Variate Generation for Accelerated Life and Proportional Hazards Models with Time Dependent Covariates

Lawrence M. Leemis

Li-Hsing Shih

Kurt Reynertson

The University of Oklahoma

School of Industrial Engineering

202 West Boyd, Room 124

Norman, OK 73019

September, 1989

SUMMARY

Variate generation algorithms for lifetimes when survival models incorporate time dependent covariates are presented. These algorithms are closed form for special cases of the function that links the covariate values to the survivor distribution. These algorithms are illustrated by several examples.

Some key words: Accelerated life model, Monte Carlo simulation, Proportional hazards model, Time dependent covariates, Variate generation.

1. Introduction

The inclusion of time dependent covariates in a survival model significantly complicates analysis. As survival models become more complex, fewer tractable solutions will be available and Monte Carlo methods will be relied on more often to estimate performance measures. The purpose of this paper is to give algorithms for generating lifetimes from models with time dependent covariates for Monte Carlo simulation analysis.

Covariates accommodate complex lifetime models which may be appropriate in certain medical or industrial situations. Analysis techniques for models containing covariates that do not change with time have been established. In a medical setting, examples of covariates that are not time dependent include gender and treatment type. Examples of covariates that may change with time include the dosage level of a drug and the caloric intake. In industrial applications, environmental factors of products such as temperature or stress may vary with time. Two models that are often used to incorporate the effects of these covariates on lifetimes are the *accelerated life* and *proportional hazards* models, which will be considered here.

2. Notation

Let $h(t)$ and $H(t) = \int_0^t h(\tau)d\tau$ be the hazard and cumulative hazard functions, respectively, for a continuous nonnegative random variable T, the lifetime of the individual under study. The $q \times 1$ vector $\mathbf{z}(t)$ contains time dependent covariates associated with a particular individual. The covariates are incorporated into the lifetime distribution by the *link function* $\psi(\mathbf{z}(t))$, which satisfies $\psi(\mathbf{O}) = 1$ and $\psi(\mathbf{z}(t)) \geq 0$ for all $\mathbf{z}(t)$. A common choice for the link function is the log linear form $\psi(\mathbf{z}(t)) = e^{\beta' \mathbf{z}(t)}$, where β is a $q \times 1$ vector of regression coefficients.

The accelerated life and proportional hazards models in the constant covariate case are described in Cox and Oakes (1984). Table 1 shows the relationships between the cumulative hazard, hazard, survivor and probability density functions for these two models, where the subscript 0 denotes the baseline distribution.

	Accelerated Life	Proportional Hazards
$H(t;\mathbf{z})$	$H_0(t\psi(\mathbf{z}))$	$\psi(\mathbf{z})\,H_0(t)$
$h(t;\mathbf{z})$	$\psi(\mathbf{z})h_0(t\psi(\mathbf{z}))$	$\psi(\mathbf{z})h_0(t)$
$S(t;\mathbf{z})$	$S_0(t\psi(\mathbf{z}))$	$[S_0(t)]^{\psi(\mathbf{z})}$
$f(t;\mathbf{z})$	$\psi(\mathbf{z})f_0(t\psi(\mathbf{z}))$	$f_0(t)\psi(\mathbf{z})[S_0(t)]^{\psi(\mathbf{z})-1}$

Table 1. Distribution relationships for regression models with constant covariates

Cox and Oakes (1984) defined the accelerated life model with time dependent covariates by

$$\frac{dt^{(\mathbf{z})}}{dt^{(\mathbf{O})}} = \frac{1}{\psi(\mathbf{z}(t^{(\mathbf{z})}))}$$

where $t^{(\mathbf{z})}$ is time for an individual having covariates $\mathbf{z}(t)$ and $t^{(\mathbf{O})}$ is time for an individual having the baseline distribution. This is a time dependent generalization of $T_1 / T_0 = 1 / \psi$ for the two sample (single binary covariate) case where $H(t;\mathbf{z}) = H_0(t\,\psi(z))$, T_0 denotes the lifetime under the baseline case and T_1 denotes the lifetime in the treatment case. Integration of this equation yields

$$t^{(\mathbf{O})} = \int_0^{t^{(\mathbf{z})}} \psi(\mathbf{z}(u))du = \Psi(t^{(\mathbf{z})})$$

where Ψ is the *cumulative link function,* and $T = \Psi^{-1}(T_0)$. The cumulative hazard function H for an individual with covariates $\mathbf{z}(t)$ can be written as

$$H(t;\,\mathbf{z}(t)) = H_0(t^{(\mathbf{O})}) = H_0(\Psi(t))$$

Table 2 shows several parallel relationships between the accelerated life model with constant and time dependent covariates.

