Estimating the Mixture of Proportional Hazards Model with the Constant Baseline Hazards Function

Jong-woon Kim, Seong-phil Eo

R&D Center, Rotem Company
Gyunggi-Do, 449-910, KOREA

Abstract

Cox's proportional hazards model (PHM) has been widely applied in the analysis of lifetime data, and it can be characterized by the baseline hazard function and covariates influencing systems' lifetime, where the covariates describe operating environments (e.g. temperature, pressure, humidity). In this article, we consider the constant baseline hazard function and a discrete random variable of a covariate. The estimation procedure is developed in a parametric framework when there are not only complete data but also incomplete one. The Expectation-Maximization (EM) algorithm is employed to handle the incomplete data problem. Simulation results are presented to illustrate the accuracy and some properties of the estimation results.

1. Introduction

Notation

s :a covariate
s_k :the kth element of a covariate
g :the number of elements of s
n :the number of uncategorized field units
c_i :the number of categorized field observations whose covariates are s_i
c_{sum} : $\sum_{i=1}^{g} c_i$
d_i :the number of experiment observations whose covariates are s_i.
m_i : $c_i + d_i \ (i = 1,\ldots,g)$
m_{sum} : $\sum_{i=1}^{g} m_i$
x_j :the failure time of the jth uncategorized field observation.
y_{ij} :the failure time of the jth observation among the categorized field observations whose covariates are s_i.
z_{ij} :the failure time of the jth experiment observation among the experiment observations whose covariates are s_i.
w_{ij} : $w_{ij} = \begin{cases} y_{ij}, & 1 \le j \le c_i \\ z_{ij}, & c_i < j \le m_i \end{cases}$
θ :a vector of lifetime distribution parameters.
η :a vector of a regression parameter
ϕ : (η, θ)
ψ : (p, η, θ)

In this article, a mixed proportional hazards model (mixed PHM) is derived and a procedure is proposed to estimate all parameters in the mixed PHM with the constant baseline hazards function, which is a mixture model of the proportional hazards model (PHM). The PHM has been considered as a useful tool to deal with the environmental factors in the analysis of lifetime data. Solomon (1984) indicated that significant effects for covariates would be obtained even in the cases where the model was not wholly appropriate, and showed that the PHM is relatively robust to departures from the proportional hazards assumption. The application of PHM to reliability data has been considered by a number of authors, for example, Ansell and Phillips (1997) and Jozwiak (1997). For a list of more recent papers, see the review paper by Kumar and Klefsjö (1994).

The failure nature of an item can be modeled by the hazard rate. The assumption imposed for the PHM, in most cases, is that the hazard rate of a system is the product of a baseline (time-dependent) hazard rate $\lambda_0(t)$ and a positive (time-independent) functional term $\omega(s,\eta)$, incorporating the effects of a number of covariates such as temperature, pressure, and changes in design. That is,

$$\lambda(t|s) = \omega(s,\eta)\lambda_0(t) \tag{1}$$

where s is a row vector consisting of the covariates and η is a column vector consisting of the regression parameters. In general, there are two ways to model the baseline hazard rate $\lambda_0(t)$: by the parametric model and by the non-parametric model.

In the parametric model, we assume a suitable theoretical function for $\lambda_0(t)$. On the other hand, in the non-parametric model, no specific distribution is assumed. Note that the non-parametric method cannot always guarantee an accurate estimation because of the lack of knowledge about the lifetime distribution. In this article, the constant function is used for $\lambda_0(t)$ and the exponential form, $\exp(s\eta)$ for $\omega(s,\eta)$.

