深度生存分析

初稿 第 1 版

1　生存分析的统计学基础

在许多应用中，预测关键事件发生之前的时间持续性被称为“时间到事件”结果。这类预测对决策至关重要。例如，电子商务公司可能希望预测用户多久进行一次购买；在医疗健康领域，医院可能对预测患者何时复发感兴趣。准确预测这些时间到事件的结果，有助于在电子商务平台上展示定向广告或促销活动，或者在医疗健康领域根据预测结果合理规划治疗，从而降低患者复发的风险。

生存分析是一种专门用于研究时间到事件数据的统计方法，尤其适用于预测某一事件（如疾病的发生、死亡或设备故障等）发生的时间。与传统回归分析不同，生存分析不仅关注事件是否发生，更加注重事件发生的时间长短。生存分析的核心概念包括生存函数和风险函数（后文将详细定义这两者）：生存函数表示在某一时间点之后仍然存活的概率，而风险函数则描述某一时刻发生事件的风险。

在实际应用中，时间到事件预测的一个显著特点是，通常在收集训练数据以学习生存分析模型时，并不能看到每个数据点的真实事件发生时间，也就是说，它们的事件发生时间是被删失的。例如，某些患者在研究期间未发生死亡事件，因此其死亡时间数据被删失，无法观察到事件的真实发生时间。这种删失现象可能会对生存分析的结果产生影响。

生存分析在临床研究中尤为重要。通过患者的生存时间及其相关因素，生存分析不仅能够帮助预测疾病的进展、评估治疗效果，还能为公共卫生政策提供数据支持。此外，生存分析广泛应用于评估产品耐用性或设备故障率，成为跨学科的强大工具，广泛服务于各个领域，从而提升各行业决策的准确性和效率。

近年来，随着深度学习技术的发展，研究人员开始将深度学习与传统生存分析方法相结合，以提高时间到事件预测的准确性和效率。

在本章中，我们假设读者熟练掌握微积分、线性代数、概率论和数理统计的基本知识，以及机器学习尤其是神经网络中的基本概念。我们将从深度学习的视角重新审视经典的生存分析方法。

1.1 问题描述

在生存分析中，我们有一组样本 $\{(X_1, Y_1, \Delta_1), \ldots, (X_n, Y_n, \Delta_n)\}$，每个样本 i 由以下三部分组成：

1. $X_i \in \mathcal{X}$ 是输入变量（例如一张图片或一段文本等），
2. $Y_i \in [0, \infty)$ 是观察到的时间（可能是实际的生存时间，也可能是删失时间），
3. $\Delta_i \in \{0, 1\}$ 是事件指示符：若 $\Delta_i = 1$，表示事件已发生，Y_i 是实际生存时间；若 $\Delta_i = 0$，表示事件未发生，Y_i 为删失时间（即我们最后观察到的时间）。

我们的目标是建立输入 X 与生存时间或删失时间之间的关系模型。

观测数据的表示可以简化为以下形式：

$$Y_i = \min(T_i, C_i), \tag{1.1}$$

其中 T_i 表示真实的生存时间，C_i 表示真实的删失时间。事件指示符 Δ_i 为：

$$\Delta_i = 1\{T_i \leq C_i\}, \tag{1.2}$$

其中 $1\{\cdot\}$ 是指示函数，当其参数为真时取值为 1，否则取值为 0。

因此，对于任意给定的输入 X，其真实生存时间为 T，真实删失时间为 C，我们得到以下观测值：

$$Y = \min(T, C), \tag{1.3}$$

事件指示符为：

$$\Delta = 1\{T \leq C\}. \tag{1.4}$$

这一统计框架被称为右删失数据模型（即当 $\Delta_i = 0$ 时，表示事件未发生）。在这种模型中，真实的生存时间 T_i 大于观察到的删失时间 C_i。此框架假设删失是独立的，也就是说，在给定输入 X 的条件下，生存时间与删失时间是相互独立的。

通过建立合适的生存分析模型，我们可以从这些观测数据中推断出生存或删失的规律，并对未来的事件进行预测。

1.2 连续时间模型

对于输入 $x \in X \subseteq \mathbb{R}^d$，我们假设在条件 $X = x$ 下，生存时间 T 是一个连续随机变量，其概率密度函数为 $f(t|x)$，累积分布函数为 $F(t|x) = \int_0^t f(u|x)\,du$；这两个函数中的任意一个都完全表征了分布 $P_{T|X}(\cdot|x)$。

