SURVIVAL ANALYSIS

AD

1. Some History of Survival Analysis

The term 'survival analysis' has been used for data involving time to a certain event such as death, onset of a disease or relapse of a condition. The development of survival analysis dates back to the 17th century with the first life table ever produced by John Graunt in 1662 (Bills of Mortality).

Throughout the centuries, survival analysis was solely linked to the investigation of mortality rates; however, in the last few decades, applications of the statistical methods for survival data analysis have been extended beyond biomedical research to other fields such as criminology, sociology, marketing, institutional research and health insurance practice.

Survival analysis plays an important role when analysing data on events observed over time, such as death, cardiac arrest, relapse of drug addiction or failure of an electronic device. Besides identifying the significant risk factors, the survival model ranks these hazards by their importance in predicting survival durations. This information is essential to surgeons, psychologists and manufacturers to address these risk factors optimally.

The contributions of Kaplan and Meier in 1958 in estimating survival probabilities and hazard rates led to ground-breaking improvements in survival analysis. The Nelson-Aalen estimator is an alternative non- parametric estimator of the cumulative hazard rate. The proportional hazard model proposed by Cox in 1972 was another significant contribution to survival analysis. This semi- parametric model consists of two parts. The first component is the baseline hazard, which is a function of time and describes how risk varies over time. The second component is an exponential function of a linear combination of the predictors and is independent of time.

In essence, the Cox model can be used to compare the relative forces of mortality of two lives or two homogeneous groups of lives. These non-parametric and semi- parametric survival models assume that the members in a population are similar and so are inappropriate in the presence of unobserved diversity.

Further developments in survival analysis include the shared and unshared frailty models introduced by Vaupel in 1979 and extended by Hougaard in 1984. These models are more appropriate in accommodating heterogeneity and random effects as they eliminate biases in estimation.

2. introduction

Survival time data measures the time to a certain event, such as failure, death, response, relapse, the development of a given disease, parole, or divorce. These times are subject to random variations, and like any random variables, form a distribution. The distribution of survival times is usually described or characterised by three functions:

(1) The Survivorship Function
(2) The Probability Density Function
(3) The Hazard Function

These three functions are mathematically equivalent - if one of them is given, the other two can be derived.

3. Survivorship Function(or Survival Function)

If T is a random variable denoting the survival time then it's Survival Function denoted by $S(t)$, is defined as the probability that an individual survives longer than t:

$$\begin{aligned} S(t) &= P(\text{an individual survives longer than t}) \\ &= P(T > t) \end{aligned}$$

From the definition of the cumulative distribution function $F(t)$ of T,

$$\begin{aligned} S(t) &= 1 - P(\text{an individual fails before t}) \\ &= 1 - F(t) \end{aligned}$$

3.1. Properties of $S(t)$.

(1) $S(0) = 1$
(2) $S(\infty) = 0$
(3) $S(t)$ is everywhere monotone non-increasing function.
(4) $S(t)$ is left continuous.

3.2. NOTE:.

(1) $S(t)$ is also called as cumulative survival rate.
(2) The graph of $S(t)$ is called the survival curve. A steep survival curve represents low survival rate or short survival time. A gradual or flat survival curve represents high survival rate or longer survival. The survival survival curve is used to find percentiles of survival time and to compare survival distributions of two or more groups.
(3) The survival function is estimated as the proportion of patients surviving longer than t:

$$\hat{S}(t) = \frac{\text{Number of patients surviving longer than t}}{\text{Total number of patients}} \tag{1}$$

4. Probability Density Function(or Density Function)

The probability density function of the survival time T is defined as the limit of the probability that an individual fails in the short interval t to $t + \Delta t$ per unit width Δt, or simply the probability of failure in a small interval per unit time. It can be expressed as:

$$f(t) = \frac{\underset{\Delta t \to 0}{lim}\, P\,[\text{an individual dying in the interval } (t, t+\Delta t)]}{\Delta t} \tag{2}$$

4.1. **Properties of $f(t)$.**

(1) $f(t)$ is a non-negative function such that

$$\begin{aligned} f(t) &\geq 0 \quad \text{for all } t \geq 0 \\ &= 0 \quad \text{for } t < 0 \end{aligned}$$

(2) The area between the density curve and the t-axis is equal to 1.

4.2. **Notes:**

(1) The graph of $f(t)$ is called the density curve.