Covariates are associated with the equipment's environmental and operational conditions. η is the effect of the covariates. The equipment's environmental and operational conditions vary even within the same field. For example, in Martorell, Sanchez and Serradell (1999), it was reported that equipments at nuclear power plants worked under very different operational and environmental conditions; For instance, some components were placed in a very stressful environment under high temperature and doses of radiation, while others remained in a stress-free environment. In addition, generally, the conditions under which a product is used cannot always be known before it is installed. We assume that the covariates are random variables for these variability and uncertainty of the covariates. When the covariates in PHM are identically and independently distributed (i.i.d.) discrete random variables and the baseline hazards function is a constant one, the failure model is reduced to the following mixture form of PHM:

$$f(t)=\sum_{i=1}^{g} p_i \lambda e^{s_i \eta} \exp\left(-\lambda e^{s_i \eta} t\right) \tag{2}$$

It is assumed in this article that the support of the random variable, s, is known.

The main purpose of this chapter is to estimate lifetime distributions of the products whose failures can be modeled by the mixture of PHM of (2).

2. Literature Review

The mixed PHM is a kind of mixture model. The extensive applicability of the mixture model has generated much research on them. The existing results were classified and introduced by Titterington, et al (1985), Everitt and Hann (1981), and McLachlan and Basford (1987). The finite mixed exponential distribution and the finite mixed Weibull distribution are good candidates to model failure time. McClean (1986) considered the mixed exponential distribution with grouped follow-up data, when the number of components is known. Lau (1998) estimated hazard rates using both the mixture of geometrics and the mixture of exponential models. Jiang and Kececioglu (1992a) and Jiang and Murthy (1995) used graphical approaches and Jiang and Kececioglu (1992b) used the method of maximum likelihood for estimation problem in the mixed Weibull distributions with censored data. Jaisingh et al (1993) considered the influence of the work environment by using a Weibull & inverse Gaussian mixture. Hirose (1997) dealt with the power-law mixture model which extends the power-law model in accelerated life testing. Sy and Taylor (2000) and Peng and Dear (2000) involved the mixture models in PHM to estimate cure rates. They assumed no specific distribution for the baseline hazard function and use nonparametric two mixture models.

Kim, Yun & Dohi (2003) considered the Weibull lifetime distribution in the mixed PHM with two types of incomplete data.

3. Estimation Method

In this section we introduce the maximum likelihood method for estimating parameters of the mixed PHM. Not only is it appealing on intuition grounds, but it also possesses desirable statistical properties. For example, under very general conditions, the estimators obtained by the method are consistent and they are asymptotically normally distributed. The starting point for any particular investigation is a consideration of the form in which data are obtained.

In most applications, the data take the form of a random sample of observations where the distribution of each observation is described by a mixed PHM of the form of (2) – we will call them uncategorized field observations. In addition to the random sample from the mixed PHM there may also be random samples available of observations of which underlying categories are known. We will call them categorized experiment observations because experimental conditions are predetermined before conducting experiments and so the covariate of an experimental unit is known before testing. Moreover, if the categorized observations can be assumed to arise independently, with incidence rates $p_1,\ldots,p_k$ for individual categories, then this provides further information about the mixing weights – we will call them categorized field observations.

We use maximum likelihood techniques and the Expectation-Maximization (EM) algorithm to estimate distribution parameters, mixing proportions and a regression parameter with uncategorized field observations, categorized field observations and categorized experiment observations.

We consider the constant baseline hazard function. Then, the likelihood and log-likelihood functions are, respectively:

$$\begin{aligned} L(\psi)=&\prod_{i=1}^{n}\sum_{j=1}^{g} p_j \lambda e^{s_j \eta} \exp\left(-\lambda e^{s_j \eta} x_i\right) \\ &\times\prod_{i=1}^{g}\prod_{j=1}^{c_i} p_i \lambda e^{s_j \eta} \exp\left(-\lambda e^{s_j \eta} y_{ij}\right) \\ &\times\prod_{i=1}^{g}\prod_{j=1}^{d_i} \lambda e^{s_j \eta} \exp\left(-\lambda e^{s_j \eta} z_{ij}\right) \end{aligned} \tag{3}$$

$$\begin{aligned} \log L(\psi)=&\sum_{i=1}^{n}\log\left\{\sum_{j=1}^{g} p_j \lambda e^{s_j \eta} \exp\left(-\lambda e^{s_j \eta} x_i\right)\right\} \\ &+\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left(\log\lambda + s_i\eta - \lambda e^{s_i \eta} w_{ij}\right)+\sum_{i=1}^{g} c_i \log p_i \end{aligned} \tag{4}$$

The problem is to obtain the estimates of $\hat{\psi}$ which maximize $L(\psi)$. However, it is not easy to find the MLEs in the traditional way of differentiating L with respect to ψ and setting it equal to zero, because the likelihood function often becomes a complex multi-modal function. An alternative way is to apply an iterative algorithm such as the EM algorithm.

