (\mathbf{1} + \alpha(D_i)) \\ &= \sum_{1 \le i \le n} \mu(D_1) \cdots \mu(D_{i-1})(\mu(D_i) - \mathbf{1} - \alpha(D_i)) \\ &\qquad \times (\mathbf{1} + \alpha(D_{i+1})) \cdots (\mathbf{1} + \alpha(D_n)),\end{aligned}$$

which by Theorem 2 is dominated by

$$\begin{aligned}&\sum_{1 \le i \le n} \mu_0(D_1) \cdots \mu_0(D_{i-1})(\mu_0(D_i) - 1 - \alpha_0(D_i)) \\ &\qquad \times (1 + \alpha_0(D_{i+1})) \cdots (1 + \alpha_0(D_n)) \\ &= \prod_{1 \le i \le n} (\mu_0(D_i)) - \prod_{1 \le i \le n} (1 + \alpha_0(D_i)) \\ &= \mu_0(D) - \prod_{\mathscr{T}} (1 + \Delta\alpha_0),\end{aligned}$$

which tends to zero by Proposition 2. □

We conclude this section with some comments on product-integration over general spaces. We shall return to this subject again in Sections 3.3 and 4.3. In the scalar (commutative) case, if α_0 is a finite signed measure on an arbitrary measurable space, one can still define its product integral by (16)—that is, as a limit over refinements—with the interval $]s,t]$ replaced by an arbitrary measurable set and a partition $\mathscr{T}$ being a finite collection of disjoint measurable subsets. The proofs of all the previous results (excepting Lemmas 2 and 3 and Proposition 3, and elsewhere replacing limits as the mesh of a partition goes to zero by limits under refinements) go through without any changes; we derive the signed measure case from the nonnegative case just as was done previously for the matrix case. The product-integral over any two disjoint sets is multiplicative.

One can also generalize the results concerning limits as the mesh of a partition goes to zero when α_0 is a signed measure on the Borel sets of a separable metric space, finite on compacts. The interval $]s,t]$ is replaced by an arbitrary measurable subset of a compact set. A partition is again a finite collection of disjoint subsets. Its mesh is the maximum of the *diameters* of its components, where the diameter of a set is the supremum of the distance between any two of its points. The proof of Lemma 2 has to be only slightly modified to deal with this situation and Lemma 3 not at all (in fact, it already used commutativity). In Lemma 2, working with balls instead of with intervals, we use the fact that by compactness the supremum over x in a compact set of the θ^c measure of an ε-ball around x tends to zero with ε.

We leave it to the reader to derive the commutative case result (Definition 4 of Section 1),

$$\prod_{]s,t]}(1 + d\alpha_0) = \prod_{u\in]s,t]}(1 + \alpha_0(\{u\}))\exp(\alpha_0^c(s,t)),$$

where α_0^c is the continuous part of α_0. This can be proved by writing the product-integral of α_0 over $]s,t]$ (with respect to limits over all measurable partitions) as the product of the product-integrals over the set of atoms of α_0 in $]s,t]$ and over its complement, respectively, and by then doing some easy analysis. One can show that in $\mathbb{R}_+^k$ the various definitions coincide—thus, that in our preliminary definition (14), the supremum on the right-hand side does not increase when taken over a wider class of partitions [and similarly for (15)]. The main element in the simple proof of this is the exponential inequality (20).

Another possibility is to let the mesh of a partition be the maximum α_0 measure of an element of the partition less its α_0-largest atom. Again, the product-limit as the mesh of the partition goes to zero exists and equals our product-integral.

2.4. *The Péano series and the Volterra integral equation.* In this section we finally derive the equivalence of Definition 1 in Section 1 with Definitions 2 and 3 (Theorems 4 and 5). Also, we prove a central result (Theorem 6) on the difference of two product-integrals, called the Duhamel equation. Let α be an additive interval function on $]0,\infty[$ with dominating measure α_0. Then each of the entries α_{ij} is a finite measure on the Borel subsets of bounded subsets of $\mathbb{R}_+$, and we can define a product matrix measure on bounded subsets of $\mathbb{R}_+^n$ starting from

$$\alpha^{(n)}(D_1 \times \cdots \times D_n) = \alpha(D_1)\cdots\alpha(D_n).$$

Note that

$$\begin{aligned}|\alpha^{(n)}(D_1 \times \cdots \times D_n)| &\le \prod_{1\le i\le n}|\alpha(D_i)| \le \prod_{1\le i\le n}\alpha_0(D_i)\\ &= \alpha_0^{(n)}(D_1 \times \cdots \times D_n).\end{aligned}$$

Thus, $\alpha^{(n)}$ is dominated by $\alpha_0^{(n)}$, the usual product measure. For an interval D let $U(D;n)$ denote the subset of $\mathbb{R}_+^n$

$$U(D;n) = \{(u_1,\dots,u_n) \in D^n : u_1 < \cdots < u_n\}.$$

Then the Péano series is defined by

$$\begin{aligned}P(D;\alpha) &= \mathbf{1} + \sum_{n=1}^{\infty}\alpha^{(n)}(U(D;n))\\ &= \mathbf{1} + \sum_{n=1}^{\infty}\int\cdots\int_{u_1<\cdots<u_n;\, u_i\in D}\alpha(du_1)\cdots\alpha(du_n).\end{aligned} \tag{36}$$

Note that the series is dominated by

$$P(D;\alpha_0) = 1 + \sum_{n=1}^{\infty} \alpha_0^{(n)}(U(D;n))$$
$$\leq 1 + \sum_{n=1}^{\infty} (\alpha_0(D))^n/n! = \exp(\alpha_0(D)),$$

which shows the convergence of the series as well as the inequalities

$$|P(D;\alpha)| \leq P(D;\alpha_0) \leq \exp(\alpha_0(D)), \tag{37}$$

$$|P(D;\alpha) - \mathbf{1}| \leq P(D;\alpha_0) - 1 \leq \alpha_0(D)\exp(\alpha_0(D)), \tag{38}$$

$$\begin{aligned}|P(D;\alpha) - \mathbf{1} - \alpha(D)| &\leq P(D;\alpha_0) - 1 - \alpha_0(D) \\ &\leq \tfrac{1}{2}(\alpha_0(D))^2 \exp(\alpha_0(D)).\end{aligned} \tag{39}$$

We write $U(s,t;n)$ for the set $U(]s,t];n)$ and $P(s,t;\alpha)$ for the interval function $P(\,]s,t];\alpha)$.

Proposition 4. *The interval function $P(s,t;\alpha)$ is multiplicative.*

Proof. Define the set in $\mathbb{R}_+^n$,

$$U(s,u,t;i;n) = \{s < u_1 < \cdots < u_i \leq u < u_{i+1} < \cdots < u_n \leq t\}, \quad i = 0,\ldots,n,$$

with the obvious modifications for $i = 0$ and $i = n$. Then

$$U(s,t;n) = \{s < u_1 < \cdots < u_n \leq t\} = \bigcup_{0\leq i\leq n} U(s,u,t;i;n).$$

Hence,

$$\alpha^{(n)}(U(s,t;n)) = \sum_{0\leq i\leq n} \alpha^{(n)}(U(s,u,t;i;n))$$
$$= \sum_{0\leq i\leq n} \alpha^{(i)}(U(s,u;n))\alpha^{(n-i)}(U(u,t;n)).$$

Summing over n gives the desired result. The exponential inequality (38) shows that $P(s,t;\alpha)$ is right continuous in s and t. □

Theorem 4. *The Péano series is equal to the product-integral*

$$P(s,t;\alpha) = \prod_{]s,t]}(\mathbf{1} + d\alpha).$$

Proof. First note that

$$\begin{aligned}P(s,t;\alpha) &- \prod_{\mathscr{T}}(\mathbf{1} + \Delta\alpha) \\ &= \prod_{1\leq i\leq n} P(D_i;\alpha) - \prod_{1\leq i\leq n}(\mathbf{1} + \alpha(D_i)) \\ &= \sum_{1\leq i\leq n} (\mathbf{1} + \alpha(D_1))\cdots(\mathbf{1} + \alpha(D_{i-1}))\cdots \\ &\qquad \times (P(D_i;\alpha) - \alpha(D_i) - \mathbf{1})P(D_{i+1};\alpha)\cdots P(D_n;\alpha),\end{aligned}$$

which is dominated by the same sum with α replaced by α_0. Now the proof of Lemma 3 goes through with $\prod(1 + d\theta)$ replaced by $P(\cdot\,; \alpha_0)$, since only multiplicativity and the exponential inequalities are used; note that $P(\{s\}; \alpha_0) = 1 + \alpha_0(\{s\})$. Thus, we find that

$$\left| P(s,t;\alpha) - \prod_{\mathscr{T}} (1 + \Delta\alpha) \right| \le \max_{1 \le i \le n} \alpha_0(D_i \setminus \{s_i\}) \alpha_0(s,t) (\exp(\alpha_0(s,t)))^2,$$

where we take s_i to be the largest atom of α_0 in the interval D_i. Now by Lemma 2 the right-hand side tends to zero with $|\mathscr{T}|$ and hence the result is proved. □

PROPOSITION 5.

$$\prod_{]s,t]} (1 + d\alpha) - 1 = \int_{]s,t]} \prod_{]s,u[} (1 + d\alpha) \alpha(du) \quad (\textit{the forward equation}) \tag{40}$$

$$= \int_{]s,t[} \alpha(du) \prod_{]u,t]} (1 + d\alpha) \quad (\textit{the backward equation}), \tag{41}$$

where the integrals are Lebesgue–Stieltjes integrals.

PROOF. Using Fubini's theorem on the $(n + 1)$th term of the Péano series, we get

$$\begin{aligned} &\alpha^{(n+1)}(U(s,t;n+1)) \\ &\quad = \int_{]s,t]} \alpha^{(n)}(\{(u_1,\dots,u_n) \colon (u_1,\dots,u_n,u) \in U(s,t;n+1)\}) \alpha(du) \\ &\quad = \int_{]s,t]} \alpha^{(n)}(U(s,u-;n)) \alpha(du). \end{aligned}$$

Summing over n gives the forward equation. The backward equation is proved in just the same way. □

THEOREM 5 (The forward and backward equations). *Let $\beta(s,t)$ be any interval function which is right continuous with left limits in both variables and which satisfies either of the two equations*

$$\beta(s,t) - 1 = \int_{]s,t]} \beta(s,u-) \alpha(du),$$

$$\beta(s,t) - 1 = \int_{]s,t]} \alpha(du) \beta(u,t).$$

Then $\beta(s,t) = \prod_{]s,t]}(1 + d\alpha)$ (and the converse is also true).

PROOF. We have already proved the converse in Proposition 5. Suppose β satisfies the forward equation (for fixed s, for all $t > s$). By the cadlag property, β is then bounded for t in a bounded interval. Let $P^{(n)}(s,t;\alpha) =$

$\mathbf{1} + \sum_{1 \le k \le n} \alpha^{(k)}(U(s,t;k))$. Then

$$P^{(n)}(s,t;\alpha) = \mathbf{1} + \int_{]s,t]} P^{(n-1)}(s,u-;\alpha)\alpha(du)$$

and hence

$$\beta(s,t) - P^{(n)}(s,t;\alpha) = \int_{]s,t]} \big(\beta(s,u-) - P^{(n-1)}(s,u-;\alpha)\big)\alpha(du).$$

From the first equation for β we find

$$|\beta(s,t) - \mathbf{1}| \le \sup_{s<u\le t} |\beta(s,u-)| \alpha_0(s,t).$$

It is not difficult from this to show by induction over n that

$$\big|\beta(s,t) - P^{(n)}(s,t;\alpha)\big| \le \sup_{s<u\le t} |\beta(s,u-)| \big(\alpha_0(s,t)\big)^n.$$

For $n \to \infty$ we get that

$$\beta(s,t) = \lim_{n\to\infty} P^{(n)}(s,t;\alpha) = P(s,t;\alpha) = \prod_{]s,t]} (\mathbf{1} + d\alpha).$$

The result on the uniqueness of the solution of the backward equation (for fixed t, for all $s < t$) is proved in just the same way. □

Note that the forward equation generalizes the fundamental identity (1), whereas the backward equation is a generalization of (2). In a similar way we can generalize (4).

THEOREM 6 (Duhamel's equation). *Let α_1 and α_2 be additive. Then*

$$\prod_{]s,t]} (\mathbf{1} + d\alpha_1) - \prod_{]s,t]} (1 + d\alpha_2) = \int_{]s,t]} \prod_{]s,u[} (\mathbf{1} + d\alpha_1)(\alpha_1 - \alpha_2)(du) \prod_{]u,t]} (\mathbf{1} + d\alpha_2). \tag{42}$$

PROOF. Consider the measure $\alpha_{1,2}^{(n,m)}$ on $\mathbb{R}_+^{n+m}$ defined by

$$\alpha_{1,2}^{(n,m)}(A_1 \times \cdots \times A_n \times B_1 \times \cdots \times B_m) = \alpha_1(A_1) \cdots \alpha_1(A_n)\alpha_2(B_1) \cdots \alpha_2(B_m).$$

By applying Fubini's theorem we obtain that

$$\begin{aligned}\alpha_{1,2}^{(n,m)}&(U(s,t;n+m)) \\ &= \int_{]s,t]} \alpha_1^{(n-1)}(U(s,u-;n-1))\alpha_1(du)\alpha_2^{(m)}(U(u,t;m)) \\ &= \int_{]s,t]} \alpha_1^{(n)}(U(s,u-;n))\alpha_2(du)\alpha_2^{(m-1)}(U(u,t;m-1)).\end{aligned}$$

Summing over $n \geq 1$ and $m \geq 1$, we get

$$\int_{]s,t[} \prod_{]s,u[} (1 + d\alpha_1)\alpha_1(du)\Big(\prod_{]u,t]} (1 + d\alpha_2) - 1\Big)$$

$$= \int_{]s,t[} \Big(\prod_{]s,u[} (1 + d\alpha_1) - 1\Big)\alpha_2(du) \prod_{]u,t]} (1 + d\alpha_2),$$

which is the desired relation by application of (40) and (41). □

One may derive Theorem 4 from Theorem 3 by showing that the additive integral of $P(\cdot, \cdot\,; \alpha) - 1$ equals α. Also, Theorems 5 and 6 can be derived directly from Lemma 1 and the limit results of Theorems 1 and 2 by dominated convergence arguments.

