URV VAL D TR BUT ON AND THE R CHARACTER T C

A CONTRIBUTION TO
THE ENCYCLOPEDIA OF BIOSTATISTICS

Jo P. K e
Med ca Co ege of W sco s
8701 Watertow P a k Road
M waukee W 53226

IN R

of T. hese classes are defined and some basic properties of these classes are discussed in the final section.

HE SURVIVAL FUNC ION

he basic quantity employed to describe time-to-event phenomena is the survival function. his function, also nown as the survivor function or survivorship function, is the probability an individual survives beyond time t. It is defined asond

. Here we have defined $S(t) = Pr[T \geq t]$ as was the case in [3] and [4]. his definition was used to ma e later formulas for the discrete case simpler. Other authors (c.f. [5] and [6]) have defined $S(t) = Pr[T > t]$ which ma es the relationship $S(t) = 1 - \quad (t)$ hold for both the discrete and continuous case.

HE HAZARD FUNC ION

basic quantity, foundational in survival analysis, is the hazard function. his function is also nown as the conditional failure rate in reliability, the force of mortality in demography, the age-specific failure rate in epidemiology, the inverse of the Mill's ratio in economics or simply as the hazard rate. he hazard rate is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr\left[t \leq T < t + \Delta t | T \geq t\right]}{\Delta t}. \tag{1}$$

he hazard rate is a non-negative function. It tells us how quic ly individuals of a given age are e periencing the event of interest. he quantity $h(t)\Delta t$ is the appro imate probability that an individual who has survived to age t will e perience the event in the interval $(t, t + \Delta t)$.

his function is particularly useful in determining the appropriate failure distributions utilizing qualitative information about the mechanism of failure and for describing the way in which the chance of e periencing the event changes with time. here are many general shapes for the hazard rate. Some generic types of hazard rates are increasing, decreasing, constant, bathtub-shaped or hump-shaped hazard rates. Models with increasing hazard rates arise when there is natural aging or wear-out. Decreasing hazard functions are much less common but find occasional use when there is a very early li elihood of failure such as in certain types of electronic devices or in patients e periencing certain types of transplants. Decreasing hazard rates often arise as models for heterogenous populations where the hazard rates of members of the population are random (See railty models). Most often a bathtub-shaped hazard is appropriate in populations followed from birth. Most population mortality data follows this type of hazard function where, during an early period, deaths result primarily from infant diseases after which the death rate stabilizes followed by an increasing hazard rate due to the natural aging process. inally, id019489.99980 Dﬂ(the) jﬂ6jﬂ3400 Dﬂ(t) jﬂ34Dﬂ(w) jﬂ670

ra2,ases0o5B4ID(ﬂDwf)[(pj-637D0ilytucess.)] Jﬂmortalit) jﬂ31140

$$(t) = \int_0^t h(\ )d\ \ = -\ln\left[S\left(t\right)\right].$$

hus for continuous lifetimes we have the following relationship:

$$S\left(t\right) = \text{e p}\left\{-\quad(t)\right\} = \text{e p}\left\{-\int_0^t h(\ )d\ \ \right\}.$$

One particular distribution, which is fle ible enough to accommodate increasing $(\alpha > 1)$, decreasing $(\alpha < 1)$, or constant hazard rates $(\alpha = 1)$, is the Weibull distribution. Hazard rates, $\mathrm{h}(x) = \alpha\,\lambda\,x^{\alpha-1}$, are plotted in igure 2 for the Weibull distribution with $\lambda = .26328, \alpha = .5; \lambda = .1, \alpha = 1;$ and $\lambda = .00208, \alpha = 3$. One can see that, though the three survival functions have the same basic shape, the three hazard functions are dramatically different.$\diamondsuit$

When T is a discrete random variable, the hazard function is

$$h\left(t_j\right) = \Pr\left(T = t_j | T \geq t_j\right) = \frac{p\left(t_j\right)}{S\left(t_j\right)},\ j = 1, 2, ...$$

Since $p\left(t_j\right) = S\left(t_j\right) - S\left(t_{j+1}\right)$ we have

$$h\left(t_j\right) = 1 - S\left(t_{j+1}\right)/S\left(t_j\right),\ j = 1, 2, ...$$

so that the survival function is related to the hazard function by

$$S\left(t\right) = \prod_{j:t_j<t}\left[1 - h\left(x_j\right)\right].$$

or discrete lifetimes the "cumulative hazard" function is defined by

$$(t) = \sum_{j:t_j<t} h\left(t_j\right). \tag{2}$$

Notice that for this definition the relationship $S(t) = \mathrm{e\ \ p}[- \quad (t)]$ no longer holds true. Some authors (Co and Oa es [3]) prefer to define the cumulative hazard for discrete lifetimes, as

$$(t) = \sum_{t_j < t} \ln\left[1 - h\left(t_j\right)\right], \tag{3}$$

Note that for this definition the relationship for continuous lifetimes, $S(t) = \mathrm{e\ \ p}[- \quad (t)]$ will then be preserved for discrete lifetimes. If the $h(t_j)$ are small, (2) will be a first order appro imation to (3).

