Chapter 2

Hazard Regression

In this chapter we define a parametric regression model for log hazard functions of censored failure times. We start with introducing definitions and notation in the first section. The regression model and the estimates are defined in Section 2.2. We show the consistency and normality of the estimates in Section 2.3 and leave the proofs of two required lemmas to Section 2.4. Finally, in Section 2.5, we define the regression model for bivariate failure times and discuss inference based on the estimates.

2.1 Definitions and Notation

2.1.1 Failure Time Distribution

All functions, unless stated otherwise, are defined over the interval $R^+ \equiv [0, \infty)$. Let T be a nonnegative continuous random variable representing the failure time of individuals in some population. Let f denote the *density function* of T and let the *distribution* *function* be

$$F(t) = P(T \le t) = \int_0^t f(u)\, du.$$

The probability of an individual surviving till time t is given by the *survivor function*

$$S(t) = P(T \ge t) = \int_t^\infty f(u)\, du = 1 - F(t).$$

The *hazard function* is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t} = \frac{f(t)}{S(t)},$$

which specifies the instantaneous rate of failure at time t, given that the individual survives until at least time t.

The functions $f, F, S,$ and h give mathematically equivalent specifications of the distribution of T. Since $f(t) = -S'(t)$, by the definition of h,

$$h(t) = -\frac{d}{dt} \log S(t).$$

Thus

$$\log S(t) - \log S(0) = -\int_0^t h(u)\, du$$

and since $S(0) = 1$,

$$S(t) = \exp\left(-\int_0^t h(u)\, du\right) \tag{2.1}$$

and

$$f(t) = h(t) \exp\left(-\int_0^t h(u)\, du\right). \tag{2.2}$$

For our log hazard regression, we denote the *log hazard function* by α, that is,

$$\alpha(t) = \log h(t). \tag{2.3}$$

It is easy to derive expressions for $S(t)$ and $f(t)$ in terms of $\alpha(t)$.

2.1.2 Censored Survival Data

Consider n individuals in some population. Let T_i be the failure time of individual i, and C_i a *censoring time*, that is, a time beyond which individual i can not be observed. The variable T_i will be observed whenever $T_i \leq C_i$. Let $X_i = \min(T_i, C_i)$ and

$$\delta_i = \begin{cases} 1 & \text{if } T_i \leq C_i, \\ 0 & \text{if } T_i > C_i. \end{cases}$$

That is, δ_i indicates whether the failure time T_i is censored or not and X_i is equal to T_i if T_i is observed, and to C_i if it is not. The data from observations on n individuals consist of the pairs (X_i, δ_i), $i = 1, \cdots, n$. For the data we first make the following assumptions.

Assumptions

(1) (X_i, δ_i) are independent random vectors with an identical distribution;

(2) the failure times T_i and the censoring times C_i are independent.

We denote the survivor functions and the density functions of failure time T_i and the censoring times C_i by S, S_c, f and f_c, respectively. By Lawless (1982), under the Assumptions (1) and (2),

$$P(X < t, \Delta = \delta) = \begin{cases} \int_0^t S(u) f_c(u) du & \delta = 0, \\ \int_0^t f(u) S_c(u) du & \delta = 1. \end{cases}$$

The probability is a measure defined on $[0, \infty) \times \{0, 1\}$. We denote it by ν. Then the density function of (X_i, δ_i) is defined by

$$f(t, \delta) = f(t)^{\delta} S(t)^{1-\delta} f_c(t)^{1-\delta} S_c(t)^{\delta}. \tag{2.4}$$

It is easy to check that f is the Radon-Nikodym derivative of ν with respect to the σ-finite cross-product measure μ defined on $[0, \infty) \times \{0, 1\}$ such that the marginal of μ on $[0, \infty)$ is the Lebesgue measure and the marginal on $\{0, 1\}$ is counting measure.

2.2 Log Hazard Regression

In this section we fit a parametric model for the log hazard function α of failure time T. First we define the model in a general form and determine the estimates that we will use. Then, in Section 2.2.2, we discuss the conditions on the model.

2.2.1 Model and Estimates

Let G be a p-dimensional linear space of real continuous functions defined on $[0, \infty)$, and let $\{b_j, j = 1, \cdots, p\}$ be a basis of G. We call G the *regression space*. The *log hazard regression* is a model for the log hazard function, that is,

$$\alpha(t|\boldsymbol{\beta}) = \sum_{j=1}^{p} \beta_j b_j(t) = \boldsymbol{\beta}^T \boldsymbol{b}(t), \tag{2.5}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ is unknown and $\boldsymbol{b}(t) = (b_1(t), \cdots, b_p(t))^T$.

Given $\boldsymbol{\beta}$, by Equations (2.3), (2.1), and (2.2), we have the hazard function, survivor function, and density function

$$\begin{aligned}
h(t|\boldsymbol{\beta}) &= \exp\Big(\sum_{j=1}^{p} \beta_j b_j(t)\Big), && (2.6)\\
S(t|\boldsymbol{\beta}) &= \exp\Big(-\int_0^t \exp\big(\sum_{j=1}^{p} \beta_j b_j(u)\big)du\Big), && (2.7)\\
f(t|\boldsymbol{\beta}) &= \exp\Big(\sum_{j=1}^{p} \beta_j b_j(t)\Big) \exp\Big\{-\int_0^t \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big)du\Big\}, && (2.8)
\end{aligned}$$

respectively.

Note that to ensure that (2.7) defines a non-degenerate survivor function, it is necessary and sufficient that

$$\int_0^t \exp\,(\sum_{j=1}^{p} \beta_j b_j(u))du < \infty \tag{2.9}$$

for some $t > 0$ and

$$\int_0^{\infty} \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big)du = +\infty. \tag{2.10}$$

So we need some conditions on $\boldsymbol{b}$ and $\boldsymbol{\beta}$. For now we assume that (2.9) and (2.10) hold, and we will postpone discussing the conditions until Section 2.2.2.

For n observations $(X_i, \delta_i), i = 1, \cdots, n,$ we make the following assumptions, in addition to Assumptions (1) and (2).

Assumptions

(3) the log hazard function α of T_i satisfies model (2.5);

(4) the distribution of the censoring times C_i does not involve $\boldsymbol{\beta}$ and the density f_C of C is bounded.

Then by (2.4), the density function of (X_i, δ_i) is

$$f(t, \delta|\boldsymbol{\beta}) = f(t|\boldsymbol{\beta})^{\delta} S(t|\boldsymbol{\beta})^{1-\delta} f_c(t)^{1-\delta} S_c(t)^{\delta}. \tag{2.11}$$

We denote the joint distribution function of (X_i, δ_i) by $L(\boldsymbol{X}|\boldsymbol{\beta})$, that is,

$$L(\boldsymbol{X}|\boldsymbol{\beta}) = \prod_{i=1}^{n} f(X_i, \delta_i|\boldsymbol{\beta}),$$

and we choose $\hat{\boldsymbol{\beta}}$ to maximize L.

Let

$$f_q(t, \delta|\boldsymbol{\beta}) = f(t|\boldsymbol{\beta})^{\delta} S(t|\boldsymbol{\beta})^{1-\delta} \tag{2.12}$$

and

$$L_q(\boldsymbol{\beta}|X) = \prod_{i=1}^{n} f_q(X_i, \delta_i|\boldsymbol{\beta}).$$

Then

$$\begin{aligned} L(\boldsymbol{X}|\boldsymbol{\beta}) &= \prod_{i=1}^{n} f_q(X_i, \delta_i|\boldsymbol{\beta}) f_c(X_i)^{1-\delta_i} S_c(X_i)^{\delta_i} \\ &= L_q(\boldsymbol{\beta}|\boldsymbol{X}) \prod_{i=1}^{n} f_c(X_i)^{1-\delta_i} S_c(X_i)^{\delta_i}. \end{aligned}$$

Since S_c and f_c do not involve $\boldsymbol{\beta}$, we estimate $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, the maximizer of the function L_q.