The EM algorithm is a broadly applicable approach to the iterative computation of maximum likelihood estimates, useful in a variety of incomplete-data problems. The EM algorithm finds estimate by iteratively performing two steps: the expectation step (E-step), and the maximization step (M-step). In the E-step, we calculate the conditional expectation of the log-likelihood function for complete data. In the M-step, we search parameter values maximizing the conditional expectation. The EM algorithm can be applied to the mixed PHM by augmenting the observed data with the unobserved indicator variables to represent the values of the covariates of field units. That is, in order to pose this problem as an incomplete-data one, we now introduce as the unobservable or missing data, vector:

$$v = \left(v_1^{\mathbf{T}}, \ldots, v_n^{\mathbf{T}}\right)^T \tag{5}$$

where v_j is a g-dimensional vector of zero-one indicator variables and where v_{ij} is one or zero according to whether the covariate for x_j is s_i or not, and v_j^T is the transpose of v_j. Then the log-likelihood for the complete data is given by:

$$\log L_C(\psi) = \sum_{i=1}^{g}\sum_{j=1}^{n} v_{ij}\left(\log p_i + \log\lambda + s_i\eta - \lambda e^{s_i\eta}x_j\right) + \sum_{i=1}^{g}\sum_{j=1}^{m_i}\left(\log\lambda + s_i\eta - \lambda e^{s_i\eta}w_{ij}\right) + \sum_{i=1}^{g} c_i \log p_i \tag{6}$$

The w-th E-step requires the calculation of the expectation of the complete data log-likelihood, $\log L_C(\psi)$, conditional on the observed data and the current fit $\psi^{(w-1)}$ for ψ.

$$\begin{aligned} Q\left(\psi;\psi^{(w-1)}\right) &= E\left\{\log L_C(\psi)\middle| x, y, \psi^{(w-1)}\right\} \\ &= \sum_{i=1}^{g}\sum_{j=1}^{n} E\left(v_{ij}\middle| x_j, \psi^{(w-1)}\right)\left(\log p_i + \log\lambda + s_i\eta - \lambda e^{s_i\eta}x_j\right) \\ &+ \sum_{i=1}^{g}\sum_{j=1}^{m_i}\left(\log\lambda + s_i\eta - \lambda e^{s_i\eta}w_{ij}\right) + \sum_{i=1}^{g} c_i \log p_i \end{aligned} \tag{7}$$

This step is affected here simply by replacing each indicator variable v_{ij} by its expectation conditional on x_j which is given by:

$$E\left(v_{ij}\middle| x_j;\psi^{(w-1)}\right) = \tau_i\left(x_j;\psi^{(w-1)}\right) \tag{8}$$

In the E-step, we calculate $E\left(v_{ij}\middle| x_j;\psi^{(w-1)}\right)$ as:

$$\hat{v}_{ij} = \frac{p_i^{(w-1)}\lambda^{(w-1)}e^{s_i\eta^{(w-1)}}\exp\left(-\lambda^{(w-1)}e^{s_i\eta^{(w-1)}}x_j\right)}{\sum_{k=1}^{g} p_k^{(w-1)}\lambda^{(w-1)}e^{s_k\eta^{(w-1)}}\exp\left(-\lambda^{(w-1)}e^{s_k\eta^{(w-1)}}x_j\right)} \tag{9}$$

On the w-th M-step, the intent is to choose the new value of ψ, say $\psi^{(w)}$, that maximize $Q\left(\psi,\psi^{(w-1)}\right)$ which, from the E step, is equal here to $\log L_C(\psi)$ with each v_{ij} replaced by $\tau_i\left(x_j;\psi^{(w-1)}\right)$. One nice feature of the EM algorithm is that the solution to the M-step often exists in a closed form. However, we can't obtain the closed form of ψ in our cases and need some numerical search techniques.