条件生存函数定义为：

$$S(t|x) := P(\text{生存超过时间 } t|\text{原始输入为 } x) = P(T > t|X = x), \tag{1.5}$$

其中 $t \geq 0$ 且 $x \in X$。在本章中，我们将条件生存函数 $S(\cdot|x)$ 简称为生存函数。

根据上述定义，有：

$$S(t|x) = P(T > t|X = x) = 1 - P(T \leq t|X = x) = 1 - F(t|x). \tag{1.6}$$

所以，生存函数具备以下性质：

1. $S(\cdot|x) = 1 - F(\cdot|x)$ 是单调递减的，从 1 递减到 0，因为累积分布函数是单调递增的，从 0 递增到 1。
2. 估计函数 $S(\cdot|x)$ 等价于估计累计分布函数 $F(\cdot|x)$，这意味着我们的目标是估计条件生存时间分布 $P_{T|X}(\cdot|x)$。。

需要注意的是，不同的生存分析预测模型对生存函数 $S(\cdot|x)$ 做出不同的假设，通常我们预测的是生存函数的变换形式，而不是直接预测生存函数 $S(\cdot|x)$。

风险函数的定义如下：

$$h(t|x) := -\frac{d}{dt}\log S(t|x), \tag{1.7}$$

由此可得：

$$h(t|x) = -\frac{d}{dt}\log S(t|x) = -\frac{d}{dt}S(t|x)\cdot\frac{1}{S(t|x)} = -\frac{d}{dt}[1 - F(t|x)]\cdot\frac{1}{S(t|x)} = \frac{f(t|x)}{S(t|x)}, \tag{1.8}$$

其中，$f(\cdot|x)$ 是分布 $P_{T|X}(\cdot|x)$ 的概率密度函数。因此，风险函数是非负的，并且可能具有任意大的正值。

如果已知风险函数 $h(\cdot|x)$，则可以通过以下方式得到生存函数 $S(\cdot|x)$：

$$h(t|x) = -\frac{d}{dt}\log S(t|x) \Rightarrow \int_0^t h(u|x)\,du = -\log S(t|x) \Rightarrow S(t|x) = \exp\left(-\int_0^t h(u|x)\,du\right). \tag{1.9}$$

累积风险函数定义为：

$$H(t|x) := \int_0^t h(u|x)\,du. \tag{1.10}$$

从上面的公式中可以看出，$S(t|x) = \exp(-H(t|x))$。因此，如果我们知道 $H(t|x)$，就能得到 $S(t|x)$。进一步地，从 $h(t|x) = \frac{d}{dt}H(t|x)$ 可得，如果我们知道 $H(t|x)$，也可以得到 $h(t|x)$。需要注意的是，虽然 $S(t|x)$ 和 $H(t|x)$ 是单调函数，但 $h(t|x)$ 不一定是单调的。

我们总结一下概率密度函数 $f(\cdot|x)$、累积分布函数 $F(\cdot|x)$、生存函数 $S(\cdot|x)$、风险函数 $h(\cdot|x)$ 和累积风险函数 $H(\cdot|x)$ 之间的关系：

$$\begin{aligned}
&\text{KRA1:}\ f(t|x) = \frac{d}{dt}F(t|x) = \frac{d}{dt}(1 - S(t|x)) = h(t|x)S(t|x),\\
&\text{KRA2:}\ F(t|x) = \int_0^t f(u|x)\,du = 1 - S(t|x),\\
&\text{KRA3:}\ S(t|x) = 1 - F(t|x) = \int_t^\infty f(u|x)\,du = e^{-H(t|x)} = e^{-\int_0^t h(u|x)\,du},\\
&\text{KRA4:}\ h(t|x) = \frac{d}{dt}H(t|x) = -\frac{d}{dt}\log S(t|x) = \frac{f(t|x)}{S(t|x)},\\
&\text{KRA5:}\ H(t|x) = -\log S(t|x) = \int_0^t h(u|x)\,du.
\end{aligned} \tag{1.11}$$