We call the partial derivative of $\log L_q$ with respect to $\boldsymbol{\beta}$ the *score function*, denoted by $\boldsymbol{U}^{(n)}(\boldsymbol{\beta}|\boldsymbol{X})$, which is a vector with p components

$$\boldsymbol{U}^{(n)}(\boldsymbol{\beta}|\boldsymbol{X})_m = \frac{\partial}{\partial \beta_m} \log L_q(\boldsymbol{\beta}|\boldsymbol{X}) = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_m} \log f_q(X_i, \delta_i|\boldsymbol{\beta}), \tag{2.13}$$

$m = 1, 2, \cdots, p$. The *observed information matrix*, denoted by $I^{(n)}(\boldsymbol{\beta}|\boldsymbol{X})$, is a $p \times p$ matrix with (ml)-th entry

$$I^{(n)}(\boldsymbol{\beta}|\boldsymbol{X})_{ml} = -\frac{\partial^2}{\partial \beta_m \partial \beta_l} \log L_q(\boldsymbol{\beta}|\boldsymbol{X}) = -\sum_{i=1}^{n} \frac{\partial^2}{\partial \beta_m \partial \beta_l} \log f_q(X_i, \delta_i|\boldsymbol{\beta}). \tag{2.14}$$

We denote the *information matrix*, in the usual way, by $I(\boldsymbol{\beta})$, which is a $p \times p$ matrix with the (ml)-th entry

$$I(\boldsymbol{\beta})_{ml} = -E\left\{\frac{\partial^2}{\partial \beta_m \partial \beta_l} \log f_q(X_1, \delta_1|\boldsymbol{\beta})\right\}. \tag{2.15}$$

By definition of f_q and Equations (2.8) and (2.7), we can write $f_q, \log f_q, \boldsymbol{U}^{(n)}$, $I^{(n)}$, and $I(\boldsymbol{\beta})$ in terms of the basis functions $b_1, \cdots, b_p$. We have

$$f_q(t, \delta|\boldsymbol{\beta}) = \Big[\exp\Big(\sum_{j=1}^{p} \beta_j b_j(t)\Big)\Big]^{\delta} \exp\Big\{-\int_0^t \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du\Big\}; \tag{2.16}$$

and

$$\log f_q(t, \delta|\boldsymbol{\beta}) = \delta \sum_{j=1}^{p} \beta_j b_j(t) - \int_0^t \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du. \tag{2.17}$$

Thus, under regularity conditions,

$$\begin{aligned}
&\frac{\partial}{\partial \beta_m} \log f_q(X_i, \delta_i|\boldsymbol{\beta}) \\
= \;& \frac{\partial}{\partial \beta_m} \left\{\delta_i \sum_{j=1}^{p} \beta_j b_j(X_i) - \int_0^{X_i} \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du\right\} \\
= \;& \delta_i b_m(X_i) - \int_0^{X_i} b_m(u) \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du;
\end{aligned}$$

and

$$\frac{\partial^2}{\partial \beta_m \partial \beta_l} \log f_q(X_i, \delta_i|\boldsymbol{\beta}) = -\int_0^{X_i} b_m(u) b_l(u) \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du.$$

Hence

$$\boldsymbol{U}^{(n)}(\boldsymbol{\beta}|\boldsymbol{X})_m = \sum_{i=1}^{n} \left\{\delta_i b_m(X_i) - \int_0^{X_i} b_m(u) \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du\right\} \tag{2.18}$$

and

$$I^{(n)}(\boldsymbol{\beta}|\boldsymbol{X})_{ml} = \sum_{i=1}^{n} \int_0^{X_i} b_m(u) b_l(u) \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du. \tag{2.19}$$

Since for any vector $\boldsymbol{v} = (v_1, \cdots, v_p)^T \neq \boldsymbol{0}$,

$$\boldsymbol{v}^T I^{(n)}(\boldsymbol{\beta}|\boldsymbol{X}) \boldsymbol{v} = \sum_{i=1}^{n} \int_0^{X_i} (\boldsymbol{v}^T \boldsymbol{b}(u))^2 \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du > 0,$$

$I^{(n)}(\boldsymbol{\beta}|\boldsymbol{X})$ is positive definite (i.e., the second partial derivative of $\log L_q$ is negative definite). Therefore, if $\hat{\boldsymbol{\beta}}$ exists, it is the unique solution of the p equations

$$\boldsymbol{U}^{(n)}(\boldsymbol{\beta}|\boldsymbol{X}) = \mathbf{0}. \tag{2.20}$$

In the next chapter, we will show how to numerically solve (2.20) for $\hat{\boldsymbol{\beta}}$.

Note: By the relationship between L_q and L,

$$\frac{\partial}{\partial \beta_m} \log L(\boldsymbol{\beta}|\boldsymbol{X}) = \frac{\partial}{\partial \beta_m} \log L_q(\boldsymbol{\beta}|\boldsymbol{X}).$$

Therefore,

$$\boldsymbol{U}^{(n)}(\boldsymbol{\beta}|\boldsymbol{X})_m = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_m} \log f(X_i, \delta_i|\boldsymbol{\beta})$$

and

$$I^{(n)}(\boldsymbol{\beta}|\boldsymbol{X})_{ml} = -\sum_{i=1}^{n} \frac{\partial^2}{\partial \beta_m \partial \beta_l} \log f(X_i, \delta_i|\boldsymbol{\beta})$$

which are the usual definitions of the score function and the observed information matrix.

2.2.2 Conditions on $\boldsymbol{b}$ and $\boldsymbol{\beta}$

As we mentioned in the last section, we need some conditions on the basis $\boldsymbol{b}$ or on $\boldsymbol{\beta}$ such that (2.9) and (2.10) are true. A sufficient condition that (2.9) and (2.10) hold is that $\boldsymbol{b}$ is *bounded*. We say a function $\boldsymbol{u} : R^+ \to R^p$ is bounded if there exists $0 < M < \infty$ such that

$$\|\boldsymbol{u}(t)\|^2 \equiv \sum_{j=1}^{p} |u_j(t)|^2 < M,$$

for all $t \in R^+$.

Proposition 2.2.1 *If $\boldsymbol{b}$ is bounded, then*

(i) *(2.9) and (2.10) hold for all t and $\boldsymbol{\beta}$ and*

(ii) *there exists a number $M(\boldsymbol{\beta}, p)$ such that*

$$f(t|\boldsymbol{\beta}) \leq M(\boldsymbol{\beta}, p) \exp\Big(-\frac{t}{M(\boldsymbol{\beta}, p)}\Big). \tag{2.21}$$

Proof. Since $\boldsymbol{b}$ is bounded, there exists a number $0 < M < \infty$, dependent on $\boldsymbol{\beta}$ and p, such that $|\sum_{j=1}^{p} \beta_j b_j(u)| < M$. Hence

$$\frac{1}{\exp(M)} \leq \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) \leq \exp(M). \tag{2.22}$$

It follows that for any t

$$\int_0^t \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du \leq \int_0^t \exp(M) du < \infty$$

and

$$\int_0^\infty \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du \geq \int_0^\infty \frac{1}{\exp(M)} du = \infty.$$

So we proved that (2.9) and (2.10) are true.

Next, let $M(\boldsymbol{\beta}, p) = \exp(M)$, then, by the definition of f from (2.8) and from (2.22),

$$\begin{aligned} f(t|\boldsymbol{\beta}) &= \exp\Big(\sum_{j=1}^{p} \beta_j b_j(t)\Big) \exp\Big\{-\int_0^t \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du\Big\} \\ &\leq M(\boldsymbol{\beta}, p) \exp\Big(-\frac{t}{M(\boldsymbol{\beta}, p)}\Big), \end{aligned}$$

which completes the proof of (ii). □

We would like the regression space G to contain the log hazard for the commonly used Weibull distribution. Recall that the density function of a Weibull distribution is

$$f_w(t) = \gamma\lambda^\gamma t^{\gamma-1} \exp\Big(-(\lambda t)^\gamma\Big), \quad t > 0, \lambda > 0, \gamma > 1.$$

The corresponding log hazard function is

$$\alpha_w(t) = \log(\gamma\lambda^\gamma) + (\gamma - 1)\log(t).$$

Thus, if $b_p(t) = \log(t)$ and $b_{p-1} \equiv 1$, then model (2.5) includes the Weibull distributions.

More generally, we consider the basis $b_1, \cdots, b_{p-1}$ bounded and $b_p(t) = \log(t + c)$, where $c \geq 0$ is a constant. We call b_p *log tail* and $\{b_j, j = 1, \cdots, p\}$ *bounded plus log tail.* This basis doesn't satisfy the assumption of Proposition 2.2.1. But we have the following result.

