Survival Probability Modeling

Question: In the self-help group I am in, we have a lot of "old-timers" (a term that not all of them appreciate), that is, people who have been coming to meetings for a long time. Having been going for a while myself, and observing the comings and goings, it seems to me that the longer a person is in a group, the more likely it is that they will stay in the group. Is there a scientific reason for this? Is there a way that this observation can be used to make predictions about the number of members that will be around in the future?

Answer: Your question sounds like it is concerned with predicting the "survival" of members over time. In particular, it sounds like you'd like to know the probability of someone sticking around based on some sort of "first principles" (scientific) foundation.

Prompted by your question, I looked into survival modeling. I discovered some remarkable results that I'd like to share with the community, hence the long and involved reply. To avoid scaring people away, I have put the mathematical details into Appendices.

Survival probability modeling is aimed at predicting the proportion of a population that is still intact, after a certain time, when it is subjected to attrition. Examples include student graduation rates, the lifetime of light bulbs and Weight Watchers.

A fundamental model for the random occurrence of events is the so-called Poisson distribution which describes the probability of n events occurring in a given time. The rate at which events occur (arrival rate), λ, is usually assumed to be constant so that we have, for instance, a certain average number of light bulbs failing in a month. The model for this case takes the form of an exponential (Appendix A).

The probability model developed here is based on the simple assumption that the tendency to drop out, i.e., the *failure rate*, is proportional to how many people have survived in the group so far. Since the number of survivors goes down in time (not counting the addition of new members, of course), this means that the failure rate, or tendency to quit, goes down as well. The goal here has been to construct a model that incorporates this simple assumption.

Many real-world applications can be modeled effectively by assuming that λ is a *constant*, such as the emission of particles due to the decay of atomic nuclei. However, there are situations where a *non-constant* failure rate may be

Copyright 2014 Randy Patton

appropriate, such as in the present case. Although one way to tackle this is to incorporate multiple parameters[1] into various distributions, in order to fit the model to empirical data, the current approach relies on a more fundamental, "first principles" model which does not introduce extra parameters.

A big surprise was that the derived model (Appendix A) turns out to be identical to a Pareto[2] distribution. The Pareto distribution was originally derived to describe the distribution of wealth in a society. It provides the simplified mathematical basis for the frequently used "80-20 rule", one example of which is that 80% of the wealth of a society is owned by 20% of the people. There are many other such examples of 80-20 distributions across many applications to support this almost mythic rule-of-thumb [3].

The length of membership in a group, the distribution of wealth in a society and many other processes can, remarkably, be considered in terms survival rates.

Survival rates, and their more moribund but equivalent corollary, failure rates, involve probabilities of the occurrence of events. A classic example is the lifetime of light bulbs. Since light bulbs last only a limited amount of time, a natural feature of interest is the probability of one burning out in the next hour, day, month, etc. In this example, "burning out" is the event of interest and the distribution of bulb lifetimes something that would be nice to know.

Figure 1 shows the probability of *failure* for the Pareto model (decreasing failure rate), along with the exponential model (constant failure rate). The Pareto distribution shows a slower failure probability with time, i.e., old timers tend to stick around longer. The dashed line indicates the "80-20" ratio typically associated with the Pareto distribution (the 20 on the time axis is fortuitous, but you get the idea).

