Parametric Survival Models

Germán Rodríguez
grodri@princeton.edu

Spring, 2001; revised Spring 2005, Summer 2010

We consider briefly the analysis of survival data when one is willing to assume a parametric form for the distribution of survival time.

1 Survival Distributions

1.1 Notation

Let T denote a continuous non-negative random variable representing survival time, with probability density function (pdf) $f(t)$ and cumulative distribution function (cdf) $F(t) = \Pr\{T \leq t\}$. We focus on the *survival function* $S(t) = \Pr\{T > t\}$, the probability of being alive at t, and the hazard function $\lambda(t) = f(t)/S(t)$. Let $\Lambda(t) = \int_0^t \lambda(u)du$ denote the cumulative (or integrated) hazard and recall that

$$S(t) = \exp\{-\Lambda(t)\}.$$

Any distribution defined for $t \in [0, \infty)$ can serve as a survival distribution. We can also draft into service distributions defined for $y \in (-\infty, \infty)$ by considering $t = \exp\{y\}$, so that $y = \log t$. More generally, we can start from a r.v. W with a standard distribution in $(-\infty, \infty)$ and generate a family of survival distributions by introducing location and scale changes of the form

$$logT = Y = \alpha + \sigma W.$$

We now review some of the most important distributions.

1.2 Exponential

The exponential distribution has constant hazard $\lambda(t) = \lambda$. Thus, the survivor function is $S(t) = \exp\{-\lambda t\}$ and the density is $f(t) = \lambda \exp\{-\lambda t\}$. It can be shown that $E(T) = 1/\lambda$ and $\text{var}(T) = 1/\lambda^2$. Thus, the coefficient of variation is 1.

The exponential distribution is related to the extreme-value distribution. Specifically, T has an exponential distribution with parameter λ, denoted $T \sim E(\lambda)$, iff

$$Y = \log T = \alpha + W$$

where $\alpha = -\log \lambda$ and W has a standard extreme value (min) distribution, with density

$$f_W(w) = e^{w - e^w}.$$

This is a unimodal density with $E(W) = -\gamma$, where $\gamma = 0.5722$ is Euler's constant, and $\text{var}(W) = \pi^2/6$. The skewness is -1.14.

The proof follows immediately from a change of variables.

1.3 Weibull

T is Weibull with parameters λ and p, denoted $T \sim W(\lambda, p)$, if $T^p \sim E(\lambda)$. The cumulative hazard is $\Lambda(t) = (\lambda t)^p$, the survivor function is $S(t) = \exp\{-(\lambda t)^p\}$, and the hazard is

$$\lambda(t) = \lambda^p p t^{p-1}.$$

The log of the Weibull hazard is a linear function of log time with constant $p \log \lambda + \log p$ and slope $p - 1$. Thus, the hazard is rising if $p > 1$, constant if $p = 1$, and declining if $p < 1$.

The Weibull is also related to the extreme-value distribution: $T \sim W(\lambda, p)$ iff

$$Y = \log T = \alpha + \sigma W,$$

where W has the extreme value distribution, $\alpha = -\log \lambda$ and $p = 1/\sigma$.

The proof follows again from a change of variables; start from W and change variables to $Y = \alpha + \sigma W$, and then change to $T = e^Y$.

1.4 Gompertz-Makeham

The Gompertz distribution is characterized by the fact that the log of the hazard is linear in t, so

$$\lambda(t) = \exp\{\alpha + \beta t\}$$

and is thus closely related to the Weibull distribution where the log of the hazard is linear in $\log t$. In fact, the Gompertz *is* a log-Weibull distribution.

This distribution provides a remarkably close fit to adult mortality in contemporary developed countries.

1.5 Gamma

The gamma distribution with parameters λ and k, denoted $\Gamma(\lambda, k)$, has density

$$f(t) = \frac{\lambda(\lambda t)^{k-1} e^{-\lambda t}}{\Gamma(k)},$$

and survivor function

$$S(t) = 1 - I_k(\lambda t),$$

where $I_k(x)$ is the incomplete gamma function, defined as

$$I_k(x) = \int_0^x \lambda^{k-1} e^{-x} dx / \Gamma(k).$$

There is no closed-form expression for the survival function, but there are excellent algorithms for its computation. (R has a function called `pgamma` that computes the cdf and survivor function. This function calls k the shape parameter and $1/\lambda$ the scale parameter.)

There is no explicit formula for the hazard either, but this may be computed easily as the ratio of the density to the survivor function, $\lambda(t) = f(t)/S(t)$. The gamma hazard

- increases monotonically if $k > 1$, from a value of 0 at the origin to a maximum of λ,
- is constant if $k = 1$
- decreases monotonically if $k < 1$, from ∞ at the origin to an asymptotic value of λ.