Proposition 2.2.2 *If $\boldsymbol{b}$ is bounded plus log tail and $\beta_p > -1$, then*

(i) *(2.9) and (2.10) hold for all t and*

(ii) *there exists a number $M(\boldsymbol{\beta}, p)$ such that*

$$f(t|\boldsymbol{\beta}) \leq M(\boldsymbol{\beta}, p)(t + c)^{\beta_p} \exp\Big(-\frac{(t + c)^{\beta_p+1}}{M(\boldsymbol{\beta}, p)}\Big) \tag{2.23}$$

Proof. Since $\boldsymbol{b}$ is bounded plus log tail, there exists M such that $\Big|\sum_{j=1}^{p-1}\beta_j b_j(u)\Big| < \log(M)$. Then

$$\frac{(u + c)^{\beta_p}}{M} \leq \exp\Big(\sum_{j=1}^{p}\beta_j b_j(u)\Big) \leq M(u + c)^{\beta_p}. \tag{2.24}$$

Therefore,

$$\int_0^t \exp\Big(\sum_{j=1}^p \beta_j b_j(u)\Big) du \le \int_0^t M(u+c)^{\beta_p} du < \infty$$

for any t and

$$\int_0^\infty \exp\Big(\sum_{j=1}^p \beta_j b_j(u)\Big) du \ge \int_0^\infty \frac{(u+c)^{\beta_p}}{\exp(M)} du = \infty$$

as $\beta_p > -1$. Then (i) is proved. To prove (ii), using the definition of f from (2.8) and using (2.24), we have

$$\begin{aligned} f(t|\boldsymbol{\beta}) &= \exp\Big(\sum_{j=1}^p \beta_j b_j(t)\Big) \exp\Big\{-\int_0^t \exp\Big(\sum_{j=1}^p \beta_j b_j(u)\Big) du\Big\} \\ &\le M(t+c)^{\beta_p} \exp\Big(\frac{c^{\beta_p+1}}{(\beta_p+1)M}\Big) \exp\Big(-\frac{(t+c)^{\beta_p+1}}{(\beta_p+1)M}\Big) \\ &\equiv C_1(t+c)^{\beta_p} \exp\Big(-\frac{(t+c)^{\beta_p+1}}{C_2}\Big) \\ &\le M(\boldsymbol{\beta},p)(t+c)^{\beta_p} \exp\Big(-\frac{(t+c)^{\beta_p+1}}{M(\boldsymbol{\beta},p)}\Big), \end{aligned}$$

where $M(\boldsymbol{\beta},p) = \max\{C_1, C_2\}$. So (ii) is true. $\square$

From now on, unless stated otherwise, we assume the n observations (X_i, δ_i), $i = 1, \cdots, n$, satisfy Assumptions (1) - (4) defined in the previous sections and either

Assumption

(5a) $\boldsymbol{b}$ is bounded and $\boldsymbol{\beta} \in R^p$

or

(5b) $\boldsymbol{b}$ is bounded plus log tail and $\boldsymbol{\beta} \in R^{p-1} \times (-1, \infty)$.

We call R^p the *parameter space*, if (5a) holds and $R^{p-1} \times (-1,\infty)$ the *parameter space* if (5b) holds.

Note 1 Since (2.9) holds for all $t > 0$, $f(t) > 0$ for all $t > 0$. So the support of f is R^+, which will be used for the proof of Theorem 2.3.1 in the next section.

Note 2 It is easy to see that if $\boldsymbol{b}$ is bounded then for any compact set B in the parameter space, there exists M dependent on B and p such that

$$f(t|\boldsymbol{\beta}) \leq M \exp\Big(-\frac{t}{M}\Big)$$

for all $\boldsymbol{\beta} \in B$. If $\boldsymbol{b}$ is bounded plus log tail, then for any compact set B in the parameter space, there exist M and $\beta'_p > -1$, dependent on B and p, such that

$$f(t|\boldsymbol{\beta}) \leq M(t+c)^{\beta'_p} \exp\Big(-\frac{(t+c)^{\beta'_p+1}}{M}\Big)$$

for any $\boldsymbol{\beta} \in B$. That is, on a compact set B in the parameter space, the density function $f(t|\boldsymbol{\beta})$ of T can be dominated by a function of t.

2.3 Consistency and Normality of $\hat{\boldsymbol{\beta}}$

In this section we use Theorem 5.1 in Lehmann and Casella (1998) to show the following theorem.

Theorem 2.3.1 *Let $(X_1, \delta_1), \cdots, (X_n, \delta_n)$ satisfy Assumptions (1) - (5). Then, for $\boldsymbol{\beta}_0$, the true parameter,*

(i) $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$ *in probability;*

(ii) $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \Longrightarrow N(\mathbf{0}, [I(\boldsymbol{\beta}_0)]^{-1})$;

(iii) $\frac{1}{n}I^{(n)}(\boldsymbol{\beta}|\boldsymbol{X}) \to I(\boldsymbol{\beta}_0)$ *in probability provided* $\boldsymbol{\beta} \to \boldsymbol{\beta}_0$ *in probability.*

We first state the conditions of their theorem.

Conditions

A. For different values of $\boldsymbol{\beta}$, the distributions P_β of the observations are distinct.

B. The distributions P_β have common support.

C. The observations $(X_1, \delta_1), \cdots, (X_n, \delta_n)$ are i.i.d. with probability density $f(t, \delta|\boldsymbol{\beta})$ with respect to a σ-finite measure μ.

D. There exists an open set ω of the parameter space Ω containing the true parameter point $\boldsymbol{\beta}_0$ such that for almost all (X, δ), the density $f(X, \delta|\boldsymbol{\beta})$ admits all third derivatives $(\partial^3/\partial\beta_j\partial\beta_k\partial\beta_l)f(X, \delta|\boldsymbol{\beta})$ for all $\boldsymbol{\beta} \in \omega$.

E. The first and second logarithmic derivatives of f satisfy the equations

$$E\Big\{\frac{\partial}{\partial\boldsymbol{\beta}}\log f(X, \delta|\boldsymbol{\beta})\Big\} = 0,$$

and

$$I(\boldsymbol{\beta})_{ml} = E\Big\{-\frac{\partial^2}{\partial\beta_m\partial\beta_l}\log f(X, \delta|\boldsymbol{\beta})\Big\} = E\Big\{\frac{\partial}{\partial\beta_m}\log f(X, \delta|\boldsymbol{\beta})\cdot\frac{\partial}{\partial\beta_l}\log f(X, \delta|\boldsymbol{\beta})\Big\},$$

$l, m \in \{1, \cdots, p\}$.

F. $I(\boldsymbol{\beta})_{jk}$ is finite and the matrix $I(\boldsymbol{\beta})$, a $p \times p$ matrix with the (ml)-th entry $I(\boldsymbol{\beta})_{ml}$, is positive definite for all $\boldsymbol{\beta} \in \omega$.

G. There exist functions M_{jkl} such that

$$\left|(\partial^3/\partial\beta_j\partial\beta_k\partial\beta_l)\log f(X,\delta|\boldsymbol{\beta})\right| \leq M_{jkl}(X)$$

for all $\boldsymbol{\beta} \in \omega$, where $E_{\boldsymbol{\beta}_0}(M_{jkl}(X)) < \infty$ for all $j,k,l \in \{1,\cdots,p\}$.

Theorem 2.3.2 *(Theorem 5.1 in Lehmann, E.L. and Casella, G, 1998). Let (X_1,δ_1), $\cdots$, (X_n,δ_n) be i.i.d., each with a density $f(t,\delta|\boldsymbol{\beta})$ which satisfies the above conditions. Then with probability tending to 1 and as $n \to \infty$, there exists a solution $\hat{\boldsymbol{\beta}}$ of the likelihood equation*

$$\sum_{i=1}^{n} \frac{\partial \log f(X_i,\delta_i|\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$$

such that

(a) *$\hat{\beta}_j$ is consistent for estimating β_j,*

(b) *$\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ is asymptotically normal with mean zero and covariance matrix $[I(\boldsymbol{\beta})]^{-1}$, and*

(c) *$\hat{\beta}_j$ is asymptotically efficient in the sense that $\sqrt{n}(\hat{\beta}_j-\beta_j)$ converges in distribution to a random normal variable with mean zero and variance $[I(\boldsymbol{\beta})_{jj}]^{-1}$.*

Proof of Theorem 2.3.1.

Suppose $\hat{\boldsymbol{\beta}}$ is the solution of the equation

$$0 = \sum_{i=1}^{n} \frac{\partial \log(f_q(t,\delta|\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{\partial \log(f(t,\delta|\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}}.$$

To show (i) and (ii), we need only check the conditions of Theorem 2.3.2.

First, by Assumptions 1 and 2 and the definition of $f(\cdot,\cdot|\boldsymbol{\beta})$ from (2.11), (X_1,δ_1), $\cdots$, (X_n,δ_n) and $f(\cdot,\cdot|\boldsymbol{\beta})$ satisfy condition C.