Cox and Oakes (1984) also defined the proportional hazards model for time dependent covariates as

$$h(t;\,\mathbf{z}(t)) = \psi(\mathbf{z}(t))\,h_0(t)$$

	Constant Covariates	Time Dependent Covariates
Link Function	$\psi(\mathbf{z})$	$\psi(\mathbf{z}(t))$
Two Samples (z binary)	$\frac{T_1}{T_0} = \frac{1}{\psi}$	$\frac{dt}{dt^{(O)}} = \frac{1}{\psi(z(t))}$
Cumulative Link Function	$\Psi(t) = t\psi(\mathbf{z})$	$\Psi(t) = \int_0^t \psi(\mathbf{z}(u))du$
Cumulative Hazard Function	$H(t;\mathbf{z}) = H_0(t\psi(\mathbf{z}))$	$H(t;\mathbf{z}(t)) = H_0(\Psi(t))$
Baseline case ($\mathbf{z}(t) = \mathbf{O}$)	$\psi(\mathbf{O}) = 1$	$\psi(\mathbf{O}) = 1$

Table 2. Accelerated life models with constant and time dependent covariates

where $h_0(t)$ is the baseline hazard function. This is a generalization of the constant covariate case, where the hazard function is $h(t;\,\mathbf{z}) = \psi(\mathbf{z})\,h_0(t)$. The cumulative hazard function is obtained by integrating both sides of the hazard function

$$H(t;\,\mathbf{z}(t)) = \int_0^t \psi(\mathbf{z}(u))\,h_0(u)du$$

3. Literature Review

Ciampi and Etezadi-Amoli (1985) defined a general hazard function by $h(t;\,\mathbf{z}) = \psi(\alpha'\mathbf{z})\,h_0(\psi(\beta'\mathbf{z})\,t)$ where the link function has log linear form, and α and β are $q \times 1$ vectors of regression coefficients. This formulation facilitates hypothesis testing for proportional hazards versus accelerated life models

$$H_0\colon\ \beta = \mathbf{O}\ \ (\textit{the proportional hazards model})$$

$$H_1\colon\ \beta = \alpha\ \ (\textit{the accelerated life model})$$

Simulated and real data are analyzed to show the effectiveness of this approach.

Several authors consider parameter estimation for regression models. Prentice and Kalbfleisch (1979) discussed estimation problems in the proportional hazards, accelerated life and competing risks models with covariates. They also consider covariates that vary with time. Dale (1985) used the proportional hazards model in the presence of time dependent covariates. Kalbfleisch and McIntosh (1977) compared the efficiency of the partial likelihood method and Weibull analysis for the Cox proportional hazards model with time dependent covariates. Petersen (1986a, 1986b) proposed an algorithm for estimating parameters by maximum likelihood in a large variety of parametric survival models by using the Gauss-Newton method. The approach allows for a flexible treatment of time dependent covariates.

Devroye (1986) has written the first textbook completely devoted to variate generation. Griffith (1982) and others have observed the cumulative hazard function, $H(t)$ has a unit exponential distribution. Nessim and Preston (1983) used Monte-Carlo techniques to test for treatment effect with censored survival data and a binary covariate. Hoffmann (1985) considered the Weibull and piecewise constant hazard functions as baseline distributions for the proportional hazards model with time-varying covariates. He gave techniques for generating random variates from these models. Smith (1987) included cost factors in a Monte Carlo evaluation of a system of components. Leemis (1987) gave algorithms to generate random lifetimes via inversion of cumulative hazard function in the constant covariates case, as shown in Table 3 (where a is the time of the last event and u is uniformly distributed between 0 and 1).

	Renewal	NHPP
Accelerated life	$t \leftarrow a + \dfrac{H_0^{-1}(-\,log(u))}{\psi(\mathbf{z})}$	$t \leftarrow \dfrac{H_0^{-1}(H_0(a\psi(\mathbf{z})) - log(u))}{\psi(\mathbf{z})}$
Proportional hazards	$t \leftarrow a + H_0^{-1}(\dfrac{-\,log(u)}{\psi(\mathbf{z})})$	$t \leftarrow H_0^{-1}(H_0(a) - \dfrac{log(u)}{\psi(\mathbf{z})})$

Table 3. Variate generation for models with constant covariates

If u were replaced by $1-u$ in Table 3, the transformations would be monotonic, although not as efficient computationally.

