Richard D. Gill   Søren Johansen

Product-Integrals and Counting Processes

Preprint
June
1987

4

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

Richard D. Gill and Søren Johansen

PRODUCT-INTEGRALS AND COUNTING PROCESSES

Preprint 1987 No. 4

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

June 1987

Product-integrals and counting processes

by Richard D. Gill and Søren Johansen

Centrum voor Wiskunde en Informatica,
Kruislaan 413
1098 SJ Amsterdam,
Netherlands.

Institute of Mathematical Statistics,
University of Copenhagen,
Universitetsparken 5,
2100 Copenhagen Ø,
Denmark.

June 1987

ABSTRACT. The basic theory of the product-integral $\prod(\mathbf{1}+\mathrm{d}X)$ is summarized and applications in probability and statistics are discussed, in particular to non-homogeneous Markov processes, counting process likelihoods and the product-limit estimator.

KEY WORDS AND PHRASES. Markov process, multiplicative integral, Volterra integral equation, intensity measure, exponential semimartingale, compact (Hadamard) differentiability, survival analysis, product-limit (Kaplan-Meier) estimator.

MATHEMATICS CLASSIFICATION. Primary: 60J27, 62G05; secondary: 60H20, 45D05.

1. Introduction.

Product-integration has a long and respectable history in pure and applied mathematics. At the same time it has many applications in statistics and probability, especially in Markov processes, survival analysis and counting process models. 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 proofs, in a notation which, apart from that of the product integral $\prod(\mathbf{1}+\mathrm{d}X)$ itself, should be familiar to our intended audience. Some of our results are new though always close to previously published results. Also we shall discuss some old and new applications of product-integration in survival analysis (the product-limit estimator!), in the theory of Markov processes (the correspondence between transition probabilities and cumulative intensities), and in likelihood expressions for counting process experiments; and discuss the connection with the theory of exponential semimartingales.

The product-integral was introduced by Volterra (1887) as the solution of one of the simplest of the class of integral equations which now bears his name. The notion was further exploited and developed (mainly for real matrices) especially by Schlesinger (1931, 1932), Rasch (1931, 1934) and Birkhoff (1938). In particular Rasch, in his 1931 thesis, introduced the pregnant notation $\prod(\mathbf{1}+\mathrm{d}X)$ for the product-integral which is now favoured by most authors (unfortunately he decided to abandon this in favour of the poorer notation of Schlesinger in his later 'official' publication). All these papers are concerned with the (absolutely) continuous case. More recently, starting with a paper by Wall (1953), a quite abstract theory of product-integration was established in a stream of papers by Mac Nerney (1963), B. W. Helton (1966), and J. C. Helton (1975a,b) among others (here we have just indicated the major contributions of these authors). Especially relevant 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 quite strange. Among other things (as was already observed by Schlesinger), this abstract theory is most naturally stated in terms of *Riemann*-Stieltjes integrals, whereas in statistical theory the *Lebesgue*-Stieltjes integral is more familiar.

The major text-book on product-integration, by Dollard & Friedman (1979), is a mine of information but has for our purposes a major defect. The last-mentioned school of product-integration theory is only summarily mentioned (though at least extensive references are given), while instead the late and only chapter on *Stieltjes* product-integration - i.e. with respect to possibly discrete measures - treats, as far as all statistical applications are concerned, the *wrong* product integral $\prod e^{\mathrm{d}X}$. The authors (who are not alone in making this wrong choice) do point out that the theory of the 'exponential integral' contains many surprises and difficulties when one is used to the continuous-case results, but neglect to mention that this is not the case for $\prod(\mathbf{1}+\mathrm{d}X)$.

Product-integration makes a very natural appearance in the theory of Markov processes. A notable early work exploiting this is by Arley (1943); see also Arley & Borschenius (1945). Another important early paper 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, at least, in the context of Markov processes. This paper has never appeared in English translation. Only Johansen (1973, 1981, 1987), Aalen & Johansen (1978), and Hjort (1984), Hjort et al. (1985) seem to have followed up these results. A further surprise is that Doléans-Dade's (1970) exponential semimartingale, which plays such an important role in stochastic analysis, is 'just' a product-integral. The same holds for Jacod's (1975) formula for the Radon-Nikodym derivative for a marked point process experiment under two different probablity measures. Less surprisingly, so is the product-limit estimator from survival analysis of Kaplan & Meier (1958). Dollard & Friedman (1979) are quite unaware of all these developments.

As we shall see the standard results on product-integrals are really rather elementary and indeed it has not been difficult for many authors to rediscover them when needed. One can alternatively consider them as special cases of rather deeper results from the theory of integral equations or even the theory of stochastic differential equations, in its modern semimartingale form.

Numerical mathematics also contains a large literature under the name of product-integration (Young, 1954a,b; de Hoog & Weiss, 1983; Brunner & van der Houwen, 1986). However the term here means something completely different (integrating the product of two functions).

We conclude this section by giving three possible (equivalent) definitions of the product integral and its most important property of multiplicativity. In Section 2 we start with a version of the first of these definitions and derive the equivalence with the others and some basic properties of product integrals. In Section 3 further properties of continuity and differentiability are derived, and in the final section we sketch a number of statistical applications.

We here define the product-integral for finite real matrix-valued measures defined on the Borel subsets $\mathcal{B}$ of the interval $]0,\tau]$, say. Let X be such a measure. Thus each component X_{ij} of X, $1\leq i, j\leq p$, 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. Thinking of X as a function from $]0,\tau]$ to $\mathbb{R}^{p\times p}$, we define functions ΔX, X_-, X^c and X^d by $\Delta X(t)=X(\{t\})$, $X_-(t)=X(]0,t[)$, $X^d(t)=\sum_{s\leq t}\Delta X(s)$, $X^c=X-X^d$. Let $\mathbf{1}$ be the identity matrix and $\mathbf{0}$ the matrix of zeros. The product-integral of X, $Y=\prod(\mathbf{1}+\mathrm{d}X)$, will in the first place 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 later that Y can also be considered as a multiplicative interval-function, just as we can consider X as an additive interval function.

DEFINITION 1: *The product-limit.*

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

where $0=t_0<t_1< \dots < t_n=t$ 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.* Y is the unique solution of the equation

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

DEFINITION 3: *The Peano series.*

$$Y(t) = \mathbf{1} + \sum_{n=1}^{\infty} \int \dots \int_{0<t_1< \dots <t_n\leq t} X(\mathrm{d}t_1) \dots X(\mathrm{d}t_n) \ .$$

Each of these definitions has its own merits and in each case 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 show below 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 for defining the product integral for semimartingales. Finally the Peano series, while perhaps intuitively unappealing, is a useful technical tool and helps one to efficiently derive the required results. In particular *its* existence problem is very easily solved. This starting point was taken by Johansen (1986).

In the case $p=1$, when the matrix-valued measure X becomes an ordinary signed measure, a fourth definition is possible. This is more generally available in the commutative case, i.e. when the matrices $X(B)$, $B \in \mathcal{B}$, all commute:

DEFINITION 4: *Commutative case only.*

$$Y(t) = \prod_{s\leq t} (\mathbf{1} + \Delta X(s)).\exp(X^{c}(t)) \ .$$

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

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 $B \in \mathcal{B}$ in the following way. Letting χ_B denote the indicator function of the set B, denote by X_B the measure $\mathrm{d}X_B = \chi_B.\mathrm{d}X$. Let Y_B be the product-integral of X_B. Finally we define

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

(We 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 $\mathrm{d}X_H = H.\mathrm{d}X$ provided H is (sum-)integrable.) We now formulate the most important multiplicative property of the product-integral:

PROPERTY 1: *Multiplicativity*. For any $s<u<t$ we have

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

As we shall see in the next section, 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. Finally, we note that the integrals in Definitions 2 and 3 are to be interpreted as being of *Lebesgue*-Stieltjes type. However Definition 1, representing the product-integral by means of *Riemann*-Stieltjes approximating finite products, suggests that a neater theory is possible in which all of the ordinary integrals are of Riemann-Stieltjes type.

The reader familiar with semimartingale theory will be impatient to see us admit that the equivalence between these definitions is well known to hold for (matrix-valued) semimartingales where the notation $\mathcal{E}(X)$ instead of $\prod(\mathbf{1}+\mathrm{d}X)$ 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'). The results are proved using sophisticated functional analysis and appropriate topologies on the space of semimartingales. An early appearance of the stochastic product-integral is in McKean (1969; §4.7, 4.8). It is a pity these authors did not connect their work to the classical theory of product-integration, and used a notation which in some ways is misleading. We do not know if it is possible to derive these results in the semimartingale case by a similar elementary approach to the one we shall use, perhaps using the discretization techniques of Helland (1982); see also Marcus (1981) and Dellacherie & Meyer (1982; §VII.43). 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 solution of the integral equation. Further results on the quadrature or 'numerical solution' of stochastic differential equations are given by Clark (1984) and Pardoux & Talay (1985).

2. Basic theory.

2.1. Preliminary definitions.

Section 2 contains a brief but complete introduction to product integration of (real matrix-valued) additive interval functions and additive integration of multiplicative interval functions. To emphasize the different point of view from Section 1, we change notation and use α for an additive and μ for a multiplicative interval function; these play the role of X and Y in the previous section respectively via the equivalences $\alpha(s,t)=X(]s,t])$, $\mu(s,t)=Y(]s,t])$. We also work on the whole line $]0,\infty[$ instead of just the interval $]0,\tau]$. We shall treat the nonnegative scalar case ($p=1$, $\alpha\geq 0$) in the next subsection and show that the general matrix case follows directly from this using some simple but important algebraic identities for matrices which are first summarized. 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 corresponding partitions. We shall, with Dobrushin (1953), start with right continuous functions and thereby connect up with the usual measure theory and also establish the stronger result that the Riemann sums and products converge (uniformly) if just the mesh of the partition, i.e. the length of the largest division interval, goes to zero.

