Product Integration

Richard D. Gill

Mathematical Institute, University of Utrecht, Netherlands
EURANDOM, Eindhoven, Netherlands

May 24, 2001

Abstract

This is a brief survey of product-integration for statisticians.

All statisticians are familiar with the sum and product symbols $\sum$ and $\prod$, and with the integral symbol $\int$. Also they are aware that there is a certain analogy between summation and integration; in fact the integral symbol is nothing else than a stretched-out capital S—the S of summation. Strange therefore that not many people are aware of the existence of the product-integral $\prod$, invented by the Italian mathematician Vito Volterra in 1887, which bears exactly the same relation to the ordinary product as the integral does to summation.

The mathematical theory of product-integration is not terribly difficult and not terribly deep, which is perhaps one of the reasons it was out of fashion again by the time survival analysis came into being in the fifties. However it *is* terribly useful and it is a pity that Kaplan and Meier (1958), the inventors of the product-limit or Kaplan-Meier estimator (the nonparametric maximum likelihood estimator of an unknown distribution function based on a sample of censored survival times), did not make the connection, as neither did the authors of the classic and papers on this estimator, Efron (1967) and Breslow and Crowley (1974). Only with the Aalen and Johansen (1978) was the connection between the Kaplan-Meier estimator and product-integration made explicit. It took several more years before the connection was put to use to derive new large sample properties of the Kaplan-Meier estimator (e.g., the asymptotic normality of the Kaplan-Meier mean) in Gill (1983).

Modern treatments of the theory of product-integration with a view toward statistical applications can be found in Andersen et al. (1993), Gill and Johansen (1990), Gill (1994).

The Kaplan-Meier estimator is the product-integral of the Nelson-Aalen estimator of the cumulative or integrated hazard function; these two estimators bear the same relation to one another as the actual survival function (one minus the distribution function) and the actual cumulative hazard function. There are many other applications of product-integration in statistics, for instance in the study of multi-state processes (connected to the theory of Markov processes) as initiated by Aalen and Johansen (1978) and in the theory of partial likelihood (cf. Cox regression model); see Andersen et al. (1993). Product-integration also turns up in extreme-value theory where again the hazard rate plays an important role, and in stochastic analysis and martingale theory (stochastic integration), where it turns up under the name Doléans-Dades exponential martingale.

Properties of integrals are often easily guessed by thinking of them as sums of many, many (usually very small) terms. Similarly, product-integration generalises the taking of products. This makes properties of product-integrals easy to guess and to understand.

Let us define product-integration at a just slightly higher level of generality than Volterra's original definition (corresponding to the transition from Lebesgue to Lebesgue-Stieltjes integration). Suppose $X(t)$ is a $p \times p$ matrix-valued function of time $t \in [0, \infty)$. Suppose also X (or if you prefer, each component of X) is right continuous with left hand limits. Let $\mathbf{1}$ denote the identity matrix. The product-integral of X over the interval $(0, t]$ is now defined as

$$\prod_{0}^{t}\big(\mathbf{1} + \mathrm{d}X(s)\big) \;=\; \lim_{\max|t_i - t_{i-1}|\to 0} \prod \Big(\mathbf{1} + \big(X(t_i) - X(t_{i-1})\big)\Big)$$

where the limit is taken over a sequence of ever finer partitions $0 = t_0 < t_1 < \cdots < t_k = t$ of the time interval $[0, t]$. For the limit to exist, X has to be of bounded variation; equivalently, each component of X is the difference of two increasing functions.

A very obvious property of product-integration is its multiplicativity. Defining the product-integral over an arbitrary time interval in the natural way, we have for $0 < s < t$

$$\prod_{0}^{t}(\mathbf{1} + \mathrm{d}X) \;=\; \prod_{0}^{s}(\mathbf{1} + \mathrm{d}X)\prod_{s}^{t}(\mathbf{1} + \mathrm{d}X).$$

We can guess (for proofs, see Gill and Johansen (1990) or preferably Gill (1994)) many other useful properties of product-integrals by looking at various simple identities for finite products. For instance, in deriving asymptotic statistical theory it is often important to study the difference between two product-integrals. Now if $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ are two sequences of numbers, we have the identity:

$$\prod(1 + a_i) - \prod(1 + b_i) \;=\; \sum_j \prod_{i<j}(1 + a_i)(a_j - b_j)\prod_{i>j}(1 + b_i).$$