(2) $f(t)$ is estimated as the proportion of patients dying in an interval per unit width.

$$\hat{f}(t) = \frac{\text{number of patients dying in the interval beginning at time t}}{(totalnumberofpatients) * (intervalwidth)} \tag{3}$$

(3) The proportion of individuals that fail in any time interval and the peaks of high frequency of failure can be found from the density function.

(4) The density function is also known as unconditional failure rate.

5. Hazard Function

The hazard function $h(t)$ of survival time T gives the conditional failure rate. This is defined as the probability of failure during a very small time interval assuming that the individual has survived to the beginning of the interval, or as the limit of the probability that an individual fails in a very short interval, $t+\Delta t$, given that the individual has survived to time t.

$$h(t) = \frac{\underset{\Delta t \to 0}{lim}\, P \begin{bmatrix} \text{an individual fails in the time interval } (t, t+\Delta t) \\ \text{given the individual has survived to t} \end{bmatrix}}{\Delta t} \tag{4}$$

The hazard function can also be defined in terms of the cumulative distribution function $F(t)$ and the probability density function $f(t)$:

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{5}$$

5.1. **Notes:**

(1) The hazard function is also known as the instantaneous failure rate, force of mortality, conditional mortality rate, and age-specific failure rate.

(2) If t in equation 4 is age, it is a measure of the proneness to failure as a function of the age of the individual in the sense that the quantity $\Delta t * h(t)$ is the expected proportion of age t individuals who will fail in the short time interval $t + \Delta t$. The hazard function thus gives the risk of failure per unit time during the aging process.

(3) The hazard function is estimated as the proportion of patients dying in an interval per unit time, given that they have survived to the beginning of the interval:

$$\hat{h}(t) = \frac{\text{number of patients dying in the interval beginning at time t}}{\text{(number of patients surviving at t)*(interval width)}}$$

$$= \frac{\text{number of patients dying per unit time in the interval}}{\text{number of patients surviving at t}} \tag{6}$$

(4) The hazard function may increase, decrease, remain constant, or indicate a more complicated process.
  (a) Patients with acute leukaemia who do not respond to treatment have an increasing hazard rate.
  (b) Soldiers wounded by bullets who undergo surgery have a decreasing hazard rate.
  (c) Healthy persons between 18 and 40 years of age whose main risks of death are accidents have a constant hazard rate.
  (d) Patients with tuberculosis have risks that increase initially, then decrease after treatment.

(5) The cumulative hazard function is defined as:

$$H(t) = \int_0^t h(x)dx \tag{7}$$

5.2. **Why study hazard rate of lifetime distributions.** The hazard function is a particularly important characteristic of a lifetime distribution. It indicates the way the risk of failure varies with age or time, and this is of interest in most applications. Prior information about the shape of the hazard function can help guide model selection. Finally, if factors affecting an individual's lifetime vary over time, it is often essential to approach modelling through the hazard function.

If individuals in a population are followed right from actual birth to death, a bathtub-shaped hazard function is often seen. We are familiar with this pattern in human populations: after an initial period in which deaths result primarily from birth defects or infant diseases, the death rate drops and is relatively constant until the age of 30 or so, after which it increases with age. This pattern also manifests itself in other biological populations and in populations of manufactured items, some of which contain defects. Distributions with increasing hazard functions are seen for individuals for whom some kind of aging or wear-out takes place. Also, populations that display a bathtub-shaped hazard function are sometimes purged of weak individuals ; leaving a reduced population with an increasing hazard function. For example, manufactures use inspection or burn-in process. in which items are subjected to a brief period of operation before being sent to customers. In this way defective or poor-quality items that would fail early are removed from the population; this frequently leaves a residual population that exhibits an increasing hazard function.

Certain types of electronic devices display a decreasing hazard as items with defects fall and are removed from the population. Roughly constant hazard functions tend to occur in stable settings where failure or death is due to random phenomena such as shocks or accidents, which are external to the individual. Sometimes a hazard rate first increases to a maximum and then decreases, is encountered in many applications, for example, in the case of survival after treatment for cancer, where some individuals are cured, and in connection with the duration of marriage.

