

Survival Analysis

Chao Cheng

November 27, 2022

Contents

1	Basic knowledge	1
1.1	Survival and hazard	1
1.2	Censor	3
1.2.1	Right censor	3
1.2.2	Left censor	3
1.2.3	Interval censor	3
2	MLE	3
2.1	Parametric MLE	4
2.1.1	One-sample setting	4
2.1.2	Two-sample setting	6
2.2	Nonparametric MLE	6
2.2.1	Discrete time points	6

1 Basic knowledge

1.1 Survival and hazard

Let T denote the time to an event that we are interested in. Then we know the c.d.f.

$$F_T(t) = P(T \leq t),$$

and the corresponding p.d.f.

$$f_T(t) = \frac{d}{dt} F_T(t).$$

Here to simplify the discussion, we assume T is a continuous random variable. In the context of survival analysis, the *event* often refers to death. Then T represents the lifespan of the subject. So $F_T(t)$ represents the probability that the death occurs before t . In another word, we know the probability that the subject survives passes t is

$$S_T(t) = 1 - F_T(t) = P(T > t).$$

$S_T(t)$ is often called the **survival function?** and clearly

$$f_T(t) = -\frac{d}{dt} S_T(t).$$

The **hazard function** $h(t)$ is defined as

$$h(t) = \lim_{\Delta \rightarrow 0} \frac{P(T \leq t + \Delta | T > t)}{\Delta} = \lim_{\Delta \rightarrow 0} \frac{F_T(t + \Delta) - F_T(t)}{\Delta \cdot S_T(t)} = \frac{f_T(t)}{S_T(t)}.$$

$h(t)$ represents the **instant hazard? unified probability?** that the subject will be dead instantly after t given the fact that it's alive at t . And the **cummulative hazard function** is

$$H(t) = \int_0^t h(x) dx = \int_0^t \frac{f_T(x)}{S_T(x)} dx = \int_0^t \frac{-dS_T(x)}{S_T(x)} = -\log(S_T(x))|_0^t = -\log(S_T(t)).$$

Proposition 1. *The random variable $H(T)$ follows unit exponential distribution $EXP(1)$.*

Proof.

$$\begin{aligned} P(H(T) \leq t) &= P(-\log S(T) \leq t) \\ &= P(1 - F(T) \geq e^{-t}) \\ &= P(T \leq F^{-1}(1 - e^{-t})) \\ &= F(F^{-1}(1 - e^{-t})) \\ &= 1 - e^{-t}, \end{aligned}$$

which is the c.d.f of $EXP(1)$. Here to simplify the deduction we make some assumptions that

- $F(t)$ is continuous.
- $F^{-1}(t)$ is well defined.

Also to simplify the notation and avoid confusion, we use $S(\cdot)$ and $F(\cdot)$ instead of $S_T(\cdot)$ and $F_T(\cdot)$ like before. \square

1. Exponential distribution: Denote $T \sim EXP(\lambda)$. Then

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t} \\ F(t) &= 1 - e^{-\lambda t} \quad S(t) = e^{-\lambda t} \\ h(t) &= \lambda \quad \text{constant hazard} \\ H(t) &= \lambda t \\ E(T) &= 1/\lambda \quad \text{Var}(T) = 1/\lambda^2 \end{aligned}$$

2. Weibull distribution: Denote $T \sim W(p, \lambda)$. Then

$$\begin{aligned} f(t) &= p\lambda^p t^{p-1} \exp(-(\lambda t)^p) \\ F(t) &= 1 - \exp(-(\lambda t)^p) \quad S(t) = \exp(-(\lambda t)^p) \\ h(t) &= p\lambda^p t^{p-1} \\ H(t) &= (\lambda t)^p \\ E(T) &= \frac{1}{\lambda} \cdot \Gamma\left(1 + \frac{1}{p}\right) \quad \text{Var}(T) = \frac{1}{\lambda^2} \left(\Gamma\left(1 + \frac{2}{p}\right) - \Gamma\left(1 + \frac{1}{p}\right)^2 \right) \\ E(T^m) &= \frac{1}{\lambda^m} \Gamma\left(1 + \frac{m}{p}\right) \end{aligned}$$

1.2 Censor

1.2.1 Right censor

- Type I: an i.i.d sample $T_1, \dots, T_n \sim F$ and a **fixed** constant c . And the observed data is (U_i, δ_i) for $i = 1, \dots, n$ where

$$U_i = \min(T_i, c)$$
$$\delta_i = 1_{T_i \leq c}.$$

So the observed data consists of a **random** number, r , of uncensored observations, all of which are less than c . And $n - r$ censored observations, all are c .

