บทที่ 2
ทฤษฎีและสถิติที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงรายละเอียดของการวิเคราะห์ความถดถอยโลจิสติคทั้งแบบสถิตย์และแบบพลวัตร การประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) ฟังก์ชันอัตราภาวะภัย (Hazard Function) ฟังก์ชันการอยู่รอด (Survival Function) ซึ่งฟังก์ชันเหล่านี้เกี่ยวข้องกับตัวแบบถดถอยโลจิสติคที่จะทำการศึกษาในการวิจัยครั้งนี้

ข้อสมมติทั่วไปของตัวแบบถดถอยโลจิสติค

จุดประสงค์ของการวิเคราะห์ความถดถอยโลจิสติค เพื่อสร้างตัวแบบถดถอยโลจิสติคที่ใช้อธิบายความสัมพันธ์ระหว่างตัวแปรตามซึ่งเป็นตัวแปรเชิงกลุ่มและกลุ่มของตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงกลุ่มหรือทั้งสองอย่าง และพยากรณ์ความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจหรือโอกาสที่หน่วยตัวอย่างจะอยู่ในกลุ่มที่สนใจ

ข้อมูลสำหรับการวิเคราะห์จะอยู่ในรูป $(y_i, \mathbf{x}_i)\,; i = 1,2,\ldots,n$ เมื่อตัวแปรตาม Y_i มีการแจกแจงแบบเบอร์นูลีด้วยพารามิเตอร์ p_i $\mathbf{x}_i$ แทน $x_{1i}, x_{2i}, \ldots, x_{pi}$ เป็นตัวแปรอิสระร่วมกัน p ตัวแปร และข้อมูลตัวอย่าง n หน่วยเป็นอิสระต่อกัน นั่นคือ

$$y_i = \begin{cases} 1 & ;\ \text{เกิดเหตุการณ์ที่สนใจ ด้วยความน่าจะเป็น}\ \ p_i \\ 0 & ;\ \text{ไม่เกิดเหตุการณ์ที่สนใจ ด้วยความน่าจะเป็น}\ \ 1 - p_i \end{cases}$$

p_i คือความน่าจะเป็นแบบมีเงื่อนไขของ $Y = 1$ เมื่อกำหนด $\mathbf{x}$ ณ ระดับที่ i นั่นคือ $p_i = P(Y_i = 1 \mid \mathbf{x}_i)$ และจะเขียนแทนด้วย $\pi(\mathbf{x}_i)$ เพราะเป็นฟังก์ชันของ $\mathbf{x}_i$ ดังนั้นสมการถดถอยโลจิสติค (Logistic Response Function) คือ

$$E[Y \mid X] = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}}} = \pi(\mathbf{x}_i) \qquad (2.1)$$

จะทำการแปลงค่า $\pi(\mathbf{x}_i)$ จากช่วง (0,1) เป็นค่าของ $\operatorname{logit}(\pi(\mathbf{x}_i))$ ที่อยู่ในช่วง $(-\infty, \infty)$ ซึ่งฟังก์ชันการแปลงโลจิสติคของ $\pi(\mathbf{x}_i)$ คือ

$$\operatorname{logit}(\pi(\mathbf{x}_i)) = \ln\left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} \qquad (2.2)$$

และ Odd Ratio ซึ่งเป็นอัตราส่วนระหว่างโอกาสที่เหตุการณ์จะเกิดกับโอกาสที่เหตุการณ์จะไม่เกิด คือ

$$Odd = \left(\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)} \right) = e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\ldots+\beta_p x_{pi}} \tag{2.3}$$

เราสามารถเขียนสมการ (2.2) ในรูปของเมทริกซ์ ได้ดังนี้

$$\mathbf{G}(\mathbf{X}) = \mathbf{X\beta} \tag{2.4}$$

โดยที่

$$\mathbf{G}(\mathbf{X}) = \begin{bmatrix} \log it(\pi(\mathbf{x}_1)) \\ \log it(\pi(\mathbf{x}_2)) \\ \vdots \\ \log it(\pi(\mathbf{x}_n)) \end{bmatrix}_{n\times 1}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix}_{n\times(p+1)}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1}$$

