Reliability and Life Testing

Xiaoqian Xun
(Notes by Shuai Wei)

October 3, 2023

Contents

Chapter 1

Lifetime Distributions

1.1 Distribution Representatives

Let T be the liftime of an item (time to failure).

1.1.1 Survial function

Definition 1.1. The survival function

$$S(t) = P(T \geqslant t) = 1 - F(t), t \geqslant 0.$$

Remark. (a) All survival function satisfy:

(1)

$$S(0) = 1.$$

(2)

$$\lim_{t \to \infty} S(t) = 0.$$

(3)

$$S(t) \text{ is nonincreasing in } t.$$

(b) $S(t)$ is useful for camping survival pattern of items. If $S_1 \geqslant S_2(t), \forall t \geqslant 0$, then item 1 is superior to item 2.

(c) The conditional survival function is

$$S_{T|T \geqslant t}(\tau) := P(T \geqslant \tau | T \geqslant t) = \frac{P(T \geqslant \tau, T \geqslant t)}{P(T \geqslant t)} = \frac{P(T \geqslant \max(\tau, t))}{P(T \geqslant t)} = \begin{cases} \frac{S(\tau)}{S(t)}, & \tau \geqslant t, \\ 1, & \tau \leqslant t. \end{cases}$$

1.1.2 PDF and CDF

Definition 1.2. The pdf is defined through $P(a \leqslant T \leqslant b) = \int_a^b f(t)dt, \forall a < b$, then we call the *integrand* $f(t)$ the pdf of T.

Remark. (a) The pdf $f(t)$ has the probability interpretation

$$f(t)\Delta t \approx P(t \leqslant T \leqslant t + \Delta t) \text{ for some small } \Delta t.$$

(b) $f(t)$ is the likelihood of failure at time t.

(c)

$$S'(t) = -f(t).$$

1.1.3 Hazard function

Hazard function gives the amount of risk associated w/ an item at time t, often referred to as "instantaneous failure rate".

Definition 1.3.

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P(t \leqslant T \leqslant t + \Delta t | T \geqslant t)}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{P(t \leqslant T \leqslant t + \Delta t)}{P(T \geqslant t)} \frac{1}{\Delta t} \\ &= \lim_{\Delta t \to 0} \frac{S(t) - S(t + \Delta t)}{S(t)} \frac{1}{\Delta t} \\ &= -\frac{1}{S(t)} \lim_{\Delta t \to 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} \\ &= -\frac{1}{S(t)} S'(t) \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

Remark. (a)

$$h(t) \geqslant 0, \forall t \geqslant 0.$$

(b)

$$\int_0^\infty h(t)dt = -\ln(S(t))|_0^\infty = \infty.$$

Example 1.4. Consider the Weibull dist w/ cdf

$$F(t) = 1 - e^{-(\lambda t)^\kappa}, t \geqslant 0,$$

where λ is a scale parameter and κ is a shape parameter. The pdf

$$f(t) = \lambda k (\lambda t)^{k-1} e^{-(\lambda t)^\kappa}, t \geqslant 0.$$

The survival function

$$S(t) = e^{-(\lambda t)^{\kappa}}, t \geqslant 0.$$

The hazard funnction is

$$h(t) = \lambda k(\lambda t)^{k-1}, t \geqslant 0,$$

where $0 < \kappa < 1$, $h(t)$ is decreasing, $\kappa = 1$, $h(t)$ is a constant. $\kappa > 1$, $h(t)$ is increasing.

Definition 1.5. Cumulative hazard function

$$H(t) = \int_0^t h(s)ds, \; t \geqslant 0.$$

Remark. Notice that $H(t)$ satisfies

(a)

$$H(0) = 0.$$

(b)

$$\lim_{t \to \infty} H(t) = \infty.$$

(c)

$$H(t) \text{ is nondecreasing in } t.$$

1.1.4 Mean residual life function

Lemma 1.6.

$$f_{T|T \geqslant t}(\tau) = \begin{cases} \frac{f(\tau)}{S(t)}, & \tau \geqslant t \\ 0, & \tau < t. \end{cases}$$

Proof. Notice here $T \geqslant t$ is an event but not a r.v. Let $\tau > t$, then

$$\begin{aligned} F_{T|T \geqslant t}(\tau) &= P(T \leqslant \tau | T \geqslant t) \\ &= \frac{P(t \leqslant T \leqslant \tau)}{P(T \geqslant t)} \\ &= \frac{S(t) - S(\tau)}{S(t)} \\ &= 1 - \frac{S(\tau)}{S(t)}. \end{aligned}$$

□

So

$$f_{T|t \geqslant t} = -\frac{S'(\tau)}{S(t)} = \frac{f(\tau)}{S(t)}.$$

Definition 1.7. The mean residual life function

$$L(t) = E\left[T - t | T \geqslant t\right], \; t \geqslant 0.$$

It is the expected remaining life given that the item has survived to time t.

Remark.

$$E[T] = L(0).$$

Proposition 1.8.

$$L(t) = \frac{1}{S(t)} \int_t^\infty \tau f(\tau) d\tau - t.$$

Proof.

$$\begin{aligned} L(t) &= E[T|T \geqslant t] - E[t|T \geqslant t] \\ &= \int_t^\infty \tau f_{T|t \geqslant t}(\tau) d\tau - t \\ &= \int_t^\infty \tau \frac{f(\tau)}{S(t)} d\tau - t \\ &= \frac{1}{S(t)} \int_t^\infty \tau f(\tau) d\tau - t. \end{aligned}$$

□

Remark. Any $L(t)$ satisfies

(a)

$$L(t) \geqslant 0.$$

(b)

$$L'(t) \geqslant -1.$$

(c)

$$\int_0^\infty \frac{1}{L(t)} dt = \infty.$$

Example 1.9. Let

$$S(t) = e^{-\lambda t},\ t \geqslant 0,$$

where λ is the failure rate. Then

$$f(t) = \lambda e^{-\lambda t},\ t \geqslant 0.$$

Hence

$$L(t) = e^{\lambda t} \int_t^\infty \tau \lambda e^{-\lambda \tau} - t = \left(t + \frac{1}{\lambda}\right) - t = \frac{1}{\lambda},\ t \geqslant 0.$$

1.2 Discrete Dist.

T takes values in $\{t_1, t_2, \cdots\}$, where $0 \leqslant t_i < t_j$ when $i < j$.
The pmfs is

$$f(t_i) = P(T = t_i),\ i = 1, 2, \cdots$$

The survival functions

$$S(t) = P(T \geqslant t) = \sum_{t_i \geqslant t} f(t_i),\ t \geqslant 0.$$

The hazard function $h(t)$ is still intepreted as the risk at each time $t_j, j = 1, 2, \cdots$, though.

$$\begin{aligned} h(t_j) &= P(T = t_j | T \geqslant t_j) \\ &= \frac{P(T = t_j)}{P(T \geqslant t_j)} \\ &= \frac{f(t_j)}{S(t_j)},\ j = 1, 2, \cdots \end{aligned}$$

1.3 Moments and Fractiles (Quantiles)

1.3.1 Moments

Proposition 1.10.

$$\begin{aligned} E[T] &= \int_0^\infty t f(t) dt \\ &= -\int_0^\infty t d[S(t)] \\ &= -tS(t)|_0^\infty + \int_0^\infty S(t) dt \\ &= \int_0^\infty S(t) dt \end{aligned}$$

if $\lim_{t \to \infty} tS(t) = 0$.

1.3.2 Fractiles

It corresponds to a specified proportion of items fail.

Definition 1.11.

$$F(t_p) = P(T \leqslant t_p) = p.$$

Then

$$t_p = F^{-1}(p).$$

At t_p, p percent of the iterms fail.

Remark.

$$t_{0.5} = \text{Median}.$$

Example 1.12. Let T be the time (miles) of the 1st failure of a engine.
Assume

$$T \sim Weibull(\lambda = 0.0000077, \kappa = 1.22).$$

Find the warranty period that would allow 1% of engines to fail during the warranty period. Let

$$F(t_{0.01}) = 1 - e^{-(\lambda t_{0.01})^\kappa} = 0.01.$$

Then

$$t_{0.01} = \frac{1}{\lambda} [-\ln(1 - 0.01)]^{\frac{1}{k}} = 2992(\text{miles}).$$

Definition 1.13. Skewness

$$\nu_3 = E\left[\left(\frac{T-\mu}{\sigma}\right)^3\right].$$

Example 1.14. Let $S(t) = e^{-\lambda t}, \ t \geqslant 0$. Then

$$\nu_3 = E\left[\left(\frac{T-\mu}{\sigma}\right)^3\right] = E\left[\left(\frac{T-\frac{1}{\lambda}}{\frac{1}{\lambda}}\right)^3\right] = 2.$$

Definition 1.15. Kurtosis

$$\nu_4 = E\left[\left(\frac{T-\mu}{\sigma}\right)^4\right] = 9.$$

1.4 System Lifetime Distribution

Assume all items work **independently**.
Let r be the realibility, which is the probability that the system works. Let p_i be the probability that the ith item is functional for $i = 1, \ldots, n$. Then the survival function of a system is

$$S(t) = r(S_1(t), \ldots, S_n(t)).$$

(a) n Series System. The realiability is

$$r(p_1, \ldots, p_n) = p_1 \cdots p_n.$$

The survival function

$$\begin{aligned} S(t) &= P(T_s \geqslant t) \\ &= P(T_1 \geqslant t, \cdots, T_n \geqslant t) \\ &= P(T_1 \geqslant t) \cdots P(T_n \geqslant t) \\ &= S_1(t) \cdots S_n(t) \end{aligned}$$

(b) n Parallel System. The reliability is

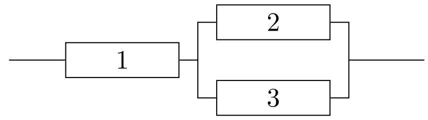
$$r(p_1, \ldots, p_n) = 1 - (1-p_1) \cdots (1-p_n).$$

Then

$$S(t) = 1 - (1 - S_1(t)) \cdots (1 - S_n(t)).$$

(c) General System.

(1) The general system 1.

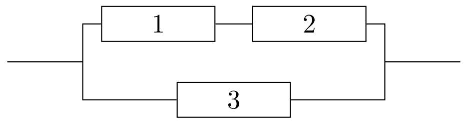


$$r(p_1,p_2,p_3) = p_1(1-(1-p_2)(1-p_3)),$$

and

$$S(t) = S_1(t)(1-(1-S_2(t))(1-S_3(t))).$$

(2) The general system 2.

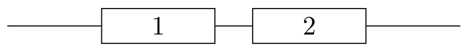


$$r(p_1,p_2,p_3) = 1-(1-p_1p_2)(1-p_3),$$

and

$$S(t) = 1-(1-S_1(t)S_2(t))(1-S_3(t)).$$

Example 1.16. Consider



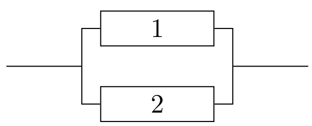
Assume $h_1(t)=1$, $h_2(t)=2,\ t\geqslant 0$. Then $S_1(t)=e^{-t},\ t\geqslant 0$, and $S_2(t)=e^{-2t},\ t\geqslant 0$. Hence

$$S(t) = S_1(t)S_2(t) = e^{-t}e^{-2t} = e^{-3t},\ t\geqslant 0.$$

Thus,

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = 3.$$

Example 1.17. Consider



$T_1 \sim \exp(1)$, $T_2 \sim \exp(2)$. Then

$$\begin{aligned} S(t) &= 1-(1-S_1(t))(1-S_2(t)) \\ &= 1-(1-e^{-t})(1-e^{-2t}) \\ &= e^{-t}+e^{-2t}-e^{-3t},\ t\geqslant 0. \end{aligned}$$

So

$$h(t) = \frac{e^{-t}+2e^{-2t}-3e^{-3t}}{e^{-t}+e^{-2t}-e^{-3t}},\ t\geqslant 0.$$

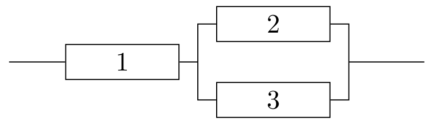
The mean time to system failure rate

$$\mu = \int_0^\infty S(t)dt = \int_0^\infty (e^{-t}+e^{-2t}-e^{-3t})dt = 1+\frac{1}{2}-\frac{1}{3} = \frac{7}{6}.$$

Or

$$\mu = \frac{1}{\mu_1+\mu_2} + \frac{\mu_1}{\mu_1+\mu_2}\frac{1}{\mu_2} + \frac{\mu_2}{\mu_1+\mu_2}\frac{1}{\mu_1} = \frac{\mu_1\mu_2+\mu_1^2+\mu_2^2}{\mu_1\mu_2(\mu_1+\mu_2)} = \frac{7}{6}.$$

Example 1.18. Consider



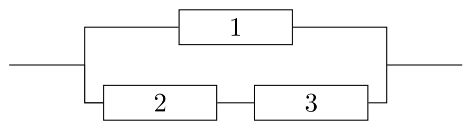
Assume $S_1(t) = S_2(t) = S_3(t)e^{-\lambda t},\ t \geqslant 0$.
Then

$$S(t) = S_1(t)(1 - (1 - S_2(t)(1 - S_3(t)))) = 2e^{-2\lambda t} - e^{-3\lambda t},\ t \geqslant 0,$$

and

$$h(t) = -\frac{S'(t)}{S(t)} = \frac{4e^{-\lambda t} - 3e^{-3\lambda t}}{2e^{-2\lambda t} - e^{-3\lambda t}},\ t \geqslant 0.$$

Example 1.19. Consider



Assume $T_1 \sim U(2, 10),\ T_2 \sim (4, 14),\ T_3 \sim U(6, 18)$. Find the 90th fractile of the distribution of the remaining time to the failure of the system that has survived to 8. It is easy to see that $f_1(t) = \frac{1}{8},\ 2 < t < 10,\ f_2(t) = \frac{1}{10},\ 4 < t < 14$ and $f_3(t) = \frac{1}{12},\ 6 < t < 18$. Hence

$$S_1(t) = \begin{cases} 1, & t \leqslant 2 \\ \frac{10-t}{8}, & 2 < t < 10 \\ 0, & t \geqslant 10 \end{cases}$$

$$S_1(t) = \begin{cases} 1, & t \leqslant 4 \\ \frac{14-t}{10}, & 4 < t < 14 \\ 0, & t \geqslant 14 \end{cases}$$

$$S_1(t) = \begin{cases} 1, & t \leqslant 6 \\ \frac{18-t}{12}, & 0 < t < 18 \\ 0, & t \geqslant 14 \end{cases}$$

Thus,

$$\begin{aligned} S(t) &= 1 - (1 - S_1(t))(1 - S_2(t)S_3(t)) \\ &= S_1(t) + S_2(t)S_3(t) - S_1(t)S_2(t)S_3(t). \end{aligned}$$

Therfore, the conditional survival function for the system that is still operating at time 8 is

$$S_{T|T\geqslant 8(t)} = \frac{S(t)}{S(8)} = \frac{S_1(t) + S_2(t)S_3(t) - S_1(t)S_2(t)S_3(t)}{\frac{5}{8}},\ t \geqslant 8.$$

From the graph of $S_{T|T\geqslant 8(t)}$, we find $t_{0.9} \in (10, 14)$, so $S(t_0.9) = 0$. Let $S_{T|T\geqslant 8(t_{0.9})} = 1 - 0.9 = 0.1$, we have

$$\frac{\frac{14-t_{0.9}}{10}\frac{18-t}{12}}{\frac{5}{8}} = 0.1.$$

Hence

$$2t_{0.1}^2 - 64t_{0.9} + 489 = 0.$$

Thus,

$$t_{0.9} = \frac{32 - \sqrt{46}}{2} \approx 12.6088.$$

Chapter 2

Parametric Lifetime Models

2.1 Parameters

- Location paramter
- Scale parameter
- Shift parameter

2.2 Exponential Distribution

Definition 2.1.

$$T \sim \exp(\lambda),$$

w/

$$S(t) = e^{-\lambda t},\ h(t) = \lambda,\ H(t) = \lambda t,\ L(t) = \frac{1}{\lambda}, \forall t \geqslant 0,$$

where λ is often called failure rate.

The following are properties of exponential distribution.

Proposition 2.2. Exponential distribution is the only continuous distribution w/ the memeoryless property.

Proposition 2.3. If $T \sim \exp(\lambda)$, then $\lambda T \sim \exp(1)$, which is *standard* exponential distribution.

If T is a nonnegative continuous r.v. w/ cumulative hazard function H, then $H(T) \sim \exp(1)$.

(a) Note this property is mathematically equivalent to

$$F(T) \sim Uni(0, 1),$$

(b) To generate a random variable T w/ $F(t)$ or H(t), you might consider

M1: Generate $U \sim Uni(0, 1)$, then

$$T \overset{d}{=} F^{-1}(U).$$

M2: Generate $U \sim Uni(0,1)$, then

$$T \stackrel{d}{=} H^{-1}(-\log(1-U)),$$

or

$$T \stackrel{d}{=} H^{-1}(-\log(U)),$$

because $1-U \stackrel{d}{=} U$ or

$$P(-\log(1-U) \leqslant x) = P\left(1-U \geqslant e^{-x}\right) = P(U \leqslant 1-e^{-x}) = 1-e^{-x}.$$

and

$$P(-\log U \leqslant x) = P\left(U \geqslant e^{-x}\right) = 1-e^{-x}.$$

Example 2.4. Assume $T \sim Weibul$ w/

$$S(t) = e^{-(\lambda t)^k}, t \geqslant 0.$$

Discuss how to generate Weibull distribution r.v. from converting $Uni(0,1)$.

$$H(t) = -\log S(t) = (\lambda t)^k.$$

Then

$$H^{-1}(y) = \frac{1}{\lambda} y^{\frac{1}{k}}.$$

Hence

$$T \stackrel{d}{=} \frac{1}{\lambda}[-\log(1-U)]^{\frac{1}{k}},$$

where $U \sim Uni(0,1)$.

Proposition 2.5. If $T \sim \exp(\lambda)$, then

$$E[T^s] = \frac{T(s+1)}{\lambda^s},\ s > -1.$$

Remark. In statistics, a pivotal quantity or pivot is a function of observations and unobservable parameters whose probability distribution does not depend on the unknown parameters (also referred to as nuisance parameters).

Proposition 2.6. If $T_i \sim \exp(\lambda)$ independent for $i = 1, \ldots, n$, then

$$2\lambda \sum_{i=1}^{n} T_i \sim \chi^2(2n),$$

where $2\lambda \sum_{i=1}^{n} T_i$ is pivot. Knowing this, it is useful to construct a confidence interval for λ.

$$\chi^2_{\frac{\alpha}{2}}(2n) < 2\lambda \sum_{i=1}^{n} T_i < \chi^2_{1-\frac{\alpha}{2}}(2n),$$

so

$$\frac{\chi^2_{\frac{\alpha}{2}}(2n)}{2\sum_{i=1}^{n} T_i} < \lambda < \frac{\chi^2_{1-\frac{\alpha}{2}}(2n)}{2\sum_{i=1}^{n} T_i}.$$

Proposition 2.7. If $T_i \sim \exp(\lambda)$ independent for $i = 1, \ldots, n$, and $\{T_{(i)}\}_{1 \leqslant n}$ is order statistic. Assume $T_{(0)} = 0$, for $i = 1, \ldots, n$, define

$$G_i = T_{(i)} - T_{(i-1)}.$$

Then

(a) The gap statatistics $G_1, \ldots, G_n$ are independent.