This can be easily proved by replacing the middle term on the right hand side of the equation, $(a_j - b_j)$, by $(1 + a_j) - (1 + b_j)$. Expanding about this difference, the right hand side becomes

$$\sum_j \Bigg(\prod_{i\le j}(1 + a_i)\prod_{i>j}(1 + b_i) \;-\; \prod_{i\le j-1}(1 + a_i)\prod_{i>j-1}(1 + b_i)\Bigg).$$

This is a telescoping sum; writing out the terms one by one the whole expression collapses to the two outside products, giving the left hand side of the identity. The same manipulations work for matrices. In general it is therefore no surprise, replacing sums by integrals and products by product-integrals, that

$$\prod_0^t (\mathbf{1}+\mathrm{d}X) - \prod_0^t (\mathbf{1}+\mathrm{d}Y) \;=\; \int_{s=0}^t \prod_0^{s-} (\mathbf{1}+\mathrm{d}X)(\mathrm{d}X(s)-\mathrm{d}Y(s)) \prod_{s+}^t (\mathbf{1}+\mathrm{d}Y).$$

This valuable identity is called the Duhamel equation (the name refers to a classical identity for the derivative with respect to a parameter of the solution of a differential equation).

As an example, consider the scalar case ($p = 1$), let A be a cumulative hazard function and $\widehat{A}$ the Nelson-Aalen estimator based on a sample of censored survival times. In more detail, we are considering the statistical problem of estimating the survival curve $S(t) = \Pr\{T \geq t\}$ given a sample of independently censored i.i.d. survival times $T_1, \ldots, T_n$. The cumulative hazard rate $A(t)$ is defined by

$$A(t) \;=\; \int_0^t \frac{\mathrm{d}S(t)}{S(t-)};$$

A is just the integrated hazard rate in the absoutely continuous case, the cumulative sum of discrete hazards in the discrete case. Let $t_1 < t_2 < \ldots$ denote the distinct times when deaths are observed; let r_j denote the number of individuals at risk just before time t_j and let d_j denote the number of observed deaths at time t_j. We estimate the cumulative hazard function A corresponding to S with the Nelson-Aalen estimator

$$\widehat{A}(t) = \sum_{t_j \leq t} \frac{d_j}{r_j}.$$

This is a discrete cumulative hazard function, corresponding to the discrete estimated hazard $\widehat{\alpha}(t_j) = d_j/r_j$, $\widehat{\alpha}(t) = 0$ for t not an observed death time. The product-integral of $\widehat{A}$ is then

$$\widehat{S}(t) \;=\; \prod_0^t (1-\mathrm{d}\widehat{A}) \;=\; \prod_{t_j \leq t} \left(1 - \frac{d_j}{r_j}\right),$$

which is nothing else than the Kaplan-Meier estimator of the true survival function S. The Duhamel equation now becomes the identity

$$\widehat{S}(t) - S(t) \;=\; \int_{s=0}^t \widehat{S}(s-)\Big(\mathrm{d}\widehat{A}(s) - \mathrm{d}A(s)\Big)\frac{S(t)}{S(s)}$$

which can be exploited to get both small sample and asymptotic results, see Gill (1980, 1983, 1994), Gill and Johansen (1990), Andersen et al. (1993). The same identity pays off in studying Dabrowska's (1988) multivariate product-limit estimator (see Andersen et al. (1993); Gill (1994); van der Laan (1996)), and in studying Aalen and Johansen's (1978) estimator of the transition matrix of an inhomogeneous Markov chain (see (Andersen et al., 1993)). It can be rewritten ((Gill, 1994)) as a so-called *van der Laan identity*, van der Laan (1996) expressing $\widehat{S} - S$ as a function-indexed empirical process, evaluated at a random argument, so that the classical large sample results for Kaplan-Meier (consistency, asymptotic normality) can be got by a two-line proof: without further calculations simply invoke the modern forms of the Glivenko-Cantelli theorem and the Donsker theorem; i.e., the functional versions of the classical law of large numbers and the central limit theorem respectively.