Second, since $\{b_1,\cdots,b_p\}$ is a basis, from (2.16), the definition of f_q in terms of the basis functions, for different value of $\boldsymbol{\beta}$, f_q is distinct. From Note 1, after Propositions 2.2.1 and 2.2.2, the support of f_q is $[0,\infty)$. By the relationship between $f(\cdot,\cdot|\boldsymbol{\beta})$ and f_q from (2.11) and (2.12), conditions A and B hold.

Then, from (2.11) and (2.12),

$$\frac{\partial \log f(X_1,\delta_1|\boldsymbol{\beta})}{\partial \beta_l} = \frac{\partial \log f_q(X_1,\delta_1|\boldsymbol{\beta})}{\partial \beta_l}, \tag{2.25}$$

$l=1,\cdots,p$, since S_c and f_c do not involve $\boldsymbol{\beta}$. Hence condition D follows from a direct calculation of the derivatives of f_q from the definition (2.16).

Next for condition F, $I(\boldsymbol{\beta})_{jk}$ is finite follows from Lemma 2.3.4 below. As shown in the calculations after (2.19), $I(\boldsymbol{\beta})$ is positive definite.

Finally, conditions E and G follow from Lemmas 2.3.4 and 2.3.5 below, respectively.

Therefore, (a) and (b) in Theorem 2.3.2 yield (i) and (ii), respectively.

To prove (iii), we need to use (i) and the bounds in Lemma 2.3.4 below. The proof is as in Theorem 3.10 of Lehmann and Cassella (1998). Thus, we have finished proving Theorem 2.3.1.

Using (ii) and (iii) in the above theorem and Slutsky's theorem, we have the following Corollary.

Corollary 2.3.3 *Let $(X_1,\delta_1),\cdots,(X_n,\delta_n)$ satisfy Assumptions (1) - (5). Then*

$$[nI^{(n)}(\hat{\boldsymbol{\beta}}|\boldsymbol{X})]^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \Rightarrow N(\mathbf{0},I),$$

where I is the $p \times p$ identity matrix.

The following two lemmas are used in the proof of Theorem 2.3.1.

Lemma 2.3.4 *Suppose that $\boldsymbol{b}$ satisfies Assumption (5a) (or Assumption (5b)) and B is a compact set in R^p (or in $R^{p-1} \times (-1, \infty)$). Then there exists $0 < M < \infty$, dependent on B and p (or on B, p, and c), such that for any $\boldsymbol{\beta} \in B$, $s_k \in \{1, \cdots, p\}; k = 1, \cdots, m; m = 0, 1, 2, 3$,*

$$\left| \frac{\partial^m \log f_q(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \beta_{s_1} \cdots \partial \beta_{s_m}} \right| < M K(X_1), \tag{2.26}$$

where

$$K(X_1) = \begin{cases} 1 + X_1 & \text{if } \boldsymbol{b} \text{ is bounded} \\ 1 + |\log(X_1 + c)| + (X_1 + c)^{\beta_p + 4} & \text{if } \boldsymbol{b} \text{ is bounded plus log tail.} \end{cases} \tag{2.27}$$

Furthermore, $E\{K(X_1)\} < \infty$.

Lemma 2.3.5 *Let $\boldsymbol{\beta}_0$ be the true values of the parameters in model (2.5). Then*

i) $E\left\{ \frac{\partial \log f_q(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}_0} \right\} = \mathbf{0}$;

ii) $E\left\{ \frac{\partial \log f_q(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \beta_m} \Big|_{\boldsymbol{\beta}_0} \frac{\partial \log f_q(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \beta_l} \Big|_{\boldsymbol{\beta}_0} \right\} = -E\left\{ \frac{\partial^2 \log f_q(X_1, \delta_1 | \boldsymbol{\beta})}{\partial \beta_m \partial \beta_l} \Big|_{\boldsymbol{\beta}_0} \right\}$.

We will give the proofs of the above lemmas in the next section.

2.4 Proofs of the Lemmas

Now we prove Lemmas 2.3.4 and 2.3.5.

For convenience, we first calculate the partial derivatives of $\log f_q$ with respect to $\boldsymbol{\beta}$ in terms of the basis functions. By the definition of $\log f_q$ from (2.17), for $s_k \in \{1, \cdots, p\}; k = 1, \cdots, m$,

$$\frac{\partial^m \log f_q(X_1, \delta_1|\boldsymbol{\beta})}{\partial\beta_{s_1}\cdots\partial\beta_{s_m}} = \begin{cases} \delta_1 \sum_{j=1}^{p} \beta_j b_j(X_1) - \int_0^{X_1} \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du, & m = 0; \\ \delta_1 b_{s_1}(X_1) - \int_0^{X_1} b_{s_1}(u) \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du, & m = 1; \\ -\int_0^{X_1} b_{s_1}(u)\cdots b_{s_m}(u) \exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) du, & m = 2, 3. \end{cases} \tag{2.28}$$

Proof of Lemma 2.3.4

If $\boldsymbol{b}$ is bounded, since B is a compact set, $\sum_{j=1}^{p} |\beta_j b_j(u)|$ is bounded on B, and so is $\exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big)$. Hence, by (2.28), we can find $0 < M < \infty$, dependent on B and p, such that (2.26) holds.

If $\boldsymbol{b}$ is bounded plus log tail, then $b_1, \cdots, b_{p-1}$ are bounded. Hence there exists $1 < M_1 < \infty$ such that for $\boldsymbol{\beta} \in B$

$$\max\left\{\Big|\sum_{j=1}^{p-1} \beta_j b_j(u)\Big|, \sum_{j=1}^{p-1} |b_j(u)|\right\} < M_1, \tag{2.29}$$

which implies

$$\exp\Big(\sum_{j=1}^{p} \beta_j b_j(u)\Big) < e^{M_1}(u+c)^{\beta_p}. \tag{2.30}$$

Then for $m = 0$, by (2.28) and (2.30), (2.26) is true. To prove (2.26) for $m = 1, 2, 3$, it suffices to show that there exists $M < \infty$ such that

$$\int_0^{X_1} |b_{s_1}(u)\cdots b_{s_m}(u)|(u+c)^{\beta_p} du < M(1 + (X_1 + c)^{\beta_p+4}), \tag{2.31}$$

$s_k \in \{1, \cdots, p\}; k = 1, \cdots, m$. This clearly holds for $s_k \in \{1, \cdots, p-1\}$, since b_j is bounded, $j = 1, \cdots, p-1$. But the log tail $\log(t + c)$ is unbounded, which makes the proof of (2.31) more complicated. So we need to investigate the properties of the integral

$$\int_0^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du,$$

$m = 1, 2, 3$. Noting $\beta_p > -1$ and $c \geq 0$, we have the following easily proven facts.

Fact 1 for any $\ell > 0$ and any compact set $B \subset (-1, \infty)$, there exists $M_1 < \infty$ such that

$$\sup_{u \in (0,1], \beta_p \in B} |\log(u+c)|^{\ell} (u+c)^{\frac{\beta_p+1}{2}} < M_1;$$

Fact 2 for any compact set $B \in (-1, \infty)$, $\sup_{\beta_p \in B} \int_0^1 (u+c)^{\frac{\beta_p - 1}{2}} du < \infty$;

Fact 3 $\log(u+c) < u + c$ for $u \in [1, \infty)$,

Fact 4 for any $\beta' > 0, \beta'' > -1$, and $M' > 0$,

$$\int_0^{\infty} (u+c)^{\beta''} \exp\Big\{ -(u+c)^{\beta'}/M' \Big\} du < \infty.$$

We will use **Facts 1 - 3** to prove that

i) there exists $M_2 < \infty$ such that for $\boldsymbol{\beta} \in B$,

$$\int_0^1 |\log(u+c)|^m (u+c)^{\beta_p} du < M_2,$$

$m = 1, 2, 3$;

ii) there exists $M_3 < \infty$ such that for $X_1 > 1$ and $\boldsymbol{\beta} \in B$,

$$\int_1^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du < M_3 (X_1 + c)^{\beta_p + 4},$$

$m = 1, 2, 3$.

If i) and ii) hold, then, when $X_1 \le 1$,

$$\begin{aligned}
& \int_0^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du \\
\le & \int_0^1 |\log(u+c)|^m (u+c)^{\beta_p} du < M_2,
\end{aligned}$$

and when $X_1 > 1$,

$$\begin{aligned}
& \int_0^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du \\
= & \int_0^1 |\log(u+c)|^m (u+c)^{\beta_p} du + \int_1^{X_i} |\log(u+c)|^m (u+c)^{\beta_p} du \\
\le & M_2 + M_3 (X_1 + c)^{\beta_p + 4} \le M(1 + (X_1 + c)^{\beta_p + 4})
\end{aligned}$$

for $M = \max\{M_2, M_3\}$. So (2.31) would hold, which implies that (2.26) would be true.