(b)

$$G_i \sim \exp((n - i + 1)\lambda).$$

Proof. The joint pdf of $T_{(1)}, \ldots, T_{(n)}$ is

$$f_{T_{(1)},\ldots,T_{(n)}}(t_1, \ldots, t_n) = n! \prod_{i=1}^{n} f_{T_1}(t) = n!\lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}.$$

Since by induction,

$$T_{(1)} = G_1,\ T_{(2)} = G_2 + T_{(1)} = G_1 + G_2,\ T_{(i)} = G_i + T_{(i-1)} = \sum_{j=1}^{i} G_j.$$

Since

$$|J| = \begin{vmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & \cdots & 1 \end{vmatrix},$$

$$\begin{aligned} f_{G_1,\ldots,G_n}(g_1, \ldots, g_n) &= n!\lambda^n e^{-\lambda\left(\sum_{i=1}^{n} \sum_{j=1}^{i} g_j\right)} \\ &= n!\lambda^n e^{-\lambda\left(\sum_{j=1}^{n} \sum_{i=j}^{n} g_j\right)} \\ &= n!\lambda^n e^{-\lambda\left(\sum_{j=1}^{n} (n-j+1) g_j\right)} \\ &= \prod_{i=1}^{n} (n - i + 1)\lambda e^{-\lambda(n-i+1)g_i} \\ &= \prod_{i=1}^{n} f_{G_i}, \end{aligned}$$

where $f_{G_i} = (n - i + 1)\lambda e^{-\lambda(n-i+1)g_i}$. □

Remark. Since G_j are iid, and

$$T_{(i)} = \sum_{j=1}^{i} G_j,$$

we have

$$E[T_{(i)}] = \sum_{j=1}^{i} \frac{1}{n-j+1},$$

and

$$\mathrm{Var}(T_{(i)}) = \sum_{j=1}^{r} \frac{1}{(n-i+1)^2}.$$

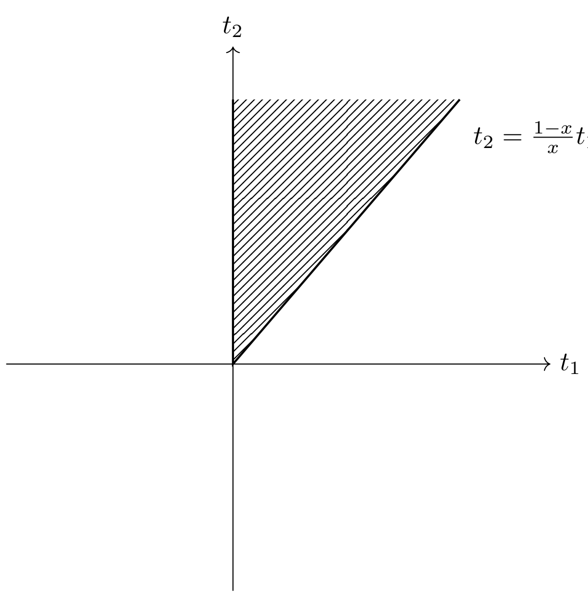
Proposition 2.8. If $T_1, T_2, \cdots$ are iid $\mathrm{Exp}(\lambda)$, denoting the inner-event times for a point process, then the number of events in $[0, t]$ is $\mathrm{Poisson}(\lambda t)$.

Proposition 2.9. If $T_1, T_2 \sim \mathrm{Exp}(\lambda)$ and are independent, then

$$\frac{T_1}{T_1 + T_2} \sim U(0,1).$$

Proof. M1. Let $x > 0$ (tricky), then

$$\begin{aligned}
P\left(\frac{T_1}{T_1+T_2} \leqslant x\right) &= P(T_1(1-x) \leqslant T_2 x) \\
&= P(T_2 \geqslant \frac{1-x}{x} T_1) \ (tricky) \\
&= \int_0^\infty \lambda e^{-\lambda t_1} dt_1 \int_{\frac{1-x}{x} t_1}^\infty \lambda e^{-\lambda t_2} dt_2 \\
&= \int_0^\infty \lambda e^{-\lambda t_1} e^{-\lambda\left(\frac{t_1}{x} - t_1\right)} dt_1 \\
&= \int_0^\infty \lambda e^{-\frac{\lambda}{x} t_1} dt_1 \\
&= x.
\end{aligned}$$



M2. Let

$$X = T_1 + T_2,\ Y = \frac{T_1}{T_1 + T_2}.$$

Then $Y \in (0,1)$ and

$$T_1 = XY,\ T_2 = X - XY.$$

Hence

$$J = \begin{vmatrix} Y & X \\ 1-Y & -X \end{vmatrix} = -X.$$

Then

$$\begin{aligned} f_{X,Y}(x,y) &= \lambda^2 e^{-\lambda(xy+x-xy)}|J| \\ &= \lambda^2 x e^{-\lambda x} \\ &= g(x)h(y), \end{aligned}$$

where assuming $h(y) = 1$. Hence

$$f_Y(y) \propto 1 = h(y).$$

Since $y \in (0,1)$, we have

$$f_Y(y) = \mathbb{1}_{\{[0,1]\}}(y).$$

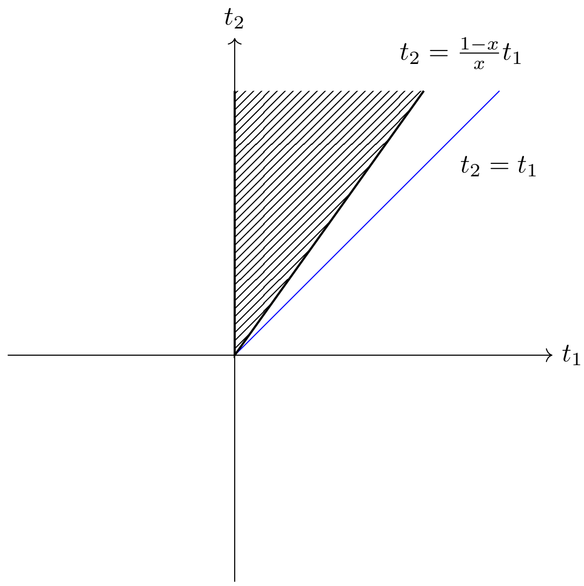
□

Proposition 2.10. If $T_1, T_2 \sim \text{Exp}(\lambda)$ and are independent, then

$$\frac{T_{(1)}}{T_{(1)} + T_{(2)}} \sim U(0, \frac{1}{2}).$$

Proof. Let $\frac{1}{2} > x > 0$, then $\frac{1-x}{x} > 1$ and then

$$\begin{aligned} P\left(\frac{T_{(1)}}{T_{(1)} + T_{(2)}} \leqslant x\right) &= P\left(T_{(2)} \geqslant \frac{1-x}{x} T_{(1)}\right) \\ &= 2! \int_0^\infty \lambda e^{-\lambda t_1} dt_1 \int_{\frac{1-x}{x}t_1}^\infty \lambda e^{-\lambda t_2} dt_2 \\ &= 2 \int_0^\infty \lambda e^{-\lambda t_1} dt_1 \left(e^{-\lambda \frac{1-x}{x} t_1}\right) \\ &= 2 \int_0^\infty \lambda e^{-\frac{\lambda}{x} t_1} \\ &= 2x. \end{aligned}$$



□

Proposition 2.11. If $T_1 \sim \exp(\lambda_1)$ and $T_2 \sim \exp(\lambda_2)$ and are independent. Then

$$P(T_1 < T_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2},$$

which is the probability that item 1's failure causes the system failure.

Remark. Exp dist.

Pro. Simple and convenient.

Con. Limited in application due to memoryless property.

2.3 Weibull distribution

Definition 2.12. T is called to have the Weibull dist (location-scale) Weibull(λ, κ) if

$$S(t) = e^{-(\lambda t)^\kappa},\ f(t) = \kappa \lambda^\kappa t^{\kappa-1} e^{-(\lambda t)^\kappa},\ h(t) = \kappa \lambda^\kappa t^{\kappa-1},\ H(t) = (\lambda t)^\kappa,\ t > 0,$$

where λ is the scale and κ is the shape argument.

The following are graphs under different κ while fixing λ.

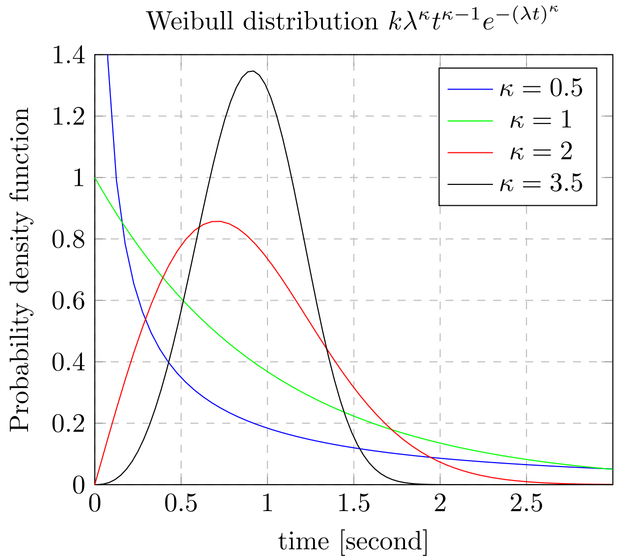


Figure 2.1: $\lambda = 1$

From these graphs, we can see that the Weibull distribution has different shapes for $\kappa > 1$ and $\kappa \leqslant 1$. When $3 < \kappa < 4$, it is close to a normal distribution.

Remark. We have the following.

- $h(t) \to 0$ as $t \to \infty$ when $\kappa < 1$.
- $h(t) \to \infty$ as $t \to \infty$ when $\kappa > 1$.
- $h(t)$ is constant as $t \to \infty$ when $\kappa = 1$.

Remark. Rayleigh distribution is a special distribution of Weibull distribution w/ $k = 2$, which has a linear hazard rate $2\lambda^2 t$.

Proposition 2.13. The mean residual life function is

$$\begin{aligned} L(t) &= \frac{1}{S(t)} \int_{\tau}^{\infty} S(\tau) d\tau \\ &= e^{(\lambda t)^{\kappa}} \int_{\tau}^{\infty} e^{-(\lambda \tau)^{\kappa}} d\tau \\ &= \frac{e^{(\lambda t)^{\kappa}}}{\lambda \kappa} \int_{(\lambda t)^{k}}^{\infty} u^{\frac{1}{\kappa}-1} e^{-u} du \\ &= \frac{e^{(\lambda t)^{\kappa}}}{\lambda \kappa} \Gamma\left(\frac{1}{\kappa}\right) \left[1 - I\left(\frac{1}{\kappa}, (\lambda t)^{\kappa}\right)\right], \end{aligned}$$

where

$$I(y, x) = \frac{1}{\Gamma(y)} \int_0^x u^{y-1} e^{-u} du$$

is the incomplete Gamma function for $x > 0, y > 0$.

Proposition 2.14. The moment of Weibull distribution is

$$E(T^r) = \frac{r}{\kappa\lambda^r}\Gamma\left(\frac{r}{\kappa}\right), \forall r = 1, 2, \cdots$$

So the mean

$$E[T] = \frac{1}{\lambda\kappa}\Gamma\left(\frac{1}{\kappa}\right).$$

The variance

$$\text{Var}(T) = \frac{1}{\lambda^2}\left[\frac{2}{\kappa}\Gamma\left(\frac{2}{\kappa}\right) - \left(\frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)\right)^2\right]$$

The coefficient of variance

$$\nu = \frac{\sigma}{\mu} = \frac{\left[\frac{2}{\kappa}\Gamma\left(\frac{2}{\kappa}\right) - \left(\frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)\right)^2\right]^{\frac{1}{2}}}{\frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)}.$$

Example 2.15. The lifetime in hours of a certain time of spring $T \sim$Weibull$(\lambda = 0.0014, \kappa = 1.28)$. Find

(a) The mean time to failure

$$E[T] = \frac{1}{0.0014 \times 1.28}\Gamma\left(\frac{1}{1.28}\right) = 661.8(\text{hours}).$$

(b) The probability that a spring will operate for 500 hours.

$$P(S(500)) = e^{-(0.0014\times 500)^{1.28}} = 0.531.$$

(c) The probability that a spring that has operated 200 hours w/o failure will operate another 500 hours.

$$S_{T|T\geqslant 200}(700) = \frac{S(700)}{S(200)} = 0.459.$$

Remark. (a) If

$$T \sim \text{Weibull}(\lambda, \kappa),$$

then $Y = \log T$ has the extreme value distribution w/

$$F_Y(y) = 1 - e^{-e^{\frac{y-u}{b}}},\ -\infty < y < \infty,$$

where $\mu = -\log\lambda$ (location), $b = \frac{1}{\kappa}$ (shape).

Proof.

$$\begin{aligned} F_Y(y) &= P(T \leqslant e^y) \\ &= 1 - e^{-(\lambda e^y)^\kappa} \\ &= 1 - e^{-e^{(y+\log\lambda)\kappa}} \\ &= 1 - e^{-e^{\frac{y-\mu}{b}}}. \end{aligned}$$

□

(b) Weibull dist also has the self-reproducing property. If $T_1, \dots, T_n$ are iid and $T_1 \sim$Weibull(λ, κ). Then

$$\min\{T_1, \dots, T_n\} \sim \text{Weibull}\left(\left(\sum_{i=1}^{n} \lambda_i^{\kappa}\right)^{\frac{1}{\kappa}}, \kappa\right).$$

(c) 3-parameter Weibull dist w/ pdf

$$f(t) = \kappa\lambda^{\kappa}(t-\mu)^{\kappa-1}e^{-\lambda(t-\mu)^{\kappa}},\ t \geqslant \mu.$$

2.4 Gamma Dist

Definition 2.16.

$$T \sim \text{Gamma}(\lambda, \kappa),$$

if its pdf

$$f(t) = \frac{\lambda}{\Gamma(\kappa)}(\lambda t)^{\kappa-1}e^{-\lambda t}, t > 0.$$

Then

$$F(t) = \int_0^t \frac{\lambda}{\Gamma(\kappa)}(\lambda\tau)^{\kappa-1}e^{-\lambda\tau}d\tau = \frac{1}{\Gamma(\kappa)}\int_0^{\lambda t} \lambda^{\kappa-1}e^{-x}dx = I(k, \lambda t), t > 0.$$

Also

$$E[T^r] = \frac{\kappa(\kappa+1)\cdots(\kappa+r-1)}{\lambda^r},\ r = 1, 2, \dots.$$

Then

$$E[T] = \frac{\kappa}{\lambda}, \text{Var}(T) = \frac{\kappa}{\lambda^2},$$

and then

$$\nu = \kappa^{-\frac{1}{2}},\ \nu_3 = 2\kappa^{-\frac{1}{2}},\ \nu_4 = 3 + \frac{6}{\kappa}.$$

Remark. It is difficult to differentiate b/w Gamma and Weibull dist based on pdf but difference become apparent when their hazard function is compared.

Remark. Special cases:

(a)

$$\text{Gamma}\left(\frac{1}{2}, \kappa\right) = \chi^2_{2\kappa}.$$

(b)

$$\text{Gamma}(\lambda = n, \kappa) = \text{Erlang}(n),$$

where $n \in \mathbb{Z}^+$.

Example 2.17. Other lifetime distributions:

(a) log-normal / pdf

$$\frac{1}{\sigma t\sqrt{2\pi}}e^{-\frac{(\log t-\mu)^2}{2\sigma^2}},\ t > 0.$$

(b) Inverse Gaussian w/ pdf

$$\sqrt{\frac{\lambda}{2\pi t^3}} e^{-\frac{\lambda}{2\mu^2 t}(t-\mu)^2},\ t > 0.$$

(c) Exponential power w/ pdf

$$e^{1-e^{-\lambda t^\kappa}} e^{-\lambda t^\kappa} \lambda \kappa t^{\kappa-1},\ t > 0.$$

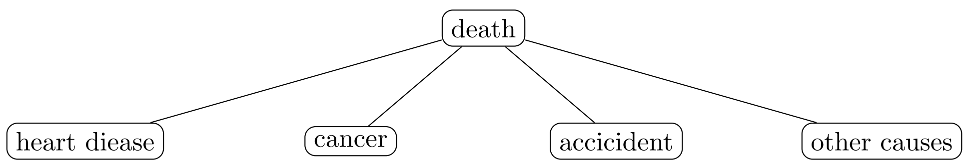
(d) Pareto w/ pdf

$$\frac{\kappa \lambda^\kappa}{t^{\kappa+1}},\ t > 0.$$

Chapter 3

Specialized models

3.1 Competing risks



Causes of failure may be grouped into k causes. In competing risk analysis, an item is assumed to be subject to k competing risks (or causes) denoted by $C_1, \ldots, C_k$.
Competing risk can be viewed as a series system of components.
Each risk can be thought of as a component in a series system in which the system failure occurs when a system component fails.

Definition 3.1 (net life)**.** Let X_j having $f_{X_j}(t), S_{X_j}(t)$, $h_{X_j}(t)$, $H_{X_j}(t)$ and corresponding risk C_j be the net life denoting the lifetime that occurs if only risk j is present,$\forall j = 1, \ldots, k$.

Remark. (a) Unless all risks except j are eliminated, X_j is NOT necessarily observed.

(b) Each net life is potential lifetime that is observed w/ certainty only if other $k - 1$ risks are eliminated.

(c) The actual observed life

$$T = \min\{X_1, \ldots, X_k\}.$$

Proposition 3.2. Assume that net lives are indep, then

$$h_T(t) = \sum_{j=1}^{k} h_{X_j}(t).$$

Proof.

$$
\begin{aligned}
H_T(t) &= -\log S_T(t) \\
&= -\log \prod_{i=1}^{k} S_{X_i}(t) \\
&= -\sum_{i=1}^{k} \log S_{X_i}(t) \\
&= \sum_{i=1}^{k} h_{X_j}(t).
\end{aligned}
$$

□

Definition 3.3. The net probability of failure in $[a, b)$ from risk j

$$
\begin{aligned}
q_j(a,b) &= P(a \leqslant X_j < b | X_j \geqslant a) \\
&= 1 - P(X_j \geqslant b | X_j \geqslant a) \\
&= 1 - \frac{P(X_j \geqslant b)}{P(X_j \geqslant a)} \\
&= 1 - \frac{S_{X_j}(b)}{S_{X_j}(a)} \\
&= 1 - \frac{e^{-H_{X_j}(b)}}{e^{-H_{X_j}(a)}} \\
&= 1 - e^{-\int_a^b h_{X_j}(t)dt}.
\end{aligned}
$$

Definition 3.4 (Crude life)**.** Let Y_j having $f_{Y_j}(t), S_{Y_j}(t)$, $h_{Y_j}(t)$, $H_{Y_j}(t)$ and corresponding C_j be the crude life denoting the lifetime conditioned on risk j being the cause of failure in the presence of all other risks.

Definition 3.5. The crude probability of failure in $[a, b)$ from causes j

$$
\begin{aligned}
Q_j(a,b) &= P(a \leqslant X_j \leqslant b, X_j < X_i, \forall i \neq j | T \geqslant a) \\
&= \int_a^b h_{X_j}(x) e^{-\int_a^x h_T(x)dx} dx, \; j \in [k].
\end{aligned}
$$

Definition 3.6. The probability of failure due to risk j,

$$\pi_j = P(X_j = T).$$

Clearly,

$$\sum_{j=1}^{k} \pi_j = 1,$$

and

$$\pi_j = Q_j(0, \infty).$$

Example 3.7. The causes of calulator failure are C_1 w/ net life X_1 and C_2 w/ net life X_2.

X_1 is the lifetime of the calculator if the only way it can fail is by being dropped.

X_2 is the lifetime of the calculator if it is bolted to a desk and can not be dropped.

Y_1 is the lifetime of the calculator that failed due to being dropped in the presence of other causes of failure.

Y_2 is the lifetime of the calculator that did not fail by being dropped, but was bolted to a desk to avoid its being dropped.