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

$$\Pi_{1\leq i\leq n}(\mathbf{1}+A_i) - \mathbf{1} = \Sigma_{1\leq i\leq n}(\mathbf{1}+A_1)\ldots(\mathbf{1}+A_{i-1})A_i, \tag{1}$$

$$\Pi_{1\leq i\leq n}(\mathbf{1}+A_i) - \mathbf{1} = \Sigma_{1\leq i\leq n}A_i\,(\mathbf{1}+A_{i+1})\ldots(\mathbf{1}+A_n), \tag{2}$$

$$\Pi_{1\leq i\leq n}(\mathbf{1}+A_i) - \mathbf{1} - \Sigma_{1\leq i\leq n}A_i = \Sigma_{1\leq i<j\leq n}A_i(\mathbf{1}+A_{i+1})\ldots(\mathbf{1}+A_{j-1})A_j, \tag{3}$$

$$\Pi_{1\leq i\leq n}(\mathbf{1}+A_i) - \Pi_{1\leq i\leq n}(\mathbf{1}+B_i) =$$

$$\Sigma_{1\leq i\leq n}(\mathbf{1}+A_1)\ldots(\mathbf{1}+A_{i-1})(A_i-B_i)(\mathbf{1}+B_{i+1})\ldots(\mathbf{1}+B_n). \tag{4}$$

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

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

but (2) implies that

$$\Pi_{1\leq i<j}(\mathbf{1}+A_i) = \mathbf{1} + \Sigma_{1\leq i<j}A_i(\mathbf{1}+A_{i+1})\ldots(\mathbf{1}+A_{j-1}).$$

Inserting this in (5) gives (3). □

We now define an additive interval function $\alpha(s,t)$, $0\leq s\leq t<\infty$, with values in the $p\times p$ matrices by the properties

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

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

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

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

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

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

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

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

$$\Sigma_{\mathcal{T}}\mathrm{d}(\mu-\mathbf{1}) = \Sigma_{1\le i\le n}(\mu(D_i)-\mathbf{1}) \tag{12}$$

and the Riemann product

$$\Pi_{\mathcal{T}}(\mathbf{1}+\mathrm{d}\alpha) = \Pi_{1\le i\le n}(\mathbf{1}+\alpha(D_i)). \tag{13}$$

We shall study the limits of these quantities as $|\mathcal{T}|\to 0$ 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 results hold for $\mu(s,t)$.

2.2 *The nonnegative scalar case.*

Let now $p=1$ and assume throughout that $\alpha_0(s,t)$ is an additive and nonnegative interval function, and that $\mu_0(s,t)$ is a multiplicative function which is ≥ 1. We initially define the product-integral of α_0 by

$$\Pi_{]s,t]}(1+\mathrm{d}\alpha_0) = \sup_{\mathcal{T}} \Pi_{\mathcal{T}}(1+\mathrm{d}\alpha_0) \tag{14}$$

and the additive integral of μ_0 by

$$\int_{]s,t]}\mathrm{d}(\mu_0-1) = \inf_{\mathcal{T}} \Sigma_{\mathcal{T}}\mathrm{d}(\mu_0-1). \tag{15}$$

THEOREM 1. *If α_0 is an additive nonnegative interval function then $(s,t) \to \Pi_{]s,t]}(1+\mathrm{d}\alpha_0)$ is a multiplicative function bounded below by $1+\alpha_0$ and above by $\exp(\alpha_0)$, and if μ_0 is a multiplicative function ≥ 1 then $(s,t) \to \int_{]s,t]}\mathrm{d}(\mu_0-1)$ is an additive function bounded above by μ_0-1 and below by $\log(\mu_0)$.*

PROOF. From the inequalities

$$1+a+b \le (1+a)(1+b) \le \exp(a+b),\ a \ge 0,\ b \ge 0 \tag{16}$$

it follows that $\Pi_{\mathcal{T}}(1+\mathrm{d}\alpha_0)$ is increasing over refinements of $\mathcal{T}$, and hence that

$$\Pi_{]s,t]}(1+\mathrm{d}\alpha_0) = \lim_{\mathcal{T}} \Pi_{\mathcal{T}}(1+\mathrm{d}\alpha_0), \tag{17}$$

where the limit is taken along refinements of $\mathcal{T}$. It follows that $\Pi(1+\mathrm{d}\alpha)$ 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). Similarly from

$$0 \le \log(ab) \le a\text{-}1+b\text{-}1 \le ab\text{-}1, \quad a \ge 1, b \ge 1 \tag{18}$$

it is seen that $\Sigma_{\mathcal{T}}\mathrm{d}(\mu_0\text{-}1)$ is decreasing in $\mathcal{T}$, that

$$\int_{]s,t]} \mathrm{d}(\mu_0\text{-}1) = \lim_{\mathcal{T}} \Sigma_{\mathcal{T}}\mathrm{d}(\mu_0\text{-}1), \tag{19}$$

and that $(s,t) \to \int_{]s,t]}\mathrm{d}(\mu_0\text{-}1)$ is a nonnegative additive interval function bounded above by $\mu_0(s,t)\text{-}1$ and below by $\log(\mu_0(s,t))$. □

From (1), (3), (16), and a further elementary inequality for 'log' and 'exp' one deduces immediately that the product and additive integrals satisfy the following inequalities (we already have (20) and (23) from Theorem 1):

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

$$\alpha_0(s,t) \le \Pi_{]s,t]}(1+\mathrm{d}\alpha_0)\text{-}1 \le \alpha_0(s,t)\exp(\alpha_0(s,t)), \tag{21}$$

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

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

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

THEOREM 2. *If*

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

then

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

and vice versa.

PROOF. Let α_0 be additive and nonnegative and define μ_0 by (25), then evaluate, by relation (4) of Lemma 1 and (20)

$$\begin{aligned} 0 &\le \Sigma_{\mathcal{T}}\mathrm{d}(\mu_0\text{-}1) - \alpha_0(s,t) = \Sigma_{1\le i\le n}\mu_0(D_i)\text{-}1\text{-}\alpha_0(D_i) \\ &\le \Sigma_{1\le i\le n}(1+\alpha_0(D_1))\ldots(1+\alpha_0(D_{i\text{-}1}))(\mu_0(D_i)\text{-}1\text{-}\alpha_0(D_i))\mu_0(D_{i+1})\ldots\mu_0(D_n) \\ &= \Pi_{1\le i\le n}\mu_0(D_i) - \Pi_{1\le i\le n}(1+\alpha_0(D_i)) \\ &= \mu_0(s,t) - \Pi_{\mathcal{T}}(1+\mathrm{d}\alpha_0). \end{aligned}$$

Now take limits over refinements of $\mathcal{T}$ and we get (26). The other relation is proved similarly: let μ_0 be multiplicative and ≥ 1, and define α_0 by (26). Then we evaluate by (4) and (23)

$$
\begin{aligned}
0 &\le \mu_0(s,t) - \Pi_{\mathcal{T}}(1+d\alpha_0) \\
&= \Pi_{1\le i\le n}\mu_0(D_i) - \Pi_{1\le i\le n}(1+\alpha_0(D_i)) \\
&= \Sigma_{1\le i\le n}(1+\alpha_0(D_1))\ldots(1+\alpha_0(D_{i-1}))(\mu_0(D_i)-1-\alpha_0(D_i))\mu_0(D_{i+1})\ldots\mu_0(D_n) \\
&\le \Sigma_{1\le i\le n}\,\mu_0(D_1)\ldots\mu_0(D_{i-1})(\mu_0(D_i)-1-\alpha_0(D_i))\,\mu_0(D_{i+1})\ldots\mu_0(D_n) \\
&\le \mu_0(s,t)\,(\Sigma_{\mathcal{T}}\,d(\mu_0-1) - \alpha_0(s,t)).
\end{aligned}
$$

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

It follows from the first part of the above proof that we have

COROLLARY 1. *If the relations* (25) *and* (26) *hold then*

$$0 \le \Sigma_{\mathcal{T}}d(\mu_0-1) - \alpha_0(s,t) \le \mu_0(s,t) - \Pi_{\mathcal{T}}(1+d\alpha_0).$$

We shall now show that we have an even stronger approximation result, namely

$$\Pi_{]s,t]}(1+d\alpha_0) = \lim_{|\mathcal{T}|\to 0} \Pi_{\mathcal{T}}(1+d\alpha_0)$$

and

$$\int_{]s,t]}d(\mu_0-1) = \lim_{|\mathcal{T}|\to 0} \Sigma_{\mathcal{T}}d(\mu_0-1),$$

(the limits are in fact, as we shall later see, uniform on bounded intervals). First note that an additive nonnegative interval function by the continuity assumption (8) determines a σ-additive measure which is finite on bounded intervals, with the property that $\alpha_0(]s,t]) = \alpha_0(s,t)$. In the following we shall think of α_0 as a measure.

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

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

PROOF. Let $a_1 \ge a_2 \ge \ldots$ be the sizes of the atoms of the measure θ in the interval $]s,t]$, 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 $\theta^c(]0,u])$ is uniformly continuous on the interval $[0,t]$, and we can hence choose a $\delta_1(\varepsilon)$ such that any interval of length $\le \delta_1(\varepsilon)$ has θ^c measure $\le \varepsilon/2$. If the interval is also chosen less than $\delta_2(\varepsilon)=\min |b_i-b_j|$ for i and $j \le n(\varepsilon)$, 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 $\leq \varepsilon/2$, by the choice of $n(\varepsilon)$. Thus for any partition $\mathcal{T}$ with $|\mathcal{T}|$ less than $\min(\delta_1(\varepsilon),\delta_2(\varepsilon))$ we have that $\max_{1\leq i\leq n} \theta(D_i\setminus\{s_i\}) \leq \varepsilon$ 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*

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

PROOF. Let $D_1 =]s,u[$ and $D_2=]u,t]$ then

$$0 \leq \Pi_D(1+d\theta)-1-\theta(D) = \Pi_{D_1}(1+d\theta)\,(1+\theta\{u\})\,\Pi_{D_2}(1+d\theta) -1 - \theta(D)$$

$$= \Pi_{D_1}(1+d\theta)\Pi_{D_2}(1+d\theta) - 1 - \theta(D_1\cup D_2) + \theta\{u\}(\Pi_{D_1}(1+d\theta)\Pi_{D_2}(1+d\theta) - 1)$$

$$\leq \tfrac{1}{2}\theta(D\setminus\{u\})^2 \exp(\theta(D\setminus\{u\}))+\theta\{u\}\theta(D\setminus\{u\})\exp(\theta(D\setminus\{u\}))$$

$$\leq \theta(D\setminus\{u\})\theta(D)\exp(\theta(D)).$$ □

THEOREM 3. *If* α_0 *is an additive nonnegative interval function then*

$$\Pi_{]s,t]}(1+d\alpha_0) = \lim_{|\mathcal{T}|\to 0} \Pi_{\mathcal{T}}(1+d\alpha_0). \tag{27}$$

If μ_0 *is a multiplicative interval function and* $\mu_0 \geq 1$, *then*

$$\int_{]s,t]}d(\mu_0 -1) = \lim_{|\mathcal{T}|\to 0} \Sigma_{\mathcal{T}}d(\mu_0-1). \tag{28}$$

PROOF. Let $M_i = \Pi_{D_i}(1+d\alpha_0)$ 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 $|M_i -N_i| \leq \alpha_0(D_i\setminus\{s_i\})\alpha_0(D_i)\exp(\alpha_0(D_i))$. We then get from (4) that

$$0 \leq \Pi_{]s,t]}(1+d\alpha_0) - \Pi_{\mathcal{T}}(1+d\alpha_0) \tag{29}$$

$$= \Pi_{1\leq i\leq n}M_i - \Pi_{1\leq i\leq n}N_i$$

$$= \Sigma_{1\leq i\leq n}M_1\cdots M_{i-1}(M_i-N_i)N_{i+1}\cdots N_n$$

$$\leq \max_{1\leq i\leq n} \alpha_0(D_i\setminus\{s_i\})\alpha_0(D)(\exp(\alpha_0(D)))^2,$$

where s_i is the position of the largest atom in D_i. Now as $|\mathcal{T}|\to 0$ this goes to zero by Lemma 2 which proves (27). It follows from Corollary 1 that (28) holds. □

2.3. *The general matrix case.*

For a $p{\times}p$ matrix A we define the norm $|A| = \max_i \sum_j |a_{ij}|$. We define an interval function β with values in the $p{\times}p$ matrices to have bounded variation on $]s,t]$ if

$$|\beta|(s,t) = \sup_{\mathcal{T}} \Sigma_{1\le i\le n} |\beta(D_i)| \le c < \infty.$$

We shall say that β has bounded variation if $|\beta|(s,t)$ is uniformly bounded in $s \le t$ for each finite t.