Proof of i): Since B is compact, using **Facts 1** and **2**, we can find $M_2', M_2'' < \infty$ such that for $\boldsymbol{\beta} \in B$,

$$|\log(u+c)|^m (u+c)^{\frac{\beta_p+1}{2}} < M_2',$$

$m = 1, 2, 3$, for $u \in (0, 1)$ and

$$\int_0^1 (u+c)^{\frac{\beta_p-1}{2}} du \le M_2''.$$

Then for $\boldsymbol{\beta} \in B$,

$$\begin{aligned}
& \int_0^1 |\log(u+c)|^m (u+c)^{\beta_p} du \\
\le & \int_0^1 |\log(u+c)|^m (u+c)^{\frac{\beta_p+1}{2}} (u+c)^{\frac{\beta_p-1}{2}} du \\
\le & M_2' \int_0^1 (u+c)^{\frac{\beta_p-1}{2}} du \le M_2' M_2'' = M_2 < \infty,
\end{aligned}$$

$m = 1, 2, 3$.

Proof of ii): By **Fact 3** and (2.29),

$$\begin{aligned}
& \int_1^{X_1} |\log(u+c)|^m (u+c)^{\beta_p} du \\
\leq & \int_1^{X_1} (u+c)^3 (u+c)^{\beta_p} du \\
= & \frac{1}{\beta_p+4}\Big((X_1+c)^{\beta_p+4} - (1+c)^{\beta_p+4}\Big) \\
\leq & M_3 (X_1+c)^{\beta_p+4}
\end{aligned}$$

$m = 1, 2, 3$, for some $M_3 > 0$.

Now that we have proven that (2.26) is true, we show that $E\{K(X_1)\} < \infty$.

First suppose that $\boldsymbol{b}$ is bounded. Then

$$E\{K(X_1)\} = E\{1 + X_1\} \leq E\{(1 + T_1)\},$$

since, by definition, $X_1 \leq T_1$. By Proposition 2.4.1 below, $E\{T_1\} < \infty$.

Now consider $\boldsymbol{b}$ bounded plus log tail. Write

$$K(X_1) = K^*(X_1) + K^{**}(X_1),$$

where

$$K^*(X_1) = 1 + (X_1+c)^{\beta_p+4} + |\log(X_1+c)| I\{X_1 + c \geq 1\}$$

and

$$K^{**}(X_1) = |\log(X_1+c)| I\{X_1 + c < 1\}.$$

Since K^* is a non-decreasing function,

$$E\{K^*(X_1)\} \leq E\{K^*(T_1)\} < E\{K(T_1)\},$$

which is finite by Proposition 2.4.2 below. To show $E(K^{**})$ is bounded, note that, by Assumption 4, f_c is bounded, and by (ii) in Proposition 2.2.2, $f(t|\boldsymbol{\beta})$ is bounded on $[0,1]$. Thus, the marginal density of X_1,

$$f_{X_1}(x) = f(x|\boldsymbol{\beta})S_C(x) + S(x|\boldsymbol{\beta})f_C(x),$$

is bounded. Then we have

$$E\Big\{K^{**}(X_1)\Big\} \le \int_0^1 |\log(x+c)|f_{X_1}(x)dx < \infty$$

since $\int_0^1 |\log(x+c)|dx < \infty$. Therefore,

$$E\{K(X_1)\} = E\{K^*(X_1)\} + E\Big\{K^{**}(X_1)\Big\} < \infty,$$

which completes the proof of Lemma 2.3.4.

The following propositions were used in the proof of Lemma 2.3.4.

Proposition 2.4.1 *If $\boldsymbol{b}$ is bounded, then for any $a > 0$, $E(T^a) < \infty$.*

Proof. Since $\boldsymbol{b}$ is bounded, by (ii) in Proposition 2.2.1, there exists $0 < M(\boldsymbol{\beta}, p) < \infty$ such that

$$f(t|\boldsymbol{\beta}) \le M(\boldsymbol{\beta}, p) \exp\Big(-\frac{t}{M(\boldsymbol{\beta}, p)}\Big).$$

From this and **Fact 4** in the proof of Lemma 2.3.4, we have

$$E(T^a) = \int_0^\infty t^a f(t|\boldsymbol{\beta})dt \le \int_0^\infty M(\boldsymbol{\beta}, p)t^a \exp\Big(\frac{-t}{M(\boldsymbol{\beta}, p)}\Big)dt < \infty.$$

□

Proposition 2.4.2 *If $\boldsymbol{b}$ is bounded plus log tail and $\beta_p > -1$, then*

(i) *for any* $a > 0, E(T^a) < \infty$, *and*

(ii) $E(|\log(T+c)|) < \infty$.

Proof. Since $b_1, \cdots, b_{p-1}$ are bounded, by (ii) in Proposition 2.2.2, there exists $0 < M(\boldsymbol{\beta}, p) < \infty$ such that (2.23) hold. Let

$$g(t, \boldsymbol{\beta}) = \exp\Big(-\frac{(t+c)^{\beta_p+1}}{M(\boldsymbol{\beta}, p)}\Big).$$

Then from (2.23) and **Fact 4** in the proof of Lemma 2.3.4,

$$E(T^a) = \int_0^\infty t^a f(t|\boldsymbol{\beta})\, dt \le M(\boldsymbol{\beta}, p) \int_0^\infty t^a (t+c)^{\beta_p} g(t, \boldsymbol{\beta})\, dt < \infty,$$

as $\beta_p > -1$. Thus (i) is proved.

To prove (ii), using (2.23) and **Facts 1 - 4** in the proof of Lemma 2.3.4, we have

$$\begin{aligned}
& E(|\log(T+c)|) = \int_0^\infty |\log(t+c)| f(t|\boldsymbol{\beta}) dt \\
\le\ & M(\boldsymbol{\beta}, p) \int_0^\infty |\log(t+c)| (t+c)^{\beta_p} g(t, \boldsymbol{\beta})\, dt \\
\le\ & M(\boldsymbol{\beta}, p) \Big\{ \int_0^1 |\log(t+c)| (t+c)^{\frac{\beta_p+1}{2}} (t+c)^{\frac{\beta_p-1}{2}} g(t, \boldsymbol{\beta})\, dt + \int_1^\infty (t+c)^{\beta_p+1} g(t, \boldsymbol{\beta})\, dt \Big\} \\
<\ & M(\boldsymbol{\beta}, p) \Big\{ M_1 \int_0^1 (t+c)^{\frac{\beta_p-1}{2}} g(t, \boldsymbol{\beta})\, dt + \int_1^\infty (t+c)^{\beta_p+1} g(t, \boldsymbol{\beta})\, dt \Big\} < \infty,
\end{aligned}$$

which finishes the proof of (ii). □

Remark: The following proposition will be used in the proof of Lemma 2.3.5 and Section 2.5. The proof is omitted since it uses the same procedures and arguments as in the proofs of Propositions 2.4.1 and 2.4.2.

Proposition 2.4.3 *Suppose that* $\boldsymbol{b}$ *satisfies the assumptions of Lemma 2.3.4. Then* $E\{[K(X_1)]^2\} < \infty$.

The proof of Lemma 2.3.5

To prove i), write

$$E\left\{\frac{\partial \log f_q(X_1,\delta_1|\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}_0}\right\} = E\left\{\frac{\partial \log f(X_1,\delta_1|\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}_0}\right\}$$
$$= \int \frac{\partial f(X_1,\delta_1|\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}_0} d\mu = \left(\frac{\partial}{\partial \boldsymbol{\beta}}\int f(X_1,\delta_1|\boldsymbol{\beta})d\mu\right)\Big|_{\boldsymbol{\beta}_0} = \frac{\partial E(\mathbf{1})}{\partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}_0} = \mathbf{0}.$$

To justify the interchange of differentiation and integration, we must prove that there exists a neighborhood B of $\boldsymbol{\beta}_0$ such that uniformly on B, the partial derivatives of $f(X_1,\delta_1|\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ can be dominated by an integrable function of (X_1,δ). Then by the Dominated Convergence Theorem, we can exchange the order of derivative and integral.