- Type II: an i.i.d sample $T_1, \dots, T_n \sim F$ and a **pre-defined** number of failure r . The observation is stopped when r failure occurs and the stopping time is c . The observed data is still the form (U_i, δ_i) for $i = 1, \dots, n$, the same as that in Type I censor. But in actuality, we observe the first r **order statistics**

$$T_{(1,n)}, \dots, T_{(r,n)}.$$

Note that here $(U_1, \delta_1), \dots, (U_n, \delta_n)$ are **dependent** whereas they are independent for Type I.

- Type III (Random censor): The underlying data is

$$c_1, \dots, c_n \text{ constant}$$
$$T_1, \dots, T_n \sim F.$$

And the observed data is (U_i, δ_i) for $i = 1, \dots, n$, where

$$U_i = \min(T_i, c_i)$$
$$\delta_i = 1_{T_i \leq c_i}.$$

Note: for inference, c_i is often treated as constant. For study design or studying the asymptotic property, they are often treated as i.i.d random variables C_1, \dots, C_n .

1.2.2 Left censor

T_i is censored when $T_i \leq l_i$.

1.2.3 Interval censor

$l_i \leq T_i \leq u_i$, but only l_i and u_i are observed.

2 MLE

There is an i.i.d survival time sample T_1, \dots, T_n with common and unknown c.d.f. $F(\cdot)$ and the observed data is (U_i, δ_i) for $i = 1, \dots, n$, where

$$U_i = \min(T_i, C_i)$$
$$\delta_i = 1(T_i \leq C_i)$$

and C_i is the (**fixed** or **random**) censoring time. Let \perp denote “is independent of”. We assume $T_i \perp C_i$ (**Non-informative censoring, the key assumption**) and (U_i, δ_i) are also i.i.d. The observed data consists of two parts. U_i is continuous while δ_i is binary.

$$\begin{aligned} (U_i, \delta_i) &= (u_i, 1) & T_i \text{ is uncensored at } u_i \\ (U_i, \delta_i) &= (u_i, 0) & T_i \text{ is censored at } u_i \end{aligned}$$

When C_i s are known constants, the likelihood for (U_i, δ_i) is

$$\begin{aligned} L_i(F) &= \begin{cases} f(u_i) & \text{if } \delta_i = 1 \\ 1 - F(u_i) & \text{if } \delta_i = 0 \end{cases} \\ &= f(u_i)^{\delta_i} (1 - F(u_i))^{1-\delta_i} \end{aligned}$$

Therefore

$$L(F) = \prod_{i=1}^n L_i(F) = \prod_{i=1}^n \left(f(u_i)^{\delta_i} (1 - F(u_i))^{1-\delta_i} \right) = \prod_{i=1}^n \left(h(u_i)^{\delta_i} S(u_i) \right). \quad (1)$$

The last equality relies on the fact that $f(t) = h(t) S(t)$.

When C_i s are i.i.d. $\sim G$, where G is continuous with p.d.f g . Then we have

$$P(U_i \leq u, \delta_i = 1) = P(T_i \leq u, T_i \leq C_i) = \int_0^u \int_t^\infty f(t) g(c) dc dt = \int_0^u f(t) (1 - G(t)) dt$$

Therefore the likelihood for $\delta_i = 1$ is

$$L_i(F, G) = f(u_i) (1 - G(u_i)) \quad \text{when } \delta_i = 1.$$

And similarly, for $\delta_i = 0$, the likelihood is

$$L_i(F, G) = g(u_i) (1 - F(u_i)) \quad \text{when } \delta_i = 0.$$

Hence the full likelihood is

$$\begin{aligned} L(F, G) &= \prod_{i=1}^n \left\{ (f(u_i) (1 - G(u_i)))^{\delta_i} ((1 - F(u_i)) g(u_i))^{1-\delta_i} \right\} \\ &= \prod_{i=1}^n \left\{ f(u_i)^{\delta_i} (1 - F(u_i))^{1-\delta_i} \right\} \cdot \prod_{i=1}^n \left\{ g(u_i)^{1-\delta_i} (1 - G(u_i))^{\delta_i} \right\} \end{aligned} \quad (2)$$

So the core to maximize $L(F, G)$ with respect to F in (2) is the same as that in (1).