LEMMA 4. *The additive interval function α is of bounded variation if and only if it is dominated by an additive nonnegative real interval function α_0, that is $|\alpha(s,t)| \le \alpha_0(s,t)$. If μ is a multiplicative interval function then $\mu\text{-}\mathbf{1}$ is of bounded variation if and only if there is a real multiplicative interval function $\mu_0 \ge 1$ such that* $|\mu(\mathrm{s,t})-\mathbf{1}| \le \mu_0(\mathrm{s,t})-1$.

PROOF. Let α be additive and of bounded variation, then we define

$$\alpha_0(s,t) = |\alpha|(s,t) = \sup_{\mathcal{T}} \Sigma_{1\le i\le n} |\alpha(D_i)|.$$

If μ is multiplicative then the function

$$(s,t) \to |\mu\text{-}\mathbf{1}|(s,t) = \sup_{\mathcal{T}} \Sigma_{1\le i\le n} |\mu(D_i)\text{-}\mathbf{1}|$$

is not additive, but super additive, i.e.

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

Then define $\alpha_0(u,t) = |\mu\text{-}\mathbf{1}|(0,t) - |\mu\text{-}\mathbf{1}|(0,u) \ge |\mu\text{-}\mathbf{1}|(u,t)$, and $\mu_0(s,t) = \Pi_{]s,t]}(1+\mathrm{d}\alpha_0)$, then

$$|\mu(s,t)\text{-}\mathbf{1}| \le |\mu\text{-}\mathbf{1}|(s,t) \le \alpha_0(s,t) \le \mu_0(s,t)\text{-}1. \quad \square$$

We shall now show the existence of product and additive integrals by reducing the problem to the nonnegative scalar case.

THEOREM 4. *Let α be additive and dominated by α_0 then $\Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha)$ exists and*

$$|\Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha) - \mathbf{1}| \le \Pi_{]s,t]}(1+\mathrm{d}\alpha_0) - 1, \tag{30}$$

$$|\Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha) - \mathbf{1} - \alpha(s,t)| \le \Pi_{]s,t]}(1+\mathrm{d}\alpha_0) - 1 - \alpha_0(s,t). \tag{31}$$

Furthermore

$$\Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha) = \lim_{|\mathcal{T}|\to 0} \Pi_{\mathcal{T}}(\mathbf{1}+\mathrm{d}\alpha)$$

uniformly in $0\le s\le t\le u$ for any fixed u.

PROOF. From (3) of Lemma 1 we find

$$\Pi_{\mathcal{T}}(\mathbf{1}+\mathrm{d}\alpha) - \mathbf{1} - \alpha(D) = \Sigma_{1\le i<j\le n} \alpha(D_i)(\mathbf{1}+\alpha(D_{i+1}))\ldots(\mathbf{1}+\alpha(D_{j-1}))\alpha(D_j).$$

Hence

$$|\Pi_{\mathcal{T}}(\mathbf{1}+d\alpha) - \mathbf{1} - \alpha(D)| \le \Pi_{\mathcal{T}}(1+d\alpha_0) - 1 - \alpha_0(D) \tag{32}$$

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

$$\Pi_{\mathcal{S}}(\mathbf{1}+d\alpha) - \Pi_{\mathcal{T}}(\mathbf{1}+d\alpha) = \Pi_{1\le i\le n}\Pi_{\mathcal{D}_i}(\mathbf{1}+d\alpha) - \Pi_{1\le i\le n}(\mathbf{1}+\alpha(D_i)).$$

Now apply (4) of Lemma 1 and the inequality (32) above and we find

$$|\Pi_{\mathcal{S}}(\mathbf{1}+d\alpha) - \Pi_{\mathcal{T}}(\mathbf{1}+d\alpha)| \le \Pi_{\mathcal{S}}(1+d\alpha_0) - \Pi_{\mathcal{T}}(1+d\alpha_0). \tag{33}$$

But the product integral of α_0 exists and equals the limit as $|\mathcal{T}|\to 0$ of the Riemann products hence the same result holds for α. Taking the limit as $|\mathcal{T}|\to 0$ of (32) gives (31). In the same way (30) is proved using (1) of Lemma 1. Combining (29) and (33) we get

$$0 \le |\Pi_{]s,t]}(\mathbf{1}+d\alpha) - \Pi_{\mathcal{T}}(\mathbf{1}+d\alpha)| \le \Pi_{]s,t]}(1+d\alpha_0) - \Pi_{\mathcal{T}}(1+d\alpha_0)$$

$$\le \max_i \alpha_0(D_i\setminus\{s_i\})\alpha_0(D)(\exp(\alpha_0(D)))^2.$$

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

THEOREM 5. *Let* μ *be multiplicative and dominated by* μ_0, *i.e.* $|\mu-\mathbf{1}|(s,t) \le \mu_0(s,t)-1$, *then the additive integral of* $\mu-\mathbf{1}$ *exists and*

$$|\textstyle\int_{]s,t]} d(\mu-\mathbf{1})| \le \int_{]s,t]} d(\mu_0-1), \tag{34}$$

$$|\mu(s,t) - \mathbf{1} - \textstyle\int_{]s,t]} d(\mu-\mathbf{1})| \le \mu_0(s,t) - 1 - \int_{]s,t]} d(\mu_0-1). \tag{35}$$

Furthermore we have

$$\textstyle\int_{]s,t]} d(\mu-\mathbf{1}) = \lim_{|\mathcal{T}|\to 0} \Sigma_{\mathcal{T}} d(\mu-\mathbf{1})$$

uniformly in $0\le s\le t\le u$ *for any fixed u.*

PROOF. Let us evaluate

$$\mu(D) - \mathbf{1} - \Sigma_{\mathcal{T}} d(\mu-\mathbf{1}) = \Pi_{1\le i\le n}\,\mu(D_i) - \mathbf{1} - \Sigma_{1\le i\le n}(\mu(D_i)-\mathbf{1}).$$

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

$$\Sigma_{1\le i<j\le n}\,(\mu(D_i)-\mathbf{1})\mu(D_{i+1})\ldots\mu(D_{j-1})(\mu(D_j)-\mathbf{1})$$

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

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

$$|\mu(D) - \mathbf{1} - \Sigma_{\mathcal{T}} d(\mu-\mathbf{1})| \le \mu_0(D) - 1 - \Sigma_{\mathcal{T}} d(\mu_0-1). \tag{36}$$

Let now $\mathcal{S}$ be a refinement of $\mathcal{T}$ and $\mathcal{D}_i$ the corresponding partition of D_i, then

$$\Sigma_{\mathcal{T}}\mathrm{d}(\mu-\mathbf{1}) - \Sigma_{\mathcal{S}}\mathrm{d}(\mu-\mathbf{1}) = \Sigma_{1\le i\le n}(\mu(D_i)+\mathbf{1} - \Sigma_{\mathcal{D}_i}\mathrm{d}(\mu-\mathbf{1})),$$

hence

$$|\Sigma_{\mathcal{T}}\mathrm{d}(\mu-\mathbf{1}) - \Sigma_{\mathcal{S}}\mathrm{d}(\mu-\mathbf{1})| \le \Sigma_{\mathcal{T}}\mathrm{d}(\mu_0-1) - \Sigma_{\mathcal{S}}\mathrm{d}(\mu_0-1)\,.$$

Now $\int\mathrm{d}(\mu_0-1)$ exists and moreover equals the limit as $|\mathcal{T}|\to 0$ of the Riemann sums which shows that the same results hold for μ. The inequality (34) now follows trivially from the definitions. Taking the limit as $|\mathcal{T}|\to 0$ in (36) gives (35). The uniformity follows from Theorem 4 and Corollary 1. □

THEOREM 6. *If α is additive and of bounded variation and μ is defined by*

$$\mu(s,t) = \Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha) \tag{37}$$

then

$$\alpha(s,t) = \int_{]s,t]}\mathrm{d}(\mu-\mathbf{1}). \tag{38}$$

Similarly if μ is multiplicative and $\mu-\mathbf{1}$ of bounded variation and α is defined by (38) *then* (37) *holds.*

PROOF. Assume α to be additive and of bounded variation and define μ by (37). Let α_0 dominate α and define μ_0 to be the product integral of α_0, then by (30) and Theorem 2 $|\mu-\mathbf{1}| \le \mu_0-1$, which shows that $\mu-\mathbf{1}$ is of bounded variation, and that $\int\mathrm{d}(\mu-\mathbf{1})$ exists. Now evaluate

$$\Sigma_{\mathcal{T}}\mathrm{d}(\mu-\mathbf{1}) - \alpha(s,t) = \Sigma_{1\le i\le n}\, \mu(D_i)-\mathbf{1}-\alpha(D_i)$$

which by (31) is dominated by

$$\Sigma_{1\le i\le n}\, \mu_0(D_i)-1-\alpha_0(D_i) \le \Sigma_{\mathcal{T}}\, \mathrm{d}(\mu_0-1) - \alpha_0(\mathrm{D}),$$

which goes to zero by Theorem 2. Similarly if μ is multiplicative and $\mu-\mathbf{1}$ dominated by μ_0-1, then by (34) and Theorem 2, $\alpha = \int\mathrm{d}(\mu-\mathbf{1})$ is dominated by $\alpha_0 = \int\mathrm{d}(\mu_0-1)$, and hence

$$\begin{aligned}\mu(D) - \Pi_{\mathcal{T}}(\mathbf{1}+\mathrm{d}\alpha) &= \Pi_{1\le i\le n}\mu(D_i) - \Pi_{1\le i\le n}(\mathbf{1}+\alpha(D_i))\\ &= \Sigma_{1\le i\le n}\, \mu(D_1)\ldots\mu(D_{i-1}).\\ &\qquad .(\mu(D_i)-\mathbf{1}-\alpha(D_i))\,(\mathbf{1}+\alpha(D_{i+1}))\ldots(\mathbf{1}+\alpha(D_n))\end{aligned}$$

which by Theorem 5 is dominated by

$$\begin{aligned}&\Sigma_{1\le i\le n}\, \mu_0(D_1)\ldots\mu_0(D_{i-1}).\\ &\qquad .(\mu_0(D_i)-1-\alpha_0(D_i))\,(1+\alpha_0(D_{i+1}))\ldots(1+\alpha_0(D_n))\\ &= \mu_0(D) - \Pi_{\mathcal{T}}(1+\mathrm{d}\alpha_0)\end{aligned}$$

which tends to zero by Theorem 2. □

In the scalar (commutative) case, if α_0 is a finite signed measure on an arbitrary measurable space one can define its product integral e.g. by (17) or even if the space is a separable metric space (i.e. a Polish space) by (27); the interval $]s,t]$ is replaced by an arbitrary measurable set and the notions of a partition $\mathcal{T}$ and (in the Polish case, when product-integrating over a totally bounded set) its mesh $|\mathcal{T}|$ are generalized appropriately. All the above results still hold, because we do not need the ordering of the real numbers when the terms in the approximating products commute. In particular, we leave it to the reader to derive the commutative case result

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

where $\alpha_0{}^c$ is the continuous part of α_0.