To do this, we choose $a>0$ such that $B=\{\boldsymbol{\beta} : \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le a\}$ is compact in the parameter space. By Lemma 2.3.4, there exists M such that for $\boldsymbol{\beta}\in B$,

$$\left|\frac{\partial \log f_q(X_1,\delta_1|\boldsymbol{\beta})}{\partial \beta_l}\right| \le MK(X_1), \tag{2.32}$$

$l=1,\cdots,p$, and $E(K(X_1))<\infty$. Hence, from (2.25) and (2.32), for $\boldsymbol{\beta}\in B$,

$$\left|\frac{\partial f(X_1,\delta_1|\boldsymbol{\beta})}{\partial \beta_l}\right| = f(X_1,\delta_1|\boldsymbol{\beta})\left|\frac{\partial \log f(X_1,\delta_1|\boldsymbol{\beta})}{\partial \beta_l}\right| \le MK(X_1)f(X_1,\delta_1|\boldsymbol{\beta}).$$

Thus it suffices to find an integrable function of (X_1,δ) to dominate $K(X_1)f(X_1,\delta_1|\boldsymbol{\beta})$.

By the definition of $f(\cdot,\cdot|\boldsymbol{\beta})$ from (2.11),

$$K(t)f(t,1|\boldsymbol{\beta}) = K(t)f(t|\boldsymbol{\beta})S_c(t) \le K(t)f(t|\boldsymbol{\beta}) \le K(t)f^*(t),$$

where f^* is the bound on f given in Note 2 after Proposition 2.2.2. It is easy to show that Kf^* is integrable.

To bound $K(X_1)f(X_1,0|\boldsymbol{\beta}) = K(X_1)S(X_1|\boldsymbol{\beta})f_C(X_1)$, it suffices to bound $K(X_1)S(X_1|\boldsymbol{\beta})$, since, by assumption 4, f_c is bounded. By the definition of $S(t|\boldsymbol{\beta})$ from (2.7) and by the proofs of Propositions 2.2.1 and 2.2.2, there exists $M > 0$ such that, for $\boldsymbol{\beta} \in B$,

$$S(t|\boldsymbol{\beta}) \leq \exp\Big(-\frac{t}{M}\Big) \equiv S^*(t)$$

if $\boldsymbol{b}$ is bounded, and

$$S(t|\boldsymbol{\beta}) \leq M\exp\Big(-\frac{(t+c)^{\beta_p'+1}}{M}\Big) \equiv S^*(t)$$

if $\boldsymbol{b}$ is bounded plus log tail, where β_p' is the lower bound of β_p for $\boldsymbol{\beta} \in B$. One easily shows that KS^* is integrable.

To prove ii), provided integration and differentiation can be interchanged, we can write

$$\begin{aligned}
0 &= \frac{\partial}{\partial\beta_m}E\left\{\frac{\partial\log f(X_1,\delta_1|\boldsymbol{\beta})}{\partial\beta_l}\right\} = \frac{\partial}{\partial\beta_m}\int\frac{\partial\log f(X_1,\delta_1|\boldsymbol{\beta})}{\partial\beta_l}f(X_1,\delta_1|\boldsymbol{\beta})\,d\mu \\
&= \int\frac{\partial}{\partial\beta_m}\left(\frac{\partial\log f(X_1,\delta_1|\boldsymbol{\beta})}{\partial\beta_l}f(X_1,\delta_1|\boldsymbol{\beta})\right)d\mu \\
&= \int\left\{\frac{\partial\log f(X_1,\delta_1|\boldsymbol{\beta})}{\partial\beta_m}\cdot\frac{\partial\log f(X_1,\delta_1|\boldsymbol{\beta})}{\partial\beta_l}+\frac{\partial^2\log f(X_1,\delta_1|\boldsymbol{\beta})}{\partial\beta_m\beta_l}\right\}f(X_1,\delta_1|\boldsymbol{\beta})d\mu \\
&= E\left\{\frac{\partial\log f_q(X_1,\delta_1|\boldsymbol{\beta})}{\partial\beta_m}\cdot\frac{\partial\log f_q(X_1,\delta_1|\boldsymbol{\beta})}{\partial\beta_l}\right\}+E\left\{\frac{\partial^2\log f_q(X_1,\delta_1|\boldsymbol{\beta})}{\partial\beta_m\partial\beta_l}\right\}.
\end{aligned}$$

Therefore,

$$E\left\{\frac{\partial}{\partial\beta_m}\log f_q(X_1,\delta_1|\boldsymbol{\beta})\frac{\partial}{\partial\beta_l}\log f_q(X_1,\delta_1|\boldsymbol{\beta})\right\} = -E\left\{\frac{\partial^2}{\partial\beta_m\partial\beta_l}\log f_q(X_1,\delta_1|\boldsymbol{\beta})\right\}.$$

To exchange integration and differentiation, we need to bound $\partial^2\log f(X_1,\delta_1|\boldsymbol{\beta})/\partial\beta_m\partial\beta_l$. By i) and Proposition 2.4.3, this can be done with an argument similar to that in the proof of i). We omit the calculations here.

2.5 Log Hazard Regression of Paired Failure Times

Now we consider the individuals with paired failure times. We first introduce the notation and define the log hazard regression model and estimates for paired failure data. Then, in Section 2.5.2, we show the consistency and normality of our estimates and give a consistent estimate of the covariance matrix of the estimates.

2.5.1 Model and Notation

Consider n subjects with paired failure times for each. For $k = 1, 2$, and $i = 1, \cdots, n$, let T_{ki} and C_{ki} be the kth failure time and censoring time of the ith individual, respectively. Let $X_{ki} = \min(T_{ki}, C_{ki})$ and

$$\delta_{ki} = \begin{cases} 1 & \text{if } T_{ki} \leq C_{ki}, \\ 0 & \text{if } T_{ki} > C_{ki}. \end{cases}$$

Then (X_{ki}, δ_{ki}) is the observation of the kth failure of the ith individual.

Note: the censoring times may be the same, i.e., $C_{1i} = C_{2i}$.

We refer to f_k, F_k, S_k, h_k, and α_k as the corresponding density, distribution, survivor, hazard, and log hazard functions, respectively. We fit a log hazard regression model for the two log hazard functions separately. That is, we suppose that

$$\alpha_1(t|\boldsymbol{\beta}_1) = \sum_{j=1}^{p_1} \beta_{1j} b_{1j}(t)$$

and

$$\alpha_2(t|\boldsymbol{\beta}_2) = \sum_{j=1}^{p_2} \beta_{2j} b_{2j}(t).$$

For $k = 1, 2$, we estimate $\boldsymbol{\beta}_k$ by $\hat{\boldsymbol{\beta}}_k$ as defined in Section 2.2.1. Referring to the univariate case, we denote the regression basis by $\boldsymbol{b}_k$. As in (2.16), we let

$$f_{kq}(t, \delta|\boldsymbol{\beta}_k) = f_k(t|\boldsymbol{\beta}_k)^{\delta} S_k(t|\boldsymbol{\beta}_k)^{1-\delta}.$$

We also let $\boldsymbol{U}_k^{(n)}(\boldsymbol{\beta}_k|\boldsymbol{X}_k)$, $I_k^{(n)}(\boldsymbol{\beta}_k|\boldsymbol{X}_k)$, and $I_k(\boldsymbol{\beta}_k)$ be the corresponding score functions, observed information matrix, and information matrix, respectively, see (2.13), (2.14) and (2.15). Similarly, we denote by K_k the corresponding functions defined in (2.27). Let

$$\underline{\boldsymbol{\beta}} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \qquad \hat{\underline{\boldsymbol{\beta}}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix}. \tag{2.33}$$

We denote the ith summand in $\boldsymbol{U}_k^{(n)}$ by $\boldsymbol{u}_{ki}(\boldsymbol{\beta}_k)$, that is,

$$\boldsymbol{u}_{ki}(\boldsymbol{\beta}_k) = \frac{\partial}{\partial \boldsymbol{\beta}_k} \log f_{kq}(X_{ki}, \delta_{ki}|\boldsymbol{\beta}_k),$$

$k = 1, 2$. Let

$$\boldsymbol{U}^{(n)}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2|\boldsymbol{X}_1, \boldsymbol{X}_2) = \begin{pmatrix} \boldsymbol{U}_1^{(n)}(\boldsymbol{\beta}_1|\boldsymbol{X}_1) \\ \boldsymbol{U}_2^{(n)}(\boldsymbol{\beta}_2|\boldsymbol{X}_2) \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} \boldsymbol{u}_{1i}(\boldsymbol{\beta}_1) \\ \boldsymbol{u}_{2i}(\boldsymbol{\beta}_2) \end{pmatrix},$$

and define

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \operatorname{cov}(\boldsymbol{u}_{11}(\boldsymbol{\beta}_1), \boldsymbol{u}_{11}(\boldsymbol{\beta}_1)) & \operatorname{cov}(\boldsymbol{u}_{11}(\boldsymbol{\beta}_1), \boldsymbol{u}_{21}(\boldsymbol{\beta}_2)) \\ \operatorname{cov}(\boldsymbol{u}_{21}(\boldsymbol{\beta}_2), \boldsymbol{u}_{11}(\boldsymbol{\beta}_1)) & \operatorname{cov}(\boldsymbol{u}_{21}(\boldsymbol{\beta}_2), \boldsymbol{u}_{21}(\boldsymbol{\beta}_2)) \end{pmatrix}. \tag{2.34}$$

2.5.2 Consistency and Normality of $\hat{\underline{\boldsymbol{\beta}}}$

In this section, we show the consistency and the normality of the estimates $\hat{\underline{\boldsymbol{\beta}}}$. The main result, given in Theorem 2.5.2, relies on the asymptotic normality of $\boldsymbol{U}^{(n)}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2|\boldsymbol{X}_1, \boldsymbol{X}_2)$ given in the following.

