Percentile Residual Life Function for a Class of Life Distributions Having the 'Setting the Clock Back to Zero' Property

B. Raja Rao¹, Jasem M. Alhumoud² and C.V. Damaraju
¹Posthumous

²Civil Engineering Department, Kuwait University, P.O. Box 5969, Safat, Kuwait.
email: jasem@kuc01.kuniv.edu.kw; jasem@civil.kuniv.edu.kw

Abstract

In the present paper, it is demonstrated that the Percentile Residual Life (PRL) function of a group of individuals in biomedical investigations can easily be obtained if the class of their life distributions possesses, what has been called the 'Setting the clock back to zero' property. It is shown that the PRL function for such a family is just its $100(1-\alpha)$ percentile except, perhaps, with a different vector of parameter values. It is proved that the Gompertzian growth process, Krane's family of distributions and the linear hazard exponential distribution have this property and a simple expression is derived for their PRL function.

As a simple application of the main result obtained in the present paper, we have utilized a large-scale serial sacrifice experiment by the National Center of Toxicological Research involving 24,000 female mice that were fed 2-AAF from infancy. The number that had developed bladder neoplasms and/or liver neoplasms is available. It is demonstrated that the tumor incidence intensity at time x gives a class of survival models which has the setting the clock back to zero property. Its PRL function is evaluated and tabulated for several choices of parameter values.

Keywords: Hazard rate; force of mortality; survival function; exponential; Rayleigh and Weibull distributions; the Gompertz growth process; time to tumor distribution; exposure to toxic materials; competing risks.

1 Introduction

Haines and Singpurwalla (1974) have introduced and studied several characterizations of classes of survival functions under the notion of stochastic wear. These characterizations are based on several possible and intuitively appealing criteria to describe the aging or non-aging of a system, a component or a biological organism. In addition to the failure rate and the survival function explain the aging process of organisms and industrial objects in biometry and actuarial sciences,

the Mean Residual Life (MRL), also known as the life expectancy is studied (Chiang, 1968). The Percentile Residual Life (PRL) function is used to supplement these quantities. The basic idea is that some percentile say the $100(1-\alpha)$ percentile, of its residual life decreases (or increases) in time, as the organism ages.

The PRL function can be effectively used in biometry, social studies, actuarial statistics and reliability. Questions such as, "what proportion of the 5-year old children in a population survive their 7th birthday? If a group of individuals have a tumor-free life of 3 years, what proportion of them can go on another 2 years tumor-free?", can be answered in terms of the PRL function such a group of individuals.

The $100(1-\alpha)$ Percentile ($0 < \alpha < 1$) Residual Life (PRL) at age defined t_0 be the $100(1-\alpha)$ percentile of the remaining life given survival up to time t_0 and is denoted by $R(t_0, \alpha)$. As an example if $\alpha=0.05$, then we get the 95% PRL function. That is, 95% of the individuals of a given age t_0 will survive another $R(t_0, 0.05)$ years. An interesting case occurs when $\alpha = 0.5$, which is referred to as the median residual lifetime.

The MRL function, also known as the life expectancy of a living organism is the expected remaining life time of the organism given that it has survived t_0 time units. In symbols

$$e_{t_0} = E(X - t_0 | X \geq t_0) \quad (1-1)$$

The PRL function, denoted by $R(t_0, \alpha)$, on the other hand, is defined as the $100(1-\alpha)$ percentile of the remaining life, given survival up to time t_0 :

$$P[X \geq t_0 + R(t_0, \alpha) | X \geq t_0] = 1 - \alpha \quad (1-2)$$

If the life time random variable X has the survival function

$$S(x) = P(X \geq x), x \geq 0 \quad (1-3)$$

then equation (1-2) becomes the survival function of the remaining life:

$$\frac{S[t_0 + R(t_0, \alpha)]}{S(t_0)} = 1 - \alpha, \quad (1-4)$$

$$\text{i.e. } S[t_0 + R(t_0, \alpha)] = (1 - \alpha)S(t_0)$$

which gives the PRL Function

$$R(t_0, \alpha) = S^{-1}\{ (1 - \alpha) S(t_0) \} - t_0. \quad (1-5)$$

Equation (1-5) shows how long $100(1-\alpha)\%$ of the individuals of age t_0 survive. Alternatively equation (1-5) is expressed in terms of the c.d.f of the life time defined by $F(x) = 1-S(x)$ as

$$R(t_0, \alpha) = F^{-1}[1 - (1 - \alpha)(1 - F(t_0))] - t_0, \quad (1-6)$$

where of course, F is assumed to be continuous and F^{-1} is defined as

$$F^{-1}(x) = \inf [t : F(t) \geq x], \quad 0 < x < 1.$$

The determination of the PRL function $R(t_0, \alpha)$ from either (1-5) or (1-6) can get quite difficult, except, of course, when the life length X has an exponential distribution, with the survival function

$$S(x) = P(X \geq x) = e^{-\theta x}, \quad \theta > 0, x > 0. \quad (1-7)$$

Equation (1-4) becomes

$$e^{-\theta[t_0 + R(t_0, \alpha)]} = (1 - \alpha)e^{-\theta t_0}$$

and

$$R(t_0, \alpha) = -\frac{1}{\theta} \ln(1 - \alpha), \quad (1-8)$$

which is exactly the $100(1-\alpha)$ percentile of the original exponential distribution defined in equation (1-7). This function is independent of the age t_0 of an object. result is, of course, due to the lack of memory property (LMP) of an exponential distribution.

The question that arises is: Are there classes of life distributions, other the class of exponential distributions, for which the PRL function is just 100(1- α) percentile of the original class, except perhaps with a different parameter or parameter vector? In the present paper, it is demonstrated that the PRL function of a class life distributions is just the 100(1- α) percentile of the original life distribution the class has the setting the clock back to zero property (Rao and Talwalker, 1990).