4. Variate Generation Methods

Generating lifetimes corresponding to individuals having time dependent covariates is more complicated than the constant covariate case. Although we consider only inversion here, more complicated techniques, such as thinning, may be adapted to variate generation for models with time dependent covariates. The following two subsections consider variate generation for the accelerated life and proportional hazards models separately.

4.1 Accelerated life model

Since the cumulative hazard function, $H(T)$, has a unit exponential distribution, it can be equated to $-\log(1-u)$ for variate generation

$$H(t; \mathbf{z}(t)) = -\log(1-u) = H_0(\Psi(t))$$

A random variate t can be generated by two consecutive inversions

$$t \leftarrow \Psi^{-1}(H_0^{-1}(-\log(1-u)))$$

where u is uniformly distributed between 0 and 1. Whether or not the variate generation is closed form depends on the tractability of the inverses of the Ψ and H_0 functions.

The exponential distribution is used as a baseline in the following three examples. Random variate generation can be performed by inverting the cumulative hazard function

$$H_0^{-1}(y) = \frac{y}{\lambda}$$

where λ is the failure rate. Substituting this into the general variate generation expression yields

$$t \leftarrow \Psi^{-1}\left(\frac{-log(1-u)}{\lambda}\right)$$

The following three examples consider three simple link functions: exponential, power and step functions.

Example 1 (exponential link function) Let the link function be

$$\psi(\mathbf{z}(t)) = e^{\beta t}$$

which corresponds to the cumulative link function

$$\Psi(t) = \int_0^t e^{\beta u} du$$

$$= \frac{1}{\beta}(e^{\beta t} - 1) \qquad t > 0$$

The inverse of cumulative link function is

$$\Psi^{-1}(y) = \frac{1}{\beta}\log(\beta y + 1)$$

Using $-\,log(1-u)$ as an argument in the variate generation expression

$$t \leftarrow \frac{1}{\beta}\log(\frac{-\beta\log(1-u)}{\lambda} + 1)$$

Example 2 (power link function) Let the link function have the form

$$\psi(z(t)) = t^k$$

so that the cumulative link function is

$$\Psi(t) = \int_0^t u^k du = \frac{t^{k+1}}{k+1}$$

and the inverse is

$$\Psi^{-1}(y) = (y(k+1))^{\frac{1}{k+1}}$$

As before, a closed form is obtained for generating lifetimes

$$t \leftarrow \left(\frac{-(k+1)\,log(1-u)}{\lambda}\right)^{\frac{1}{k+1}}$$

Example 3 (step link function) Let the link function be a step function

$$\psi(z(t)) = \begin{cases} c_1 & 0 < t < b \\ c_2 & t \geq b \end{cases}$$

This is the case where an individual is given dosage c_1 of a drug up to time b, then dosage c_2 thereafter.

The cumulative link function is

$$\Psi(t) = \begin{cases} c_1 t & 0 < t < b \\ bc_1 + c_2(t-b) & t \geq b \end{cases}$$

The inverse of Ψ is

$$\Psi^{-1}(y) = \begin{cases} y/c_1 & 0 < y < bc_1 \\ (y - bc_1)/c_2 + b & y \geq bc_1 \end{cases}$$

The general expression for variate generation is then closed form

$$t \leftarrow \begin{cases} \dfrac{-\,log(1-u)}{\lambda c_1} & 0 < \dfrac{-\,log(1-u)}{\lambda} < bc_1 \\ -\dfrac{1}{c_2 \lambda}\log(1-u) - \dfrac{bc_1}{c_2} + b & \dfrac{-\,log(1-u)}{\lambda} \geq bc_1 \end{cases}$$

These three examples of how to generate lifetimes in the accelerated life model have been chosen because of their tractability. Numerical methods are required if the covariates, the link function or the baseline distribution are more complicated.

4.2 Proportional hazards model

In the proportional hazards model, random variates can be generated by inverting the cumulative hazard function

$$H(t;\mathbf{z}(t)) = \int_0^t \psi(\mathbf{z}(u))h_0(u)du$$

A closed form equation for variate generation requires inversion of $H(t;\mathbf{z}(t))$.

Example 4 (Exponential baseline distribution) If the baseline function is exponential, then $h_0(t) = \lambda$, and the cumulative hazard function with the link function ψ is

$$H(t;\mathbf{z}(t)) = \lambda \int_0^t \psi(\mathbf{z}(u))du = \lambda\,\Psi(t)$$

where Ψ is the cumulative link function. Since the cumulative hazard function is unit exponentially distributed, a lifetime variate t is generated by

$$t \leftarrow \Psi^{-1}\left(\frac{-\log(1-u)}{\lambda}\right)$$

which is identical to the accelerated life model.