Proposition 2.5.1 *Suppose, marginally, each of the failure times satisfies Assumptions (1) through (5). Then*

$$\frac{1}{\sqrt{n}}\boldsymbol{U}^{(n)}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2|\boldsymbol{X}_1, \boldsymbol{X}_2) \Rightarrow N(0, \Sigma).$$

Proof. Since $\boldsymbol{u}_{ki}(\boldsymbol{\beta}_k), i = 1, \cdots, n$, are i.i.d. and, by i) in Lemma 2.3.5, $E(\boldsymbol{u}_{ki}(\boldsymbol{\beta}_k)) = 0$, using the multivariate central limit theorem, we have

$$\frac{1}{\sqrt{n}}\boldsymbol{U}^{(n)}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2|\boldsymbol{X}_1, \boldsymbol{X}_2) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{u}_{1i}(\boldsymbol{\beta}_1) \\ \boldsymbol{u}_{2i}(\boldsymbol{\beta}_2) \end{pmatrix} \Rightarrow N(0, \Sigma).$$

□

Theorem 2.5.2 *Suppose, marginally, each of the failure times satisfies Assumptions (1) through (5). Then*

(i) $\hat{\underset{\sim}{\boldsymbol{\beta}}} \to \underset{\sim}{\boldsymbol{\beta}}$ *in probability;*

(ii) $\sqrt{n}(\hat{\underset{\sim}{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}) \Rightarrow N(0, Q)$, *where*

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = \begin{pmatrix} I_1(\boldsymbol{\beta}_1)^{-1}\Sigma_{11}I_1(\boldsymbol{\beta}_1)^{-1} & I_1(\boldsymbol{\beta}_1)^{-1}\Sigma_{12}I_2(\boldsymbol{\beta}_2)^{-1} \\ I_2(\boldsymbol{\beta}_2)^{-1}\Sigma_{21}I_1(\boldsymbol{\beta}_1)^{-1} & I_2(\boldsymbol{\beta}_2)^{-1}\Sigma_{22}I_2(\boldsymbol{\beta}_2)^{-1} \end{pmatrix}. \tag{2.35}$$

Proof. (i) For any $\epsilon > 0$, by Theorem 2.3.1,

$$P(\|\underset{\sim}{\boldsymbol{\beta}} - \hat{\underset{\sim}{\boldsymbol{\beta}}}\| > \epsilon) \leq P(\|\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1\| > \epsilon) + P(\|\boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2\| > \epsilon) \to 0,$$

as $n \to \infty$.

(ii) By expanding $\boldsymbol{U}^{(n)}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2|\boldsymbol{X}_1, \boldsymbol{X}_2)$ in a Taylor series about $\underset{\sim}{\boldsymbol{\beta}}$, we have

$$\boldsymbol{U}^{(n)}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2|\boldsymbol{X}_1, \boldsymbol{X}_2) = \begin{pmatrix} \boldsymbol{U}_1^{(n)}(\boldsymbol{\beta}_1|\boldsymbol{X}_1) \\ \boldsymbol{U}_2^{(n)}(\boldsymbol{\beta}_2|\boldsymbol{X}_2) \end{pmatrix} - \begin{pmatrix} I_1^{(n)}(\boldsymbol{\beta}_1^*|\boldsymbol{X}_1)(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \\ I_2^{(n)}(\boldsymbol{\beta}_2^*|\boldsymbol{X}_2)(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \end{pmatrix},$$

where $\underset{\sim}{\boldsymbol{\beta}}^* \equiv (\boldsymbol{\beta}_1^*, \boldsymbol{\beta}_2^*)^T$ is on a line segment between $\hat{\underset{\sim}{\boldsymbol{\beta}}}$ and $\underset{\sim}{\boldsymbol{\beta}}$. Since $\boldsymbol{U}^{(n)}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2 | \boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{0}$,

$$\begin{pmatrix} \boldsymbol{U}_1^{(n)}(\boldsymbol{\beta}_1|\boldsymbol{X}_1) \\ \boldsymbol{U}_2^{(n)}(\boldsymbol{\beta}_2|\boldsymbol{X}_2) \end{pmatrix} = \begin{pmatrix} I_1^{(n)}(\boldsymbol{\beta}_1^*|\boldsymbol{X}_1) & 0 \\ 0 & I_2^{(n)}(\boldsymbol{\beta}_2^*|\boldsymbol{X}_2) \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix},$$

i.e.,

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n} I_1^{(n)}(\boldsymbol{\beta}_1^*|\boldsymbol{X}_1) & 0 \\ 0 & \frac{1}{n} I_2^{(n)}(\boldsymbol{\beta}_2^*|\boldsymbol{X}_2) \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}} \boldsymbol{U}_1^{(n)}(\boldsymbol{\beta}_1|\boldsymbol{X}_1) \\ \frac{1}{\sqrt{n}} \boldsymbol{U}_2^{(n)}(\boldsymbol{\beta}_2|\boldsymbol{X}_2) \end{pmatrix}.$$

By (iii) in Theorem 2.3.1,

$$\Big[\frac{1}{n} I_k^{(n)}(\boldsymbol{\beta}_k^*|\boldsymbol{X}_k)\Big]^{-1} \to I_k(\boldsymbol{\beta}_k)^{-1}$$

in probability, since $\boldsymbol{\beta}_k^* \to \boldsymbol{\beta}_k$ in probability, $k = 1, 2$. Using this and Proposition 2.5.1, by Slutsky's theorem, we have

$$\sqrt{n}(\hat{\underset{\sim}{\boldsymbol{\beta}}} - \underset{\sim}{\boldsymbol{\beta}}) \Rightarrow N(0, Q).$$

$\square$

To estimate Q, let

$$\hat{Q}_{kl} = \Big[\frac{1}{n} I_k^{(n)}(\hat{\boldsymbol{\beta}}_k|\boldsymbol{X}_k)\Big]^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{u}_{ki}(\hat{\boldsymbol{\beta}}_k)\boldsymbol{u}_{li}(\hat{\boldsymbol{\beta}}_l)^T\right) \Big[\frac{1}{n} I_l^{(n)}(\hat{\boldsymbol{\beta}}_l|\boldsymbol{X}_l)\Big]^{-1}, \tag{2.36}$$

$k, l = 1, 2$, and

$$\hat{Q} = \begin{pmatrix} \hat{Q}_{11} & \hat{Q}_{12} \\ \hat{Q}_{21} & \hat{Q}_{22} \end{pmatrix}.$$

In the next theorem we show that $\hat{Q}$ is a consistent estimate of Q.