3. Further properties of the product-integral.

3.1. *Continuity and differentiability.* From the Duhamel equation (42) and the exponential inequality (20) it is clear that the product-integral is a continuous functional from additive to multiplicative interval functions, where continuity is with respect to the variation norm on bounded intervals. It is rather less clear that the functional is also continuous and even differentiable with respect to the supremum norm, provided the variation is uniformly bounded. By differentiability we mean here Hadamard or compact differentiability, which is intermediate between the more familiar Gâteaux (directional) and Fréchet (bounded) differentiability and exactly atuned to the functional version of the δ-method, a basic and, in principle, elementary tool of large-sample statistical theory [see Reeds (1976) and Gill (1989)]. This will be illustrated in several parts of Section 4.

To begin with we give definitions of the two norms. We work on a fixed bounded interval $]0, \tau]$. As in Section 2, α is an additive and μ a multiplicative matrix-valued interval function. The variation norm of an interval function β is simply its variation over this interval (see Section 2.3),

$$\|\beta\|_v = |\beta|(0, \tau),$$

while the supremum norm is given by

$$\|\beta\|_\infty = \sup_{0 \leq s \leq t \leq \tau} |\beta(s,t)|.$$

We also recall the integration by parts formula for cadlag (right continuous with left-hand limits) bounded variation matrix functions H and K,

$$(HK)(]s,t]) = H(t)K(t) - H(s)K(s) = \int_{]s,t]} H_-(dK) + \int_{]s,t]} (dH)K.$$

If one of H or K does not have bounded variation, we use this formula to define the integral with respect to it in terms of the other integral. Considering H also as an interval function via $H(s,t) = H(t) - H(s)$, the formula can be applied to prove the following lemma.

LEMMA 5. *Let H, K and U be cadlag matrix-valued functions such that K and U are of bounded variation, whereas H may have unbounded variation. Then*

$$\left\| \int (dH) K \right\|_\infty \le 2\|H\|_\infty \|K\|_v, \tag{43}$$

$$\left\| \int U(dH) \right\|_\infty \le 2\|H\|_\infty \|U\|_v, \tag{44}$$

and hence

$$\left\| \int U(dH) K \right\|_\infty \le 4\|H\|_\infty \|U\|_v \|K\|_v. \tag{45}$$

PROOF. The integrals with respect to H are defined by partial integration. So

$$\left\| \int (dH) K \right\|_\infty \le \|HK\|_\infty + \|H\|_\infty \|K\|_v \le 2\|H\|_\infty \|K\|_v,$$

since $\|K\|_\infty \le \|K\|_v$. The relations (44) and (45) are proved similarly. □

We shall now establish an expansion of the product integral as a function of the integrand.

PROPOSITION 6. *Let α and β be additive interval functions of bounded variation. Then*

$$\begin{aligned}
\prod_{]s,t[} (1 + d(\alpha+\beta)) &- \prod_{]s,t[} (1 + d\alpha) \\
&= \sum_{m=1}^{n} \int \cdots \int_{s<u_1<\cdots<u_m<u_{m+1}=t} \prod_{]s,u_1[} (1 + d\alpha) \\
&\qquad \times \prod_{i=1}^{m} \left(\beta(du_i) \prod_{]u_i,u_{i+1}[} (1 + d\alpha) \right) \\
&\quad + \int \cdots \int_{s<u_0<\cdots<u_n<u_{n+1}=t} \prod_{]s,u_0[} (1 + d(\alpha+\beta)) \\
&\qquad \times \prod_{i=0}^{n} \left(\beta(du_i) \prod_{]u_i,u_{i+1}[} (1 + d\alpha) \right).
\end{aligned}$$

In particular, we get for $n = 0$ the Duhamel equation,

$$\begin{aligned}
&\prod_{]s,t[} (1 + d(\alpha+\beta)) - \prod_{]s,t[} (1 + d\alpha) \\
&\qquad = \int_{s<u<t} \prod_{]s,u[} (1 + d(\alpha+\beta)) \beta(du) \prod_{]u,t[} (1 + d\alpha),
\end{aligned} \tag{46}$$

and for $n = 1$ we get the first-order expansion with a remainder term,

$$\prod_{]s,t[}(1 + d(\alpha + \beta)) - \prod_{]s,t[}(1 + d\alpha)$$
$$- \int_{s<u<t} \prod_{]s,u[}(1 + d\alpha)\beta(du) \prod_{]u,t[}(1 + d\alpha)$$
$$= \iint_{s<u<v<t} \prod_{]s,u[}(1 + d(\alpha + \beta))\beta(du) \prod_{]u,v[}(1 + d\alpha)\beta(dv) \prod_{]v,t[}(1 + d\alpha).$$

PROOF. The relations follow by repeated application of the Duhamel equation. □

Continuity and differentiability of any order of the product-integral in the variation norm follow easily from (46) and (47) but we shall prove much stronger results in the supremum norm.

THEOREM 7 (Continuity of the product-integral in supremum norm). *Let $\alpha^{(n)}$, $n = 1, 2, \dots$ be a sequence of additive interval functions on $]0, \tau]$ such that*

$$\|\alpha^{(n)} - \alpha\|_\infty \to 0 \quad as\ n \to \infty,$$
$$\limsup \|\alpha^{(n)}\|_v < \infty,$$

for some interval function α which is consequently also additive and of bounded variation. Then, defining $\mu^{(n)} = \prod(1 + d\alpha^{(n)})$, $\mu = \prod(1 + d\alpha)$, we have

$$\|\mu^{(n)} - \mu\|_\infty \to 0 \quad as\ n \to \infty.$$

PROOF. For $\beta = \alpha^{(n)} - \alpha$ we get from (46), using the inequality (45), that

$$\|\mu^{(n)} - \mu\|_\infty \le 4\|\mu^{(n)}\|_v \|\mu\|_v \|\alpha^{(n)} - \alpha\|_\infty.$$

Now the Kolmogorov equation (40) and the exponential inequality (20) imply that

$$\|\mu^{(n)}\|_v \le \|\mu^{(n)}\|_\infty \|\alpha^{(n)}\|_v \le \exp(\|\alpha^{(n)}\|_\infty)\|\alpha^{(n)}\|_v,$$

which are bounded by assumption. This shows that $\|\mu^{(n)} - \mu\|_\infty \to 0$ as $n \to \infty$, and hence that the product-integral is continuous in the supremum norm. □

We further refine this result to a differentiability result as follows: Let $\alpha^{(n)}$, α, $\mu^{(n)}$ and μ be as in Theorem 7 and suppose, moreover, that $\alpha^{(n)} = \alpha + t_n h_n$, where t_n is a sequence of real positive numbers, $t_n \to 0$ as $n \to \infty$, and where h_n is a sequence of additive interval functions converging in supremum norm to an interval function h which must also be additive but may not be of bounded variation. So we have

$$\limsup \|\alpha^{(n)}\|_v < \infty \quad \text{and} \quad t_n^{-1}(\alpha^{(n)} - \alpha) = h_n \to h \quad \text{in supremum norm.}$$

The mapping $\mathscr{P}$: $\alpha \to \mu = \prod(1 + d\alpha)$, restricted to a set of α with uniformly bounded *variation* norm, is then *supremum norm* Hadamard differentiable if we correspondingly have

$$(48)\qquad t_n^{-1}\big(\mathscr{P}(\alpha^{(n)}) - \mathscr{P}(\alpha)\big) \to d\mathscr{P}(\alpha)\cdot h \quad \text{in supremum norm},$$

where $d\mathscr{P}(\alpha)$, the derivative of $\mathscr{P}$ at the point α, is some continuous linear mapping from the space of additive matrix-valued interval functions on $]0, \tau]$ to the space of interval functions, both endowed with the supremum norm. [See Gill (1989), Lemma 1, for the propriety of making this restriction on the domain of $\mathscr{P}$.] To find out what the derivative will be [its form is actually already given for special h as the last term on the left-hand side of (47)], we evaluate the left-hand side of (48) with the Duhamèl equation and obtain

$$(49)\quad \big(t_n^{-1}(\mathscr{P}(\alpha^{(n)}) - \mathscr{P}(\alpha))\big)(s,t) = \int_{]s,t]} \mu^{(n)}(s, u-)h_n(du)\mu(u,t),$$

We expect (49) to converge to $\int_{]s,t]}\mu(s, u-)h(du)\mu(u,t)$, where the integral with respect to h will have to be defined by application (twice) of the integration-by-parts formula; recall that h itself may not be of bounded variation. This supposed limit then does define, at the given point α, a continuous linear function of h, which we denote, in anticipation of the desired result, by $(d\mathscr{P}(\alpha)\cdot h)(s,t)$; thus, $(d\mathscr{P}(\alpha)\cdot h)$ is again an additive interval function.

THEOREM 8 (Compact differentiability of the product-integral with respect to the supremum norm). *Consider the product-integral as a mapping $\mathscr{P}$ from the set of additive interval functions on $]0, \tau]$ with variation bounded by the constant c to the space of interval functions on $]0, \tau]$, both domain and range endowed with the supremum norm. Let α be given and define $\mu = \mathscr{P}(\alpha) = \prod(1 + d\alpha)$. Then $\mathscr{P}$ is compactly differentiable at α with derivative $d\mathscr{P}(\alpha)$ given by*

$$(d\mathscr{P}(\alpha)\cdot h)(s,t) = \int_{]s,t]} \mu(s, u-)h(du)\mu(u,t),$$

where the integral with respect to h is defined by the integration-by-parts formula.

PROOF. We just need to prove the convergence of (49) to the limit previously described. The inequality (45) and the relation (47) show that

$$(50)\quad \begin{aligned} &\|\mu^{(n)} - \mu - \int \mu\, d(\alpha^{(n)} - \alpha)\mu\|_\infty \\ &\quad \le 4\|\mu^{(n)}\|_v \|\mu\|_v \left\| \iint_{s<u<v<t} \beta(du)\mu(u, v-)\beta(dv) \right\|_\infty, \end{aligned}$$

where $\beta = \alpha^{(n)} - \alpha$. In order to continue the evaluation we need the following expression for the term $\mu(u, v-)$, which is found by application of the forward and the backward Kolmogorov equations,

$$\mu(u,v-) = \mathbf{1} + \int_{]u,v[} d\alpha + \iint_{u<x<y<v} \alpha(dx)\mu(x,y-)\alpha(dy). \tag{51}$$

Now we insert (51) into the last integral of (50) and find

$$\begin{aligned}
&\iint_{s<u<v<t} \beta(du)\mu(u,v-)\beta(dv) \\
&\quad = \iint_{s<u<v<t} \beta(du)\beta(dv) + \iiint_{s<u<x<v<t} \beta(du)\alpha(dx)\beta(dv) \\
&\qquad + \iiiint_{s<u<x<y<v<t} \beta(du)\alpha(dx)\mu(x,y-)\alpha(dy)\beta(y,t-).
\end{aligned}$$

This shows that the following bound holds:

$$\left\|\iint d\beta\,\mu\,d\beta\right\|_\infty \le \left\|\iint d\beta\,d\beta\right\|_\infty + \|\beta\|_\infty^2\|\alpha\|_v + \|\beta\|_\infty^2\|\mu\|_\infty\|\alpha\|_v^2. \tag{52}$$

Finally,

$$\iint d\beta\,d\beta = \beta(s,t-)^2 - \sum (\Delta\beta)^2. \tag{53}$$

Now apply the representation $\beta = \alpha^{(n)} - \alpha = t_n h_n$. By combining (50), (52) and (53) we find $\|t_n^{-1}(\mu^{(n)} - \mu) - \int \mu\, dh_n \mu\|_\infty$ can be bounded by $c_1 t_n(\sum(\Delta h_n)^2 + c_2)$ for some constants c_1 and c_2. The second term tends to zero since $t_n \to 0$. If h_n has uniformly bounded variation, then the first term also can be bounded by $t_n 2\|h_n\|_\infty\|h_n\|_v$, which also tends to zero. But, more importantly, we have to consider the situation where the variation of h_n tends to infinity but still $t_n\|h_n\|_v$ is bounded. In this case we write

$$t_n \int \Delta h_n\, dh_n = \int h_n\, d\alpha^{(n)} - \int h_n\, d\alpha - \int h_{n-}\, d\alpha^{(n)} + \int h_{n-}\, d\alpha.$$

For the first two terms on the right-hand side, we have

$$\begin{aligned}
\left\|\int h_n\, d(\alpha^{(n)} - \alpha)\right\|_\infty &\le \left\|\int (h_n - h_m)\, d(\alpha^{(n)} - \alpha)\right\|_\infty + \left\|\int h_m\, d(\alpha^{(n)} - \alpha)\right\|_\infty \\
&\le \|h_n - h_m\|_\infty\|\alpha^{(n)} - \alpha\|_v + 2\|h_m\|_v\|\alpha^{(n)} - \alpha\|_\infty.
\end{aligned}$$

Now let first $n \to \infty$ and then $m \to \infty$, and let these terms converge to zero. Precisely the same argument shows that the remaining terms $\int h_{n-}\, d(\alpha^{(n)} - \alpha)$ also converge to zero in supremum norm. □

The last lines of the proof actually form a version of the proof of the Helly–Bray lemma; see Gill [(1989), Lemma 3] for a complete and simple proof in a similar context.

The theorem also applies directly to the mapping $X \to Y = \prod(\mathbf{1} + dX)$ from the usual Skorohod space $D(0,\tau)^{p\times p}$ to itself under the supremum norm when we represent α by the cadlag function $X(t) = \alpha(0, t)$, since we have [for X with $X(0) = \mathbf{0}$]

$$\|X\|_\infty \le \|\alpha\|_\infty \le 2\|X\|_\infty$$

and similarly for μ and Y provided Y is bounded away from zero.

3.2. *Gronwall inequalities, inhomogeneous equations, anticipating integrands.* In this section we briefly summarize some further useful results on product-integrals, starting with a version of Gronwall's inequality [Gronwall (1919)]. [See Beesack (1975) for a general survey of the topic of Gronwall inequalities, especially Sections 11 and 12, and see B. W. Helton (1969) and J. C. Helton (1977) for results in the context of product-integration.] Consider the nonnegative scalar case and recall that $\mu_0(t) = \prod_{]0,t]}(1 + d\alpha_0)$ satisfies the Volterra integral equation

$$\mu_0(t) = 1 + \int_{]0,t]} \mu_0(s-)\alpha_0(ds).$$

The basic Gronwall inequality is now the following.