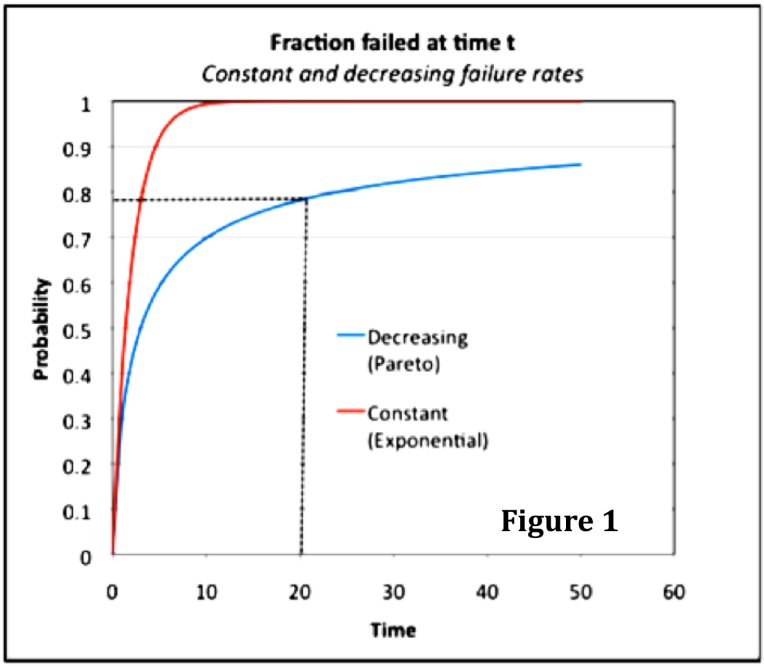


Figure 1

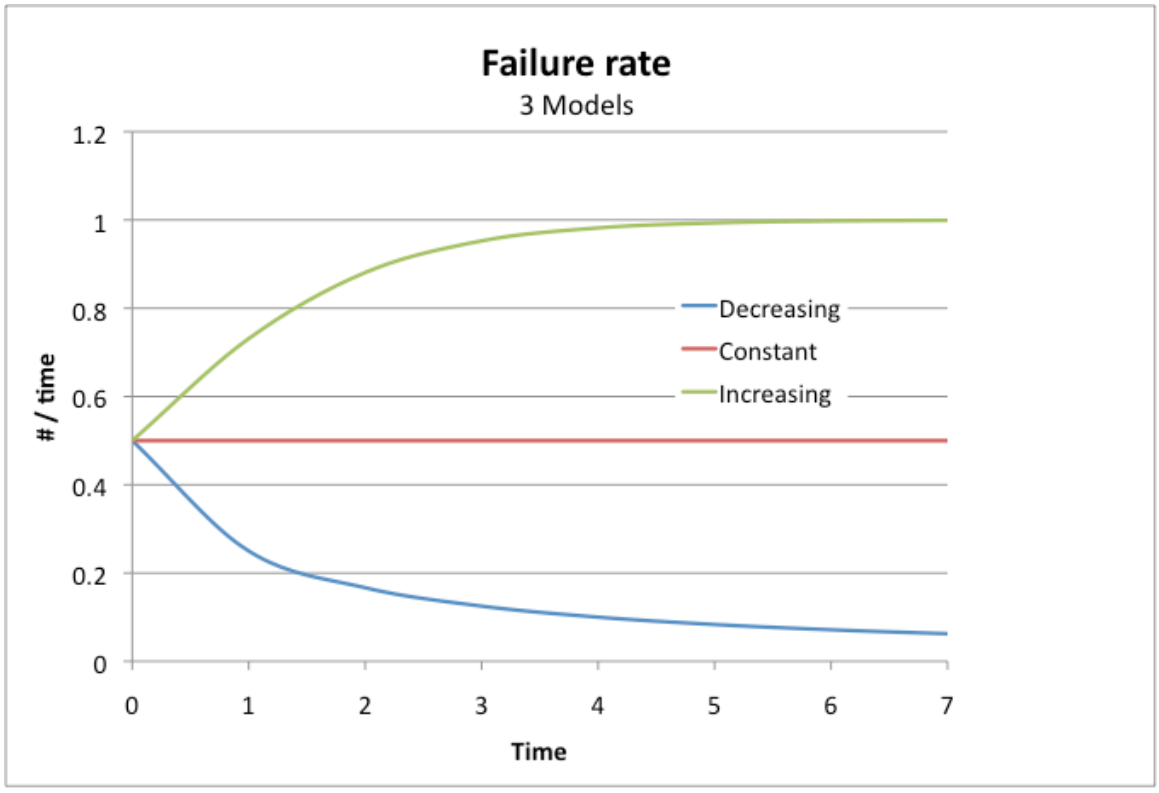
[1] such as the Weibull or Pareto Type III and IV distributions
[2] *Wilfried Fritz Pareto*; 1848 - 1923) http://en.wikipedia.org/wiki/Pareto_principle
[3] http://management.about.com/cs/generalmanagement/a/Pareto081202.htm

Increasing Failure Rate: In the spirit of the above derivation, it seemed natural to ask what sort of distribution would ensue if the failure rate λ *increases* with time.