In the M-step, we find the new values of ψ, say $\psi^{(w)}$, that maximize $Q\left(\psi,\psi^{(w-1)}\right)$. Differentiating the function Q with respect to λ and η in turn and equating to zero, we obtain the maximizing equations:

$$\frac{\partial Q}{\partial\lambda} = \frac{1}{\lambda}\left(n + m_{sum}\right) - \sum_{i=1}^{g}\sum_{j=1}^{n}\hat{\tau}_{ij}x_j e^{s_i\eta} - \sum_{i=1}^{g}\sum_{j=1}^{m_i} w_{ij}e^{s_i\eta} = 0 \tag{10}$$

$$\frac{\partial Q}{\partial\eta} = \sum_{i=1}^{g}\sum_{j=1}^{n}\hat{\tau}_{ij}\left(s_i - \lambda s_i x_j e^{s_i\eta}\right) + \sum_{i=1}^{g}\sum_{j=1}^{m_i}\left(s_i - \lambda s_i w_{ij}e^{s_i\eta}\right) = 0 \tag{11}$$

***Theorem* 1**: *For fixed* (p,η), *the function* Q *in* (7) *is concave with respect to* λ. *For fixed* (p,λ), *the function* Q *is concave with respect to* η.

Proof : The second order conditions for the parameters λ and η are derived as:

$$\frac{\partial^2 Q}{\partial\lambda^2} = -\frac{1}{\lambda^2}\left(n + m_{sum}\right) \tag{12}$$

$$\frac{\partial^2 Q}{\partial\eta^2} = -\sum_{i=1}^{g}\sum_{j=1}^{n}\hat{\tau}_{ij}\lambda s_i^2 x_j e^{s_i\eta} - \sum_{i=1}^{g}\sum_{j=1}^{m_i}\lambda s_i^2 w_{ij}e^{s_i\eta} \tag{13}$$

They are negative in λ and η, respectively. □

By eliminating λ from these two equations and simplifying them, we get:

$$\sum_{i=1}^{g}\sum_{j=1}^{n}\hat{\tau}_{ij}s_i+\sum_{i=1}^{g}\sum_{j=1}^{m_i}s_i = (n+c_{sum})\frac{\left(\sum_{i=1}^{g}\sum_{j=1}^{n}\hat{\tau}_{ij}s_i x_j e^{s_i\eta}+\sum_{i=1}^{g}\sum_{j=1}^{m_i}s_i w_{ij}e^{s_i\eta}\right)}{\left(\sum_{i=1}^{g}\sum_{j=1}^{n}\hat{\tau}_{ij}x_j e^{s_i\eta}+\sum_{i=1}^{g}\sum_{j=1}^{m_i}w_{ij}e^{s_i\eta}\right)} \quad (14)$$

$\eta^{(w)}$ can be obtained from (14) using a line search. Theorem 1 guarantees the accuracy and effectiveness of the line search technique.

Using $\eta^{(w)}$, the renewed parameter $\lambda^{(w)}$ is obtained from (15) as follows:

$$\lambda^{(w)}=\left(n+m_{sum}\right)\Big/\left(\sum_{i=1}^{g}\sum_{j=1}^{n}\hat{\tau}_{ij}x_j e^{s_i\eta^{(w)}}+\sum_{i=1}^{g}\sum_{j=1}^{m_i}w_{ij}e^{s_i\eta^{(w)}}\right) \quad (15)$$

$p_k^{(w)}$ can be calculated as (16) by the method similar to the generic mixture distributions.(Refer to Kim et al(2003))

$$\hat{p}_k=\left(\sum_{j=1}^{n}\hat{\tau}_{kj}+c_k\right)\Big/\left(n+c_{sum}\right) \quad for\ k=1,\ldots,g \quad (16)$$

Note that (14), (15), and (16) do not give the estimators explicitly; instead they must be solved using the general EM iterative procedure.