6. Relationships of the Survival Functions

6.1. **Expressing $S(t)$ in terms of $f(t)$ and $h(t)$.** : We know that,

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= 1 - \int_{x=0}^{t} f(x)dx \\ &= \int_{x=t}^{\infty} f(x)dx \end{aligned} \tag{8}$$

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= -\frac{S'(t)}{S(t)} \left[\text{Since,} S(t) = 1 - F(t) \text{ and } F'(t) = f(t)\right] \\ &= -\frac{d}{dt} log(S(t)) \end{aligned} \tag{9}$$

Integrating both sides w.r.t. "t" over the ranges 0 to x, we have,

$$\begin{aligned} \textstyle\int_{t=0}^{x} h(t)dt &= -\int_{t=0}^{x} \frac{d}{dt} log(S(t))dt \\ \Rightarrow \quad \textstyle\int_{t=0}^{x} h(t)dt &= -\left[log(S(t))\right]_{t=0}^{x} \\ \Rightarrow \quad -\textstyle\int_{t=0}^{x} h(t)dt &= log(S(x)) - log(S(0)) \\ \Rightarrow \quad -\textstyle\int_{t=0}^{x} h(t)dt &= log(S(x)) \quad \text{[Since, S(0)=1]} \\ \Rightarrow \quad S(x) &= exp\left(-\int_{t=0}^{x} h(t)dt\right) \end{aligned} \tag{10}$$

6.2. **Expressing $f(t)$ in terms of $S(t)$ and $h(t)$.** :

$$\begin{aligned} f(t) &= \frac{d}{dt} F(t) \\ &= -\frac{d}{dt} S(t) \left[Since, S(t) = 1 - F(t)\right] \end{aligned} \tag{11}$$

Differentiating both sides of equation

$$\begin{aligned} S'(t) &= \tfrac{d}{dt}\left(exp\left(-\textstyle\int_{t=0}^{x} h(t)dt\right)\right) \\ \Rightarrow \quad f(t) &= -\tfrac{d}{dt}\left(-exp\left(\textstyle\int_{t=0}^{x} h(t)dt\right)\right) \end{aligned} \tag{12}$$

6.3. **Expressing $h(t)$ in terms of $f(t)$ and $S(t)$.** From equation

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

Again,

$$\begin{aligned} h(t) &= -\frac{d}{dt} log(S(t)) \\ &= -\frac{d}{dt} log\left(\int_{x=t}^{\infty} f(x)\right) \text{From equation} \end{aligned} \tag{13}$$

7. Some Survival Distributions

Distribution	**PDF ($f(t)$)**	Survival Function ($S(t)$)	Hazard Rate ($h(t)$)
Exponential	$\lambda * exp(-\lambda t)$	$exp(-\lambda t)$	λ
Gamma	$\frac{\lambda}{\Gamma(k)}(\lambda t)^{k-1} exp(-\lambda t)$		
Weibull	$\alpha\lambda^{\alpha} t^{\alpha-1} exp(-(\lambda t)^{\alpha})$	$exp(-(\lambda t)^{\alpha})$	$\alpha\lambda^{\alpha} t^{\alpha-1}$
Rayleigh	$\frac{t}{\sigma^2} exp\left(-\frac{t^2}{2\sigma^2}\right)$	$exp\left(-\frac{t^2}{2\sigma^2}\right)$	$\frac{t}{\sigma^2}$
Log-Normal	$\frac{1}{\sqrt{2\pi}\tau t} exp\left(-\frac{(ln(t)-v)^2}{2\tau^2}\right)$	$\Phi\left(\frac{v-ln(t)}{\tau}\right)$	$\frac{\phi\left(\frac{v-ln(t)}{\tau}\right)/\tau t}{\Phi\left(\frac{v-ln(t)}{\tau}\right)}$

7.1. **The Exponential Distribution.** The most important one parameter family of life distributions is the family of exponential distributions. This importance is partly due to the fact that several of the most commonly used families of life distributions are two or three parameter extensions of the exponential distributions. Exponential distributions, with their constant hazard rates, form a baseline for evaluating other families. Because they have only one parameter, they are quite simple to describe and are exceptionally amenable to statistical analysis.