THEOREM 9. *Suppose α_0 is a nonnegative scalar additive interval function on $]0, \tau]$ with $\mu_0(t) = \prod_{]0,t]}(1 + d\alpha_0)$; suppose μ is a cadlag nonnegative real function such that*

$$\mu(t) \le 1 + \int_{]0,t]} \mu(s-)\alpha_0(ds) \quad \text{for all } t \le \tau. \tag{54}$$

Then $\mu(t) \le \mu_0(t)$ for all $t \le \tau$.

PROOF. On repeatedly substituting the inequality for μ in the right-hand side of (54) we see the Péano series for μ_0 appearing together with a remainder which converges to zero by the implicit boundedness of μ. □

There are also inequality versions of inhomogeneous Volterra integral equations. The proof of the following very useful theorem is left to the reader.

THEOREM 10 (The inhomogeneous equation). *Let α be an additive interval function of bounded variation and ψ a cadlag matrix-valued function on $]0, \infty[$. Then ϕ satisfies*

$$\phi(t) = \psi(t) + \int_{]0,t]} \phi(s-)\alpha(ds) \quad \text{for all } t \tag{55}$$

iff

$$\phi(t) = \psi(t) + \int_{]0,t]} \psi(s-)\alpha(ds) \prod_{]s,t]} (\mathbf{1} + d\alpha) \quad \text{for all } t. \tag{56}$$

The related Gronwall inequality is of course that in the nonnegative scalar case, (55) with " = " replaced by " ≤ " implies the same modification of (56).

Finally, we mention a slight variant of the Volterra integral equation. Suppose that ϕ and α are as before and that ϕ satisfies the equation

$$\phi(t) = \mathbf{1} + \int_{]0,t]} \phi(s)\alpha(ds) \quad \text{for all } t,$$

that is, the integrand is the "anticipating" $\phi(s)$ instead of the nonanticipating $\phi(s-)$, which would have led to the product-integral of α as the solution. It turns out that this equation has as unique solution,

$$\phi(t) = \prod_{]0,t]} (\mathbf{1} + d\alpha)^{-1} = \left(\prod_{[t,0[} (\mathbf{1} - d\alpha) \right)^{-1},$$

provided that the inverse on the right-hand side exists. We refer to J. C. Helton (1977, 1978) for a complete collection of results combining all these kinds of integral equations and corresponding Gronwall inequalities.

3.3. *Product-integration of semimartingales.* A semimartingale X is a real-valued cadlag stochastic process which can be written as the sum of a process A with paths locally of bounded variation and a local square-integrable martingale M. The latter process may have infinite variation on compacts, but it does have finite quadratic variation: Taking a sequence of partitions $\mathscr{T}_m$ of $]0,\infty[$ such that $\mathscr{T}_m \cap]0,\tau]$ is finite for each $\tau < \infty$ and with mesh $|\mathscr{T}_m \cap]0,\tau]| \to 0$ as $m \to \infty$, we have

$$\sum_{\mathscr{T}_m \cap]0,t]} (\Delta M)^2 \to_P [M](t) \tag{57}$$

for a nondecreasing finite process $[M]$. Because finite variation implies finite quadratic variation, we also have

$$\sum_{\mathscr{T}_m \cap]0,t]} (\Delta X)^2 \to_P [X](t) \tag{58}$$

for a nondecreasing finite process $[X]$, called the (optional) quadratic variation of X.

The limits in (57) and (58) are actually uniform in t in $]0,\tau]$ for each finite τ, in probability. Note that $[A] = \Sigma(\Delta A)^2$, that is $[A](t)$ is the sum of the squares of the jumps of A in $[0,t]$. If M is continuous, then $[M] = \langle M \rangle$, the predictable variation process of M, and $[X] = [M] + [A]$. (A semimartingale allowing such a decomposition is called *special*.)

For two semimartingales X_1 and X_2 one can define the optional covariation process $[X_1, X_2]$ by replacing $(\Delta X)^2$ by $(\Delta X_1)(\Delta X_2)$ in (58); thus the notation $[X]$ is really just shorthand for $[X, X]$. If X is now a *vector* of semimartingales, then $[X]$ becomes the *matrix* of optional covariation processes of the components of X. If X is $p \times p$ matrix-valued, then $[X]$ is a $p^2 \times p^2$ matrix.

The solution of the Volterra (stochastic) integral equation (the forward equation) for X was first described by Doléans-Dade (1970) and is usually referred to as her exponential semimartingale. More generally it turns out that both $\prod(\mathbf{1} + dX)$ and $\prod(e^{dX})$ can be defined by (suitable modifications of) any of the Definitions 1–4 as we mentioned already in Section 1—see Karandikar (1983) and Hakim-Dowek and Lépingle (1986) for the most recent results of this type. [In fact, McKean (1960) already implicitly used $\prod(e^{dX})$ to construct Brownian motion on a Lie group.] Now, even when Definition 4 (the commutative case) is applicable, $\prod(\mathbf{1} + dX)$ and $\prod(e^{dX})$ are generally different. As we stated in Section 1, the equivalence between the definitions has been established [Emery (1978)] by the very clever construction of appropriate topologies on the space of semimartingales and then modifying the abstract approach to the theory of integral equations (based on the fixed-point theorem for contractive mappings on a Banach space).

Now it turns out that the finite quadratic variation property (58) alone is sufficient for establishing one of the central results in stochastic analysis: Itô's celebrated formula,

$$f(X_t) = f(X_0) + \int_0^t f'(X_{s-})\, dX_s + \int_0^t f''(X_{s-})\, d[X]_s, \tag{59}$$

for a twice continuously differentiable function f, with the obvious vector and matrix analogues. This is done by Föllmer (1981) using a quite straightforward pathwise (deterministic) analysis, establishing (59) by looking at sequences of partitions for which (58) holds almost surely. [This is based on a result due to Doléans-Dade (1969); more precisely, sequences are considered in which there is weak convergence of the discrete measures $\mu_m = \sum_{t_i \in \mathscr{T}_m} \delta_{t_i}(X_{t_i} - X_{t_{i-1}})^2$ to the measure with distribution function $[X]_t$.] Since the Itô formula involves the stochastic integral $\int f'(X_-)\, dX$ with respect to the possibly infinite variation process X, part of the proof is dedicated to establishing a meaning for this integral as a limit of Riemann sums; again, *just* the finite quadratic variation property is needed to do this.

It is well known that the stochastic integral of a predictable process with respect to a semimartingale is an in-probability limit of Riemann sums, though the proof of this, as well as that of (58), is based on a delicate stochastic analysis. In fact, if the partitions are chosen appropriately—in fact randomly, using stopping times—one can get almost-sure convergence. However [see, e.g., Bichteler (1981)], the partitions are now chosen dependently on the integrand while in Föllmer's construction they are fixed in advance and work for all C^2 functions f simultaneously. [Alternatively, one can make the existence of Riemann-approximating sums the starting point for the theory of stochastic integration, and only at the end show that the class of natural integrating processes is the class of semimartingales; see Protter (1989).]

All these results lead one to hope that the stochastic aspects of the product-limit definition of $\prod(\mathbf{1} + dX)$ for a semimartingale X can be completely ignored: One should simply try to define $\prod(\mathbf{1} + dX)$ for deterministic $p \times p$ matrix functions of finite quadratic covariation, that is, functions such that a sequence of partitions $\mathscr{T}_m$ exists satisfying

$$\sum_{\mathscr{T}_m \cap]0, t]} (\Delta X_{ij}\, \Delta K_{kl}) \to [X_{ij}, X_{kl}](t)$$

for a finite $p^2 \times p^2$ matrix function $[X]$. Such a theory, if we want to include the equivalent Definitions 2 and 3 (the forward equation and the Péano series) would have to include a "deterministic" theory of integration of the left-continuous version of such a function with respect to another.

Unfortunately, the latter seems to be very hard in general and we have not even succeeded in showing the existence as a product-limit (Definition 1) of $\mathcal{T}\!\!\mathcal{T}(1 + dX)$. The programme seems to fail because of the asymmetry of $\int X_{1-}\,dX_2$, which makes its definition rather delicate. Even if we restrict attention to the product-limit, we still seem implicitly to need this integral. However, we can easily define $\int X_-\,dX$ "symmetrically" for real-valued X by the trick $\int X_-\,dX = \frac{1}{2}(X^2 - [X, X])$. This explains why we can get much of the way in the scalar case, as we now explain.

Let X_t, $t \in [0, 1]$, be a cadlag *real* function with jumps strictly less than 1 in absolute value. We consider partitions $\mathscr{T}$ of $[0, \tau]$ and corresponding finite products $\prod_{\mathscr{T}}(1 + \Delta X)$. Taking logarithms and expanding, we find

$$\begin{aligned}\log \prod_{\mathscr{T}} (1 + \Delta X) &= \sum_{\mathscr{T}} \Delta X - \tfrac{1}{2} \sum_{\mathscr{T}} \Delta X^2 + \tfrac{1}{3} \sum_{\mathscr{T}} \Delta X^3 + \cdots \\ &= X_\tau - \tfrac{1}{2} \sum_{\mathscr{T}} \Delta X^2 + \tfrac{1}{3} \sum_{\mathscr{T}} \Delta X^3 + \cdots . \end{aligned} \tag{60}$$

Consider a sequence of partitions with mesh converging to zero and $\sum_{\mathscr{T}} \Delta X^2 \to [X]_\tau$. Note that $|\sum_{\mathscr{T}} \Delta X^k| \le (\max_{\mathscr{T}} |\Delta X|)^{k-2} \sum_{\mathscr{T}} \Delta X^2$ and that $\sum_{\mathscr{T}} \Delta X^k \to \sum_t (\Delta X_t)^k$ for $k > 2$, and $|\sum_t (\Delta X_t)^k| \le (\max_t |\Delta X_t|)^{k-2} [X]_\tau$, where $\Delta X_t = X(\{t\})$. By absolute convergence of the series for $\log(1 + x)$ in $|x| < 1$ this gives convergence of $\log \prod_{\mathscr{T}}(1 + \Delta X)$ to

$$X_\tau - \frac{1}{2}[X]_\tau + \sum_{k=3}^{\infty} \frac{(-1)^{k+1}}{k} \sum_t (\Delta X_t)^k.$$

We now simply define $\mathcal{T}\!\!\mathcal{T}_{[0,\tau]}(1 + dX)$ as the exponential of this expression.

One also has the Péano series representation for $\mathcal{T}\!\!\mathcal{T}_{[0,\tau]}(1 + dX)$, including the existence of each of its terms as $\lim \sum \cdots \sum_{i_1 < \cdots < i_k} \Delta_{i_1} X \cdots \Delta_{i_k} X$, where $\Delta_i X$ denotes the increment of X over the ith element of the partition. This is proved, as pointed out to us by Atma Mandrekar, by noting that

$$\prod_{\mathscr{T}} (1 + \Delta X) = 1 + \sum_{k=1}^{\infty} \sum_{i_1 < \cdots < i_k} \cdots \sum \Delta_{i_1} X \cdots \Delta_{i_k} X \tag{61}$$

and $\sum \cdots \sum_{i_1 < \cdots < i_k} \Delta_{i_1} X \cdots \Delta_{i_k} X$ is a polynomial of degree k in $\sum_{\mathscr{T}} \Delta X$, $\sum_{\mathscr{T}} \Delta X^2, \dots, \sum_{\mathscr{T}} \Delta X^k$, each of which converges as $|\mathscr{T}| \to 0$. These polynomials are called Newton's polynomials [see Avram and Taqqu (1986), Appendix]. So each term of the Péano series does exist as a limit of iterated Riemann sums and we just need to bound the tail of the series (61) (summing over $k > k_0$, say) by something arbitrarily small, uniformly in $\mathscr{T}$, for sufficiently large k_0. This can be done quite easily by taking the power series for the exponential of the right-hand side of (60) and rearranging terms.

It is not at all clear what can be said about a version of the forward equation characterisation of this product-integral (Definition 2), since it is very hard to give a meaning to the integrals with respect to X in this definition, and more precisely to describe the class of possible integrands (which must depend on X). This may be possible, as pointed out to us by Hans Föllmer, by means of the characterization of $\int Y_- \, dX - \int X_- \, dY$ as twice the area enclosed by the curve obtained by plotting Y_t against X_t and joining its endpoints with a straight line (counting the area enclosed by closed loops as many times—positive or minus—as the curve goes clockwise or anticlockwise, respectively, round the loop). Since we already can define $\int X_- \, dY + \int Y_- \, dX$ by the symmetrization trick as $XY - [X, Y]$, this yields a definition of $\int Y_- \, dX$ whenever the area just described is well defined; see Lévy [(1948), Section VII.55] for the case when X and Y are independently Brownian motions.

Possibly all this could lead to a general theory for the matrix case too; however, this brings us rather far afield from our primary interests and it is therefore not pursued here.

A different topic is the question as to whether it is possible to derive a Duhamel equation for semimartingales; as we saw, all our other major product-integral results are valid for semimartingales. The answer to this question is that a Duhamel-like equation can be obtained for semimartingales by applying the Itô formula to $f(X_t, Y_t) = X_t Y_t^{-1}$, and then by multiplying throughout by Y_t. Since the stochastic integral is only defined for predictable integrands, arriving at the Duhamel equation depends on making an appropriate definition of the integrals involved. We leave the details to the interested reader.

3.4. *Product-integration over* $\mathbb{R}_+^k$. As we mentioned in Section 2.3, the product-integral of a *real* signed measure over bounded subsets of $\mathbb{R}_+^k$ can be defined as a limit over partitions, where partitions are now chosen more freely. One can ask how much of the "equivalent definitions" and "further properties" of Sections 2.4, 3.1 and 3.2 are preserved under this extension. Note that the Kolmogorov forward and backward equations, the Péano series and the Duhamel equation, all make strong use of the total ordering of the real numbers in $\mathbb{R}_+^1$. Two sets of results now follow, depending on whether we introduce a *total* or only a *partial* ordering on $\mathbb{R}_+^k$.

