The Covariate Order Method for Nonparametric Censored Exponential Regression

Bo Henry Lindqvist — bo@math.ntnu.no
Jan Terje Kvaløy — jtk@math.ntnu.no

Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim

1. Introduction and Summary

In reliability engineering a first analysis of failure data is often based on the assumption of independent and identically exponentially distributed lifetimes. An often more realistic assumption is to assume that the lifetimes Z are exponentially distributed with hazard rate λ varying from unit to unit according to individual differences, for example due to varying environmental conditions. In many cases these differences can be represented by an observable covariate (vector) $\mathbf{x}$, leading to a model where $\lambda = \lambda(\mathbf{x})$ is a function of $\mathbf{x}$. Estimating the hazard rate function $\lambda(\mathbf{x})$ from observed data is called exponential regression. If some of the observations are right censored we have censored exponential regression. In this case there is a (possibly random) censoring time C such that only $T = \min(Z, C)$ and $\delta = I(Z \le C)$ are observed. Thus in addition to the observed time T we know whether there was a failure ($\delta = 1$) or a censoring ($\delta = 0$). Assume that we have observations from n independent units, where in addition to T, δ we observe the corresponding covariate vector $\mathbf{X}$. This gives the data $(T_1, \delta_1, \mathbf{X}_1), \ldots, (T_n, \delta_n, \mathbf{X}_n)$.

As a simple example of the use in reliability studies, suppose that times Z between failures of some equipment are assumed to be exponentially distributed with hazard rate $\lambda(\mathbf{x})$ depending on covariates such as age, duration of previous repair, type of repair etc. Censoring could here be used to represent for example planned maintenance.

Lindqvist et al (1988) considered lifetimes Z of subsea valves for oil/gas production. A typical covariate vector would then be $\mathbf{x} = (x_1, x_2, x_3)$, where x_1 is the gas/oil ratio, x_2 is the CO_2- content and x_3 is the speed of oil through the valve. Censoring did here correspond to planned testing and maintenance operations (for example due to governmental regulations). A parametric model was used for the hazard rate $\lambda(\mathbf{x})$, namely $\lambda(\mathbf{x}) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$. This is an example of a generalized linear model and the parameters can be estimated by ordinary maximum likelihood using standard computer programs. A seemingly more realistic model would be to replace the linear terms $\beta_i x_i$ by general functions $g_i(x_i)$ and then estimate the functions g_i by some nonparametric technique like spline smoothing, kernel smoothing or local polynomial smoothing. The purpose of this paper is to present a new nonparametric method for estimation in such exponential regression models. For the case of a one-dimensional covariate, the method is based on arranging the observations according to increasing order of the covariate, and using kernel estimation on a specific point process defined by the corresponding cumulative observation times. The method is then generalized to models with vector valued covariates.

2. Local Likelihood Methods

Before presenting the covariate order method we briefly review some well known competing methods. Let the model and data be as outlined in the previous section, but assume now that X is a one-dimensional covariate. Thus, for given $X = x$, Z is assumed to be exponentially distributed with an unknown hazard rate $\lambda(x)$, that is $f_Z(t|x) = \lambda(x)\exp(-\lambda(x)t)$. Further, C is distributed according to some unknown censoring distribution $f_C(t|x)$ which may depend on x, and C is assumed to be independent of Z given X.

Suppose first that $\lambda(x) \equiv \lambda$, i.e. the hazard rate does not depend on the covariate. Then the likelihood of the data is given by $\prod_{i=1}^n \lambda^{\delta_i} \exp(-\lambda T_i)$ so the log-likelihood is

$$\sum_{i=1}^{n} (\delta_i \log \lambda - \lambda T_i) \qquad (1)$$

This is maximized by $\hat{\lambda} = \sum \delta_i / \sum T_i$, which is the usual "occurrence/exposure ratio" estimator for λ.

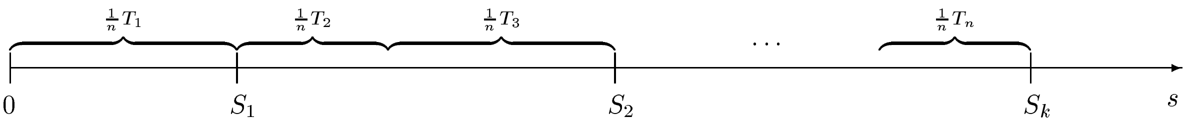

Figure 1: Construction of covariate order process.