Appendix B provides a derivation of the probability of survival when the failure rate is proportional to the probability of *failure* instead of being proportional to the probability of *survival* (as in the first model). This means that the failure rate, λ, increases with time.

The results of the derivation show that this new model is *not* equivalent to a Pareto distribution. The derivation, however, revealed another surprise, namely that the failure rate is described by the *logistics function*[4] often used in modeling changing populations.

Comparison of Models: Figure 2 shows a comparison of the failure rates for the 3 models: constant, decreasing and increasing.



[4] http://en.wikipedia.org/wiki/Logistic_function

Copyright 2014 Randy Patton

Figures 3 shows the probabilities of failure (dropping out) after a time *t*, for the 3 models. The curves for the constant and decreasing failure cases are as before (figure 1, rescaled). The curve for the *increasing* failure rate model approaches 1 much more quickly than the other two, which again is expected since this case models what happens if the tendency to quit increases as more members are seen quitting. This can be called the "rats leaving the sinking ship" scenario. It is also reminiscent of panic in the stock market.

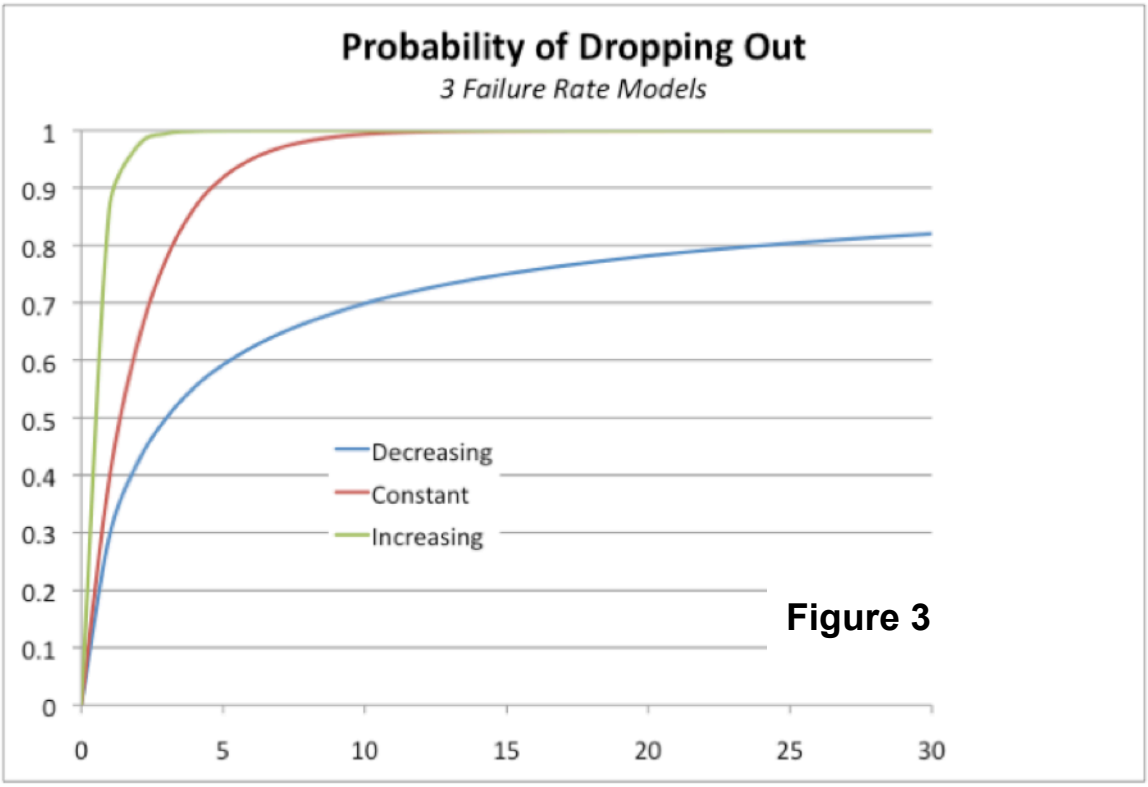


Figure 3

From the figure, it is satisfying to note that the constant failure rate case is bracketed by the two models for the time-dependent failure rates.