ฟังก์ชันภาวะน่าจะเป็น (Likelihood Function) ของความถดถอยโลจิสติค

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} (\pi(\mathbf{x}_i))^{y_i} (1-\pi(\mathbf{x}_i))^{1-y_i} \tag{2.5}$$

$$\begin{aligned} L(\boldsymbol{\beta}) &= \ln l(\boldsymbol{\beta}) \\ &= \sum_{i=1}^{n} \{y_i \ln(\pi(\mathbf{x}_i)) + (1-y_i)\ln(1-\pi(\mathbf{x}_i))\} \\ &= \sum_{i=1}^{n} \{y_i \ln(\pi(\mathbf{x}_i)) + \ln(1-\pi(\mathbf{x}_i)) - y_i \ln(1-\pi(\mathbf{x}_i))\} \end{aligned}$$

$$= \sum_{i=1}^{n} \{ y_i \{ \ln(\pi(\mathbf{x}_i)) - \ln(1-\pi(\mathbf{x}_i)) \} + \ln(1-\pi(\mathbf{x}_i)) \}$$

$$= \sum_{i=1}^{n} \left\{ y_i \ln\left(\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)} \right) + \ln(1-\pi(\mathbf{x}_i)) \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i \ln\left(\frac{e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}} \cdot \frac{1}{1-\dfrac{e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}}} \right) + \ln\left(1 - \frac{e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}} \right) \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i \ln\left(e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}} \right) + \ln\left(\frac{1}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}} \right) \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i \ln\left(e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}} \right) + \ln 1 - \ln\left(1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}} \right) \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i \left(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} \right) - \ln\left(1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}} \right) \right\} \tag{2.6}$$

การประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด

สำหรับการประมาณค่าพารามิเตอร์หรือค่าสัมประสิทธิ์ความถดถอยในสมการถดถอยโลจิสติค $\boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots, \beta_p)$ จะใช้วิธีภาวะน่าจะเป็นสูงสุด โดยการหาตัวประมาณพารามิเตอร์ที่ทำให้ $L(\boldsymbol{\beta})$ ในสมการ (2.6) มีค่ามากที่สุด โดยการหาอนุพันธ์เทียบกับพารามิเตอร์แต่ละตัวแล้วให้ผลลัพธ์เท่ากับศูนย์ ซึ่งจะได้สมการที่ไม่เป็นเส้นตรง $p+1$ สมการ และสามารถหาค่าประมาณพารามิเตอร์ $\hat{\boldsymbol{\beta}}' = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)$ ที่สอดคล้องได้ด้วยวิธีทำซ้ำ (Iterative Method) ซึ่งเรียกตัวประมาณนี้ว่า "ตัวประมาณภาวะน่าจะเป็นสูงสุด" (Maximum Likelihood Estimator : MLE)

จากสมการ (2.6) เมื่อหาอนุพันธ์เทียบกับพารามิเตอร์ $\beta_0, \beta_1, \ldots, \beta_p$ จะได้ว่า

$$\left. \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \sum_{i=1}^{n} \left\{ y_i - \left(\frac{1}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}} \cdot e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}} \right) \right\}$$

$$= \sum_{i=1}^{n} (y_i - \pi(\mathbf{x}_i)) = 0$$

$$\left. \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \sum_{i=1}^{n} \left\{ x_{1i} y_i - \left(\frac{1}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}} \cdot e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}} (x_{1i}) \right) \right\}$$

$$= \sum_{i=1}^{n} x_{1i} (y_i - \pi(\mathbf{x}_i)) = 0$$

$$\vdots$$

$$\left.\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_p}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \sum_{i=1}^{n}\left\{x_{pi}y_i - \left(\frac{1}{1+e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}} \cdot e^{\beta_0+\beta_1 x_{1i}+\ldots+\beta_p x_{pi}}(x_{pi})\right)\right\}$$

$$= \sum_{i=1}^{n} x_{pi}\left(y_i - \pi(\mathbf{x}_i)\right) = 0 \tag{2.7}$$

ซึ่งจะสามารถหาค่าประมาณพารามิเตอร์ $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ ที่สอดคล้องกับสมการ (2.7) ได้ด้วยวิธีทำซ้ำ

ตัวแบบความถดถอยโลจิสติคแบบสถิตย์และแบบพลวัตรจะมีสมการอยู่ในรูปแบบเดียวกัน และการประมาณค่าสัมประสิทธิ์ความถดถอยในสมการจะใช้วิธีภาวะน่าจะเป็นสูงสุดเช่นเดียวกัน สำหรับส่วนที่แตกต่างกันของตัวแบบทั้งสอง คือ การพิจารณาข้อมูลสำหรับใช้ในการประมาณค่าพารามิเตอร์ในสมการซึ่งจะแสดงอยู่ใน Likelihood Function ดังจะกล่าวต่อไปนี้ Shumway, Tyler. (2001)

การวิเคราะห์ความถดถอยโลจิสติคแบบสถิตย์ (Static Logistic Regression Analysis)

