Chapter 5.
ESTIMATION PROCEDURES

In this chapter, we present estimation procedures to the DwACM. Under the frequentist framework, we sketch an EM algorithm to approximate the MLE, and a bootstrap method to approximate the standard deviation of the MLE. For Bayesian estimation, we apply MCMC with Gibbs sampler to approximate the posterior distribution of the parameters of interest.

5.1 Frequentist Inferences

From frequentist point of view, the likelihood summarizes all information about the collected data, and the MLE is one of the most common-used estimator. Since the likelihood function (3.3) involves high dimension integrals, and is difficult to get a closed form, deriving the MLE from the likelihood directly is almost impossible. That motivates us applying the EM-algorithm, with some modifications, to approximate the MLE.

Recall that we consider N experimental units. For unit i, we take measures for an auxiliary continuous measurement at n_i prespecified times, $t_{i1}, \cdots, t_{in_i}$. Let Y_{ij} denote the result and Z_i be the final go-nogo dichotomous outcome. The two iterations of the EM-algorithm are as follows:

E-Step:

Compute expectations of complete data log-likelihood over the latent parts, T's, with respect to the conditional distribution given the observed parts, which is

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = Q_1(\alpha_0, \alpha_1|\boldsymbol{\theta}^{(p)}) + Q_2(\beta_0, \beta_1, \sigma^2|\boldsymbol{\theta}^{(p)}) - \sum_{i=1}^{N} n_i \log \sqrt{2\pi},$$

where

$$Q_1(\alpha_0, \alpha_1|\boldsymbol{\theta}^{(p)}) = N \log \alpha_0 + \alpha_1 \sum_{i=1}^{N} W_i - \alpha_0 \sum_{i=1}^{N} E_{\boldsymbol{\theta}^{(p)}}[T_i|\mathbf{Y}_i, Z_i] \exp(\alpha_1 W_i), \tag{5.1}$$

and

$$\begin{aligned} Q_2(\beta_0, \beta_1, \sigma^2|\boldsymbol{\theta}^{(p)}) &= \sum_{i=1}^{N}\sum_{j=1}^{n_i} \left\{ -\frac{1}{2}\log(\sigma^2) - \frac{(Y_{ij}-\beta_0)^2}{2\sigma^2} \right. \\ &+ \left. \frac{\beta_0\beta_1 Y_{ij} E_{\boldsymbol{\theta}^{(p)}}[\frac{t_{ij}}{T_i}|\mathbf{Y}_i, Z_i]}{\sigma^2} - \frac{\beta_1^2 E_{\boldsymbol{\theta}^{(p)}}[(\frac{t_{ij}}{T_i})^2|\mathbf{Y}_i, Z_i]}{2\sigma^2} \right\}. \end{aligned} \tag{5.2}$$

In our example, the latent parts appear in both $Q_1(\alpha_0, \alpha_1|\boldsymbol{\theta}^{(p)})$ and $Q_2(\beta_0, \beta_1, \sigma^2|\boldsymbol{\theta}^{(p)})$ are linear, which means the conditional expected complete log-likelihood can be obtained by replacing the latent parts by their conditional expectations. However, there are no closed forms for those expectations, and hence, the ordinary EM-algorithm does not work.

If it was easy to sample from the conditional distributions, the conditional expectations could be approximated by using simple Monte Carlo methods according to the following procedure. At first, generate $T_i^1, \cdots, T_i^B$ from $f_i(T|\mathbf{Y}_i, Z_i)$. Then approximate the conditional expectations of a function appearing in log-likelihood, say $g(T_i)$, by

$$E_{\boldsymbol{\theta}^{(p)}}[g(T_i)|\mathbf{Y}_i, Z_i] \approx \frac{1}{B}\sum_{s=1}^{B} g(T_i^s).$$

Such a method is known as Monte Carlo EM (MCEM), cf. Wei and Tanner (1900).