Alternative Derivation: In an effort to try and tie together the above modeling results into a more fundamental formulation, the simple first order ordinary differential equation (ODE) describing the constant event rate model was reconsidered. A time-dependent event rate suggests including higher order rates-of-change in the time dependence of the population itself. Thus, the basic model was extended to a 2^{nd}-order ODE, and to keep it simple, with constant coefficients. This resulted in a successful duplication of the above results, as detailed in Appendix C.

This extended result provides a simple mathematical backbone for understanding the connection between statistical distributions describing 1-dimensional population dynamics.

Summary: The results show that the tendency to stay with or abandon a group can be modeled using straightforward assumptions about a person's attitude about the group. This interpretation is summarized for the three types of failure rates:

Constant: Length of membership doesn't matter:
- Situation similar to random failure of light bulbs

Decreasing: Tendency to quit decreases as a person spends more time in the group
- The individual reacts positively to support of group

Increasing: Seeing the failure of others increases a persons tendency to want to leave
- Individual wants to get out while the getting is good.

Another situation, particularly relevant in this era of the Great Recession, is the persistence of unemployment. A recent article[5] describes how a combination of a low supply of jobs and discrimination by employers contributes to long-term unemployment. This reflects a decrease in the rate at which people get hired and results in a Pareto distribution, as derived above.

No doubt the reader can come up with other examples of physics or psychologically based situations that can be modeled in this way.

[5] "Few of long-term jobless find work", LA Times, 3/21/14

Copyright 2014 Randy Patton

Appendix A – Survival Probabilities: Constant and Decreasing Failure Rates

Constant Failure Rate Model: A fundamental model for the random occurrence of events is the Poisson process describing the probability of n events occurring in a given time.

$$P(n) = \frac{e^{-\lambda}\lambda^{n}}{n!} \qquad \text{Eq. A.1}$$

The occurrence, or arrival, rate of events is given by the parameter:

λ = number of events *n* occurring per unit time.

The expected number of events in time t is given by the product λt. Based on this model, the probability of an event occurring within time period *t* is given by the exponential distribution[6]

$$P_f(t) = 1 - e^{-\lambda t} \qquad \text{Eq. A.2}$$

with the parameter λ interpreted as a failure rate. Note that at the start of the process, $t = 0$, which means we have $e^0 = 1$, so that $P_f(0) = 0$. This says that, as expected, the number of failures is zero since no time has elapsed in which a failure can occur. Conversely, for $t \rightarrow \infty$, $P_f(t) \rightarrow 1$, meaning that eventually everything fails.

Decreasing Failure Rate Model: Rather than incorporating lots of parameters to model time-dependent survival rates[7], the current derivation focuses on determining a model-based form for a non-constant failure rate. The germ of the model is the observation that in self-help groups, long-term members have a tendency to remain in the group longer, i.e., the failure rate (likelihood of dropping out of the group) decreases with time. Now, since the overall number of survivors in any group decreases with time, the model can be stated as

- Failure rate is proportional to the probability of *surviving* until time t.

Mathematically, this can be written[8]

$$\lambda(t) \propto P_{surv}(t) = e^{-\lambda t}. \qquad \text{Eq. A.3}$$

[6] The subscript *f* has been used to denote the fact that, in this study, an event is considered as the occurrence of a *failure*.

[7] Given enough parameters, *aka* degrees of freedom, and a flexible functional form, just about any wiggly curve derived from data can be fit by a model, though the underlying structure of the process is obscured.