The exponential, the linear hazard exponential, the Gompertz distributions belong to this class and the PRL function is explicitly derived for such distributions. The Weibull distribution does not belong to this class and the PRL function is not easily derivable. A recent model proposed by Chiang and Conforti (1989) leads to a class of life distributions, and this class has been shown to the setting the clock back to zero property. The PRL function is evaluated tabulated for several choices of the parameter values. As pointed out by Schmitlein and Morrison (1981), many of the practical shortcomings associated with the MRL function are not present in the function. Also, if the data are censored, the empirical MRL function can not be evaluated. Even if the data are not censored, a few long time survivors can affect the MRL function. The PRL is used in such situations. The estimation of the PRL function $R(t_0, \alpha)$ with complete and censored data is discussed by many researchers (Barbas et al., 1986; Csorgo and Csorgo, 1987; Chung, 1989). The empirical c.d.f. $F(x)$ is used with complete samples and

Kaplan-Meier estimator is used with censored samples. Feng and Kulasekera (1991) have proposed a smooth nonparametric estimator for the PRL function using Kernel density estimators.

2 Some Preliminaries

Let us suppose that the data available for analysis are a sample from a population, for which the proportion of items placed in service and surviving at age x is given by an unknown differentiable function of non-negative ages. This function will be called the survival function $S(x, \beta)$. Here β may be a single parameter or a vector of parameters. When a parametric space Ω , this defines a family of survival functions $[S(x, \beta), \beta \in \Omega]$. This function is required to satisfy the following conditions

$$S(0, \beta) = 1, S(\infty, \beta) = 0 \text{ and } S'(x, \beta) \leq 0, x \geq 0. \quad (2-1)$$

The mortality function or the probability density function p.d.f. $f(x, \beta), \beta \in \Omega$, of the life length X of the organism is the instantaneous rate of decrease of $S(x, \beta)$, i.e.

$$f(x, \beta) dx = -dS(x, \beta) \quad (2-2)$$

The force of mortality, or the hazard rate, $\gamma(x, \beta)$ is the proportional rate of decrease of $S(x, \beta)$, i.e.,

$$\gamma(x, \beta) = \frac{f(x, \beta)}{S(x, \beta)} \quad (2-3)$$

This shows that the survival function can be expressed in terms of the hazard rate as

$$S(x, \beta) = \exp \left[- \int_0^x \gamma(u, \beta) du \right] = \exp[-H(x, \beta)], \quad (2-4)$$

where the function $H(x, \beta)$ is called the cumulative hazard function. Also, the mortality function can be expressed as

$$f(x, \beta) = \gamma(x, \beta) \exp[-H(x, \beta)]. \quad (2-5)$$

These results are well-mown and have been stated here for ease of presentation. Several useful applications of the function $\gamma(x, \beta)$ have been given in literature under different names. Actuaries call it the force of mortality, with reference to a specific response or life time distribution, which describes distribution of the lifetimes of individuals over a population of individuals, i.e., $\gamma(x, \beta) dx$ represents the probability that an individual of age x will die in interval $(x, x+dx)$. In other words, $\gamma(x, \beta) dx$ is the conditional probability of death in the interval $(x, x + dx)$ given survival up to time x . It is clear from equation (2-3) that $\gamma(x, \beta)$ uniquely determines the p.d.f. $f(x, \beta)$. In the theory of life-testing, (Barlow and Proschan, 1965; Barlow

and Proschan, 1975) several forms of the function $\gamma(x, \beta)$ are used. The following choices give closed forms for the survival and force of mortality functions.

a) $\gamma(x, \beta) = a$ constant θ , which gives an exponential distribution for the length with the p.d.f.

$$f(x, \beta) = \theta e^{-\theta x}, \quad x \geq 0, \quad \theta > 0. \quad (2-6)$$

b) $\gamma(x, \beta) = px^{p-1}$, $p > 0$, which gives the Weibull distribution for the length with the p.d.f

$$f(x, \beta) = p x^{p-1} e^{-x^p}, \quad x \geq 0, \quad p > 0 \quad (2-7)$$

When $p=2$, the resulting distribution is called a Rayleigh distribution.

c) $\gamma(x, \beta) = a+bx$, where a and b are parameters such that $\gamma(x, \beta) \geq 0$, $x \geq 0$. This gives the linear hazard exponential distribution with the p.d.f.

$$f(x, \beta) = (a + bx) \exp \left[- \left(ax + \frac{b}{2} x^2 \right) \right] \quad (2-8)$$

$$d) \gamma(x, \beta) = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m \quad (2-9)$$

where the constants $a_0, a_1, a_2, \dots, a_m$ are parameters such that $\gamma(x, \beta) \geq 0$, $x \geq 0$. This choice gives the general Krane family of life distributions (Krane, 1963; Roa, 1990a). Some of its special cases are

- i) $m = 0$ gives an exponential distribution.
- ii) $m = 1$ gives a linear hazard exponential distribution.
- iii) $a_0 = a_1 = a_2 = \dots = a_{m-1} = 0$ gives a Weibull distribution. Of course, the parameter m for a Weibull distribution need not be an integer.

e) $\gamma(x, \beta) = \tau e^{-\theta x}$, which gives the Gompertz distribution for the life length, with the p.d.f

$$f(x, \beta) = \tau e^{-\theta x} \exp \left[- \frac{\tau}{\theta} (1 - e^{-\theta x}) \right], \quad \tau \geq 0, \quad \theta > 0. \quad (2-10)$$

The exponential distribution has been used by Epstein and Sobel (1953) in industrial life testing. Zelen (1966) applied this model to analyze survival data in animal tumor systems and acute leukemia. The Weibull model has been used in several applications in engineering, industry and cancer research (Bain, 1978; Laird, 1965). The Gompertz model for $\gamma(x, \beta)$ has been used by Garg et al. (1970), who have studied its properties and obtained maximum likelihood estimates of its parameters. These distributions can, of course, be generalized by replacing x by $x-b$, where b represents, what is called the 'guarantee' time in industry, where no item can fail before b units of time have elapsed. In epidemiological or biomedical applications, the parameter b might

represent, the 'latent' period of some disease. This period may be simply defined as the time elapsed between first exposure to an agent and the appearance of a symptom.

