Chapter 3.
DwACM Models

In this chapter, we construct a model to explain two types of data which are a dichotomous outcome about if a tested unit is active at the end of the experiment and a set of observation over time from a continuous measurement. We denote the model as DwACM, which is an abbreviation of "Dichotomous-data with Auxiliary Continuous Measurements". The continuous measurement is regarded as a candidate of a degradation measurement which may contain some information about the reliability of experimental units. If the relationship of these two types of data is strong, the continuous measurement can be considered as a good degradation measurement.

3.1 Modeling a Degradation Measurement

An ideal degradation measurement has two characteristics. First, it has to be monotone in time. Otherwise, it is difficult to tell the real condition of a product. Second, there is a clear critical point, the threshold. Then, the state of a product can be classified by whether the outcome of the measurement is above or below the given threshold. For simplicity, we assume all the degradation measurements are decreasing. For those increasing ones, they can be multiplied by -1 to satisfy the assumption.

To capture the monotone property, we assume that the response of the measurement is a linear function of time plus a random error. The mean degradation paths (the linear functions) of products have the same intercept and various slopes. It means that the degradation measurement of all units start at the same value and decay with different speed because of quality inconsistency. Consider N products. For one specific measurement Y and each individual i, $i = 1, \cdots, N$, $Y_{i1}, \cdots, Y_{in_i}$ are measured at prespecified times, $t_{i1}, \cdots, t_{in_i}$, respectively. Let

$$Y_{ij} = \beta_0 + \beta_i^R t_{ij} + \varepsilon_{ij}, \tag{3.1}$$

for $j = 1, \cdots, n_i$, where β_0 is the fixed common intercept and β_i^R's are random slopes. Assume that the errors, ε_{ij}'s follow independent normal distribution with mean 0 and variance σ^2.

This linear assumption between Y and t might be too strong. However, they can be transformed in order to fit the linear relationship more. Some basic models for modeling empirical degradation measurements, such as

$$\begin{aligned} \text{Exponential:} \quad & Y = \beta_0 \exp[\beta_1 t]\varepsilon_1, \\ \text{Power:} \quad & Y = \beta_0 t^{\beta_1}\varepsilon_2, \\ \text{Logarithmic:} \quad & Y = \beta_0 + \beta_1 \log t + \varepsilon_3, \end{aligned}$$

are all contained in our simple linear model after transformation (taking logarithm to both sides of exponential and power model), provided ε_1 and ε_2 follow log-normal distributions and ε_3 follows a normal distribution. Besides, it is well known that for any monotone function there exists a transformation which transforms the monotone function into linear. Although there is no guarantee that the transformation can transform all the degradation paths into linear simultaneously, the transformed paths should have similar pattern if they are from similar models. It can be observed, in practice, that the degradation paths usually decay in parallel. Another reason for using a simple model is that a complex model is hard to interpret and may not make any sense when there is no further physics or chemistry knowledge about the measurement. Therefore, the model (3.1) is applicable.

If Y's are observations of a good degradation measurement, the lifetime of product i will be the time t at which $\beta_0 + \beta_i^R t$ reaches the threshold (the soft failure). According to the model assumption, the degradation paths of all products have the same starting values. The difference among slopes, β_i^R's is the only way to distinguish different paths and, of course, the reason why $\beta_0 + \beta_i^R t$ as well as the lifetime of product i, T_i, varies from item to item. Therefore, β_i^R should be related to T_i. To derive the relationship, we assume that the lifetime is the time when its corresponding degradation path reaches the threshold. Then, β_i^R has to be limited to satisfy the condition that

$$\beta_0 + \beta_i^R T_i = \tau,$$

for all $i = 1, \cdots, N$, where τ denotes the threshold. From the equation above, it can be derived that the product, $\beta_i^R T_i$ is a constant, $\tau - \beta_0$. As the result,

$$\beta_i^R = \frac{\beta_1}{T_i},$$

where β_1 equals the fixed constant, $\tau - \beta_0$. Replacing β_i^R in (3.1) with $\frac{\beta_1}{T_i}$, it can be rewritten as

$$Y_{ij} = \beta_0 + \beta_1 \frac{t_{ij}}{T_i} + \varepsilon_{ij},$$

which states that once the lifetime is known the mean degradation process will be determined uniquely. The threshold of the degradation is $\beta_0 + \beta_1$.

