Chapter 2 & Appendix B

PubH 7450

©Wei Pan

Email: weip@biostat.umn.edu

Http: www.biostat.umn.edu/~weip

Outline

- Chapter 2: Basic quantities
- Section 3.5: Likelihood for survival data
- Appendix B: Reveiw of maximum likelihood theory

Chapter 2: Basic quantities

- Recall X: time of event of interest.
  we observe some copies of X, possibly censored/truncated; because X is ..., we'd like to summarize information in the form of ... , based on which we can do 2-sample comparisons & regression.

- §2.2. Survival function

$$S(x) = Pr(X > x) = 1 - Pr(X \leq x) = 1 - F(x).$$

  X: r.v.
  x: any observed/given value
  F: CDF

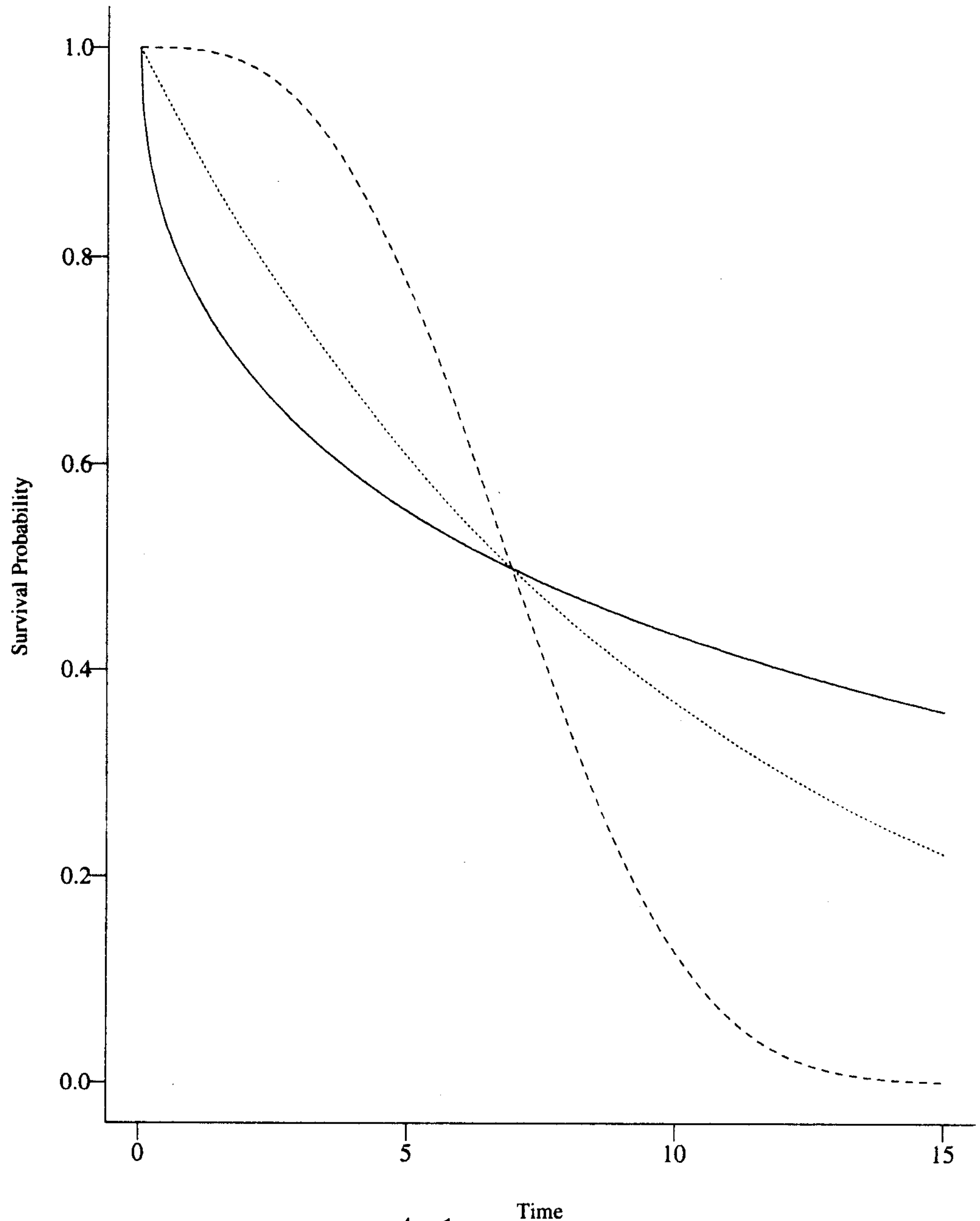
- $0 \leq S(x) \leq 1$, $S(x)$ is non-increasing in x.

- For continuous X,
  $S(x) = \int_x^\infty f(t)dt$, where PDF $f(x) = -\frac{dS(x)}{dx}$.