It is difficult to sample the Monte Carlo replicates from the conditional distributions directly in our case. Appreciate to the conditional independence of $\mathbf{Y}, \boldsymbol{Z}$ given $\boldsymbol{T}$, we have

$$f(T|\mathbf{Y}, \boldsymbol{Z}) = \frac{f(\boldsymbol{T}|\boldsymbol{Z})f(\mathbf{Y}|\boldsymbol{T}, \boldsymbol{Z})}{f(\mathbf{Y}|\boldsymbol{Z})} \propto f(\boldsymbol{T}|\boldsymbol{Z})f(\mathbf{Y}|\boldsymbol{T}).$$

By this property, we will describe an efficient Monte Carlo simulation. Before doing so, we first prove that all those conditional expectations exist. Note that the conditional expectation can be written as

$$E[g(T_i)|\boldsymbol{Y}_i, Z_i] = \frac{\int_{\Omega_{Z_i}} g(T_i)f(T_i)f(\boldsymbol{Y}_i|T_i)dT_i}{\int_{\Omega_{Z_i}} f(T_i)f(\boldsymbol{Y}_i|T_i)dT_i},$$

where $g(u)$ is taken as $1, u, 1/u$ and $1/u^2$ in the computation of the conditional expected complete-data log-likelihood. This leads to

$$\begin{aligned}
\int_{\Omega_{Z_i}} g(T_i)f(T_i)f(\boldsymbol{Y}_i|T_i)dT_i &\leq \int_0^\infty g(T_i)f(T_i)f(\boldsymbol{Y}_i|T_i)dT_i \\
&= \int_0^\infty g(\frac{1}{S_i})f(\frac{1}{S_i})\frac{1}{\sigma_i' S_i^2}\phi(\frac{S_i-\mu_i'}{\sigma_i'})dS_i \\
&\leq C(g,\boldsymbol{\theta})E[g(\frac{1}{S_i})\frac{1}{S_i^2}\exp(-\frac{\alpha_i}{S_i})] \\
&< E[g(\frac{1}{S_i})\frac{1}{S_i^2}] \\
&< \infty,
\end{aligned}$$

where the expectation is taken with respect to the normal distribution with mean, μ_i' and variance $\sigma_i'^2$, and

$$(\mu_i', \sigma_i'^2) = \Big(\frac{\sum_{j=1}^{n_i}(Y_{ij}-\beta_0)t_{ij}}{\beta_1\sum_{j=1}^{n_i}t_{ij}^2}, \frac{\sigma^2}{\beta_1^2\sum_{j=1}^{n_i}t_{ij}^2}\Big).$$

Here $\alpha_i = \alpha_0\exp(\alpha_1 W_i)$ is the failure rate of the lifetime distribution, and $C(g,\boldsymbol{\theta})$ is a constant depends on the function g and the parameters. Therefore all the integrals we need in the E-Step exist.

To implement it, we first generate B random samples $T_i^1, \cdots, T_i^B$ from the conditional distribution of T_i given Z_i, which is a truncated exponential distribution. Then, the expected value of a function $g(T_i)$ can be approximated by

$$E_{\boldsymbol{\theta}^{(p)}}[g(T_i)|\mathbf{Y}_i] \approx \frac{\sum_{s=1}^{B} g(T_i^s) \prod_{j=1}^{n_i} \phi(\frac{Y_{ij} - \beta_0^{(p)} - \beta_1^{(p)} \frac{t_{ij}}{T_i^s}}{\sigma^{(p)}})}{\sum_{s=1}^{B} \prod_{j=1}^{n_i} \phi(\frac{Y_{ij} - \beta_0^{(p)} - \beta_1^{(p)} \frac{t_{ij}}{T_i^s}}{\sigma^{(p)}})}. \tag{5.3}$$

Note that approximation (5.3) may work poorly when the number of measuring time is large. The reason behind this and a modification of (5.3) will be given in Chapter 6.

M-Step:

Choosing $\boldsymbol{\theta} = \boldsymbol{\theta}^{(p+1)}$ to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$. By using conditional independence assumption, the maximization in each iteration can be proceeded by maximizing $Q_1(\alpha_0, \alpha_1|\boldsymbol{\theta}^{(p)})$ and $Q_2(\beta_0, \beta_1, \sigma^2|\boldsymbol{\theta}^{(p)})$ separately. Let $(\alpha_0^{(p+1)}, \alpha_1^{(p+1)})$ be the solution of

$$\frac{\partial}{\partial(\alpha_0, \alpha_1)} Q_1(\alpha_0, \alpha_1|\boldsymbol{\theta}^{(p)}) = \mathbf{0},$$

which is equivalent to the following system of equations,

$$\begin{cases} \frac{N}{\alpha_0} - \sum_{i=1}^{N} E_{\boldsymbol{\theta}^{(p)}}[T_i|\boldsymbol{Y}_i, Z_i] \exp(\alpha_1 W_i) = 0, \\ \sum_{i=1}^{N} W_i - \alpha_0 \sum_{i=1}^{N} E_{\boldsymbol{\theta}^{(p)}}[T_i|\boldsymbol{Y}_i, Z_i] W_i \exp(\alpha_1 W_i) = 0. \end{cases} . \tag{5.4}$$