A total ordering gives stronger and more results, but always formulated in terms of the somewhat arbitrary choice of "linearization" of space. For instance, one is free to define

$$x < y \quad \Leftrightarrow \quad \text{for some } j = 1, \dots, k, \, x_j < y_j, \, x_i = y_i, \, i > j,$$

$$x \le y \quad \Leftrightarrow \quad x < y \text{ or } x = y.$$

Now any two x, y satisfy just one of the three possibilities $x < y$, $x = y$ or $x > y$. This total ordering corresponds when $k = 2$ to scanning row-wise (from bottom to top and, within rows, from left to right) as on a video screen.

Now, defining

$$]0,s] = \{u: 0 < u \le s\},$$
$$]0,s[= \{u: 0 < u < s\},$$
$$]s,t] = \{u: s < u \le t\} =]0,t] \setminus]0,s]$$

and so on, we find, by simply inspecting the proofs in Section 2.4, that for any bounded B the following hold:

$$\prod_{B}(1+d\alpha) = 1 + \int_{s\in B} \prod_{B\cap]0,s[}(1+d\alpha)\alpha(ds)$$
$$= 1 + \int_{s\in B}\alpha(ds)\prod_{B\cap]s,\infty[}(1+d\alpha) \quad \text{(Kolmogorov)};$$

$$\prod_{B}(1+d\alpha) = 1 + \sum_{n=1}^{\infty}\int\cdots\int_{s_1<\cdots<s_n;\, s_i\in B}\alpha(ds_1)\cdots\alpha(ds_n) \quad \text{(Péano)};$$

$$\prod_{B}(1+d\alpha) - \prod_{B}(1+d\beta)$$
$$= \int_{s\in B}\prod_{B\cap]0,s[}(1+d\alpha)(\alpha(ds)-\beta(ds))\prod_{B\cap]s,\infty[}(1+d\beta) \quad \text{(Duhamel)}.$$

Hence, we can also derive continuity and differentiability results, where the proof goes just as in Section 3.1 but using now the integration-by-parts formula,

$$U(B)V(B) = \int_{s\in B}\int_{t\in B}U(ds)V(dt)$$
$$= \iint_{s,t\in B,\, s<t}U(ds)V(dt) + \iint_{s,t\in B,\, s\ge t}U(ds)V(dt)$$
$$= \int_{t\in B}U(]0,t[\,\cap B)V(dt) + \int_{s\in B}V(]0,s]\cap B)U(ds),$$

for measures U and V on $\mathbb{R}_+^k$. Suppose we restrict attention to sets B which are rectangles aligned with the axes (not intervals now!) and define the variation norm as usual but let the supremum norm of a measure be the supremum over the absolute value of the measure assigned to rectangles (equivalent to the supremum norm of the corresponding distribution function). Then the set $]0,s]\cap B$ appearing at the end of the integration-by-parts formula is a union of up to k rectangles, and consequently we can bound $|\int_{t\in B}U(]0,t[\cap B)V(dt)|$ by a constant $(k+1)$ times the product of the variation norm of U with the supremum norm of V. In the differentiability proof this constant will appear repeatedly, but the result is that the product-integral of a real measure over $\mathbb{R}_+^k$ is compactly differentiable with exactly the same derivative as in Theorem 8 provided we interpret norms, intervals, integration by parts, and so on, in the appropriate way.

What about partial orders? An alternative theory, not of the product-integral but of the Péano series $P(B;\alpha)$, follows by defining for an arbitrary *partial* order $\ll$ on $\mathbb{R}_+^k$ and arbitrary bounded B

$$P(B;\alpha) = 1 + \sum_{n=1}^{\infty} \int\cdots\int_{u_1 \ll \cdots \ll u_n,\, u_i \in B} \alpha(du_1) \cdots \alpha(du_n).$$

For instance, one could take the partial order

$$x \ll y \quad \Leftrightarrow \quad x_i < y_i \;\forall\, i.$$

Define also $x \leq y \;\Leftrightarrow\; x_i \leq y_i \;\forall\, i$. One can again verify that the proofs of the Kolmogorov and Duhamel equations go through without change for the Péano series with respect to any given partial order, just defining

$$\begin{aligned}]0,s] &= \{u\colon 0 \ll u \leq s\}, \\ ]0,s[&= \{u\colon 0 \ll u \ll s\}, \\ ]s,t] &= \{u\colon s \ll u \leq t\} \neq]0,t] \setminus]0,s], \end{aligned}$$

we get

$$\begin{aligned} P(B;\alpha) &= 1 + \int_{s \in B} P(B \cap]0,s[;\alpha)\alpha(ds) \\ &= 1 + \int_{s \in B} d\alpha(s) P(B \cap]s,\infty[;\alpha) \quad \text{(Kolmogorov)}, \end{aligned}$$

$$\begin{aligned} &P(B;\alpha) - P(B;\beta) \\ &\quad = \int_{s \in B} P(B \cap]0,s[;\alpha)(\alpha(ds) - \beta(ds)) P(B \cap]s,\infty[;\beta) \quad \text{(Duhamel)}. \end{aligned}$$

The key to continuity and differentiability is the Duhamel equation plus integration by parts, which looks almost the same:

$$\begin{aligned} U(B)V(B) &= \int_{s \in B} \int_{t \in B} U(ds)V(dt) \\ &= \iint_{s,t \in B,\, s \ll t} U(ds)V(dt) + \iint_{s,t \in B,\, s \not\ll t} U(ds)V(dt) \\ &= \int_{t \in B} U(]0,t[\cap B)V(dt) + \int_{s \in B} V(B \setminus]s,\infty[)U(ds). \end{aligned}$$

Still, if B is a rectangle and $\ll$ is the just described partial order, $B \setminus]s,\infty[$ is just the difference of two rectangles. So again this leads to identical-looking continuity and differentiability results; only the interpretation of interval and order are different.

We do not have a product-limit interpretation of the Péano series. Though one might define a partition of B along points $0 = t_0 \leq t_1 \leq \cdots \leq t_n = \infty$ as the collection of sets $B \cap (]0,t_i] \setminus]0,t_{i-1}])$, two partitions do not generally have a common refinement when working with partially ordered time (partitions are only *partially* ordered too). Thus many different product-limits along refinements can and will exist, one of them sometimes being *the* product-integral, though none being the Péano series with respect to the partial order. In fact, for the usual partial order, all product-limits of a real measure α are the same and are the ordinary product integral if and only if whenever positive mass is assigned to an affine subspace parallel to the coordinate axes, it all lies on one atom.

These results have applications in multivariate censored data and explain why there are actually so many different generalizations of the Kaplan–Meier estimator of the univariate case. Two of these generalizations are briefly discussed in Section 4.3.

4. Applications.

4.1. *The application of product-integrals to survival and hazard functions.* Let T be a positive random variable, the survival time of a cancer patient, say. We define the survival function $S(t) = 1 - F(t) = \Pr\{T > t\}$. If S is positive and differentiable, then we can define the hazard rate

$$\lambda(t) = -\frac{d\log(S(t))}{dt} = \lim_{h\downarrow 0}\frac{1}{h}\Pr\{T \le t + h | T > t\} \tag{62}$$

and we have the well-known relation

$$S(t) = \exp\left[-\int_{]0,t]}\lambda(u)\,du\right]. \tag{63}$$

The measure Λ defined by $\Lambda(]0,t]) = \int_{]0,t]}\lambda(u)\,du$ is called the intensity or hazard measure. We shall generalize the relations (62) and (63) to an arbitrary survival function S and a correspondingly more general hazard measure Λ. The following theorem characterizes those intensity measures Λ which can arise and shows that the generalized relation between $-\Lambda$ and S is simply that between a measure and its product-integral or between the additive interval function $-\Lambda(s,t) = -\Lambda(]s,t])$ and the multiplicative interval function $S(s,t) = \Pr(T \ge t | T \ge s) = S(t)/S(s)$. This relationship can of course be exploited and also further generalized as we shall see in later sections. We let $\tau \le \infty$ be the upper endpoint of the support of T. The two cases (a) and (b) in the theorem correspond to the cases that τ itself has positive or zero probability, respectively.

THEOREM 11. *Let Λ be a nonnegative measure on $]0,\tau]$ which is finite on $]0,s]$ and such that $\Lambda(\{s\}) < 1$ for all $s < \tau$ and which satisfies either*

(a) $$\Lambda(]0,\tau[) < \infty, \qquad \Lambda(\{\tau\}) = 1$$

or

(b) $$\Lambda(]0,\tau[) = \infty, \qquad \Lambda(\{\tau\}) = 0.$$

Then, defining

$$S(t) = \prod_{]0,t]} (1 - d\Lambda), \tag{64}$$

S is the survival function of a random variable T with upper support endpoint τ. Conversely, if T is a positive random variable with survival function S and upper support endpoint τ satisfying either

$$S(\tau -) > 0 \tag{a'}$$

or

$$S(\tau -) = 0, \tag{b'}$$

then Λ, defined by

$$\Lambda(]0,t]) = -\int_{]0,t]} \frac{S(du)}{S(u-)} \tag{65}$$

$$= -\int_{]0,t]} d(S - 1), \tag{65'}$$

has the properties just described, where in (65′) *S is interpreted as the multiplicative interval function* $S(s,t) = S(t)/S(s)$.

Equation (65′) states that $\Lambda(t) = \lim_{|\mathscr{T}|\to 0} \sum_i \Pr\{T \le t_i | T > t_{i-1}\}$, where $\mathscr{T}$ denotes a partition of $]0,t]$. The probabilistic interpretation of the atoms of Λ is $\Lambda(\{t\}) = \Pr\{T = t | T \ge t\}$. Strictly speaking, in the case (b) we first need to define the product-integral of an unbounded measure, that is, by the usual definition of an indefinite Riemann integral. Before we prove the theorem we need a technical lemma.

LEMMA 6. *Let Λ satisfy the assumptions in Theorem* 11 *and let S be defined by* (64). *Then* $S(t) > 0$ *for all* $t < \tau$ *and also, in case* (a), *for* $t = \tau$.

PROOF. Let $1 - 2\eta$ be the size of the largest atom of Λ in $]0,t]$ and choose a partition of $]0,t]$ such that $\Lambda(]t_i, t_{i+1}[) < \eta$ for all i. Then

$$1 - \Lambda(]t_i, t_{i+1}]) = 1 - \Lambda(]t_i, t_{i+1}[) - \Lambda(\{t_{i+1}\})$$
$$\ge 1 - \eta - (1 - 2\eta) = \eta$$

and hence by the inequality $1 - x \ge \exp(-c(\eta)x)$ for all $0 \le x \le 1 - \eta$, where $c(\eta) = -\log(\eta)/(1-\eta) < \infty$,

$$\log(1 - \Lambda(]t_i, t_{i+1}])) \ge -c(\eta)\Lambda(]t_i, t_{i+1}]).$$

Summing over i and letting $|\mathscr{T}| \to \infty$ gives $S(t) = \prod_{]0,t]}(1 - d\Lambda) \ge \exp(-c(\eta)\Lambda(]0,t])) > 0$. □

PROOF OF THEOREM 11. Here we show the equivalence of (64), (65) and (65′); the fine details are left to the reader. Suppose first Λ is given and S is defined by (64). Then by the forward equation (40) (Theorem 5) we have $S(t) = 1 - \int_{]0,t]} S(u-)\Lambda(du)$. Since, by Lemma 6, $S(u-) > 0$ we get (65). If on the other hand Λ satisfies (65), then S solves the forward equation and hence by the uniqueness of its solution (Theorem 6) it is given by the product-integral (64). The equivalence between (64) and (65′) follows simply from Theorem 3 by considering the multiplicative interval function $S(s,t) = \Pr(T \geq t | T \geq s) = S(t)/S(s)$. □

Because we are in the commutative case we can rewrite (64) in the far less intuitive but very well known form,

$$S(t) = \prod_{s \leq t} (1 - \Lambda(\{s\}))\exp(-\Lambda^c(t)).$$

See, for instance, Cox [(1972), page 172], where the term product-integral is used [following Arley and Borschenius (1945), as David Cox has informed us]. Beran [(1981), Theorem 2.1] gives a sketch of a nice direct proof. Wellner (1985) points out the connection with Doléans-Dade's exponential semimartingale [Doléans-Dade (1970)].

In the next section we will use the Duhamel equation in order to discuss functional differentiation of the mapping from hazard measures to survival functions. Since the mapping is one-to-one, the inverse mapping must have a derivative which is also the inverse, in the appropriate sense, of the forward derivative. Efron and Johnstone (1990) and Ritov and Wellner (1988) discuss the intimately related derivatives of the mappings between hazard rate λ and probability density f, whose special isometric properties turn out to be connected with the information identity (mentioned again briefly at the end of Section 4.6)

$$E_\theta\left(\frac{\partial \log f(T;\theta)}{\partial \theta}\right)^2 = E_\theta\left(\frac{\partial \log \lambda(T;\theta)}{\partial \theta}\right)^2.$$

4.2. *The product-limit estimator*. We saw in the Section 4.1 that a survival function is the product-integral of its intensity or hazard measure. Given censored observations from a life distribution, the natural estimator of the survival function is the product-limit or Kaplan–Meier estimator [Kaplan and Meier (1958)], and it turns out to be the product-integral of the equally well known empirical cumulative hazard function, or Nelson–Aalen estimator [Nelson (1969), Aalen (1975) and Johansen (1978)]. This puts the machinery of product-integration at our disposal in order to derive various properties of these estimators, as we shall now sketch. The key ingredient is the Duhamel equation (42), which expresses the difference between survival function estimator and estimand in terms of the difference between the corresponding empirical and true hazard measures.