接下来，假设 $g(Y_i;\varphi)$ 表示删失过程的概率密度函数，$G(Y_i;\varphi)$ 是删失过程的生存函数，且假设 T_i 和 C_i 是独立事件。我们可以推导出以下的似然函数：

$$\begin{aligned}
L &= \prod_{i=1}^n P(T_i \in [Y_i, Y_i + \Delta t_1), C_i > Y_i)^{\Delta_i} \cdot P(T_i > Y_i, C_i \in [Y_i, Y_i + \Delta t_2))^{1-\Delta_i}\\
&= \prod_{i=1}^n [f(Y_i;\theta)\Delta t_1 G(Y_i;\varphi)]^{\Delta_i} \cdot [S(Y_i;\theta)g(Y_i;\varphi)\Delta t_2]^{1-\Delta_i}\\
&= \prod_{i=1}^n [f(Y_i;\theta)]^{\Delta_i}[S(Y_i;\theta)]^{1-\Delta_i}[\Delta t_1 G(Y_i;\varphi)]^{\Delta_i}[g(Y_i;\varphi)\Delta t_2]^{1-\Delta_i}.
\end{aligned} \tag{1.12}$$

所以，似然函数可表示为：$L \propto \prod_{i=1}^n [f(Y_i;\theta)]^{\Delta_i}[S(Y_i;\theta)]^{1-\Delta_i}$。进而，我们重新定义似然函数 L：

$$L := \prod_{i=1}^n \left[f(Y_i|X_i)^{\Delta_i} S(Y_i|X_i)^{1-\Delta_i}\right]. \tag{1.13}$$

通过使用关系 (1.11)，可以将其重写为：

$$
\begin{aligned}
L &= \prod_{i=1}^{n} \left[f(Y_i|X_i)^{\Delta_i} S(Y_i|X_i)^{1-\Delta_i} \right] \\
&\overset{\text{KRA1}}{=} \prod_{i=1}^{n} \left[h(Y_i|X_i)^{\Delta_i} S(Y_i|X_i)^{\Delta_i} \right] \overset{\text{KRA3}}{=} \prod_{i=1}^{n} \left[h(Y_i|X_i)^{\Delta_i} \exp\left(-\int_0^{Y_i} h(u|X_i) du \right) \right].
\end{aligned} \tag{1.14}
$$

进一步，将 $h(t|x) = h(t|x;\theta)$ 代入上式，我们得到：

$$
L(\theta) = \prod_{i=1}^{n} \left[h(Y_i|X_i;\theta)^{\Delta_i} \exp\left(-\int_0^{Y_i} h(u|X_i;\theta) du \right) \right]. \tag{1.15}
$$

其中θ 为生存分析模型中风险函数的参数。对似然函数 (1.15) 取对数，得到：

$$
\begin{aligned}
\log L(\theta) &= \log\left(\prod_{i=1}^{n} \left[h(Y_i|X_i;\theta)^{\Delta_i} \exp\left(-\int_0^{Y_i} h(u|X_i;\theta) du \right) \right] \right) \\
&= \sum_{i=1}^{n} \left[\Delta_i \log h(Y_i|X_i;\theta) - \int_0^{Y_i} h(u|X_i;\theta) du \right].
\end{aligned} \tag{1.16}
$$

在经典统计学中，通常通过极大化对数似然函数 $\max_\theta \log L(\theta)$ 来估计统计学参数 θ. 然而, 在机器学习中，通常通过极小化负对数似然函数：

$$
L_{\text{HNLL}}(\theta) := -\frac{1}{n} \log L(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \left[\Delta_i \log h(Y_i|X_i;\theta) - \int_0^{Y_i} h(u|X_i;\theta) du \right]. \tag{1.17}
$$

将其作为神经网络的损失函数。极大化对数似然函数与极小化负对数似然函数是等价的。

下面，我们以风险比例模型为例，展示如何构造风险比例模型的负对数似然函数。在比例风险模型中，假设风险函数可以分解为如下形式：

$$
h(t|x) = h_0(t;\theta) e^{f(x;\theta)} \quad \text{对于} \quad t \geq 0, x \in X, \tag{1.18}
$$

其中 $h_0(\cdot;\theta) : [0,\infty) \to [0,\infty)$ 和 $f(\cdot;\theta) : X \to \mathbb{R}$ 是带参数向量 θ 的函数。特别地：