Example 5 (Weibull baseline distribution) Assume the baseline distribution is Weibull with hazard function $h_0(t) = \beta\lambda^\beta t^{\beta-1}$. The cumulative hazard function with link function ψ is

$$H(t;\mathbf{z}(t)) = \beta\lambda^\beta \int_0^t \psi(\mathbf{z}(u))u^{\beta-1}du$$

Tractability of the inverse of H depends on the choice of link function. For example, if the link function has the power form, $\psi(\mathbf{z}(t)) = t^k$, then the cumulative hazard function H is

$$H(t;\mathbf{z}(t)) = \beta\lambda^\beta \int_0^t u^{\beta+k-1}du = \frac{\beta\lambda^\beta t^{\beta+k}}{\beta+k}$$

Therefore, a variate can be generated by

$$t \leftarrow \left(\frac{-\log(1-u)(\beta+k)}{\beta\lambda^\beta}\right)^{\frac{1}{\beta+k}}$$

On the other hand, if the link function is a step function

$$\psi(\mathbf{z}(t)) = \begin{cases} c_1 & 0 \le t < b \\ c_2 & t \ge b \end{cases}$$

the cumulative hazard function is

$$H(t;\mathbf{z}(t)) = \begin{cases} c_1(\lambda t)^\beta & 0 \le t < b \\ c_1(\lambda b)^\beta + c_2\lambda^\beta(t^\beta - b^\beta) & t \ge b \end{cases}$$

A lifetime variate t can be generated by

$$t \leftarrow \begin{cases} \frac{1}{\lambda}\left(\frac{-\log(1-u)}{c_1}\right)^{\frac{1}{\beta}} & 0 \le u < 1 - e^{-c_1(b\lambda)^\beta} \\ \left(\frac{-\log(1-u) - c_1(\lambda b)^\beta}{c_2\lambda^\beta} + b^\beta\right)^{\frac{1}{\beta}} & 1 - e^{-c_1(b\lambda)^\beta} \le u \le 1 \end{cases}$$

5. Conclusions

Lifetime variate generation for Monte Carlo simulation is complicated by the presence of time dependent covariates. When particular forms of the covariates, link function and baseline distribution are chosen and there is a single covariate, variate generation is closed form. Numerical methods must be relied on when there are several time dependent covariates, the cumulative link function is not invertible or when the cumulative hazard function is not invertible.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of a Biomedical Research Support Grant from The University of Oklahoma with funds provided from the National Institutes for Health in preparing this paper. The authors also thank a referee for his helpful suggestions.

References

Ciampi, A. and Etezadi-Amoli, J. (1985), A General Model for Testing the Proportional Hazards and the Accelerated Failure Time Hypotheses in the Analysis of Censored Survival Data with Covariates, *Communications in Statistics: Theory and Methods* **14,** 651-667.

Cox, D.R. and D. Oakes (1984), *Analysis of Survival Data* (Chapman and Hall, London).

Dale, C.J. (1985), Application of the Proportional Hazards Model in the Reliability Field, *Reliability Engineering* **10,** 1-14.

Devroye, L. (1986), *Non-Uniform Random Variate Generation* (Springer-Verlag, New York).

Griffith, W. (1982), Representation of Distributions Having Monotone or Bathtub Shaped Hazard Functions, *IEEE Transactions on Reliability* **R-31,** 95-96.

Hoffman, R.G. (1985), Simulated Temporal Models for Proportional Hazards Models with Time Varying Covariates, American Statistical Association, Statistical Computing Section, 390-395.

Kalbfleisch, J.D. and A.A. McIntosh (1977), Efficiency in Survival Distributions with Time-Dependent Covariables, *Biometrika* **64,** 47-50.

Leemis, L.M. (1987), Variate Generation for Accelerated Life and Proportional Hazards Models, *Operations Research* **35,** 892-894.

Nessim, S. and D.L. Preston (1983), A Monte-Carlo comparison of several methods for the analysis of censored survival data with treatment and covariate effects, *Computational Statistics and Data Analysis* **1,** 183-190.

Petersen, T. (1986a), Fitting Parametric Survival Models with Time-Dependent Covariates, *Applied Statistics* **35,** 281-288.

Petersen, T. (1986b), Estimating Fully Parametric Hazard Rate Models with Time-Dependent Covariates, *Sociological Methods & Research* **14,** 219-246.

Prentice, R.L. and J.D. Kalbfleisch (1979), Hazard Rate Models with Covariates, *Biometrics* **35,** 25-39.

Smith, J.R. (1987), Reliability Analysis by Simulation, ASQC 1987 Annual Technical Conference Transactions, 654-662.