2.4. *The Peano series and the Volterra integral equation.*

Let α be an additive interval function on $]0,\infty]$ with dominating measure α_0. Then each of the entries $\alpha_{i\,j}$ is a finite measure on (the Borel subsets of) bounded subsets of $\mathbb{R}^+$, and we can define product measure on bounded subsets of $(\mathbb{R}^+)^n$ starting from

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

Note that

$$|\alpha^{(n)}(D_1\times\ldots\times D_n)| \ \le\ \Pi_{1\le i\le n}\,|\alpha(D_i)|$$

$$\le\ \Pi_{1\le i\le n}\,\alpha_0(D_i)\ = \alpha_0{}^{(n)}(D_1\times\ldots\times D_n).$$

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

$$A(n,D) = \{(u_1,\ldots,u_n) \in (D)^n : u_1<\ldots<u_n\ \}$$

then the Peano series is defined by

$$P(\alpha,D) = \mathbf{1} + \Sigma_{1\le n<\infty}\ \alpha^{(n)}(A(n,D)). \tag{39}$$

Note that the series is dominated by

$$P(\alpha_0,D)\ =\ 1 + \Sigma_{1\le n<\infty}\ \alpha_0{}^{(\mathrm{n})}(A(n,D))$$

$$\le\ 1 + \Sigma_{1\le \mathrm{n}<\infty}\ \alpha_0(D)^n/n!\ =\ \exp(\alpha_0(D))$$

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

$$|P(\alpha,D)|\ \le\ P(\alpha_0,D)\ \le\ \exp(\alpha_0(D)) \tag{40}$$

$$|P(\alpha,D) - \mathbf{1}|\ \le\ P(\alpha_0,D) - 1$$

$$\le\ \alpha_0(D)\exp(\alpha_0(D)) \tag{41}$$

$$|P(\alpha,D) - \mathbf{1} - \alpha(D)|\ \le\ P(\alpha_0,D) - 1 - \alpha_0(D)$$

$$\le\ \tfrac{1}{2}\alpha_0(D)^2\ \exp(\alpha_0(D)) \tag{42}$$

We write $A(n,s,t)$ and $P(\alpha,s,t)$ for $A(n,]s,t])$ and $P(\alpha,]s,t])$ respectively.

THEOREM 7. *The interval function $P(\alpha,s,t)$ is multiplicative.*

PROOF. Define the set in $\mathbb{R}^n$

$$A(i,n,s,u,t) = \{s<u_1 <\ldots<u_i \leq u<u_{i+1}<\ldots<u_n \leq t\},$$

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

$$A(n,s,t) = \{s<u_1 <\ldots<u_n \leq t\}$$

$$= \cup_{0\leq i\leq n} A(i,n,s,u,t).$$

Hence

$$\alpha^{(n)}(A(n,s,t)) = \Sigma_{0\leq i\leq n}\, \alpha^{(n)}(A(i,n,s,u,t))$$

$$= \Sigma_{0\leq i\leq n}\, \alpha^{(i)}(A(i,s,u))\, \alpha^{(n-i)}(A(n-i,u,t)).$$

Summing over n gives the desired result. The exponential inequality (41) shows that $P(\alpha,s,t)$ is right continuous in s and t. □

THEOREM 8. *The Peano series is equal to the product integral*

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

PROOF. First note that

$$P(\alpha,s,t) - \Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha) = \Pi_{1\leq i\leq n}P(\alpha,D_i) - \Pi_{1\leq i\leq n}(\mathbf{1}+\alpha(D_i))$$

$$= \Sigma_{1\leq i\leq n}(\mathbf{1}+\alpha(D_1))\ldots(\mathbf{1}+\alpha(D_{i-1})).$$

$$.(\mathbf{1}-P(\alpha,D_i)+\alpha(D_i))P(\alpha,D_{i+1})\ldots P(\alpha,D_n)$$

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

$$|P(\alpha,s,t) - \Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha)|$$

$$\leq \max_{1\leq i\leq n} \alpha_0(D_i \setminus \{s_i\})\alpha_0(s,t)(\exp(\alpha_0(s,t)))^2$$

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 $|\mathcal{T}|$ and hence the result is proved. □

THEOREM 9. *(The forward equation)*

$$\Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha) - \mathbf{1} = \int_{]s,t]} \Pi_{]s,u[}(\mathbf{1}+\mathrm{d}\alpha)\ \alpha(\mathrm{d}u) \tag{43}$$

where the integral is a Lebesgue-Stieltjes integral.

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

$$\alpha^{(n+1)}(A(n+1,s,t))$$

$$= \int_{]s,t]} \alpha^{(n+1)}(A(n+1,s,t) \mid u_{n+1} = u)\alpha(\mathrm{d}u)$$

$$= \int_{]s,t]} \alpha^{(n)}(A(n,s,u-))\alpha(\mathrm{d}u).$$

Summing over n gives the desired result. □

Similarly we can prove

THEOREM 10. *(The backward equation)*

$$\Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha) - \mathbf{1} = \int_{]s,t]}\alpha(\mathrm{d}u)\ \Pi_{]u,t]}(\mathbf{1}+\mathrm{d}\alpha). \tag{44}$$

THEOREM 11. *Let $\beta(s,t)$ be any interval function which is right continuous with left limits in both variables and of bounded variation, and which satisfies*

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

or

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

Then $\beta(s,t) = \Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha)$.

PROOF. Let $S_n(s,t) = \mathbf{1} + \Sigma_{1\le k\le n}\alpha^{(k)}(A(k,s,t))$ then

$$S_n(s,t) = \mathbf{1} + \int_{]s,t]} S_{n-1}(s,u-)\alpha(\mathrm{d}u)$$

and hence

$$\beta(s,t) - S_n(s,t) = \int_{]s,t]} (\beta(s,u-)-S_{n-1}(s,u-))\alpha(\mathrm{d}u).$$

It is not difficult to show by induction over n that

$$|\beta(s,t)-S_n(s,t)| \le \sup_{u\le t} |\beta|(u,t)\alpha_0^{(n)}(s,t).$$

For $n\to\infty$ we get that

$$\beta(s,t) = \lim_{n\to\infty} S_n(s,t) = \Pi_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha). \quad \square$$

Note that the forward equation generalises the fundamental relation (1) whereas the backward equation is a generalisation of (2). In a similar way we can generalise (4):

THEOREM 12. (*Duhamel's equation*). *Let α_1 and α_2 be additive then*

$$\prod_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha_1) - \prod_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha_2)$$

$$= \int_{]s,t]} \prod_{]s,u[}(\mathbf{1}+\mathrm{d}\alpha_1)\ (\alpha_1-\alpha_2)(\mathrm{d}u)\ \prod_{]u,t]}(\mathbf{1}+\mathrm{d}\alpha_2). \tag{45}$$

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

$$\alpha_{1,2}{}^{(n,m)}\ (A_1\times\ldots\times A_n\times B_1\times\ldots\times B_m)$$

$$= \alpha_1(A_1)\ldots\alpha_1(A_n)\alpha_2(B_1\ )\ldots\alpha_2(B_m).$$

By applying Fubini's theorem we obtain

$$\alpha_{1,2}{}^{(n,m)}(A(n+m,s,t))$$

$$= \int_{]s,t]} \alpha_1{}^{(n-1)}(A(n-1,s,u-)\ \alpha_1(\mathrm{d}u)\ \alpha_2{}^{(\mathrm{m})}(A(m,u,\ t))$$

$$= \int_{]s,t]}\alpha_1{}^{(n)}(A(n,s,u-)\ \alpha_2(\mathrm{d}u)\ \alpha_2{}^{(m-1)}(A(m-1,u,t)).$$

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

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

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

which is the desired relation. □

Theorem 8 can alternatively be derived from Theorem 6 by showing that the additive integral of P-$\mathbf{1}$ equals α. Also Theorems 9, 10 and 12 can be derived directly from Lemma 1 and the limit results of Theorems 4 and 5 by dominated convergence arguments.

3. Further properties of the product-integral.

3.1. Continuity and differentiability.

From Duhamel's equation (45) 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. Less clear is 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 Gateaux ('directional') and Frechet ('bounded') differentiability and exactly atuned to the functional version of the delta-method, a basic and elementary tool of large-sample statistical theory (see Reeds (1976) and Gill (1987)). This will be illustrated in 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 subsection 2.3),

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

while the supremum norm is given by

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

We also recall the integration by parts formula for cadlag (right continuous with left hand limits) bounded variation functions U and V,

$$\mathrm{d}(UV) = U_{-}(\mathrm{d}V) + (\mathrm{d}U)V.$$

Let $\mu_i = \prod(\mathbf{1}+\mathrm{d}\alpha_i)$, i=1,2. Duhamel's equation then becomes

$$(\mu_1-\mu_2)(s,t) = \int_{]s,t]} \mu_1(s,u-)\,(\alpha_1-\alpha_2)(\mathrm{d}u)\,\mu_2(u,t)$$

$$= \int_{]s,t]} \mu_1(s,u-)\,\nu(\mathrm{d}u) \tag{46}$$

where (keeping s and t fixed for the moment)

$$\nu(x) = \int_{]s,x]} (\alpha_1-\alpha_2)(\mathrm{d}u)\,\mu_2(u,t) \tag{47}$$

Now integration by parts and the backward equation applied to (47) gives

$$\| \nu \|_\infty \le \| \alpha_1 - \alpha_2 \|_\infty . (2+ \| \alpha_2 \|_V) . \| \mu_2 \|_\infty,$$

uniformly in s and t. Again integrating by parts but now applying the forward equation to (46) gives

$$\| \mu_1 - \mu_2 \|_\infty \le \| \mu_1 \|_\infty . (2+ \| \alpha_1 \|_V) . \| \nu \|_\infty$$

$$\le \| \alpha_1 - \alpha_2 \|_\infty . \| \mu_1 \|_\infty . (2+ \| \alpha_1 \|_V) . \| \mu_2 \|_\infty . (2+ \| \alpha_2 \|_V) .$$

(48)

Noting that $\| \mu_i \|_\infty \le \exp(\| \alpha_i \|_\infty)$, we get the required continuity result:

THEOREM 13. (*Continuity of the product integral in supremum norm*). *Let* $\alpha^{(n)}$, n=1,2,... *be a sequence of additive interval functions on*]0,τ] *such that*