1. 指数型比例风险模型：$h(t|x;\theta) := e^{\beta^\top x+\psi}$. 对于 $t \geq 0, x \in X$。此时，基准风险函数和协变量影响函数分别为：

$$
h_0(t;\theta) = e^{\psi}, \quad f(x;\theta) = \beta^\top x, \tag{1.19}
$$

其中参数 $\theta = (\beta,\psi) \in \mathbb{R}^d \times \mathbb{R}$。将这个风险函数 $h(t|x;\theta) = e^{\beta^\top x+\psi}$ 代入负对数似然函数的公式 (1.17)，我们得到：

$$
\begin{aligned}
&L_{\mathrm{HNLL}}(\beta,\psi) \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i(\beta^\top X_i+\psi)-\int_0^{Y_i} e^{\beta^\top X_i+\psi}du\right] = -\frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i(\beta^\top X_i+\psi)-Y_i e^{\beta^\top X_i+\psi}\right].
\end{aligned}
\tag{1.20}
$$

2. 威布尔型比例风险模型：$h(t|x;\theta) := t^e e^{\varphi-e^{\beta^\top x}} e^{\varphi+\psi+\varphi}$. 对于 $t \geq 0, x \in X$。此时，基准风险函数和协变量影响函数分别为：

$$
h_0(t;\theta) = t^{e^{\varphi-1}} e^{\psi+\varphi}, \quad f(x;\theta) = (\beta^\top x)e^{\varphi}, \tag{1.21}
$$

其中参数 $\theta = (\beta, \psi, \varphi) \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}$。将风险函数 $h(t|x;\theta) := te^{\varphi-e^{\beta^\top x}} e^{\varphi+\psi+\varphi}$ 代入负对数似然函数公式 (1.17)，我们得到：

$$
L_{\mathrm{HNLL}}(\beta,\psi,\varphi) = -\frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i\left((e^{\varphi}-1)\log Y_i + (\beta^\top X_i)e^{\varphi}+\psi+\varphi\right) - Y_i e^{\varphi} e^{(\beta^\top X_i)e^{\varphi}+\psi}\right]. \tag{1.22}
$$

接下来，我们介绍更一般的 Cox 风险比例模型。根据 Cox [1] 和Breslow [2] 的工作，基准风险函数 $h_0(t;\theta)$ 可以被表示为分段常数函数，其定义如下：

$$
h_0(t;\theta) := \begin{cases} \lambda_l, & \text{如果 } \tau_{l-1} < t \leq \tau_l,\ l \in [L], \\ 0, & \text{如果 } t > \tau_L, \end{cases} \tag{1.23}
$$

其中 $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_L) \in [0, \infty)$ 是基准风险函数的参数，而 $\tau(1), \tau(2), \ldots, \tau(L)$ 是事件发生的特定时间点，通常这些时间点表示生存分析中的观测事件时间。此外，定义 $\tau_0 := 0$。

将基准风险函数和风险函数的定义代入负对数似然函数公式 (1.17)，可以得到：

$$
\begin{aligned}
\log L(\theta,\lambda) &= \sum_{i=1}^{n}\left[\Delta_i \log h(Y_i|X_i;\theta) - \int_0^{Y_i} h(u|X_i;\theta)\,du\right] \\
&= \sum_{i=1}^{n}\left[\Delta_i \log\left(h_0(Y_i;\theta)e^{f(X_i;\theta)}\right) - \int_0^{Y_i} h_0(u;\theta)e^{f(X_i;\theta)}\,du\right] \\
&= \sum_{m=1}^{L} D[m]\log\lambda_m + \sum_{i=1}^{n}\Delta_i f(X_i;\theta) - \sum_{m=1}^{L}(\tau_m-\tau_{m-1})\lambda_m \sum_{j=1}^{n} I\left(Y_j \geq m\right)e^{f(X_j;\theta)}.
\end{aligned}
\tag{1.24}
$$

通过对 λ_l 进行求导，并设定 $\left.\frac{d\log L(\theta)}{d\lambda_l}\right|_{\lambda_l=\widetilde{\lambda_l}} = 0$，我们可以得到基准风险函数的最大似然估计：

$$\widetilde{\lambda}_l = \frac{D[l]}{(\tau_l - \tau_{l-1}) \sum_{j=1}^{n} I\left(Y_j \geqslant l\right) e^{f(X_j;\theta)}}. \tag{1.25}$$