Clearly, the obverved lifetime

$$T = \min\{X_1, X_2\}.$$

Assume $X_1 \sim \exp(1)$ and $X_2 \sim \exp(2)$, are independent. Then

$$h_{X_1}(t) = 1, t > 0, \ h_{X_2}(t) = 2, t > 0.$$

Then

$$q_1(a,b) = 1 - e^{-\int_a^b 1dt} = 1 - e^{-(b-a)},$$

and similarly,

$$q_2(a,b) = 1 - e^{-\int_a^b 2dt} = 1 - e^{-2(b-a)}.$$

Then

$$\begin{aligned} Q_1(a,b) &= P(a \leqslant X_1 < b, X_1 < X_2 | T \geqslant a) \\ &= \frac{P(a \leqslant X_1 < b, X_1 < X_2)}{P(X_1 > a, X_2 > a)} \\ &= \frac{\int_a^b \int_{x_1}^\infty e^{-x_1} 2e^{-2x_2} dx_1 dx_2}{\int_a^\infty \int_a^\infty e^{-x_1} 2e^{-2x_2} dx_1 dx_2} \\ &= \frac{\frac{1}{3}\left(e^{-3a} - e^{-3b}\right)}{e^{-3a}} \\ &= \frac{1}{3}\left[1 - e^{-3(b-a)}\right], \end{aligned}$$

and similarly

$$Q_2(a,b) = \frac{2}{3}\left[1 - e^{-3(b-a)}\right].$$

Then

$$\pi_1 = P(\text{failure due to risk 1}) = P(X_1 = T) = P(X_1 < X_2) = \int\int_{X_1 \leqslant X_2} f(x_1, x_2) dx_1 dx_2 = \frac{1}{3}.$$

Alternative:

$$\pi_1 = Q_1(0, \infty) = \frac{1}{3}.$$

Similarly,

$$\pi_2 = \frac{2}{3}.$$

Now consider the distribution of the crude life Y_1 and Y_2,

$$\begin{aligned}
S_{Y_1}(y_1) &= P(T \geqslant y_1 | X_1 = T) \\
&= \frac{P(T \geqslant y_1, X_1 = T)}{P(X_1 = T)} \\
&= \frac{P(X_1 \geqslant y_1, X_1 < X_2)}{\pi_1} \\
&= \frac{\int_y^\infty \int_{x_1}^\infty e^{-x_1} 2e^{-2x_2} dx_1 dx_2}{\frac{1}{3}} \\
&= e^{-3y_1}, \ y_1 > 0.
\end{aligned}$$

Then

$$f_{Y_1}(y) = 3e^{-3y_1}.$$

Hence

$$Y_1 \sim \exp(3).$$

Similarly,

$$Y_2 \sim \exp(3).$$

Remark. There are two ways to generate T:

M_1: generate an exp(3).

M_2: generate an exp(1) and an exp(2), take the minimum (for each). (Then $\frac{1}{3}$ from exp(2) and $\frac{2}{3}$ from exp(2).) (Not related to mean.)

3.2 Compute Marginal from Joint

Let $X_1, X_2, \dots, X_k$ be cont. net lives and

$$T = \min\{X_1, \dots, X_k\}$$

be the observed failure time of the item. Assume the joint pdf is

$$f(x_1, \dots, x_k).$$

Then the joint survival function is

$$\begin{aligned}
S(x_1, \dots, x_k) &= P(X_1 \geqslant x_1, \dots, X_k \geqslant x_k) \\
&= \int_{x_1}^\infty \cdots \int_{x_k}^\infty f(t_1, \dots, t_k) dt_1 \cdots dt_k.
\end{aligned}$$

The marginal net survival function is

$$S_{X_j}(x_j) = P(X_j \geqslant x_j) = S(0, \dots, 0, x_j, 0, \dots, 0).$$

The sur. fcn for the observed time T is

$$S_T(t) = P(T \geqslant t) = P(X_1 \geqslant t, \dots, X_k \geqslant t) = S(t, \dots, t).$$

Also,

$$-\frac{\delta}{\delta X_j}S(x_1,\ldots,x_k)=\lim_{\Delta x\to 0}\frac{S(x_1,\ldots,x_j,\ldots,x_k)-S(x_1,\ldots,x_j+\Delta x,\cdots x_k)}{\Delta x}.$$

We can show that the prob. of failure from risk j is

$$\pi_j=\int_0^\infty\left[-\frac{\delta}{x_j}S(x_1,\ldots,x_k)\Big|_{x_1=\cdots=x_n=t}\right]dt,\forall j=1,\ldots,k.$$

Now consider the sur. fcn. of crude lifetime $Y_1,\ldots,Y_k$. Let $r.v.$ J be index of the cause of failure, then

$$\begin{aligned}P(T\geqslant t,J=j)&=P(X_j\geqslant t,X_j\leqslant X_i,\forall i\neq j)\\&=\int_t^\infty\int_{x_j}^\infty\cdots\int_{x_j}^\infty f(x_1,\ldots,x_k)dx_1\cdots dx_{j-1}d_{x_{j+1}}\ldots dx_kd_j.\end{aligned}$$

Then the sur. fcn. for T is

$$S_T(t)=P(T\geqslant t)=\sum_{j=1}^k P(T\geqslant t,J=j).$$

Also,

$$\pi_j=P(X_j=T)=P(J=j)=P(T\geqslant 0,J=j),\forall j\in[n].$$

The sur. fcn. of the crude life Y_j is

$$\begin{aligned}S_{Y_j}(y_j)&=P(T\geqslant y_j|J=j)\\&=\frac{P(T\geqslant y_j,J=j)}{P(J=j)}\\&=\frac{P(T\geqslant y_j,T=j)}{\pi_j},\forall j=1,\ldots,k.\end{aligned}$$

Example 3.8 (Cont Example 3.7)**.** Since

$$S_{X_1}(t)=e^{-t},\ S_{X_2}(t)=e^{-2t},\ t>0,$$

we have

$$S(x_1,x_2)=S_{X_1,X_2}(x_1,x_2)=e^{-x_1-2x_2},\ x_1>0,x_2>0.$$

Hence

$$\pi_1=\int_0^\infty\left[-\frac{\delta}{\delta x_1}S(x_1,x_2)\Big|_{X_1=X_2=t}\right]dt=\int_0^\infty e^{-3t}dt=\frac{1}{3},$$

which is the same as our previous result. The survival prob. conditioning on risk 1 being the cause of failure

$$\begin{aligned}P(T\geqslant t,J=1)&=P(X_1>t,X_1<X_2)\\&=\int_t^\infty\int_{x_1}^\infty f(x_1,x_2)dx_2dx_1\\&=\frac{1}{3}e^{-3t},\ t>0.\end{aligned}$$

Similarly,

$$P(T \geqslant t, J = 2) = \frac{2}{3}e^{-3t},\ t > 0.$$

Thus, the sur. fcn. for the 1st crude lifetime is

$$S_{Y_1}(y_1) = \frac{P(T \geqslant y_1, J = 1)}{\pi_1} = e^{-3y_1},\ y_1 > 0,$$

and similarly,

$$S_{Y_2}(y_2) = e^{-3y_2},\ y_2 > 0.$$

Remark. net lives $X_1, \ldots, X_k$ (potential lifetimes) $\Rightarrow$ crude lives $Y_1, \ldots, Y_k$ (observed lifetimes), but we don not have the converse direction unless under the assumption of independence of net lives.

Theorem 3.9. *Assume* $\pi_j = P(J = j)$ *and* $S_{Y_j}(t), \forall j \in [n]$ *is known.*
When the risks are indep., the dist of net life of X_j *is*

$$h_{X_j}(t) = \frac{\pi_j f_{Y_j}(t)}{\sum_{i=1}^{k} \pi_i S_{Y_i}(t)},\ t > 0, \forall j = 1, \ldots, k.$$

Example 3.10 (Cont Example 3.7)**.** Since

$$\pi_1 = P(X_1 = J) = P(J = 1) = \frac{1}{3},$$

and

$$\pi_2 = P(J = 2) = \frac{2}{3},$$

and

$$S_{Y_1}(t) = e^{-3t},\quad S_{Y_2}(t) = e^{-3t},\ t > 0,$$

we have

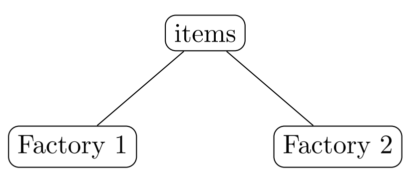
$$h_{X_1}(t) = \frac{\frac{1}{3} \cdot 3e^{-3t}}{\frac{1}{3}e^{-3t} + \frac{2}{3}e^{-3t}} = 1,\ t > 0,$$

and similarly,

$$h_{X_2}(t) = 2,\ t > 0.$$

3.3 Mixture Distributions

Motivations:



but not certain which one the item came from. eg: For factory 1,

$$T_1 \sim f_1(t), \text{ eg. } N(\mu_1, \sigma_1^2),$$

for factory 2,

$$T_2 \sim f_2(t), \text{ eg. } N(\mu_2, \sigma_2^2)).$$

Items $\pi_1(60\%)$ from f_1 and $1-\pi(40\%)$ from f_2. Then what is the distribution of the lifetime of an item? Let

$$S = \begin{cases} 1, & \text{an item is from } F_1 \\ 2, & \text{an item is from } F_2. \end{cases}$$

Then

$$\begin{aligned} F_T(t) &= P(T \leqslant t) \\ &= P(T \leqslant t|S=1)P(S=1) + P(T \leqslant t|S=2)P(S=2) \\ &= P(T_1 \leqslant t)P(S=1) + P(T_2 \leqslant t)P(S=2) \\ &= F_{T_1}(t)\pi + F_{T_2}(t)(1-\pi), \ t > 0 \end{aligned}$$

or

$$f_T(t) = \pi f_{T_1}(t) + (1-\pi) f_{T_2}(t).$$

3.3.1 A finit mixture

When items can be divided into m populations by a characteristic of the item (eg: manufacturing), then

$$f(t) = \sum_{i=1}^{m} p_i f_i(t|\theta_i),$$

where $f_i(t|\theta_i)$ is the pdf of the ith population, $i = 1, \ldots, m$ and p_i are mixture parameter w/ $p_i \geqslant 0$ and $\sum_{i=1}^m p_i = 1$.

Example 3.11. $m = 2$.

$$T_1 \sim \exp(1), \ T_2 \sim \exp(2).$$

$$p_1 = \frac{1}{3}, \ p_2 = \frac{2}{3}.$$

Then the pdf of the lifetime of an item, whose manufacturing site is unknown,

$$f(t) = \frac{1}{3}e^{-t} + \frac{2}{3}e^{-2t}, \ t > 0.$$

The model is a special case of hyper-exponential distributions, which is the finite mixture of m exponential distributions.

Remark. (a) An identifiability problem exists in finite mixture. It is often impossible to determine the component distribution from the distribution of the population.

Example 3.12. (!!!)

$$f_{T_1}(t) = \frac{2}{3} f_{V_1}(t) + \frac{1}{3} f_{V_2}(t),\ 0 \leqslant t \leqslant 1.$$

$$f_{T_2}(t) = \frac{1}{2} f_{V_3}(t) + \frac{1}{2} f_{V_4}(t),\ 0 \leqslant t \leqslant 1,$$

where

$$V_1 \sim U(0,1),\ V_2 \sim U(1/4, 3/4),$$

$$V_3 \sim U(0, 3/4),\ V_4 \sim U(1/4, 1).$$

Clearly, T_1 and T_2 have the same distribution w/ pdf

$$f_{T_1}(t) = f_{T_2}(t) = \begin{cases} 2/3, & 0 < t \leqslant 1/4, \\ 4/3, & 1/4 < t \leqslant 3/4, \\ 2/3, & 3/4 < t \leqslant 1. \end{cases}$$

(b) Combining competing risks and finite mixture theory, we have

$$f_T(t) = \sum_{i=1}^{m} p_i \left[\sum_{j=1}^{k_i} h_{ij}(t) e^{-\int_0^t \sum_{j=1}^{k_i} h_{ij}(t)} \right],$$

where m is the number of populations, $\sum_{i=1}^m p_i = 1$, k_i is the number of risks acting within the ith population. $h_{ij}(t)$ is the hazard function for the jth risk within $i$the population with $i = 1, \ldots, m$ and $j = 1, \ldots, k_i$.

Example 3.13. In casualty insurance application.
$m = 3$ of dwelling (single family, condo and apartment), each is subjected to and insured for $k_1 = k_2 = k_3 = 5$ risks (fire, float, tornado, earthquake, bomb.)

3.3.2 Continuous mixture

$$f(t) = \int_\Theta f(t|\theta) p(\theta) d\theta,$$

where θ is calling the mixture parameter. $p(\theta)$ indicates the distribution pf mixture parameter. Note that if $p(\theta)$ is a finite pmf, then the above becomes the finite mixture. Another representation

$$F(t) = \int_\Theta F(t|\theta) dG(\theta)$$

is often called the mixture of

$$F = \{F(t|\theta) | \theta \in \Theta\} \text{ w.r.t. } G,$$

where G can be discrete or continuous since a *d.f.* always exists.

Remark. It may be worthy noting the above representation need not be proper even though $F(t|\theta)$ is proper for all θ.

Example 3.14. For example, $F(t|\theta)$ is an exponential distribution and $G(\theta)$ is a Poisson distribution, then we can show that $\lim_{t\to\infty} F(t) < 1$.

The corresponding survival function is

$$S(t) = 1 - F(t) = \int_{\Theta} (1 - F(t|\theta))dG(\theta) = \int_{\Theta} S(t|\theta)dG(\theta).$$

The corresponding pdf if exists, is

$$f(t) = \int_{\Theta} f(t|\theta)dG(\theta).$$

The hazard function if the pdf exists, is

$$h(t) = \frac{f(t)}{S(t)} = \frac{\int_{\Theta} f(t|\theta)dG(\theta)}{\int_{\Theta} S(t|\theta)dG(\theta)}.$$

Proposition 3.15. Let

$$h(t|\theta) = \frac{f(t|\theta}{S(t|\theta)}.$$

If

$$h_l \leqslant h(t|\theta) \leqslant h_u, \forall \theta \in \Theta,$$

then

$$h_l \leqslant h(t) \leqslant h_u, \forall t.$$

Example 3.16. Let's consider the special case in which G puts mass at but two points.

$$F(t) = \pi F_1(t) + (1 - \pi)F_2(t).$$

Let I be a r.v., independent of taking only values 0,1 for which

$$P(I = 1) = \pi, \ P(I = 0) = 1 - \pi.$$

Then the mixture

$$T \overset{d}{=} IT_1 + (1 - I)T_2,$$

where, be careful, I is a r.v..

Proof.

$$\begin{aligned}
&P(IT_1 + (1 - I)T_2 \leqslant t) \\
=& P(IT_1 + (1 - I)T_2 \leqslant t|I = 1)P(I = 1) + P(IT_1 + (1 - I)T_2 \leqslant t|I = 0)P(I = 0) \\
=& P(T_1 \leqslant t)\pi + P(T_2 \leqslant t)(1 - \pi) \\
=& \pi F(t) + (1 - \pi)F(t) \\
=& F(t).
\end{aligned}$$

□

Note that

$$\begin{aligned}
h(t) &= \frac{\pi f_1(t) + (1 - \pi)f_2(t)}{\pi S_1(t) + (1 - \pi)S_2(t)} \\
&= w(t)h_1(t) + (1 - w(t))h_2(t),
\end{aligned}$$

where

$$w(t) = \frac{\pi S_1(t)}{\pi S_1(t) + (1-\pi)S_2(t)}.$$

It follows that

$$\min\{h_1(t), h_2(t)\} \leqslant h(t) \leqslant \max\{h_1(t), h_2(t)\},$$

since $h_1(t), h_2(t)$ are bounded.

Proposition 3.17. Suppose that $h(t)$ and $w(t)$ are given above. If $h_1(t) \leqslant h_2(t)$, then

$$\frac{d}{dt}w(t) \geqslant 0,$$

w/ strict inequality of $S_1(t)S_2(t) > 0$.

Proof. Note that

$$\frac{d}{dt}w(t) = \frac{\pi(1-\pi)S_1(t)S_2(t)\left[h_2(t) - h_1(t)\right]}{\left[\pi S_1(t) + (1-\pi)S_2(t)\right]^2} \geqslant 0.$$

□

3.3.3 Mixture and Minima

Suppose $T = \min\{T_1, T_2\}$ and T_i has distribution G_i, $i = 1, 2$.
If T_1 and T_2 are independent, then survival function of T is

$$S(t) = S_1(t)S_2(t).$$

This is a simple version of the competing risk model discussed before. Here T_1, T_2 can be regarded as potential waiting time to failure due to two different causes and the actual failure time is their minimum. Since there are two different possible causes of failure, the distribution can be thought of as a mixture.

$$S(t) = S_1(t)S_2(t) = \pi S_1^*(t) + (1-\pi)S_2^*(t)$$

can be solved to obtain

$$S_1^*(t) = \frac{\int_t^\infty S_2(t)dG_1(t)}{\int_0^\infty S_2(t)dG_1(t)},$$

$$S_2^*(t) = \frac{\int_t^\infty S_1(t)dG_2(t)}{\int_0^\infty S_1(t)dG_2(t)};$$

and

$$\pi = \int_0^\infty S_2(\tau)dG_1(\tau),$$

$$1 - \pi = \int_0^\infty S_1(\tau)dG_2(\tau).$$

Example 3.18. Suppose

$$T \sim \exp(\lambda_1),$$

and assume

$$\lambda_1 \sim \exp(\lambda_2).$$

Then

$$\begin{aligned} f_{T|\Lambda_1}(t|\lambda_1) &= \lambda_1 e^{-\lambda_1 t},\ t > 0, \\ f_{\Lambda_1}(\lambda_1) &= \lambda_2 e^{-\lambda_2 \lambda_1},\ \lambda_1 > 0. \end{aligned}$$

The unconditional pdf

$$\begin{aligned} f_T(t) &= \int_0^\infty f(T_1|\Lambda_1)(t|\lambda_1) f_{\Lambda_1}(\lambda_1) d\lambda_1 \\ &= \lambda_2 \int_0^\infty \lambda_1 e^{-\lambda_1(t+\lambda_2)} d\lambda_1 \quad (!!!) \\ &= \frac{\lambda_2}{(\lambda_2 + t)^2},\ t > 0, \end{aligned}$$

by regarding it as a Gamma integral.

Example 3.19. Let

$$T \sim \mathrm{Poi}(\lambda)$$

w/ pmf

$$f_{T|\Lambda}(t|\lambda) = \frac{\lambda^t e^{-\lambda}}{t!},\ t = 0, 1, \cdots.$$

Suppose Λ is a r.v. having the Gamma distribution w/ shape κ and scale δ w/ pdf

$$p_\Lambda(\lambda) = \frac{\delta^\kappa}{\Gamma(\kappa)} \lambda^{\kappa-1} e^{-\delta\lambda},\ \lambda > 0,\ (\kappa > 0, \delta > 0).$$

The unconditional pdf is

$$\begin{aligned} f_T(t) &= \int_0^\infty \frac{\lambda^t e^{-\lambda}}{t!} \frac{\delta^\kappa}{\Gamma(\kappa)} \lambda^{\kappa-1} e^{-\delta\lambda} d\lambda \\ &= \frac{\delta^\kappa}{\Gamma(\kappa) t!} \int_0^\infty \lambda^{t+\kappa-1} e^{-(1+\delta)\lambda} d\lambda \\ &= \frac{\Gamma(t+\kappa)\delta^\kappa (1+\delta)^{-(t+\kappa)}}{\Gamma(\kappa) t!},\ t = 0, 1, \ldots, \end{aligned}$$

which is called Gamma-Poisson distribution.

Example 3.20. Student t-dsitribution.

$$X \sim t(\nu)$$

if pdf is

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma} \left[1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right]^{-\frac{\nu+1}{2}},\ x \in \mathbb{R},$$

where μ : location, σ : scale, ν : shape.

Remark. (a) When $\nu = 1$,

$$f(x) = \frac{1}{\pi\sigma\left[1 + \frac{(x-\mu)^2}{\sigma^2}\right]}, \ x \in \mathbb{R},$$

which is the pdf of the Cauchy distribution.

(b) When $\nu \to \infty$,

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ x \in \mathbb{R}.$$

(c)

$$\int_0^1 N\left(x|\mu, (\lambda\tau)^{-1}\right) \text{Gamma}\left(\tau \middle| \frac{\nu}{2}, \frac{\nu}{2}\right) d\tau = t_\nu\left(x|\mu, \lambda^{-1}\right). \quad (\lambda = \sigma^2)$$

Chapter 4

Bayesian Inference

Recall mixture pdf

$$f(t) = \int_\Theta f(t|\theta)p(\theta)d\theta,$$

we can regard it as a joint pdf.

Definition 4.1 (MLE)**.** Likelihood inference (classical or frequentist inference)

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x|\theta),$$

where $f(x|\theta)$ is pdf or pmf on the probability space

$$(\Omega, \mathcal{F}, P_\theta)$$

and

$$\theta \in \Theta.$$

The sampling distribution

$$f(x_1, \ldots, x_n|\theta) = f(\underset{\sim}{x}|\theta)$$

is the distribution of the observed data conditional on its parameters. This is also termed the likelihood, especially when viewed as a function of the parameter(s),

$$L(\theta|x_1, \ldots, x_n) = f(x_1, \ldots, x_n|\theta).$$

Write it as

$$L(\theta|\underset{\sim}{x}) = f(\underset{\sim}{x}|\theta) = \prod_{i=1}^n f(x_i|\theta),$$

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta \in \Theta} L(\theta|\underset{\sim}{x}).$$

Definition 4.2 (Fisher information)**.** Fisher information of a random variable X with probability measure P_θ from the family $\{P_\theta : \theta \in \Theta\}$ is defined by

$$I(\theta_0) := E_{\theta_0}\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\bigg|_{\theta=\theta_0}\right)^2 \overset{\text{exp}}{=} -E_{\theta_0}\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\bigg|_{\theta=\theta_0}\right).$$

Remark. (a)

$$l(\theta|\underset{\sim}{x}) = \log L(\theta|\underset{\sim}{x}) = \sum_{i=1}^{n} \log f(x_i|\theta).$$

Set

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta\in\Theta} l(\theta|\underset{\sim}{x}),$$

since log is an increasing function.