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

$$\limsup \| \alpha^{(n)} \|_V < \infty,$$

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

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

We can further refine this result to a differentiability result as follows. Let $\alpha^{(n)}$, α, $\mu^{(n)}$ and μ be as in Theorem 13 and suppose that we actually have $\alpha^{(n)} = \alpha + t_n h_n$ 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

$$t_n^{-1}(\,\alpha^{(n)} - \alpha\,) = h_n \to h \quad \text{in supremum norm.}$$

The mapping $\mathfrak{P} : \alpha \to \mu = \prod(\mathbf{1}+\mathrm{d}\alpha)$ will be Hadamard differentiable if we correspondingly have

$$t_n^{-1}(\,\mathfrak{P}(\alpha^{(n)}) - \mathfrak{P}(\alpha)\,) \to \mathrm{d}\mathfrak{P}(\alpha).h \quad \text{in supremum norm,} \tag{49}$$

where $\mathrm{d}\mathfrak{P}(\alpha)$, the derivative of $\mathfrak{P}$ at the point α, is a 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. (Note that we have restricted the domain of $\mathfrak{P}$ to a subset of uniformly-bounded-variation additive interval functions, see Gill (1987) Lemma 1 for the propriety of such a restriction). Now by evaluating the left hand side of (49) with Duhamel's equation, we obtain

$$\{t_n^{-1}(\,\mathfrak{P}(\alpha^{(n)}) - \mathfrak{P}(\alpha)\,)\}(s,t) \;=\; \int_{]s,t]} \mu^{(n)}(s,u-)h_n(\mathrm{d}u)\,\mu(u,t)$$

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

$$\{t_n^{-1}(\,\mathfrak{P}(\alpha^{(n)}) - \mathfrak{P}(\alpha)\,)\}(s,t) \;-\; (\mathrm{d}\mathfrak{P}(\alpha).h\,)(s,t)$$
$$= \int_{]s,t]} \mu^{(n)}(s,u-)(h_n-h)(\mathrm{d}u)\,\mu(u,t)$$

(cf. (46)), (48) gives

$$\|t_n^{-1}(\,\mathfrak{P}(\alpha^{(n)}) - \mathfrak{P}(\alpha)\,) \;-\; \mathrm{d}\mathfrak{P}(\alpha).h\;\|_\infty$$
$$\le \|\,h_n-h\,\|_\infty \,.\, \|\,\mu^{(n)}\,\|_\infty \,.\, (\,2+\|\,\alpha^{(n)}\,\|_{\mathrm{V}}\,) \,.\, \|\,\mu\,\|_\infty \,.\, (\,2+\|\,\alpha\,\|_{\mathrm{V}}\,)\,.$$

We can summarize this result in the following theorem:

THEOREM 14. (*Compact differentiability of the product integral with respect to the supremum norm*). *Consider the product integral as a mapping* $\mathfrak{P}$ *from the set of additive interval functions on* $]0,\tau]$ *with variation bounded by the constant c to the space of interval functions on* $]0,\tau]$ (*the domain considered as a subset of the range*), *both endowed with the supremum norm. Let* α *be given and define* $\mu = \mathfrak{P}(\alpha) = \prod(\mathbf{1}+\mathrm{d}\alpha)$. *Then* $\mathfrak{P}$ *is compactly differentiable at* α *with derivative* $\mathrm{d}\mathfrak{P}(\alpha)$ *given by*

$$(\mathrm{d}\mathfrak{P}(\alpha).h\,)(s,t) = \int_{]s,t]} \mu(s,u-)\;h(\mathrm{d}u)\;\;\mu(u,t)$$

where the integral with respect to h is defined by formal application of the integration by parts formula.

This theorem also applies directly to the mapping $X \to Y = \prod(\mathbf{1}+\mathrm{d}X)$ 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 $X(0)=\mathbf{0}$)

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

and similarly for μ and Y.

4.2. Gronwall inequalities, inhomogeneous equations, anticipating integrands.

In this subsection we briefly summarize some further useful results on product integrals, starting with a version of Gronwalls's (1919) inequality. (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 $\phi_0(t) = \prod_{]0,t]}(1+\mathrm{d}\alpha_0)$ satisfies the Volterra integral equation

$$\phi_0(t) = 1 + \int_{]0,t]}\phi_0(s-)\alpha_0(\mathrm{d}s).$$

The basic Gronwall inequality is now the following:

THEOREM 15. *Suppose* α_0 *is a nonnegative scalar additive interval function on* $]0,\tau]$ *with* $\phi_0(t) = \prod_{]0,t]}(1+\mathrm{d}\alpha_0)$; *suppose* ϕ *is a cadlag nonnegative real function such that*

$$\phi(t) \leq 1 + \int_{]0,t]}\phi(s-)\alpha_0(\mathrm{d}s) \quad \textit{for all } t\leq\tau. \tag{50}$$

Then $\phi(t) \leq \phi_0(t)$ *for all* $t\leq\tau$.

PROOF. On repeatedly substituting the inequality for ϕ in the right hand side of (50) we see the Peano series for μ_0 appearing with a remainder which converges to zero by boundedness of ϕ. □

There are also inequality versions of inhomogenous Volterra integral equations. The proof of the following theorem is left to the reader:

THEOREM 16. (*The inhomogenous equation*). *Let* α *be an additive integral function of bounded variation and* ψ *a cadlag matrix-valued function on* $]0,\infty[$. *Then* ϕ *satisfies*

$$\phi(t) = \psi(t) + \int_{]0,t]}\phi(s-)\alpha(\mathrm{d}s) \quad \textit{for all } t \tag{51}$$

iff

$$\phi(t) = \psi(t) + \int_{]0,t]}\psi(s-)\alpha(\mathrm{d}s)\prod_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha) \quad \textit{for all } t. \tag{52}$$

The related Gronwall inequality is of course that in the nonnegative scalar case, (51) with '=' replaced by '≤' implies the same for (52).

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

$$\phi(t) \;=\; \mathbf{1} + \int_{]0,t]}\phi(s)\alpha(\mathrm{d}s) \qquad \text{for all } t,$$

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

$$\phi(t) \;=\; \prod_{]0,t]}(\mathbf{1}\text{-d}\alpha)^{-1} = \big(\prod_{[t,0[}(\mathbf{1}\text{-d}\alpha)\big)^{-1}$$

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. Some stochastic Gronwall inequalities are derived by Valkeila (1982) and Nikunen & Valkeila (1984).

4. Applications.

4.1. The application of product integrals to survival and hazard functions.

Let T be a positive random variable, the life length of a lightbulb, say. We define the survival function $S(t) = 1\text{-}F(t) = \mathrm{P}\{T>t\}$. If S is positive and differentiable then we can define the hazard rate

$$\lambda(t) = -\,\mathrm{d}\log(S(t))/\mathrm{d}t \;=\lim_{h\downarrow 0}h^{-1}\mathrm{P}\{\, T<t+h \;\mid T\geq t\;\} \tag{53}$$

and we have the well known relation

$$S(t) = \exp(-\int_{]0,t]}\lambda(u)\mathrm{d}u\,). \tag{54}$$

The measure Λ defined by $\Lambda(]0,t]) = \int_{]0,t]}\lambda(u)\mathrm{d}u$ is called the intensity or hazard measure. We shall generalise the relations (53) and (54) 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 S and Λ is that between a measure and its product integral. We let $\tau\leq\infty$ 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 17. *Let Λ be a nonnegative measure on $]0,\tau]$ which is finite on $]0,\sigma]$ and such that $\Lambda(\{\sigma\})< 1$ for all $\sigma<\tau$, and satisfies either*

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

or

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

Then defining

$$S(t) =\prod_{]0,t]}(1\text{-d}\Lambda), \tag{55}$$

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 satisfying either

$$S(\tau\text{-}) > 0,\;\; S(\tau) = 0 \tag{a'}$$

or

$$S(t) > 0 \textit{ for all } t<\tau, \;\; S(\tau-) = 0 \tag{b'}$$

then Λ *defined by*

$$\Lambda(]0,t]) = -\int_{]0,t]} S(\mathrm{d}u)/S(u-) \tag{56}$$

has the properties just described.

The probabilistic intepretation of the atoms of Λ is the following: $\Lambda(\{t\}) = \mathrm{P}\{T{=}t \mid T{\geq}t\}$. Strictly speaking in the case (b) we first need to define the product integral of an unbounded measure, e.g. by the usual definition of an indefinite Riemann integral. Before we prove the theorem we need a technical lemma,

LEMMA 5. *Let* Λ *satisfy the assumptions in* Theorem 17, *then* $S(t) > 0$ *for all* $t<\tau$.

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

$$1\text{-}\Lambda(]t_i,t_{i+1}]) = 1 - \Lambda(]t_i,t_{i+1}[) - \Lambda(\{t_{i+1}\}) \;\geq 1 - \varepsilon - (1\text{-}2\varepsilon\,) = \varepsilon$$

and hence by the inequality $1\text{-}x \;\geq\; \exp(-c(\varepsilon)x)$ for all $0{\leq}x{\leq}\,\varepsilon$ where $c(\varepsilon) = -\log(1\text{-}\varepsilon)/\varepsilon < \infty$

$$\log(1\text{-}\Lambda(]t_i,t_{i+1}])) \;\geq\; -c(\varepsilon)\,\Lambda(]t_i,t_{i+1}]).$$

Now summing over i and passing to the limit gives $S(t) = \prod_{]0,t]}(1\text{-}\mathrm{d}\Lambda) \geq C > 0$. □

PROOF OF THEOREM 17. Here we show the equivalence of (55) and (56); the rest of the proof is left to the reader. Suppose first S is given by (55). Then by the forward equation (43) (Theorem 9) we have $S(t) = 1 - \int_{]0,t]} S(u-)\Lambda(\mathrm{d}u)$. Since by Lemma 5 $S(u-) > 0$ we get (56). If on the other hand Λ satisfies (56) then S solves the forward equation and hence by the uniqueness of its solution (Theorem 11) it is given by the product integral (55). □

Because we are in the commutative case we can rewrite (55) as

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

in which form it is quite well known; see for instance Cox (1972) p. 172, where the term product integral is used (following Arley & Borschenius (1945), as D. R. Cox (personal communication) has informed us). Wellner (1985) points out the connection with Doléans-Dade's (1970) exponential semimartingale.