2.1 Parametric MLE

2.1.1 One-sample setting

Suppose T_1, \dots, T_n are i.i.d. $Exp(\lambda)$, and subject to noninformative right censoring. Then (1) becomes

$$L = L(\lambda) = \prod_{i=1}^n \left\{ (\lambda e^{-\lambda u_i})^{\delta_i} (e^{-\lambda u_i})^{1-\delta_i} \right\} = \lambda^{\sum_{i=1}^n \delta_i} e^{-\lambda \sum_{i=1}^n u_i} = \lambda^r e^{-\lambda W},$$

where $r = \sum_{i=1}^n \delta_i$ is the number of observed events and $W = \sum_{i=1}^n u_i$ is the total of observed time. Therefore $\log L = r \log \lambda - \lambda W$ and the MLE for λ is

$$\hat{\lambda} = \frac{r}{W}.$$

Furthermore, we know that

$$\begin{cases} \frac{\partial \log L}{\partial \lambda} = \frac{r}{\lambda} - W \\ \frac{\partial^2 \log L}{\partial \lambda^2} = -\frac{r}{\lambda^2} \end{cases}.$$

Based on properties of fisher information ([See the notes about fisher information for more details.](#)), we know that at the **true underlying value** λ , it must satisfy

$$\begin{cases} E \frac{\partial \log L}{\partial \lambda} = \frac{Er}{\lambda} - EW = 0 \\ I(\lambda) = -E \frac{\partial^2 \log L}{\partial \lambda^2} = \frac{Er}{\lambda^2} \\ I^*(\lambda) = \frac{1}{n} I(\lambda) = \frac{Er}{n\lambda^2} \end{cases}. \quad (3)$$

Note that in (3), r and W are random variables. And the probability to observe an event is

$$p = P(\delta_i = 1) = P(U_i \leq \infty, \delta_i = 1) = \int_0^\infty f(t) (1 - G(t)) dt.$$

Therefore $r \sim \text{binomial}(n, p)$, $Er = np$. And from property of MLE, we can write

$$\frac{\sqrt{n}(\hat{\lambda} - \lambda)}{\sqrt{I^*(\lambda)^{-1}}} = \frac{(\hat{\lambda} - \lambda)}{\sqrt{I(\lambda)^{-1}}} \xrightarrow{D} N(0, 1),$$

which means approximately

$$\hat{\lambda} \stackrel{\text{apx}}{\sim} N(\lambda, I(\lambda)^{-1}) = N\left(\lambda, \frac{\lambda^2}{np}\right).$$

Unfortunately, both λ and p (essentially $G(\cdot)$) are unknown. We plug in the estimation $\hat{\lambda} = r/W$ and $\hat{p} = r/W$ and apply Slutsky's theorem. This means for the purpose of estimation, we use

$$\begin{cases} \hat{\lambda} = \frac{r}{W} \\ I(\hat{\lambda}) = \frac{r}{\hat{\lambda}^2}, \quad I^*(\hat{\lambda}) = \frac{r}{n\hat{\lambda}^2} \end{cases} \quad (4)$$

Not that unlike (3), here in (4), r and W are observations. And we have

$$\hat{\lambda} \stackrel{\text{apx}}{\sim} N\left(\lambda, \frac{r}{W^2}\right). \quad (5)$$

Note that it turns out that a better approximation is to assume $\log \hat{\lambda}$ is normal. Using the delta method, this gives

$$\log \hat{\lambda} \stackrel{\text{apx}}{\sim} N\left(\log \lambda, \frac{1}{np}\right) \approx N\left(\log \lambda, \frac{1}{r}\right). \quad (6)$$

Now based on (5) or (6), we can construct CI on λ , which also means we can perform hypothesis testing about λ .

2.1.2 Two-sample setting

For two samples x_1, \dots, x_n and y_1, \dots, y_m , both follow exponential distribution with parameters λ_1 and λ_2 . Assume noninformative censoring in each group, using same tech in Section 2.1.1 we can get

$$Z = \frac{\log \hat{\lambda}_1 - \log \hat{\lambda}_2}{\sqrt{\frac{1}{r_1} + \frac{1}{r_2}}} \stackrel{\text{apx}}{\sim} N(0, 1).$$

2.2 Nonparametric MLE

The NPMLE of survivor function $S(\cdot)$ based on i.i.d. survival time and non-informative right censoring is often known as Kaplan-Meier estimator or the Product-Limit Estimator. Here we provide some heuristic development, but formal proofs will be deferred to other notes.

2.2.1 Discrete time points

To begin with,

References