For the sake of definiteness we first work in the classical random censorship model. Let $T_1, \ldots, T_n$ be i.i.d. positive lifetimes from the distribution F with survival function S and let, independently thereof, $C_1, \ldots, C_n$ be i.i.d. positive censoring variables from a distribution with survival function H. Both S and H may have a discrete component and may put positive mass on $t = +\infty$. Let $\tilde{T}_i = T_i \wedge C_i$ and $D_i = 1\{T_i \le C_i\}$, $i = 1, \dots, n$, be the data actually observed. Define

$$\begin{aligned} N_n(t) &= n^{-1}\#\{i: \tilde{T}_i \le t,\, D_i = 1\}, \\ Y_n(t) &= n^{-1}\#\{i: \tilde{T}_i \ge t\}, \\ \hat{\Lambda}_n(t) &= \int_{]0,t]} \frac{dN_n}{Y_n}, \\ \hat{S}_n(t) &= \prod_{]0,t]} (1 - d\hat{\Lambda}_n). \end{aligned}$$

So $\hat{S}_n$ is the Kaplan–Meier estimator of S and $\hat{\Lambda}_n$ is the Nelson–Aalen estimator of Λ. One finds easily

$$\begin{aligned} EN_n(t) &= \int_{]0,t]} H(s-)F(ds), \\ EY_n(t) &= H(t-)S(t-), \\ \Lambda(t) &= \int_{]0,t]} \frac{d(EN_n)}{EY_n}, \quad \text{provided } EY_n(t) > 0, \\ S(t) &= \prod_{]0,t]} (1 - d\Lambda). \end{aligned}$$

The Duhamel equation applied to $\hat{S}_n - S$ gives

$$\hat{S}_n(t) - S(t) = \int_{]0,t]} \hat{S}_n(s-)(\Lambda - \hat{\Lambda}_n)(ds)\frac{S(t)}{S(s)},$$

or

$$\frac{\hat{S}_n(t)}{S(t)} - 1 = -\int_{]0,t]} \frac{\hat{S}_n(s-)}{S(s)}(\hat{\Lambda}_n - \Lambda)(ds). \tag{66}$$

This key equality was first established in the more general context of inhomogeneous Markov processes by Aalen and Johansen (1978) and later exploited by Gill (1980a, 1983) to derive small-sample results (unbiasedness, variance) and large-sample results (consistency, weak convergence) for the Kaplan–Meier estimator using martingale methods: Namely, one has in (66) (if one replaces t by $t \wedge \max_i \tilde{T}_i$) that the integrating function $\hat{\Lambda}_n - \Lambda$, stopped at the largest observation, is a square-integrable martingale while the integrand is a predictable process, so the left-hand side of the equation is a square-integrable martingale too. These martingale techniques are available in a wider class of censoring models than just the random censoring described previously; for instance, under models for random truncation [Keiding and Gill (1990)].

All the same, many important models do not have this martingale structure (a few are mentioned later) and a less delicate approach is needed. Here we take an approach restricted to i.i.d. models (but not requiring martingale structure) and use Theorems 7 and 8 (continuity and differentiability of the product-integral) to derive strong consistency and weak convergence of the Kaplan–Meier estimator. Note that we proved these two theorems using the Duhamel equation so this approach is also based on (66) but in a disguised form. A law of the iterated logarithm can also be derived in this way.

We work on a fixed interval $[0, \tau]$, where τ satisfies $EY_n(\tau) = S(\tau-)H(\tau-) > 0$. Note that N_n and Y_n and their expectations EN and EY (since these do not depend on n we have dropped the subscript) are bounded monotone functions in $D[0, \tau]$ or $D_-[0, \tau]$, the cadlag and "caglad" functions, respectively, on $[0, \tau]$. By monotony and boundedness they actually have uniformly bounded variation. The variation of $\hat{\Lambda}_n$ is not uniformly bounded but there exists a constant which it only exceeds with probability tending to 0 as n tends to infinity. We endow these spaces with the supremum norm and their product with the max supremum norm, both to be denoted by $\|\cdot\|_\infty$. We can now consider $\hat{S}_n$ as the result of applying three mappings one after the other:

$$(N_n, Y_n) \to \left(N_n, \frac{1}{Y_n}\right) \to \hat{\Lambda}_n \to \hat{S}_n,$$

that is, going through the spaces

$$D[0,\tau] \times D_-[0,\tau] \to D[0,\tau] \times D_-[0,\tau] \to D[0,\tau] \to D[0,\tau],$$

and corresponding to inversion of one component, then integrating one with respect to the other and finally product-integrating. Now, we have already showed that the last mapping is continuous and even compactly differentiable when we restrict its domain to a set of elements of $D[0, \tau]$ of uniformly bounded variation. The same is true for the central mapping (integration) by Gill [(1989)], Lemma 3]; this is essentially the Helly–Bray lemma. The first mapping is trivially continuous and differentiable when we restrict it to elements of $D_-[0, \tau]$ bounded uniformly away from zero. Now, by the Glivenko–Cantelli theorem we have $\|(N_n - EN, Y_n - EY)\|_\infty \to 0$ a.s. as $n \to \infty$. This gives us consistency of the Kaplan–Meier estimator,

$$\|\hat{S}_n - S\|_\infty \to 0 \quad \text{a.s. as } n \to \infty.$$

Weak convergence of $n^{1/2}(\hat{S}_n - S)$ follows directly from weak convergence of $n^{1/2}(N_n - EN, Y_n - EY)$ and the compact differentiability of the three mappings and hence of their composition. Here we use the weak convergence theory of Dudley (1966) as nicely described in Pollard (1984), which allows us to work in $D[0, \tau] \times D_-[0, \tau]$ with the uniform topology (associated with the max supremum norm). This is accomplished by endowing this space with a smaller σ-field than the Borel σ-field (generated by all open sets), namely, the σ-field generated by the coordinate projections or the open balls, which in fact turns out [Billingsley (1968), Theorem 14.5] to be just the usual Skorohod σ-algebra. Since not every continuous function is now measurable, the classical definition of weak convergence needs slight modification. However, it turns out that all the classical theory still goes through. Whenever the limit process is continuous, weak convergence in the new sense is equivalent to the more familiar notion of weak convergence with respect to the Skorohod topology, as described in Billingsley (1968); otherwise the result is slightly stronger.

The other side of this story, the use of compact differentiability, is simply the functional version of the δ-method. This approach is described in Gill (1989), following Reeds (1976). The new weak convergence theory fits beautifully with the differentiability theory and allows more natural extension to multivariate time indexed processes. The even further generalized weak convergence theory due to Hoffman-Jørgensen fits just as nicely [see van der Vaart and Wellner (1990) and Sheehy and Wellner (1990a, 1990b)].

Again, we fix τ such that $EY(\tau) > 0$. It is clear that

$$n^{1/2}(N_n - EN, Y_n - EY) \to_{\mathscr{D}} (Z_N, Z_Y) \quad \text{in } (D[0,\tau] \times D_-[0,\tau], \|\cdot\|_\infty) \quad \text{as } n \to \infty,$$

where (Z_N, Z_Y) is a bivariate Gaussian process with zero mean and the same covariance structure as the process on the left-hand side (the same for all n). Now, the three mappings

$$\phi\colon (x,y) \to (x,u) = \left(x, \frac{1}{y}\right),$$

$$\psi\colon (x,u) \to v = \int_{[0,\cdot]} u\,dx,$$

$$\xi\colon v \to z = \prod_{[0,\cdot]} (1 - dv),$$

are all compactly differentiable at the relevant point $x = EN$, $y = EY$, $u = 1/EY$, $v = \Lambda$, $z = S$ with derivatives

$$d\phi(x,y)\cdot(h,k) = \left(h, -\frac{k}{y^2}\right) = (h,j),$$

$$d\psi(x,u)\cdot(h,j) = \int_{[0,\cdot]} j\,dx + \int_{[0,\cdot]} u\,dh = l,$$

$$d\xi(v)\cdot l = -z\int_{[0,\cdot]} \frac{z_-}{z}\,dl.$$

By composition of these derivative mappings (the chain rule), evaluated at the point $x = EN$, $y = EY$, $u = 1/EY$, $v = \Lambda$, $z = S$ and acting on $(h,k) = (Z_N, Z_Y)$, we obtain the required weak convergence result,

$$\begin{aligned} n^{1/2}(\hat{S}_n - S) &\to_{\mathscr{D}} S\int_{[0,\cdot]} \frac{1}{(1-\Delta\Lambda)}\left(-\frac{Z_Y}{(EY)^2}\,dEN + \frac{1}{EY}\,dZ_N\right) \\ &= -S\int_{[0,\cdot]} \frac{dZ_N - Z_Y\,d\Lambda}{(1-\Delta\Lambda)EY} \quad \text{in } (D[0,\tau], \|\cdot\|_\infty) \text{ as } n \to \infty. \end{aligned}$$

Direct calculation of the covariance structure of $Z_N - \int_{[0,\cdot]} Z_Y \, d\Lambda$ (the same as its counterpart for $n = 1$) shows that this zero mean Gaussian process has independent increments with variance function

$$\int_{[0,\cdot]} EY(1 - \Delta\Lambda) \, d\Lambda.$$

Thus,

$$n^{1/2}(\hat{S}_n - S) \to_{\mathscr{D}} SW\left(\int_{[0,\cdot]} \frac{d\Lambda}{(1 - \Delta\Lambda)EY}\right),$$

where W is a standard Wiener process.

Though this result also follows perhaps more easily from the martingale approach mentioned previously, it has great significance as being essentially the same proof as the first proof of weak convergence of the Kaplan–Meier estimator ever given, due to Breslow and Crowley (1974). That proof seemed very long, complicated and ad hoc at the time [Burke, Csörgő and Horváth (1981) even claimed it was incorrect], but one can now recognize in it a standard δ-method argument. Our contribution is to isolate and name the components of the proof simply as checks of the differentiability of the various functionals involved. This modularization allows us to use the same components in many different settings and for other purposes, a few of which are mentioned later.

We note that continuity of survival distribution or of censoring distribution has not been required, nor have any martingale properties been used, just the weak convergence of the underlying empiricals. The proof also gives a weak Bahadur representation or asymptotic linearity for $n^{1/2}(\hat{S}_n - S)$ (a linear representation in terms of the underlying empirical distribution functions), and it shows that the bootstrap works for the Kaplan–Meier estimator [see Gill (1989) and Sheehy and Wellner (1990a, 1990b)]. Asymptotic efficiency properties of the Kaplan–Meier estimator follow directly from asymptotic efficiency of the empirical distribution functions in the random censorship model and compact differentiability [see van der Vaart (1988); see also Keiding and Gill (1990) for similar arguments in the model of *random truncation*]. One can go on and use compact differentiability of the relevant functionals in order to show that a large number of statistics derived from the Kaplan–Meier estimator also converge weakly; for instance, its quantile function, the quantile residual lifetime function, the total time on test plot, and so on.

The differentiability-based proof of weak convergence is available in any i.i.d. situation as long as EN and EY are related via $dEN = Y d\Lambda$. The previously mentioned martingale property need not be available; an example of this is when the observations come from censored observation of a renewal process [see Gill (1980b, 1981). An amusing application is the estimation of the length distribution of the line segments of a Poisson line-segment process observed through a finite window [Laslett (1982) and Denby and Vardi (1985)]. The lengths of line segments which extend outside the window are censored by the boundary of the observational window. The Kaplan–Meier estimator cannot be used in the naive way, but it can be used if one only includes line segments of which at least one endpoint is inside the window, with uncensored line segments (both endpoints visible) duplicated (i.e., counted as two observations). To put it another way, each observed endpoint provides one observation, censored at the first crossing of the boundary of the window (which need not even be convex). Asymptotics are as the intensity of the Poisson process converges to infinity, keeping the window fixed. The relation $dEN = Y d\Lambda$ holds because it is true when one restricts attention to the two (overlapping) subsamples: the line segments of which the "northern endpoint" is visible and those of which the "southern endpoint" is visible. For these subsamples, conditioning on the visibility of the endpoint and using standard properties of the Poisson line-segment process, the distance from the relevant endpoint to the boundary of the window in the direction of the line-segment is independent of the length of the line, and therefore we have a classical random censoring model for each subsample apart.

4.3. *Multivariate Kaplan–Meier estimators.* The problem of multivariate censored data has resisted attack for a long time. The NPMLE is not unique and is computationally intractable; the pathwise product-limit estimator based on writing the multivariate survival function as a product of univariate marginal and conditional survival functions depends on the arbitrary choice of first, second, and so forth, coordinates and, moreover, usually assigns negative mass somewhere; estimators based on multivariate density estimation are very complicated and are based on many more arbitrary choices [Muñoz (1980), Campbell (1981), Campbell and Földes (1982) and Tsai, Leurgans and Crowley (1986)].

An important recent proposal is Dabrowska's multivariate product-limit estimator [Dabrowska (1988, 1989)], based on a beautiful representation of a multivariate survival function in terms of its conditional multivariate hazard measures. In fact, ordinary product-integrals (over rectangles) of fairly simple functionals of these are all that are involved, and indeed consistency, asymptotic normality and correctness of the bootstrap (both nonparametric, resampling from the data, and semiparametric, resampling from the estimated model) all hold by routine applications of our continuity and differentiability results (as extended in Section 3.4) and the functional δ-method.

For vectors, $\gg$ denotes coordinatewise $>$ and $\geq$ denotes coordinatewise $\geq$. Let $\bar{F}(t) = \Pr(T \gg t)$, $t \in \mathbb{R}_+^k$, be the survival function of a k-variate survival time T. Dabrowska's estimator is based on the representation

$$\bar{F}(t) = \prod_{C \subseteq S} \underset{s_C \leq t_C}{\pi} \left(\prod_{B \subseteq C} \left(1 + \sum_{\varnothing \subset A \subseteq B} (-1)^{|A|} \Lambda(ds_A | s_C) \right)^{(-1)^{|C \setminus B|}} \right).$$

Here, for given subsets $\varnothing \subset A \subseteq B \subseteq C \subseteq S = \{1, \ldots, k\}$ we denote by $\Lambda(ds_A|s_C)$ the conditional multivariate hazard measure $\Lambda(ds_A|s_C) = \Pr(T_A \leq s_A + ds_A | T_C \geq s_C)$, and for $s \in \mathbb{R}_+^k$, s_C is the vector $(s_i\colon i \in C)$.

Estimation is done by replacing the conditional hazard measures by their natural empirical analogues, setting $\Lambda(ds_A|s_C)$ equal to the point mass: number of observations with T_A *known* to equal s_A, divided by the number of observations with T_C *known* to exceed or equal s_C. The formula can be understood by approximating the product-integrals by finite products with respect to partitions of $]0, t]$ into small rectangles, replacing $(1 + \sum_{\varnothing \subset A \subseteq B}(-1)^{|A|}\Lambda(ds_A|s_C))$ by the ratio $\Pr\{T_B \geq s_B + ds_B\}/\Pr\{T_C \geq s_C\}$, followed by massive cancellations. For given C, the product-integral is equal to the multiplicative interval function

$$S_C(s_C, t_C) = \prod_{\varnothing \subseteq B \subseteq C} \bar{F}_C((t_{C \setminus B}, s_B))^{(-1)^{|B|}}$$

generalizing Theorem 11 of Section 4.1.