Probability Density Function.

$$f(t) = exp(-\lambda t), \quad t \geq 0$$

Survival Function.

$$S(t) = exp(-\lambda t), \quad t \geq 0$$

Hazard Function.

$$h(t) = \lambda, \quad t \geq 0$$

Cumulative Density Function.

$$F(t) = 1 - exp(-\lambda t), \quad t \geq 0$$

Odd's ratio.

$$\Phi(t) = \frac{S(t)}{F(t)} = \frac{1}{e^{\lambda t} - 1}$$

Total Time on Test Transform.

$$\psi(p) = \frac{p}{\lambda}, \quad 0 \leq p \leq 1$$

Lorenz Curve.

$$L(p) = p + (1-p)log(1-p), \quad 0 \leq p \leq 1$$

Gini's Index.

$$G = 1/2$$

Raw Moments.

$$\mu_r = E(T^r) = \int_0^\infty t^r \lambda e^{-\lambda t} dt = \frac{\Gamma(r+1)}{\lambda^r}, \quad r > -1$$

Normalised Moments.

$$\lambda_r = \frac{\mu_r}{\Gamma(r+1)} = \frac{1}{\lambda^r}$$

Mean and Variance.

$$\mu = \frac{1}{\lambda}$$
$$\sigma^2 = \frac{1}{\lambda^2}$$

Cumulative Density Function.

$$F(t) = 1 - exp(-\lambda t), \quad t \geq 0$$

Co-efficient of Variation.

$$CV(T) = \frac{\mu}{\sigma} = 1$$

7.2. **Characterization of the Exponential Distribution.**

7.2.1. *Lack of Memory Property.* A distribution F is exponential if and only if:

$$S(x+t) = S(x)S(t) \quad \text{for all } x, t \geq 0$$

7.2.2. *Constant Hazard Rate.* A distribution has a constant hazard rate if and only if it is an exponential.

7.2.3. *Mean Residual Life.* A distribution F has a mean residual life independent of age if and only if it is an exponential.

7.2.4. A distribution F is an exponential distribution if and only if any one of the following holds:

(1) F has support $[0, \infty)$ and has a density that is both log concave and log convex on its support.
(2) F is in the intersection of the classes IHR and DHR distributions.
(3) F is in the intersection of the classes of IHRA and DHRA distributions.
(4) F is in the intersection of the classes of NBUE and NWUE distributions.

7.2.5. A distribution F is exponential if and only if

$$(S(t))^k = S(kt), \quad k = 1, 2, \ldots, t \geq 0$$

7.3. **Some Basic Properties of Exponential Distribution.**

7.3.1. *Closure under Minima.* If X_1, X_2 are independent random variables following exponential distribution with parameters $\lambda_1, \lambda_2 (> 0)$, then $min(X_1, X_2)$ will follow an exponential distribution with parameter $\lambda_1 + \lambda_2$

7.3.2. *Infinite Divisibility.*

7.3.3. *Poisson Process.* Let, E be a particular event and $t \geq 0$ be a non-negative real which denotes time. Let, $N(t)$ denote the number of times event E occurs in the time interval $[0, t)$. Then $N(t)$ is a counting process taking values in the set of non-negative integers.

Definition 1. If $N(t)$ satisfies the following conditions then $\{N(t); t \geq 0\}$ is called a Poisson Process.

(1) $N(0) = 0, N(t) \geq 0, \forall t \geq 0$ and $N(t+h) - N(t)$ has the same distribution as $N(h), h > 0$.
(2) $N(t+h) - N(t)$ is independent of $\{N(s); s \leq t\}$, $\forall t > 0, h > 0$
(3) For each, $\lambda > 0$, such that $P(N(h) = 1) = \lambda h + o(h)$ as $h \to 0$ and $P(N(h) > 1) = o(h)$ as $h \to 0$

Definition 2. If $\{N(t); t \geq 0\}$ is a Poisson Process satisfying the above conditions then,

$$P(N(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad , k = 0, 1, 2, \ldots$$

Exponential Distribution. If T_i denotes the inter-arrival time between $(i-1) - th$ and $i - th$ occurances of event E, then $\{T_i, i \geq i\}$ is a sequence of iid random variables having an exponential distribution with parameter $\lambda > 0$.

Connection between Exponential and Geometric Distribution. Suppose, Y_n have the geometric distribution with parameter $p = \lambda/n$, where $n \geq \lambda$. Then,

$$\begin{aligned} S_n(x) &= P\left[\frac{Y_n}{n} > x\right] \\ &= P\left[Y_n > nx\right] \\ &= P\left(1 - \frac{\lambda}{n}\right)^{[nx]} \end{aligned}$$

Clearly,

$$lim_{n\to\infty} S_n(x) = e^{-\lambda x}, \quad x \geq 0$$

Random Sums.