If $k = 1$ the gamma reduces to the exponential distribution, which can be described as the waiting time to one hit in a Poisson process. If k is an integer $k > 1$ then the gamma distribution is called the Erlang distribution and can be characterized as the waiting time to k hits in a Poisson process. The distribution exists for non-integer k as well.

The gamma distribution can also be characterized in terms of the distribution of log-time. By a simple change of variables one can show that $T \sim \Gamma(\lambda, k)$ iff

$$\log T = Y = \alpha + W,$$

where W has a *generalized* extreme-value distribution with density

$$f_w(w) = \frac{e^{kw - e^w}}{\Gamma(k)},$$

controlled by a parameter k. This density reduces to the ordinary extreme value distribution when $k = 1$.

1.6 Generalized Gamma

Stacy has proposed a generalized gamma distribution that fits neatly in the scheme we are developing, as it simply adds a scale parameter in the expression for $\log T$, so that

$$Y = \log T = \alpha + \sigma W,$$

where W has a generalized extreme value distribution with parameter k. The density of the generalized gamma distribution can be written as

$$f(t) = \frac{\lambda p (\lambda t)^{pk-1} e^{-(\lambda t)^p}}{\Gamma(k)},$$

where $p = 1/\sigma$.

The generalized gamma includes the following interesting special cases:

- gamma, when $p = 1$,
- Weibull, when $k = 1$,
- exponential, when $p = 1$ and $k = 1$.

It also includes the log-normal as a special limiting case when $k \to \infty$.

1.7 Log-Normal

T has a lognormal distribution iff

$$Y = \log T = \alpha + \sigma W,$$

where W has a standard normal distribution.

The hazard function of the log-normal distribution increases from 0 to reach a maximum and then decreases monotonically, approaching 0 as $t \to \infty$.

As $k \to \infty$ the generalized extreme value distribution approaches a standard normal, and thus the generalized gamma approaches a log-normal.

1.8 Log-Logistic

T has a log-logistic distribution iff

$$Y = \log T = \alpha + \sigma W,$$

where W has a standard logistic distribution, with pdf

$$f_W(w) = \frac{e^w}{(1+e^w)^2},$$

and cdf

$$F_W(w) = \frac{e^w}{1+e^w}.$$

The survivor function is the complement

$$S_W(w) = \frac{1}{1+e^w}.$$

Changing variables to T we find that the log-logistic survivor function is

$$S(t) = \frac{1}{1+(\lambda t)^p},$$

where we have written, as usual, $\alpha = -\log\lambda$ and $p = 1/\sigma$. Taking logs we obtain the (negative) integrated hazard, and differentiating w.r.t. t we find the hazard function

$$\lambda(t) = \frac{\lambda p(\lambda t)^{p-1}}{1+(\lambda t)^p}.$$

Note that the *logit* of the survival function $S(t)$ is linear in $\log t$. This fact provides a diagnostic plot: if you have a non-parametric estimate of the survivor function you can plot its logit against log-time; if the graph looks like a straight line then the survivor function is log-logistic.

The hazard itself is

- monotone decreasing from ∞ if $p < 1$,
- monotone decreasing from λ if $p = 1$, and
- similar to the log-normal if $p > 1$.

1.9 Generalized F

Kalbfleisch and Prentice (1980) consider the more general case where

$$Y = \log T = \alpha + \sigma W$$

and W is distributed as the log of an F-variate (which adds two more parameters).

The interesting thing about this distribution is that it includes *all* of the above distributions as special or limiting cases, and is therefore useful for testing different parametric forms.

1.10 The Coale-McNeil Model

The Coale-McNeil model of first marriage frequencies among women who will eventually marry is closely related to the extreme value and gamma distributions.

The model assumes that the density of first marriages at age a among women who will eventually marry is given by

$$g(a) = g_0\left(\frac{a-a_0}{k}\right)\frac{1}{k},$$

where a_0 and k are location and scale parameters and $g_0(.)$ is a standard schedule based on Swedish data. This standard schedule was first derived empirically, but later Coale and McNeil showed that it could be closely approximated by the following analytic expression:

$$g_0(z) = 1.946e^{-0.174(z-6.06)-e^{-0.288(z-6.06)}}.$$

It will be convenient to write a somewhat more general model with three parameters:

$$g(x) = \frac{\lambda}{\gamma(\alpha/\gamma)}e^{-\alpha(x-\theta)-e^{-\lambda(x-\theta)}}.$$

This is a form of extreme value distribution. In fact, if $\alpha = \lambda$ it reduces to the standard extreme value distribution that we discussed before. This more general case is known as a (reversed) generalized extreme value.