A completely different proposal due to Peter Bickel and using *marginal* multivariate hazards is also briefly mentioned in Dabrowska's (1988) paper. This is based on the fact that the survival function is the solution of the following *inhomogeneous* Volterra integral equation involving the multivariate hazard $\Lambda(dt) = \Lambda(dt_S|t_S)$, with boundary conditions involving the marginal distribution functions in strictly lower dimensions:

$$\bar{F}(t) = -\sum_{\varnothing \subseteq C \subset S} (-1)^{|S \setminus C|}\bar{F}_C(t_C) + (-1)^{|S|}\int_{s \leq t} \bar{F}(s-)\Lambda(ds).$$

The last integral reduces to $F(t)$ and a simple application of the inclusion–exclusion principle yields the required equality (taking $\bar{F}_\varnothing \equiv 1$).

We therefore get, by Theorem 10, a recursive expression for the survival function in terms of the (lower-dimensional) marginal multivariate survival functions and the Péano series for the multivariate hazard, with respect to the given partial order. The marginal multivariate survival functions depend in exactly the same way on the corresponding marginal multivariate hazards $\Lambda(dt_C) = \Lambda(dt_C|t_C)$ together with even lower dimensional marginals. The marginal hazards (a subset of the set of conditional hazards) can be estimated as before, and the basic asymptotic properties of the corresponding Volterra estimator are derivable form the second set of continuity and differentiability results of Section 3.4.

Both of these estimators are symmetric in the coordinates $\{1, \dots, k\}$, though unfortunately neither is usually, strictly speaking, a survival function: They will assign negative mass to some regions. Nor is either asymptotically efficient —the first estimator seems to use more of the information in the data, but does not combine it optimally—though each could be used as the starting point of a one-step Newton–Raphson iteration on the log-likelihood for an appropriately finely discretized model, leading presumably to an asymptotically efficient estimator [cf. Bickel, Ritov and Wellner (1991) for this approach in a somewhat simpler setting]. Aad van der Vaart (private communication) conjectures that no efficient estimator is a differentiable functional of the empirical data; if this is true, one must choose between using one of several different smoothly behaved but inefficient estimators and an essentially unique but unstable efficient one.

More on the derivation of these estimators and their large sample properties will be given in Bickel, Gill and Wellner (1990). It appears that the Dabrowska estimator is usually rather good and the Volterra estimator surprisingly rather poor, especially at large t. The pathwise estimator is of intermediate quality. At independence of all survival times and censoring times the Dabrowska estimator is amazingly but quite coincidentally asymptotically efficient.

4.4. *Markov processes*. In this section we generalize the results of Sections 4.1 and 4.2 in another direction: to multiple states, rather than to multidimensional time. The survival analysis model described by a single (censored) survival time can also be considered as a very simple Markov process model, with just two states—alive and dead—and a transition between them at the survival time T. On top of this process, censoring is superimposed as a kind of nuisance factor; alternatively, considering the censoring on an equal footing (a competing risk), the process can be described as a three-state Markov process: alive, censored and dead. It also turns out [Keiding and Gill (1990)] that another important model from survival analysis, the random truncation or delayed entry model, can also be considered as a three-state Markov model, the three states being preentry, entered but not yet failed, and dead.

More complex Markov models are also frequently used in survival analysis, for instance, to model progression between various stages of a disease: the three-state illness–death–recovery model; four-state models for studying the interaction between two life history events, and so on. On top of these models delayed entry ("immigration") and censoring ("emigration") can be imposed. The book of Andersen, Borgan, Gill and Keiding (1990) will contain a very thorough study of practical and theoretical aspects of such statistical models.

It turns out that the analysis of such models can be completely identified with the product-integral–based treatment of Sections 4.1 and 4.2 by simply taking the step from the scalar case to the $p \times p$ matrix case, where p is the number of states in the model. The statistical aspects of this were first worked on by Aalen and Johansen (1978), building on the probabilistic work of Dobrushin (1953) and his predecessors. Here we shall give some of the details behind all this work as well as sketching some further results. Consider a time-inhomogeneous Markov process X_t, $t \in [0, \infty[$ on a finite state space E, with transition probabilities

$$p_{ij}(s,t) = \Pr\{X_t = j | X_s = i\}.$$

It is well known that the transition probabilities satisfy the Chapman–Kolmogorov equations; that is, if we define $P(s,t) = \{p_{ij}(s,t): i, j \in E\}$, then

$$P(s,t) = P(s,u)P(u,t), \qquad 0 \le s \le u \le t < \infty, \tag{67}$$

$$P(s,s) = \mathbf{1}, \qquad 0 \le s < \infty. \tag{68}$$

If $P(\cdot,\cdot)$ is differentiable in both arguments, one can prove that

$$Q(t) = \left.\frac{\partial P(s,t)}{\partial t}\right|_{t=s} = -\left.\frac{\partial P(s,t)}{\partial s}\right|_{s=t} \tag{69}$$

and that $P(\cdot, \cdot)$ satisfies the forward and backward Kolmogorov equations,

$$\frac{\partial P(s,t)}{\partial t} = P(s,t)Q(t), \tag{70}$$

$$\frac{\partial P(s,t)}{\partial s} = -Q(s)P(s,t), \tag{71}$$

with initial conditions $P(s, s) = \mathbf{1}$. The solution of either of these equations is unique.

It is clear that in this case $P(s, t)$ is a multiplicative matrix-valued interval function, and the differential equations for P (with the uniqueness of their solution) are special cases of the forward and backward equations of Theorem 5. The solution is given by the product-integral of the matrix-valued additive interval function, or measure,

$$\alpha(s,t) = \int_{]s,t]} Q(u)\, du. \tag{72}$$

With this formulation one can say that the problem of determining the transition probabilities from the intensities is just the problem of product-integrating the intensity measure. Similarly, the problem of determining the intensities or intensity measure is solved by the additive integration of the interval function $P(s, t) - \mathbf{1}$. In other words, the differentiation (70) followed by the integration (72) can be replaced by the single process of integrating an interval function.

These relations have very simple difference-equation analogues for a discrete time Markov chain. If X_t is a time-inhomogeneous Markov process which only jumps at integer times (s, t, etc.), then one can let $Q(s) = P(s - 1, s) - \mathbf{1}$ and recover P from the discrete intensity measure $\alpha(s, t) = \sum_{s < u \le t} Q(u)$ by the difference equation [cf. (70)] $P(s, t) - P(s, t - 1) = P(s, t - 1)Q(t)$ with initial condition (68).

Thus the well-known relations (67)–(72) are special cases of the general concepts of product-integrals and sum-integrals. This is especially important for statistical applications because we then often meet product-limit estimators as solutions of equations like (70) or (71), in which the intensity measure α is now a discrete empirical measure. The general concepts treated in the previous sections show that one can construct the solution to the equation for given Q or rather α in just the same way whether one is working with absolutely continuous or with discrete α, so that the statistical calculations become identical to the probabilistic calculations.

One can also use product-integral theory to derive properties of the estimators. In particular, the Duhamel equation gives a stochastic differential equation for the estimators [the matrix generalization of (66) of Section 4.2] that allows the theory of martingales to be applied and hence one can easily find the asymptotic distribution of the estimators (one could also, in fact, use the δ-method). Such a programme is carried out by Aalen and Johansen (1978). Here we mainly discuss the probabilistic part of the problem, that is, the existence of an intensity measure for an arbitrarily given Markov process and vice-versa. We also pay some attention to the martingale connection: The intensity measure provides the deterministic part of the stochastic intensity of the counting processes counting direct moves or jumps between states.

First, we define an intensity measure as a matrix-valued measure or additive interval function α on the Borel sets of $]0, \infty[$ such that $\alpha(s,t) = \alpha(]s,t])$ is finite on bounded sets and such that

$$\alpha_{ii}(s,t) \leq 0, \qquad \alpha_{ij}(s,t) \geq 0, \qquad i \neq j, \tag{73}$$

$$\sum_{j \in E} \alpha_{ij}(s,t) = 0. \tag{74}$$

It is also necessary to assume that

$$\alpha_{ii}(\{t\}) \geq -1 \quad \text{for all } t. \tag{75}$$

Note that α is dominated by the real measure

$$\alpha_0(s,t) = -2 \operatorname{trace} \alpha(s,t),$$

which is of bounded variation on finite intervals. We then define

$$P(s,t) = \prod_{]s,t]} (\mathbf{1} + d\alpha) \tag{76}$$

and obtain the following result.

THEOREM 12. *The function P defined by* (76) *satisfies*

$$P(s,t) \quad \text{is a stochastic matrix} \tag{77}$$

$$P(s,t) = P(s,u)P(u,t), \qquad 0 \leq s \leq u \leq t < \infty, \tag{78}$$

$$P(s,s) = \mathbf{1}, \qquad 0 \leq s < \infty, \tag{79}$$

$$P(s,t) \to \mathbf{1} \qquad \text{as } t \downarrow s. \tag{80}$$

PROOF. The assumptions (73)–(75) about α imply that $P(s,t)$ is stochastic, being a limit of products of stochastic matrices, and (78), (79) and (80) follow from the properties of the product-integral. □

Note that if α has an atom at the point t, then P will have a discontinuity at t and $P(s,t) \to \mathbf{1} + \alpha(\{t\})$ as $s \uparrow t$.

THEOREM 13. *The function P defined by* (76) *satisfies the Kolmogorov equations*

$$\frac{dP(s,\cdot)}{d\alpha_0}(t) = P(s,t-)\frac{d\alpha}{d\alpha_0}(t) \quad \text{for } \alpha_0\text{-almost all } t \in]s,\infty[, \tag{81}$$

$$\frac{dP(\cdot,t)}{d\alpha_0}(s) = -\frac{d\alpha}{d\alpha_0}(s)P(s,t) \quad \text{for } \alpha_0\text{-almost all } s \in [0,t[. \tag{82}$$

PROOF. From the forward equation of Theorem 5 we have

$$P(s,t) = \mathbf{1} + \int_{]s,t]} P(s,u-)\alpha(du),$$

and hence the function $t \to P(s,t)$ is absolutely continuous with respect to α_0. Taking Radon–Nikodym derivatives, we get (72). Relation (73) follows similarly from the backward equation. □

We finally note the following.

THEOREM 14. *The interval function $P - \mathbf{1}$ is of bounded variation.*

PROOF. This is just Theorem 1. □

We shall now discuss the inverse problem. Let Markov transition probabilities P be given.

THEOREM 15. *If the transition probabilities $P(s,t)$ are right continuous and if $P - \mathbf{1}$ is of bounded variation, then P is the product-integral of the intensity measure given by*

$$\alpha(s,t) = \int_{]s,t]} d(P - \mathbf{1}).$$

PROOF. This is Theorems 2 and 3. □

Note that the Péano series representation (Theorem 4) is not the series representation of the minimal solution as given by Feller (1940). This is most easily seen by comparing the first term which for the Péano series is just $\mathbf{1}$ whereas the minimal solution starts with the matrix ${}_0P$ with elements

$${}_0p_{ij}(s,t) = \delta_{ij} \prod_{]s,t]} (1 - d\alpha_{ii}).$$

The terms of the Péano series need not even be positive whereas the nth term of the series for the minimal solution has an interpretation as the probability of jumping from i to j in exactly n steps. Informally, the meaning of the bounded variation assumption is that if any particular state is kept artificially occupied by introducing a new particle into it whenever another one leaves, then the expected number of jumps out of the state during any finite interval is finite.

We now briefly discuss the construction of the underlying process X_t. Under the equivalent assumptions of Theorem 12 or 15 there exists a Markov process (with the given transition matrices) with piecewise-constant sample paths which are right continuous and have finitely many jumps on finite intervals. The process is well defined, starting from any given state at any given time point. The crucial assumption that makes the construction of these processes possible is the assumption of bounded variation. The idea of the construction is that from the time t_0 of entering state i, the intensity measure of the random time at which the process leaves the state is given by $-\alpha_{ii}$ restricted to $]t_0, \infty[$, and given that a jump from i occurs at time t, then the probability that the new state is j is $-(d\alpha_{ij}/d\alpha_{ii})(t)$. The process is started at time zero in an arbitrary given state. So one simply constructs the process from a sequence of alternating jump times and jump states, each having the just-specified conditional distribution given its predecessors. In fact, this sequence of pairs of random variables is itself a discrete time Markov process.

The mathematical task is to show that this construction actually defines X_t for all t; in other words, that almost surely only a finite number of jumps occur in any finite time interval. Also one must show that the process X is indeed Markov, with the same transition probabilities and intensities as one started with. Most of these ingredients can be found in Jacobsen (1972) and Johansen (1986). Actually Jacobsen (1972) does not allow atoms of size -1 in the diagonal elements of α, meaning a certain jump out of the relevant state or states at that time. However, it is quite easy to put together the pieces of his construction from one atom to the next—the finiteness of α means there are only finitely many pieces to be considered in finite time intervals.

One can now go on to show, by a straightforward use of the theory of marked point processes of Jacod (1975) or Jacobsen (1982), that the counting process

$$N_{ij}(t) = \sum_{s \le t} 1\{X_{s-} = i, X_s = j\}, \qquad i \ne j,$$

has a predictable compensator given by

$$\Lambda_{ij}(t) = \int_{]0,t]} 1\{X_{s-} = i\} \alpha_{ij}(ds). \tag{83}$$

(Full details of all this are given in an earlier version of this paper, available from the authors.)

Since jumps of the processes N_{ij} do not occur simultaneously, $\{M_{ij} = N_{ij} - \Lambda_{ij}\}$ is a collection of orthogonal square-integrable martingales. In other words,

$$N_{ij} = \Lambda_{ij} + M_{ij} \tag{84}$$

is the Doob–Meyer decomposition of N_{ij}. Note that $EN_{ij}(t) = E\Lambda_{ij}(t) \le \alpha_{ij}(0, t) < \infty$ for all $t < \infty$.