(b) $\hat{\theta}_{\mathrm{MLE}}$ usually satisfies consistency and asymptotic normality.

- *Consitency.*

  $$\hat{\theta}_{\mathrm{MLE}} \xrightarrow{p} \theta_0,$$

  for some $\theta_0 \in \mathbb{R}$.
- *Asymptotic Normality.*

  $$\sqrt{n}(\hat{\theta}_{\mathrm{MLE}} - \theta_0) \overset{n\to\infty}{\sim} N(0, I^{-1}(\theta_0)),$$

  where $I(\theta_0)$ is the fisher information.

(c) The above can be used to construct a $100(1-\alpha)\%$ interval for θ.

$$\sqrt{n} I^{\frac{1}{2}}(\theta_0)(\hat{\theta}_{\mathrm{MLE}} - \theta_0) \sim N(0,1).$$

$$P\left(-z_{1-\frac{\alpha}{2}} \leqslant \sqrt{n} I^{\frac{1}{2}}(\theta_0)(\hat{\theta}_{\mathrm{MLE}} - \theta_0) \leqslant z_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

So

$$\theta_0 \in \left[\hat{\theta}_{\mathrm{MLE}} - z_{1-\frac{\alpha}{2}} n^{-\frac{1}{2}} I^{-\frac{1}{2}}(\theta_0),\ \hat{\theta}_{\mathrm{MLE}} + z_{1-\frac{\alpha}{2}} n^{-\frac{1}{2}} I^{-\frac{1}{2}}(\theta_0)\right],$$

where θ_0 on the right side can be replaced by $\hat{\theta}_{\mathrm{MLE}}$. Then $100(1-\alpha)\%$ CI

$$\theta_0 \in \left[\hat{\theta}_{\mathrm{MLE}} - z_{1-\frac{\alpha}{2}} n^{-\frac{1}{2}} I^{-\frac{1}{2}}(\hat{\theta}_{\mathrm{MLE}}),\ \hat{\theta}_{\mathrm{MLE}} + z_{1-\frac{\alpha}{2}} n^{-\frac{1}{2}} I^{-\frac{1}{2}}(\hat{\theta}_{\mathrm{MLE}})\right].$$

Example 4.3.

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta, 1).$$

Then the pivot quantity

$$\sqrt{n}(\overline{X} - \theta) \sim N(0,1).$$

So $100(1-\alpha)\%$ CI for θ

$$\left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{n}},\ \overline{X} + z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{n}}\right],$$

where we can regard the CI a random variable, since the population mean θ is a constant, there is 95% probability that the CI will cover the population mean θ. For each sample, the corresponding CI will cover θ or not cover the true mean θ, about 95% of the observed intervals $\left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{n}},\ \bar{x} + z_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{n}}\right]$ will cover θ.

4.1 Bayesian inference

The likelihood fcn is viewed as the conditional dist of data given the parameter, and assume the parameter θ is a r.v. following some dist called the prior. So we have the Bayesian model

$$\begin{cases} X_1, \ldots, X_n|\theta \sim f(x_1, \ldots, x_n|\theta) \\ \theta \sim \pi(\theta). \end{cases}$$

Then we have the marginal dist of the data

$$m(\underset{\sim}{x}) = \int_\Theta f(\underset{\sim}{x}|\theta)\pi(\theta)d\theta,$$

which is referred to as the *prior predictive dist* of the data. If we do not observe any data, the prior predictive dist is the relevant dist for making probability statements about the *unknown value of the data*. Similarly, the prior $\pi(\theta)$ is the relevant dist to use in making probability statements about θ before we observe the data. Then Bayesian inference is based on the conditional dist of θ given the data.

- Likelihood fcn:
$$\underset{\sim}{x}|\theta \sim f(\underset{\sim}{x}|\theta).$$
- Prior:
$$\pi \sim \pi(\theta).$$
- Posterior dist:
$$\pi(\theta|\underset{\sim}{x}) = \frac{f(\underset{\sim}{x}|\theta)\pi(\theta)}{m(\underset{\sim}{x})},$$
where
$$m(\tilde{x}) = \int_\Theta f(\underset{\sim}{x}|\theta)\pi(\theta)d\theta,$$
and is called prior predictive dist or mixture dist,
and
$$\int_\Theta \pi(\theta|\underset{\sim}{x})d\theta = 1.$$

Theorem 4.4 (Bayesian Theorem)**.**

$$\pi(\theta|\underset{\sim}{x}) = \frac{f(\underset{\sim}{x}|\theta)\pi(\theta)}{m(\underset{\sim}{x})} = \frac{f(\underset{\sim}{x}|\theta)\pi(\theta)}{\int_\Theta f(\underset{\sim}{x}|\theta)\pi(\theta)d\theta}.$$

Then the key of Bayesian inference is updating the prior distribution $\pi(\theta)$ using the posterior distribution $\pi(\theta|\underset{\sim}{x})$.

Remark. It can be shown that for the two data set $\underset{\sim}{x_1}$ and $\underset{\sim}{x_2}$, if we use them together to compute our posterior dist $\pi(\theta|\underset{\sim}{x}_1, \underset{\sim}{x}_2)$ given the prior $\pi(\theta)$, which has the same dist as posterior dist $\pi(\theta|\underset{\sim}{x}_2)$ given the prior $\pi(\theta|\underset{\sim}{x}_1)$, where $\pi(\theta|\underset{\sim}{x}_1)$ is the posterior dist given the prior $\pi(\theta)$.

Example 4.5.

$$\begin{cases} X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli } \theta \\ \theta \sim \text{Beta}(\alpha, \beta) \text{ w/ known } \alpha, \beta. \end{cases}$$

(If α is unknown, we can use hierachy mixture dist.)
We know

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

The marginal

$$\begin{aligned} m(\underset{\sim}{x}) &= \int_0^1 \prod_{i=1}^n \theta^{x_i}(1-\theta_i)^{1-x_i} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta \\ &= \frac{1}{B(\alpha,\beta)} \int_0^1 \theta^{\sum_{i=1}^n x_i + \alpha - 1}(1-\theta)^{n-\sum_{i=1}^n x_i + \beta - 1} d\theta \\ &= \frac{B\left(\sum_{i=1}^n x_i + \alpha,\ n - \sum_{i=1}^n x_i + \beta\right)}{B(\alpha,\beta)}. \end{aligned}$$

So the posterior dist of θ

$$\begin{aligned} \pi(\theta|\underset{\sim}{x}) &= \frac{f(\underset{\sim}{x}|\theta)\pi(\theta)}{m(\underset{\sim}{x})} \\ &= \frac{\prod_{i=1}^n \theta^{x_i}(1-\theta_i)^{1-x_i} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}}{\frac{B\left(\sum_{i=1}^n x_i+\alpha,\ n-\sum_{i=1}^n x_i+\beta\right)}{B(\alpha,\beta)})} \\ &= \frac{\theta^{\sum_{i=1}^n x_i+\alpha-1}(1-\theta)^{n-\sum_{i=1}^n x_i+\beta-1}}{B\left(\sum_{i=1}^n x_i+\alpha,\ n-\sum_{i=1}^n x_i+\beta\right)}. \end{aligned}$$

Thus,

$$\theta|x \sim \text{Beta}\left(\sum_{i=1}^n x_i + \alpha, n - \sum_{i=1}^n x_i + \beta\right).$$

Let

$$\alpha = \beta = 1,$$

then the prior

$$\theta \sim U(0,1).$$

Let

$$n = 40,\ \sum_{i=1}^n x_i = 10.$$

$$\theta|\underset{\sim}{x} \sim \text{Beta}(11, 31).$$

If we plot the graph, we get it approximately concentrated at $\frac{10}{40} = 0.25$.
We can also make hypothesis about, for example, $\theta = 1$ or $\theta \neq 1$.

Proposition 4.6.

$$\pi(\theta|\underset{\sim}{x}) \propto f(\underset{\sim}{x}|\theta)\pi(\theta).$$

Example 4.7.

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \tau_0^2).$$

$$\mu \sim N(\mu_0, \tau_0^2).$$

Find the posterior dist of u.
(indexing by 0 denotes it is known.)

$$\begin{aligned}
\pi(\mu|\underset{\sim}{x}) &\propto f(\underset{\sim}{x}|\mu)\pi(\mu) \\
&\propto \prod_{i=1}^{n} e^{-\frac{(x_i-\mu)^2}{2\sigma_0^2}} e^{-\frac{(\mu-\mu_0)^2}{2\tau_0^2}} \\
&= e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma_0^2} - \frac{(\mu-\mu_0)}{2\tau_0^2}} \\
&\propto e^{-\frac{n\mu^2-2n\bar{x}\mu}{2\sigma_0^2} - \frac{\mu^2-2\mu_0\mu}{2\tau_0^2}} \\
&= e^{-\frac{1}{2}\left[\left(\frac{n}{\sigma_0^2}+\frac{1}{\tau_0^2}\right)\mu^2 - 2\left(\frac{n}{\sigma_0^2}\bar{x}+\frac{\mu_0}{\tau_0^2}\right)\mu\right]} \\
&\propto exp\left\{-\frac{\left(\mu - \frac{\frac{n}{\sigma_0^2}\bar{x}+\frac{\mu_0}{\tau_0^2}}{\frac{n}{\sigma_0^2}+\frac{1}{\tau_0^2}}\right)^2}{2\left(\frac{n}{\sigma_0^2}+\frac{1}{\tau_0^2}\right)^{-1}}\right\},
\end{aligned}$$

so

$$\mu|\underset{\sim}{x} \sim N\left(\mu^*, \sigma^{*2}\right),$$

w/

$$u^* = \frac{\frac{n}{\sigma_0^2}\bar{x} + \frac{\mu_0}{\tau_0^2}}{\frac{n}{\sigma_0^2} + \frac{1}{\tau_0^2}},$$

and

$$\sigma^{*2} = \left(\frac{n}{\sigma_0^2} + \frac{1}{\tau_0^2}\right)^{-1}.$$

Then

$$m(\underset{\sim}{x}) = \frac{f(\tilde{x}|\mu)\pi(\mu)}{\pi(\mu|\tilde{x})} = \cdots.$$

We have the precision

$$\frac{1}{\sigma^{*2}} = \frac{n}{\sigma_0^2} + \frac{1}{\tau_0^2} = \frac{1}{\frac{\sigma_0^2}{n}} + \frac{1}{\tau_0^2}. = \text{precision of } \bar{x} + \text{ presision of } \mu.$$

The no info about μ, then always

$$\mu \propto 1.$$

If $\tau_0 \to \infty$, the graph is flat, then

$$\mu^* \to \bar{x},$$

and

$$\sigma^{*2} \to \frac{\sigma_0^2}{n}.$$

eg:

$$\sigma_0^2 = 1,\ \tau_0^2 = 2, \mu_0 = 0,$$
$$n = 10,\ \bar{x} = 1.2.$$

Then

$$\mu|\underset{\sim}{x} \sim N(1.1429, 0.9524),$$

w/ 1.1429 close to 1.2.

Bayesian sufficient statistic is the same as the classical sufficient statistic. So if we use $\bar{x}$ instead of $\underset{\sim}{x}$, we get the same result. Next, make inference for unknown parameter from the posterior dist. Note

$$\theta|\underset{\sim}{x} \sim \pi(\theta|\underset{\sim}{x}).$$

Then

$$\hat{\theta} = E[\theta|\underset{\sim}{x}] = \int_\Theta \theta\pi(\theta|\underset{\sim}{x})d\theta.$$

For our example,

$$\hat{\mu} = E[\mu|\underset{\sim}{x}] = \frac{\frac{n}{\sigma_0^2}\bar{x} + \frac{\mu_0}{\tau_0^2}}{\frac{n}{\sigma_0^2} + \frac{1}{\tau_0^2}} = \mu^*.$$

median: $t_{0.5}$. mode:

$$\hat{\theta} = \arg\max \pi(\theta|\underset{\sim}{x})$$

Definition 4.8. We use *credible interval* to find the interval estimation for θ.
(Use condidence interval to make hypothesis.)

Remark. (a) We can also predict the future obs. in Bayesian stat.

$$x^*|\theta \sim f(x^*|\theta),$$
$$\theta|\underset{\sim}{x} \sim \pi(\theta|\underset{\sim}{x}).$$

Then

$$p(x^*|\underset{\sim}{x}) = \int_\Theta f(x^*|\theta)\pi(\theta|\underset{\sim}{x})d\theta,$$

where is the posterior predictive dist.

(b) Use monte Carlo simulation.

Chapter 5

Statistical inference

Definition 5.1. A *statistic* is any quantity whose value can be calculated from sample data. A statistic is a random variable denoted by an uppercase letter; a lowercase letter is used to represent the calculated or observed value. of the statistic.

Example 5.2. Both sample mean $\overline{X}$ and sample variance S^2 are in fact random variables since the are the results of ramdom experiments, and since their value can be calculated from sample data, they are statistics. Both the sample mean and variance vary from one sample to the next, and hence they have a distribution!

Definition 5.3. The probability distribution of a statistics is called *sampling distribution*, which describes how the statistic varies in value across all samples that might be selected.

Definition 5.4. The *pivot quantity* is a function of observations and unobservable parameters whose probability distribution does not depend on the unknown parameters. Note that a pivot quantity need not be a statistic. If it is a statistic, then it is known as an ancillary statistic, where *ancillary statistic* is a statistic whose sampling distribution does not depend on the parameters of the model.

Definition 5.5. Then r.v.'s $X_1, \ldots, X_n$ are said to form a random sample of size n if

(a) The X_i's are independent r.v.'s.
Selection of any one (student, chip, light bulb, etc.) has no influence on the others.

(b) Every X_i has the same probability distribution.
Each unit in the sample has equal chance of being chosen.

Remark.

- These two conditions can be combined by saying that X_i's are iid.
- These conditions can be satisfied exactly only if population size is infinite or the sampling is done with replacement. Otherwise, if sample size n is at most 5% of population size N, we can usually assume that X_is are iid.

We typically obtain the sampling distribution through a simulation experiment for which we need to determine

- The statistic of interest (the mean and variance of the weight, defect rate, etc.)

- The population distribution (is it normal, uniform, Poisson?)
- The sample size n (how many subjects/objects are there in each sample).
- The number of replications k (how many samples are obtained).

To do this by random sampling

- Obtain k different random samples from your population.
- Calculate the value(s) of the sought statistic(s) for each of the k samples.
- Construct a histogram of k such numbers (i.e., statistics obtained from k samples)
  * The histogram gives an approximate sampling distribution of the statistic.
  * The approximation will approach true distribution as $k \to \infty$.

5.1 Confidence interval

Confidence interval-the probability is not that the parameter is in a particular interval, but that the intervals in repeated experiments will contain the parameter.

Take polling data for elections. When it's reported that a political party is currently getting a specified level of support (say 37%), with an accuracy of plus or minus some amount (say 2%), they normally state that the results are true 19 times out of 20 (that's a 95% confidence level). This means that if they were to repeat the polling 20 times, the true level of support for that political party would fall within 19 intervals out of 20. It does not mean that there's a 95% chance that the true level of support for that political party is within the range of support being quoted (35 to 39%) in that specific poll. Let's say you're building a confidence interval of the mean. The population mean is an unknown constant, not a random variable. The random variables are the sample mean and sample variance used to build the interval, which vary between experiments. In other words it is the interval that varies and which can be considered a "random variable" of sorts. Once values for the sample mean and sample variance have been calculated for an interval, it's not correct to make probability statements about the population mean-that would imply that it's a random variable. The population mean is a constant that either is or isn't in the interval.

5.2 Sufficient statistics

Definition 5.6. Given a set X of of independent identically distributed data conditioned on an unknown parameter θ, a sufficient statistic is a function $T(X)$ whose value contains all the information needed to compute any estimate of the parameter (e.g. a MLE). Due to the factorization theorem (see below), for a sufficient statistic $T(X)$, the joint distribution can be written as $p(X) = h(X)g(\theta, T(X))$. From this factorization, it can easily be seen that the MLE of will interact with X only through $T(X)$.

More generally, the "unknown parameter" may represent a vector of unknown quantities or may represent everything about the model that is unknown or not fully specified. In such a case, the sufficient statistic may be a set of functions, called a jointly sufficient statistic. Typically, there are as many functions as there are parameters. For example, for a Gaussian distribution with unknown mean and variance, the jointly sufficient statistic, from which maximum likelihood estimates of both parameters can be estimated, consists of two functions, the sum of all data points and the sum of all squared data points (or equivalently, the sample mean and sample variance).

Fisher's factorization theorem or factorization criterion provides a convenient characterization of a sufficient statistic. If the probability density function is $f_\theta(x)$, then T is sufficient for θ if and only if nonnegative functions g and h can be found such that

$$f_\theta(x) = h(x)\, g_\theta(T(x)),$$

i.e. the density f can be factored into a product such that one factor, h, does not depend on θ and the other factor, which does depend on θ, depends on x only through $T(x)$.
It is easy to see that if $F(t)$ is a one-to-one function and T is a sufficient statistic, then $F(T)$ is a sufficient statistic. In particular we can multiply a sufficient statistic by a nonzero constant and get another sufficient statistic.

Chapter 6

Accelerated Life Models

$$T_1, \ldots, T_n \overset{iid}{\sim} f(t).$$

Now consider $Z = (Z_1, \ldots, Z_q)^T$ contains q covariates (explanatory variables, they might be treatments, stress, intrinsic, property of an item,etc).
For example,

$$Z_i = \begin{cases} 1, & \text{an item is in the treatment group,} \\ 0, & \text{an item is in the control group.} \end{cases}$$

or Z_i is the dosage in medical setting or turning speed in manufacturing setting. The covariates influence the lifetime dist. They affect the rate of which the item ages. The survival fcn for T in the accelerated life model

$$S(t) = S_0\left(t\psi(z)\right),\ t \geqslant 0,$$

where $S_0(\cdot)$ is a baseline survival fcn and $\psi(t)$ is a link fcn. The baseline dist. corresponding to all the covariates equal to 0. In reliability setting, this is typically the normal operating and for the item, other covariates vectors are often used for accelerated environment. A popular choice of $\psi(z)$ is the log linear link fcn

$$\psi(z) = e^{\beta^T z}.$$

where we have the linear model

$$\log\left(\psi(z)\right) = \beta_1 z_1 + \cdots + \beta_q z_q$$

So accelerated if $\psi(z) > 1$, decelerated if $\psi(z) < 1$. Another way of viewing the accelerated life model is to denote the lifetime as

$$T = T_o/\psi(z),$$

where T_0 is the lifetime under the baseline condition ($z = 0$.)

Example 6.1. $\psi(z) = 2$, then this item moves through time at twice the rate of an item under the baseline condition. If the log linear link fcn is used, then

$$T_0/T_1 = \psi(z) = e^{\beta^T z},$$

or

$$\log T_0 - \log T_1 = \beta_1 z_1 + \cdots + \beta_q z_q.$$

Remark.

$$H(t) = -\log S(t) = -\log S_0\left(t\psi(z)\right) = H_0\left(t\psi(z)\right).$$

$$h(t) = H'(t) = \psi(z)h_0\left(t\psi(z)\right),\ t \geqslant 0.$$

$$f(t) = S(t)h(t) = S_0\left(t\phi(z)\right)\psi(z)h_0\left(t\psi(z)\right) = \phi(z)f_0(t\phi(z)),$$

which is clearly a pdf.

Example 6.2. Consider T_0 ~Weibull(κ, λ), the link fcn $\psi(z) = e^{\beta^T z}$. Find the survival fcn and the prob. hat an item w/ covariates z fails by time 1000.