การวิเคราะห์ความถดถอยโลจิสติคแบบสถิตย์หรือการวิเคราะห์แบบปกติ เป็นการพิจารณาข้อมูลของแต่ละหน่วยตัวอย่างเพียง 1 ช่วงเวลาหรือ 1 ชุดสำหรับตัวแปรต่างๆ ที่เกี่ยวข้อง ที่จะใช้ในการประมาณค่าพารามิเตอร์ $\boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots, \beta_p)$ ในตัวแบบ โดยจะพิจารณาข้อมูลของตัวแปรอิสระที่จุดเริ่มต้นของช่วงเวลาที่ศึกษาและพิจารณาเหตุการณ์หรือค่าสังเกตของตัวแปรตามที่จุดสิ้นสุดของการศึกษา ดังนั้นเราจะมีจำนวนข้อมูลทั้งหมดเท่ากับ n ชุดสำหรับ n หน่วยตัวอย่าง ซึ่งจะได้ Likelihood Function และ Log Likelihood Function ของความถดถอยโลจิสติคแบบสถิตย์สำหรับช่วงเวลา $t = 1,2,\ldots,T$ ดังนี้

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} F(t_i, \mathbf{x}_i; \boldsymbol{\beta})^{y_i}\left[1 - F(t_i, \mathbf{x}_i; \boldsymbol{\beta})\right]^{1-y_i}$$

$$\begin{aligned} L(\boldsymbol{\beta}) &= \ln l(\boldsymbol{\beta}) \\ &= \sum_{i=1}^{n} y_{i1}\ln\left[P(y_{i1}=1 \mid \mathbf{x}_{i1}, \boldsymbol{\beta})\right] \\ &= \quad + y_{i2}\ln\left[P(y_{i2}=1 \mid \mathbf{x}_{i2}, \boldsymbol{\beta})\right] \\ &= \quad + \ldots + y_{iT}\ln\left[P(y_{iT}=1 \mid \mathbf{x}_{iT}, \boldsymbol{\beta})\right] \\ &= \quad + (1-y_{iT})\ln\left[P(y_{iT}=0 \mid \mathbf{x}_{iT}, \boldsymbol{\beta})\right] \end{aligned}$$

จาก Log likelihood function จะเห็นว่าไม่มีการพิจารณาหรือสนใจความน่าจะเป็นที่หน่วยตัวอย่างไม่เกิดเหตุการณ์ในช่วงเวลาก่อนที่จะเกิดเหตุการณ์ขึ้น

การวิเคราะห์ความถดถอยโลจิสติคแบบพลวัตร (Dynamic Logistic Regression Analysis)

การวิเคราะห์ความถดถอยโลจิสติคแบบพลวัตร จะเป็นการพิจารณาข้อมูลของตัวอย่างในหลายช่วงเวลา เพื่อใช้การประมาณค่าพารามิเตอร์ $\boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots, \beta_p)$ ในสมการ โดยจะทำการแบ่งช่วงเวลาของการศึกษาออกเป็นหลายช่วงเวลาย่อยและพิจารณาค่าของตัวแปรต่างๆ ที่เปลี่ยนแปลงไปในแต่ละช่วงเวลาย่อย นั่นคือ แต่ละหน่วยตัวอย่างจะมีข้อมูลของตัวแปรต่างๆ ที่เกี่ยวข้องสำหรับหลายช่วงเวลา Likelihood Function และ Log Likelihood Function สำหรับช่วงเวลา $t = 1,2,\ldots,T$ ของความถดถอยโลจิสติคแบบพลวัตรจะเกี่ยวข้องกับ Survival function และ Hazard function โดยมีรูปแบบดังนี้ [1]

$$\begin{aligned} l(\boldsymbol{\beta}) &= \prod_{i=1}^{n}\left(F(t_i, \mathbf{x}_i; \boldsymbol{\beta})^{y_i} \prod_{j<t_i}[1 - F(j, \mathbf{x}_i; \boldsymbol{\beta})]\right) \\ &= \prod_{i=1}^{n}\left(\lambda(t_i, \mathbf{x}_i; \boldsymbol{\beta})^{y_i} \prod_{j<t_i}[1 - \lambda(j, \mathbf{x}_i; \boldsymbol{\beta})]\right) \\ &= \prod_{i=1}^{n} \lambda(t_i, \mathbf{x}_i; \boldsymbol{\beta})^{y_i} S(t_i, \mathbf{x}_i; \boldsymbol{\beta}) \end{aligned} \tag{2.8}$$

$$\begin{aligned} L(\boldsymbol{\beta}) &= \ln l(\boldsymbol{\beta}) \\ &= \sum_{i=1}^{n}\left(y_{it} \ln[P(y_{it} = 1 \mid \mathbf{x}_{it}, \boldsymbol{\beta})] + \sum_{j<t_i} \ln[P(y_{ij} = 0 \mid \mathbf{x}_{ij}, \boldsymbol{\beta})]\right) \end{aligned} \tag{2.9}$$

