

# Product Integration

Richard D. Gill

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

August 9, 2001

## Abstract

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

## 1 Product-Integration

Product-integration was introduced more than 110 years ago by the Italian mathematician Vito Volterra, as a tool in the solution of a certain class of differential equations. It was studied intensively by mathematicians for half a century, but finally the subject became unfashionable and lapsed into obscurity. That is a pity, since ideas of product-integration make a very natural appearance in survival analysis, and the development of this subject (in particular, of the Kaplan-Meier estimator) could have been a lot smoother if product-integration had been a familiar topic from the start. The Kaplan-Meier estimator is the product-integral of the Nelson-Aalen estimator of the cumulative hazard function; these two estimators bear the same relation to one another as the actual survival function and the actual cumulative hazard function. There are many other applications of product-integration in survival analysis, for instance in the study of multi-state processes (connected to the theory of Markov processes), and in the theory of partial likelihood.

Ordinary integration is a generalisation of summation, and properties of integrals are often easily guessed by thinking of them as sums of very, very many terms (all or most of them being very small). Similarly, product-integration generalises the taking of products; a product integral is a product of many, many terms (all or most of them being very close to the number 1). Thinking of product-integrals in this simplistic way is actually very helpful. Properties of product-integrals are easy to guess and to understand. The theory of product-integration can be a great help in studying the statistical properties of statistical quantities which explicitly or implicitly are defined in terms of product-integrals.

Before defining product-integrals in general and exhibiting some of their properties, we will discuss the relation, in survival analysis, between survival

function and hazard function. This will lead us naturally to the notion of product-integration in the most simple possible of contexts.

Consider a survival time  $T$  with survival function  $S(t) = \Pr(T > t)$ ,  $t \geq 0$ ;  $S(0) = 1$ . Suppose  $T$  is continuously distributed with a density  $f(t)$  and a hazard rate  $\alpha(t)$ . These two functions have intuitive probabilistic meanings: for a small time interval  $t, t+h$ , the unconditional probability  $\Pr(t \leq T \leq t+h) \approx f(t).h$ ; while the conditional probability  $\Pr(t \leq T \leq t+h | T \geq t) \approx \alpha(t).h$ . In fact, the probability density  $f(t) = -(d/dt)S(t)$  while the hazard rate  $\alpha(t) = f(t)/S(t)$ . One can mathematically recover the distribution function  $F(t) = 1 - S(t)$  from the density by integration;  $F(t) = \int_0^t f(s)ds$ . Also one can recover the survival function from the hazard rate. Noting that  $\alpha(t) = -(d/dt) \log S(t)$  one finds by integration (and using that  $S(0) = 1$ , hence  $\log S(0) = 0$ ), that  $-\log S(t) = \int_0^t \alpha(s)ds$  hence  $S(t) = \exp(-\int_0^t \alpha(s)ds)$ . This is simple enough, but neither the result nor its derivation have a probabilistic interpretation.

If the survival time  $T$  had a discrete distribution, one would introduce the discrete density  $f(t) = \Pr(T = t)$  and the discrete hazard  $\alpha(t) = \Pr(T = t | T \geq t) = f(t)/S(t-)$ . Still the survival function can be recovered from both density and hazard, but the formula in the latter case now seems quite different:  $S(t) = \prod_0^t (1 - \alpha(s))$ . The continuous case formula  $S(t) = \exp(-\int_0^t \alpha(s)ds)$  has therefore two major defects: firstly, it does not have any intuitive interpretation; and secondly; it gives the wrong generalisation to the discrete case.