The exponential distribution, apart from having several useful applications in reliability theory and theoretical and applied statistics, has a remarkable property, that is, the distribution remains unchanged under the following three operations (i) truncating the original distribution at some point $x_0 \geq 0$, (ii) considering the observable distribution for life times $X \geq x_0$ and (iii) changing the origin by means of the transformation given by $X_1 = X - x_0$, so that $X_1 \geq 0$ i.e., setting the clock back to zero.

3 'Setting the Clock Back to Zero' Property and the PRL Function

An alternative definition: A family of life distributions $\{S(x, \beta), x \geq 0, \beta \in \Omega\}$ is said to have the "setting the clock back to zero" property or be 'invariant' (Rao and Talwalker, 1990), if for each $\beta \in \Omega$ and $t_0 \geq 0$, the survival function satisfies the equation

$$\frac{S(x + t_0, \beta)}{S(t_0, \beta)} = S(x, \beta^*), \quad \text{with } \beta = \beta^*(t_0) \in \Omega. \quad (3-1)$$

Here β may be a single parameter or a vector of parameters and Ω is the parametric space. Also β^* is the new parameter (or vector of parameters) as a function of t_0 . Equation (3-1) can be alternatively stated as

$$P(X \geq x + t_0 \mid X \geq t_0) = P(X^* \geq x), \quad (3-2)$$

where the random variable X^* has the same distribution as that of X , except that the vector of parameters β is replaced by the vector β^* .

This invariance property means that the conditional distribution of additional length of survival of an organism, given that it has survived t_0 time units, remains in the family. Clearly this property generalizes the lack of memory property of the exponential distribution in the sense that for the exponential distribution, the conditional distribution of additional survival time is exactly the same as the original distribution. To see this, note that if $\beta = \beta^*$ in equation (3-1), we get the functional equation

$$S(x + t_0, \beta) = S(t_0, \beta)S(x, \beta^*) \quad (3-3)$$

which implies that $S(x, \beta)$ is an exponential function (Karlin, 1966). Setting the clock back to zero property may be described as 'the next best thing' (or is it 'the next worst thing?') to the lack of memory property of the exponential distribution. The main result of the present paper is contained in the following theorem.

Theorem (3.1): Suppose that a family of life distributions $\{S(x, \beta), x \geq 0, \beta \in \Omega\}$ has the setting the clock back to zero property. Then its $100(1-\alpha)$ PRL function is its $100(1-\alpha)$ percentile, except perhaps with a different vector $\beta^* \in \Omega$ of parameters.

Proof: The PRL function for such a family of life distributions will be denoted by the symbol $R(t_0, \alpha, \beta)$ to show its dependence on the vector $\beta \in \Omega$. This is a solution of the equation

$$P[X \geq t_0 + R(t_0, \alpha) | X \geq t_0] = 1 - \alpha,$$

i.e.,

$$\frac{S[t_0 + R(t_0, \alpha, \beta), \beta]}{S(t_0, \beta)} = 1 - \alpha. \quad (3-4)$$

But since the family of life distributions has the setting the clock back to zero property, this equation becomes

$$S[R(t_0, \alpha, \beta^*), \beta] = 1 - \alpha, \quad \beta^* \in \Omega.$$

whose solution is uniquely determined as

$$R(t_0, \alpha, \beta^*) = S^{-1}(1 - \alpha). \quad (3-5)$$

Thus equation (3-5) shows that if a family of life distributions has the setting the clock back to zero property, then its PRL function is just the $100(1-\alpha)$ percentile of the family, except perhaps with a different vector of parameters, $\beta^* \in \Omega$. The converse of this theorem is easy to prove. The case of the exponential distribution is discussed in equation (1-8).

Example (1): The linear hazard exponential distribution, see eq. (2-8) has the setting the clock back to zero property (Kodlin, 1966), since its survival function is

$$S(x, \beta) = \exp\left[-\left(ax + \frac{b}{2}x^2\right)\right], \quad \beta = (a, b) \in \Omega$$

and

$$\frac{S(x + t_0, \beta)}{S(t_0, \beta)} = \exp\left[-\left(a'x + \frac{b}{2}x^2\right)\right], \quad \beta = (a', b) \in \Omega$$

where $a' = a + bt_0$. Its PRL function is simply the positive root of the quadratic equation

$$a'x + \frac{b}{2}x^2 = -\ln(1 - \alpha).$$

Example (2): More generally, the general Krane family of distributions (Krane, 1963; Rao, 1990) given by $\{S(x, \beta), x \geq 0, \beta \in \Omega\}$ where $S(x, \beta) = \exp(a_0x + a_1x^2 + \dots + a_mx^{m+1})$ $\beta = (a_0, a_1, \dots, a_m) \in \Omega$ (equation 2-9), has the setting the clock back to zero property and its PRL is simply the positive root of the equation (3-5).