 he hazard rate is well-defined quantity for the case where T has both discrete and continuous components. In this case the hazard function defined by (1) will have a continuous part, $h_c(t)$ and a discrete part with mass h_j at time $t_1 < t \quad < \ldots$. he survival function in this case can be e pressed as

$$S(t) = \mathrm{e\ \ p}\left\{ - \int_0^t h_c(\quad) d \quad \right\} \prod_{j: t_j < t} (1 - h_j)$$

 or any survival function one can e press the relationship between the hazard rate and the survival function by the using the notion of a product integral. or a function, G(), define the product integral of $1 - d \quad (\ )$ over the range a to b by

$$P_a^b[1 - d \quad (\ )] = \lim \prod_{k=1}^{r} \{1 - [\quad (\ _k) - \quad (\ _{k-1})]\},$$

where $a = \ _1 < \ldots < \ _r = b$ and the limit is ta en as $r \to \infty$ and $kjjcj$ $tjjjb$

the right and have finite left hand limits. If we define the cumulative hazard rate as

$$(t) = \int_0^t h_c(\ )d \quad + \sum_{j:t_j<t} h_j$$

then the survival function in the continuous, discrete or mi ed case is given by

$$S(t) = P_0^t \left[1 - d \quad (\ )\right].$$

Because of this property the product integral plays an important role in survival analytic techniques.

HE MEAN RESIDUAL LIFE FUNC ION

he fourth basic parameter of interest is the mean residual life at time t. his parameter measures, for individuals of age t, their e pected remaining lifetime. It is defined as

$$mrl(t) = E(T - t|T \geq t).$$

It can be shown, using integration by parts or a partial summation formula, that the mean residual life is the area under the survival curve to the right of t divided by $S(t)$. Note that the mean life, $\mu = mrl(0)$, is the total area under the survival curve.

or a continuous random variable we have

$$mrl(t) = \frac{\int_{t}^{\infty} (\quad - t) f(t) d}{S(t)} = \int_{t}^{\infty} S(\quad)$$

$$\frac{S(mdrl(t))}{S(t)} = .5.$$

he population median is simply the median residual life at time 0.

o illustrate these quantities consider the three Weibull distributions considered earlier. igure 3 shows the mean residual life function for the Weibull models with $\alpha = 0.5, 1.0$ and 3.0. s the figure shows the

$$= \quad exp\left\{\right.$$

$$
\begin{aligned}
S(t) &= \sum_{j:t_j \geq t} p(t_j) \\
&= \prod_{j:t_j < t} [1 - h(t_j)].
\end{aligned}
$$

If T is an integer valued random variable with mean residual lif7l first aging class is the class of increasing hazard rate (IHR) distributions and the dual class ofdual definitions of a IHR class. Since (4) implies that $S^{1/t}(t)$ is increasing in t we have that T is in the IHR class if and only if $S(\theta t) \geq S^{\theta}(t)$. second characterization of the IHR class is that if T is in the IHR class then for any $\lambda > 0$ the quantity $S(t) - e^{-\lambda t}$ has at most one change of sign and if it does have a change in sign then it is from + to −. he class of

c Dﬂﬁ:saijﬂ850 Dﬂ(IHR) jﬂ

rom this second definition we see that T has an NBU distribution if the probability an individual of age t lives an additional x time units is smaller than the probability an individual of age 0 survives to age x.   his aging class includes all the IHR   distributions.

  fifth aging class is the class of new better (worse) that new in e pectation, NBUE (NWUE) distributions.   distribution is in the NBUE (NWUE) class if its mean, μ, is finite and

$$\int_t^\infty S(\quad)d\quad \leq (\geq)\, \mu S(t) \text{ for all } t.$$

he NBUE class is one where the mean residual life of an individual of age t is less that the mean of an individual of age 0.

  final aging class is the class of harmonic new better (worse) than used in e pectation, HNBUE (HNWUE) distributions.   distribution is said to be in the HNBUE (HNWUE) class if its mean is finite and

$$\int_t^\infty S(\quad)d\quad \leq \mu\, \text{e}\ \ \text{p}(-t/\mu).$$

n equivalent definition for the HNBUE class is

$$\left\{\frac{1}{t}\int_0^t \frac{dx}{mrl(x)}\right\}^{-1} \leq mrl(0).$$

his means that for a HNBUE distribution the integral harmonic value of the residual life of an individual of age t is smaller than the same quantity for a newly born individual.

he aging classes are ordered as follows:

$$I \quad R \Longrightarrow I \quad RA \Longrightarrow NBU$$

6. Klein, J.P. and Moeschberger, M. . *Survival Analy i : Method for en ored and Truncated Data*. Springer–Verlag, New Yor , 1997.

Figure 2
Comparison of Weibull Hazard Func

Figure 3

Comparison of Weibull Mean Residual Life Funct