我们将公式 (1.25) 代入公式 (1.24)：

$$\begin{aligned} \log L(\theta) &= \sum_{m=1}^{L} D[m] \log \frac{D[l]}{(\tau_{(m)} - \tau_{(m-1)}) \sum_{j=1}^{n} \mathbb{I}\left\{Y_j \geqslant m\right\} e^{f(X_j;\theta)}} + \sum_{i=1}^{n} \Delta_i f(X_i;\theta) - \sum_{m=1}^{L} D[m] \\ &= \sum_{i=1}^{n} \Delta_i \left[f(X_i;\theta) - \sum_{i=1}^{n} \log \left(\sum_{j=1}^{n} \mathbb{I}\left\{Y_j \geqslant Y_i\right\} e^{f(X_j;\theta)} \right) \right] + \text{常数}. \end{aligned} \tag{1.26}$$

最终得到 Cox 模型的负对数似然函数：

$$L_{\mathrm{HNLL}}(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \Delta_i \left[f(X_i;\theta) - \sum_{i=1}^{n} \log \left(\sum_{j=1}^{n} \mathbb{I}\left\{Y_j \geqslant Y_i\right\} e^{f(X_j;\theta)} \right) \right]. \tag{1.27}$$

这个公式为Cox比例风险模型的负对数似然函数，涵盖了基准风险的估计和协变量的影响。在实际应用中，这一负对数似然函数常用于模型参数的估计。

1.3 离散时间模型

在离散时间模型中，假设时间被离散化为自定义的 L 个时间点 $\tau_{(1)}, \tau_{(2)}, \ldots, \tau_{(L)} \in [0, \infty)$，其中 $\tau_{(1)} < \tau_{(2)} < \cdots < \tau_{(L)}$。假设所有训练值 Y_i 已经离散化为从 $\tau_{(1)}, \tau_{(2)}, \ldots, \tau_{(L)}$ 选取的值。分布 $P(T \mid X = x)$ 的概率质量函数(PMF)和累计分布函数(CDF)的定义如下：

$$\text{PMF} : f[l \mid x] := P(T = \tau_{(l)} \mid X = x), \text{对应} l \in [L],$$
$$\text{CDF} : F[l \mid x] := P(T \leq \tau_{(l)} \mid X = x) = \sum_{m=1}^{l} f[m \mid x].$$

其中概率质量函数 $f[\cdot \mid x]$ 满足：对于所有 $l \in [L]$，有 $f[l \mid x] \geq 0$; $\sum_{l=1}^{L} f[l \mid x] = 1$。另外，概率质量函数和累计分布函数之间的关系是：$f[l \mid x] = F[l \mid x] - F[l-1 \mid x]$，其中 $F[0 \mid x] := 0$。

对于任意 $x \in X$，在时间索引 $l \in [L]$ 处的离散时间生存函数定义为：

$$S[l \mid x] := P(T > \tau_{(l)} \mid X = x) = 1 - F[l \mid x] = 1 - \sum_{m=1}^{l} f[m \mid x]. \tag{1.28}$$

此外，概率质量函数 $f[l \mid x]$ 可以通过公式 (1.28) 表示为：

$$f[l \mid x] = F[l \mid x] - F[l-1 \mid x] = (1 - S[l \mid x]) - (1 - S[l-1 \mid x]) = S[l-1 \mid x] - S[l \mid x] l \in [L], \tag{1.29}$$

其中 $S[0 \mid x] := 1$。

离散时间风险函数 $h[l \mid x]$ 的定义是 $h[l \mid x] := P(T = \tau_{(l)} \mid X = x, T \geqslant \tau_{(l)})$。那么风险函数 $h[l \mid x]$ 可以写成

$$\begin{aligned} h[l \mid x] &= \frac{P(T = \tau_{(l)}, T \geqslant \tau_{(l)} \mid X = x)}{P(T > \tau_{(l-1)} \mid X = x)} = \frac{P(T = \tau_{(l)} \mid X = x)}{P(T \geqslant \tau_{(l)} \mid X = x)} \\ &= \frac{P(T = \tau_{(l)} \mid X = x)}{P(T > \tau_{(l-1)} \mid X = x)} = \frac{f[l \mid x]}{S[l-1 \mid x]} = \frac{S[l \mid x] - S[l-1 \mid x]}{S[l-1 \mid x]} \end{aligned} \tag{1.30}$$