If it is known that $\lambda(x)$ varies slowly with x, then the above estimator suggests the following local estimator at a point x_0: $\hat\lambda(x_0) = \sum_{i\in N(x_0)} \delta_i / \sum_{i\in N(x_0)} T_i$, where $N(x_0)$ is a set of indices i for which x_i is sufficiently near x_0 (where we have a choice regarding membership requirements for $N(x_0)$). A formal way of producing such estimators is through a local likelihood approach, where we estimate $\lambda(x_0)$ at a fixed x_0 by maximizing

$$\sum_{i=1}^{n} (\delta_i \log \lambda(x_0) - \lambda(x_0)T_i) \frac{1}{h} K\left(\frac{x_0 - x_i}{h}\right) \tag{2}$$

where $K(\cdot)$ a positive kernel function with integral 1, usually a unimodal density function, symmetric about 0 (e.g. the standard normal density). Note that the local likelihood (2) is similar to (1), but that the ith term is weighted according to the distance between x_i and x_0. Maximizing (2) leads to the estimator $\hat\lambda(x_0) = \sum w_i\delta_i / \sum w_i T_i$ where $w_i = K((x_0 - x_i)/h)$. The value of h determines the smoothness of the estimator and is chosen by for example cross validation (see Section 3).

Finally we shall briefly mention local polynomial exponential regression, where in (2) we replace $\lambda(x_0)$ by a local polynomial $\beta_0 + \beta_1(x_i - x_0) + \beta_2(x_i - x_0)^2 + \cdots$, then estimate the parameters β_i by maximizing (2) and finally put $\hat\lambda(x_0) = \hat\beta_0$.

3. The Covariate Order Method

Let the situation be as in the previous section. The covariate order method consists in first arranging the observations $(T_1, \delta_1, X_1), \ldots, (T_n, \delta_n, X_n)$ such that $X_1 \le X_2 \le \cdots \le X_n$. Next, for convenience, we divide the observation times by n, the total number of observations. Then let the scaled observation times $T_1/n, \ldots, T_n/n$, irrespectively if they are censored or not, be subsequent intervals on an artificial time axis s. Let points which are endpoints of an interval corresponding to an uncensored observation ($\delta = 1$) be considered as events, occurring at times denoted $S_1, S_2, \ldots$. This is visualized in Figure 1, for an example where the ordered observations are $(T_1, \delta_1 = 1), (T_2, \delta_2 = 0), (T_3, \delta_3 = 1), \ldots, (T_{n-1}, \delta_{n-1} = 0), (T_n, \delta_n = 1)$.

Formally $S_i = \sum_{j=1}^{k(i)} T_j/n$ where $k(i) = \min\{r \mid \sum_{j=1}^{r} \delta_j = i\}$. Note that if $\lambda(x) \equiv \lambda$ is constant in x, then the process $S_1, S_2, \ldots$ is a homogeneous Poisson process. Indeed, given $X_1, \ldots, X_n$ we can regard the process on the s-axis as putting into operation sequentially n units, with i.i.d. exponential lifetimes, and where replacement takes place at either failure or censoring of the unit. Then at each point in time there is a constant hazard λ of experiencing a failure.

If $\lambda(x)$ varies reasonably smoothly and not too much with x, then intuitively the process $S_1, S_2, \ldots$ could be regarded as nearly a nonhomogeneous Poisson process, for which the intensity function can be estimated by ordinary kernel density estimation based on the points $S_1, S_2, \ldots$. An estimate of $\lambda(x)$ then results by transforming this kernel estimate in an appropriate way to be described next.

Define $N(s)$ for $s > 0$ to be the number of events S_i on the s-axis in the interval $(0, s]$. Then the conditional intensity of the process $S_1, S_2, \ldots$ at a point s is $n\lambda(X_{N(s)+1})$. This is because at time s the unit "in operation" is the one numbered $N(s) + 1$. The basic idea of the covariate order method is to estimate this intensity from the process $S_1, S_2, \ldots$ by for example a kernel estimator, as motivated above, yielding the estimator $\hat\rho(s)$. Then finally we invert the relation $n\hat\lambda(X_{N(s)+1}) = \hat\rho_n(s)$ to obtain an estimate $\hat\lambda(x)$ at given points x. The key here is the relationship between the process $X_1, X_2, \ldots$ on the "covariate axis" and the process $S_1, S_2, \ldots$ on the s-axis. A possible way of estimating a general relationship between points on the two axes is by plotting points $(X_1, T_1/n), \ldots, (X_n, \sum_{i=1}^{n} T_i/n)$ and drawing straight lines between successive points, in other words using the function

$$\tilde{s}(x) = \frac{1}{n}\sum_{i=1}^{J-1} T_i + \frac{x - X_{J-1}}{X_J - X_{J-1}}\frac{1}{n}T_J, \quad X_{J-1} < x \leq X_J \tag{3}$$

