A Constrained Quadratic Spline as a Model for the Cumulative Hazard Function

V.V. Krivtsov^{*}, I.V. Kolmanovsky, T.P. Davis Ford Motor Co., MD 606, 14555 Rotunda Drive, Dearborn, MI 48120

Abstract

Probability distributions traditionally used in reliability analysis (e.g., exponential, Weibull, lognormal, normal, gamma and inverse Gaussian) do not always provide enough flexibility to model real world lifetime data. Considered in this paper is a quadratic spline with a single free knot as a model of the cumulative hazard function for increasing failure rate distributions. A parsimonious (essentially closed form) solution with a simple numerical search is proposed for estimating spline parameters. Automotive data examples are discussed as an illustration.

Keywords: cumulative hazard function, spline models.

Introduction

Most parametric modeling work in lifetime data analysis relates to (log-)location-scale distributions and some other suitable distributions such as gamma and inverse Gaussian. Despite (or, arguably, *because*) of their parsimonious structure, these distributions do not provide enough flexibility in some real world data scenarios. Models, in which the hazard function (or some transform of it) is a low-order polynomial, may provide a more reasonable fit to data in these situations. Notable particular cases of the polynomial hazard models include the Rayleigh distribution with the linear hazard function and the Gompertz distribution with the log-linear hazard function.

An alternative way to achieve flexibility in modeling hazard functions is to use so-called *spline functions*, which are comprised of polynomial pieces connected at special points referred to as *knots*. Often, spline parameters need to be constrained in order to keep the hazard function non-negative. Rosenberg (1995) considers various spline models for hazard functions and Kooperberg et al. (1995) for

^{*}Corresponding author. Phone/Fax: 313-323-8711.

E-mail: VKrivtso@ford.com (Vasiliy Krivtsov)

the logarithm of the hazard function. Lawless (2003) discusses applications of cubic splines for a wide range of lifetime data.

Working with automotive warranty data, we noticed that a spline of the form

$$h(t) = \begin{cases} 2b_1 t, & 0 \le t \le k \\ 2b_1 t + 2b_2 (t-k), t > k \end{cases}$$

with b_1 , b_2 , k > 0 fits quite well for many data sets exhibiting a linearly increasing hazard function. As an example, in Figure 1, consider the cumulative hazard function for an automotive chassis component data. It is evident from the figure that the proposed spline provides a better fit to the data than some common probability distributions estimated via a non-linear least square method.



Figure 1. Nonparametric estimate of cumulative hazard function versus normal, lognormal, Weibull and the quadratic spline fits.

For estimation convenience, we will proceed with the respective *cumulative* hazard function, which takes the form of a quadratic spline with a single free knot:

$$H(t) = \begin{cases} b_1 t^2, & 0 \le t \le k \\ b_1 t^2 + b_2 (t-k)^2, & t > k \end{cases}$$

where b_1 , b_2 and k > 0 need to be estimated to best fit the nonparametric estimates of H(t) available in the form of $(t_i; y_i)$, i = 1, ..., N; $0 \le t_1 < t_2 < ... < t_N$. Considered in this paper is a restricted least-

squares procedure limited to point estimation only. (Due to considerably large sample sizes of field data, the respective interval estimation is of a lesser interest.)

The rest of the paper is structured as follows. First, the objective function that corresponds to the sum of squared residuals is analyzed and decomposed into a sum of fixed cost (which is not affected by the values of b_1 , b_2 and k) and variable cost (which is affected by the values of b_1 , b_2 and k). Then, we analyze the monotonicity and convexity constraints and translate them into conditions on b_1 , b_2 and k. The optimization is reduced to a two-stage process, where in the inner stage b_1 and b_2 are optimized using Lagrange multiplier rule (the closed form solution to the inner stage problem is given in a separate section) and in the outer stage the value of k is optimized via one dimensional grid search. In the final section, we consider practical examples illustrating applications examples of the proposed model.

Objective Function

Let $n(k)=\max\{i:t_i \le k\}$. In the least-squares fit, our objective is to minimize the sum of squares:

$$f = \sum_{i=1}^{n(k)} (b_1 t_i^2 - y_i)^2 + \sum_{i=n(k)+1}^{N} (b_1 t_i^2 + b_2 (t_i - k)^2 - y_i)^2 \to \min_{i=n(k)+1} (b_1 t_i^2 - y_i)^2 \to \max_{i=n(k)+1} (b_1 t_i$$

After algebraic transformations we obtain,

$$f = f_f + f_v ,$$

where

$$f_f = \sum_{i=1}^N y_i^2$$
 ,

is the fixed cost (independent of k, b_1 , b_2) and

$$f_{v} = \beta_{1}b_{2}^{2} + \beta_{2}b_{1}b_{2} + \beta_{3}b_{1} + \beta_{4}b_{1} + \beta_{5}b_{2}$$
$$\beta_{1} = \sum_{i=1}^{N} t_{i}^{4} ,$$
$$\beta_{2} = \sum_{i=n(k)+1}^{N} (t_{i} - k)^{4} ,$$
$$\beta_{3} = 2\sum_{i=n(k)+1}^{N} t_{i}^{2} (t_{i} - k)^{2} ,$$
$$\beta_{4} = -2\sum_{i=1}^{N} y_{i}t_{i}^{2} ,$$

$$\beta_5 = -2 \sum_{i=n(k)+1}^N y_i (t_i - k)^2$$
,

is the variable cost.