4.2. The product-limit estimator.

Just as a survival function is the product integral of its intensity measure, in a statistical situation with censored observations from a life distribution we have that the natural estimator of the survival function, the product-limit or Kaplan-Meier (1958) estimator, is the product integral of the empirical cumulative hazard function, the so called Nelson-Aalen estimator (Nelson, 1969, Aalen, 1975). 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 Duhamel's equation (45) 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 work with the classical random censorship model. So let T_1, ..., T_n be i.i.d. positive lifetimes from the distribution F with survival function S and let, independent thereof, $C_1, ..., C_n$ 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 $D_i=1\{T_i \leq C_i\}$, $i=1,...n$, be the actually observed data. Define

$$N_n(t) = n^{-1} \#\{ i : \tilde{T}_i \leq t, D_i = 1 \},$$

$$Y_n(t) = n^{-1} \#\{ i : \tilde{T}_i \geq t \},$$

$$\Lambda_n(t) = \int_{]0,t]} Y_n^{-1} \mathrm{d}N_n,$$

$$S_n(t) = \prod_{]0,t]} (1-\mathrm{d}\Lambda_n).$$

So S_n is the Kaplan-Meier estimator of S and Λ_n is the Nelson-Aalen estimator of Λ. One finds easily

$$\mathrm{E}N_n(t) = \int_{]0,t]} H(s-)F(\mathrm{d}s),$$

$$\mathrm{E}Y_n(t) = H(t-)S(t-),$$

$$\Lambda(t) = \int_{]0,t]} (\mathrm{E}Y_n)^{-1} \mathrm{d}(\mathrm{E}N_n) \quad \text{provided } \mathrm{E}Y_n(\mathrm{t}) > 0,$$

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

Duhamel's equation applied to S_n-S gives:

$$S_n(t) - S(t) = \int_{]0,t]} S_n(s-) (\Lambda_n - \Lambda)(\mathrm{d}s) (S(t)/S(s))$$

or

$$S_n(t)/S(t) - 1 = \int_{]0,t]} (S_n(s-)/S(s)) (\Lambda_n - \Lambda)(\mathrm{d}s). \tag{57}$$

This key equality was first established in the more general context of inhomogeneous Markov processes by Aalen & Johansen (1978) and later exploited by Gill (1980a, 1983a) to derive small-sample results (unbiasedness, variance) and large-sample results (consistency, weak convergence) for the Kaplan-Meier estimator using martingale methods: one has namely in (57) (if one replaces t by $t \wedge \max_i \tilde{T}_i$) that the integrating function $\Lambda_n - \Lambda$, 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 far wider class of censoring models than just the random censoring described above; for instance, under models for random truncation (Keiding & Gill, 1987).

Here we take an alternative approach restricted to i.i.d. models and use Theorems 13 and 14 (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 Duhamel's equation so this approach is also based on (57) 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 $\mathrm{E}Y(\tau) = S(\tau-)H(\tau-) > 0$. Note that N_n and Y_n and their expectations $\mathrm{E}N$ and $\mathrm{E}Y$ (since these do not depend on n we have dropped the subscript) are bounded monotone functions in $\mathrm{D}[0,\tau]$ or $\mathrm{D}_-[0,\tau]$ and hence even have uniformly bounded variation. The variation of Λ_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 norms. We can now consider S_n as the result of applying three mappings one after the other:

$$(N_n,Y_n) \to (N_n,1/Y_n) \to \Lambda_n \to S_n$$

i.e. going through the spaces

$$\mathrm{D}[0,\tau]\times\mathrm{D}_-[0,\tau] \to \mathrm{D}[0,\tau]\times\mathrm{D}_-[0,\tau] \to \mathrm{D}[0,\tau] \to \mathrm{D}[0,\tau],$$

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 $\mathrm{D}[0,\tau]$ of uniformly bounded variation. The same is true for the central mapping (integration) by Gill (1987, Lemma 3); this is essentially the Helly-Bray lemma. The first mapping is trivially continuous and differentiable when we restrict to elements of $\mathrm{D}_-[0,\tau]$ uniformly bounded away from zero. Now by the Glivenko-Cantelli theorem we have $\max(\|N_n - \mathrm{E}N\|,\|Y_n - \mathrm{E}Y\|) \to 0$ a.s. as $n\to\infty$, where $\|\cdot\|$ denotes the supremum norm on $\mathrm{D}[0,\tau]$ or $\mathrm{D}_-[0,\tau]$. This gives us consistency of the Kaplan-Meier estimator:

$$\|S_n - S\| \to 0 \text{ a.s. as } n\to\infty.$$

Weak convergence of $n^{\frac{1}{2}}(S_n - S)$ follows directly from weak convergence of $n^{\frac{1}{2}}(N_n - \mathrm{E}N, Y_n - \mathrm{E}Y)$ and the compact differentiability of the three mappings and hence of their composition. Here we use the weak convergence theory of Pollard (1984) which allows us to work in $(\mathrm{D}[0,\tau]\times\mathrm{D}_-[0,\tau], \|\cdot\|)$ without introducing the Skorohod topology by uncoupling the sigma-algebra needed to define random elements of this space (which becomes that generated by the coordinate projections or equivalently by the open balls with respect to the given sup-norm) from the Borel sigma-algebra (generated by the open sets), which is too large. Whenever the limit process is continuous such a weak convergence result 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 the theory, the use of compact differentiability, is simply the functional version of the delta-method or in other words a first order von Mises expansion. This approach is described in Gill (1987) following Reeds (1976).

Again we fix τ such that $\mathrm{E}Y(\tau)<\infty$. It is clear that

$$n^{\frac{1}{2}}(N_n - \mathrm{E}N, Y_n - \mathrm{E}Y) \to^{\mathcal{D}} (Z_N, Z_Y) \quad \text{in } (\mathrm{D}[0,\tau]\times\mathrm{D}_{-}[0,\tau], \|\cdot\|) \text{ as } n\to\infty,$$

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 mappings

$$\phi : y \to u = 1/y\,,$$

$$\psi : (x,u) \to v = \int_{[0,\cdot]} u\mathrm{d}x,$$

$$\xi : v \to z = \prod_{[0,\cdot]}(1-\mathrm{d}v),$$

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

$$\mathrm{d}\phi\,(y).k \;=\; -k/y^{2} = j,$$

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

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

By composition of these mappings, substitution of x=EN, y=EY, u=1/EY, $v=\Lambda$, $z=S$ and (h,k) = (Z_N,Z_Y) we obtain the required weak convergence result:

$$n^{\frac{1}{2}}(S_n - S) \;\to^{\mathcal{D}}\; -S\int_{[0,\cdot]}(1-\Delta\Lambda)^{-1}(-(Z_Y/(\mathrm{E}Y)^2)\mathrm{dE}N + (1/\mathrm{E}Y)\mathrm{d}Z_N)$$

$$= \; -S\int_{[0,\cdot]}((1-\Delta\Lambda)\mathrm{E}Y)^{-1}\,(\mathrm{d}Z_N - Z_Y\mathrm{d}\Lambda\,) \qquad \text{in } (\mathrm{D}[0,\tau], \|\cdot\|) \text{ as } n\to\infty.$$

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

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

Thus

$$n^{\frac{1}{2}}(S_n - S) \;\to^{\mathcal{D}}\; -S.W(\int_{[0,\cdot]}(\mathrm{E}Y\,(1-\Delta\Lambda))^{-1}\mathrm{d}\Lambda\,)$$

where W is a standard Wiener process.

This result also follows perhaps more easily from the martingale approach mentioned above. Its main significance is that this line of proof is really the same as the first proof of weak convergence of the Kaplan-Meier estimator ever, given by Breslow & Crowley (1974). That proof seemed complicated and ad hoc but one can now recognise in it a standard delta-method argument. We note that continuity of survival distribution or of censoring distribution has not been required.

The proof via compact differentiability also shows that one has a weak Bahadur representation for $n^{\frac{1}{2}}(S_n - S)$ and that the bootstrap works for the Kaplan-Meier estimator; see Gill (1987). Efficiency properties of the Kaplan-Meier estimator are shown to follow from compact differentiability by van der Vaart (1987); see also Keiding & Gill (1987). The proof of weak convergence also goes through in i.i.d. situations when the above mentioned martingale property is not available, e.g. when the observations come from censored observation of a renewal process, see Gill (1980b, 1981). Finally, 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.

4.3. Markov processes.

Consider a time-inhomogeneous Markov process X_t, $t \in [0,\infty[$ on a finite state space E, with transition probabilities

$$p_{ij}(s,t) = \mathrm{P}\{X_t = j \mid X_s = i\}.$$

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

$$P(s,t) = P(s,u)P(u,t), \quad 0 \le s \le u \le t < \infty, \tag{58}$$

$$P(s,s) = \mathbf{1}, \quad 0 \le s < \infty. \tag{59}$$

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

$$Q(t) = \partial P(s,t)/\partial t = -\partial P(s,t)/\partial s \tag{60}$$

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

$$\partial P(s,t)/\partial t = P(s,t)Q(t) \tag{61}$$

$$\partial P(s,t)/\partial s = -Q(s)P(s,t) \tag{62}$$

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 (with the uniqueness of their solution) are special cases of Theorems 9 to 11. The solution is given by the product integral of the matrix valued additive interval function, or measure,

$$\alpha(s,t) = \int_{]s,t]} Q(u)\mathrm{d}u. \tag{63}$$

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 $P(s,t)$-$\mathbf{1}$. Notice how the differentiation (61) followed by the integration (63) can be replaced by the process of integrating an interval function.

Thus the well known relations (58) to (63) are special cases of the general concepts of product integrals and additive integrals. This is especially important for statistical applications because we then often meet product-limit estimators as solutions of equations like (61) or (62), 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 α in the same way whether one is working with absolutely continuous or discrete α, so that the statistical calculations become the same as the probabilistic calculations. Also one can use product-integral theory to get properties of the estimators. In particular one can obtain a stochastic differential equation (Duhamel's equation, Theorem 12) for the estimators that allows the theory of martingales to be applied and hence one can quite easily for instance find the asymptotic distribution of the estimators. Such a programme is carried out by Aalen & Johansen (1978). Here we just discuss the probabilistic part of the problem, i.e. the existence of an intensity measure for an arbitrary given Markov process and vica-versa.

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

$$\alpha_{ii}(s,t) \leq 0, \ \ \alpha_{ij}(s,t) \geq 0, \ \ i{\neq}j, \tag{64}$$

$$\Sigma_{j\in E}\, \alpha_{ij}(s,t) = 0. \tag{65}$$

It is also necessary to assume that

$$\alpha_{ii}(\{t\}) \geq -1 \quad \text{for all } t. \tag{66}$$

Note that α is dominated by the real measure

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

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

$$P(\mathrm{s,t}) = \Pi_{]s,t]}\, (\mathbf{1}+\mathrm{d}\alpha) \tag{67}$$

and we find the following result:

THEOREM 17. *The function $P(\cdot,\cdot)$ defined by* (67) *satisfies*

$$P(s,t) \text{ is a stochastic matrix} \tag{68}$$

$$P(s,t) = P(s,u)\, P(u,t), \quad 0 \leq s \leq u \leq t < \infty, \tag{69}$$

$$P(s,s) = \mathbf{1}, \qquad 0 \leq s < \infty, \tag{70}$$

$$P(s,t) \to \mathbf{1} \text{ as } t{\downarrow}s. \tag{71}$$

PROOF. The assumption (63) about α implies that $P(s,t)$ is stochastic and (69), (70) and (71) 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 18. *The function $P(\cdot,\cdot)$ defined by (67) satisfies the Kolmogorov equations*

$$\partial P(s,t)/\partial\alpha_0(t) = P(s,t\text{-})(\mathrm{d}\alpha/\mathrm{d}\alpha_0)(t) \quad a.s.[\alpha_0] \text{ in } t\in\,]s,\infty[, \tag{72}$$

$$\partial P(s,t)/\partial\alpha_0(s) = (\mathrm{d}\alpha/\mathrm{d}\alpha_0)(s)P(s,t) \quad a.s.[\alpha_0] \text{ in } s\in[0,t[. \tag{73}$$

PROOF. From Theorem 9 we have

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

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 Theorem 10. □

We finally note

THEOREM 19. *The interval function $(s,t) \to P(s,t)\text{-}\mathbf{1}$ is of bounded variation.*
PROOF. See for instance Theorem 6. □

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

THEOREM 20. *If the transition probabilities $P(s,t)$ are right continuous and of bounded variation then the corresponding intensity measure is given by*

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

PROOF. That $P\text{-}\mathbf{1}$ is integrable follows from Theorem 5 and that the measure $\int\mathrm{d}(P\text{-}\mathbf{1})$ is the intensity measure for P follows from Theorem 6. □

Note that the Peano series representation (Theorem 8) 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 Peano series is just $\mathbf{1}$ whereas the minimal solution starts with the matrix ${}_0P$ with elements

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

The terms of the Peano series need not even be positive whereas the n'th term of the series for the minimal solution has an interpretation as the probability of jumping from i to j in exactly n steps.

