Relationships between different parameters

Summary: continuous case

If T is a continuous non-negative random variable,

- Cumulative distribution function
  $$F(t) = P(T \leq t)$$
- Probability density function
  $$\mathrm{f}(t) = \frac{dF(t)}{dt} = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$
- Survival function
  $$\mathrm{S}(t) = P(T > t)$$
- Hazard function
  $$\mathrm{h}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$
- Cumulative hazard function
  $$H(t) = \int_0^t h(u)du$$

Relationship

- There is a one-to-one relationship between these parameters.
- If we know any one of the functions, we can calculate the rest functions.
- They are equivalent ways to summarize information in a random variable.

- Result 1:

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

- If f(t) is known, we can find $h(t)$.
  - First find S(t)=$\int_t^{+\infty} f(u)du$.
  - Then use it to find $h(t)$.

An Example

If $T \sim \exp(\lambda)$ with density function

$$f(t) = \lambda e^{-\lambda t} \quad \text{for } t \geq 0.$$

Then

$$S(t) = \int_t^{+\infty} \lambda e^{-\lambda u} du = e^{-\lambda t},$$

and

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

- Result 2:

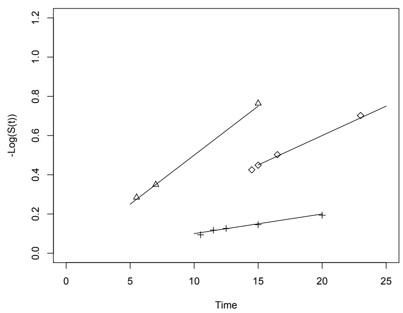
$$h(t) = -\frac{d\log[S(t)]}{dt}$$

- If S(t) is known, we can use it to find $h(t)$.
- Interesting comparison:

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

$$\mathrm{f}(t) = -\frac{dS(t)}{dt}$$

An Example



Straight lines
$\Longrightarrow$Exponential distribution

Difference slopes
$\Longrightarrow$ Difference hazard rates

- Result 3:
  $H(t) = -\log(S(t))\,, S(t) = \exp(-H(t))$
- Note that

$\int_0^{+\infty} h(u)du \quad H(+\infty) = -\log(S(+\infty)) = +\infty.$

- Recall that $f(t)$ with $\int_0^{+\infty} f(u)du = 1.$
- Hazard function is not a probability density.

Summary: discrete case

If T is a discrete non-negative random variable,

- Probability distribution(mass) function
  $\mathrm{f}(t_j) = P(T = t_j)$ for $j = 1,2,\cdots$
- Cumulative distribution function
  $\mathrm{F}(t) = P(T \le t) = \sum_{t_j \le t} \mathrm{f}(t_j)$
- Survival function
  $\mathrm{S}(t) = P(T > t) = \sum_{t_j > t} \mathrm{f}(t_j)$
- Hazard function
  $\mathrm{h}(t_j) = \frac{f(t_j)}{S(t_{j-1})}$ for $j = 1,2,\cdots$
- Cumulative hazard function
  $H(t) = \sum_{t_j \le t} \mathrm{h}(t_j)$

Results

- Similarly we have the following results for discrete random variables.
- Result 1:

  $\mathrm{h}(t_j) = \frac{f(t_j)}{S(t_{j-1})}=\frac{S(t_{j-1})-S(t_j)}{S(t_{j-1})}=1-\frac{S(t_j)}{S(t_{j-1})}$
- Result 2:

  $S(t) = \prod_{t_j \le t}(1-\mathrm{h}(t_j))$
- Why?

  $\prod_{t_j \le t}(1-\mathrm{h}(t_j))=\prod_{t_j \le t}\frac{S(t_j)}{S(t_{j-1})}$=S($\tilde{t}$)=S(t),

  where $\tilde{t}$ is the largest t_j such that $t_j \le t$.

Summary

- We explored how different parameters are related.
- These results are useful later in the estimation.
- For example, if we know how to estimate the hazard function, we can obtain an estimator for the survival function.