[8] Using the fact that *Probability of survival = 1 – Probability of failure*

The key to the present derivation is to allow the λ in the exponent to be a function of time (i.e., non-constant), so that

$$\lambda(t) = \alpha e^{-\frac{1}{\lambda_0}\int_0^t \lambda(t')dt'} \quad . \qquad \text{Eq. A.4}$$

with the proportionality coefficient α being the failure rate at time t = 0.

Solving for λ(t) by taking the log of both sides and then the derivative *wrt* t gives

$$\frac{d\ln\left(\frac{\lambda(t)}{\lambda_0}\right)}{dt} = \frac{\lambda'}{\lambda} = -\frac{d}{dt}\left(-\frac{1}{\lambda_0}\int_0^t \lambda(t')dt'\right) = \frac{-\lambda(t)}{\lambda_0}$$

$$\lambda' = -\frac{\lambda^2}{\lambda_0}$$

$$\lambda(t) = \frac{\lambda_0}{1+t} \qquad \text{Eqs. A.5}$$

with limits

$$\lambda(0) = \alpha = \lambda_0 \text{ and } \lambda(\infty) = 0. \qquad \text{Eq. A.6}$$

As can be seen, the failure rate λ(t) decreases with time, as expected.

Solving for the probability of failure after time t, we have[9]

$$\lambda_0 \int_0^t \frac{dt'}{1+t} = \lambda_0 \ln(1+t')\Big|_0^t = \lambda_0 \ln(1+t) \qquad \text{Eq. A.7}$$

$$P_f(t) = 1 - e^{-\lambda_0 \ln(1+t)} = 1 - \frac{1}{(1+t)^{\lambda_0}} \qquad \text{Eq. A.8}$$

This has the form of a classic Pareto distribution. It is interesting to note that the parameter λ_0, which is the initial failure rate of the process, is equal to the so-called shape parameter of a Pareto distribution as described in the literature. I am not aware of this remarkable equivalence being noted in the literature.

[9] with help from Leibnitz' Rule

Copyright 2014 Randy Patton

Appendix B – Survival Probability Based on Increasing Failure Rate

For this case, in analogy with Eq. A.3, let λ be proportional to the probability of *failure*:

$$\lambda(t) \propto P_{fail}(t) = \alpha\left(1 - e^{-\lambda t}\right) + \lambda_0 \qquad \text{Eq. B.1}$$

where again $\lambda_0 = \lambda(0)$ and α is a new parameter. It can be immediately noticed that, since $e^0 =$, indeed $\lambda_0 = \lambda(0)$ and that $\lambda(\infty) = \alpha + \lambda_0$.

To derive the probability, again write the failure rate as a function of time

$$\lambda(t) = \alpha\left[1 - e^{-\int_0^t \lambda(t')dt'}\right] + \lambda_0 . \qquad \text{Eq. B.2}$$

Let

$$u(t) = \int_0^t \lambda(t')dt' \qquad \text{Eq. B.3}$$

and rearrange to get

$$\frac{\alpha + \lambda_0}{\alpha} - \frac{\lambda}{\alpha} = e^{-u} \quad . \qquad \text{Eq. B.4}$$

Take the log

$$\ln\left[\frac{\alpha + \lambda_0}{\alpha} - \frac{\lambda}{\alpha}\right] = -u \qquad \text{Eq. B.5}$$

and then the derivative *wrt* to t

$$\frac{d\lambda}{dt} = \lambda(\alpha + \lambda_0 - \lambda) \quad \text{with} \quad \frac{du}{dt} = \lambda \ . \qquad \text{Eq. B.6}$$

The next step is to use separation of variables to write

$$\frac{d\lambda}{\lambda\left(\alpha + \lambda_0 - \lambda\right)} = dt \qquad \text{Eq. B.7}$$

and expand using partial fractions

$$\left(\frac{1}{(\alpha+\lambda_0)\lambda}+\frac{1}{(\alpha+\lambda_0)(\alpha+\lambda_0-\lambda)}\right)d\lambda = dt \quad \text{Eq. B.8}$$

Integrate both sides (and let t + constant → t for simplicity)

$$\frac{1}{(\alpha+\lambda_0)}\left(\ln\lambda - \ln(\alpha+\lambda_0-\lambda)\right) = t \quad \text{Eq. B.9}$$

and rearrange

$$\ln\left(\frac{\lambda}{\alpha+\lambda_0-\lambda)}\right) = (\alpha+\lambda_0)t \quad \text{Eq. B.10}$$

Exponentiating gives

$$\lambda = \frac{\alpha+\lambda_0}{1+e^{-(\alpha+\lambda_0)t}} \quad \text{Eq. B.11}$$

To satisfy the initial condition $\lambda(t) = \lambda_0$, we need $\alpha = \lambda_0$.