การประมาณค่าพารามิเตอร์ในสมการถดถอยโลจิสติคแบบพลวัตรจะใช้วิธีภาวะน่าจะเป็นสูงสุดเช่นเดียวกับตัวแบบสถิตย์ แต่สถิติของการทดสอบที่ได้จากการวิเคราะห์จะยังไม่ถูกต้องสำหรับตัวแบบพลวัตร เพราะข้อมูลที่ใช้ในการวิเคราะห์แบบพลวัตรเป็นข้อมูลที่เกิดขึ้นของหน่วยตัวอย่างในหลายช่วงเวลาหรือมีค่าสังเกตหลายค่าในหน่วยตัวอย่างเดียว ซึ่งทำให้ข้อมูลมีความบกพร่องในคุณสมบัติของความเป็นอิสระต่อกันระหว่างช่วงเวลาย่อย เพราะหน่วยตัวอย่างที่เกิดเหตุการณ์ในช่วงเวลา $t-1$ แล้ว จะไม่สามารถเกิดเหตุการณ์ในช่วงเวลา t ได้อีก เช่นเดียวกันกับหน่วยตัวอย่างที่มีชีวิตอยู่หรือไม่เกิดเหตุการณ์ในช่วงเวลา t ก็ไม่สามารถเกิดเหตุการณ์ในช่วงเวลา $t-1$ ได้ จึงทำให้ค่าสังเกตไม่เป็นอิสระต่อกัน Shumway, Tyler. (2001) ได้แนะนำถึงวิธีการแก้ปัญหานี้ โดยจะต้องทำการปรับแก้ค่าสถิติทดสอบ (Likelihood ratio test) ซึ่งประมาณด้วยการ

แจกแจงไคสแควร์ (χ^2) โดยการหารด้วยจำนวนเฉลี่ยของจำนวนตัวอย่างที่ใช้ในการวิเคราะห์แบบพลวัตร (จำนวนค่าสังเกตในหลายช่วงเวลา) ต่อจำนวนตัวอย่างที่ใช้ในการวิเคราะห์แบบสถิตย์ (จำนวนตัวอย่างจริงๆ) ดังนี้

$$\chi^2 \Big/ \left(n_{Dynamic} / n_{Static} \right) \tag{2.10}$$

เมื่อ $[-2LL_0]-[-2LL_1] \sim \chi^2_p$

$-2LL_0 = -2 \times \log likelihood$ ของฟังก์ชันที่มีเพียงค่าคงที่

$-2LL_1 = -2 \times \log likelihood$ ของฟังก์ชันที่มีตัวแปรอิสระจำนวน p ตัว

ซึ่งจากการปรับจำนวนตัวอย่าง n ในตัวแบบพลวัตรจึงเป็นการเหมาะสมแล้ว ดังนั้นจึงไม่จำเป็นต้องมีข้อสมมติที่ว่าตัวอย่างที่ใช้ในการวิเคราะห์ต้องเป็นอิสระต่อกัน

Survival Function หรือ ฟังก์ชันการอยู่รอด

เป็นฟังก์ชันความน่าจะเป็นที่ค่าสังเกตจะอยู่รอด (ไม่เกิดเหตุการณ์) ต่อไปจากช่วงเวลา t โดยที่อยู่รอดมาแล้วจนถึงช่วงเวลา t มีรูปแบบสมการดังนี้

$$S(t,\mathbf{x};\boldsymbol{\beta}) = P(T \geq t) = 1 - \sum_{j<t} f(j,\mathbf{x};\boldsymbol{\beta}) = 1 - F(j,\mathbf{x};\boldsymbol{\beta}) \tag{2.11}$$

สำหรับช่วงเวลา $t = 1,2,\ldots$

โดยที่ $f(t,\mathbf{x};\boldsymbol{\beta})$ เป็นฟังก์ชันความน่าจะเป็น (Probability Function) ที่ช่วงเวลา t

$F(t,\mathbf{x};\boldsymbol{\beta})$ เป็นฟังก์ชันการแจกแจงสะสม (Cumulative Distribution Function) ที่ช่วงเวลา t นั่นคือ