- Example 2.1 and Fig 2.1: Weibull distribution.

$$S(x) = \exp(-\lambda x^{\alpha}), \qquad \lambda > 0, \qquad \alpha > 0.$$

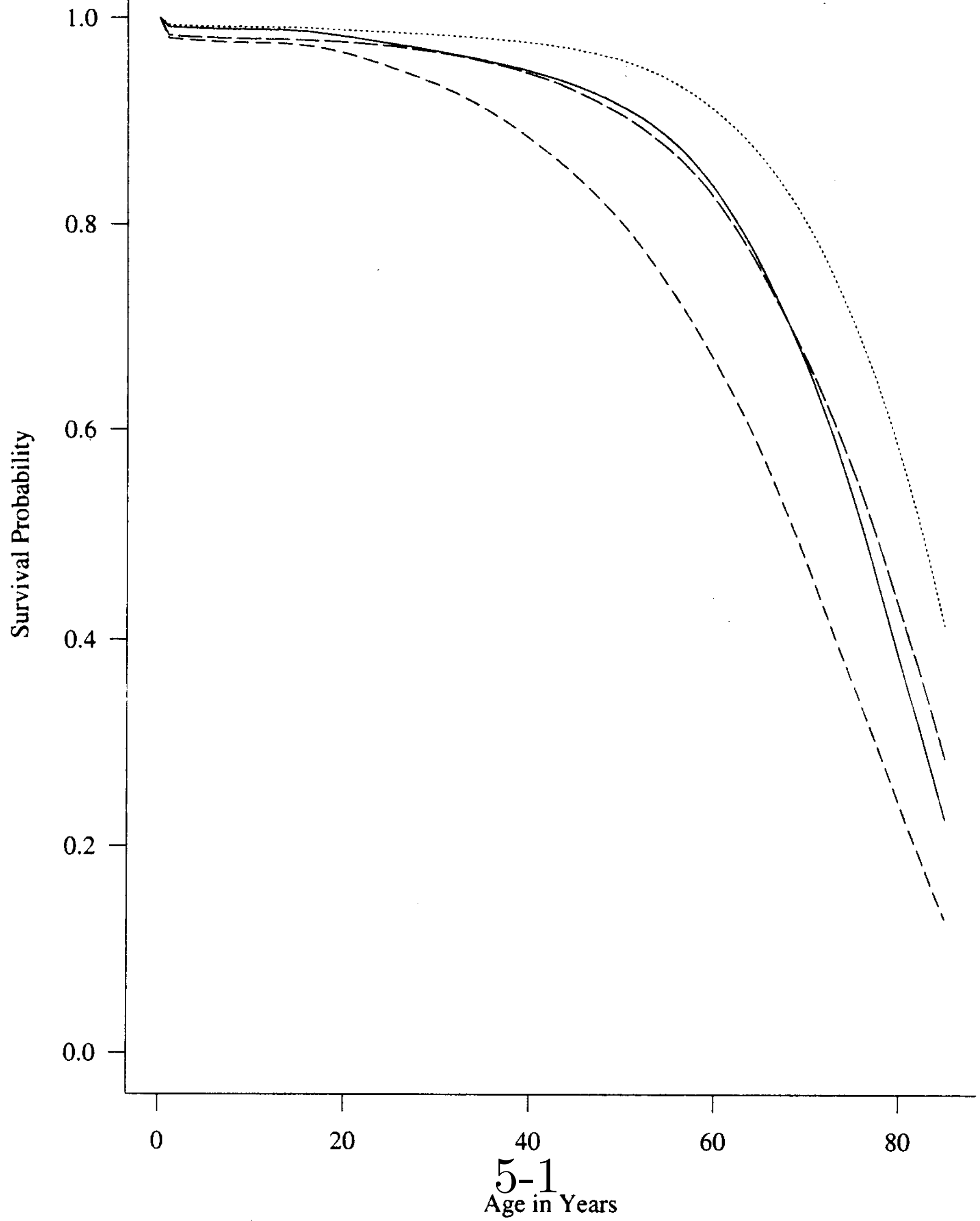
$\alpha = 1 \Longrightarrow$ exponential distr.



4-1

Figure 2.1 *Weibull Survival functions for* $\alpha = 0.5$, $\lambda = 0.26328$ (————); $\alpha = 1.0$, $\lambda = 0.1$ (······); $\alpha = 3.0$, $\lambda = 0.00208$ (------).

- Example 2.2 and Fig 2.2: yearly survival curves for all causes of mortality for US and each of the 50 states, by race and sex.