We are now in position to write down an expression for the covariate order estimator of $\lambda(x)$. Let $K(\cdot)$ be a positive kernel as defined earlier, and let h_s be a smoothing parameter which is either constant or varying along the s-axis. Assume that $h_s \to 0$ as $n \to \infty$ for all s. Further assume that there is a sequence h_n such that $h_s \geq h_n$ for all s, n where $nh_n \to \infty$ as $n \to \infty$. Then the estimator

$$\hat{\lambda}(x) = \hat{\rho}(\tilde{s}(x)) \equiv \frac{1}{nh_s}\sum_{i=1}^{k} K\left(\frac{\tilde{s}(x) - S_i}{h_s}\right) \tag{4}$$

can be shown to be a uniformly consistent estimator of $\lambda(x)$. Here $k = \sum_{i=1}^{n}\delta_i$ is the number of uncensored observations.

In practice the estimated correspondence function (3) may be replaced by more sophisticated estimators, improving on the smoothness of the estimator (4).

To avoid the estimator $\hat{\lambda}(x)$ to be seriously downward biased near the endpoints special care must be taken at the boundaries. Viewed only as a problem on the s-axis the estimator (4) is simply density estimation on the s-axis, and techniques for handling boundary problems in density estimation can be adopted. A common technique is to reflect the data points around both endpoints, see for example Silverman (1986).

In order to choose the "best" value of the smoothing parameter h_s, a likelihood cross-validation criterion has been used, see for example Simonoff (1996).

4. Several Covariates

The mentioned methods for nonparametric exponential regression can, at least in principle, be directly generalized to multidimensional covariates with, say, m components. For the local methods the generalization is in principle immediate, by using m-dimensional kernels. For the covariate order method it is less clear, however, how to generalize to multidimensional $\mathbf{x}$.

In any case direct generalizations will soon be beaten by the so called "curse of dimensionality". A better approach in many cases is to reduce the dimension of the problem by assuming some structure on the covariate space. Such an approach is also more appropriate for generalizing the covariate order method.

One way to proceed is to assume that the hazard rate can be written in the form of the generalized additive model

$$\lambda(\mathbf{x}) = \exp(\alpha + g_1(x_1)+, \ldots, +g_m(x_m)), \tag{5}$$

where $\mathbf{x} = (x_1, \ldots, x_m)$, and where $g_1(\cdot), \ldots, g_m(\cdot)$ are unspecified smooth functions. These functions can be estimated iteratively by using the one dimensional covariate order method repeatedly. The algorithm below is similar to the backfitting algorithm of Hastie and Tibshirani (1990).

Let $\hat{g}_i^{(d)}(x_i)$ be the estimate of $g_i(x_i)$ after the dth iteration of the algorithm, and define $\hat{g}_i^{(0)}(x_i) = 0$. Then the algorithm can be written as follows:

1. Repeat 2-3 for $d = 1, \ldots, D$ where $|\hat{\alpha}^{(D-1)} - \hat{\alpha}^{(D)}|/\hat{\alpha}^{(D-1)} < \epsilon$.
2. Calculate $\hat{\alpha}^{(d)} = \log\left(\sum_{i=1}^{n}\delta_i / \sum_{i=1}^{n} T_i \exp(\hat{g}_1^{(d-1)}(X_{i1}) + \ldots + \hat{g}_m^{(d-1)}(X_{im}))\right)$.
3. Repeat 4-5 for $j = 1, \ldots, m$.
4. Transform $T_1, \ldots, T_n$ to $L_{1j}, \ldots, L_{nj}$ where
$L_{ij} = T_i \exp(\hat{\alpha}^{(d)} + \hat{g}_1^{(d)}(X_{i1}) + \cdots + \hat{g}_{j-1}^{(d)}(X_{i,j-1}) + \hat{g}_{j+1}^{(d-1)}(X_{i,j+1}) + \cdots + \hat{g}_m^{(d-1)}(X_{im}))$, $i = 1, \ldots, n$.
5. Calculate the estimate $\tilde{g}_j^{(d)}(x_j)$ from $(L_{1j}, \delta_1, X_{1j}), \ldots, (L_{nj}, \delta_n, X_{nj})$ using the covariate order method, and let $\hat{g}_j^{(d)}(x_j) = \tilde{g}_j^{(d)}(x_j) - \sum_{i=1}^{n}\tilde{g}_j^{(d)}(X_{ij})/n$.