Example (3): The PRL for the family of Pareto distributions $\{S(x, \beta), x \geq K, \beta \in \Omega\}$ can be obtained easily. Here

$$S(x, \beta) = \frac{K^a}{x^a}, \quad x \geq K, \quad \beta = (a, K)$$

Also for $t_0 \geq k$, we have

$$P(X \geq R + t_0 \mid X \geq t_0) = \frac{S(t_0 + R)}{S(t_0)} = \left(1 + \frac{R}{t_0}\right)^{-a} = 1 - \alpha$$

so that

$$R(t_0, \alpha) = R = t_0[(1 - \alpha)^{-1/a} - 1].$$

Observe that the PRL function is a linear function of t_0 .

Example (4): The family of Weibull distributions $\{S(x, \beta), x \geq 0, \beta \in \Omega\}$, where $S(x, \beta)$ is its survival function

$$S(x, \beta) = \exp(-x^p), \quad x \geq 0, \quad p > 0,$$

does not have the setting the clock back to zero property, since the ratio

$$\frac{S(x + x_0, \beta)}{S(x_0, \beta)} = \exp[-(x + x_0)^p + x_0^p]$$

$$\frac{S(x + x_0, \beta)}{S(x_0, \beta)} \neq S(x, \beta^*)$$

Its PRL function is

$$R(x_0, \alpha) = \{ -[\ln(1 - \alpha) + x^p] \}^{1/p} - x_0$$

and by theorem (3.1), this is not the $100(1 - \alpha)$ percentile of the original distribution.

Example (5): The family of Gompertz distributions $\{S(x, \beta), x \geq 0, \beta = (\tau, \theta)\}$. The parametric space is $\Omega = \{(\tau, \theta): 0 \leq \tau \leq \infty, 0 \leq \theta \leq \infty\}$. This family has the setting the clock back to zero property, since the ratio

$$\frac{S(x + t_0, \beta)}{S(t_0, \beta)} = \exp\left[-\frac{\tau'}{\theta}(1 - e^{-\theta x})\right] = S(x, \beta^*)$$

where $\beta^* = (\tau, \theta)$ and $\tau' = \tau e^{-\theta t_0}$. Clearly $\beta^* \in \Omega$. Its PRL function is simply a root of the equation

$$S(x, \beta^*) = 1 - \alpha$$

i.e.

$$R(t_0, \alpha, \beta) = S^{-1}(1 - \alpha) = \frac{-1}{\theta} \ln\left[1 + \frac{\theta}{\tau} e^{\theta t_0} \ln(1 - \alpha)\right].$$

Observe that $R(t_0, \alpha, \beta)$ is a decreasing function of t_0 . In this operation of setting the clock back to zero, the parameter e remains unchanged and the parameter τ becomes τ' . In view of this property, truncating a Gompertz distribution at time X_0 and setting the origin at t_0 (i.e., in terms of time, "Setting the clock back to zero") leaves the form of the distribution unaltered except for the value of t , which changes from τ to τ' .

Laird (1965 and 1969) have utilized this property of the Gompertzian growth process and have presented several applications in the dynamics of normal, embryonic and tumor growth in animals and mammals. The parameter e , which does not undergo any change under this 'setting the clock back to zero' transformation, has been called the normalizing constant. Laird (1965 and 1969) remark that the specific growth rate for many organisms decays approximately exponentially as the animal ages, and derive the Gompertzian growth equation, supplemented by an arithmetic growth curve which fits the accretionary growth displayed by many animals at early maturity. In these considerations biological time is seen to be exponential process rather than the linear process one customarily assumes.

4 An Application to Cancer Risk Assessment: Survival/Sacrifice Experimental Study

Let us now apply the model to study the time to tumor in a serial sacrifice experiment conducted by the National Center for Toxicological Research (NCTR). The basic variable, in this application, is not the life time of an individual but the time to tumor (Farmer et al, 1980) and Littlefield et al, 1980). The NCTR conducted a large scale experiment on 24,000 female mice that from infancy were continuously fed 2-AAF (acetyarninofluorence). There were seven dose levels ranging from 30 to 150 ppm of 2-AAF, plus a control. The mice were sacrificed from 9 to 33 months from the initiation of feeding. For each group of mice sacrificed, the number that had developed bladder neoplasms and/or liver neoplasms were determined. Few mice were found to

have developed tumor at low levels of 2-AAF. Enough bladder tumor cases were found only at the highest dose level, to ensure the reliability of the incidence rates. Thus only the data on bladder cancer incidence for the dose level of 150 ppm of 2-AAF are used for illustration of the method of analysis.

When an individual is exposed to toxic materials, his body's metabolic and other biological reactions render most of the absorbed molecules inactive, but a few metabolites remain. The survival of the individual is directly affected by the cumulative effect of the interaction between exposure to toxins and biological reaction. The force of mortality or the hazard rate at a particular age is a function of the level of toxic material and the length of exposure. Chiang and Conforti (1989) introduced a stochastic model in which the mortality intensity, or the hazard rate is a function of the accumulated effect of an individual's continuous exposure to toxic material in the environment (absorbing coefficient) and his biological reaction to the toxin absorbed (discharging coefficient). Some special cases are also considered where there is a change in exposure level or exposure is discontinued or exposure is discrete in time. The random variable is the time to tumor.

A simple model is proposed for the force of mortality or the instantaneous failure rate

$$\lambda(x) = \mu(a+x) + \lambda(x, \delta, v), \quad (4-1)$$

where the internal factor $\mu(a+x)$ is a function of age $a+x$ and the external factor $\lambda(x, \delta, v)$ is a function of x, δ, v , derived in their paper. The external factor is assumed to be a function of the accumulated dose. It is assumed that the external factor dominates the mortality intensity.