Taking Y identically equal to zero in the Duhamel equation yields the formula

$$\prod_{0}^{t}(\mathbf{1} + \mathrm{d}X) - \mathbf{1} \;=\; \int_{s=0}^{t} \prod_{0}^{s-}(\mathbf{1} + \mathrm{d}X)\mathrm{d}X(s).$$

This is the integral version of Kolmogorov's forward differential equation from the theory of Markov processes, and it is the type of equation:

$$Y(t) \;=\; \mathbf{1} + \int_{0}^{t} Y(s-)\mathrm{d}X(s)$$

(in unknown Y, given X), which motivated Volterra to invent product-integration. $Y(t) = \prod_{0}^{t}(\mathbf{1} + \mathrm{d}X)$ is the unique solution of this equation.

Notes on the References

Aalen and Johansen (1978) introduced simultaneously counting process theory and product-integration to the study of nonparametric estimation for Markov processes; the relevance to the Kaplan-Meier estimator was noted by the authors but not noticed by the world!

The book Andersen et al. (1993) contains a 'users' guide' to product-integration in the context of counting processes and generalised survival analysis.

Breslow and Crowley (1974) gave the first rigourous large-sample results for Kaplan-Meier using the then recently developed Billingsley-style theory of weak convergence.

Dabrowska (1988) invented or discovered a beautiful generalization of the product-limit characterization of the Kaplan-Meier estimator to higher dimensions. Other characterizations, e.g., nonparametric maximum likelihood, lead to other estimators; see Gill (1994), van der Laan (1996).

The classic paper Efron (1967) introduced the redistribute-to-the-right and self-consistency properties of the Kaplan-Meier estimator and claimed but did not prove weak convergence of the Kaplan-Meier estimator on the whole line in order to establish asymptotic normality of a new Wilcoxon generalization, results finally established in Gill (1980).

Gill and Johansen (1990) give the basic theory, some history, and miscellaneous applications. Gill (1994) contains some improvements and further applications.

Gill (1980) emphasized the counting process approach to survival analysis, using some product-limit theory from Aalen and Johansen (1978) but not highlighting this part of the theory.

Gill (1994) is perhaps cryptically brief in parts, but a yet more polished treatment of product-integration and its applications in survival analysis.

The classic Kaplan and Meier (1958) is actually number 2 in the list of most cited ever papers in mathematics, statistics and computer science (fourth place is held by Cox, 1972; first place by Duncan, 1955, on multiple range tests). The authors never met but submitted simultaneously their independent inventions of the product-limit estimator to *J. Amer. Statist. Assoc.*; the resulting joint paper was the product of postal collaboration.

van der Laan (1996) Contains a beautiful identity for the nonparametric maximum likelihood estimator in a missing data problem: estimator minus estimand equals the empirical process of the optimal influence curve evaluated at the estimator, $\widehat{F} - F = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IC}_{\mathrm{opt}}(X_i; \widehat{F})$; applications to the bivariate censored data problem as well as treatment of other estimators for the same problem.

References

Aalen, O. and S. Johansen (1978). An empirical transition matrix for non-homogenous Markov chains based on censored observations. *Scandinavian Journal of Statistics 5*, 141–150.

Andersen, P., Ø. Borgan, R. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.

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

Dabrowska, D. (1988). Kaplan-meier estimate on the plane. *Annals of Statistics 16*, 1475–1489.

Efron, B. (1967). The two sample problem with censored data. In L. LeCam and J. Neyman (Eds.), *Proc. 5th Berkeley Symp. Math. Strat. Prob.*, Berkeley, Ca., pp. 831–853. Univ. Calif. Press.

Gill, R. (1980). *Censoring and Stochastic Integrals*. MC Tract **124**. Amsterdam: Centre for Mathematics and Computer Science (CWI).

Gill, R. (1983). Large sample behaviour of the product limit estimator on the whole line. *Annals of Statistics*, 44–58.

Gill, R. (1994). Lectures on survival analysis. In P. Bernard (Ed.), *Lectures on Probability Theory (Ecole d'Été de Probabilités de Saint Flour XXII - 1992)*, Berlin, pp. 115–241. Springer-Verlag (SLNM 1581).

Gill, R. and S. Johansen (1990). A survey of product-integration with a view towards application in survival analysis. *Annals of Statistics 18*, 1501–1555.

Kaplan, E. and P. Meier (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association 53*, 457–481, 562–563.

van der Laan, M. (1996). *Efficient and Inefficient Estimation in Semiparametric Models*, Volume 114 of *CWI Tract*. Amsterdam: Centre for Mathematics and Computer Science.