The baseline survival fcn

$$S(t) = e^{-(\lambda t)^\kappa}.$$

So the sur. fcn. w/ cov. z

$$S(t) = e^{-(\lambda t\psi(t))^\kappa} = e^{-\left(\lambda e^{\beta^T z} t\right)^\kappa},\ t > 0.$$

So

$$T \sim \text{Weibull}\left(\kappa, \lambda e^{\beta^T z}\right).$$

Thus,

$$P(T \leqslant 1000) = 1 - S(1000) = 1 - e^{-\left(1000\lambda e^{\beta^T z}\right)^\kappa}.$$

Example 6.3. Consider $\psi(z) = e^{\beta^T z}$ in an accelerated life model w/ a log logistic baseline fcn. Find the survival fcn as an expression for the mean time to failure for an item w/ convariance z.

The baseline survival fcn

$$S_0(t) = \frac{1}{1 + (\lambda t)^\kappa},\ t \geqslant 0.$$

The sur. fcn w/ covariates z

$$S(t) = S_0\left(t\psi(t)\right) = \frac{1}{1 + \left(\lambda\psi(z)t\right)^\kappa} = \frac{1}{1 + \left(\lambda e^{\beta^T z} t\right)^\kappa}.$$

So T follows a log logistic dist. w/ scale $\lambda e^{\beta^T z}$ and shape κ. Therefore,

$$\begin{aligned}
\mu &= \int_0^\infty S(t)dt \\
&= \int_0^\infty \frac{1}{1 + \left(\lambda e^{\beta^T z} t\right)^\kappa} dt \\
&= \frac{\pi}{\lambda\kappa e^{\beta^T z} \sin\left(\frac{\pi}{\kappa}\right)},\ \text{for } \kappa > 1.
\end{aligned}$$

6.1 Proportional Hazard Model

The propotional hazard model is defined by

$$h(t) = \psi(z)h_0(t),\ t > 0.$$

$\psi(t) > 1$ increases hazard rate; $\psi(t) < 1$ decreases hazard rate. The popular choice of $\psi(z)$ is

$$\psi(z) = e^{\beta^T z}.$$

We can get other dist. representation. Or example,

$$H(t) = \int_0^t h(\tau)d\tau = \int_0^t \psi(z)h_0(\tau)d\tau = \psi(z)\int_0^t h_0(\tau)d\tau = \psi(z)H_0(t).$$

Table 6.1: Comparison

	Accelerated life model	Proportional Hazard Model
$S(t)$	$S_0\left(t\psi(z)\right)$	$\left(S_0(t)\right)^{\psi(z)}$
$f(t)$	$\psi(z)f\left(t\psi(z)\right)$	$f_0(t)\psi(z)\left(S_0(t)\right)^{\psi(z)-1}$
$h(t)$	$\psi(z)H_0\left(t\psi(z)\right)$	$\psi(z)h_0(t)$
$H(t)$	$H_0\left(t\psi(z)\right)$	$\psi(z)H_0(t)$

Example 6.4. Consider a Weibull baseline dist. in a propotional hazard model. Find the hazard and sur. fcn.

$$h_0(t) = \kappa\lambda^\kappa t^{\kappa-1},\ t \geqslant 0.$$

Hence

$$h(t) = \psi(z)h_0(t) = \psi(z)\kappa\lambda^\kappa t^{\kappa-1},\ t > 0,$$

and

$$S(t) = \left(S_0(t)\right)^{\psi(t)} = \left(e^{-(\lambda t)^\kappa}\right)^{\psi(z)},\ t > 0.$$

Thus,

$$T \sim \text{Weibull}\left(\lambda\psi(z)^{\frac{1}{\kappa}}, \kappa\right).$$

Chapter 7

Lifetime Data Analysis

Data

$$t_1, \dots, t_n \sim f(x; \theta),$$

where $f(x; \theta)$ is the population distribution. Ususally just observe $t_{(1)}, \dots, t_{(r)}$ with $r \ll n$. Goal: To make inference on θ, and the statistical inference involves

- Point estimation
- Interval estimation
- Hypothesis testing

7.1 Point estimation

A *point estimator* $\hat{\theta}$ is a **statistic** w.r.t some quantity, for example, the distribution mean.

A *statistic* is a function of data values that does not depend on any unknown parameter. For example, sample mean and sample variance. The statistic is a random variable when we think about all the values it could take on based on all the different samples we could collect. But once we collect a single sample, we calculate a specific value of the statistic.

7.1.1 Unbiased

$\hat{\theta}$ is an unbiased est. of θ if

$$E(\hat{\theta}) = \theta.$$

Example 7.1.

$$t_1, \dots, t_n \sim \exp\left(\frac{1}{\theta}\right).$$

Then

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

is an unbiaed est. of θ since $E(\hat{\theta}) = \theta$.

Definition 7.2.

$$\text{bias} = E(\hat{\theta}) - \theta.$$

If $\hat{\theta}$ is a biased of θ but

$$\lim_{n\to\infty} E(\hat{\theta}) = \theta,$$

then we call $\hat{\theta}$ as an *asymptotically unbiased* est. of θ since $\hat{\theta}$ can be regard a fcn of n.

To compare two unbiased est. $\hat{\theta}_1$ and $\hat{\theta}_2$, we define

$$E = \frac{\text{Var}(\hat{\theta}_1)}{\text{Var}(\hat{\theta}_2)}$$

as the efficiency of $\hat{\theta}_1$ related to $\hat{\theta}_2$. When $E < 1$, $\hat{\theta}_1$ is better than θ_2. When $E > 1$, $\hat{\theta}_1$ is worse than θ_2.

Example 7.3. Assume

$$t_1, t_2, t_3 \overset{iid}{\sim} \exp(1/\theta).$$

Consider

$$\hat{\theta}_1 = \frac{t_1 + t_2 + t_3}{3},$$
$$\hat{\theta}_2 = \frac{t_1 + 4t_2 + t_3}{6}.$$

Clearly, $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased of θ.

$$\text{Var}(\hat{\theta}_1) = \frac{\theta^2}{3},$$
$$\text{Var}(\hat{\theta}_2) = \frac{\theta^2 + 16\theta^2 + \theta^2}{36} = \frac{\theta^2}{2}.$$

Hence

$$E = \frac{\hat{\theta}_1}{\hat{\theta}_2} = \frac{1/3}{1/2} = \frac{2}{3} < 1.$$

Thus, $\hat{\theta}_1$ is better than $\hat{\theta_2}$.

Remark. Consider

$$\hat{\theta} = \frac{\alpha_1 t_1 + \alpha_2 t_2 + \alpha_3 t_3}{\alpha_1 + \alpha_2 + \alpha_3}, \alpha_1, \alpha_2, \alpha_3 \geqslant 0 \text{ and } \alpha_1 + \alpha_2 + \alpha_3 > 0,$$

then the best choice is

$$\alpha_1 = \alpha_2 = \alpha_3 = 1.$$

Theorem 7.4 (Cramer-Rao inequality)**.** *Let $T_1, \ldots, T_n$ be iid random lifetimes from a population w/ pdf $f(t)$, where domain of suppport does not depend on any unknown parameters. Assume that the cont. first-order exists. Let $\hat{\theta}$ be an unbiased est. of θ, then*

$$\mathit{Var}(\theta) \geqslant \frac{1}{n \cdot E\left[(\partial \log f(T)/\partial\theta)^2\right]} =: \textit{lower bound.}$$

Definition 7.5. $\hat{\theta}$ is called as a *minimum variance unbiased est.* (MVUE) of θ if

$$\mathrm{Var}(\theta) = \text{lower bound}.$$

Definition 7.6.

$$E\left[(\partial \log f(T)/\partial\theta)^2\right]$$

is called *fisher information* of the unknown quantity of population.

Example 7.7.

$$t_1, \ldots, t_n \overset{iid}{\sim} \exp\left(\frac{1}{\theta}\right).$$

Then

$$\hat{\theta} = \frac{1}{n}\sum_{n=1}^{n} t_n$$

is MVUE of θ.

Proof. Since

$$f(t) = \frac{1}{\theta}e^{-\frac{1}{\theta}t},\ t > 0.$$

$$\begin{aligned}(\partial \log f(T)/\partial\theta)^2 &= \left(\partial\left(-\log\theta - \frac{t}{\theta}\right)/\partial\theta\right)^2 \\ &= \left(-\frac{1}{\theta} + \frac{t}{\theta^2}\right)^2 \\ &= \frac{1}{\theta^2} - \frac{2t}{\theta^3} + \frac{t^2}{\theta^4}.\end{aligned}$$

Then

$$I(\theta) = E\left[\frac{1}{\theta^2} - \frac{2t}{\theta^3} + \frac{t^2}{\theta^4}\right] = \frac{1}{\theta^2} - \frac{2}{\theta^2} + \frac{2}{\theta^2} = \frac{1}{\theta^2}.$$

Hence

$$\text{lower bound} = \frac{1}{n\frac{1}{\theta^2}} = \frac{\theta^2}{n} = \mathrm{Var}(\hat{\theta}).$$

Thus, $\hat{\theta}$ is a MVUE of θ. □

Remark. Estimator has small bias or big variation. Trade off often needed. So we often use mean square error (MSE)

$$E(\hat{\theta} - \theta)^2 = \mathrm{Var}(\hat{\theta}) + \left(E(\hat{\theta} - \theta)\right)^2 = \mathrm{Var}(\hat{\theta}) + bias^2,$$

where $(\hat{\theta} - \theta)^2$ is the square loss of function.
(Refer to the decision theory for general loss fcn.)

7.1.2 Consistency

$\hat{\theta}$ is a consistent est. of θ if

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}-\theta\right| < \epsilon\right) = 1, \forall \epsilon > 0,$$

i.e.,

$$\hat{\theta} \xrightarrow{p} \theta.$$

For any $\epsilon > 0$, when n is sufficient large,

$$\hat{\theta} \in (\theta - \epsilon, \theta + \epsilon).$$

If we directly compute

$$P\left(\left|\hat{\theta}-\theta\right| < \epsilon\right),$$

we will use the joint dist. with n fold intervals (hyperplane). So we often show consistency through showing

$$\lim_{n\to\infty} \operatorname{Var}(\hat{\theta}) = 0,$$

since

$$0 = P\left(\left|\hat{\theta}-\theta\right| \geqslant \epsilon\right) \leqslant \frac{1}{\epsilon^2}\int \left|\hat{\theta}-\theta\right|^2 = \frac{\operatorname{Var}(\hat{\theta})}{\epsilon^2}.$$

Example 7.8 (Example 7.7 Continued)**.**

$$\theta = \frac{1}{n}\sum_{i=1}^{n} t_i,$$

is also a consistency est. of θ since

$$\operatorname{Var}(\hat{\theta}) = \frac{\theta^2}{n} \to 0.$$

7.2 Interval Estimation

An interval est. contains more info than a point est.

$$P(L \leqslant \theta \leqslant U) = 1 - \alpha,$$

where L and U are statistics and we call $[l, u]$ (exactly) $100(1-\alpha)\%$ confidence interval of θ. Similarly,

$$P(L \leqslant \theta \leqslant U) \geqslant 1 - \alpha$$

are also used w/ equ. holding for some θ.

Remark. (a) Intepretation of CI.

(b) Find a $(1-\alpha)100\%$ CI through using pivot, which is a r.v..

Example 7.9. Assume

$$t_1, \ldots, t_n \overset{i.i.d.}{\sim} \exp(1/\theta).$$

Find a CI for θ. Note

$$\frac{2}{\theta}\sum_{i=1}^{n} t_i \sim \chi^2(2n).$$

Then

$$P\left(\chi^2_{\frac{\alpha}{2},2n} \leqslant \frac{2}{\theta}\sum_{i=1}^{n} t_i \leqslant \chi^2_{1-\frac{\alpha}{2},2n}\right) = 1-\alpha,$$

or

$$P\left(\frac{2\sum_{i=1}^{n} t_i}{\chi^2_{1-\frac{\alpha}{2},2n}} \leqslant \theta \leqslant \frac{2\sum_{i=1}^{n} t_i}{\chi^2_{\frac{\alpha}{2},2n}}\right) = 1-\alpha.$$

Then L, U is a function of $\sum_{i=1}^{n} t_i$. So $(1-\alpha)\%$:

$$\left[\frac{2\sum_{i=1}^{n} t_i}{\chi^2_{1-\frac{\alpha}{2},2n}}, \frac{2\sum_{i=1}^{n} t_i}{\chi^2_{\frac{\alpha}{2},2n}}\right].$$

If $t \sim \exp(1/\theta)$,

$$f(t) = \frac{1}{\theta}e^{-\frac{1}{\theta}t} = \frac{t^{1-1}}{\theta^1\Gamma(1)}e^{-\frac{1}{\theta}t} \sim \text{Gamma}(1,\theta).$$

Hence

$$\sum_{i=1}^{n} t_i \sim \text{Gamma}(n,\theta).$$

The pdf of $\sum_{i=1}^{n} t_i$ is

$$f_1(t) = \frac{t^{n-1}}{\theta^n\Gamma(n)}e^{-\frac{1}{\theta}t}.$$

Then the pdf of $\frac{2}{\theta}\sum_{i=1}^{n} t_i$ is

$$f_2(t) = \frac{t^{n-1}}{2^n\Gamma(n)}e^{-\frac{1}{2}t}.$$

Then

$$\frac{2}{\theta}\sum_{i=1}^{n} t_i \sim \text{Gamma}(n,2),$$

which is $\chi^2(2n)$.

Remark. (a)

$$\sum_{i=1}^{n} t_i \sim \text{Erlang}(n, 1/\theta).$$

(b) An asympotically exact CI

$$\lim_{n\to\infty} P(L \leqslant \theta \leqslant U) = 1-\alpha.$$

(c) CI is NOT unique.

How to calculate the prob. that the $100(1-\alpha)\%$ CI contains one value θ_0?
Assume that (L, U) has joint pdf $f_{L,U}(l, u)$, then (!!!)

$$\begin{aligned} P(L \leqslant \theta_0 \leqslant U) &= \int\int_{L\leqslant\theta_0\leqslant U} f(l,u)dldu \\ &= \int_{-\infty}^{\theta_0}\int_{-\theta_0}^{\infty} f(l,u)dldu. \end{aligned}$$

Note that the above equality $= 1-\alpha$ when $\theta_0 = \theta$. In practise, we expect the above prob is as small as possible Also we can calculate the mean of interval width.

$$E[W] = E[U-L] = \int_{-\infty}^{\infty}\int_{l}^{\infty}(u-l)f(l,u)dldu.$$

The variance of the interval width

$$V(W) = \int_{-\infty}^{\infty}\left((u-l)^2 - (E[W])^2\right)f(l,u)dldu.$$

We expect that $E[W]$ is as small as possible. We can also use one-sided CI, similarly. Good CI satisfies

(a) Covergence of true value (close to $1-\alpha$.)

(b) Covergence of false value θ_0 as small as possible.

(c) Expectation of the interval width is small.

7.3 Likelihood Theory

Suppose

$$t_1, \ldots, t_n \overset{iid}{\sim} f(t; \underset{\sim}{\theta}),$$

with $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)^T$. The likelihood

$$L(\underset{\sim}{t}; \underset{\sim}{\theta}) = f(\underset{\sim}{t}; \underset{\sim}{\theta}) = \prod_{i=1}^{n} f(t_i; \underset{\sim}{\theta}),$$

which is a function of θ or the joint pdf. The MLE

$$\hat{\theta} = \arg\max_{\theta\in\Theta} L(\underset{\sim}{t}; \underset{\sim}{\theta}).$$

In practice, it is often to maximize the log likelihood.

$$l(\underset{\sim}{t}; \underset{\sim}{\theta}) = \log L(\underset{\sim}{t}; \underset{\sim}{\theta}) = \sum_{i=1}^{n} \log f(t_i; \underset{\sim}{\theta}).$$

Note the $l(\underset{\sim}{t};\theta)$ is asymptotically normal distributed by CLT. Since

$$\int_0^\infty \cdots \int_0^\infty L(\underset{\sim}{t};\underset{\sim}{\theta})d\underset{\sim}{t} = 1,$$

we have

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta}\int_0^\infty \cdots \int_0^\infty L(\underset{\sim}{t};\underset{\sim}{\theta})dt \\
&= \int_0^\infty \cdots \int_0^\infty \frac{\partial L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_i}dt \\
&= \int_0^\infty \cdots \int_0^\infty \frac{\partial \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_i}L(\underset{\sim}{t};\underset{\sim}{\theta})dt \\
&= E\left[\frac{\partial \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_i}\right] \\
&= E[u_i(\tilde{\theta})],
\end{aligned}$$

where

$$u_i(\underset{\sim}{\theta}) = \frac{\partial \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_i},\ i = 1,\ldots,p,$$

contains data t_i's and

$$\underset{\sim}{u}(\underset{\sim}{\theta}) = (u_1(\underset{\sim}{\theta}),\ldots,u_p(\underset{\sim}{\theta}),$$

is called as the score vector and

$$E[\underset{\sim}{u}(\underset{\sim}{\theta})] = \underset{\sim}{0}.$$

Remark. $\hat{\theta}_{\text{MLE}}$ can be obtained by setting

$$\underset{\sim}{u}(\underset{\sim}{\theta}) = 0.$$

Next,

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta_j}\int_0^\infty \cdots \int_0^\infty \frac{\partial \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_i}L(\underset{\sim}{t};\underset{\sim}{\theta})dt \\
&= \int_0^\infty \cdots \int_0^\infty \left(\frac{\partial^2 \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_i\partial\theta_j}L(\underset{\sim}{t};\underset{\sim}{\theta}) + \frac{\partial \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_i}\frac{\partial \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_j}L(\underset{\sim}{t};\underset{\sim}{\theta})\right)dt \\
&= E\left[\frac{\partial^2 \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_i\partial\theta_j}\right] + E[u_i(\underset{\sim}{\theta})u_j(\underset{\sim}{\theta})],\ i,j = 1,\ldots,p.
\end{aligned}$$

hence

$$E\left[-\frac{\partial^2 \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_i\partial\theta_j}\right] = E[u_i(\underset{\sim}{\theta})u_j(\underset{\sim}{\theta})] = \text{Cov}(u_i(\underset{\sim}{\theta}),u_j(\underset{\sim}{\theta})).$$

Thus,

$$E\left[-\frac{\partial^2 \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial \theta_i^2}\right] = E[u_i^2(\underset{\sim}{\theta})] = \text{Var}(u_i(\underset{\sim}{\theta}))$$
$$= E\left[\left(\frac{\partial \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial \theta_i}\right)^2\right].$$

Note that

$$I(\theta) = \left[E\left[-\frac{\partial^2 \log L(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial \theta_i \partial \theta_j}\right]\right]_{p\times p} = \left[\text{Cov}\left(u_i(\underset{\sim}{\theta}), u_j(\underset{\sim}{\theta})\right)\right]_{p\times p}.$$

Example 7.10. Suppose

$$t_1, \ldots, t_n \overset{iid}{\sim} \exp(1/\theta).$$

Find the score vector and Fisher information.

$$L(\underset{\sim}{t};\theta) = \prod_{i=1}^n \frac{1}{\theta} e^{-\frac{t_i}{\theta}} = \theta^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^n t_i}.$$

Then

$$\log L(\underset{\sim}{t};\theta) = -n\log\theta - \frac{1}{\theta}\sum_{i=1}^n t_i.$$

Hence

$$\mu(\theta) = \frac{d \log L(\underset{\sim}{t};\theta)}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n t_i.$$

Clearly,

$$E[u(\theta)] = 0.$$

Set $u(\theta) = 0$, we have

$$\hat{\theta}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^n t_i.$$

Note that

$$\frac{d^2 \log L(\underset{\sim}{t};\theta)}{d\theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^n t_i.$$

Then

$$\begin{aligned} I(\theta) &= E\left[-\frac{\partial \log L(\tilde{t};\theta)}{d\theta^2}\right] \\ &= -\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^n E[t_i] \\ &= -\frac{n}{\theta^2} + \frac{2n}{\theta^2} \\ &= \frac{n}{\theta^2}. \end{aligned}$$

Also,

$$\begin{aligned}\operatorname{Var}(u(\theta)) &= \frac{1}{\theta^4}\sum_{i=1}^{n}\operatorname{Var}(t_i)\\ &= \frac{1}{\theta^4}n\theta^2\\ &= \frac{n}{\theta^2}\\ &= I(\theta),\ \text{verified.}\end{aligned}$$

Moreover, assume $\theta = 1$, then

$$u(\theta) = -n + \sum_{i=1}^{n} t_i,$$

which has a shifted Erlang$(n, 1)$ distribution w/ pdf

$$f(u) = \frac{1}{(n-1)!}(u+n)^{n-1}e^{-(u+n)},\ u > -n.$$

Remark. (a) Observed information matrix

$$O(\hat{\theta}) = I(\hat{\theta}).$$

(b) If we are interested in $\phi = g(\theta)$, then the score of $\phi =$ the score of $\theta/g'(\theta)$. The Fisher information of $\phi=$ the Fisher info of $\theta/g'^2(\theta)$.