From the derivation above, it can be seen that the mean path contains only one random coefficient, the slope, is essential. If there are more than one random coefficient, such as both the slope and the intercept are random, then the one-to-one correspondence between each random coefficient and the lifetime does not exist. In that case, it is very difficult to establish the relationship between each random coefficient and the lifetime.

Note that in this dissertation, we use the capital letters to denote both the random variables and their outcomes. Y is used also to denote the measurement itself depending on the contexts. If more than one measurements are considered at the same time, we use different superscripts to distinguish them. The superscript as well as the subscripts will be dropped out, if there is no confusion.

3.2 The Linkage of Two Types of Data

In our study, the lifetime, T_i, can not be observed. Let us assume T_i follow a distribution with probability density function (pdf) $f_i(T_i;\boldsymbol{\theta})$ and cumulative distribution function (cdf) $F_i(T_i;\boldsymbol{\theta})$. Many distributions are used to model lifetime data, for example, exponential, Weibull, log-normal, etc.. They may also contain covariates. The proportional hazards model is one possibility in this respect, cf. Lawless (1982).

In this dissertation, we assume lifetime is distributed exponentially. Denote W_i as the covariate of unit i and consider the following proportional hazards model:

$$f_i(T_i; W_i, \boldsymbol{\theta}) = \alpha_0 \exp(\alpha_1 W_i) \exp\{-[\alpha_0 T_i \exp(\alpha_1 W_i)]\}. \qquad (3.2)$$

Here, we consider the model with only one covariate. In (3.2) α_0 is the hazard rate of the baseline distribution ($W_i = 0$), and α_1 is the effect caused by of one unit increasing of W_i. It is that T_i given W_i follows an exponential distribution with mean $1/[\alpha_0 \exp(\alpha_1 W_i)]$.

In our study, we can also observe if product i is active at terminal experimental time, t_{in_i}, in addition to the continuous measurements. Denote Z_i as 1 if product i still performs well at time t_{in_i}; 0 otherwise. Thus, if T_i was given, Z_i is determined. Recall that we assume the continuous measurement, Y_{ij}, follows a normal distribution with mean $\beta_0 + \beta_1 \frac{t_{ij}}{T_i}$ and variance σ^2.

By using the same idea as that of Sammel, Ryan and Legler (1997), given unobserved latent variables, we assume that the conditional independence among the continuous and discrete variables. We apply similar assumption to link T_i, Z_i and Y_i in one model. Specifically, if T_i's were observed, according to the conditional independence assumption, the complete data log-likelihood for product i would be obtained easily as

$$\begin{aligned}\log L_i^c(\boldsymbol{\theta}; T_i, \mathbf{Y}_i) &= \log f_i(T_i; \boldsymbol{\theta}) + \log f_{\boldsymbol{\theta}}(\mathbf{Y}_i | T_i) \\ &= \log \alpha_0 + \alpha_1 W_i - \alpha_0 T_i \exp(\alpha_1 W_i) \\ &\quad -\tfrac{1}{2} n_i \log 2\pi\sigma^2 - \tfrac{1}{2} \textstyle\sum_{j=1}^{n_i} (\frac{Y_{ij} - \beta_1 - \beta_2 \frac{t_{ij}}{T_i}}{\sigma})^2,\end{aligned}$$

where $\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{in_i})$ and $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_0, \beta_1, \sigma^2)$. If the responses from different products are independent, we can get the complete data likelihood

$$L^c(\boldsymbol{\theta}; \boldsymbol{T}, \mathbf{Y}) = \prod_{i=1}^{N} L_i^c(\boldsymbol{\theta}; T_i, \mathbf{Y}_i),$$