In the sequel, we analyze the problem of minimizing f_v with respect to b_1 and b_2 for a fixed k, subject to the monotonicity constraints discussed in the following section. This optimization problem can be solved essentially in closed form, resulting in the optimal cost $f_{min}(k)$, see Section 4. The function f_{min} is then minimized with respect to k via one-dimensional grid search to find an optimal value of k. The grid search can be replaced by a more efficient, interval sub-division algorithm (Fibonacci division or Golden Section), but this assumes that f_{min} is unimodal, which may not always be the case.

Constraints

The constraint that H(t) be monotonically non-decreasing in interval [0, k_u] (where k_u may be $+\infty$) translates into non-negativity condition for the derivative:

$$2b_1 t \ge 0$$
; if $t \le k$,
 $2b_1 t + 2b_2 (t-k) \ge 0$, if $t > k$

Since the derivative is a piecewise linear function in t, it is sufficient to enforce the constraint at t=0, k, k_{μ} that results in

$$b_1 \ge 0, b_1 k_u + b_2 (k_u - k) \ge 0,$$

If the monotonicity is desired on an infinite time interval $[0, +\infty]$ (i.e., $k_u \rightarrow +\infty$), then the constraints become

$$b_1 \ge 0, b_1 + b_2 \ge 0$$

To summarize both of these cases, the constraints can be written in the form:

$$b_1 \ge 0, \ \alpha_1 b_1 + \alpha_2 b_2 \ge 0,$$

where $\alpha_1 = k_u$, $\alpha_2 = (k_u - k)$ if $k_u < +\infty$, $k_u > k > 0$; and $\alpha_1 = 1$, $\alpha_2 = 1$ if $k_u = +\infty$.

The constraint that H(t) be convex translates into a requirement that the second derivative be nonnegative, i.e., $b_1 \ge 0$, $b_1 + b_2 \ge 0$. This is the same as the monotonicity requirement on the infinite time interval and thus can be enforced by setting $k_u = +\infty$.

Lagrange Multiplier Rule Based Optimization of *b*₁ and *b*₂ (Inner Stage)

The problem of minimizing f_v with respect to b_1 and b_2 for a fixed k subject to the constraints $b_1 \ge 0$,

 $\alpha_1 b_1 + \alpha_2 b_2 \ge 0$ is a constrained optimization problem (in fact, it is a quadratic programming problem). To solve it we use the Lagrange multiplier rule. The Lagrangian function has the form

$$L = \lambda_0 f_v + \lambda_1 (-b_1) + \lambda_2 (-\alpha_1 b_1 - \alpha_2 b_2),$$

where λ_0 , λ_1 , λ_2 are Lagrange multipliers. The optimal b_1^* , b_2^* must satisfy the following conditions for *some* $\lambda_0 \in \{0, 1\}, \lambda_1 \ge 0, \lambda_2 \ge 0, \lambda_0 + \lambda_1 + \lambda_2 > 0$,

$$\frac{\partial L}{\partial b_1}|_{b_1=b_1^*,b_2=b_2^*}=0, \ \frac{\partial L}{\partial b_2}|_{b_1=b_1^*,b_2=b_2^*}=0,$$

and

$$\lambda_1(-b_1^*) = 0, \lambda_2(\alpha_1b_1^* + \alpha_2b_2^*) = 0,$$

Note that

$$\frac{\partial L}{\partial b_1} = \lambda_0 (2\beta_1 b_1 + \beta_3 b_2 + \beta_4) - \lambda_1 - \alpha_1 \lambda_2,$$
$$\frac{\partial L}{\partial b_2} = \lambda_0 (2\beta_{2b_2} + \beta_3 b_1 + \beta_5) - \alpha_2 \lambda_2.$$

In the sequel, we analyze in more detail different cases that these conditions generate. Each of the cases results in a potential candidate solution; the final solution is selected from these candidate solutions. It is the one that minimizes f_v while satisfying the constraints $b_1 \ge 0$, $\alpha_1 b_1 + \alpha_2 b_2 \ge 0$.

Case 1: $\lambda_0 = 1$; $\lambda_1 > 0$; $\lambda_2 > 0$

In this case,

$$b_1^* = 0, b_2^* = 0$$

is the candidate solution.