需要注意的是，尽管连续时间的风险函数 $h(t \mid x)$ 可以为非负值并且可能超过 1，但在离散时间中，$h[l \mid x]$ 作为一个概率，不能超过 1。

另外，从公式 (1.30) 可以得到：

$$h[l \mid x] = \frac{S[l \mid x] - S[l-1 \mid x]}{S[l-1 \mid x]} \Leftrightarrow S[l \mid x] = S[l-1 \mid x](1 - h[l \mid x]). \tag{1.31}$$

于是，生存函数 $S[l \mid x]$ 可以表示为：

$$S[l \mid x] = \prod_{m=1}^{l} (1 - h[m \mid x]), \quad l \in [L], \tag{1.32}$$

上述公式 (1.32) 还说明了如何利用风险函数 $h[\cdot \mid x]$ 来估计生存函数 $S[\cdot \mid x]$。

离散时间累计风险函数定义为：

$$H[l \mid x] := \sum_{m=1}^{l} h[m \mid x], \tag{1.33}$$

根据上述公式 (1.33)，有以下关系：

$$h[l \mid x] = H[l \mid x] - H[l-1 \mid x], \tag{1.34}$$

其中 $H[0 \mid x] := 0$。

需要注意的是，在连续时间模型中，关系式 $-\log S[l \mid x] = H[l \mid x]$ 成立。然而，在离散时间中，相应的表达式为：

$$-\log S[l \mid x] = H[l \mid x] + \sum_{m=1}^{l} \sum_{p=2}^{\infty} \frac{(h[m \mid x])^p}{p}, \quad l \in [L]. \tag{1.35}$$

总结一下，概率质量函数 PMF $f(\cdot|x)$、累积分布函数 CDF $F(\cdot|x)$、生存函数 $S(\cdot|x)$、风险函数 $h(\cdot|x)$ 和累计风险函数 $H(\cdot|x)$ 之间的关系：

$$\begin{aligned}
&\text{KRB1}: \quad f[l \mid x] = F[l \mid x] - F[l-1 \mid x] = S[l-1 \mid x] - S[l \mid x] = h[l \mid x] S[l-1 \mid x] \\
&\text{KRB2}: \quad F[l \mid x] = \sum_{m=1}^{l} f[m \mid x] = 1 - S[l \mid x] \\
&\text{KRB3}: \quad S[l \mid x] = 1 - F[l \mid x] = \sum_{m=l+1}^{L} f[m \mid x] = \prod_{m=1}^{l} (1 - h[m \mid x]) \\
&\text{KRB4}: \quad h[l \mid x] = H[l \mid x] - H[l-1 \mid x] = \frac{S[l-1 \mid x] - S[l \mid x]}{S[l-1 \mid x]} = \frac{f[l \mid x]}{S[l-1 \mid x]} \\
&\text{KRB5}: \quad H[l \mid x] = \sum_{m=1}^{l} \frac{S[m-1 \mid x] - S[m \mid x]}{S[m-1 \mid x]} = \sum_{m=1}^{l} h[m \mid x]
\end{aligned} \tag{1.36}$$

其中 $F[0 \mid x] = 0, \quad S[0 \mid x] = 1, \quad H[0 \mid x] = 0$。

与连续时间模型的似然函数 (1.13)类似，离散时间模型的似然函数表达为：

$$L := \prod_{i=1}^{n} [f[\kappa(Y_i) \mid X_i]]^{\Delta_i} [S[\kappa(Y_i) \mid X_i]]^{1-\Delta_i}, \tag{1.37}$$

其中，$\kappa(Y_i)$ 表示与观测时间 Y_i 对应的时间索引，Y_i 被离散化为 $\tau_{(1)}, \tau_{(2)}, \ldots, \tau_{(L)}$ 中的一个值。函数 $f[\kappa(Y_i) \mid X_i]$ 表示在时间 $\kappa(Y_i)$ 发生事件的概率质量，条件是协变量 X_i，而 $S[\kappa(Y_i) \mid X_i]$ 表示在相应时间索引下，条件为 X_i 的生存函数。对于第 i 个个体，指示符 Δ_i 如果事件发生则取值 1（即 $\Delta_i = 1$），如果事件被删失则取值 0（即 $\Delta_i = 0$）。