7.4. **The Gamma Distribution.**

Probability Density Function.

$$f(t|\alpha,\beta) = \frac{\lambda^v t^{v-1} e^{-\lambda t}}{\Gamma(v)}, \quad t \geq 0$$

(1) $f(t)$ is completely montone, log convex and decreasing for $0 < v < 1$.
(2) $f(t)$ is log concave and unimodal, for $v \geq 1$, with mode at the point

$$x = \frac{v-1}{\lambda}$$

Survival Function.

$$S(t|\lambda, v) = \int_t^\infty f(x|\alpha,\beta)dx$$

When v is an integer then the survival function can be written in a closed form:

$$S(t|\lambda, v) = \sum_{k=0}^{v-1} e^{-\lambda t}(\lambda t)^k/k!$$

(1) $S(t)$ is log concave on $[0,\infty)$ for $v \geq 1$ and
(2) $S(t)$ is log convex on $[0,\infty)$ for $v \leq 1$

Hazard Function. It does not have a convenient form.

(1) $h(t)$ is increasing in $t \geq 0$ when $v > 1$
(2) $h(t)$ i constant λ when $v = 1$
(3) $h(t)$ is decreasing in $t \geq 0$ when $0 < v < 1$
(4) Also,

$$\lim_{x\to 0} h(x) = \begin{cases} 0, & \text{for } v > 1 \\ \lambda, & \text{for } \quad v = 1 \\ \infty, & \text{for } 0 < v < 1 \end{cases}$$

Cumulative Function.

$$F(t) = 1 - S(t|\lambda, v), \quad t \geq 0$$

It is log concave for all v.

Raw Moments.

$$E(T^r) = \frac{\Gamma(r+v)}{\Gamma(v)\lambda^r}, \quad r > -v$$

Mean and Variance.

$$E(T) = \frac{v}{\lambda}$$
$$V(T) = \frac{v}{\lambda^2}$$

Co-efficient of Variation.

$$CV = \frac{1}{\sqrt{v}}$$

7.4.1. *Derivation of Gamma Distribution.*

From Poisson Process. If X is the waiting time for the v-th jump in the Poisson Process $\{N(t), t \geq 0\}$ with parameter λ then X is a gamma random variable with parameter λ and v(an integer).

By introducing Moment Parameter. BY introduction of a moment parameter β in the exponential density, we have the gamma density.

7.5. **Some Properties of Gamma Distribution.**

Suppose that X_1 and X_2 are independent, positive non-degenerate random variables. Then $X_1 + X_2$ and X_1/X_2 are independent if and only if X_1 and X_2 have gamma distribution with the same scale paramerter.

Limit of Survival Function.

$$lim_{v \to \infty} S\left(\frac{vx}{\lambda}|\lambda, v\right) = \begin{cases} 1, & x < 1 \\ 0, & x > 1 \end{cases}$$

If $Y = (X - v/\lambda)/(\sqrt{v}/\lambda)$, then the limiting distribution of Y as $v \to \infty$. is a normal distribution with mean 0 and variance 1.

Suppose that $X_1, \ldots, X_n$ are iid exponential variables with parameter λ then the random variable $\sum X_i$ has a gamma distribution with shape parameter n.

7.6. **The Weibull Distribution.**

Probability Density Function.

$$f(t|\lambda,\alpha) \quad = \quad \alpha\lambda(\lambda t)^{\alpha-1} exp\left(-(\lambda t)^{\alpha}\right), \quad t \geq 0$$

$logf(t)$ is convex for $0 < \alpha < 1$, linear for $\alpha = 1$ and concave for $\alpha > 1$. The distribution is unimodal for $\alpha \geq 1$ and the mode is $\frac{1}{\lambda}\left(1 - 1/\alpha\right)^{1/\alpha}$

Survival Function.

$$S(t|\lambda,\alpha) = exp\left(-(\lambda t)^{\alpha}\right)$$

Hazard Function.

$$h(t|\lambda,\alpha) \quad = \quad \alpha\lambda(\lambda t)^{\alpha-1}, \quad t \geq 0$$

(1) $h(t)$ is increasing in $t \geq 0$ when $\alpha > 1$
(2) $h(t)$ i constant λ when $\alpha = 1$
(3) $h(t)$ is decreasing in $t \geq 0$ when $0 < \alpha < 1$
(4) Also,

$$\lim_{x\to\infty} h(x) \quad = \begin{cases} 0, & \text{for } \alpha < 1 \\ \infty, & \text{for } \alpha > 1 \end{cases}$$