Case 2: $\lambda_0 = 1$; $\lambda_1 = 0$; $\lambda_2 > 0$

In this case, $\alpha_1 b_1^* + \alpha_2 b_2^* = 0$. Since $\alpha_1 > 0$, $b_1^* = -\frac{\alpha_2 b_2^*}{\alpha_1}$. From $\frac{\partial L}{\partial b_i} = 0$, i = 1, 2, and $\lambda_1 = 0$, we obtain a

system of two linear equations in b_2^* , λ_2 :

$$(2\beta_1(-\frac{\alpha_2}{\alpha_1})+\beta_3)b_2^*-\alpha_1\lambda_2=-\beta_4,$$
$$(2\beta_2(-\frac{\alpha_2}{\alpha_1})-\beta_3\frac{\alpha_2}{\alpha_1})b_2^*-\alpha_2\lambda_2=-\beta_5.$$

These linear equations are solved for b_2^* , λ_2 and then $b_1^* = -\frac{\alpha_2 b_2^*}{\alpha_1}$.

When solving this system of linear equations, it may turn out (as a rare possibility) that the determinant is zero (i.e., the system is singular). Although the analysis can be continued, in the actual code, we simply discard the candidate solution for this case.

Case 3: $\lambda_0 = 1$, $\lambda_1 > 0$, $\lambda_2 = 0$

In this case, $b_1^* = 0$ and the two linear equations

$$\beta_3 b_2^* - \lambda_1 = -\beta_4,$$

$$2\beta_2 b_2^* + 0\lambda_1 = -\beta_5,$$

are solved for b_2^* , λ_0 . Clearly, $b_1^* = 0$, $b_2^* = -\frac{\beta_5}{2\beta_2}$ is the candidate solution.

Case 4: $\lambda_0 = 1$, $\lambda_1 = 0$, $\lambda_2 = 0$

In this case, the following two linear equations are solved for b_1^* , b_2^*

$$2\beta_{1}b_{1}^{*} + \beta_{3}b_{2}^{*} = -\beta_{4},$$

$$2\beta_{2}b_{2}^{*} + \beta_{3}b_{1}^{*} = -\beta_{5},$$

to obtain a candidate solution.

Case 5: $\lambda_0 = 0$, $\lambda_1 > 0$, $\lambda_2 > 0$

In this case, $b_1^* = 0$, $\alpha_1 b_1^* + \alpha_2 b_2^* = 0$ implying $b_2^* = 0$ form the candidate solution.

Case 6: $\lambda_0 = 0$, $\lambda_1 = 0$, $\lambda_2 > 0$

In this case $b_1^* = -\frac{\alpha_2 b_2^*}{\alpha_1}$ and $\frac{\partial L}{\partial b_1} = -\lambda_1 - \alpha_1 \lambda_2 = 0$. Hence, $\lambda_2 = 0$ which is not possible, since

 $\lambda_0 + \lambda_1 + \lambda_2 > 0$. This case does not provide candidate solutions.

Case 7: $\lambda_0 = 0$, $\lambda_1 > 0$, $\lambda_2 = 0$

In this case, $\frac{\partial L}{\partial b_1} = 0$ and, hence, $\lambda_1 = 0$, which is not possible. This case does not provide candidate

solutions.

Case 8: $\lambda_0 = 0$, $\lambda_1 = 0$, $\lambda_2 = 0$

This case is not possible, since $\lambda_0 + \lambda_1 + \lambda_2 > 0$.

Practical Examples

We would like to first consider a fleet battery data set (Davis, 2003). This example pertains to the ability of the battery to have enough power to turn over the starter motor and start the engine. Note that

time in service is used as the life variable. This is considered as a surrogate for engine starting cycles, which might have been more appropriate, but the investigating engineer did not know the number of times each vehicle had been started. Time in service was considered a more suitable surrogate than, perhaps, mileage.

The cumulative hazard function estimated from data collected from a large fleet of captive vehicles is illustrated in Figure 2. The quadratic spline parameter estimates are: $\hat{b}_1 = 1.51 \cdot 10^{-4}$, $\hat{b}_2 = 2.88 \cdot 10^{-5}$, $\hat{k} = 13.97$. Weibull distribution also provides a reasonable approximation for this data set, although, with a larger value of the cost function (the sum of residual squares) than the quadratic spline.



Figure 2. A fleet battery example.

Another example we would like to share involves an axle seal component, for which mileage is believed to be a better life variable for obvious engineering reasons. The data are obtained from a warranty database system. The estimated quadratic spline ($\hat{b}_1 = 5.65 \cdot 10^{-12}$, $\hat{b}_2 = 1.43 \cdot 10^{-11}$,

 $\hat{k} = 20,468$) shows a better approximation to the actual data than some conventionally used probability distributions.



Figure 3. An axle seal example.

References

1. Davis, T.P. (2003). "Reliability improvement in automotive engineering." In *Global Vehicle Reliability - Prediction and Optimization Techniques*, edited by JE Strutt and PL Hall. Professional Engineering Publications, IMechE, London.

2. Kooperberg, C.; Stone, C.; Truong, Y. (1995). "Hazard regression." *Journal of American Statistical Association*, 90, pp.78-94.

3. Lawless, J.F. (2003). Statistical models and methods for lifetime data, Wiley, New Jersey.

4. Rosenberg, P.S. (1995). "Hazard function estimation using B-splines." *Biometrics*, 51, pp. 874-887.