where $\boldsymbol{T} = (T_1, \cdots, T_N)$ and $\mathbf{Y} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_N)$. Integrating out the latent part, the real likelihood for all products is

$$L(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{Z}) = \int_{\Omega_{\mathbf{Z}}} L^c(\boldsymbol{\theta}; \boldsymbol{T}, \mathbf{Y}) d\boldsymbol{T}, \tag{3.3}$$

where $\boldsymbol{Z} = (Z_1, \cdots, Z_N)$ and

$$\Omega_{\mathbf{Z}} = \prod_{i \in \{i|Z_i=1\}} [t_{in_i}, \infty) \times \prod_{i \in \{i|Z_i=0\}} (0, t_{in_i}).$$

When there is no auxiliary measure, the DwACM can be reduced to the binary response model,

$$P(Z_i = 1) = 1 - F(t_{in_i}, \boldsymbol{\alpha}).$$

However, if there is no binary response, the DwACM can not be reduced to a regular degradation measurement model. The reason is that in a regular degradation measurement model there must be a known threshold. It can not be determined in the DwACM if Z is not available. Therefore, the dichotomous response plays an important role. It provides information about the threshold. We may also say that Z identifies the model. To illustrate this, we can rewrite the continuous response as

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} u_i + \varepsilon_{ij},$$

where the reciprocal of u_i follows an gamma distribution with mean $1/\alpha$ and variance $1/\alpha^2$. It is obvious that the models with parameter $\boldsymbol{\theta} = (\alpha, \beta_0, \beta_1, \sigma^2)$ and $\boldsymbol{\theta}^* = (\alpha^*, \beta_0, \beta_1^*, \sigma^2)$, respectively, are identical when

$$\alpha^* \beta_1^* = \alpha \beta_1.$$

Hence, the model is not identifiable. However, if Z_i is also taken into consideration, then

$$\begin{aligned} f_{\boldsymbol{\theta}^*}(\boldsymbol{Y}_i, Z_i = 1) &= \int_t^\infty \prod_{j=1}^{n_i} \phi(\frac{Y_{ij} - \beta_0 - \beta_1^* t_{ij} u_i}{\sigma}) f_{\alpha^*}(u_i) du_i \\ &= \int_{\frac{\alpha^*}{\alpha} t}^\infty \prod_{j=1}^{n_i} \phi(\frac{Y_{ij} - \beta_0 - \beta_1 t_{ij} u_i}{\sigma}) f_{\alpha}(u_i) du_i \\ &\neq f_{\boldsymbol{\theta}}(\boldsymbol{Y}_i, Z_i = 1), \end{aligned}$$

and so is the case of $Z_i = 0$. Therefore, the model becomes identifiable. Intuitively, we can get the information about the common intercept and the random slopes from those responses of the continuous measurement. However, we can not distinguish β_1 from $\frac{1}{T_i}$ if we do not have the dichotomous results. The dichotomous results provide the information about the threshold, which is $\beta_0 + \beta_1$. Once both $\beta_0 + \beta_1$ and β_0 are estimated, β_1 can be estimated easily by simple subtraction. Then, we can tell the fixed coefficient, β_1, from $\frac{1}{T_i}$. Hence, for a product, it is not necessary to have both the binary result and the auxiliary measures. As long as some of the testing products are with the dichotomous outcomes which can provide the information about the threshold, the DwACM works. The DwACM therefore is very flexible.

Most of the time, the likelihood (3.3) does not have a closed form. It is almost impossible to derive the MLE or the posterior distribution by using classical methods. Thus, some computational based methods, which are briefly introduced in Chapter 2, are used. In frequentist setting, the EM algorithm provides a possible approach to obtain the MLE. If T is also regarded as a unknown parameter, the hierarchical structure can be found. Hence, we can also consider doing Bayesian analysis, and use MCMC to approximate the posterior distribution. See Chapter 5 for details.