(c) The MLE has the invariant property,

$$\hat{\phi} = g(\phi_{\text{MLE}}).$$

Example 7.11. Suppose

$$t_1, \ldots, t_n \overset{iid}{\sim} \text{Inverse Gauss}(\lambda, \mu),$$

w/ pdf

$$f(t) = \sqrt{\frac{\lambda}{2\pi}}t^{-\frac{3}{2}}e^{-\frac{\lambda(t-\mu)^2}{2\mu^2 t}},\ t > 0.$$

The likelihood fcn is

$$\begin{aligned}L(\underset{\sim}{t}; \lambda, \mu) &= \prod_{i=1}^{n} f(t_i)\\ &= \lambda^{\frac{n}{2}}(2\pi)^{-\frac{n}{2}}\prod_{i=1}^{n} t_i^{-\frac{3}{2}}e^{-\frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(t_i-\mu)^2}{t_i}}.\end{aligned}$$

The log likelihood fcn is

$$\begin{aligned}\log L(\underset{\sim}{t}; \lambda, \mu) &= \frac{n}{2}\log\lambda - \frac{n}{2}\log 2\pi + \log\prod_{i=1}^{n} t_i^{-\frac{3}{2}} - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(t_i-\mu)^2}{t_i}\\ &= \frac{n}{2}\log\lambda - \frac{n}{2}\log 2\pi + \log\prod_{i=1}^{n} t_i^{-\frac{3}{2}} - \lambda\sum_{i=1}^{n}\left(\frac{t_i}{2\mu^2} - \frac{1}{\mu} + \frac{1}{2t_i}\right).\end{aligned}$$

Then

$$u_\lambda(\lambda,\mu) = \frac{\partial \log(\underset{\sim}{t},\lambda,\mu)}{\partial \lambda} = \frac{n}{2\lambda} - \frac{1}{2\mu^2}\sum_{i=1}^{n}\frac{(t_i-\mu)^2}{t_i},$$

$$u_\mu(\lambda,\mu) = -\lambda\sum_{i=1}^{n}\left(-\frac{t_i}{\mu^3}+\frac{1}{\mu^2}\right) = \frac{\lambda}{\mu^3}\left(\sum_{i=1}^{n} t_i - n\mu\right).$$

Seting $\underset{\sim}{u}(\lambda,\mu) = (u_\lambda(\lambda,\mu), u_\mu(\lambda,\mu)) = 0$, we have

$$\hat{\mu}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} t_i,$$

$$\hat{\lambda} = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{t_i} - \frac{n}{\sum_{i=1}^{n} t_i}\right]^{-1}.$$

Next, we compute $I(\lambda,\theta)$.

$$\frac{\partial^2 \log L}{\partial \lambda^2} = -\frac{n}{2\lambda^2},$$

$$\frac{\partial^2 \log L}{\partial \lambda \partial \mu} = \frac{1}{\mu^3}\sum_{i=1}^{n} t_i - \frac{n}{\mu^2},$$

$$\frac{\partial^2 \log L}{\partial \mu^2} = -\frac{3\lambda\sum_{i=1}^{n} t_i}{\mu^4} + \frac{2n\lambda}{\mu^3}.$$

Note $E[t_i] = \mu$, so

$$I(\lambda,\mu) = \begin{pmatrix} E\left[-\frac{\partial^2 \log L}{\partial \lambda^2}\right] & E\left[-\frac{\partial^2 \log L}{\partial \lambda \partial \mu}\right] \\ E\left[-\frac{\partial^2 \log L}{\partial \lambda \partial \mu}\right] & E\left[-\frac{\partial^2 \log L}{\partial \mu^2}\right]. \end{pmatrix} = \begin{pmatrix} \frac{n}{2\lambda} & 0 \\ 0 & \frac{n\lambda}{\mu^3} \end{pmatrix}.$$

Proposition 7.12 (Asymptotically properties)**.**

$$\begin{aligned} u_i(\underset{\sim}{\theta}) &= \frac{\partial \log L}{\partial \theta_i} \\ &= \frac{\partial}{\partial \theta_i}\sum_{k=1}^{n} f(t_k;\underset{\sim}{\theta}) \\ &= \sum_{i=1}^{n}\frac{\partial}{\partial \theta_i}\log f(t_k;\underset{\sim}{\theta}), \end{aligned}$$

which is a sum of iid r.v.'s. So the score vector $u(\underset{\sim}{\theta})$ is asymtotically normal with mean 0 and covariance matrix $I(\theta)$. Therefore,

$$\underset{\sim}{u}^T(\underset{\sim}{\theta})I^{-1}(\underset{\sim}{\theta})\underset{\sim}{u}(\underset{\sim}{\theta}) \overset{n\to\infty}{\longrightarrow} \chi_p^2,$$

which can be used to determine CI and perform hypothesis testing on θ.

Chapter 8

Parametric Estimate for Models Without Covariants

8.1 Censoring

Often, it is impossible or impractise to observe lifetime of all items in a test. A censored observation occurs when only a bounded is known on the time of failure.

Definition 8.1. Complete data: data w/o sensoring or all failure items are observed. Right censoring: One or more items have only a lower bound known on the lifetime.

Example 8.2. (a) Life testing, 12 machines are put into service on Jan 1. Suppose 7 of them have failed by Dec 31. This is the case of 7 failures and 5 right-censored observations.

(b) Medical study, patients can still be alive or the researchers can lost contact w/ them at the end of the study.

8.1.1 Right censoring

(a) Type II censoring / order statistic censoring.

Definition 8.3. Stop the test when the number of failure of items reaches the assigned value r. The likelihood is the jpdf of $X_{(1)}, \ldots, X_{(r)}$.

(b) Type I censoring (time censoring)

Definition 8.4. Terminate the test at a particular time.

(c) Random censoring

Definition 8.5. Each individual item are withdrawn from the test at any time during the study, often assume that the i-th lifetime t_i, and the i-th censoring time C_i are independent random variables.

8.1.2 Left censoring

Less frequently than right censoring. For example, forget to record.

8.1.3 Interval censoring

Lifetime falls into a interval due to checking items periodically, e.g. once a week, i.e.,

$$[L, R]$$

where L, R can be random variable.

How to handle problems of censoring data?

(a) Ignoring censored obs. (Bad)

(b) Wait for all right-censored obs to fail. (Bad)

(c) Analysis on data directly. Let

$$T_1, \ldots, T_n$$

be indep (real) lifetimes (may not be observed). Let

$$C_1, \ldots, C_n$$

be indep. right censoring time. Define

$$U = \{1 \leqslant i \leqslant n \mid t_i \leqslant C_i\},$$

which is uncensored obs,

$$C = \{1 \leqslant i \leqslant n \mid t_i > C_i\},$$

which is censored obs..

Analysis based on info from both U and C.

Example 8.6.

$$U = \{1, 2, 4\},$$

$$C = \{3, 5\}.$$

The usual form for lifetime data is given by the pairs

$$(x_i, \delta_i)$$

w/

$$x_i = \min(t_i, C_i) \text{ (observed)}$$

and

$$\delta_i = \begin{cases} 1, & x_i = t_i \text{ item failed/uncensored,} \\ 0, & x_i = C_i \text{ observed not fail/censored.} \end{cases}$$

The likelihood function is

$$\begin{aligned} L(\underset{\sim}{x}, \underset{\sim}{\theta}) &= \prod_{i=1}^{n} f^{\delta_i}(x_i; \underset{\sim}{\theta}) S^{1-\delta_i}(x_i; \underset{\sim}{\theta}) \\ &= \prod_{i \in U} f(t_i; \underset{\sim}{\theta}) \cdot \prod_{i \in C} S(C_i; \underset{\sim}{\theta}) \\ &= \prod_{i=1}^{n} h^{\delta_i}(x_i; \underset{\sim}{\theta}) S(x_i; \theta). \end{aligned}$$

Hence

$$\begin{aligned} \log L(\underset{\sim}{\theta}) &= \sum_{i=1}^{n} \log h^{\delta_i}(x_i, \underset{\sim}{\theta}) - \sum_{i=1}^{n} H(x_i; \underset{\sim}{\theta}) \\ &= \sum_{i \in U} \log h(x_i, \underset{\sim}{\theta}) - \sum_{i=1}^{n} H(x_i, \underset{\sim}{\theta}). \end{aligned}$$

8.1.4 Progressive censoring

Remark. Progressive Type II right-censoring. Unde the censoring scheme, n units are placed on life test at time zero. Immediately following the first failure, R_1 surviving items are to be removed from the test at random. Immediately following the first failure, R_2 surviving items are to be removed from the test at random. The process continues until at the time of m-th observed failure, the remaining $R_m = n - R_1 - \cdots - R_{m-1} - m$ units are all to be removed from the test ($R_0 = 0$). If

$$R_1 = \cdots = R_{m-1} = 0,$$

then

$$R_m = n - m,$$

which implies it reduces to the Type II right-censoring If

$$R_1 = \cdots = R_m = 0,$$

then we have complete data.

Use EM algorithm if S hard to find.

8.2 Exponential Distribution

Quite popular due to its tractability for parametric est. and inference.

$$\begin{aligned} f(t; \lambda) &= \lambda e^{-\lambda t}, \\ S(t; \lambda) &= e^{-\lambda t}, \\ h(t; \lambda) &= \lambda, \\ H(t; \lambda) &= \lambda t, \end{aligned}$$

where λ is the failure rate and its mean is $1/\lambda$.

8.2.1 Complete data

$$T_1, \ldots, T_n \overset{iid}{\sim} f(t; \lambda).$$

$$L(\lambda) = \prod_{i=1}^{n} f(t_i; \lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}.$$

$$l(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} t_i.$$

Score

$$u(\lambda) = dl/d\lambda = n/\lambda - \sum_{i=1}^{n} t_i.$$

Setting $u(\lambda) = 0$,

$$\hat{\lambda}_{MLE} = \frac{1}{\bar{t}} = \frac{n}{\sum_{i=1}^{n} t_i}.$$

Remark. (a) Note that $\sum_{i=1}^{n} t_i$ is often referred to the total time on the test, which is a complete sufficient statistic.

(b) Fisher info matrix

$$i(\lambda) = E\left[-\frac{\partial^2 \log L(\lambda)}{\partial \lambda^2}\right] = E\left[\frac{n}{\lambda^2}\right] = \frac{n}{\lambda^2},$$

and thus the observed Fisher info. matrix (also asympotically normally distributed)

$$O(\hat{\lambda}) = \frac{n}{\hat{\lambda}^2_{\text{MLE}}} = \frac{\left(\sum_{i=1}^{n} t_i\right)^2}{n}.$$

To get a CT on λ, we use the sampling distribution of $\sum_{i=1}^{n} t_i$,

$$2\lambda \sum_{i=1}^{n} t_i = \frac{2n\lambda}{\hat{\lambda}_{\text{MLE}}} \sim \chi^2(2n).$$

So w/ prob. $1 - \alpha$,

$$\chi^2_{(1-\alpha/2, 2n)} < \frac{2n\lambda}{\hat{\lambda}_{\text{MLE}}} < \chi^2_{(\alpha/2, 2n)}.$$

Thus, $100(1 - \alpha)\%$ CI

$$\frac{\hat{\lambda}_{\text{MLE}} \chi^2_{(1-\alpha/2, 2n)}}{2n} < \lambda < \frac{\hat{\lambda}_{\text{MLE}} \chi^2_{(\alpha/2, 2n)}}{2n}.$$

The p-value for testing

$$H_0 : \lambda = \lambda_0 \text{ v.s. } H_1 : \lambda \neq \lambda_0,$$

is

Remark. (a) MLE of μ

$$\hat{\mu}_{\text{MLE}} = \frac{1}{\hat{\lambda}_{\text{MLE}}} = \bar{t}.$$

(b) $100(1-\alpha)\%$ CI (replace λ and λ_{MLE} simultaneously.)

$$\frac{2n\hat{\mu}_{\text{MLE}}}{\chi^2_{(\alpha/2,2n)}} < \mu < \frac{2n\hat{\mu}_{\text{MLE}}}{\chi^2_{(1-\alpha/2,2n)}}.$$

(c) For any fix time t, the MLE of $S(t) = e^{-\lambda t}$

$$\hat{S}(t) = e^{-\hat{\lambda} t},\ t > 0.$$

Its interval estimation

$$e^{-ut} < S(t) < e^{-lt},$$

where (l, u) is a $100(1-\alpha)\%$ CI for λ.

Now we consider Bayesian inference for λ.

$$f(\underset{\sim}{t}|\lambda) = \lambda^n e^{-\left(\sum_{i=1}^n t_i\right)\lambda}.$$

Prior distribution:

$$\lambda \sim \text{Gamma}(\alpha, \beta).$$

Then

$$\pi(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}.$$

So the posterior distribution

$$\begin{aligned}\pi(\lambda|\underset{\sim}{t}) &\propto f(\underset{\sim}{t}|\lambda)\pi(\lambda)\\ &\propto \lambda^{n+\alpha-1} e^{-\left(\sum_{i=1}^n t_i+\beta\right)\lambda}.\end{aligned}$$

Thus,

$$\lambda|\underset{\sim}{t} \sim \text{Gamma}\left(n+\alpha, \sum_{i=1}^n t_i + \beta\right).$$

Hence conjugate prior. Then the posterior mean of λ

$$\hat{\lambda}_{\text{pos}} = E[\lambda|\underset{\sim}{t}] = \frac{n+\alpha}{\sum_{i=1}^n t_i + \beta}.$$

The posterior mean of μ

$$\begin{aligned}\hat{\mu}_{\text{pos}} &= E[\mu|\underset{\sim}{t}]\\ &= E[1/\lambda|\underset{\sim}{t}]\\ &= \int_0^\infty f_{\Lambda|\underset{\sim}{t}}(\lambda|\underset{\sim}{t})d\lambda\\ &= \int_0^\infty \frac{1}{\lambda}\frac{\left(\sum_{i=1}^n t_i+\beta\right)^{n+\alpha}\lambda^{n+\alpha-1}e^{-\left(\sum_{i=1}^n t_i+\beta\right)\lambda}}{\Gamma(n+\alpha)}\\ &= \frac{\Gamma(n+\alpha-1)}{\Gamma(n+\alpha)}\cdot\left(\sum_{i=1}^n t_i+\beta\right)\\ &= \frac{\sum_{i=1}^n t_i+\beta}{n+\alpha-1}.\end{aligned}$$

So it is not invariant. How to choose α and β?

(a) historical record:

$$\alpha/\beta = \bar{x}.$$

(b) default choice:

$$\pi(\lambda) \propto I(\lambda)^{1/2}.$$

e.g.

$$I(\lambda) \propto \lambda^{-1}.$$

(c) empirical metod

$$\underset{\sim}{\theta} = (\alpha, \beta).$$

$$\hat{\theta} = \arg\max L(\alpha, \beta | \underset{\sim}{t}).$$

$$\left.\begin{array}{c} f(\underset{\sim}{t}|\lambda), \\ \pi(\lambda|\alpha,\beta) \end{array}\right\} \Rightarrow f(\underset{\sim}{t}|\alpha,\beta) = L(\alpha,\beta|\underset{\sim}{t}).$$

(d) Consider prior on α, β.

8.2.2 Type II censored data

Complete data is a special case of type II-right censored data w/ $r = n$. As before,

$$T_1, \dots, T_n \overset{iid}{\sim} f(t, \lambda).$$

$$C_1 = \dots = C_n = t_{(r)},$$

which is the censoring times and

$$x_i = \min\{t_i, C_i\}.$$

The likelihood function

$$\begin{aligned} L(\lambda) &= \prod_{i \in U} f(x_i; \lambda) \cdot \prod_{i \in C} S(x_i; \lambda) \\ &= f(t_{(1)}, \dots, t_{(r)} | \lambda) \end{aligned}$$

and thus the log likelihood function

$$\begin{aligned} \log L(\lambda) &= \sum_{i \in U} \log h(x_i; \lambda) - \sum_{i=1}^{n} H(x_i; \lambda) \\ &= r \cdot \log \lambda - \lambda \sum_{i=1}^{n} x_i. \end{aligned}$$

The total time on test

$$\sum_{i=1}^{n} x_i = \sum_{i \in U} t_i + \sum_{i \in C} C_i = \sum_{i=1}^{r} t_{(i)} + (n-r) t_{(r)},$$

where

$$t_{(1)} \leqslant \cdots \leqslant t_{(r)}$$

are r order statistic of observed failures. The score function

$$u(\lambda) = \frac{d \log L(\lambda)}{d\lambda} = \frac{r}{\lambda} - \sum_{i=1}^{n} x_i.$$

Setting $u(\lambda) = 0$,

$$\hat{\lambda}_{\text{MLE}} = \frac{r}{\sum_{i=1}^{n} x_i}.$$

Note that

$$\begin{aligned} I(\lambda) &= E\left[-\frac{d^2 \log L(\lambda)}{d\lambda}\right] \\ &= E\left[\frac{r}{\lambda^2}\right] = \frac{r}{\lambda^2}. \end{aligned}$$

To get an exact CI for λ, we need the dist. of $\sum_{i=1}^{n} X_i$.

$$2\lambda \sum_{i=1}^{n} X_i = \frac{2r\lambda}{\hat{\lambda}_{\text{MLE}}} \sim \chi^2(2r),$$

where $2\lambda \sum_{i=1}^{n} X_i$ is the space statistic exp. So a $100(1-\alpha)\%$ CI for λ is

$$\frac{\hat{\lambda}}{2r} \chi^2_{(2r, 1-\alpha/2)} \leqslant \lambda \leqslant \frac{\hat{\lambda}}{2r} \chi^2_{(2r, \alpha/2)}.$$

Hypo. test (p-value)

$$H_0 : \lambda = \lambda_0 \ vs \ H_1 : \lambda \neq \lambda_0.$$

Bayesian inference on λ is quite similar to that of complete data.

8.2.3 Type I censored data

A life test of n items is terminated at time c. Then

$$C_1 = \cdots = C_n = c.$$

The total time on test is

$$\begin{aligned} \sum_{i=1}^{n} x_i &= \sum_{i \in U} t_i + \sum_{i \in C} c \\ &= \sum_{i=1}^{r} t_{(i)} + (n-r)c, \end{aligned}$$

where r is a random variable and

$$c \in [t_{(r)}, t_{(r+1)}].$$

$$I(\lambda) \approx \frac{r}{\lambda^2}.$$

The log likelihood function

$$\begin{aligned}\log L(\lambda) &= \sum_{i\in U} \log h(x_i;\theta) - \sum_{i=1}^{n} H(x_i;\theta) \\ &= \sum_{i\in U} \log\lambda - \lambda\sum_{i=1}^{n} x_i \\ &= r\log\lambda - \lambda\sum_{i=1}^{n} x_i.\end{aligned}$$

Let

$$d\frac{\log L(\lambda)}{d\lambda} = \frac{r}{\lambda} - \sum_{i=1}^{n} x_i = 0,$$

we have

$$\hat{\lambda}_{\text{MLE}^*} = \frac{r}{\sum_{i=1}^{n} x_i},$$

!!!!!

where r is a random variable (maximum index r such that $t_{(r)} \leqslant c$.) For the same value r, type I censoring has a large total time on test than that of type II because of the gap

$$c - t_{(r)},$$

and so

$$(n-r)c \geqslant (n-r)t_{(r)}.$$

Remark. The sample dist. of $\sum_{i=1}^{n} x_i$ is NOT tractable. So we can only get CI using approximating method

$$2\lambda\sum_{i=1}^{n} x_i \mathrel{\dot\sim} \chi^2_{2r+1},$$

based on the fact that if $c = t_{(r)}$, then

$$2\lambda\sum_{i=1}^{n} x_i \sim \chi^2_{(2r)};$$

if $c = t_{(r+1)}$, then

$$2\lambda\sum_{i=1}^{n} x_i \sim \chi^2_{(2r+2)}.$$

8.2.4 Random censored data sets

The total time on test

$$\sum_{i=1}^{n} x_i = \sum_{i \in U} t_i + \sum_{i \in C} c_i.$$

The sampling distribution of $\sum_{i=1}^{n} x_i$ becomes more complicated in this case. We will use some approximating to get a CI on λ.