根据关系 (1.36)，我们将公式 (1.37) 重新写为：

$$
\begin{aligned}
L &= \prod_{i=1}^{n} \left[f[\kappa(Y_i) \mid X_i]\right]^{\Delta_i} \left[S[\kappa(Y_i) \mid X_i]\right]^{1-\Delta_i} \\
&\overset{\text{KRB1}}{=} \prod_{i=1}^{n} \left[(h[\kappa(Y_i) \mid X_i] S[\kappa(Y_i) - 1 \mid X_i])^{\Delta_i} S[\kappa(Y_i) \mid X_i]^{1-\Delta_i}\right] \\
&\overset{\text{KRB3}}{=} \prod_{i=1}^{n} \left[h[\kappa(Y_i) \mid X_i] \left(\prod_{m=1}^{\kappa(Y_i)-1} (1 - h[m \mid X_i])\right)^{\Delta_i} \left(\prod_{m=1}^{\kappa(Y_i)} (1 - h[m \mid X_i])\right)^{1-\Delta_i}\right] \\
&= \prod_{i=1}^{n} \left[h[\kappa(Y_i) \mid X_i]^{\Delta_i} (1 - h[\kappa(Y_i) \mid X_i])^{1-\Delta_i} \left(\prod_{m=1}^{\kappa(Y_i)} (1 - h[m \mid X_i])\right)\right].
\end{aligned}
\tag{1.38}
$$

然后，对公式 (1.38) 两边取对数，得到对数似然函数：

$$
\log L = \sum_{i=1}^{n} \left[\Delta_i \log\left(h[\kappa(Y_i) \mid X_i]\right) + (1 - \Delta_i) \log\left(1 - h[\kappa(Y_i) \mid X_i]\right)\right] + \sum_{m=1}^{\kappa(Y_i)-1} \log\left(1 - h[m \mid X_i]\right) \tag{1.39}
$$

在经典统计学中，通常通过极大化对数似然函数 (1.39) 求得统计学参数 ξ，而在机器学习中等价地通过最小化负对数似然函数：

$$
L_{\text{HNLL}}(\xi) = -\frac{1}{n} \sum_{i=1}^{n} \left[\Delta_i \log\left(h[\kappa(Y_i) \mid X_i; \xi]\right) + (1 - \Delta_i) \log\left(1 - h[\kappa(Y_i) \mid X_i; \xi]\right)\right] + \sum_{m=1}^{\kappa(Y_i)-1} \log\left(1 - h[m \mid X_i; \xi]\right). \tag{1.40}
$$

通过这些公式和推导，离散时间模型中的生存分析和相关量的估计得到了有效描述。

1.4 本章小结

1. 本章给出了连续时间模型下的概率密度函数 $f(\cdot|x)$、累积分布函数 $F(\cdot|x)$、生存函数 $S(\cdot|x)$、风险函数 $h(\cdot|x)$ 和累积风险函数 $H(\cdot|x)$ 之间的关系 (1.11)，以及离散时间模型下的概率质量函数 PMF $f(\cdot|x)$、累积分布函数 CDF $F(\cdot|x)$、生存函数 $S(\cdot|x)$、风险函数 $h(\cdot|x)$ 和累计风险函数 $H(\cdot|x)$ 之间的关系 (1.36)。

2. 本章还给出了连续和离散时间模型下的基本似然函数 (1.13) 和 (1.37)。

3. 本章进一步给出了连续和离散时间模型下的基本非负对数似然函数 (1.17) 和 (1.40)。

需要说明的是，以上三点为后续章节的基础。

1.5 参考文献

[1] D. Cox, "Regression models and life-tables(with discussion)," *Journal of the Royal Statistical Society. Series B*, vol. 34(2), pp. 187–220, 1972.

[2] N. Breslow, "Discussion of the paper by d. r. cox," *Journal of the Royal Statistical Society (B)*, vol. 34, pp. 216–217, 1972.