The total amount of chemical absorbed during the time interval $(0, x)$ that will be potentially carcinogenic at time x is represented by the quantity $(\delta/v)(1 - e^{-vx})$, where δ is the absorbing coefficient corresponding to the dose level and v is the discharging coefficient reflecting the intensity of the metabolic activity. Accounting for a latent period of x_0 months, this amount is divided into two components:

$$\frac{\delta}{v}(1 - e^{-vx}) = \frac{\delta}{v}(e^{-vx_0} - e^{-vx}) + \frac{\delta}{v}(1 - e^{-vx_0}),$$

the first component being the amount of chemical absorbed during the interval $(0, x-x_0)$ that may be carcinogenic at time x and the second being the amount absorbed during $(x-x_0, x)$ that is potentially carcinogenic at x . The tumor incidence intensity at time x is assumed as a function of the first component only, that is

$$\lambda(x, \delta, v) = b \frac{\delta}{v} e^{-vx_0} (1 - e^{-v(x-x_0)})$$

$$\lambda(x, \delta, v) = \frac{\beta}{v} e^{-vx_0} (1 - e^{-v(x-x_0)}), \quad x \geq x_0,$$

The cumulative hazard function is then

$$\Lambda(x, x_0, \delta, \nu) = \int_{x_0}^x \lambda(t, \delta, \nu) dt$$

$$\Lambda(x, x_0, \delta, \nu) = \frac{\beta'}{\nu} \left[(x - x_0) - \frac{1}{\nu} (1 - e^{-\nu(x-x_0)}) \right], \quad \beta' = \beta e^{-\nu x_0}$$

and the survival function of the time to turnor is

$$s(x, \beta, \nu) = \exp \left\{ -\frac{\beta'}{\nu} \left[(x - x_0) - \frac{1}{\nu} (1 - e^{-\nu(x-x_0)}) \right] \right\}, \quad x \geq x_0. \quad (4-2)$$

In this paper, we shall propose a slightly more general form for the force of mortality function and show that the generalized family of life distributions has the 'setting the clock back to zero' property. As a simple application of this result, the PRL function is derived for such a family. Some further results are also discussed. In the present section, we shall propose a more general form of the hazard rate or the force of mortality function given by

$$\lambda(x, \alpha, \beta, \nu) = \frac{\beta'}{\nu} (1 - \phi e^{-\nu x}), \quad \phi > 0 \quad (4-3)$$

so that $\phi=1$ would give the Chiang-Conforti model. In equation (4-3), it is understood that the hazard rate $\lambda(x, \phi, \beta', \nu) \geq 0$. Since $\phi > 0$, $\beta' > 0$ and $\nu > 0$, this is equivalent to $1 - \phi e^{-\nu x} \geq 0$ or $x \geq x_1$, where

$$x_1 = \frac{1}{\nu} \ln \phi$$

Then $\lambda(x, \phi, \beta', \nu) \geq 0$ for $x \geq x_1$. But if $0 < \phi \leq 1$, then the hazard rate $\lambda(x, \phi, \beta', \nu)$ is non-negative for $x \geq 0$.

The vector of parameters is $\theta = (\phi, \beta', \nu)$ and the parametric space is $\Omega = \{\theta = (\phi, \beta', \nu): \phi > 0, \beta' > 0, \nu > 0\}$. When $\theta \in \Omega$, the survival function corresponding to the more general form given in eq(4-3) is

$$S(x, \theta) = \exp \left\{ -\frac{\beta'}{\nu} \left[x - \frac{\phi}{\nu} (1 - e^{-\nu x}) \right] \right\} \quad (4-4)$$

We shall demonstrate now that the family of survival distributions described in eq (4-4) has the 'setting the clock back to zero' property and that the PRL at any given time X_0 can be obtained in a simple manner. From eq. (4-4), the probability of survival to age to is

$$S(t_0, \theta) = \exp \left\{ -\frac{\beta'}{\nu} \left[t_0 - \frac{\phi}{\nu} (1 - e^{-\nu t_0}) \right] \right\} \quad (4-5)$$

We also find the ratio (see our equation [3-1])

$$\begin{aligned}\frac{S(x+t_0, \theta)}{S(t_0, \theta)} &= \exp \left\{ -\frac{\beta'}{\nu} \left[x - \frac{\phi e^{-\nu t_0}}{\nu} (1 - e^{-\nu x}) \right] \right\} \\ \frac{S(x+t_0, \theta)}{S(t_0, \theta)} &= \exp \left\{ -\frac{\beta'}{\nu} \left[x - \frac{\phi'}{\nu} (1 - e^{-\nu x}) \right] \right\} \\ \frac{S(x+t_0, \theta)}{S(t_0, \theta)} &= S(x, \theta^*)\end{aligned}\tag{4-6}$$

where $\phi' = \phi e^{-\nu t_0}$. In the notation of our equation (3-1), the vector of parameters $\theta^* = (\phi', \beta', \nu) \in \Omega$. This invariance property implies that the conditional distribution of additional length of survival of the organism, given that it has survived to time units, remains in the family. Thus the family of survival distributions described in equation (4-3) has the 'setting the clock back to zero' property.

The PRL function of a group of organisms at age t_0 is the $100(1-\alpha)$ percentile of the remaining life of the organisms. In other words, the PRL function $R(t_0, \alpha)$ shows how long an organism of age t_0 would survive with a high probability $(1-\alpha)$. In our present application of time to tumor incidence, this would mean that if no tumor has been found until time t_0 , $100(1-\alpha)$ percent of the organisms can go on for another $R(t_0, \alpha)$ time units of tumor-free life. This is simply the unique root of the equation

$$S[R(t_0, \alpha), \theta^*] = 1 - \alpha.$$

From equation (4-6), writing R for $R(t_0, \alpha, \theta^*)$ for simplicity, we get the equation

$$\exp \left\{ -\frac{\beta'}{\nu} \left[R - \frac{\phi'}{\nu} (1 - e^{-\nu R}) \right] \right\} = 1 - \alpha.\tag{4-7}$$

Taking natural logarithms of both sides, it will be seen that R is the unique root of the equation

$$R - \frac{\phi'}{\nu} (1 - e^{-\nu R}) = -\frac{\nu}{\beta'} \ln(1 - \alpha), \quad \phi' = \phi e^{-\nu t_0}.\tag{4-8}$$

Table 3 tabulates the values of the PRL function $R = R(t_0, \alpha, \theta^*)$ for some values of t_0 , a , β' and ν .