I : Based on an approx to a result from type II censoring case

$$2\lambda \sum_{i=1}^{n} x_i \dot{\sim} \chi^2(2r).$$

II : Based on the likelihood ratio test

$$2\left[\log L(\hat{\lambda}) - \log L(\lambda)\right] \dot{\sim} \chi^2(1).$$

III : Based on MLE

$$\hat{\lambda}_{\text{MLE}} \sim N\left(\lambda, I^{-1}(\lambda)\right),$$

where $I^{-1}(\lambda)$ can be replaced by $I^{-1}(\hat{\lambda}_{\text{MLE}})$.

8.2.5 Comparing two exponential distribution

Let

$$x = \min(t, c).$$

We have two iid sets

$$x_{11}, x_{12}, \ldots, x_{1n} \sim \exp(\lambda_1);$$

$$x_{21}, x_{22}, \ldots, x_{2n} \sim \exp(\lambda_2).$$

Assume type II censoring w/ $r_1 > 0$ failures observed in the first sample and w/ $r_2 > 0$ failures observed in the second sample. We want to construct CI for

$$\frac{\lambda_1}{\lambda_2},$$

which similar to construct CI for

$$\frac{\sigma_1^2}{\sigma_2^2}$$

in the normally distributed case. Note

$$2\lambda_1 \sum_{i=1}^{n} x_i \sim \chi^2(2r_1),$$

$$2\lambda_2 \sum_{i=1}^{n} x_i \sim \chi^2(2r_2),$$

where we have the two test statistics are indep.

If a pivot is a statistic, then it is known as an ancillary statistic. Pivotal quantities are fundamental to the construction of test statistics, as they allow the statistic to not depend on parameters-for example, Student's t-statistic is for a normal distribution with unknown variance (and mean). They also provide one method of constructing confidence intervals, and the use of pivotal quantities improves performance of the bootstrap. In the form of ancillary statistics, they can be used to construct frequentist prediction intervals (predictive confidence intervals).

Then

$$\frac{2\lambda_1 \sum_{i=1}^n x_i/2r_1}{2\lambda_2 \sum_{i=1}^n x_i/2r_2} = \frac{\lambda_1}{\lambda_2}\frac{\hat{\lambda_2}}{\hat{\lambda_1}} \sim F(2r_1, 2r_2).$$

So w/ probability $1-\alpha$,

$$f_{2r_1,2r_2,1-\alpha/2} \leqslant \frac{\lambda_1}{\lambda_2}\frac{\hat{\lambda_2}}{\hat{\lambda_1}} \leqslant f_{2r_1,2r_2,\alpha/2}.$$

w/ CI, we can easily draw the conclusion for testing

$$H_0 : \lambda_1 = \lambda_2 \ vs \ H_1 : \lambda_1 \neq \lambda_2,$$

e.g. we check whether 1 is in CI or not!

8.2.6 Prediction

Consider

$$f(x) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}, \ x > 0.$$

Let

$$t_{(r+1)} \leqslant \cdots \leqslant t_{(n-s)},$$

where

$$1 \leqslant r+1 \leqslant n-s \leqslant n-1.$$

be a doubly type II censored sample. The MLE of θ does not have a closed form unless $r=0$. The BLUE of θ is given by

$$\hat{\theta} = \frac{1}{k}\sum_{i=r+1}^{n-s} a_i t_{(i)},$$

where

$$a_i = \begin{cases} \frac{\sum_{l=n-r}^n \frac{1}{l}}{\sum_{l=n-r}^n \frac{1}{l^2}} - (n-r-1), & i = r+1 \\ 1, & i = r+2, \dots, n-s-1 \\ s+1, & i = n-s \end{cases}$$

and

$$K = (n-r-s-1) + \frac{\left(\sum_{l=n-r}^n \frac{1}{l}\right)^2}{\sum_{l=n-r}^n \frac{1}{l^2}}.$$

Sketch:

$$\Theta = \left\{ \hat{\theta} \,\middle|\, \hat{\theta} = \sum_{i=r+1}^{n-s} b_i t_{(i)}, \ E[\hat{\theta}] = \theta \right\}.$$

Then

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \mathrm{Var}(\theta).$$

Let It is well-known that normalized spacing statistics assuming $t_{(0)} = 0$,

$$S_i = (n-i+1)\left(T_{(i)} - T_{(i-1)}\right), \forall i \in [n]$$

are iid exp r.v.'s. Let us denote

$$S_i^* = (n-i+1)\left(t_{(i)} - t_{(i-1)}\right)/\theta = \frac{S_i}{\theta} \sim \exp(1).$$

Then the BLUE of θ can be rewritten as a linear function of S_i^* since there is a 1-1 relationship between the ordering statistics and spacing. Suppose we are interested in predicting the lth order statistic with $n-s < l \leqslant n$.
It is natural to consider the following pivot

$$Z_1 = \frac{t_{(l)} - t_{(n-s)}}{\hat{\theta}},$$

where $\hat{\theta}$ is from the BLUE.
It is clearly that we need to find the exact value t such that

$$P\left(Z_1 = \frac{t_{(l)} - t_{(n-s)}}{\hat{\theta}} > z\right) = \alpha.$$

In order to construct the exact PI for $t_{(l)}$. Note that the above prob. can be rewritten as

$$P\left(\sum_{i=1}^{l} c_i S_i^* > 0\right),$$

where

$$c_i = \begin{cases} -\frac{t}{K(n-i+1)}\sum_{j=r+1}^{n-s} a_j, & i = 1,\ldots,r+1 \\ -\frac{t}{n-i+1}\sum_{j=i}^{n-s} a_j, & j = r+2,\ldots,n-s \\ \frac{1}{n-i+1}, & i = n-s+1,\ldots,l. \end{cases}$$

Using the algorithm in Huffer and Lin (2001), we can find the exact z satisfing the above equation. More precisely, given α, we can find z_1, z_2 such that

$$P(Z_1 > z_1) = \frac{\alpha}{2},\ P(Z_1 > z_2) = 1 - \frac{\alpha}{2}.$$

Thus an exact $100(1-\alpha)\%$ PI for $t_{(l)}$ is

$$(t_{(n-s)} + z_2\hat{\theta}, t_{(n-s)} + z_1\hat{\theta}).$$

8.3 Two parameter exponential distribution

If

$$T \sim \exp(\mu, \theta),$$

then

$$f(t;\mu,\theta) = \frac{1}{\theta}e^{-\frac{t-u}{\theta}},\ t \geqslant u.$$

$$F(t;\mu,\theta) = 1 - e^{-\frac{t-u}{\theta}},\ t \geqslant u.$$

8.3.1 Complete data set

$$\begin{aligned} L(\mu,\theta) &= \prod_{i=1}^{n} f(t_i;\mu,\theta) \\ &= \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n}(t_i-\mu)} \mathbb{1}_{\{t_i \geqslant \mu\}} \\ &= \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n}(t_i-\mu)} \mathbb{1}_{\{t_{(1)} \geqslant \mu\}}. \end{aligned}$$

Hence

$$T(\underset{\sim}{t}) = (\bar{t}, t_{(1)})$$

is a sufficient statistic for

$$(\theta,\mu),$$

and

$$\hat{\mu}_{\text{MLE}} = t_{(1)}.$$

Then

$$L(\hat{\mu}_{\text{MLE}},\theta) = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n}(t_i-\hat{\mu}_{\text{MLE}})}.$$

Setting

$$\frac{dL(\hat{\mu}_{\text{MLE}},\theta)}{d\theta} = 0,$$

we have

$$\hat{\theta}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(t_i - \hat{\mu}_{\text{MLE}}) = \bar{t} - t_{(1)}.$$

Note

$$\begin{aligned} n\left(\bar{t} - t_{(1)}\right) &= \sum_{i=1}^{n} t_i - n t_{(1)} \\ &= \sum_{i=1}^{n} t_{(i)} - n t_{(1)} \\ &= \sum_{i=1}^{n} \left(t_{(i)} - t_{(1)}\right) \\ &= \sum_{i=2}^{n} \left(t_{(i)} - t_{(1)}\right) \sim \text{Gamma?} \end{aligned}$$

and

$$t_{(1)} \sim \exp\left(\mu, \frac{\theta}{n}\right),$$

$$t_{(i)} - t_{(i-1)} \sim \exp\left(\mu, \frac{\theta}{n-i+1}\right),$$

and

$$t_{(i)} - t_{(1)} = \sum_{k=2}^{i} \left(t_{(k)} - t_{(k-1)} \right).$$

Next, note

$$\hat{\mu}_{\text{MLE}} = t_{(1)} \sim \exp\left(\mu, \frac{\theta}{n}\right),$$

and

$$\hat{\theta}_{\text{MLE}} \perp\!\!\!\perp \hat{\mu}_{\text{MLE}}???$$

Then

$$E[\hat{\mu}_{\text{MLE}}] = \mu + \frac{\theta}{n},$$

$$\text{Var}(\hat{\mu}_{\text{MLE}}) = \frac{\theta^2}{n^2},$$

and

$$E[\hat{\theta}_{\text{MLE}}] = (\mu + \theta) - \left(\mu + \frac{\theta}{n}\right) = \theta\left(1 - \frac{1}{n}\right).$$

Since

$$\begin{aligned} \text{Var}(\bar{t}) &= \frac{\text{Var}(t_1)}{n} \\ &= \frac{\theta^2}{n}. \end{aligned}$$

Since

$$\hat{\theta}_{\text{MLE}} \perp\!\!\!\perp \hat{\mu}_{\text{MLE}},$$

we have

$$\begin{aligned} \frac{\theta^2}{n} = \text{Var}(\bar{t}) &= \text{Var}(\bar{t} - t_{(1)} + t_{(1)}) \\ &= \text{Var}\left(\bar{t} - t_{(1)}\right) + \text{Var}\left(t_{(1)}\right) \\ &= \text{Var}\left(\bar{t} - t_{(1)}\right) + \frac{\theta^2}{n^2}. \end{aligned}$$

Thus,

$$\text{Var}\left(\bar{t} - t_{(1)}\right) = \theta^2\left(\frac{1}{n} - \frac{1}{n^2}\right),$$

or

$$\text{Var}\left(\bar{t} - t_{(1)}\right) = \theta^2\left(\frac{1}{n} + \frac{1}{n^2} - \frac{2}{n^3}\right).$$

Remark. Note that when μ is known,

$$\hat{\theta}_{\text{MLE}} = \bar{t} - \mu.$$

When θ is known,

$$\hat{\theta}_{\text{MLE}} = t_{(1)}.$$

Theorem 8.7. *Since*

$$T(\underset{\sim}{t}) = (\bar{t}, t_{(1)})$$

is a sufficient statistic for (θ, μ)*, we have*

$$T_2(\underset{\sim}{t}) = (\hat{\theta}_{MLE}, \hat{\mu})$$

is a sufficient satistic for (θ, μ).

Theorem 8.8.

$$T_2(\underset{\sim}{t}) = (\hat{\theta}_{MLE}, \hat{\mu})$$

is a complete statistic for (θ, μ).

Proof. Use definition and take derivative for a vector. □

Example 8.9.

$$\frac{T_1}{\sum_{i=1}^n T_i} \sim \text{Beta}(1, n-1),$$

Proof.

$$T_1 \sim \exp(\theta).$$

$$\sum_{i=1}^n T_i \sim \text{Gamma}\,(n, \theta)\,.$$

□

Example 8.10. Assume we are interested in the reliability function $R(x_0)$, the probability that lifetime exceed a value t_0,

$$R(t_0) = e^{-\frac{t_0 - \mu}{\theta}}.$$

Assume

$$\mu = 0.$$

The MLE of $R(t_0)$ is

$$\hat{R}(t_0) = e^{-\frac{t_0}{\hat{\theta}}} = e^{-\frac{nt_0}{\sum_{i=1}^n t_i}}.$$

It is biased, but a MVUE can be obtained using Blackwell-rao theorem. Note

$$W = \begin{cases} 1, & T_1 > t_0 \\ 0, & T_1 \leqslant t_0 \end{cases}$$

is unbiased of $R(t_0)$ b/c

$$E[W] = P(T_1 > t_0) = R(t_0).$$

Since $\sum_{i=1}^n T_i$ is a sufficient of θ (if both arguments are unknown, use both), a MVUE of $R(t_0)$ is

$$E\left[W \middle| \sum_{i=1}^n T_i\right] = E\left[\mathbb{1}_{\{T_1 > t_0\}} \middle| \sum_{i=1}^n T_i\right] = P\left(T_1 > t_0 \middle| \sum_{i=1}^n T_i\right),$$

is a UMVUE of $R(t_0)$. Method1: Compute the conditional pdf. Method 2: Note

$$\frac{T_1}{\sum_{i=1}^n T_i} \sim \text{Beta}(1, n-1),$$

which is from the scale family and so it is an ancillary statistic and thus, by Basu's theorem, it is independent of the complete and minimal sufficient statistic

$$\sum_{i=1}^{n} T_i.$$

Therefore,

$$\begin{aligned}
E\left[W \middle| \sum_{i=1}^{n} T_i\right] &= P\left(T_1 > t_0 \middle| \sum_{i=1}^{n} T_i\right) \\
&= P\left(\frac{T_1}{\sum_{i=1}^{n} T_i} > \frac{t_0}{\sum_{i=1}^{n} T_i} \middle| \sum_{i=1}^{n} T_i\right) \\
&= \int_{\min\left\{\frac{t_0}{\sum_{i=1}^{n} T_i},1\right\}}^{1} (n-1)(1-z)^{n-2} dz \\
&= \begin{cases} \left(1 - \frac{t_0}{\sum_{i=1}^{n} T_i}\right)^{n-1}, & t_0 \leqslant \sum_{i=1}^{n} T_i \\ 0, & t_0 > \sum_{i=1}^{n} T_i \end{cases}
\end{aligned}$$

is the MVUE of $R(t_0)$, Rargh (1963).

8.3.2 Type II right censored data

$$T_{(1)} \leqslant \cdots \leqslant T_{(r)}.$$

The joint pdf (integrating on the pdf of n order statistics.)

$$\begin{aligned}
L(\mu, \theta) &= f\left(t_{(1)}, \ldots, t_{(r)}; \mu, \theta\right) \\
&= \frac{n!}{(n-r)!} f\left(t_{(1)}; \mu, \theta\right) \cdots f\left(t_{(r)}; \mu, \theta\right) \left[1 - F\left(t_{(r)}; \mu, \theta\right)\right]^{n-r} \\
&\propto \prod_{i=1}^{r} \frac{1}{\theta} e^{-\frac{t_{(i)}-\mu}{\theta}} \left[e^{-\frac{t_{(r)}-\mu}{\theta}}\right]^{n-r} \\
&= \frac{1}{\theta^r} e^{-\frac{1}{\theta}\left[\left(\sum_{i=1}^{r} t_{(i)}\right) - n\mu + (n-r)t_{(r)}\right]}.
\end{aligned}$$

So we get

$$\hat{\mu}_{\text{MLE}} = t_{(1)},$$

$$\hat{\theta}_{\text{MLE}} = \frac{1}{r}\left[\sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)} - n\hat{\mu}_{\text{MLE}}\right].$$

8.3.3 Double censored type II data

$$T_{(r+1)} \leqslant \cdots \leqslant T_{(n-s)}.$$

MLEs do not have explicit expressions unless

$$r = 0.$$

The BLUEs of μ and θ are

$$\hat{\mu} = \sum_{i=r+1}^{n-s} a_i t_{(i)},$$

$$\hat{\theta} = \sum_{i=r+1}^{n-s} b_i t_{(i)},$$

where

$$a_i = \begin{cases} 1 + \frac{n-r-1}{n-r-s} \sum_{l=n-r}^{n} \frac{1}{l}, & i = r+1 \\ \frac{1}{n-r-s-1} \sum_{l=n-r}^{n} \frac{1}{l}, & i = r+2, \ldots, n-s-1 \\ -\frac{s+1}{n-r-s-1} \sum_{l=n-s}^{n} \frac{1}{l}, & i = n-s \end{cases}$$

and

$$b_i = \begin{cases} -\frac{n-r-1}{n-r-s-1}, & i = r+1 \\ \frac{1}{n-r-s-1}, & i = r+2, \ldots, n-s-1 \\ \frac{s+1}{n-r-s-1}, & i = n-s \end{cases}$$

Predicting the lth order statistic with $n - s < l \leqslant n$, it is natural to consider the pivot

$$z = \frac{t_{(l)} - t_{(n-s)}}{\hat{\theta}},$$

where $\hat{\theta}$ is from the BLUE. We need to find the exact value z_0 such that

$$P\left(z = \frac{t_{(l)} - t_{(n-s)}}{\hat{\theta}} > z_0\right) = \alpha.$$

Similarly, the above prob. can be rewritten as

$$P\left(\sum_{i=1}^{n} d_i S_i^* > 0\right)$$

w/

$$d_i = \begin{cases} 0, & i = 1, \ldots, r+1, \\ -\frac{z_0}{n-r-s-1}, & i = r+2, \ldots, n-s, \\ \frac{1}{n-i+1}, & i = n-s+1, \ldots, n. \end{cases}$$

Using the algorithm in Huffer and Lin (2001), we can find z_0 satisfying the above equation. So we may determine z_1, z_2 such that

$$P(Z_1 > z_1) = \frac{\alpha_1}{2},\ P(Z_1 > z_2) = 1 - \frac{\alpha}{2}.$$

So the exact $(1-\alpha)\%$ PI for $t_{(l)}$ is

$$\left(t_{(n-s)} + z_2\hat{\theta},\ t_{(n-s)} + z_1\hat{\theta}\right).$$

8.3.4 Weibull distribution

Example 8.11. X has a Weibull distribution if it has the pdf

$$f(x;\mu,\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{x-\mu}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-\mu}{\alpha}\right)^{\beta}},\ x > \mu,$$

cdf

$$F(x;\mu,\alpha,\beta) = 1 - e^{-\left(\frac{x-\mu}{\alpha}\right)^{\beta}},\ x > \mu,$$

hazard function

$$h(x;\mu,\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{x-\mu}{\alpha}\right)^{\beta-1},\ x > \mu,$$

which is increasing if $\beta > 1$ and decreasing if $0 < \beta < 1$.

Remark. (a) Exponential distribution is a special case of Weibull distribution w/ $\beta = 1$.

(b) If $X \sim$ Weibull(μ,α,β), then

$$Y = \left(\frac{X-\mu}{\alpha}\right)^{\beta} \sim \exp(1).$$

Then to generate a Weibull(μ,α,β),

1) Generate

$$U \sim U(0,1).$$

2) Then

$$Y = -\log(1-U) \sim \exp(1).$$

3)

$$X = \mu + \alpha Y^{1/\beta} \stackrel{d}{=} \mu + \alpha(-\log U)^{1/\beta}.$$

Or we can get the inverse function $\mu + \alpha(-\log(1-U))^{1/\beta}$ directly.

(c) If $\beta > 1$, the Weibull(μ,α,β) density goes to 0 as $x \to \mu$ and there is a single mode (max) at

$$x = \alpha\left(\frac{\beta-1}{\beta}\right)^{1\beta} + \mu.$$

If $0 < \beta < 1$, the mode is at μ.