The mean of this distribution is

$$\mu = \theta - \frac{1}{\lambda}\psi(\alpha/\lambda),$$

where $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the *digamma* function (or derivative of the log of the gamma function).

The Swedish standard derived by Coale and McNeil corresponds to the case

$$\alpha = 0.174, \qquad \lambda = 0.288, \qquad \text{and} \qquad \theta = 6.06,$$

which gives a mean of $\mu = 11.36$.

By a simple change of variables, it can be seen that the more general case with parameters a_0 and k corresponds to

$$\alpha^* = \frac{0.174}{k}, \qquad \lambda^* = \frac{0.288}{k}, \qquad \text{and} \qquad \theta^* = a_0 + 6.06k.$$

Thus, X has the (more general) Coale-McNeil distribution with parameters α, λ and θ iff

$$X = \theta - \frac{1}{\lambda} \log Y,$$

where Y has a gamma distribution with shape parameter $p = \alpha/\lambda$.

In other words, age at marriage is distributed as a linear function of the logarithm of a gamma random variable.

In particular, the Swedish standard can be obtained as

$$X = 6.06 - \frac{1}{0.288} \log Y,$$

where Y is gamma with $p = \alpha/\lambda = 0.174/0.288 = 0.604$.

The case with parameters a_0 and k can be obtained as

$$X = a_0 + 6.06k - \frac{k}{0.288} \log Y,$$

where Y is again gamma with $p = 0.604$.

The Coale-McNeil models holds the ratio $p = \alpha/\lambda$ fixed at 0.604, but along the way we have generalized the model and could entertain the notion of estimating p rather than holding it fixed.

The main significance of these results is computational:

- we can calculate marriage schedules as long as we have a function to compute the incomplete gamma function (or even chi-squared)
- we can fit nuptiality models using software for fitting gamma models.

For further details see my 1980 paper with Trussell. In that paper we used the mean and standard deviation as the parameters of interest, instead of a_0 and k. I have also written a set of R/S functions to compute marriage schedules, and these are documented separately.

2 Models With Covariates

There are four approaches to modelling survival data with covariates:

- Parametric Families
- Accelerated Life
- Proportional Hazards
- Proportional Odds

We describe each in turn.

2.1 Parametric Families

A general approach is to pick one of the parametric distributions that we have discussed and let the *parameters* of that distribution depend on covariates. For example,

- In an exponential distribution we could let the parameter λ depend on a vector of covariates x, for example using a log-linear model where

$$\log \lambda = x'\beta$$

- In a Weibull distribution we could use a similar model for λ while holding p fixed, or we could let p depend on covariates as well, for example as

$$\log p = x'\gamma$$

- In the Coale-McNeil model using the Rodríguez-Trussell parametrization, one could use a linear model for the mean

$$\mu = x'\beta$$

  while holding the standard deviation σ constant (as usually done in linear models). Alternatively, we could let the dispersion depend on covariates as well, using

$$\log \sigma = x'\gamma,$$

  with parameters γ. In the most general case, we could let the proportion that eventually marries depend on yet another set of parameters.

In general, with k groups one could give each group its own distribution in a family. This is a workable approach, but it is not exactly parsimonious and doesn't lend itself to easy interpretations.

2.2 Accelerated Life Models

Consider an ordinary regression model for log survival time, of the form

$$Y = \log T = -x'\beta + \sigma W,$$

where the error term W has a suitable distribution, e.g. extreme value, generalized extreme value, normal or logistic. This leads to Weibull, generalized gamma, log-normal or log-logistic models for T.

For example if W is extreme value then T has a Weibull distribution with

$$\log \lambda = x'\beta \qquad \text{and} \qquad p = \frac{1}{\sigma}.$$

Note that λ depends on the covariates but p is assumed the same for everyone.

This model has an *accelerated life* interpretation. In this formulation we view the error term σW as a standard or reference distribution that applies when $x = 0$. It will be convenient to translate the reference distribution to the time scale by defining $T_0 = \exp\{\sigma W\}$. The probability that a reference subject will be alive at time t, which will be denoted $S_0(t)$, is

$$S_0(t) = \Pr\{T_0 > t\} = \Pr\{W > \log t/\sigma\}.$$

Consider now the effect of the covariates x. In this model T is distributed as $T_0 e^{-x'\beta}$, so the covariates act multiplicatively on survival time. What is the probability that a subject with covariate values x will be alive at time t?

$$S(t, x) = \Pr\{T > t|x\} = \Pr\{T_0 e^{-x'\beta} > t\} = \Pr\{T_0 > te^{x'\beta}\} = S_0(te^{x'\beta}).$$

In words, the probability that a subject with covariates x will be alive at time t is the same as the probability that a reference subject will be alive at time $t \exp\{x'\beta\}$. This may be interpreted as time passing more rapidly (or people aging more quickly) by a factor $\exp\{x'\beta\}$, for example twice as fast or half as fast. (The analogy to 'dog years' should no go unnoticed.)