We now want to discuss the construction of the underlying process $X(t)$. We want to show that under the assumptions of Theorem 20 there exists a Markov process (with these transition matrices) with piecewise constant sample paths which are right continuous and have finitely many jumps on finite intervals. Moreover we show that the intensity measure plays an important role in the Doob-Meyer decomposition of the processes counting each type of jump. The crucial assumption that makes the construction of these processes possible is the assumption of bounded variation. But the atoms of size -1 in the diagonal elements of α play a special role. The idea of the construction is that when in state i, the intensity measure of the time at which the process leaves the state is given by $-\alpha_{ii}$; and given that a jump from i occurs at time t then the probability that the new state is j is $-(\mathrm{d}\alpha_{ij}/\mathrm{d}\alpha_{ii})(t)$. The process is started at time zero in an arbitrary state.

We can construct the process $X(t)$ using the theory of multivariate counting processes, see Jacobsen (1982), or the theory of marked point processes, see Jacod (1975). Suppose first that α has no atoms of size -1. Let $0 < T_1 < T_2 < \ldots < \infty$ be random times and let $X_1, X_2, \ldots$ be random variables taking values in the jump space $\{(i,j), i \neq j, i \in E, j \in E\}$. Let X_0 be a random initial state of the form (i,i) with an arbitrary distribution. We then specify for $n \geq 0$

$$\mathrm{P}\{\ T_{n+1} > t \ \mid T_1, \ldots, T_{n-1}, T_n = s,\ X_0, \ldots, X_{n-1}, X_n = (i_n, i)\ \}$$

$$= \prod\nolimits_{]s,t]}(1+\mathrm{d}\alpha_{ii}) \tag{74}$$

whenever $X_1, \ldots, X_n$ form a chain, i.e. the last coordinate of X_i is equal to the first coordinate of X_{i+1}. Also we specify

$$\mathrm{P}\{\ X_{n+1} = (i,j) \ \mid T_1, \ldots, T_n,\ T_{n+1} = t,\ X_0, \ldots, X_{n-1},\ X_n = (i_n, i)\ \}$$

$$= -(\mathrm{d}\alpha_{ij}/\mathrm{d}\alpha_{ii})(t). \tag{75}$$

These relations define the joint distribution of all (T_n, X_n). It follows from Jacod (1975) that the counting process

$$N_{ij}(t) = \Sigma_{1 \leq n < \infty}\, 1\{\ T_n \leq \mathrm{t},\ X_n = (i,j)\}$$

has predictable compensator given by the formula

$$\int_{]0,t]} \Sigma_{1 \leq n < \infty}\, 1\{T_n < s \leq T_{n+1}\} G_n(\mathrm{d}s, x) / \int_{[s,\infty]} (\Sigma_x\, G_n(\mathrm{d}v, x)) \tag{76}$$

in Jacod's notation. Here x is the mark (i,j) and G_n is the joint distribution of T_{n+1} and X_{n+1} given the past values of T and X. In the present case we find from (75) and (76) that

$$G_n(\mathrm{d}s, (i,j)) \ = \ -\,\mathrm{d}\{\prod\nolimits_{]u,s]}(1+\mathrm{d}\alpha_{ii})\}.[-\mathrm{d}\alpha_{ij}/\mathrm{d}\alpha_{ii}](s) \tag{77}$$

if the last value of T was u and the last value of X was (k,i) for some k in E. From (43) we find (for fixed u) that

$$\mathrm{d}\prod\nolimits_{]u,s]}(1+\mathrm{d}\alpha_{ii}) \ = \ -\prod\nolimits_{]u,s[}(1+\mathrm{d}\alpha_{ii})\}\ \alpha_{ii}(\mathrm{d}s). \tag{78}$$

The denominator of (76) is

$$\Sigma_{j \neq i} G_n([s,\infty], (i,j)) \ = \ \prod\nolimits_{]u,s[}(1+\mathrm{d}\alpha_{ii}). \tag{79}$$

Combining (77), (78) and (79) we find that the predictable compensator A_{ij} of N_{ij} is given by

$$\int_{]0,t]} \sum_{1\le n<\infty} 1\{ T_n < u \le T_{n+1}, X_n = (k,i) \text{ for some } k\} \alpha_{ij}(\mathrm{d}u)$$
$$= \int_{]0,t]} 1\{X(u-) = i\}\alpha_{ij}(\mathrm{d}u),$$

where the process $X(t)$ is defined by

$$\{ X(t) = i\} \;=\; \cup_{n=0,\, k\in E}^{\infty}\{X_n = (k,i) \text{ and } T_n \le t \le T_{n+1} \}.$$

Thus we can rewrite N_{ij} as

$$N_{ij}(t) \;=\; \#\{s \le t \;: X(s-) = i\,, X(s) = j \;\}$$

and we have shown that $\{ M_{ij}=N_{ij}-A_{ij} \}$ is a collection of orthogonal square integrable martingales. In other words, $N_{ij}=A_{ij}+M_{ij}$ is the Doob-Meyer decomposition of N_{ij}. Since $\mathrm{E}N_{ij}(t) = \mathrm{E}A_{ij}(t) \le \alpha_{ij}(0,t) < \infty$ for all $t<\infty$ we have a finite number of jumps in bounded intervals and the process X is well-defined. Jacobsen (1972) shows that $X(t)$ is a Markov process on $[0,\infty[$ and that the transition probabilities $P(s,t)$ satisfy certain integral equations. It is shown by Johansen (1986) that these are equivalent to the generalised Kolmogorov equations, and hence that the transition probabilities are given by the product integral $P(s,t) = \prod_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha)$. Jacobsen's (1972) proof requires that for any $s<t$ and any state i, the probability of not leaving i in $]s,t]$, given the process is there at time s, is positive. This requirement is satisfied precisely by ruling out the case of atoms of size -1 of any α_{ii}, because such an atom means that a jump out of i at this time is certain.

But as long as no α_{ii} has atoms of size -1 on $[0,a[$ the above construction defines a Markov process with intensity measure α on that interval. Now suppose in general that $a_1<a_2<\ldots$ are the positions of the atoms of size -1 of all the measures α_{ii}. We have just shown how to construct X on $[0,a_1[$. The fact that $X(a_1-)$ exists means that we can generate $X(a_1)$ by a jump from $X(a_1-)$ according to the transition matrix

$$P(a_1-,a_1) \;=\; \mathbf{1}+\alpha(\{a_1\}).$$

On the interval $[a_1,a_2[$ we repeat the construction defining a Markov process with state space E starting at $X(a_1)$ and with transition probabilitites $P(s,t)$, $a_1\le s \le t<a_2$. This is repeated to generate a process defined for all t. The martingale property of $N_{ij}-A_{ij}$ stays preserved in this procedure, as well as the property $P(s,t) = \prod_{]s,t]}(\mathbf{1}+\mathrm{d}\alpha)$.

If ever we arrive in an absorbing state the construction terminates. Moreover if for some time t and some subset of states absorption in this subset by this time is certain, the transition probabilities from states outside the subset become undefined (by reference to the process X) from this time onwards. However in the way described above we can, for any time t, define a Markov process starting at time t according to any initial distribution over the state space E. For all such constructed processes the transition probabilities (from states i and at times t with $\mathrm{P}\{X(t)=i\}>0$) are given by the elements of the matrices P.

We close this subsection with some remarks on statistical applications of these ideas. We have already mentioned the paper of Aalen & Johansen (1978) on nonparametric estimation of the transition probabilities of an inhomogeneous Markov process, based on censored observations from the process, which makes much use of product integration. Gill (1983b) uses product-integral methods to derive the asymptotic distribution of the processes N_{ij} themselves. Johansen (1981) considers nonparametric estimation of a Markov branching process. Since we now have an infinite state-space the transition probabilities cannot be handled as we have described above. However the Markovian nature of the process ensures that other multiplicative interval functions exist, in particular functions related to the mean and variance of the number of offspring per original parent over each time interval. Hjort (1984) gives a Bayesian treatment of the nonparametric estimation of the intensity measure of a Markov process and Hjort, Natvig & Funnemark (1985) give a reliability application, using the product-integral discrete approximation to the transition matrix to derive results on the association between states in time.

4.4. Likelihoods for counting process experiments.

In this subsection we give Jacod's (1975) formula for the Radon-Nikodym derivative of two probability measures on the measurable space generated by a multivariate counting process. 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 $N = (N_1,\ldots,N_k)$ be a multivariate counting process with compensator $A = (A_1,\ldots,A_k)$ with respect to a probability measure P and a filtration of the special form

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

Let P' be another probability measure, dominated by P, and suppose $\mathcal{F}_0$ contains all subsets of P'-null sets of $\mathcal{F} = \mathcal{F}_\tau$ (so the same holds for P too). Under P' N is still a multivariate counting process with respect to this filtration but its compensator is generally different, A' say. Recall that A can be interpreted as an integrated conditional intensity by the heuristic

$$\mathrm{d}A_i(t) = \mathrm{P}\{ \mathrm{d}N_i(t)=1 \mid \mathcal{F}_{t-} \}. \tag{80}$$

So given P on $\mathcal{F}_0$ one should be able to reconstruct P on $\mathcal{F}_\tau$ by multiplication of conditional probabilities. The next theorem makes this idea rigorous. Indeed, the distribution of N (given $\mathcal{F}_0$) is determined by its compensator in the way just indicated.