4. Experimental Results

Simulation experiments are carried out to investigate the accuracy of the estimation. The failure times of a unit in the kth group (component), are generated as follows:

$$t=-\log U/\lambda\exp(s_k\eta) \quad (17)$$

where U is a uniform (0,1) random variable.

The failure times of uncategorized field units are generated from (17) with the covariate generated from $P(s=s_k)=p_k$, (k = 1... g). The number of simulation runs is set to be 200 for each case. Mean squared error (*MSE*) is calculated for evaluating the accuracy.

4.1 Analysis with the Only Uncategorized Field Observations

The effect of the number of the uncategorized field observations is investigated when any categorized observations are not collected. The input values of the parameters are set to be $\lambda=1.5$, $\eta=2$, p_1= 0.2, p_2= 0.5, p_3= 0.3, s_1 = 0.1, s_2 = 0.5 and s_3 = 1. We calculate *MSE* of $\hat{\lambda}$ and $\hat{\eta}$, and average of the *MSE*s of p_1, p_2 and p_3. Table 1 shows that a lot of data are needed for accurate estimation when there are only uncategorized field observations. However, the monotone decreasing property of *MSEs* for all the estimates of parameters roughly implies the consistency of the estimation.

Table 1. MSE as the number of uncategorized field observations

n	10	30	100	1000
MSE$(\hat{\lambda})$	6.4292	1.8408	0.3079	0.0381
MSE$(\hat{\eta})$	2.9727	1.4281	0.3652	0.0357
MSE$(\hat{p})$	0.0363	0.0207	0.0093	0.0048

4.2 Effect of the Number of Two Types of Field Observations

The second experiment is performed to investigate which information is more important between the number of uncategorized field observations and the number of categorized field ones. The same input values of the parameter as the first experiment are used and nine combinations (n, c) are considered. Table 2 represents *MSE*s of the estimates. The accuracy of the estimation is higher as both the number of complete observations and the number of incomplete observations increase. However, it is noticeable that the precision of $\hat{\lambda}$, $\hat{\eta}$ and $\hat{p}$ are more sensitive to the number of the categorized field observations.

Table 2. *MSE* as the number of the uncategorized and categorized field observations.

n	c_{sum}	$MSE(\hat{\lambda})$	$MSE(\hat{\eta})$	$MSE(\hat{p})$
10	10	1.9240	1.3748	0.0169
	30	0.2690	0.3577	0.0063
	100	0.0980	0.0870	0.0020
30	10	0.5091	0.5942	0.0176
	30	0.2300	0.2114	0.0061
	100	0.0767	0.0777	0.0020
100	10	0.2741	0.2567	0.0100
	30	0.1621	0.1722	0.0054
	100	0.0656	0.0734	0.0018

4.3 Effect of the Number of Uncategorized Field and Categorized Experiment Observations

The same type of experiment in Section 4.2 is performed with the uncategorized field observations and the experiment ones. The same input values of the parameters as Section 4.2 are used. Table 3 shows *MSE* of the estimates. The accuracy of the estimation becomes higher as both the number of uncategorized field observations and the number of categorized experimental ones increases. It is noticeable that the precision of $\hat{\lambda}$ and $\hat{\eta}$ is more sensitive to the number of experiment observations than the number of uncategorized field ones. On the other hand, that of $\hat{p}$ is much more sensitive to the number of uncategorized filed observations.