5-1

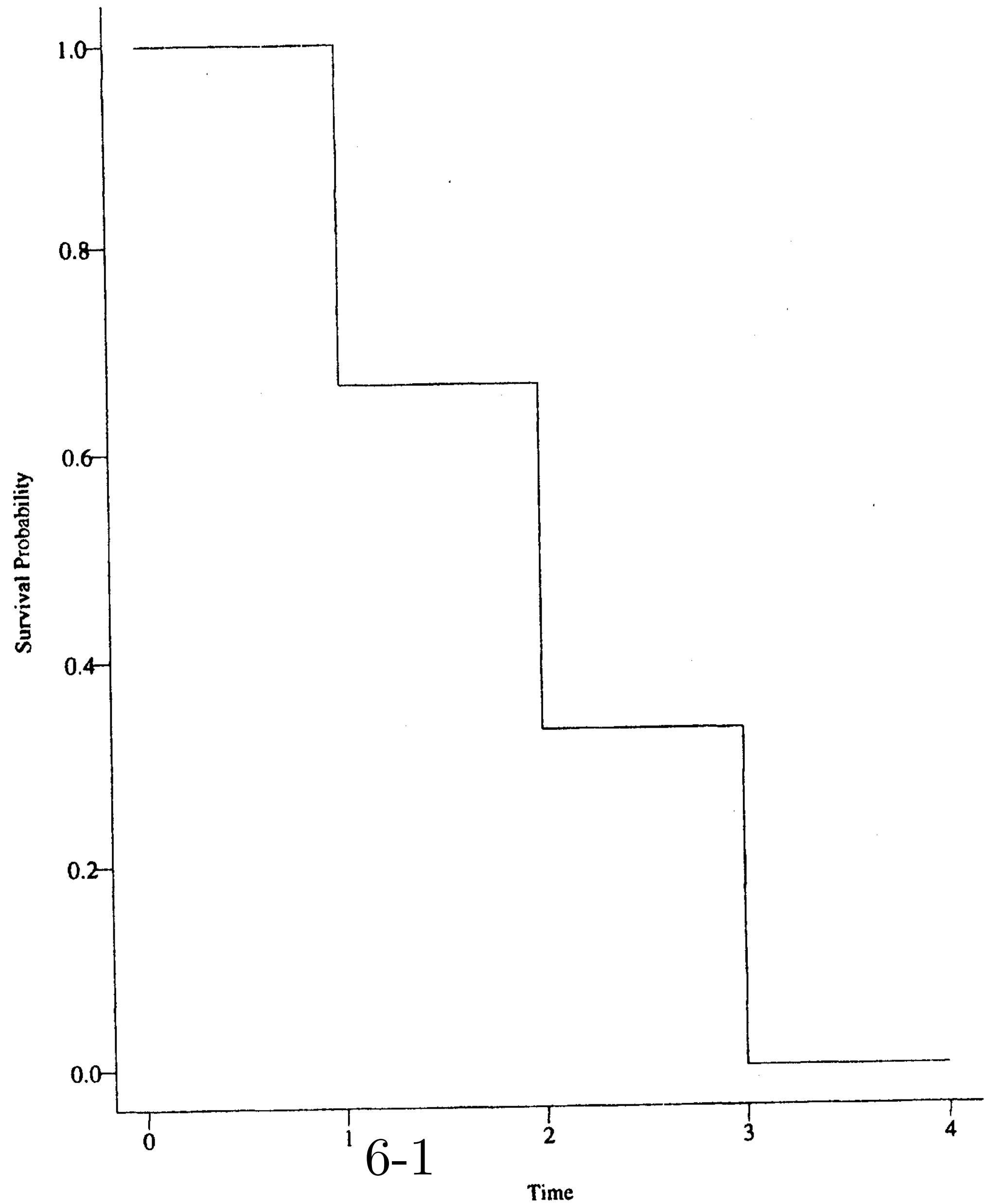
Figure 2.2 *Survival Functions for all cause mortality for the US population in 1989. White males (————); white females (······); black males (------); black females (——— ———).*

- X: discrete
  $p(x_j) = Pr(X = x_j), \quad , j = 1, 2, ... \quad , x_1 < x_2 < ...$
  $S(x) = \sum_{x_j > x} p(x_j)$.

- Example 2.3 and Fig 2.3: $p(x_j) = Pr(X = j) = 1/3$ for $j = 1, 2, 3$.

$$S(x) = \begin{cases} 1 & \text{if } 0 \leq x \\ 2/3 & \text{if } 1 \leq x < 2 \\ 1/3 & \text{if } 2 \leq x < 3 \\ 0 & \text{if } 3 \leq x \end{cases}$$

  E.g. $S(1) = Pr(X > 1) = \sum_{x_j > 1} p(x_j) = p(x_2) + p(x_3) = 2/3$.



6-1

Figure 2.3 *Survival function for a discrete random lifetime*

- §2.3. Hazard function/rate

$$h(x) = \lim_{\Delta x \to 0} \frac{P(x \le X < x + \Delta x | X \ge x)}{\Delta x}$$

$$\Longrightarrow P(x \le X < x + \Delta x | X \ge x) \approx h(x) \cdot \Delta x.$$

- X: continuous

$$h(x) = \lim_{\Delta x \to 0} \frac{P(x \le X < x + \Delta x)/\Delta x}{S(x)} = \frac{f(x)}{S(x)} = -\frac{d \log S(x)}{dx}.$$