$F(t,\mathbf{x};\boldsymbol{\beta}) = P(T \leq t)$

$S(t,\mathbf{x};\boldsymbol{\beta})$ มีคุณสมบัติดังนี้

1. เป็นฟังก์ชันไม่เพิ่ม (Nonincreasing function)
2. เป็นฟังก์ชันต่อเนื่องของช่วงเวลา t
3. $S(t,\mathbf{x};\boldsymbol{\beta}) = 1$ เมื่อ $t = 0$ และ $S(t,\mathbf{x};\boldsymbol{\beta}) = 0$ เมื่อ $t = \infty$

Hazard function หรือ ฟังก์ชันอัตราภาวะภัย

เป็นอัตราส่วนความน่าจะเป็นที่ค่าสังเกตจะสูญเสีย (เกิดเหตุการณ์) ในช่วงเวลาสั้นๆ t เมื่อค่าสังเกตมีการอยู่รอดมาแล้วจนถึงเวลา t มีรูปแบบสมการดังนี้

$$\lambda(t,\mathbf{x};\boldsymbol{\beta}) = \frac{f(t,\mathbf{x};\boldsymbol{\beta})}{S(t,\mathbf{x};\boldsymbol{\beta})} \tag{2.12}$$

ซึ่ง $\lambda(t,\mathbf{x};\boldsymbol{\beta}) \geq 0$ เมื่อ $\boldsymbol{\theta} >$ และ $S(t,\mathbf{x};\boldsymbol{\beta}) = \prod_{j<t}[1-\lambda(j,\mathbf{x};\boldsymbol{\beta})]$

และจะได้ว่า $\lambda(t,\mathbf{x};\boldsymbol{\beta}) = F(t,\mathbf{x};\boldsymbol{\beta})$ สำหรับการความถดถอยโลจิสติคที่ช่วงเวลาไม่ต่อเนื่อง

ในการวิจัยครั้งนี้จะทำการศึกษากรณี Hazard function หรือ Hazard model สำหรับช่วงเวลาไม่ต่อเนื่องและตัวแปรอิสระ X ที่มีค่าเปลี่ยนไปตามเวลานั้นมาจากการแจกแจงแบบเอกซโพเนนเชียล (Exponential Distribution) ด้วยพารามิเตอร์ λ ซึ่งมีรูปแบบดังนี้

$$f(t,x;\lambda) = \lambda e^{-\lambda t} \quad ;\lambda > 0, t \geq 0$$
$$F(t,x;\lambda) = 1 - e^{-\lambda t}$$
$$S(t,x;\lambda) = e^{-\lambda t}$$
$$\lambda(t,x;\lambda) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

จะได้ค่าเฉลี่ยและความแปรปรวนดังนี้

$$E[X] = \frac{1}{\lambda} \quad ; \quad V[X] = \frac{1}{\lambda^2}$$

ซึ่งจะพบว่า $\lambda(t,\mathbf{x};\boldsymbol{\beta})$ เมื่อ X มีการแจกแจงแบบเอกซโพเนนเชียลจะไม่ขึ้นอยู่กับเวลา t นั่นคือ ในแต่ละช่วงเวลาที่เปลี่ยนไปอัตราการสูญเสีย $\lambda(t,\mathbf{x};\boldsymbol{\beta})$ จะยังคงที่

การประมาณค่า Hazard Function และ Survival Function

สำหรับช่วงเวลาไม่ต่อเนื่อง $t = 1,2,\ldots,T$ การประมาณค่า $\lambda(t,\mathbf{x};\boldsymbol{\beta})$ และ $S(t,\mathbf{x};\boldsymbol{\beta})$ ในแต่ละช่วงเวลา t สามารถคำนวณได้จากสมการดังต่อไปนี้ [2]

$$\hat{\lambda}(t,\mathbf{x};\boldsymbol{\beta}) = \frac{h_t}{n_t} \text{ และ } \hat{S}(t,\mathbf{x};\boldsymbol{\beta}) = \prod_{t=1}^{t}\frac{(n_t - h_t)}{n_t} = \prod_{t=1}^{t}\left[1-\hat{\lambda}(t,\mathbf{x};\boldsymbol{\beta})\right] \tag{2.13}$$

เมื่อ h_t คือ จำนวนหน่วยตัวอย่างที่เกิดเหตุการณ์ที่สนใจหรือสูญเสียในช่วงเวลา t

n_t คือ จำนวนหน่วยตัวอย่างที่ถูกสังเกตหรือหน่วยตัวอย่างที่เสี่ยงจะเกิดเหตุการณ์ที่สนใจในช่วงเวลา t ซึ่งมีค่าเท่ากับผลรวมของจำนวนหน่วยตัวอย่างที่เกิดเหตุการณ์และไม่เกิดเหตุการณ์ในช่วงเวลา t