(5) The hazard rate is concave when $1 \leq \alpha \leq 2$ and otherwise convex.

$$\lim_{x\to 0} h(x) \quad = \begin{cases} \infty, & \text{for } 0 < \alpha < 1 \\ \lambda, & \text{for } \alpha = 1 \\ 0, & \text{for } \alpha > 1 \end{cases}$$

Cumulative Function.

$$F(t|\lambda,\alpha) = 1 - exp\left(-(\lambda x)^{\alpha}\right)$$

Raw Moments.

$$E(T^r) \quad = \quad \frac{\Gamma\left(\frac{r}{\alpha} + v\right)}{\lambda^r}, \quad r > -\alpha$$

Mean and Variance.

$$E(T) \quad = \quad \frac{\Gamma\left(\frac{1}{\alpha} + 1\right)}{\lambda}$$

$$V(T) \quad = \quad \frac{\Gamma\left(\frac{2}{\alpha} + 1\right) - \Gamma^2\left(\frac{1}{\alpha} + 1\right)}{\lambda^2}$$

Co-efficient of Variation.

$$CV \quad = \quad \left[\frac{2\alpha}{B(1/\alpha, 1/\alpha)} - 1\right]^{1/2}$$

Gini Index.

$$Gini(\alpha) \quad = \quad 1 - \frac{1}{2^{1/\alpha}}$$

Total Time on Test Transform.

$$\psi(p) = \int_0^{F^{-1}(p)} z^{(1/\alpha)-1} e^{-z} dz$$

7.7. **The Rayleigh Distribution.**

Probability Density Function.

$$f(t) = \frac{t}{\sigma^2} exp\left(-\frac{t^2}{2\sigma^2}\right), \quad t \geq 0$$

(1) The density is unimodal with mode at σ

Survival Function.

$$S(t) = exp\left(-\frac{t^2}{2\sigma^2}\right), \quad t \geq 0$$

Hazard Function.

$$h(t) = \frac{t}{\sigma^2}, \quad t \geq 0$$

Cumulative Function.

$$F(t) = 1 - \frac{x}{\sigma^2} exp\left(-\frac{t^2}{2\sigma^2}\right), \quad t \geq 0$$

Raw Moments.

$$E(T^r) = \frac{2^{r/2}\sigma^r \Gamma\left(1+\frac{r}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}$$

Mean and Variance.

$$E(T) = \frac{2^{1/2}\sigma \Gamma\left(1+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}$$

$$V(T) = \frac{2^1\sigma^2 \Gamma\left(1+\frac{2}{2}\right) - \left(2^{1/2}\sigma\Gamma\left(1+\frac{1}{2}\right)\right)^2}{\Gamma^2(1/2)}$$

Co-efficient of Variation.

$$CV = \left[\frac{2\alpha}{B(1/\alpha, 1/\alpha)} - 1\right]^{1/2}$$

Gini Index.

$$Gini(\alpha) = 1 - \frac{1}{2^{1/\alpha}}$$

Total Time on Test Transform.

$$\psi(p) = \int_0^{F^{-1}(p)} z^{(1/\alpha)-1} e^{-z} dz$$

NOTES. Suppose that X and Y are $N(0,1)$ random variables with mean 0 and variance σ^2. And let,

$$Z = \left[X^2 + Y^2\right]^{1/2}$$

Then Z^2 has an exponential distribution with scale parameter $1/\sqrt{(2\sigma^2)}$ and Z follows a Weibull distribution with scale parameter $\frac{1}{\sqrt{2\sigma^2}}$ and shape parameter 2. This distribution was derived by Rayleigh and has come to be known as Rayleigh distribution.

7.8. **Log-Normal Distribution.**

Probability Density Function.

$$\begin{aligned} f(t) &= \frac{1}{x\sqrt{2\pi\sigma^2}} exp\left\{-\frac{1}{2\sigma^2}(log(x)-\mu)^2\right\}, \quad x>0 \\ &= \frac{1}{x}\phi\left(\frac{log(x)-\mu}{\sigma}\right) \end{aligned}$$

(1) For any real θ,

$$lim_{x\to\infty} x^\theta f(x) = lim_{x\to 0} x^\theta f(x) = 0$$

(2) Also for the k-th derivative of f ,

$$x^\theta f^{(k)}(x) = lim_{x\to 0} x^\theta f(x) = 0$$

(3) The log-normal density is unimodal, with mode at the point,

$$exp\left(\mu-\sigma^2\right) = \lambda^{-1} exp\left(-1/\alpha^2\right) = exp\left((\beta-1)/\alpha^2\right)$$