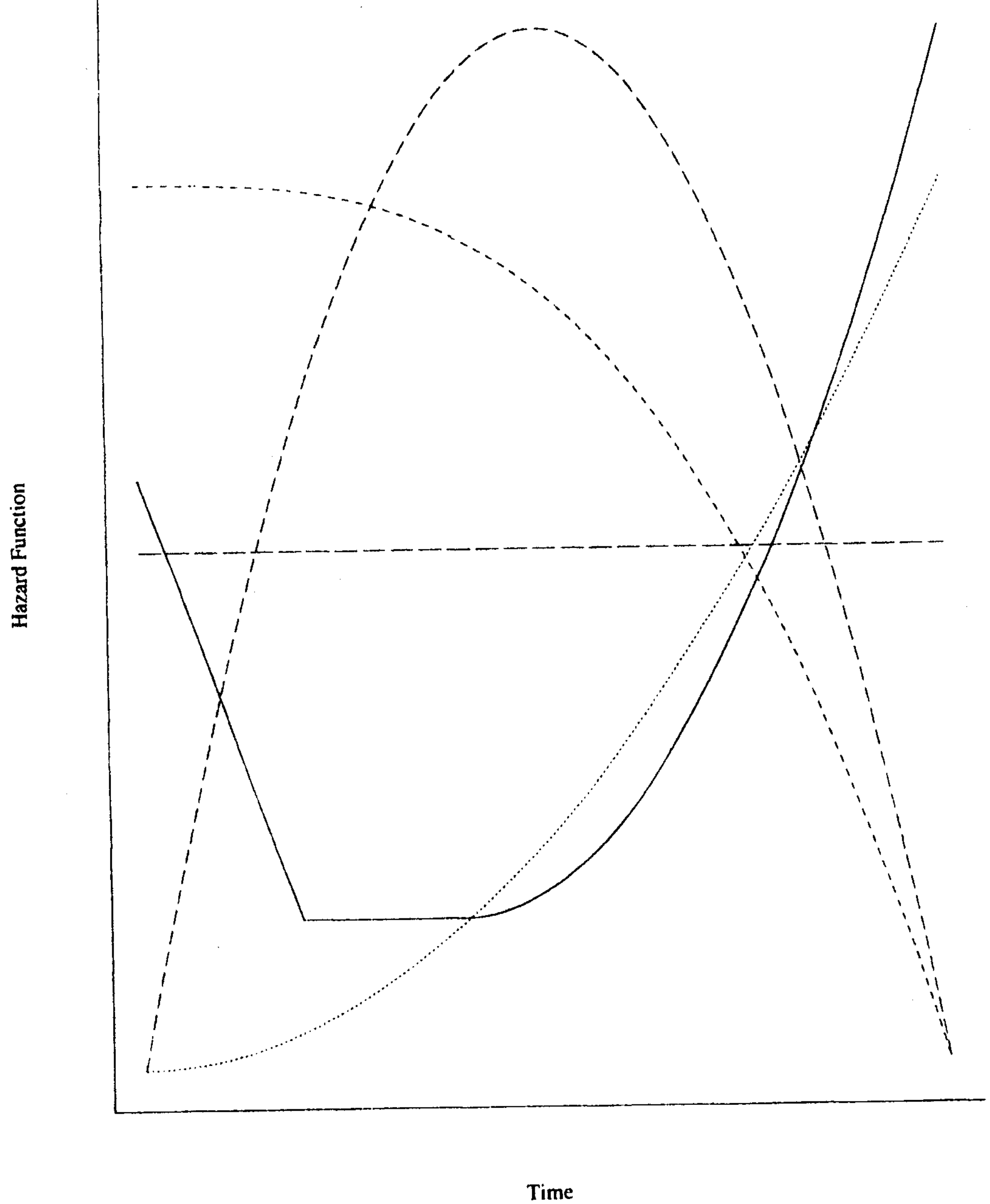
- Cumulative hazard function:

$$H(x) = \int_0^x h(t)dt = -\log S(x).$$

$$S(x) = \exp[-H(x)] = \exp\left(-\int_0^x h(t)dt\right).$$

- $h(x) \ge 0$, $H(x) \ge 0$

  Why? —-



7-1

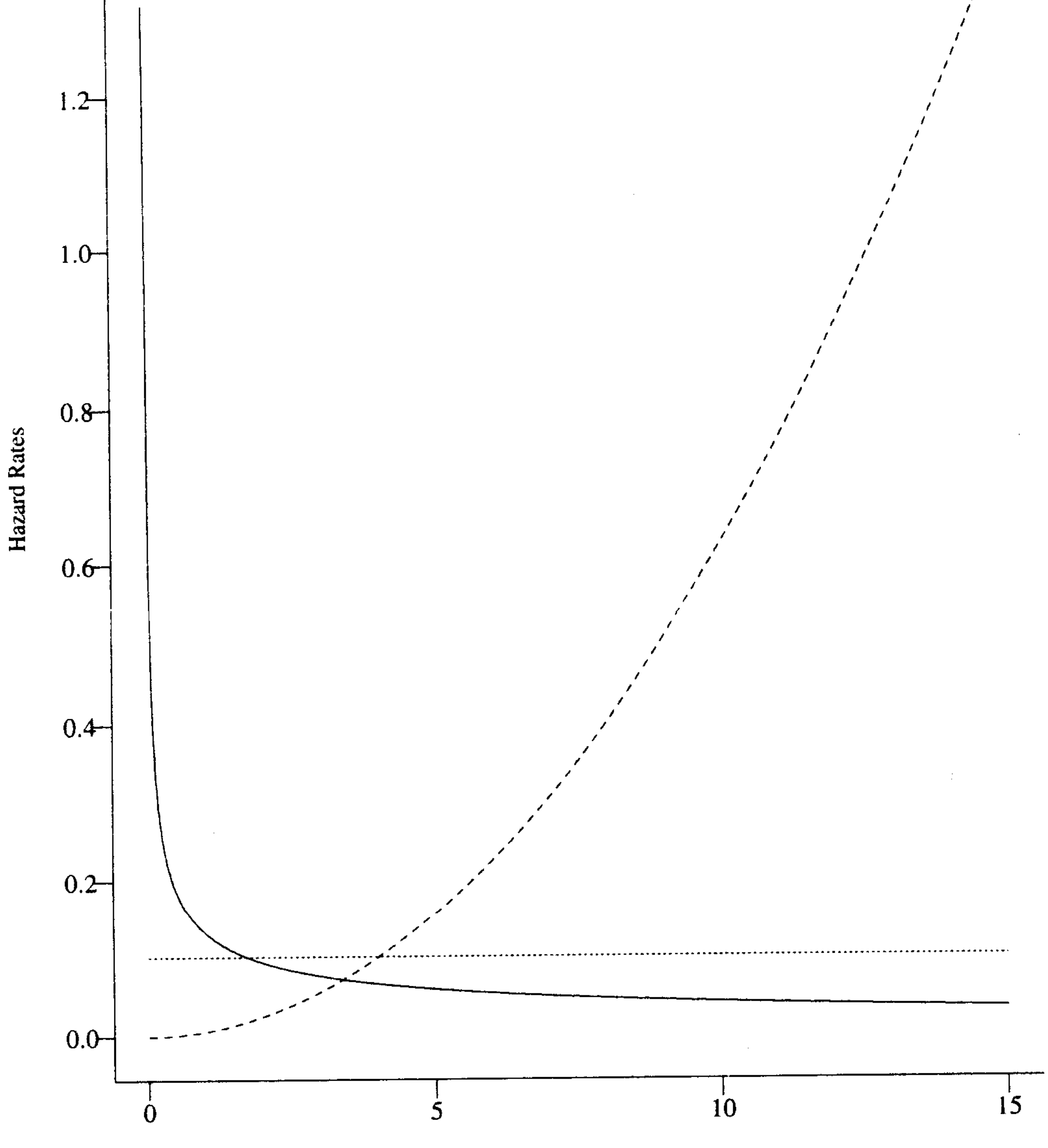
Figure 2.4 *Shapes of hazard functions. Constant hazard (——— ———); increasing hazard (------); decreasing hazard (- - - - - -); bathtub shaped (————); humpshaped (——–—).*

- Example 2.1. Weibull distribution

  $S(x) = \exp(-\lambda x^{\alpha})$, $\lambda > 0$, $\alpha > 0$

  $\Longrightarrow h(x) = \alpha\lambda x^{\alpha-1}$.

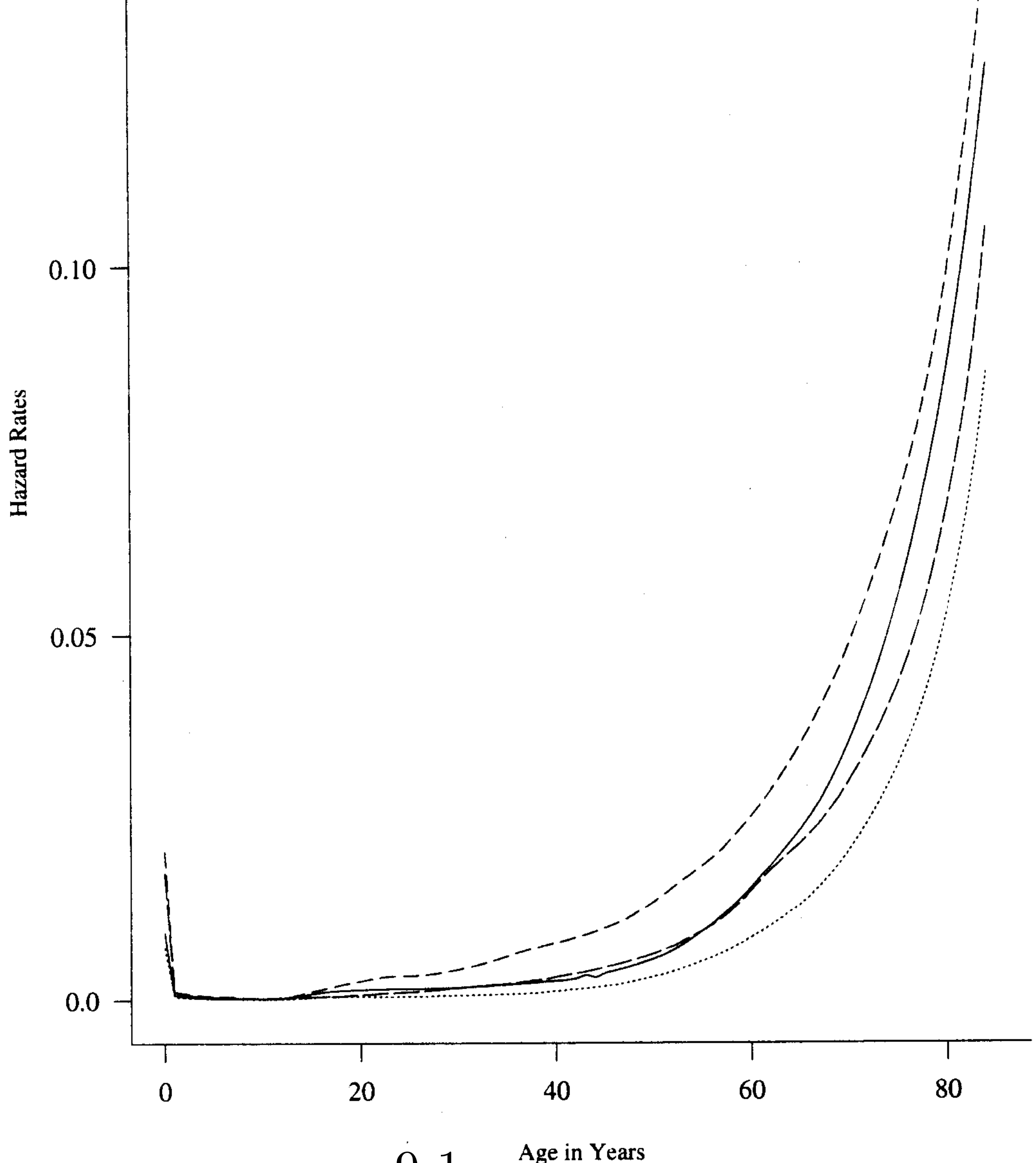
  How does $h(x)$ look like?