4.1 Estimation of the Parameters and the Survival Function

The estimates β and ν and their sample variances are derived by an iterative method in the Statistical Analysis System (SAS). Using β and ν we found an estimate of the survival function

$S(x, \beta, \nu)$. Table 1 summarizes the estimates of the NCIR data and table 2 shows some selected percentiles of the distribution of time to tumor for bladder neoplasms for mice in the NCTR data. The percentile, x_a as given in Chiang and Conforti (1989) is a solution of the equation, (see equation 4-2) :

$$P(X \geq x_0) = S(x_0, \beta, \nu) = 1 - \alpha$$

$$i.e., \exp \left\{ -\frac{\beta}{\nu} \left[(x_a - t_0) - \frac{1}{\nu} (1 - e^{-\nu(x - t_0)}) \right] \right\} = 1 - \alpha .$$

Table 1: Application of the Model to the NCTR Serial Data -A Summary^a

1-Total number of dead, morbund and sacrificed mice that were necropsied	N=1089
2-Total number of mice diagnosed with tumor	N=496
3- Latent period (0, x_0)	X_0 = 10 months ^b
4- Maximum lifetime	x = 34 months
5- Estimates of the parameters	β =0.051, SE(β)=0.011 ν =0.404, SE(ν)=0.130
6- Goodness of fit	χ^2 =5.44, $\Pr(\chi^2 \geq 5.44) = 0.365$
7- Time to tumor appearance, X	
Mean	X=19.51 months
Standard deviation	S=6.89 months
Median	Md=17.89

^a Bladder neoplasms in mice continuously fed 2-AAF at 150 ppm.

^b Time since initiation of feeding of 2-AAF.

Table 2: Selected Percentiles of the Distribution of Time to Tumor for Bladder Neoplasms *for* mice continuously Fed 2-AAF at 150 ppm

Distribution function at x_α ; $F(x_\alpha)$	Percentile (months ^a) x_α	PRL (months) $R(x_0, \alpha)$
0.05	11.57	0.41
0.10	12.36	0.85
0.25	14.34	2.31
0.50	17.89	5.53
0.75	23.51	11.02
0.90	30.81	18.28

^a Time since the initiation of feeding 2-AAF.

That is, x_α is an upper $100(1-\alpha)$ percentile of the original population of mice. In other words, $100(1-\alpha)$ percent of all mice can go on beyond the latent period of $t_0=10$ months tumor-free. x_α is not the $100(1-\alpha)$ percentile of the remaining life RPL of the mice. As the authors mention, "some mice died before their sacrifice times, and others were moribund and sacrificed before their sacrifice times. These mice were also necropsied for bladder and liver tumors". In other words, some mice developed bladder and liver neoplasms even before the $t_0=10$ month latent period. The data on these mice "are included with the group of mice sacrificed nearest to their death, (Farmer et al, 1980; Littlefield et al, 1980).

Figure 1. Relationship between PRL function and proportions surviving.

The PRL function $R(t_0, \alpha)$, on the other hand, is defined by

$$P\{X \geq t_0 + R(X_0, \alpha) \mid X \geq t_0\} = 1 - \alpha.$$

This shows that a proportion $(1 - \alpha)$ of the t_0 -month-old mice can go on for another R months tumour-free. Values of $R(t_0, \alpha, \beta^*, v^*)$ are shown in Figure 1.

5 Conclusions, Some Closure Theorems and the PRL Function

The Lack of Memory Property of an exponential distribution is well known. As remarked previously, setting the clock back to zero property of a family of life distributions may be described as the 'next-best thing' (or is it the 'next-worst thing?') to the lack of memory property and is shared not only by the exponential distribution, but also by a few other families of life distributions. In this section, we shall discuss some closure theorems, which mimic the following useful property of an exponential distribution, that is,

Theorem (5.1): Suppose that X_1 and X_2 are independent random variables having exponential distributions with the survival functions

$$S_1(x, \theta_1) = e^{-\theta_1 x}, \quad S_2(x, \theta_2) = e^{-\theta_2 x}$$

$x \geq 0$, $\theta_1 > 0$ and $\theta_2 > 0$. Then the random variable $Z = \min(X_1, X_2)$ also has an exponential distribution, with the survival function

$$S(x, \theta_1, \theta_2) = e^{-(\theta_1 + \theta_2)x}.$$

The theorem (5.1) extends itself to any number of random variables $X_1, X_2, \dots, X_n$. In competing risk theory, if $X_1, X_2, \dots, X_n$ are the hypothetical life times of n competing risks acting on a family of organisms, then $Z = \min(X_1, X_2, \dots, X_n)$ is the actually observed life time of an organism. The survival function of the observed life length Z is the product

$$S(z, \theta_1, \theta_2, \dots, \theta_n) = P(Z \geq z) = e^{-z \left(\sum_{i=1}^n \theta_i \right)}, \quad z \geq 0. \quad (5-1)$$

As remarked in the Introduction, see equations (1-7) and (1-8), suppose that an organism has survived z_0 units of time under the risk of death due to n diseases or causes. The i th risk has the PRL function

$$R_i(z_0, \alpha_i) = -\frac{1}{\theta_i} \ln(1 - \alpha_i), \quad i = 1, 2, \dots, n. \quad (5-2)$$

This is, of course, the $100(1 - \alpha)$ percentile of the i th risk. What is the PRL function of the observed life length Z ? Since Z has an exponential distribution with the survival function (5-1), its PRL function $R(Z_0, \alpha)$ satisfies

$$e^{-\left(\sum_{i=1}^n \theta_i\right) R(z_0, \beta)} = \prod_{i=1}^n (1 - \alpha_i) .$$

This gives the PRL function

$$R(z_0, \alpha) = - \frac{\sum_{i=1}^n \ln(1 - \alpha_i)}{\sum_{i=1}^n \theta_i} . \quad (5-3)$$

$R(Z_0, \alpha)$ is the $100(1-\alpha)$ PRL function of Z , where α is given by $\alpha = 1 - \prod (1-\alpha_i)$.