Here is how both formulas can be unified and made intuitively interpretable. Define the cumulative hazard  $A(t)$  by, in the continuous case,  $A(t) = \int_0^t \alpha(s)ds$ , and in the discrete case,  $A(t) = \sum_0^t \alpha(s)$ . (These two formulas are special cases of the completely general expression  $A(t) = \int_0^t dS(s)/S(s-)$ .) Now we can write, both in the continuous and the discrete case,

$$S(t) = \prod_0^t (1 - dA(s)) \tag{1}$$

which can be interpreted as the product over many small time intervals  $s, s+ds$  making up the interval  $[0, t]$ , of the probability  $(1 - dA(s))$ . Since the hazard  $dA(s)$  can be thought of as the probability of dying in the interval from  $s$  to  $s+ds$  given survival up to the beginning of that time interval, 1 minus the hazard is the probability of surviving through the small time interval given survival up to its start. Multiplying over the small time intervals making up  $[0, t]$  yields the unconditional probability of surviving past  $t$ ; in other words, equation (1) is just the limiting form of the equality

$$\Pr(T > t) = \prod_{i=1}^k \Pr(T > t_i | T > t_{i-1}) = \prod_{i=1}^k \left(1 - \Pr(T \leq t_i | T > t_{i-1})\right)$$

where  $0 = t_0 < t_1 < \dots < t_k = t$  is a partition of the time interval  $[0, t]$ .

Consider now the statistical problem of estimating the survival curve  $S(t)$  given a sample of independently censored survival times. Let  $t_1 < t_2 < \dots$

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  $\widehat{\alpha}(t_j) = d_j/r_j$ ,  $\widehat{\alpha}(t)$  zero for  $t$  not an observed death time. The product-integral of  $\widehat{A}$  is then

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

which is nothing else than the Kaplan-Meier estimator.

The actual definition of the product-integral in (1) is the following:

$$\prod_0^t (1 - dA(s)) = \lim_{\max |t_i - t_{i-1}| \rightarrow 0} \prod (1 - (A(t_i) - A(t_{i-1})))$$

where the limit is taken over a sequence of ever finer partitions  $0 = t_0 < t_1 < \dots < t_k = t$  of the time interval  $[0, t]$ .

From this point we can choose either to study properties of the product-integral or define it in greater generality. Both aspects are important in applications. Let us first give a more general definition. The important generalisation is that we will define product-integrals of matrix-valued functions, rather than just scalar valued functions. The concept now really comes into its own, because when we multiply a sequence of matrices together the result will generally depend on the order in which the matrices are taken. Even in the continuous case there will not be a simple exponential formula expressing the result in terms of an ordinary integral. Multiplying products of matrices turns up in the theory of Markov processes, and this connects directly to the statistical analysis of multi-state models in survival analysis.

Suppose  $X(t)$  is a  $p \times p$  matrix-valued function of time  $t$ . Suppose also  $X$  (or if you like, each component of  $X$ ) is right continuous with left hand limits. Let  $I$  denote the identity matrix. The product-integral of  $X$  over the interval  $[0, t]$  is now defined as

$$\prod_0^t (I + dX(s)) = \lim_{\max |t_i - t_{i-1}| \rightarrow 0} \prod (I + (X(t_i) - X(t_{i-1})))$$

where as always the limit is taken over a sequence of ever finer partitions  $0 = t_0 < t_1 < \dots < t_k = t$  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.

## 2 Application to Markov processes

We briefly sketch the application of product-integration to Markov processes. Suppose an individual moves between  $p$  different states as time proceeds, staying in each state for some random length of time and then jumping to another. Suppose the individual has intensity  $\alpha_{ij}(t)$  of jumping from state  $i$  to state  $j$  at time  $t$ , given the whole past history (in other words, the process is Markov: the intensity only depends on the present time and the present state). Define cumulative intensities  $A_{ij}(t) = \int_0^t \alpha_{ij}(s) ds$  and negative total cumulative intensity of leaving a state  $A_{ii} = -\sum_{j \neq i} A_{ij}$ . Collect these into a square matrix valued function of time  $A$ . Then one can show that the matrix of transition probabilities  $P(0, t)$ , whose  $ij$  component is the probability of being in state  $j$  at time  $t$  given the individual started at time 0 in state  $i$ , is given by a product-integral of  $A$ :

$$P(0, t) = \prod_0^t (I + dA(s)).$$

This formula generalises the usual formula for transition probabilities of a discrete time Markov chain, since the matrix  $(I + dA(s))$  can be thought of as the transition probability matrix for the small time interval  $s, s + ds$ .