Now integrate to obtain the exponent

$$u(t) = \int_0^t \lambda(t')dt' = (\alpha+\lambda_0)\int_0^t \frac{dt'}{1+e^{-(\alpha+\lambda_0)t'}}$$
$$= (\alpha+\lambda_0)\left[t + \frac{1}{\alpha+\lambda_0}\ln(1+e^{-(\alpha+\lambda_0)t'})\right] - \ln 2 \quad \text{Eq. B.12}$$

And finally substitute into the expression for the probability

$$P_{surv}(t) = e^{-u} = \frac{e^{-(\alpha+\lambda_0)t}}{2}\cdot\left(1+e^{-(\alpha+\lambda_0)t}\right) \quad \text{Eq. B.13}$$

Copyright 2014 Randy Patton

The increasing failure rate model corresponds to the situation where members of the population react to the increasing number of failures by increasing their tendency to "defect". This could be called the "rats leaving the ship" scenario

An even more interesting aspect of the increasing failure rate model is the differential equation associated with it, namely

$$\frac{d\lambda}{dt} = \lambda(\alpha + \lambda_0 - \lambda) \quad . \qquad \text{Eq. B.14}$$

This first-order differential equation is used to model population growth limited by various mechanisms. The population grows exponentially in the beginning, i.e., when $\lambda < \alpha + \lambda$, and then declines as λ^2 for larger values, for instance when resources needed for growth are depleted.

This differential equation can also be used to model epidemics. In this case, the population is divided into a fraction *x* of individuals who are infected and the fraction *y* of those uninfected. The model assume that the disease in transmitted by contact between the *x* and *y* groups so that the rate of infection is proportional to *x* times *y*, or

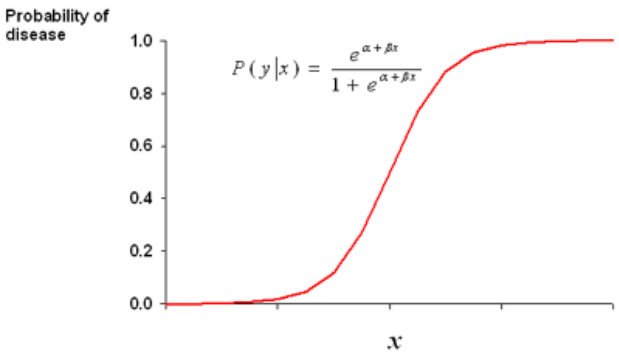
$$\frac{dx}{dt} \propto xy = x(1-x) \qquad \text{Eq. B.15}$$

since $x + y = 1$. With appropriate parameters, this has the form of Eq. 7.

As shown in the derivation, this differential equation for λ results in the function

$$\lambda = \frac{\alpha + \lambda_0}{1 + e^{-(\alpha+\lambda_0)t}} \qquad \text{Eq. B.16}$$

which is the same as that for a so-called logistic function[10] used to model a wide variety of phenomena. It is used, for instance, to provide the commonly used "sigmoid function" (with an appropriate minus sign for negative t) which provides a smooth curve from ±1 for $-\infty < t < +\infty$, as in the figure.