8-1

Figure 2.5 *Weibull hazard functions for* $\alpha = 0.5$, $\lambda = 0.26328$ (———); $\alpha = 1.0$, $\lambda = 0.1$ (------); $\alpha = 3.0$, $\lambda = 0.00208$ (— — —).

- Example 2.2. 1989 US mortality hazard rates.



9-1

Figure 2.6 *Hazard functions for all cause mortality for the US population in 1989. White males (———); white females (······); black males (- - - - - -); black females (—— ——).*

- Interpretation of hazard function:
  $h(x) \cdot \Delta x \approx P(x \leq X < x + \Delta x | X \geq x)$, instantaneous probability of an event at time x^+, given ...

- $h(x) \propto h(x) \cdot \Delta x$
  Is it possible to have $h(x) > 1$?

- Example: Fig 2.6; suppose $h(0.1) = 0.02$, $h(10) \approx 0$, $h(20) = 0.005$, $h(70) = 0.05$. Given you four people at ages 0.1, 10, 20 and 70 respectively, order their chance of dying in the next moment.

  ...

- $h(x)$: expected number of events per person-unit time, if $h(.)$ stay constant within the interval.
  E.g., h(20)=0.000001, if the hazard stays the same within the next two year, what is the expected # of deaths within 2 yrs for 1000,000 people at 20?

...

Why?

- X: discrete
- for $j = 1, 2, ...,$

$$h(x_j) = P(X = x_j | X \geq x_j) = \frac{p(x_j)}{S(x_{j-1})} = 1 - \frac{S(x_j)}{S(x_{j-1})}.$$

$$S(x) = \prod_{x_j \leq x} \frac{S(x_j)}{S(x_{j-1})} = \prod_{x_j \leq x} [1 - h(x_j)].$$

- Example 2.3: $p(x_j) = Pr(X = j) = 1/3$ for $j = 1, 2, 3.$

$$S(x) = \begin{cases} 1 & \text{if } 0 \leq x \\ 2/3 & \text{if } 1 \leq x < 2 \\ 1/3 & \text{if } 2 \leq x < 3 \\ 0 & \text{if } 3 \leq x \end{cases}$$

$$h(x_j) = \begin{cases} 1/3 & \text{for } j = 1 \\ (1/3)/(2/3) = 1/2 & \text{for } j = 2 \\ (1/3)/(1/3) = 1 & \text{for } j = 3 \\ 0 & \text{o/w} \end{cases}$$

- $H(x) = \sum_{x_j \le x} h(x_j)$
  $\Longrightarrow S(x) \ne \exp[-H(x)]$.

- Another definition:
  $H(x) = -\sum_{x_j \le x} \log[1 - h(x_j)]$
  $\Longrightarrow S(x) = \exp[-H(x)]$.
  Note: $H(x) \approx \sum_{x_j \le x} h(x_j)$ if ...

- §2.4. Mean residual life function & median life
- Mean residual life at time x:

$$mrl(x) = E(X - x|X > x) = \frac{\int_x^\infty (t-x)f(t)dt}{S(x)} = \frac{\int_x^\infty S(t)dt}{S(x)}.$$

  See p.35 for deriving the last equality.
  Numerator:

- Mean lifetime
  $\mu = mrl(0) = E(X) = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt.$
- pth quantile (=100pth percentile) of the distr of X is the smallest x_p s.t. $S(x_p) \le 1 - p$; i.e.

$$x_p = \inf\{t : S(t) \le 1 - p\}.$$

  If X is continuous r.v., then x_p is obtained by solving $S(x_p) = 1 - p$.
- Median lifetime is $x_{0.5}$

- §2.5. Common parametric models

  Exp.