Theorem 2.5.3 *Suppose, marginally, each of the failure times satisfies Assumptions (1) - (5) for $k = 1, 2$. Then $\hat{Q}$ is a consistent estimate of Q.*

Proof. Let $\hat{\underline{\boldsymbol{\beta}}}_0$ be the true values of the parameter. From Theorem 2.3.1,

$$\frac{1}{n} I_k^{(n)}(\hat{\boldsymbol{\beta}}_k | \boldsymbol{X}_k) \to I_k(\boldsymbol{\beta}_{k0})$$

in probability, $k = 1, 2$. Therefore, we need only prove that

$$\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{u}_{ki}(\hat{\boldsymbol{\beta}}_k) \boldsymbol{u}_{li}(\hat{\boldsymbol{\beta}}_l)^T \to \Sigma_{kl} \tag{2.37}$$

in probability, $k, l = 1, 2$. By the law of large numbers,

$$\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{u}_{ki}(\boldsymbol{\beta}_{k0}) \boldsymbol{u}_{li}(\boldsymbol{\beta}_{l0})^T \to E(\boldsymbol{u}_{k1}(\boldsymbol{\beta}_{k0}) \boldsymbol{u}_{l1}(\boldsymbol{\beta}_{l0})^T) = \Sigma_{kl},$$

in probability. Hence, to prove (2.37), it is sufficient to prove that

$$\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{u}_{ki}(\hat{\boldsymbol{\beta}}_k) \boldsymbol{u}_{li}(\hat{\boldsymbol{\beta}}_l)^T - \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{u}_{ki}(\boldsymbol{\beta}_{k0}) \boldsymbol{u}_{li}(\boldsymbol{\beta}_{l0})^T \to 0$$

in probability.

Let $u_{kis}(\boldsymbol{\beta}_k)$ be the sth component of $\boldsymbol{u}_{ki}(\boldsymbol{\beta}_k)$, i.e.,

$$u_{kis}(\boldsymbol{\beta}_k) = \frac{\partial}{\partial \beta_{ks}} \log f_{kq}(X_{ki}, \delta_{ki} | \boldsymbol{\beta}_k),$$

$s = 1, \cdots, p_k; k = 1, 2; i = 1, \cdots, n$. We need to prove

$$\frac{1}{n} \sum_{i=1}^{n} u_{kis}(\hat{\boldsymbol{\beta}}_k) u_{lij}(\hat{\boldsymbol{\beta}}_l) - \frac{1}{n} \sum_{i=1}^{n} u_{kis}(\boldsymbol{\beta}_{k0}) u_{lij}(\boldsymbol{\beta}_{l0}) \to 0, \tag{2.38}$$

in probability, for $s = 1, \cdots, p_k; j = 1, \cdots, p_l; k, l = 1, 2$. Using the mean value theorem applied to $\sum_{i=1}^{n} u_{kis}(\hat{\boldsymbol{\beta}}_k) u_{lij}(\hat{\boldsymbol{\beta}}_l)$ and $\sum_{i=1}^{n} u_{kis}(\boldsymbol{\beta}_{k0}) u_{lij}(\boldsymbol{\beta}_{l0})$, we have

$$\left| \frac{1}{n} \sum_{i=1}^{n} u_{kis}(\hat{\boldsymbol{\beta}}_k) u_{lij}(\hat{\boldsymbol{\beta}}_l) - \frac{1}{n} \sum_{i=1}^{n} u_{kis}(\boldsymbol{\beta}_{k0}) u_{lij}(\boldsymbol{\beta}_{l0}) \right|$$

$$
\begin{aligned}
&\le \frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{p_k}\left|\frac{\partial}{\partial \boldsymbol{\beta}_{km}}(u_{kis}(\boldsymbol{\beta}_k)u_{lij}(\boldsymbol{\beta}_l))\Big|_{\underline{\boldsymbol{\beta}}^*}\right| |\hat{\boldsymbol{\beta}}_{km} - \boldsymbol{\beta}_{k0m}| \\
&\qquad + \frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{p_l}\left|\frac{\partial}{\partial \boldsymbol{\beta}_{lm}}(u_{kis}(\boldsymbol{\beta}_k)u_{lij}(\boldsymbol{\beta}_l))\Big|_{\underline{\boldsymbol{\beta}}^*}\right| |\hat{\boldsymbol{\beta}}_{lm} - \boldsymbol{\beta}_{l0m}| \\
&\le \frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{p_k}\left|\frac{\partial}{\partial \boldsymbol{\beta}_{km}}(u_{kis}(\boldsymbol{\beta}_k)u_{lij}(\boldsymbol{\beta}_l))\Big|_{\underline{\boldsymbol{\beta}}^*}\right| \sum_{m=1}^{p_k} |\hat{\boldsymbol{\beta}}_{km} - \boldsymbol{\beta}_{k0m}| \\
&\qquad + \frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{p_l}\left|\frac{\partial}{\partial \boldsymbol{\beta}_{lm}}(u_{kis}(\boldsymbol{\beta}_k)u_{lij}(\boldsymbol{\beta}_l))\Big|_{\underline{\boldsymbol{\beta}}^*}\right| \sum_{m=1}^{p_k} |\hat{\boldsymbol{\beta}}_{lm} - \boldsymbol{\beta}_{l0m}|
\end{aligned}
$$

where $\underline{\boldsymbol{\beta}}^*$ is on a line segment between $\hat{\underline{\boldsymbol{\beta}}}$ and $\underline{\boldsymbol{\beta}}_0$. Thus, since $\hat{\underline{\boldsymbol{\beta}}} \to \underline{\boldsymbol{\beta}}_0$ in probability, to prove (2.38), it suffices to show that $\frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{p_k}\left|\frac{\partial}{\partial \boldsymbol{\beta}_{km}}(u_{kis}(\boldsymbol{\beta}_k)u_{lij}(\boldsymbol{\beta}_l))\right|$ is bounded in probability around $\underline{\boldsymbol{\beta}}_0$. That means we need to find $M < \infty$ and a neighborhood of $\underline{\boldsymbol{\beta}}_0$ such that

$$
P\left\{\sup_{\beta\in B}\frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{p_k}\left|\frac{\partial}{\partial \boldsymbol{\beta}_{km}}(u_{kis}(\boldsymbol{\beta}_k)u_{lij}(\boldsymbol{\beta}_l))\right| < M\right\} \to 1
$$

when $n \to \infty$, $k, l = 1, 2$.

We choose a number $a > 0$ such that $B = \{\underline{\boldsymbol{\beta}} : \|\underline{\boldsymbol{\beta}} - \underline{\boldsymbol{\beta}}_0\| \le a\}$ is a compact set in the parameter space. By Lemma 2.3.4, there exist M_k and M_l such that for $\underline{\boldsymbol{\beta}} \in B, k, l = 1, 2, k \neq l, m = 1, \cdots, p_k$,

$$
\begin{aligned}
&\left|\frac{\partial}{\partial \boldsymbol{\beta}_{km}}(u_{kis}(\boldsymbol{\beta}_k)u_{lij}(\boldsymbol{\beta}_l))\right| = \left|u_{lij}(\boldsymbol{\beta}_l)\frac{\partial}{\partial \boldsymbol{\beta}_{km}}u_{kis}(\boldsymbol{\beta}_k)\right| \\
= &\left|\Big(\frac{\partial}{\partial \boldsymbol{\beta}_{lj}}\log f_{lq}(X_{li}, \delta_{li}|\boldsymbol{\beta}_l)\frac{\partial^2}{\partial \beta_{ks}\partial \boldsymbol{\beta}_{km}}\log f_{kq}(X_{ki}, \delta_{ki}|\boldsymbol{\beta}_k)\Big)\right| \\
\le &\ M_k M_l K_k(X_{ki})K_l(X_{li}).
\end{aligned}
$$

For $k = l$,

$$
\left|\frac{\partial}{\partial \boldsymbol{\beta}_{km}}(u_{kis}(\boldsymbol{\beta}_k)u_{lij}(\boldsymbol{\beta}_l))\right| \le 2M_k^2[K_k(X_{ki})]^2.
$$

Thus,

$$
\begin{aligned}
& P\left\{\sup_{\beta\in B}\frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{p_k}\left|\frac{\partial}{\partial\boldsymbol{\beta}_{km}}(u_{kis}(\boldsymbol{\beta}_k)u_{lij}(\boldsymbol{\beta}_l))\right| < M\right\} \\
\geq \quad & P\left\{2p_k M_k M_l \frac{1}{n}\sum_{i=1}^{n} K_k(X_{ki})K_l(X_{li}) \leq M\right\}
\end{aligned}
$$

which will converge to 1 for M sufficiently large, since

$$\frac{1}{n}\sum_{i=1}^{n} K_k(X_{ki})K_l(X_{li}) \to E(K_k(X_{ki})K_l(X_{li})) < \infty$$

by Proposition 2.4.3. □

As in the univariate case, we have the following Corollary.

Corollary 2.5.4 *If* $\boldsymbol{b}_1$ *and* $\boldsymbol{b}_2$ *satisfy assumptions 1 through 5, then*

$$\sqrt{n}\hat{Q}^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow N(\mathbf{0}, I),$$

where I *is* $(p_1 + p_2) \times (p_1 + p_2)$ *identity matrix.*