We can show that similar results hold with respect to the setting the clock back to zero. The following theorems give some closure properties of the setting the clock back to zero property, as applied to competing risks.

Theorem (5-2): Let X and Y denote the life times of an organism due to risks 1 and 2 respectively. Let $U = \min(X, Y)$ denote the actual (observed) life time of the organism. If the survival distributions of the life times X and Y have the, 'setting the clock back to zero' property, so does the survival distribution of the life time U , under the assumption of independent risks.

Proof: Let the hazard rates of the two risks of death be denoted by $\lambda_1(x, \beta)$ and $\lambda_2(x, \nu)$ so that the individual survival functions are

$$S_1(x, \beta) = \exp\left[-\int_0^x \lambda_1(u, \beta) du\right] = e^{-H_1(x, \beta)}$$

$$S_2(x, \nu) = \exp\left[-\int_0^x \lambda_2(u, \nu) du\right] = e^{-H_2(x, \nu)}$$

where $H_1(x, \beta)$ and $H_2(x, \nu)$ are the cumulative hazard functions for the two risks of death, where $\beta \in \Omega_1$ and $\nu \in \Omega_2$. Here $\sim$ and u are parameters or vectors of parameters. We are given that these families of survival functions have the setting the clock back to zero property, that is,

$$\frac{S_1(x + x_0, \beta)}{S_1(x_0)} = S_1(x, \beta^*)$$

and

$$\frac{S_2(x + x_0, \nu)}{S_2(x_0)} = S_2(x, \nu^*)$$

where $\beta^* = \beta^*(x_0) \in \Omega_1$ and $\nu^* = \nu^*(x_0) \in \Omega_2$. This implies that

$$H_1(x + x_0, \beta) - H_1(x_0, \beta) = H_1(x, \beta^*)$$

$$H_2(x + x_0, \nu) - H_2(x_0, \nu) = H_2(x, \nu^*)$$

The actual (observed) life time of the organism is $U = \min(X, Y)$. Under the assumption of independent risks of death, the survival function of the life time U is the product

$$S(x, \theta) = S_1(x, \beta)S_2(x, \nu).$$

Here $\theta = (\beta, \nu)$ is the vector of parameters for the distribution of the observed life length U . The parametric space is the cartesian product $\Omega = \Omega_1 \times \Omega_2$. If we now form the ratio

$$\frac{S(x + x_0, \theta)}{S(x_0, \theta)} = \frac{S_1(x + x_0, \beta)S_2(x + x_0, \nu)}{S_1(x_0, \beta)S_2(x_0, \nu)}$$

$$\frac{S(x + x_0, \theta)}{S(x_0, \theta)} = \exp\{-[H_1(x, \beta^*) + H_2(x, \nu^*)]\}$$

$$\frac{S(x + x_0, \theta)}{S(x_0, \theta)} = S_1(x, \beta^*)S_2(x, \nu^*) = S(x, \theta^*), \quad \theta^* = (\beta^*, \nu^*) \in \Omega.$$

This shows that the family of life distributions of the actual (observed) life length of an organism also has the setting the clock back to zero property. Observe that the hazard rate for the observed life time U is given by the sum

$$\lambda(x, \beta, \nu) = \lambda_1(x, \beta) + \lambda_2(x, \nu).$$

Remark. Incidentally, it will be seen that the theorem is true for the more general situation in which

$$\lambda(x, \beta, \nu) = \lambda_1(x, \beta) \pm \lambda_2(x, \nu).$$

Of course, it is assumed that $\lambda_1(x, \beta)$ and $\lambda_2(x, \nu)$ are such that $\lambda(x, \beta, \nu) \geq 0$, $x \geq x_1 > 0$. It is also seen that $\lambda(x, \beta, \nu)$ is the hazard rate of some random variable.

The Converse of Theorem (5-2) : That the converse of the theorem is not true is seen from the following example. Let X and Y be two independent life time variables with the hazard functions

$$\lambda_1(x, \beta) = a_0 + a_1x, \quad \lambda_2(x, \nu) = a_2x^2.$$

Observe that these are the hazard rates respectively of a linear hazard exponential distribution and a Weibull distribution. The former family has, and the latter does not have, the setting the clock back to zero property. But $U=\min(X,Y)$ has the hazard rate

$$\lambda(x, \beta, \nu, \mu) = a_0 + a_1x + a_2x^2 + a_3x^3.$$

This is the hazard rate of the General Krane Family. A similar result holds for the sum or difference of any number of hazard rates. As another simple application of the theorem in the more general situation, let us choose

$$\lambda_1(x, \beta, \nu) = \frac{\beta}{\nu}, \quad \lambda_2(x, \beta, \nu) = \frac{\alpha\beta}{\nu} e^{-\nu x}$$

so that their difference becomes the Chiang-Conforti hazard rate:

$$\lambda(x, \beta, \nu) = \frac{\beta}{\nu} (1 - \alpha e^{-\nu x}).$$

Observe that $\lambda_1(x, \beta, \nu)$ is the hazard rate of an exponential distribution and $\lambda_2(x, \beta, \nu)$ is the hazard rate of a Gompertz distribution. These families are shown to have the setting the clock back to zero property. As a further example, let us choose

$$\lambda_1(x, \beta, \nu) = \frac{\beta}{\nu} e^{-\nu x}, \quad \lambda_2(x, \beta, \mu) = \frac{\beta}{\mu} e^{-\mu x}.$$