  Weibull

  Gamma

  Lognormal

  Pareto

  Gompertz

Hazard Rates, Survival Functions, Probability Density Functions, and Expected Lifetimes for Some Common Parametric Distributions

Distribution	*Hazard Rate* $h(x)$	*Survival Function* $S(x)$	*Probability Density Function* $f(x)$	*Mean* $E(X)$
Exponential $\lambda > 0, x \geq 0$	λ	$\exp[-\lambda x]$	$\lambda \exp(-\lambda x)$	$\frac{1}{\lambda}$
Weibull $\alpha, \lambda > 0,$ $x \geq 0$	$\alpha\lambda x^{\alpha-1}$	$\exp[-\lambda x^{\alpha}]$	$\alpha\lambda x^{\alpha-1} \exp(-\lambda x^{\alpha})$	$\frac{\Gamma(1 + 1/\alpha)}{\lambda^{1/\alpha}}$
Gamma $\beta, \lambda > 0,$ $x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - I(\lambda x, \beta)$*	$\frac{\lambda^{\beta} x^{\beta-1} \exp(-\lambda x)}{\Gamma(\beta)}$	$\frac{\beta}{\lambda}$
Log normal $\sigma > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - \Phi\left[\frac{\ln x - \mu}{\sigma}\right]$	$\frac{\exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right]}{x(2\pi)^{1/2}\sigma}$	$\exp(\mu + 0.5\sigma^2)$
Log logistic $\alpha, \lambda > 0, x \geq 0$	$\frac{\alpha x^{\alpha-1}\lambda}{1 + \lambda x^{\alpha}}$	$\frac{1}{1 + \lambda x^{\alpha}}$	$\frac{\alpha x^{\alpha-1}\lambda}{[1 + \lambda x^{\alpha}]^2}$	$\frac{\pi \mathrm{Csc}(\pi/\alpha)}{\alpha\lambda^{1/\alpha}}$ if $\alpha > 1$
Normal $\sigma > 0,$ $-\infty < x < \infty$	$\frac{f(x)}{S(x)}$	$1 - \Phi\left[\frac{x-\mu}{\sigma}\right]$	$\frac{\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}{(2\pi)^{1/2}\sigma}$	μ
Exponential power $\alpha, \lambda > 0, x \geq 0$	$\alpha\lambda^{\alpha} x^{\alpha-1} \exp\{[\lambda x]^{\alpha}\}$	$\exp\{1 - \exp[(\lambda x)^{\alpha}]\}$	$\alpha e \lambda^{\alpha} x^{\alpha-1} \exp[(\lambda x)^{\alpha}] - \exp\{\exp[(\lambda x)^{\alpha}]\}$	$\int_0^{\infty} S(x)dx$
Gompertz $\theta, \alpha > 0, x \geq 0$	$\theta e^{\alpha x}$	$\exp\left[\frac{\theta}{\alpha}(1 - e^{\alpha x})\right]$	$\theta e^{\alpha x} \exp\left[\frac{\theta}{\alpha}(1 - e^{\alpha x})\right]$	$\int_0^{\infty} S(x)dx$
Inverse Gaussian $\lambda \geq 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$\Phi\left[\left(\frac{\lambda}{x}\right)^{1/2}\left(1 - \frac{x}{\mu}\right)\right] - e^{2\lambda/\mu}\Phi\left\{-\left[\frac{\lambda}{x}\right]^{1/2}\left(1 + \frac{x}{\mu}\right)\right\}$	$\left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left[\frac{\lambda(x-\mu^2)}{2\mu^2 x}\right]$	μ
Pareto $\theta > 0, \lambda > 0$ $x \geq \lambda$	$\frac{\theta}{x}$	$\frac{\lambda^{\theta}}{x^{\theta}}$	$\frac{\theta\lambda^{\theta}}{x^{\theta+1}}$	$\frac{\theta\lambda}{\theta - 1}$ if $\theta > 1$
Generalized gamma $\lambda > 0, \alpha > 0,$ $\beta > 0, x \geq 0$	$\frac{f(x)}{S(x)}$	$1 - I[\lambda x^{\alpha}, \beta]$	$\frac{\alpha\lambda^{\beta} x^{\alpha\beta-1} \exp(-\lambda x^{\alpha})}{\Gamma(\beta)}$	$\int_0^{\infty} S(x)dx$

* $I(t, \beta) = \int_0^t u^{\beta-1} \exp(-u)du/\Gamma(\beta)$.

§3.5 Likelihood Construction

- X: $f(x)$ or $p(x)$, $S(x)$, $h(x)$
- exact, $X_i = x_i$: $L_i = f(x_i)$.
- right censored, $X_i > c_i$: $L_i = S(c_i)$.
- left censored, $X_i \le c_i$: $L_i = 1 - S(c_i)$.
- interval censored, $X_i \in (B_i, U_i]$: $L_i = S(B_i) - S(U_i)$.
- left truncation, (y_i, x_i): $L_i = f(x_i)/S(y_i)$.
- Total likelihood
  $L = \prod_{i \in E} f(x_i) \prod_{i \in RC} S(c_i) \prod_{i \in LC} [1 - S(c_i)] \prod_{i \in IC} [S(B_i) - S(U_i)] ...$
- For right-censored data: (T_i, δ_i), $i = 1, ..., n$
  $L_i = f(t_i)$ if $\delta_i = 1$;

