# Chapter 2. Functions of survival time

As we mentioned in Chapter 1, the interested failure time can be defined as a non-negative random variable. Other than the usual distribution functions (probability density function (p.d.f.) and cumulative density function (c.d.f.)) that we use in classical probability concept, we frequently discuss survival function, hazard function and cumulative hazard function in survival analysis. In Section 2.1, we give two tables to represent the definition of the distribution functions in continuous and discrete cases, respectively. The relationship between them will be discussed in Section 2.2.

## 2.1 Definitions:

| | $T$ is a continuous r.v taking positive values |
|---|---|
| (i) Survival function | $S(t) = \Pr(T \geq t)$ where $S(0) = 1$ and $S(\infty) = 0$ |
| (ii) Cumulative density function | $F(t) = 1 - S(t) = \Pr(T < t)$ |
| (iii) Probability density function | $f(t) = F^{'}(t) = -S^{'}(t) = \lim\limits_{\Delta \to o^{+}} \Pr(t \leq T < t + \Delta)/\Delta$ |
| (iv) Hazard function | $h(t) = f(t)/S(t)$ |
| (v) Cumulative hazard function | $H(t) = \int_0^t h(\mu) d\mu$ |

| | $T$ is a discrete r.v taking values $t_1 < t_2 < \cdots$ |
|---|---|
| (i) Probability density function | $f(t) = \begin{cases} f_i, & \text{if } t = t_i \\ \hline 0, & \text{otherwise} \end{cases}$ where $f_i = \Pr(T = t_i)$ |
| (ii) Survival function | $S(t) = \sum f_i I(t_i \geq t)$ where $I(A) = \begin{cases} 1, & \text{if } A \text{ is true} \\ \hline 0, & \text{otherwise} \end{cases}$ |
| (iii) Cumulative density function | $F(t) = 1 - S(t) = \sum f_i I(t_i < t)$ |
| (iv) Hazard function | $h(t) = f(t)/S(t)$ |
| (v) Cumulative hazard function | $H(t) = \sum h(t_i) I(t_i \leq t)$ |

## 2.2 The relationship between functions:

If $T$ is continuous, then

(i) $S(t) = 1 - F(t) = \int_t^\infty f(x)dx$

(ii) $-S^{'}(t) = f(t) = F^{'}(t)$

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

(iv) $S(t) = \exp\{-H(t)\} = \exp\{-\int_0^t h(x)dx\}$

(v) $f(t) = h(t)\exp\{-\int_0^t h(x)dx\} = h(t)\exp\{-H(t)\}$

If $T$ is discrete, then

(i) $h(t_i) = f(t_i)/S(t_i) = f(t_i)/[f(t_i) + f(t_{i+1}) + \cdots]$, this implies $1 - h(t_i) = S(t_{i+1})/S(t_i)$.

(ii) $S(t_i) = \Pr(T \geq t_i) = \Pr(T \geq t_i | T \geq t_{i-1}) \Pr(T \geq t_{i-1}) = [S(t_i)/S(t_{i-1})]S(t_{i-1}) = [1 - h(t_{i-1})]S(t_{i-1}) = \prod_{j=1}^{i-1}[1 - h(t_j)]$

(iii) $f(t_i) = h(t_i)S(t_i) = h(t_i)\prod_{j=1}^{i-1}[1 - h(t_j)]$

Up to now, we know that given one of the five functions, the other four functions can be derived. In the following, we will introduce four important parameters (mean, variance, percentile and mean residual life) and give the relationship of them to the survival function, respectively.

Mean:

$\mu = E(T) = \int_0^\infty tf(t)dt = -\int_0^\infty tdS(t) = -tS(t)|_0^\infty + \int_0^\infty S(t)dt = \int_0^\infty S(t)dt$

Variance:

$\sigma^2(T) = \int_0^\infty t^2 f(t)dt - \mu^2 = 2\int_0^\infty tS(t)dt - [\int_0^\infty S(t)dt]^2$

Percentile

The $100p$th percentile of the distribution of $T$ is the value $X_p$ such that $F(X_p^+) \geq p$ and $F(X_p^-) \leq p$, where $F(a^+) = \lim_{\Delta \to 0^+} F(a + \Delta)$ and $F(a^-) = \lim_{\Delta \to 0^+} F(a - \Delta)$ Or equivalently, $S(X_p^-) \geq 1 - p$ and $S(X_p^+) \leq 1 - p$

Note: If $F$ is continuous, then $F(X_p) = p$. median $= X_{0.5}$.

Mean residual life at time $x$ :

$\text{mrl}(x) = E(T - x | T \geq x) = \int_x^\infty (t - x)f(t)dt/S(x) = \int_x^\infty S(t)dt/S(x)$

Note: $\mu = \text{mrl}(0)$.

Since the hazard function is being discuss in many survival analysis problem, we give a more detail review of its statistical meaning in the following. At the end of this Chapter, we point out some extensions of the hazard functions of bivariate cases.

Statistical meaning of the hazard function:

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} = \lim_{\Delta \to 0^+} \frac{\Pr(t \le T < t+\Delta)/\Delta}{S(t)} \\ &= \lim_{\Delta \to 0^+} \frac{\Pr(t \le T < t+\Delta, T \ge t)/\Delta}{\Pr(T \ge t)} = \frac{\Pr(t \le T < t+\Delta | T \ge t)}{\Delta} \end{aligned}$$

The extension of hazard function of bivariate non-negative r.v's.

Let $S(s,t) = \Pr(T_1 \ge s, T_2 \ge t)$, and $f(s,t)$ = joint probability density function. The following are three different extensions of hazard rate.

(i) $h(s,t) = f(s,t)/S(s,t)$

(ii) $h(s,t) = (h_1(s,t), h_2(s,t))^T = (\frac{\partial ln(S(s,t))}{\partial s}, \frac{\partial ln(S(s,t))}{\partial t})^T$

(iii) $h_1(s) = \lim_{\Delta \to 0^+} \Pr(s \le T_1 < s+\Delta | T_1 \ge s, T_2 \ge s)/\Delta$

$h_2(t) = \lim_{\Delta \to 0^+} \Pr(t \le T_2 < t+\Delta | T_1 \ge t, T_2 \ge t)/\Delta$

$h_{12}(s,t) = \lim_{\Delta \to 0^+} \Pr(s \le T_1 < s+\Delta | T_1 \ge s, T_2 = t)/\Delta$, for $s > t$

$h_{21}(s,t) = \lim_{\Delta \to 0^+} \Pr(t \le T_2 < t+\Delta | T_1 = s, T_2 \ge t)/\Delta$, for $t > s$