We can apply a Newton-Raphson method to find a numerical solution of (5.4). The $(\beta_0^{(p+1)}, \beta_1^{(p+1)}, \sigma^{2(p+1)})$ is the solution of

$$\frac{\partial}{\partial(\beta_0, \beta_1, \sigma^2)} Q_2(\beta_0, \beta_1, \sigma^2|\boldsymbol{\theta}^{(p)}) = \mathbf{0},$$

and it can be solved analytically as

$$
\begin{aligned}
\beta_0^{(p+1)} &= \frac{(\sum_{i=1}^N \sum_{j=1}^{n_i} Y_{ij})(\sum_{i=1}^N \sum_{i=1}^N E_i[(\frac{t_{ij}}{T_i})^2|.]) - (\sum_{i=1}^N \sum_{j=1}^{n_i} E_i[\frac{t_{ij}}{T_i}|.])(\sum_{i=1}^N \sum_{i=1}^N E_i[\frac{t_{ij}}{T_i} Y_{ij}|.])}{(\sum_{i=1}^N n_i)(\sum_{i=1}^N \sum_{j=1}^{n_i} E_i[(\frac{t_{ij}}{T_i})^2|.]) - (\sum_{i=1}^N \sum_{j=1}^{n_i} E_i[\frac{t_{ij}}{T_i}|.])^2}, \\
\beta_1^{(p+1)} &= \frac{(\sum_{i=1}^N n_i)(\sum_{i=1}^N \sum_{j=1}^{n_i} E_i[\frac{t_{ij}}{T_i} Y_{ij}|.]) - (\sum_{i=1}^N \sum_{j=1}^{n_i} E_i[\frac{t_{ij}}{T_i}|.])(\sum_{i=1}^N \sum_{j=1}^{n_i} Y_{ij})}{(\sum_{i=1}^N n_i)(\sum_{i=1}^N \sum_{j=1}^{n_i} E_i[(\frac{t_{ij}}{T_i})^2|.]) - (\sum_{i=1}^N \sum_{j=1}^{n_i} E_i[\frac{t_{ij}}{T_i}|.])^2},
\end{aligned}
$$

and

$$
\sigma^{2(p+1)} = \frac{\sum_{i=1}^N \sum_{j=1}^{n_i} (Y_{ij} - \beta_0^{(p+1)} - \beta_1^{(p+1)} E_i[\frac{t_{ij}}{T_i}|.])^2}{\sum_{i=1}^N n_i},
$$

where $E_i[g(T)|.] = E_{\boldsymbol{\theta}^{(p)}}[g(T)|\boldsymbol{Y}_i, Z_i]$.

To evaluate the accuracy of the MLE, the standard deviation is estimated by the following bootstrap method. The MLE, $\widehat{\theta}$, is calculated at first according to the above EM algorithm. Then, B bootstrap samples are generalized with respect to the distributions of which parameter, θ, is replaced by $\widehat{\boldsymbol{\theta}}$. For each bootstrap sample, the parameters, $\widehat{\theta}_1^*, \cdots, \widehat{\theta}_B^*$ are calculated via the same above EM algorithm. Then, the estimator of the standard deviation is given by

$$
\sqrt{\frac{1}{B} \sum_{b=1}^{B} (\widehat{\theta}_b^* - \widehat{\theta})^2}.
$$

For more materials about the bootstrap method, the reader is referred to Chapter 6 of Efron and Tibshirani (1993) for details.