The motivation for the algorithm is the observation that if Z is exponentially distributed with parameter $\exp(\alpha + g_1(x_1) + \ldots + g_m(x_m))$, then $Z\exp(\alpha + g_1(x_1) + \ldots + g_{j-1}(x_{j-1}) + g_{j+1}(x_{j+1}) + \ldots + g_m(x_m))$ is exponentially distributed with parameter $\exp(g_j(x_j))$. This explains step 4 and 5 in the algorithm. In step 5 all the estimated functions are normalized to have mean 0 to avoid possible problems caused by constant factors jumping back and forth between the estimated functions. Finally the estimate $\hat{\alpha}^{(d)}$ in step 2 is the maximum likelihood estimate based on the sample $(L_1, \delta_1), \ldots, (L_n, \delta_n)$ where $L_i = T_i \exp(\hat{g}_1^{(d-1)}(X_{i1}) + \ldots + \hat{g}_m^{(d-1)}(X_{im}))$

Notice that the algorithm can easily be extended to allow some of the g-functions to be parametric (for example for discrete covariates) and then estimate the parameters of these functions in step 2 of the algorithm using maximum likelihood.

5. Nonparametric Tests for Covariate Effect

It was noted in Section 3 that if $\lambda(x) \equiv \lambda$, i.e. there is no covariate effect, then the process $S_1, S_2, \ldots$ is a homogeneous Poisson process (HPP). This observation suggests that statistical tests for the null hypothesis of an HPP versus various non-HPP alternatives can be applied as tests for covariate effect in this situation. The further presumption that the process $S_1, S_2, \ldots$ can be considered as being approximately an NHPP when $\lambda(x)$ depends on x, suggests adopting exisitng tests for HPP versus NHPP, see for instance Ascher and Feingold (1984) and references therein. A detailed presentation of this approach for testing covariate effects is given in Kvaløy (1999b).

6. Concluding Remarks

A comprehensive simulation study (reported in Kvaløy and Lindqvist, 1998) shows that the covariate order method is generally competitive when compared to the local methods reviewed in Section 2. In particular the covariate order method seems to be doing better than the other methods in "difficult" cases with few observations or large degree of censoring. Also, the covariate order method has a large flexibility due to the possibility of utilizing any method for density estimation.

It is finally noted that the covariate order method can be adapted for use in the Cox-regression setting. The idea here is that Cox-regression can be reduced to exponential regression by transforming the lifetimes by the estimated cumulative baseline hazard function. Thus our method is also a method for nonparametric estimation in Cox-models. This has been further explored by Kvaløy (1999a).

References

1. Ascher, H. and Feingold, H. (1984). *Repairable Systems Reliability.* New York: Marcel Dekker.

2. Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* London: Chapman & Hall.

3. Kvaløy, J. T. (1999a). Nonparametric estimation in Cox-models: Time transformation versus partial likelihood methods. Statistics No. 3/99, Department of Mathematical Sciences, Norwegian University of Science and Technology.

4. Kvaløy, J. T. (1999b). Covariate order tests for covariate effect. Statistics No. 7/99, Department of Mathematical Sciences, Norwegian University of Science and Technology.

5. Kvaløy, J. T. and Lindqvist, B. H. (1998). The covariate order method for censored exponential regression. Statistics No. 10/98, Department of Mathematical Sciences, Norwegian University of Science and Technology.

6. Lindqvist, B., Molnes, E. and Rausand, M. (1988). Analysis of SCSSV performance data. *Reliability Engineering and System Safety*, **20**, 3-17.

7. Silverman, B. W. (1986). *Density Estimation.* London: Chapman & Hall.

8. Simonoff, J. S. (1996). *Smoothing Methods in Statistics.* New York: Springer Verlag.