THEOREM 21 (Jacod, 1975). *Let* $L_t = (\mathrm{d}P'/\mathrm{d}P)|_{\mathcal{F}_t}$. *Almost surely* $A_i' << A_i$ *for each i and*

$$L_t = L_0 \cdot \prod_{n:T_n \le t} (\mathrm{d}A_{J_n}'/\mathrm{d}A_{J_n})(T_n) \cdot \prod_{s \ne T_n,\, s \le t} (1-\Delta\bar{A}'(s))/(1-\Delta\bar{A}(s)) \,.$$
$$. \exp(-\bar{A}'_c(t)+\bar{A}_c(t)). \qquad (81)$$

Here $0<T_1<T_2<$ ... are the jump times of N, J_1, J_2, ... are the corresponding jump types (i.e. $\Delta N_{J_n}(T_n) = 1$), $\bar{N}=\Sigma_i N_i$, similarly for $\bar{A}$, $\bar{A}_c$ is the continuous part of $\bar{A}$, etc. By using some new notational conventions we may rewrite (81) as follows:

$$L_t = L_0 \cdot \prod_{s \le t}\{(1-\mathrm{d}\bar{A}'(s))^{1-\mathrm{d}\bar{N}(s)} \prod_i (\mathrm{d}A_i'(s))^{\mathrm{d}Ni(s)}\} \,/$$
$$\prod_{s \le t}\{(1-\mathrm{d}\bar{A}(s))^{1-\mathrm{d}\bar{N}(s)} \prod_i (\mathrm{d}A_i(s))^{\mathrm{d}Ni(s)}\}. \qquad (82)$$

This expression should be evaluated by the use of some formal algebra together with the conventions

$$\prod_s(\,.\,)^{\,\mathrm{d}Ni(s)} = \prod_{s:\Delta N_i(s)=1}(\,.\,),$$
$$\prod_s(\,.\,)^{\,1-\mathrm{d}\bar{N}(s)} = \prod_{s:\Delta\bar{N}(s)=0}(\,.\,),$$
$$\mathrm{d}A_i'(s) \,/\, \mathrm{d}A_i(s) = (\mathrm{d}A_i'/\mathrm{d}A_i)(s),$$

and the product-integral notation. We may even restate the theorem as

THEOREM 22. *For* P'<<P *the Radon-Nikodym derivative or likelihood ratio*

$$L_t\,(P';\,P) = (\mathrm{d}P'/\mathrm{d}P)|_{\mathcal{F}_t} = L_t(P') \,/\, L_t(P)$$

is given by

$$L_t(P) = L_0(P) \,.\, \prod_{s \le t}\{(1-\mathrm{d}\bar{A}(s))^{1-\mathrm{d}\bar{N}(s)} \prod_i (\mathrm{d}A_i(s))^{\mathrm{d}Ni(s)}\}. \qquad (83)$$

Note the probabilistic interpretation of this equation, corresponding exactly to (80): the likelihood function is formed by multiplying together conditional likelihoods for the experiments in which $\mathrm{d}N(t)$ is generated by choosing component i to equal 1 with conditional probability (given $\mathcal{F}_{t-}$) $\mathrm{d}A_i(t)$, the other components are then equal to zero; all components are zero (so $1\text{-}\mathrm{d}\bar{N}(s)=1$) with probability $1\text{-}\mathrm{d}\bar{A}(s)$.

Expressions such as (83) are common in heuristic calculations in survival analysis, see for instance Kalbfleisch & Prentice (1980). The fact that they also have an exact mathematical interpretation allows one (see Andersen, Borgan, Gill & Keiding, 1988) 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 & Haara

(1984) can be clarified in this way. Also martingale properties of partial likelihoods and their associated score-functions can be easily derived, by using the Volterra integral equation characterization of the product integral and the formula for the compact derivative of the product integral. Johansen (1983) uses the product-integral formulation in giving an interpretation of Cox's estimator in the Cox (1972) regression model as a nonparametric maximum likelihood estimator.

The product-integral representation of the likelihood function can be generalized in the obvious way to the case of a marked point process; i.e. the number of different kinds of jumps of N need not be finite.

5. References.

O. O. Aalen (1975), *Statistical inference for a family of counting processes*, Ph.D. thesis, Berkeley; reprinted (1976) at Institute of Mathematical Statistics, University of Copenhagen.

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

P. K. Andersen, Ø. Borgan, R. D. Gill & N. Keiding (1988), *Statistical models for counting processes*, (in preparation).

E. Arjas & P. Haara (1984), A marked point process approach to censored failure data with complicated covariates, *Scand. J. Statist.* **11**, 193-209.

N. Arley (1943), *On the theory of stochastic processes and their application to the theory of cosmic radiation*, G. E. C. Gads Forlag, Copenhagen.

N. Arley & V. Borschenius (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.

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

P. Billingsley (1968), *Weak convergence of probability measures*, Wiley, New York.

H. Brunner & P. J. van der Houwen (1986), *The Numerical Solution of Volterra Integral Equations*, North Holland, Amsterdam.

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

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

J. M. C. Clark (1984), Asymptotically optimal quadrature formulae for stochastic integrals, *Proc. 23rd Conf. on Decision and Control* (2), 712-715 (IEEE).

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

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

C. Dellacherie & J.-P. Meyer (1982), *Probability and Potentials* (*B*), North-Holland, Amsterdam.

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

J. D. Dollard & C. N. Friedman (1979), *Product integration with applications to differential equations* (*with an appendix by P. R. Masani*), Addison-Wesley, Reading, Massachusetts.

M. Emery (1978), Stabilité des solutions des équations différentielles stochastiques: application aux intégrales multiplicatives stochastiques, *Z. Wahrsch. Verw. Gebiete* **41**, 241-262.

W. Feller (1940), On the integro-differential equations of purely discontinuous Markoff processes, *Trans. Amer. Math. Soc.* **48**, 488-515; errata, ibid. (1945) **58**, 474.

R. D. Gill (1980a), *Censoring and stochastic integrals*, MC Tract **124**, Mathematisch Centrum, Amsterdam.

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

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

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

R. D. Gill (1983b), Discussion of invited papers on dependent central limit theory, *Bull. Inst. Internat. Statist.* **50**(3), 239-243.

R. D. Gill (1987), *Non- and semi-parametric maximum likelihood estimators and the von Mises method* (*Part 1; revised version*), CWI report MS-R8709, Centrum voor Wiskunde en Informatica, Amsterdam (to appear in *Scand. J. Statist.*).

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

I. S. Helland (1982), Central limit theorems for martingales with discrete or continuous time, *Scand. J. Statist.* **9**, 79-94.

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

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

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

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

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

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

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

N. L. Hjort, B. Natvig, & E. Funnemark (1985), The association in time of a Markov process with application to multistate reliability theory, *J. Appl. Probab.* **22**, 473-479.

F. R. de Hoog & R. Weiss (1973), Asymptotic expansions for product intgration, *Math. Comp.* **27**, 295-306.

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

M. Jacobsen (1982), *Statistical analysis of counting processes*, Lecture Notes in Statistics **12**, Springer-Verlag, New York.

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

S. Johansen (1973), A central limit theorem for finite semigroups and its application to the imbedding problem for finite state Markov chains, *Z. Wahrsch. Verw. Gebiete* **26**, 171-190.

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

S. Johansen (1981), *The statistical analysis of a Markov branching process*, Preprint no. 5, Institute of Mathematical Statistics, Univ. of Copenhagen.

S. Johansen (1983), An extension of Cox's regression model, *Int. Statist. Rev.* **51**, 165-174.

S. Johansen (1986), Product integrals and Markov processes, *CWI Newsletter* (12) 3-13; originally appeared (1977) as: Preprint no. 3, Institute of Mathematical Statistics, Univ. of Copenhagen.

J. D. Kalbfleisch & R. L. Prentice (1980), *The statistical analysis of failure time data*, Wiley, New York.

E. L. Kaplan & P. Meier (1958), Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.* **53**, 457-481.

N. Keiding & R. D. Gill (1987), *Random truncation models and Markov processes*, Report MS-R8702, Centrum voor Wiskunde en Informatica, Amsterdam.

H. P. McKean (1969), *Stochastic Integrals*, Academic Press, New York.

J. S. Mac Nerney (1963), Integral equations and semigroups, *Illinois J. Math.* **7**, 148-173.

S. I. Marcus (1981), Modeling and approximation of stochastic differential equations driven by semimartingales, *Stochastics* **4**, 223-245.

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

M. Nikunen & E. Valkeila (1984), A note on one-dimensional distances between two counting processes, *Theory Probab. Appl.* **29**, 558-561.

E. Pardoux & D. Talay (1985), Discretization and simulation of stochastic differential equations, *Acta Appl. Math.* **3**, 23-47.

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

D. Pollard (1984), *Convergence of stochastic processes*, Springer-Verlag, New York.

G. Rasch (1931), *Om matrixregning og dens anvendelse paa differens- og differentialligninger*, Levin & Munksgaards Forlag, Copenhagen.

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

J. A. Reeds (1976), *On the definition of von Mises functionals* (Ph.D. thesis), Research report S-44, Dept. of Statistics, Harvard University.

L. Schlesinger (1931), Neue Grundlagen für einen Infinitesimalkalkul der Matrizen, *Math. Zeit.* **33**, 33-61.

L. Schlesinger (1933), Weitere Beiträge zum Infinitesimalkalkul der Matrizen, *Math. Zeit.* **35**, 485-501.

A. W. van der Vaart (1987), *Statistical estimation for large parameter spaces*, Ph.D. thesis, University of Leiden (to appear in series CWI tracts, Centrum voor Wiskunde en Informatica, Amsterdam).

E. Valkeila (1982), A general Poisson approximation theorem, *Stochastics* **7**, 159-171.

V. Volterra (1887), Sulle equazione differzeniali lineari, *Rendiconti Accademia dei Lincei* **3**, 393-396.

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

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

A. Young (1954a), Approximate product integration, *Proc. Roy. Soc. London Ser. A*, **224**, 552-561.

A. Young (1954b), The application of approximate product integration to the numerical solution of integral equations, *Proc. Roy. Soc. London Ser. A*, **224**, 561-572.

PREPRINTS 1986

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Jespersen, N.C.B.: On the Structure of Simple Transformation Models.

No. 2 Dalgaard, Peter and Johansen, Søren: The Asymptotic Properties of the Cornish-Bowden-Eisenthal Median Estimator.

No. 3 Jespersen, N.C.B.: Dichotomizing a Continuous Covariate in the Cox Regression Model.

No. 4 Asmussen, Søren: On Ruin Problems and Queues of Markov-Modulated M/G/1 Type.

No. 5 Asmussen, Søren: The Heavy Traffic Limit of a Class of Markovian Queueing Models.

No. 6 Jacobsen, Martin: Right Censoring and the Kaplan - Meier and Nelson - Aalen Estimators.

PREPRINTS 1987

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Jensen, Søren Tolver and Johansen, Søren: Estimation of Proportional Covariances.

No. 2 Rootzén, Holger: Extremes, Loads, and Strengths.

No. 3 Bertelsen, Aksel: On the Problem of Testing Reality of a Complex Multivariate Normal Distribution.

No. 4 Gill, Richard D. and Johansen, Søren: Product-Integrals and Counting Processes.