5.2 Bayesian Inferences

It is more or less risky by only using a statistical method to find out the degradation measurement. Therefore, follow-up experiments are often suggested to confirm the degradation model, which is built upon according to previous experiments. It is natural to update the information under the Bayesian framework. Here we consider the case in which there is no covariate. Specifically,

$$T_i|\alpha \sim Exp(\alpha) \quad \text{independently},$$

$$Y_{ij}|(T_i, \beta_0, \beta_1, \sigma^2) \sim N(\beta_0 + \beta_1 \frac{t_{ij}}{T_i}, \sigma^2) \quad \text{independently},$$

and

$$Z_i|T_i = I_{(t_{in_i}, \infty)}(T_i).$$

If the prior is $\pi(\boldsymbol{\theta})$, then the joint distribution can be written as

$$f(\alpha, \beta_0, \beta_1, \sigma^2, \boldsymbol{T}, \boldsymbol{Y}, \boldsymbol{Z}) \;\; = \;\; \pi(\boldsymbol{\theta}) f(\boldsymbol{T}|\alpha) f(\boldsymbol{Y}|\boldsymbol{T}, \beta_0, \beta_1, \sigma^2) f(\boldsymbol{Z}|\boldsymbol{T}).$$

Assuming that the arguments are mutually independent, then we can write the prior $\pi(.)$ as

$$\pi(\boldsymbol{\theta}) \propto p(\alpha)p(\beta_0)p(\beta_1)p(\sigma^2).$$

To avoid the improper posterior, we choose those proper priors of α, β_0, β_1 and σ^2 to be

$$\begin{aligned} \alpha &\sim G(\tau_1, \tau_2), \\ \beta_0 &\sim N(\mu_{\beta_0}, \sigma^2_{\beta_0}), \\ \beta_1 &\sim N(\mu_{\beta_1}, \sigma^2_{\beta_1}), \end{aligned}$$

and

$$\sigma^2 \sim IG(\lambda_1, \lambda_2),$$

where $G(\tau_1, \tau_2)$ denotes the gamma distribution with density function

$$f(x) \propto \exp\{-\frac{x}{\tau_2}\} x^{\tau_1 - 1},$$

and $IG(\lambda_1, \lambda_2)$ denotes the inverse gamma distribution whose reciprocal follows $G(\lambda_1, \lambda_2)$. The super parameters setting can be done according to experiences. If there is no information available, we can choose super parameters to make the prior variance large so that the effect of the prior is moderated.

Under the prior distributions, the posterior distribution can not be derived analytically. The Gibbs sampler provides a solution to find the multi-dimensional posterior distribution. To implement it, we first derive the full conditional distributions, which are

$$
\begin{aligned}
\alpha|. &\sim G(N+\tau_1, \frac{\tau_2}{1+\tau_2\sum_{i=1}^N T_i}),\\
\beta_0|. &\sim N(\frac{\sigma_{\beta_0}^2\sum_{i=1}^N\sum_{j=1}^{n_i}(Y_{ij}-\beta_1\frac{t_{ij}}{T_i})+\sigma^2\mu_{\beta_0}}{\sigma_{\beta_0}^2\sum_{i=1}^N n_i+\sigma^2}, \frac{\sigma_{\beta_0}^2\sigma^2}{\sigma_{\beta_0}^2\sum_{i=1}^N n_i+\sigma^2}),\\
\beta_1|. &\sim N(\frac{\sigma_{\beta_1}^2\sum_{i=1}^N\sum_{j=1}^{n_i}(Y_{ij}-\beta_0)(\frac{t_{ij}}{T_i})+\sigma^2\mu_{\beta_1}}{\sigma_{\beta_1}^2\sum_{i=1}^N\sum_{j=1}^{n_i}(\frac{t_{ij}}{T_i})^2+\sigma^2}, \frac{\sigma_{\beta_1}^2\sigma^2}{\sigma_{\beta_1}^2\sum_{i=1}^N\sum_{j=1}^{n_i}(\frac{t_{ij}}{T_i})^2+\sigma^2}),\\
\sigma^2|. &\sim IG(\frac{\sum_{i=1}^N n_i}{2}+\lambda_1, \frac{2\lambda_2}{2+\lambda_2\sum_{i=1}^N\sum_{j=1}^{n_i}(Y_{ij}-\beta_0-\beta_1\frac{t_{ij}}{T_i})^2}),
\end{aligned}
$$

and

$$T_i|. \propto \exp(-\alpha T_i + \frac{b_i}{T_i} + \frac{c_i}{T_i^2}) I_{z_i}(T_i),$$

where $b_i = \frac{\beta_1}{\sigma^2}\sum_{j=1}^{n_i}(Y_{ij}-\beta_0)t_{ij}, c_i = -\frac{\beta_1^2}{2\sigma^2}\sum_{j=1}^{n_i} t_{ij}^2$. All the full conditional distributions but $T_i|.$'s are standard from which it is easy to sample. To sample T_i, we can use an one-iteration Metropolis-Hastings algorithm with proposal kernel $\exp(-\alpha T_i)$ instead of sampling from $\exp(-\alpha T_i + \frac{b_i}{T_i} + \frac{c_i}{T_i^2})$ directly. Such a hybrid algorithm is called the "Metropolis-Hastings-within-Gibbs" sampler. The procedure is described as follows.