Table 3. *MSE* as the number of uncategorized and experimental observations

n	d	MSE$(\hat{\lambda})$	MSE$(\hat{\eta})$	MSE$(\hat{p})$
	10	0.2141	0.2211	0.0571
10	30	0.1037	0.0805	0.0643
	100	0.0234	0.0234	0.0625
	10	0.2752	0.2033	0.0317
30	30	0.0612	0.0660	0.0299
	100	0.0276	0.0275	0.0296
	10	0.1367	0.1556	0.0111
100	30	0.0806	0.0644	0.0080
	100	0.0194	0.0229	0.0076

5. CONCLUDING REMARKS

We dealt with an estimation problem for the mixed proportional hazards model with the constant hazards function and proposed the estimation method based on the EM algorithm. It was shown in the simulation studies that the accuracy of the estimation was improved as the number of two types of field observations and experimental observations increased. The precision of $\hat{\lambda}$ and $\hat{\eta}$ was shown to be more sensitive to the number of two types of categorized observations than the uncategorized field observations. However, that of $\hat{p}$ was much more sensitive to the number of the field observations.

References

[1] Everitt, B.S., & Hand, D.J.(1981), *Finite Mixture Distributions*, Chapman & Hall.

[2] McLachlan, G.J., & Basford, K.E.(1987), *Mixture Models ; Inference and applications to clustering*, Marcel Dekker.

[3] Titterington, D.M., Smith, A.F.M., & Markov, U.E.(1985), *Statistical analysis of finite mixture distribution*, John Wiley & Sons.

[4] Ansell, J.I., & Phillips, M.J.(1997), Practical aspects of modeling of repairable systems data using proportional hazard models, *Reliability Engineering and System Safety*, Volume 58, 165-171.

[5] Hirose, H.(1997), Mixture model of the power law, *IEEE Transactions on Reliability*, Volume 46, 146-153.

[6] Jaisingh, L.R., Dey, D.K., & Griffith, W.S.(1993), Properties of a multivariate survival distribution generated by a Weibull & inverse Gaussian mixture, *IEEE Transactions on Reliability*, Volume 42; 618-622.

[7] Jiang, R., & Murthy, D.N.P.(1995), Modeling failure-data by mixture of 2 Weibull distributions: a graphical approach, *IEEE Transactions on Reliability*, Volume 44, 477-488.

[8] Jiang, S., & Kececioglu, D.(1992), Graphical representation of two mixed-Weibull distributions, *IEEE Transactions on Reliability*, Volume 41, 241-247.

[9] Jiang, S., & Kececioglu, D.(1992), Maximum likelihood estimates from censored data for mixed-Weibull distributions, *IEEE Transactions on Reliability*, Volume 41, 248-255.

[10] Jozwiak, I.J.(1997), An introduction to the studies of reliability of systems using the Weibull proportional hazards model, *Microelectronics and Reliability*, Volume 37, 915-918.

[11] Kim, J.W., Yun, W.Y., & Dohi, T.(2003), Estimating the mixture of proportional hazards model with incomplete data, *Journal of Quality in Maintenance Engineering*, Volume 9, 265-278.

[12] Kumar, D., & Klefsjo, B.(1994), Proportional hazards model: a review. *Reliability Engineering and System Safety*, Volume 44, 177-188.

[13] Lau, T.S.(1998), Mixture models in hazard rates estimation. *Commun.Statist-Theory Meth.*, Volume 27, 1395-1407.

[14] Martorell, S., Sanchez, A., & Serradell, V.(1999), Age-dependent reliability model considering effects of maintenance and working conditions, *Reliability Engineering and System Safety*, Volume 64, 19-31.

[15] McClean, S.(1986), Estimation for the mixed exponential distribution using grouped follow-up data, *Applied Statistics*, Volume 36, 31-37.

[16] Peng, Y., & Dear, K.B.G.(2000), A nonparametric mixture model for cure rate estimation, *Biometrics*, Volume 59, 237-243.

[17] Solomon, P.J.(1984), Effect of misspecification of regression model in the analysis of survival data, *Biometrica*, Volume 71, 291-293.

[18] Sy, J.P., & Taylor, J.M.G.(2000), Estimation in a Cox proportional hazards cure model, *Biometrics*, Volume 56, 227-236.