We close this section with some remarks on statistical applications of these ideas. We have already mentioned the paper of Aalen and Johansen (1978) on nonparametric estimation of the transition probabilities of an inhomogeneous Markov process, based on, say, n censored observations from the process. The idea is to estimate the intensity measure α_{ij} as the discrete measure, with atoms at each time t of observed jumps from i to j, equal in size to the number of processes observed to make this jump at time t, divided by the number of processes observed to be in state i at time $t-$. Using the

Doob–Meyer decomposition of the counting processes N_{ij} just described and the Duhamel equation leads to a derivation of the large sample properties of the corresponding estimator of P, completely parallel to (and containing as a special case) that of the Kaplan–Meier estimator. Many applications of these estimators together with further results are given by Andersen, Borgan, Gill and Keiding (1990). We also mention Hjort (1984), who gives a Bayesian treatment of the nonparametric estimation of the intensity measure of a Markov process while Hjort, Natvig and Funnemark (1985) give a reliability application, using the product-limit approximation to the transition matrix in order to derive results on the association between states in time.

One can also use product-integral methods to derive the asymptotic distribution of the processes $N_{ij}^{(n)}$ counting jumps between pairs of states aggregated over n (uncensored) realizations of the process X. This can be needed in statistical situations with incomplete data; see Borgan and Ramlau-Hansen (1985) and Gill (1986) for two very different applications. Here is a sketch of the idea, due to Odd Aalen. Let $Y_i^{(n)}(t)$ be the number of processes in state i at time $t-$. Note that

$$Y_i^{(n)}(t) = Y_i^{(n)}(0) + \sum_{j \neq i} N_{ji}^{(n)}(t-) - \sum_{j \neq i} N_{ij}^{(n)}(t-);$$

counting net entries to and exits from state i up to time t gives us the number there at t. For simplicity, suppose all the processes start at time zero in the same fixed state.

Collect the elements $N_{ij}^{(n)}$, $i \neq j$, into a row vector $N^{(n)}$. Adding n copies of (83) and (84), these two equations can be rewritten in the form

$$dN^{(n)} = N_-^{(n)}\,dA + n\,dB + dM^{(n)},$$

where A and B are $\frac{1}{2}p(p-1) \times \frac{1}{2}p(p-1)$ matrices containing $\pm\alpha_{ij}$ and 0 as elements in appropriate places.

Subtracting expectations and multiplying by $n^{-1/2}$ gives the equation

$$dZ^{(n)} = Z_-^{(n)}\,dA + dW^{(n)}, \tag{85}$$

where $Z^{(n)} = n^{-1/2}(N^{(n)} - EN^{(n)})$ and $W^{(n)} = n^{-1/2}M^{(n)}$ is a row vector of martingales which can be shown by the martingale central limit theorem [see, e.g., the version for stochastic integrals of counting processes with possibly discrete compensators of Gill (1980a)] to converge in distribution for each $\tau < \infty$ in $((D[0,\tau])^{(1/2)p(p-1)}, \|\cdot\|_\infty)$ to a row vector of Gaussian martingales Z, with variance functions

$$\operatorname{var} W_{ij}(t) = \int_{]0,t]} EY_i(s)\big(1 - \alpha_{ij}(\{s\})\big)\alpha_{ij}(ds),$$

independent over i, and if the intensity measures are continuous also over j, but otherwise with covariance functions

$$\operatorname{cov}(W_{ij}, W_{ij'}) = -\int_{]0,t]} EY_i(s)\alpha_{ij}(\{s\})\alpha_{ij'}(ds).$$

Equation (85), an inhomogeneous equation as in Theorem 10 of Section 3.2 (the matrices ϕ and ψ there need not be square and are now in fact taken to be row vectors), has the explicit solution

$$\begin{aligned}Z^{(n)}(t) &= W^{(n)}(t) + \int_{]0,t]} W^{(n)}(s-)A(ds)\prod_{]s,t]}(1+dA)\\ &= W^{(n)}(t) + \int_{]0,t]} W_{-}^{(n)}\, d\prod_{]\cdot,t]}(1+dA)\\ &= \int_{]0,t]} dW^{(n)}\prod_{]\cdot,t]}(1+dA).\end{aligned}$$

Thus we see that since the intensity measure is assumed finite on finite intervals, $Z^{(n)}$ is a (supremum norm) continuous (indeed, a linear) function of $W^{(n)}$ on such intervals, and therefore weak convergence of $W^{(n)}$ is carried over to $Z^{(n)}$. Moreover, and especially useful in statistical applications, the covariance structure of the process $Z^{(n)}$ can be very easily numerically calculated for given intensities by iterating the natural approximate difference equations derived from a discrete approximation to (85); see Gill and Keilman (1990) for an example. (Earlier formulas involved very complicated multiple integrals but now only single integrals are involved thanks to the nice covariance structure of W.)

This result could have been derived directly from the central limit theorem for semimartingales of Jacod and Shiryaev [(1978), Theorem IX.3.39, cf. also IX.4.31].

4.5. *Markov branching processes.* Arley (1943), with a view to applications in Markov models for cosmic ray showers, investigated the product-integral (with respect to Lebesgue measure) of continuous matrix functions $A(t)$ with countably many rows and columns and satisfying the condition that

$$\exp\Big(\sup_{a<t<b} A_{ij}(t)\Big)$$

is finite. We shall not discuss the theory in this generality but consider the special case of an inhomogeneous-time Markov branching process, that is, a *countable state space* Markov chain, with the extra structure that the nth row of the transition matrix is the n-fold convolution of the first row. Such a process represents the growth of a population by simple independent splitting of individuals at their random death times in a time-varying environment. For such processes it turns out that the product-integral defined in the previous sections can be applied to study the moments of the process, and this has useful consequences for the statistical estimation of the parameters of the process—the integrated death rates and the time-dependent probability of the family size produced on death.

Consider therefore a Markovian population of individuals developing independently of each other and who at death produce offspring. We assume that initially, that is, $t = 0$, we have $X(0)$ individuals, and denote by $G(s,t)$ the probability that an individual, alive at time s, will survive time t. An individual dying at time t will produce j offspring with probability $p_j(t)$, $j = 0, 1, \dots$. We write $G(t) = G(0, t)$, the survival function of an individual born at time zero, which is right continuous and nonincreasing. We shall assume that $G(t) > 0$ for all $t < \tau \leq \infty$ and $G(\tau) = 0$, so that all surviving individuals die at time $t = \tau$ and, by the Markov property, $G(s, t) = G(0, t)/G(0, s)$ for $s < t \leq \tau$.

The process can most easily be described by the counting processes $N_j(t)$, which counts the number of j-births before or at time t. Then $N = \sum N_j$ is the number of deaths before or at time t, and $X(t) = X(0) + \sum(j-1)N_j(t)$ is the number of individuals alive at time t.

We shall discuss the functions (assumed finite for all $t \leq \tau$)

$$m(s,t) = E(X(t)|X(s) = 1),$$
$$v(s,t) = \mathrm{Var}(X(t)|X(s) = 1)$$

and show how the relationships between G, p_j, m and v can be nicely described in product-integral notation. This allows one to derive maximum likelihood estimators for these functions and to apply the Duhamel equation to find the properties of the estimators, as is done in more detail by Johansen (1981).

It is not difficult to show, using the Markov property, that m is multiplicative,

$$m(s,t) = m(s,u)m(u,t),$$

and the corresponding additive interval function is

$$\mu_1(s,t) = \int_{]s,t]} \sum (j-1)p_j(u)\frac{G(du)}{G(u-)},$$

so that

$$m(s,t) = \prod_{]s,t]} (1 + d\mu_1). \tag{86}$$

The statistical model we get by considering the infinite-dimensional parameters $G(s,t)$ and $p_j(t)$, $j = 0, 1, \dots,$ is a semiparametric model since the transition probabilities $P_{ij}(s,t)$ can no longer vary freely as in the previous section. It is shown in Johansen (1981) that the nonparametric maximum likelihood estimators of p and G are given by

$$\hat{p}_j(u) = \frac{dN_j}{dN}(u)$$

and

$$\hat{G}(s,t) = \prod_{]s,t]} \left(1 - \frac{dN}{X_-}\right).$$

It follows that

$$\hat{\mu}_1(s,t) = \int_{]s,t]} \sum (j-1) \frac{dN_j}{X_-}$$

and

$$\hat{m}(s,t) = \frac{X(t)}{X(s)}.$$

In order to find an expression for the variance, we introduce the generating function

$$\phi(s,t;z) = \int_{]s,t]} \sum (z^j - z) p_j(u) \frac{G(du)}{G(u-)}.$$

Then $\mu_1(s,t) = \phi'(s,t;1)$, and we also define $\mu_2(s,t) = \phi''(s,t;1)$ and

$$\mu_3(s,t) = \int_{]s,t]} \left(\left(1 + \frac{d\mu_1}{dG}(u) G(\{u\}) \right)^2 - 1 \right) G(\{u\})^{-1} G(du),$$

where the integrand is interpreted by continuity if $G(\{u\}) = 0$. Finally, $\mu_4 = \mu_1 + \mu_2 - \mu_3$. Then one can show that for

$$M = \begin{pmatrix} m & v \\ 0 & m^2 \end{pmatrix} \quad \text{and} \quad \mu = \begin{pmatrix} \mu_1 & \mu_4 \\ 0 & \mu_3 \end{pmatrix},$$

we have

$$M(s,t) - \mathbf{1} = \prod_{]s,t]} (\mathbf{1} + d\mu).$$

The upper left-hand corner of this relation is just (86), and m^2 has a similar representation in terms of μ_3. If we insert

$$\hat{\phi}(s,t;z) = \int_{]s,t]} \sum (z^j - z) \frac{dN_j}{X_-},$$

one can find $\hat{\mu}$ and finally $\hat{v}$. It turns out that

$$\hat{v}(s,t) = \frac{X(t)^2}{X(s)} \int_{]s,t]} \sum j(j-1) \frac{dN_j}{(X_- + j - 1)^2} - \left(\frac{X(t)}{X(s)} \right)^2 + \frac{X(t)}{X(s)},$$

which expresses $\hat{v}$ as a stochastic integral. Applying the product-integral representation for M, one can find a Duhamel equation that expresses M as a stochastic integral with respect to the "innovations" $\hat{\mu} - \mu$. In fact, this latter process, stopped at the time of death of the last living individual, is a sum over j of stochastic integrals with respect to the counting process martingales $N_j(t) - \int_{]0,t]} p_j(s) X(s-) G(ds)/G(s-)$. (One could alternatively apply the differentiability results of Section 3.1 to derive the properties of $\hat{v}$.) Full details can be found in Johansen (1981).

4.6. *Counting process likelihoods.* In this section we give Jacod's (1975) formula for the Radon–Nikodym derivative of two probability measures on the filtered space generated by a multivariate counting process. Following Johansen (1983), we show that this extremely important but intuitively unappealing formula can be given a natural probabilistic interpretation by recasting it in terms of product-integrals. Let $N = (N_1, \dots, N_k)$ be a multivariate counting process with compensator $A = (A_1, \dots, A_k)$ with respect to a probability measure P and a filtration of the special form

$$\mathscr{F}_t = \mathscr{F}_0 \vee \sigma\{N(s): s \le t\}, \qquad t \in \mathscr{T} = [0, \tau].$$

[As explained in Jacod and Shiryaev (1987), the usual assumption of completeness of the filtration is superfluous.] Let P' be another probability measure, dominated by P. Under P', N is still a multivariate counting process with respect to this filtration but its compensator is generally different; let us take it to be A' then. Recall that A can be interpreted as an integrated conditional intensity by the heuristic

$$dA_i(t) = P\{dN_i(t) = 1 | \mathscr{F}_{t-}\}. \tag{87}$$

So, given P on $\mathscr{F}_0$, one should be able to reconstruct P on $\mathscr{F}_\tau$ by multiplication of conditional probabilities. The next theorem makes this idea rigorous. Indeed, the distribution of N (given $\mathscr{F}_0$) is determined by its compensator, though in the first instance not so obviously in the way just indicated.

THEOREM 16 [Jacod (1975)]. *Let $L_t = (dP'/dP)|_{\mathscr{F}_t}$. Almost surely-$(P)$, $A'_i \ll A_i$ for each i and*

$$\begin{aligned} L_t = L_0 &\prod_{n: T_n \le t} \left(\frac{dA'_{J_n}}{dA_{J_n}} \right) (T_n) \\ &\times \prod_{s \ne T_n,\, s \le t} \frac{1 - \Delta \bar{A}'(s)}{1 - \Delta \bar{A}(s)} \exp(-\bar{A}^{c\prime}(t) + \bar{A}^c(t)). \end{aligned} \tag{88}$$

Here $0 < T_1 < T_2 < \cdots$ are the jump times of N and $J_1, J_2, \dots$ are the corresponding jump types, that is, $\Delta N_{J_n}(T_n) = 1$. Furthermore, $\bar{N} = \sum_i N_i$, $\bar{A} = \sum_i A_i$, $\Delta \bar{N}$ denotes the jumps of $\bar{N}$, A^c is the continuous part of A, and so on.

To better understand the theorem, we rewrite (88) using product-integral notation:

$$L_t = L_0 \frac{\prod_{s \le t} \left((1 - d\bar{A}'(s))^{1 - \Delta \bar{N}(s)} \prod_i (dA'_i(s))^{\Delta N_i(s)} \right)}{\prod_{s \le t} \left((1 - d\bar{A}(s))^{1 - \Delta \bar{N}(s)} \prod_i (dA_i(s))^{\Delta N_i(s)} \right)}. \tag{89}$$

This expression should be evaluated by writing the product-integrals as the products of an ordinary product over the finite set of jump times and a product-integral over the interval $[0, t]$ less the jump times, together with the convention

$$\frac{dA'_i(s)}{dA_i(s)} = \frac{dA'_i}{dA_i}(s).$$

Note the probabilistic interpretation of numerator and denominator of equation (89), corresponding exactly to (87): The likelihood function is formed by multiplying together conditional likelihoods (given $\mathscr{F}_{t-}$) for infinitesimal experiments in which $dN(t)$ is generated by letting component i equal 1 and all other components equal 0 with conditional probability $dA_i(t)$, and letting *all* components be 0, so $1 - d\overline{N}(s) = 1$, with the complementary probability $1 - d\overline{A}(s)$. Expressions such as the numerator and denominator of (89) are common in heuristic calculations in survival analysis [see, for instance, Kalbfleisch and Prentice (1980)]. The fact that they also have an exact mathematical interpretation allows one [see Andersen, Borgan, Gill and Keiding (1990)] to construct a rigorous but at the same time transparent derivation of partial likelihood functions and the notions of noninformative and independent censoring. In particular, the important results of Arjas and Haara (1984) can be clarified in this way. The result can be immediately generalized to marked point processes.