The Algorithm:

1. Start a chain at $(\alpha^{(0)}, \boldsymbol{T}^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}, \sigma^{2(0)})$.

2. Update $(\alpha^{(p)}, \boldsymbol{T}^{(p)}, \beta_1^{(p)}, \beta_2^{(p)}, \sigma^{2(p)})$ by

   (2.1) Sampling the $\alpha^{(p+1)}$ from $G(N+\tau_1, \frac{\tau_2}{1+\tau_2\sum_{i=1}^{N} T_i^{(p)}})$.

   (2.2) Sampling $T_i^{(p+1)}$ by the following procedure:

   (2.2.1) Sampling a candidate T from $Exp(\alpha^{(p+1)})$ truncated on either $(0, t_{in_i}]$ or (t_{in_i}, ∞), depending on if $Z_i = 0$ or 1.

   (2.2.2) Generate a point U from $U(0,1)$.

   (2.2.3) If $U \leq \min[1, \exp(b_i^{(p)}(\frac{1}{T} - \frac{1}{T_i^{j(p)}}) + c_i^{(p)}(\frac{1}{T^2} - \frac{1}{T_i^{j(p)2}}))]$, set $T_i^{(p+1)} = T$; otherwise keep $T_i^{(p+1)}$ at $T_i^{(p)}$, where $b_i^{(p)} = \frac{\beta_1^{(p)}}{\sigma^{2(p)}}\sum_{j=1}^{n_i}(Y_{ij} - \beta_0^{(p)})t_{ij}$ and $c_i^{(p)} = -\frac{\beta_1^{(p)2}}{2\sigma^{2(p)}}\sum_{j=1}^{n_i} t_{ij}^2$.

   (2.3) Sampling the $\beta_0^{(p+1)}$ from

$$N\Big(\frac{\sigma_{\beta_0}^2 \sum\sum(Y_{ij} - \beta_1^{(p)}\frac{t_{ij}}{T_i^{(p+1)}}) + \sigma^{2(p)}\mu_{\beta_0}}{\sigma_{\beta_0}^2\sum_{i=1}^{N} n_i + \sigma^{2(p)}}, \frac{\sigma^{2(p)}\sigma_{\beta_0}^2}{\sigma_{\beta_0}^2\sum_{i=1}^{N} n_i + \sigma^{2(p)}}\Big).$$

   (2.4) Sampling the $\beta_1^{(p+1)}$ from

$$N\Big(\frac{\sigma_{\beta_1}^2 \sum\sum(Y_{ij} - \beta_0^{(p+1)})(\frac{t_{ij}}{T_i^{(p+1)}}) + \sigma^{2(p)}\mu_{\beta_1}}{\sigma_{\beta_1}^2\sum\sum(\frac{t_{ij}}{T_i^{(p+1)}})^2 + \sigma^{2(p)}}, \frac{\sigma_{\beta_1}^2\sigma^{2(p)}}{\sigma_{\beta_1}^2\sum\sum(\frac{t_{ij}}{T_i^{(p+1)}})^2 + \sigma^{2(p)}}\Big).$$

   (2.5) Sampling the $\sigma^{2(p+1)}$ from

$$IG\Big(\frac{\sum_{i=1}^{N} n_i}{2} + \lambda_1, \frac{2\lambda_2}{2 + \lambda_2\sum\sum(Y_{ij} - \beta_0^{(p+1)} - \beta_1^{(p+1)}\frac{t_{ij}}{T_i^{(p+1)}})^2}\Big).$$

3. Repeat 2. until $p = M_0$.

By using the standard burn-in procedure for the simulation, we only keep last M-runs for doing estimation. The M samples can be used to estimate the posterior distribution, and the posterior mean and the posterior median of the parameter can then be carried out easily.