Given possibly censored observations from a Markov process, one can estimate the elements of the matrix of cumulative intensities  $A$  by Nelson-Aalen estimators. The corresponding estimate of the transition probabilities, the Aalen-Johansen estimator, is found by taking the product-integral of  $\hat{A}$ .

## 3 Mathematical properties

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  $0 < s < t$

$$\prod_0^t (I + dX) = \prod_0^s (I + dX) \prod_s^t (I + dX).$$

We can guess many other useful properties of product-integrals by looking at various simple identities for finite products. For instance, it is often important to study the difference between two product-integrals. Now if  $a_1, \dots, a_k$  and  $b_1, \dots, b_k$  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,  $(a_j - b_j)$ , by  $(1 + a_i) - (1 + b_i)$ . Expanding about this difference, the right hand side becomes

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

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

$$\prod_0^t (I + dX) - \prod_0^t (I + dY) = \int_{s=0}^t \prod_0^{s-} (I + dX)(dX(s) - dY(s)) \prod_{s+}^t (I + dY).$$

This valuable identity is called the Duhamel equation.

As an example, consider the scalar case, let  $A$  be a cumulative hazard function and  $\hat{A}$  the Nelson-Aalen estimator based on a sample of censored survival times. Let  $S$  be the corresponding survival function and  $\hat{S}$  the Kaplan-Meier estimator. The Duhamel equation then becomes the identity

$$\hat{S}(t) - S(t) = \int_{s=0}^t \hat{S}(s-)(d\hat{A}(s) - dA(s))(S(t)/S(s))$$

which can be exploited to get both small sample and asymptotic results for the Kaplan-Meier estimator. We illustrate one other important identity in a similar manner. Note that

$$\prod_{i \leq j} (1 + a_i) - \prod_{i \leq j-1} (1 + a_i) = \prod_{i \leq j-1} (1 + a_i) a_j.$$

Adding over  $j$  from 1 to  $k$  gives us

$$\prod_{i \leq k} (1 + a_i) - 1 = \sum_j \prod_{i \leq j-1} (1 + a_i) a_j.$$

Now we can guess the identity

$$\prod_0^t (I + dX) - I = \int_{s=0}^t \prod_0^{s-} (I + dX) dX(s).$$

This is essentially Kolmogorov's forward equation from the theory of Markov processes, and it is the type of equation—solve  $Y(t) = I + \int_0^t Y(s-)dX(s)$  for unknown  $Y$ , given  $X$ —which originally motivated Volterra to invent product-integration.  $Y(t) = \prod_0^t (I + dX)$  is the unique solution of this equation. (It is also just a special case of the Duhamel equation when we take the second integrand  $Y$  identically equal to zero).

## 4 Concluding remarks

The product-integral seems first to have been used as a fundamental tool in modern survival analysis by Aalen and Johansen (1978), though it also appears in a more informal context in Cox (1972) and in Kalbfleisch and Prentice

(1980). Surveys of the theory of product-integration are given by Gill and Johansen (1990), Gill (1994). The former paper also pays attention to the earlier history of the subject. In particular, it is worth mentioning that a large variety of notations has been used for the product-integral, including large curly P's, product-symbols, and the ordinary integral sign embellished with a half circle over the top.

As well as playing a role in the theory of the Kaplan-Meier and the Aalen-Johansen estimators, the product-integral is also a useful way to write likelihoods and partial-likelihoods in survival analysis, since these can be usefully thought of as continuous products of conditional likelihoods for the data in each new infinitesimal time interval given the past. The product-integral is also useful in multivariate survival analysis. In particular, Dabrowska's 1988 multivariate product-limit estimator is based on a representation of a multivariate survival function in terms of product-integrals of a collection of higher-dimensional joint and conditional hazard functions. The book Andersen et al. (1993) gives a brief survey of the theory and many detailed applications, covering all the topics mentioned above.

## 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.
- Cox, D. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* 34, 187–220.
- Dabrowska, D. (1988). Kaplan-meier estimate on the plane. *Annals of Statistics* 16, 1475–1489.
- 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.
- Kalbfleisch, J. and R. Prentice (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.