Survival Function.

$$\begin{aligned} S(t) &= \Phi\left(\frac{log(x)-\mu}{\sigma}\right), \quad \infty<\mu, \sigma>0 \\ &= \Phi\left(log(\lambda x)^\alpha\right), \quad \lambda, \alpha>0 \\ &= \Phi\left((\alpha log(x)) - \left(\frac{\beta}{\alpha}\right)^2\right), \quad \alpha>0, \beta<\infty \end{aligned}$$

;where Φ is the std normal distribution function. $\mu = -\log\lambda, \sigma = \frac{1}{\alpha}$, and $\beta = \alpha^2\mu = \mu/\sigma^2$

Hazard Function. The hazard rate of the log-normal distribution cannot be expressed in closed form. However, using the hazard rate may be written in terms of the hazard rate h_{Normal} of the standard normal distribution as follows:

$$h(t) = \frac{h_{Normal(w)}}{\sigma} exp(-\sigma w - \mu)$$

;where $w = ((logx)-\mu)/\sigma$

(1) For all real θ,

$$lim_{x\to 0} x^\theta h(x) = 0$$

(2) Also for the k-th derivative of h ,

$$lim_{x\to 0}x^{\theta}h(x) = 0$$

(3) The hazard rate of the lognormal distribution is unimodal, with mode at $exp(\sigma z + \mu)$, where z is the unique solution of the equation

$$h_{Normal}(z) = z + \sigma$$

Cumulative Function.

$$F(t) = 1 - S(t), \quad t \geq 0$$

Raw Moments.

$$\begin{aligned} E(T^r) &= \frac{1}{\lambda^r} exp\left(\frac{r^2}{2\alpha^2}\right) \\ &= exp\left(\frac{r\beta}{\alpha^2} + \frac{r^2}{2\alpha^2}\right) \\ &= exp\left(r\mu + \frac{r^2\sigma^2}{2}\right) \end{aligned}$$

(1) The LogNormal distribution is not uniquely determined by its moments. Feller, gave an example of the density function

$$f(x) = \frac{1}{x\sqrt{w\pi}} exp\left(-\frac{1}{2}(logx)^2\right)\left[1 + \alpha sin(2\pi log(x))\right], \quad -1 \leq \alpha \leq 1, x \geq 0$$

Mean and Variance.

$$\begin{aligned} E(T) &= \frac{1}{\lambda} exp\left(\frac{1}{2\alpha^2}\right) \\ &= exp\left(\frac{\beta}{\alpha^2} + \frac{1}{2\alpha^2}\right) \\ &= exp\left(\mu + \frac{\sigma^2}{2}\right) \\ V(T) &= exp\left(2\mu + \sigma^2\right)\left[e^{\sigma^2 - 1}\right] \\ &= e^{1/\alpha^2}\left(1 - e^{1/\alpha^2}\right)/\lambda^2 \\ &= exp\left((1 - 2\beta)/\alpha^2\right)\left[e^{1/\alpha^2} - 1\right] \end{aligned}$$

Co-efficient of Variation.

$$\begin{aligned} CV &= \left[e^{\sigma^2} - 1\right]^{1/2} \\ &= \left[e^{1/\alpha^2} - 1\right]^{1/2} \end{aligned}$$

NOTES.

(1) The normal distribution arises as a limiting distribution for sums, the lognormal distribution arises as a limiting distribution for products. If $U_1, U_2, U_3, \ldots$ is a sequence of independent and identically distributed random variables with finite expectation μ and variance σ^2. Then, if $V_i = exp(U_i), i = 1, 2, 3, \ldots$ then it follows that $[\Pi_i V_i]^{1/\sqrt{n}}$ has a limiting "standard lognormal distribution".
(2) At both 0 and ∞, the density of the lognormal distribution tends to 0 faster than any power of x. This fact may partially explain the usefulness of the distribution in fitting some types of data.
(3) The lognormal distribution is "stretched to the right" as

$$Mode < Median < Mean$$
$$exp\left(\mu - \sigma^2\right) < e^{\mu} < exp\left(\mu + \frac{\sigma^2}{2}\right)$$

•

8. Bathtub Hazard Rates

A distribution is said to have a bathtub hazard rate if for some $0 \le a \le b$, the hazard rate $h(t)$ is decreasing in $t, 0 \le t \le a$, is constant in the interval $a \le t \le b$, and is increasing in $t, t \ge b$.