(d) The median of the distribution is

$$\alpha(\log 2)^{1/\beta} + \mu.$$

The most usual situation is μ known and we use $\mu = 0$, and w/ hazard and cumulative hazard function

$$h(x;\lambda,\kappa) = \kappa\lambda(\lambda t)^{\kappa-1},\ t > 0.$$

$$H(x;\lambda,\kappa) = (\lambda t)^{\kappa},\ t > 0.$$

Suppose that $T_1, \ldots, T_n$ are failure times and $C_1, \ldots, C_n$ are censoring time.

$$X_i = \min(T_i, C_i), \forall i = 1, \ldots, n.$$

When there are observed r failures,

$$\begin{aligned}
\log(\lambda, \kappa) &= \sum_{i \in U} \log h(x_i; \lambda, \kappa) - \sum_{i=1}^{n} H(x_i; \lambda, \kappa) \\
&= \sum_{i \in U} (\log \kappa + \kappa \log \lambda + (\kappa - 1) \log x_i) - \sum_{i=1}^{n} (\lambda x_i)^\kappa \\
&= r \log \kappa + r\kappa \log \lambda + (\kappa - 1) \sum_{i \in U} \log x_i - \lambda^k \sum_{i=1}^{n} x_i^\kappa.
\end{aligned}$$

So the score vector has two elements

$$u_1(\lambda, \kappa) = \frac{\partial \log L}{\partial \lambda} = \frac{\kappa r}{\lambda} - \kappa \lambda^{\kappa - 1} \sum_{i=1}^{n} x_i^\kappa,$$

$$u_2(\lambda, \kappa) = \frac{\partial \log L}{\partial \kappa} = \frac{r}{\kappa} + r \log \lambda + \sum_{i \in U} \log x_i - \sum_{i=1}^{n} (\lambda x_i)^\kappa \log(\lambda x_i).$$

Setting $u_1(\lambda, \kappa) = 0$, we have

$$\lambda = \left(\frac{r}{\sum_{i=1}^{n} x_i^\kappa} \right)^{1/\kappa}.$$

Put it into $u_2(\lambda, \kappa) = 0$, we get a equation that must be solved numerically. One may consider Newton-Raphson procedure:

$$k_{i+1} = k_i - \frac{g(k_i)}{g'(k_i)}.$$

One can also consider the multivariate version of the Newton-Raphson procedure using the desired Fisher information matrix.(Appendix D).

8.3.5 Bayesian Inference

One paper by Kundu (2008), "Bayesian Inference and Life Testing Plan for the Weibull Distribution in Presence of Progressive Censoring".
Consider a Weibull distribution w/ pdf

$$f(t; \lambda, \kappa) = \lambda \kappa t^{\kappa - 1} e^{-(\lambda t)^\kappa}, \ t > 0.$$

Assume there are n units on a test and we consider the progressive scheme and we observe data

$$\{(t_1, R_1), \ldots, (t_m, R_m)\}$$

and

$$R_1 + R_m + m = n,$$

where m and $R_1, \ldots, R_m$ are pre-set. The likelihood function of the observed sample $\{(t_1, R_1), \ldots, (t_m, R_m)\}$ is

$$L(\lambda, \kappa) \propto \kappa^m \lambda^m \prod_{i=1}^{m} t_i^{\kappa-1} \cdot e^{-\lambda \sum_{i=1}^{m} (R_i+1) t_i^{\kappa}}.$$

Case I: the shape parameter κ is known. Consider the conjugate

$$\lambda | a, b \sim \text{Gamma}(a, b),$$

w/ pdf

$$\pi(\lambda | a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}, \ \lambda > 0.$$

The posterior distribution of λ is

$$\lambda | \text{data} \sim \text{Gamma}\left(a + m, b + \sum_{i=1}^{m} (R_i + 1) t_i^{\kappa}\right).$$

Therefore, the Bayesian estimator of λ is

$$\hat{\lambda} = E[\lambda | \text{data}] = \frac{a + m}{b + \sum_{i=1}^{m} (R_i + 1) t_i^{\kappa}},$$

and credible interval can also be obtained from Gamma distribution.

Remark. (a) Loss function

1) d_2 metric

$$L(\hat{\lambda}, \lambda) = (\hat{\lambda} - \lambda)^2.$$

So the Bayesian estimate is the posterior mean $E[\lambda | \text{data}]$.

2) d_1 metric

$$L(\hat{\lambda}, \lambda) = \left| \hat{\lambda} - \lambda \right|.$$

So the Bayesian estimate is the posterior median.

3) Discrete metric

$$L(\hat{\lambda}, \lambda) = \mathbb{1}_{\{\hat{\lambda} \neq \lambda\}}.$$

So the Bayesian estimate is the posterior mode. Refer to Berger J. 1980, "Statistical decision theory".

Chapter 9

Parametric Estimate For Models with Covariants

Data set

$$\underset{\sim}{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \delta = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_n \end{pmatrix},$$

$$Z = \begin{pmatrix} z_{11} & \cdots & z_{1q} \\ \vdots & \cdots & \vdots \\ z_{n1} & \cdots & z_{nq} \end{pmatrix} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

Then we use the symbol

$$S(\underset{\sim}{t}, \underset{\sim}{z}; \underset{\sim}{\theta}, \underset{\sim}{\beta}),$$
$$F(\underset{\sim}{t}, \underset{\sim}{z}; \underset{\sim}{\theta}, \underset{\sim}{\beta}),$$
$$h(\underset{\sim}{t}, \underset{\sim}{z}; \underset{\sim}{\theta}, \underset{\sim}{\beta}),$$
$$H(\underset{\sim}{t}, \underset{\sim}{z}; \underset{\sim}{\theta}, \underset{\sim}{\beta}),$$

where

$$\theta = (\theta_1, \ldots, \theta_p)'$$

consisting of unknown parameter w/ the baseline distribution. The likelihood function can be written as

$$L(\underset{\sim}{\theta}, \underset{\sim}{\beta}) = \prod_{i \in U} f(x_i, \underset{\sim}{z_i}; \underset{\sim}{\theta}, \underset{\sim}{\beta}) \cdot \prod_{i \in C} S(x_i, \underset{\sim}{z_i}; \underset{\sim}{\theta}, \underset{\sim}{\beta}).$$

Then the log-likelihood function becomes

$$\begin{aligned} \log L(\underset{\sim}{\theta}, \underset{\sim}{\beta}) &= \sum_{i \in U} \log f(x_i, \underset{\sim}{z_i}; \underset{\sim}{\theta}, \underset{\sim}{\beta}) + \sum_{i \in C} \log S(x_i, \underset{\sim}{z_i}; \underset{\sim}{\theta}, \underset{\sim}{\beta}) \\ &= \sum_{i \in U} \log h(x_i, \underset{\sim}{z_i}; \underset{\sim}{\theta}, \underset{\sim}{\beta}) - \sum_{i=1}^{n} H(x_i, \underset{\sim}{z_i}; \underset{\sim}{\theta}, \underset{\sim}{\beta}). \end{aligned}$$

Then we can obtain the MLE of $\underset{\sim}{\theta}$ and $\underset{\sim}{\beta}$ by maximizing the above function.

9.1 Accelerated life model

$$S(\underset{\sim}{t}, \underset{\sim}{z}) = S_0(t\psi(\underset{\sim}{z})),\ t \geqslant 0,$$

w/

$$\psi(\underset{\sim}{z}) = e^{\underset{\sim}{\beta}' \underset{\sim}{z}}.$$

(a) The exponential baseline function

$$h_0(t) = \lambda.$$

The hazard function associated w/ covariates $\underset{\sim}{z}$ is

$$h(t, \underset{\sim}{z}, \underset{\sim}{\theta}, \underset{\sim}{\beta}) = \psi(\underset{\sim}{z}) h_0(t\psi(z)) = \lambda e^{\underset{\sim}{\beta}' \underset{\sim}{z}}.$$

Often, rewrite

$$\lambda = e^{\beta_0 z_0}$$

w/ $z_0 = 1$. Then

$$h(t, \underset{\sim}{z}, \underset{\sim}{\beta}) = e^{\underset{\sim}{\beta}' \underset{\sim}{z}},$$

where

$$\underset{\sim}{\beta} = (\beta_0, \beta_1, \ldots, \beta_q)',$$

$$\underset{\sim}{z} = (z_0, z_1, \ldots, z_q)'.$$

By integration, the corresponding cummulative hazard fun

$$H(\underset{\sim}{t}, \underset{\sim}{z}, \underset{\sim}{\beta}) = t e^{\underset{\sim}{\beta}' \underset{\sim}{z}}.$$

The log-likelihood function is

$$\begin{aligned} \log L(\beta) &= \sum_{i \in U} \log h(x_i, \underset{\sim}{z_i}; \underset{\sim}{\beta}) - \sum_{i=1}^{n} H(x_i, \underset{\sim}{z_i}; \underset{\sim}{\beta}) \\ &= \sum_{i \in U} \beta' z_i - \sum_{i=1}^{n} x_i e^{\beta' z_i}. \end{aligned}$$

Note

$$\frac{\partial \log L(\beta)}{\partial \beta_j} = \sum_{i \in U} z_{ij} - \sum_{i=1}^{n} x_i z_{ij} e^{\beta' z_i}, \forall j = 0, \ldots, n.$$

More details on accelated life models can be seen in Collett (2003).

(b) Weibull distribution.

9.2 Proportional Hazard Models

$$h(t, \underset{\sim}{z}; \underset{\sim}{\theta}, \underset{\sim}{\beta}) = \psi(\underset{\sim}{z}) h_0(t) = e^{\underset{\sim}{\beta}' \underset{\sim}{z}} h_0(t).$$

$$H(t, \underset{\sim}{z}; \underset{\sim}{\theta}, \underset{\sim}{\beta}) = \psi(\underset{\sim}{z}) H_0(t) = e^{\underset{\sim}{\beta}' \underset{\sim}{z}} H_0(t).$$

Then

$$\begin{aligned} \log L(\underset{\sim}{\theta}, \underset{\sim}{\beta}) &= \sum_{i \in U} \log h(x_i, \underset{\sim}{z_i}, \underset{\sim}{\theta}, \underset{\sim}{\beta}) - \sum_{i=1}^{n} H(x_i, \underset{\sim}{z_i}, \underset{\sim}{\theta}, \underset{\sim}{\beta}) \\ &= \sum_{i \in U} \left[\log h_0(x_i) + \underset{\sim}{\beta}' \underset{\sim}{z_i} \right] - \sum_{i=1}^{n} H_0(x_i) e^{\underset{\sim}{\beta}' \underset{\sim}{z_i}}. \end{aligned}$$

9.3 Assesising Model Adequacy

Goodness of fit test for testing

(a) χ^2-test, the fit of a discrete distribution to a data set. It can be adapted for use on a continuous distribution. Assume that the range of T can be partitioned into k nonoverlapping subintervals

$$[0, a_1), [a_1, a_2), \ldots, [a_{k-1}, a_k)$$

w/

$$a_0 = 0, a_k = \infty.$$

So

$$p_i = P(a_{i-1} \leqslant T < a_i) = \int_{a_{i-1}}^{a_i} f_0(t) dt, \forall i = 1, \ldots, k.$$

Consider

$$\begin{aligned} &H_0 : p_1 = p_{10},\ p_2 = p_{20},\ \cdots, p_k = p_{k0} \\ &H_1 not H_0, \end{aligned}$$

where $p_{10}, \ldots, p_{k0}$ are precribed probability that sum to 1. Fisher proposed the following

Theorem 9.1. *Let $(X_1, \ldots, X_k)$ be a multinormal r.v. w/ parameters*

$$n, p_1, \ldots, p_k$$

w/

$$\sum_{i=1}^{k} p_i = 1.$$

Then

$$Q = \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{np_i} \sim \chi^2(k-1).$$

The general form of χ^2-goodness of fit statistics is

$$Q = \sum \frac{(observed - expected)^2}{expected}.$$

(b) Kolmogorov-Smirnov test Let $F_0(t)$ be a hypothesis or fitted CDF.

$$H_0 : F(t) = F_0(t),$$

$$H_1 : F(t) \neq F_0(t).$$

For the complete data set, the test statistic is

$$D_n = \sup_t \left| \hat{F}(t) - F_0(t) \right|,$$

where $\hat{F}(t)$ is the empirical CDF from the data set. For the right-censored data set

$$D_{n,r} = \sup_{0 \leqslant t \leqslant t_{(r)}} \left| \hat{F}(t) - F_0(t) \right|$$

and the distribution of $D_{n,r}$ depends only on n and r under H_0. If there are some unknown parameters involved under H_0, we will estimate each unknown parameters and then plug it in.

9.4 Competing Risks

Definition 9.2 (Formal definition)**.** One observes the pair (T, C), where $T > 0$ is the failure time and

$$C \in \{1, \ldots, k\}$$

represents the type of failure. One intuitive way of describing a competing riks situation w/ k riks is to assume that each risk is associated with a failure time $T_j, j = 1, \ldots, k$. Thes k times are thought as latent failure times. When all risks are presenting, the time to failure of the system is the smallest of these failure times along with the actual cause. Thus,

$$T = \min\{T_1, T_2, \ldots, T_k\},$$

and if $T = T_c$, then

$$C = c.$$

Remark. Traditionally, competing risks were analyzed as if they were independent of each other. But we can consider dependent risks.

9.4.1 Model Specification

The joint distribution of the pair (T, C) from an individual is completely specified by the sub-distribution functions

$$F_j(t) = P(T \leqslant t, C = j) \text{ (joint)}.$$

Let

$$f_j(t) = F_j'(t).$$

The marginal distribution of T is

$$F(t) = P(T \leqslant t) = \sum_{j=1}^{k} F_j(t).$$

Or the survival function

$$S(t) = 1 - F(t).$$

The sub-survival function becomes

$$S_j(t) = P(T > t, C = j).$$

The marginal distribution of C is

$$\pi_j = P(C = j) = F_j(\infty).$$

Note

$$F_j(t) + S_j(t) = \pi_j.$$

The distribution of (T, C) can be specified by the sub-hazard function

$$\begin{aligned}
\lambda_j(t) &= \lim_{\Delta t \to 0} \frac{P(T \leqslant t + \Delta t, C = j | T \geqslant t)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{P(t \leqslant T \leqslant t + \Delta t, C = j)}{P(T \geqslant t)} \frac{1}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{S_j(t) - S_j(t + \Delta t)}{S(t)} \frac{1}{\Delta t} \\
&= -\frac{1}{S(t)} \lim_{\Delta t \to 0} \frac{S_j(t + \Delta t) - S_j(t)}{\Delta t} \\
&= -\frac{1}{S(t)} S_j'(t) \\
&= \frac{f_j(t)}{S(t)}.
\end{aligned}$$

Hence

$$H_j(t) = \int_0^t \lambda_j(u) du,$$

$$F_j(t) = \int_0^t \lambda_j(u) S(u) du.$$

Also, the hazard function of T is

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\sum_{j=1}^k f_j(t)}{S(t)} = \sum_{j=1}^k \frac{f_j(t)}{S(t)} = \sum_{j=1}^k \lambda_j(t).$$

Then

$$H(t) = \sum_{j=1}^k H_j(t).$$

The survival function of T is

$$S(t) = e^{-H(t)} = e^{-\sum_{j=1}^k H_j(t)} = \prod_{j=1}^k G_j^*(t),$$

where

$$G_j^* = e^{-H_j(t)}, \forall j = 1, \ldots, k.$$

Note $G_j^*(t)$ is a survival function, but it is NOT in general the distribution of any observable random variables. However, $G_j^*(t)$ is the survival function of T_j under the model with independent latent failure time.

Remark. $\lambda_j(t)$ can be interpreted as the failure rate from a specific cause, i.e., crude hazard rate.

(a) Latent failure time representation.
The joint survival function of $T_1, \ldots, T_n$ is

$$S(t_1, \ldots, t_k) = P(T_1 > t_1, \ldots, T_k > t_k).$$

Thus the survival function of T is

$$S(t) = P(T > t) = S(t, \ldots, t).$$

Also,

$$f_j(t) = -\left.\frac{\partial S(t_1, \ldots, t_k)}{\partial t_j}\right|_{t_1 = \cdots = t_k = t},$$

$$\lambda_j(t) = \frac{f_j(t)}{S(t)} = -\left.\frac{\partial \log S(t_1, \ldots, t_k)}{\partial t_j}\right|_{t_1 = \cdots = t_k = t}.$$

Typically, we are interested in how to get the joint distribution or marginal distribution of the latent failure times $T, \ldots, T_k$ from (T, C).

(b) The identification. For $N(u, \sigma^2)$, $N(\alpha + \beta, \sigma^2)$ is not identificable. In general, the joint distribution or marginal distribution of T_i can NOT be determined from the distribution of (T, C). See Tsiatis (1975). So assumptions on $T_1, \ldots, T_n$ are needed. How to deal with the identifiability problem.

(1) Assuming independent risks. Assume that the latent failure times $T_1, \ldots, T_k$ are independent. Tsiatis (1975) showed that the marginal distribution of T_j can now be computed from the subdistribution function $F_j(t)$. In practise, this means that the marginal distributions can be estimated in a consistent manner from the competing risks data. Also,

$$S(t_1, \ldots, t_k) = \prod_{i=1}^{n} S_i(t_i).$$

So

$$\overline{G_j}^*(t) = S_j(t),$$

which is the marginal survival function of T_j.

(2) Assuming a known copula, which is a multivariate probability distribution for which the marginal probability distribution of each variable is uniform, for the latent failure times.
Zheng and Klein (1995) generalized the above result, proving that the marginal distributions are identifiable when the independence is given by a known copula.

Consider $k = 2$. Let K be joint distribution function of (T_1, T_1) and G_1, G_2 be the marginals. Then the copula of (T_1, T_2) is defined by

$$C(u_1, u_2) = K\left(G_1^{-1}(u), G_2^{-1}(u)\right),$$
$$(u_1, u_2) \in [0,1] \times [0,1].$$

For the independent case, Zheng and Klein (1995) proved that if the copula $C(\cdot, \cdot)$ is known, then the marginal distribution is also known. Then G_1, G_2 is uniquely determined by the sub-distributions F_1 and F_2.

Remark. The above discussion are in the non-parametric sense. If a parameter model is specified for the latent failure times, the identifiability problem is different since it now has to do with identification of a finite set of parameters.

(c) Modeling competing risks.

(1) Modeling sub-distribution.

i. Mixture model. Specify a sub-distribution function of the form

$$F_j(t) = \pi_j Q_j(t)$$

for a given (parametric) distribution function $Q_j(t)$, e.g.,

$$Q_j(t) = 1 - \exp\left\{-\left(\frac{t}{Q_j}\right)^{\alpha_j}\right\}$$

ii. Modling the sub-hazard function

$$h_j(t, \alpha_j, Q_j) = \frac{\alpha_j}{Q_j}\left(\frac{t}{Q_j}\right)^{\alpha_j - 1}.$$

iii. Regression models.
*Proportional hazards.

$$h_j(\underset{\sim}{t}, \underset{\sim}{x}) = \psi_j(\underset{\sim}{x}) h_{0j}(t),$$

where h_{0j} is the baseline hazard function, e.g.,

$$\psi_j(x) = \exp\{\underset{\sim}{\beta}' \underset{\sim}{x}\}.$$

**Accelarated life model

$$F_j(\underset{\sim}{t}, \underset{\sim}{x}) = F_{0j}(\phi_j(\underset{\sim}{x}) t),$$

where F_{0j} is the baseline sub-distribution function corresponding to $\phi_j(\cdot) = 1$.

(2) Modeling latent varaibles. Often through specification of the joint distribution function of $T_1, \ldots, T_k$, e.g., $k = 2$ Gumbel (1960) considers the bivariate exponential distribution

$$S(t_1, t_2) = \exp\{-\lambda_1 t_1 - \lambda_2 t_2 - \nu t_1 t_2\},$$
$$h_j(t) = \lambda_j + \nu t_j, \forall j = 1, 2.$$

Also, a class of models, called frailty models, are obtained by assuming $T_1, \dots, T_k$ are conditionally independent, given a random "frailty" Z, e.g.,

$$S(t_1, \dots, t_k) = \int_0^\infty \exp\left\{-Z \sum_{j=1}^k H_j(t_j)\right\} dG_Z(z),$$

where G_Z is the cdf of Z.

Let $X_1, \dots, X_n$ denote a random sample (no censoring) of size $n > 2$ from a population with the probability density function given by

$$f(x|\lambda) = \lambda x^{\lambda-1}, 0 < x < 1,$$

where

$$\lambda > 0.$$

Then the MLE

$$\hat{\lambda}_M = -\frac{n}{\sum_{i=1}^n \log x_i}.$$

Since $\prod_{i=1}^n X_i$ is a complete and sufficient statistic from the exponential family. Let

$$Y_i = -\log X_i.$$

Then the pdf of Y_i is

$$g(y) = \lambda e^{-\lambda y}, \ y > 0.$$

Hence

$$Z = -\sum_{i=1}^n \log X_i = \sum_{i=1}^n Y_i \sim \text{Gamma}(n, \lambda).$$

Then

$$E[\hat{\lambda}_M] = nE[1/Z] = \text{constant} \cdot \lambda.$$

When we do hypothesis testing, using the pivot

$$\lambda Z \sim \text{Gamma}(n, 1).$$