$L_i = S(t_i)$ if $\delta_i = 0$. Thus,

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^{n} h(t_i)^{\delta_i} S(t_i).$$

Review: Maximum Likelihood Thoery (Appendix B)

- Given data $Y_1, ..., Y_n$ with distribution parameter θ
$\Longrightarrow \log L(\theta)$
$\Longrightarrow$ MLE $\widehat{\theta} = argmax_\theta \log L(\theta)$. usually by solving the *score equation*:
$$U(\theta) = \frac{\partial \log L}{\partial \theta} = 0.$$
- Score function: $U(\theta)$.
- Observed Fisher's information:
$$I(\theta) = -\frac{\partial^2 \log L}{\partial \theta^2}.$$
- Expected Fisher's information:
$$i(\theta) = E(I(\theta)) = E(U(\theta)U(\theta)'),$$
which may be hard to get. And $I(\theta)$ may be preferred over $i(\theta)$.

- Asymptotic normality of MLE:

$$\widehat{\theta} \overset{a.}{\sim} N(\theta, I^{-1}(\theta)).$$

  In practice, use $I^{-1}(\widehat{\theta})$.

- Test H_0: $\theta = \theta_0$ vs H_1: $\theta \neq \theta_0$.
  - Wald test. Under H_0, $\widehat{\theta} \overset{a.}{\sim} N(\theta_0, I^{-1}(\theta_0)) \Longrightarrow$
    $\chi^2_W = (\widehat{\theta} - \theta_0)' I(\widehat{\theta})(\widehat{\theta} - \theta_0) \overset{a.}{\sim} \chi^2_p$, $p = dim(\theta)$.
  - Score test. Under H_0, $U(\theta) \overset{a.}{\sim} N(0, I(\theta_0)) \Longrightarrow$
    $\chi^2_S = U(\theta_0)' I^{-1}(\theta_0) U(\theta_0) \overset{a.}{\sim} \chi^2_p$
    Note:
  - Likelihood ratio test.
    $\chi^2_{LR} = 2 \log L(\widehat{\theta}) - 2 \log L(\theta_0) \overset{a.}{\sim} \chi^2_p$
  - Computational cost:
  - Performance:
- Invert a test to get CI.

Wald 95% CI of θ with $p = 1$:

$$\widehat{\theta} \pm 1.96\sqrt{\widehat{Var}(\widehat{\theta})}$$

where $\widehat{Var}(\widehat{\theta}) = I^{-1}(\widehat{\theta})$ or $i^{-1}(\widehat{\theta})$.

- Example B.1.

  Data: (T_i, δ_i), $i = 1, ..., n$.

  Model: $X_i \sim Exp(\lambda)$.

- Likelihood...

$$
\begin{aligned}
L &= \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \\
&= \prod_{i=1}^{n} \left(\lambda e^{-\lambda t_i}\right)^{\delta_i} \left(e^{-\lambda t_i}\right)^{1-\delta_i} \\
&= \prod_{i=1}^{n} \lambda^{\delta_i} e^{(-\lambda t_i)\delta_i + (-\lambda t_i)(1-\delta_i)} \\
&= \prod_{i=1}^{n} \lambda^{\delta_i} e^{-\lambda t_i} \\
&= \lambda^{\sum_{i=1}^{n} \delta_i} e^{-\lambda \sum_{i=1}^{n} t_i}.
\end{aligned}
$$

$$\log L = \sum_{i=1}^{n} \delta_i \log \lambda - \lambda \sum_{i=1}^{n} t_i.$$

$$\frac{\partial \log L}{\partial \lambda} = \sum_{i=1}^{n} \delta_i / \lambda - \sum_{i=1}^{n} t_i = 0 \Longrightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} t_i}.$$

$I = -\frac{\partial^2 \log L}{\partial \lambda^2} = \frac{\sum_{i=1}^{n} \delta_i}{\lambda^2}.$

- Test H_0: $\lambda = 1$
- Wald test:
  $\chi^2_W = \left(\frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} t_i} - 1 \right)^2 \frac{\sum_{i=1}^{n} \delta_i}{\left(\frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} t_i} \right)^2} = \frac{\left(\sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} t_i \right)^2}{\sum_{i=1}^{n} \delta_i} \overset{a.}{\sim} \chi^2_1$

  under H_0.
- Score test:
  $\chi^2_S = \left(\frac{\sum_{i=1}^{n} \delta_i}{1} - \sum_{i=1}^{n} t_i \right)^2 \cdot \frac{1}{\sum_{i=1}^{n} \delta_i} = \chi^2_W \overset{a.}{\sim} \chi^2_1$ under H_0.

- LRT:

$$\begin{aligned}\chi^2_{LR} &= 2\left(\sum_{i=1}^{n}\delta_i \log\hat{\lambda} - \hat{\lambda}\sum_{i=1}^{n} t_i - (\sum_{i=1}^{n}\delta_i \log 1 - 1\cdot\sum_{i=1}^{n} t_i)\right) \\ &= 2\left(\sum_{i=1}^{n}\delta_i \log\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n} t_i} - \sum_{i=1}^{n}\delta_i + \sum_{i=1}^{n} t_i\right) \overset{a.}{\sim} \chi^2_1\end{aligned}$$

under H_0.