[10] http://en.wikip

Table 1 - Summary of Model Attributes

Failure rate model	**λ**	**DE**	**P_{surv}**
Constant	$\lambda = \lambda_0$	$\lambda' = 0$	$e^{-\lambda_0 t}$
Decreasing	$\lambda = \lambda_0/(1 + t)$	$\lambda' = -\lambda^{2/}\lambda_0$	$1/(1+t)^{\lambda_0}$
Increasing	$\lambda = (\alpha+\lambda_0)/\{1+\exp[-(\alpha+\lambda_0)t]\}$	$\lambda' = \lambda(\alpha+\lambda_0 - \lambda)$	$\exp[-(\alpha+\lambda_0)t]$ $\cdot\{1+\exp[-(\alpha+\lambda_0)t]\}/2$

Included in Table 1 is a column containing the differential equation (DE) which is satisfied by λ.

Copyright 2014 Randy Patton

Appendix C - Alternate derivation of distributions using an ODE

The foregoing approach to treating the decay rate λ as a function of time results in simple ordinary differential equations (ODEs) for λ as summarized in Table 1 (Appendix B). The following is an alternative derivation of these ODEs based on an linear ODE for the survival function and a heuristic form for its solution.

For *P(t)* representing a population, the classic model for population dynamics is a growth/decay process given by the simple ordinary differential equation

$$\frac{dP}{dt} = -\lambda P \rightarrow P(t) = P_0 e^{-\lambda t} \qquad \text{Eq. C.1}$$

where λ is a constant typically interpreted as a decay or growth rate (depending on sign). In either case, it can be referred to as an *event* rate. The solution *P(t)* can be called the survival function for real positive λ.

As before, this model can be extended by considering λ to be a function of time. Rather than just specifying an arbitrary function for λ, a family of solutions can be generated by assuming *P(t)* has an exponential form reminiscent of the above solution

$$P(t) \propto e^{-u(t)} \qquad \text{Eq. C.2}$$

where *u(t)* replaces *λt* in Eq. 1 and leads naturally to the definition

$$u(t) = \int_0^t \lambda(t')dt' \rightarrow \frac{du}{dt} = \lambda(t) \qquad \text{Eq. C.3}$$

.

The key to this development is to further extend the model by including the 2nd derivative of *P(t)* to form the more general version of Eq. C.1

$$\alpha \frac{d^2P}{dt^2} + \beta \frac{dP}{dt} + \gamma P = 0 \qquad \text{Eq. C.4}$$

.

where α, β and γ are constants.

This extension considers not only the rate of increase or decrease in a population (Eq. C.1) but also that the *acceleration* in population numbers (the obvious similarity to Newton's *F = ma* is noted).

Two examples of the motivation for including this 2nd order response (for clarity, let λ be called the decay rate for now)

a) Decreasing decay rate: Longevity in a group increases the longer individuals are in the group (e.g. therapy is working)
b) Increasing decay rate: Movement in population numbers triggers increased movement (e.g., investors joining in a sell off).

Eqs. C.3 and C.4 can be manipulated to give functional forms for $\lambda(t)$. Substituting Eq. C.3 into Eq. C.4 and using the form of the solution Eq. C.2 gives

$$\exp[u(t)]\left\{\alpha\left[\left(\frac{du}{dt}\right)^2+\frac{d^2u}{dt^2}\right]+\beta\frac{du}{dt}+\gamma\right\}=0$$

. Eq. C.5

Substuting Eq. 3 gives an ordinary differential equation (ODE) for the event rate λ,

$$\alpha\left\{\lambda^2+\frac{d\lambda}{dt}\right\}+\beta\lambda+\gamma=0$$

. Eq. C.6

Various varieties of an ODE for λ can now be easily obtained by setting combinations of the constants equal to zero. The three cases for constant, decreasing and increasing event rate are derived as follows.

The solution for a constant λ (Eq. 1) is obtained by simply setting $\alpha = 0$ so that $\lambda = -\gamma/\beta$. This then gives the familiar exponential solution of Eq. C.1.

Setting $\beta = \gamma = 0$, so that only α is non-zero, gives the ODE for λ associated with the Pareto distribution, interpreted as a survival function as in Eq. A.5 and A.8.

The third case is obtained by setting only $\gamma = 0$ which results in the logistic function for $\lambda(t)$, as in Eq B.6 and the survival function B.13.

Copyright 2014 Randy Patton