If $\beta > 0$, $\mu > 0$, and $\nu > 0$, these are respectively the hazard rates of two Gompertzian processes which are shown to have the setting the clock back to zero property. Their difference is

$$\lambda(x, \beta, \mu, \nu) = \lambda_1(x, \beta, \nu) - \lambda_2(x, \beta, \mu)$$

$$\lambda(x, \beta, \mu, \nu) = \frac{\beta}{\nu} (e^{-\nu x} - \alpha e^{-\mu x}) \quad (5-4)$$

where the new parameter α is

$$\alpha = \frac{\nu}{\mu}.$$

Clearly $\alpha > 0$. As before, the hazard rate in equation (5-4) is positive for those values of x satisfying $x > x_1$, where $x_1 = \log \alpha / (\mu - \nu)$, $\mu > \nu$.

It again follows from our theorem that the family of survival distributions described by the hazard rate in equation (5-4) has the setting the clock back to zero property. Its PRL can be found. In this paper, it can be shown that it is not necessary to assume that the external factor

dominates the mortality intensity. Instead, we shall use the general model (4.1) coupled with (4.3). In other words, we shall use the form where the internal factor $\mu(a+t)$ has a polynomial form of some degree, say

$$\mu(a+t) = a + a_1t + a_2t^2 + \dots + a_mt^m.$$

Then clearly the general family has a hazard rate of the form

$$\lambda(t, \theta) = a + a_1t + a_2t^2 + \dots + a_mt^m + \frac{\beta}{\nu} [1 - \alpha e^{-\nu t}]$$

and this family has the setting the clock back to zero property, by our Theorem (5-2) where the vector pf parameters is $\theta = (a, a_1, \dots, a_m, \alpha, \beta, \nu)$.

References

- [1] Bain, L.J. Statistical Analysis of Reliability and Life-Testing Models. Holt, Rinehart and Winston, New York, 1978.
- [2] Barabas, B., Csorgo, M., HorVath, L and Yandell, B.S. Bootstrapped Confidence Bands for Percentile Lifetime. Ann. Inst. Statistics Mathematics **38 (A)** (1986), 429-438.
- [3] Barlow, R.E. and Proschan, F. Statistical Theory of Reliability and Life Testing, Holt, Rinehart and Winston, New York, 1975.
- [4] Chiang, C.L. Introduction to Stochastic Processes in Biostatistics, Wiley, New York, 1968.
- [5] Chiang, C.L. and P.M. Conforti (1989). A Survival Model and Estimation of Time to Tumor. Mathematical Biosciences, **94** (1989), 1-29.
- [6] Chung, C.F. Confidence Bands for Percentile Residual Lifetime under Random censoring Model. Journal of Multivariate Analysis, **29** (1989), 94-126.
- [7] Csorgo, M and Csorgo, S. Estimation of Percentile Residual Life. Operations Research, **35 (4)** (1987), 598-605.
- [8] Epstein, B. and Sobel, M. Life Testing. Journal American Statistics Assn, **48** (1953), 486-502.

- [9] Farmer, J.H., Kodell, R.L., Greenman, D.L., and shaw, G.W. Dose and time response models for the incidence of bladder and liver neoplasms in mice fed 2-AAF continuously. *Journal of Environmental Pathology & Toxicology*, **3** (1980), 55-67.
- [10] Feng, Z and Kulasekera, K.B. Nonparametric Estimation of the Percentile Residual Life Function, *Communications in Statistics* **20** (1) (1991), 87-105.
- [11] Garg, M.L., Rao, B. Raja and Redmond, CK. Maximum Likelihood Estimation of the Parameters of the Gompertz Survival Function. *Journal of Royal Statistics Society C (Applied Statistics)*, **19** (1970), 152-159.
- [12] Haines, A.L. and Singpurwalla, N.D. Some Contributions to the Stochastic Characterization of Wear. *Reliability and Biometry: Statistical Analysis of Life Length* (eds. F. Proschan and R.F. Serfling), SIAM, Philadelphia, 1984.
- [13] Karlin, S. *A First Course in Stochastic Processes*, Academic Press, New York, 1966.
- [14] Laird, A.K. Dynamics of Normal Growth, *Growth*, **29** (1965), pp 233-248.
- [15] Laird, A.K. Dynamics of growth in tumors and in normal organisms. *Natl. Cancer Inst. Monogr*, (1969), 15-28.
- [16] Littlefield, N.A., Farmer, I.H., Gaylor, D.W. and Sheldon, W. Effects of dose and time in a long-term low-dose carcinogenic study. *Journal of Environmental Pathology & Toxicology*, **3** (1980), 17-34.
- [17] Rao, B. R. and Talwalker, S. Setting the clock back to zero property of a class of life distributions. *The Journal of Statistical Planning and Inference*, **24** (1990), 347-352.
- [18] Rao, B. R. A simple time to tumor distribution and setting the clock back to zero propeny. *Communications in Statistics, Theory Methods* **19** (2) (1990), 583-598.
- [19] Schmittlein, D.e. and Monison, D.G. The Median Residual Lifetime: A characterization Theorem and an Application. *Operation Research*, **29** (2) (1981), 392-399.
- [20] Zelen, M. Application of Exponential Model to Problems in Cancer Research, *Journal of Royal Statistics Society, A*, **129** (1966), 368-398.

Received: November 6, 2005