From at least two points of view, distributions with a bathtub hazard rate have considerable intuitive appeal. The first and most commonly given idea is based upon the assumption that the device or organism under consideration comes from a mixture of individuals of varying in- herent strength. Those individuals with life threatening defects at birth suffer a high rate of early mortality, but as a device ages without failure, the conditional probability that a life-threatening defect is present di- minishes and so the hazard rate decreases. There comes a time a when deaths due to birth defects no longer occur and accidents become the only significant cause of death, so the hazard rate becomes constant. But eventually, at time b, the adverse effects of age begin to take their toll and the hazard rate begins to rise. This concept of a mixture is related to a Bayesian concept in which the mixture is not real, but is treated mathematically as such due to uncertainty about the un- derlying distribution.

A second intuitive basis for bathtub hazard rates applies primarily to biological organisms. When young, such organisms may have im- mature immune systems, they may have difficulty competing for food, and they may suffer from a number of other disadvantages that diminish as the organism grows and matures. During the period of matu- ration, the hazard rate decreases. But eventually, the organism fully matures and again, the adverse effects of age take effect and cause the hazard rate to increase. This idea was already apparent in the writings of Price (1771) who wrote that human life, from birth up- wards, grows gradually stronger until the age of 10 years, then slowly loses strength until the age of 50, then more rapidly loses strength until, at 70 or 75, it is brought back to all the weakness of the first month. Bathtub hazard rates for human life lengths are explicitly discussed by Wittstein (1883) who based his ideas on studies of mortality tables.

The just described origin of bathtub hazard rates for biological or- ganisms has its counterpart for mechanical systems. A new system may suffer from “bugs”, that is, from errors of design or of construction. Moreover, the operators of the system may be initially inexperienced. As the system ages, the potential for bugs or human error diminishes, causing the hazard rate to decrease. But after a while, the effects of aging cause the hazard rate to rise. As noted in the New York Times Magazine, July 18, 2006, p. 56, David Lochbaum of the Union of Con- cerned Scientists has pointed out that the bathtub curve applies to the safety of nuclear power plants; they are most dangerous when first brought on line, or at the end of their life cycle. More recent examples where bathtub hazard rates were found useful for fitting data are given by Rajarshi and Rajarshi (1988). For a survey of such hazard rates, see Lai, Xie and Murthy (2001).

Bathtub hazard rates motivate a process called “burn-in” for man- ufactured items. The idea is to place the device in a simulated service (or more stressful) environment to discover defects before the device is introduced into actual service. For a review of this common practice, see Block and Savits (1997) and the references contained therein. From a Bayesian point of view, burn-in is related to belief; see Lynn and Singpurwalla (1997).

8.1. **Distributions having Bath-tub shaped hazard rate.**

Example 1. Suppose X_0 and X_2 are two independent random variables Weibull distributions having survival functions $S_0 = exp(-x^{\beta})$ and $S_2 = exp(-(\lambda x)^{\alpha})$. Then if X_1 is a random variable such that $X_1 = min(X_0, X_2)$ and X is such that

$$X = \begin{cases} X_1 & \text{w.p. } \pi \\ X_2 & \text{w.p. } (1-\pi) \end{cases}$$

Then the hazard rate of the random variable X is bathtub shaped provided that $\alpha \geq 2$ and $\beta \leq 1$.

Example 2. Suppose X_0 and X_2 are two independent random variables having survival functions $S_0 = (1-x)^4, 0 \leq x \leq 1$ and $S_2 = exp(-x^2), x \leq 0$. Then if X_1 is a random variable such that $X_1 = min(X_0, X_2)$ and X is such that

$$X = \begin{cases} X_1 & \text{w.p. } 0.5 \\ X_2 & \text{w.p. } 0.5 \end{cases}$$

Then the hazard rate of the random variable X is bathtub shaped.

Example 3.

$$h(x) = \begin{cases} \frac{1}{2\sqrt{x}} - \frac{1}{2\sqrt{a}} + \lambda \ , & x \leq a \\ \lambda \ , & a \leq x \leq b \\ \lambda x^2/b \ , & x \geq b \end{cases}$$