Rewriting the ratio in (89) as a single product-integral allows one to make use of the Volterra equation characterization of the product-integral. This yields nothing more than the fact, well-known in the counting process literature, that the likelihood ratio process is the Doléans-Dade exponential of the P-martingale Z with

$$dZ = \sum_i \frac{dA'_i}{dA_i}\, dM_i + \frac{1 - \Delta\overline{A}'}{1 - \Delta\overline{A}}\, d\overline{M},$$

where M_i is the counting process martingale $N_i - A_i$.

One can also consider (89) exactly as it is written, that is, as the ratio of two product-integrals. Subtracting 1 from each side of the equation allows us to rewrite the right-hand side as a Duhamel equation. Replacing P and P' with $P_{\theta_0 + d\theta}$ and P_{θ_0} and taking a limit produces finally a (well-known) stochastic integral equation for the score function in a parametric counting process model, in fact, as an integral over the infinitesimal experiments described previously of the conditional score for each experiment given the preceding ones. Since the expected conditional score given the past is zero, this stochastic integral is also a martingale. Further relationships can be obtained relating the integral of the expected conditional Fisher information in each small experiment and the total information. See Andersen, Borgan, Gill and Keiding (1990) for consequences of these results for partial-likelihood–based statistical analysis, and see Efron and Johnstone (1990) for the same result in the context of the simple survival analysis model, that is, a univariate counting process which can make at most one jump. Our main point here is simply that many of these originally algebraically involved results can be simply understood and written using product-integral notation in intuitively very appealing and fruitful ways.

REFERENCES

AALEN, O. O. (1975). Statistical inference for a family of counting processes. Ph.D. dissertation, Univ. California, Berkeley. [Reprinted (1976) at Inst. Math. Statist., Univ. Copenhagen.]

AALEN, O. O. and JOHANSEN, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scand. J. Statist.* **5** 141–150.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1990). Statistical models based on counting processes. Unpublished.

ARJAS, E. and HAARA, P. (1984). A marked point process approach to censored failure data with complicated covariates. *Scand. J. Statist.* **11** 193–209.

ARLEY, N. (1943). *On the Theory of Stochastic Processes and their Application to the Theory of Cosmic Radiation*. G. E. C. Gads Forlag, Copenhagen.

ARLEY, N. and BORSCHENIUS, V. (1945). On the theory of infinite systems of differential equations and their application to the theory of stochastic processes and the perturbation theory of quantum mechanics. *Acta Math.* **76** 261–322.

AVRAM, F. and TAQQU, S. (1986). Symmetric polynomials of random variables attracted by an infinitely divisible law. *Probab. Theory Related Fields* **71** 491–500.

BEESACK, P. R. (1975). *Gronwall Inequalities. Carleton Math. Lecture Notes* **11**. Carleton Univ., Ottawa.

BERAN, R. (1981). Nonparametric regression with randomly censored data. Unpublished.

BICHTELER, K. (1981). Stochastic integration and L^p-theory of semimartingales. *Ann. Probab.* **9** 49–89.

BICKEL, P. J., GILL, R. D. and WELLNER, J. A. (1990). Inefficient estimates for three bivariate models. Unpublished.

BICKEL, P. J., RITOV, Y. and WELLNER, J. A. (1991). Efficient estimation of linear functionals of a probability measure P with known marginal distributions. *Ann. Statist.* **19**. To appear.

BILLINGSLEY, P. (1968). *Weak Convergence of Probability Measures*. Wiley, New York.

BIRKHOFF, G. D. (1938). On product integration. *J. Math. Phys.* **16** 104–132.

BORGAN, Ø. and RAMLAU-HANSEN, H. (1985). Demographic incidence rates and the estimation of intensities with incomplete information. *Ann. Statist.* **13** 564–582.

BRESLOW, N. and CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–453.

BURKE, M. D., CSÖRGŐ, S. and HORVÁTH, L. (1981). Strong approximation of some biometric estimators under random censorship. *Z. Wahrsch. Verw. Gebiete* **56** 87–112.

CAMPBELL, G. and FÖLDES, A. (1982). Large-sample properties of nonparametric bivariate estimators with censored data. In *Nonparametric Statistical Inference* (B. V. Gnedenko, M. L. Puri and I. Vincze, eds.) 103–122. North-Holland, Amsterdam.

COX, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–200.

DABROWSKA, D. (1988). Kaplan–Meier estimation on the plane. *Ann. Statist.* **16** 1475–1489.

DABROWSKA, D. (1989). Kaplan–Meier on the plane: Weak convergence, LIL and the bootstrap. *J. Multivariate Anal.* **29** 308–325.

DELLACHERIE, C. and MEYER, P.-A. (1982). *Probabilities and Potential B: Theory of Martingales*. North-Holland, Amsterdam.

DENBY, L. and VARDI, Y. (1985). A short-cut method for estimation of renewal processes. *Technometrics* **27** 361–373.

DOBRUSHIN, R. L. (1953). Generalization of Kolmogorov's equations for Markov processes with a finite number of possible states. *Mat. Sb.* (*N.S.*) **33** 567–596. (In Russian.)

DOLÉANS-DADE, C. (1969). Variation quadratique des martingales continues à droite. *Ann. Math. Statist.* **40** 284–289.

DOLÉANS-DADE, C. (1970). Quelques applications de la formule de changement de variables pour les semimartingales. *Z. Wahrsch. Verw. Gebiete* **16** 181–194.

DOLLARD, J. D. and FRIEDMAN, C. N. (1979). *Product Integration with Applications to Differential Equations* (with an appendix by P. R. Masani). Addison-Wesley, Reading, Mass.

DUDLEY, R. M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

EFRON, B. and JOHNSTONE, I. (1990). Fisher's information in terms of the hazard rate. *Ann. Statist.* **18** 38–62.

EMERY, M. (1978). Stabilité des solutions des équations différentielles stochastiques: Application aux integrales multiplicatives stochastiques. *Z. Wahrsch. Verw. Gebiete* **41** 241–262.

FELLER, W. (1940). On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc.* **48** 488–515. [Erratum (1945) **58** 474.]

FÖLLMER, H. (1981). Calcul d'Itô sans probabilités. *Séminaire de Probabilités XV. Lecture Notes in Math.* **850** 143–150. Springer, Berlin.

GILL, R. D. (1980a). *Censoring and Stochastic Integrals.* MČ Tract **124**. Mathematisch Centrum, Amsterdam.

GILL, R. D. (1980b). Nonparametric estimation based on censored observation of a Markov renewal process. *Z. Wahrsch. Verw. Gebiete* **53** 97–116.

GILL, R. D. (1981). Testing with replacement and the product-limit estimator. *Ann. Statist.* **9** 853–860.

GILL, R. D. (1983). Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

GILL, R. D. (1986). On estimating transition intensities of a Markov process with aggregate data of a certain type: "Occurrences but no exposures." *Scand. J. Statist.* **13** 113–134.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method I. *Scand. J. Statist.* **16** 97–128 (with discussion); II. Unpublished.

GILL, R. D. and KEILMAN, N. (1990). On the estimation of multidimensional demographic models with population registration data. *Math. Pop. Studies* **2** 119–143.

GRONWALL, T. H. (1919). Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Ann. of Math.* **20** 292–296.

HAKIM-DOWEK, M. and LÉPINGLE, D. (1986). L'exponentielle stochastique des groupes de Lie. *Séminaire de Probabilités XX. Lecture Notes in Math.* **1204** 352–374. Springer, Berlin.

HELTON, B. W. (1966). Integral equations and product integrals. *Pacific J. Math.* **16** 297–322.

HELTON, B. W. (1969). A product integral representation for a Gronwall inequality. *Proc. Amer. Math. Soc.* **23** 493–500.

HELTON, J. C. (1975a). Mutual existence of sum and product integrals. *Pacific J. Math.* **56** 495–516.

HELTON, J. C. (1975b). Product integrals and the solution of integral equations. *Pacific J. Math.* **58** 87–103.

HELTON, J. C. (1977). Two generalizations of the Gronwall inequality by product integration. *Rocky Mountain J. Math.* **7** 733–749.

HELTON, J. C. (1978). Solution of two Volterra integral equations. *Rocky Mountain J. Math.* **8** 547–554.

HJORT, N. L. (1984). Nonparametric Bayes estimators of cumulative intensities in models with censoring. Report 762, Norwegian Computing Centre, Oslo.

HJORT, N. L., NATVIG, B. and FUNNEMARK, E. (1985). The association in time of a Markov process with application to multistate reliability theory. *J. Appl. Probab.* **22** 473–479.

JACOBSEN, M. (1972). A characterization of minimal Markov jump processes. *Z. Wahrsch. Verw. Gebiete* **23** 32–46.

JACOBSEN, M. (1982). *Statistical Analysis of Counting Processes. Lecture Notes in Statist.* **12**. Springer, Berlin.

JACOD, J. (1975). Multivariate point processes: Predictable projection, Radon–Nikodym derivatives, representation of martingales. *Z. Wahrsch. Verw. Gebiete* **31** 235–253.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, Berlin.

JOHANSEN, S. (1978). The product limit estimator as maximum likelihood estimator. *Scand. J. Statist.* **5** 195–199.

JOHANSEN, S. (1981). The statistical analysis of a Markov branching process. Preprint 5, Inst. Math. Statist., Univ. Copenhagen.

JOHANSEN, S. (1983). An extension of Cox's regression model. *Internat. Statist. Rev.* **51** 165–174.

JOHANSEN, S. (1986). Product integrals and Markov processes. *CWI Newsletter* (*12*) 3–13 [Originally appeared (1977) as Preprint 3, Inst. Math. Statist., Univ. Copenhagen.]

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

KARANDIKAR, R. L. (1983). Multiplicative stochastic integration. In *Prediction Theory and Harmonic Analysis* (V. Mandrekar and H. Salehi, eds.) 191–199. North-Holland, Amsterdam.

KEIDING, N. and GILL, R. D. (1990). Random truncation models and Markov processes. *Ann. Statist.* **18** 582–602.

LASLETT, G. M. (1982). The survival curve under monotone density constraints with applications to two-dimensional line segment processes. *Biometrika* **69** 153–160.

LÉVY, P. (1948). *Processus Stochastiques et Mouvement Brownien*. Gauthier-Villars, Paris.

MAC NERNEY, J. S. (1963). Integral equations and semigroups. *Illinois J. Math.* **7** 148–173.

MASANI, P. (1947). Multiplicative Riemann integration in normed rings. *Trans. Amer. Math. Soc.* **16** 147–192.

MASANI, P. (1966). Wiener's contributions to generalized harmonic analysis, prediction theory and filter theory. *Bull. Amer. Math. Soc.* **72** 73–125.

MASANI, P. (1979). Appendix II. In *Product Integration with Application to Differential Equations* (J. D. Dollard and C. N. Friedman, eds.). Addison-Wesley, Reading, Mass.

MCKEAN, H. P. (1960). Brownian motions on 3-dimensional rotation groups. *Mem. Coll. Sci. Kyoto Univ.* **33** 25–38.

MUÑOZ, A. (1980). Nonparametric estimation from censored bivariate observations. Technical Report, Stanford Univ.

NELSON, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Tech.* **1** 27–52.

PÉANO, G. (1888). Intégration par séries des équations différentielles linéaires. *Math. Ann.* **32** 450–456.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, Berlin.

PROTTER, P. (1989). *Stochastic Integration and Differential Equations: A New Approach*. Springer, Berlin.

RASCH, G. (1930). *Om Matrixregning og dens Anvendelse paa Differens- og Differential-ligninger*. Levin and Munksgaards Forlag, Copenhagen.

RASCH, G. (1934). Zur Theorie und Anwendung des Productintegrals. *J. Reine Angew. Math.* **171** 65–119.

REEDS, J. A. (1976). On the definition of von Mises functionals. Research Report S-44, Dept. Statist., Harvard Univ. (Ph.D. dissertation.)

RITOV, Y. and WELLNER, J. A. (1988). Censoring, martingales and the Cox model. *Contemp. Math.* **80** 191–220.

SCHLESINGER, L. (1931). Neue Grundlagen für einen Infinitesimalkalkul der Matrizen. *Math. Z.* **33** 33–61.

SCHLESINGER, L. (1932). Weitere Beiträge zum Infinitesimalkalkul der Matrizen. *Math. Z.* **35** 485–501.

SCHMIDT, E. J. P. G. (1971). On multiplicative Lebesgue integration and families of evolution operators. *Math. Scand.* **29** 113–133.

SHEEHY, A. and WELLNER, J. A. (1990a). Uniform Donsker classes of functions. Technical Report 189, Univ. Washington.

SHEEHY, A. and WELLNER, J. A. (1990b). The delta method and the bootstrap for nonlinear functions of empiricals via Hadamard derivatives. Technical Report, Univ. Washington.

TSAI, W. J., LEURGANS, S. and CROWLEY, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* **14** 1351–1365.

VAN DER VAART, A. W. (1988). *Statistical Estimation for Large Parameter Spaces*. CWI Tract **44**. Centre for Mathematics and Computer Science, Amsterdam.

VAN DER VAART, A. W. and WELLNER, J. A. (1990). Prohorov and continuous mapping theorems in the Hoffmann–Jørgensen weak convergence theory, with applications to convolution and asymptotic minimax theorems. Technical Report 157, Univ. Washington.

VOLTERRA, V. (1887). Sulle equazioni differenziali lineari. *Rend. Accad. Lincei* (Series 4) **3** 393–396.

VOLTERRA, V. and HOSTINSKY, B. (1938). *Opérations Infinitésimales Linéaires. Applications aux Equations Différentielles et Fonctionelles*. Gauthier-Villars, Paris.

WALL, H. S. (1953). Concerning harmonic matrices. *Arch. Math.* **5** 160–167.

WELLNER, J. A. (1985). A heavy censoring limit theorem for the product limit estimator. *Ann. Statist.* **13** 150–162.

MATHEMATICAL INSTITUTE
UNIVERSITY OF UTRECHT
P.O. BOX 80.010
3508 TA UTRECHT
THE NETHERLANDS

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN
UNIVERSITETSPARKEN 5
2100 COPENHAGEN Ø
DENMARK