We can also write the density and hazard functions for any subject in terms of the baseline or reference density and hazard:

$$f(t) = f_0(te^{x'\beta})e^{x'\beta},$$

and

$$\lambda(t) = \lambda_0(te^{x'\beta})e^{x'\beta}.$$

We see that a simple relationship between the survivor functions for different x's (just a stretching of the time axis), translates into a more complex relationship when viewed in terms of the pdf or the hazard function.

Consider for example a multiplier of two for a subject with covariates x. In terms of survival, this means that the probability that the subject would be alive at any given age is the same as the probability that a reference subject would be alive at twice the age. In terms of risk, it means that our subject is exposed at any given age to double the risk of a reference subject twice as old.

Note also that if we start with a given distribution and stretch the time axis we may well end up with a distribution in a completely different family. Stretching a Weibull produces another Weibull, but not all families are closed under acceleration of time. (Is the Coale-McNeil an accelerated life model?)

2.3 Proportional Hazards

An alternative approach to modelling survival data is to assume that the effect of the covariates is to increase or decrease the *hazard* by a proportionate amount at all durations. Thus

$$\lambda(t, x) = \lambda_0(t)e^{x'\beta},$$

where $\lambda_0(t)$ is the *baseline hazard*, or the hazard for a reference individual with covariate values 0, and $\exp\{x'\beta\}$ is the *relative risk* associated with covariate values x.

Obviously the cumulative hazards would follow the same relationship, as can be seen by integrating both sides of the previous equation. Exponentiating minus the integrated hazard we find the survivor functions to be

$$S(t, x) = S_0(t)^{e^{x'\beta}},$$

so the survivor function for covariates x is the baseline survivor raised to a power. If a subject is exposed to twice the risk of a reference subject at every age, then the probability that the subject will be alive at any given age is the square of the probability that the reference subject would be alive at the same age. In this model a simple relationship in terms of hazards translates into a more complex relationship in terms of survival functions.

These equations define a *family* of models. Picking a different parametric form for the baseline hazard leads to a different model in the proportional hazards family. Suppose we start with a Weibull baseline hazard, so

$$\lambda_0(t) = \lambda p(\lambda t)^{p-1},$$

and we then multiply this by a relative risk $e^{x'\beta}$. You should be able to show that the resulting hazard is again Weibull,

$$\lambda(t, x) = \lambda^* p(\lambda^* t)^{p-1},$$

with the same p as before but $\lambda^* = \lambda e^{x'\beta/p}$.

Thus, the Weibull family is closed under proportionality of hazards, but this is not true for other distributions. If T_0 is log-logistic, for example, and we multiply the *hazard* by a relative risk $e^{x'\beta}$, the resulting distribution is not log-logistic.

2.4 Proportional Hazards and Accelerated Life

Do the proportional hazard and accelerated life models ever coincide? More precisely, if we start with a hazard and multiply by a relative risk, and someone else starts with another hazard and stretches time, do we ever end up with the same distribution? The condition just formulated may be stated as

$$\lambda_0(t)e^{x'\beta} = \lambda_0^*(te^{x'\beta^*})e^{x'\beta^*}$$

for all x and t. The stars indicate that we do not necessarily start with the same baseline hazard or end up with the same parameters reflecting the effects of the covariates.

If this condition is to be true for all x then it must be true for $x = 0$, implying

$$\lambda_0(t) = \lambda_0^*(t),$$

so the baseline hazards must be the same. Let us try to find this hazard.

The trick here is to consider a very special value of the vector of covariates x, where we set the first element to $-\log t/\beta_1^*$ and the others to zero, so

$$x = (-\log t/\beta_1^*, 0, \ldots, 0).$$

Multiplying by β^* we find that $x'\beta^* = -\log t$, while multiplying by β gives $x'\beta = -\log t\beta_1/\beta_1^*$. Using these results on the condition we obtain

$$\lambda_0(t)e^{-\log t\beta_1/\beta_1^*} = \lambda_0(te^{-\log t})e^{-\log t}.$$

Using the fact that $e^{b\log a} = a^b$ we can simplify this expression to

$$\lambda_0(t)\left(\frac{1}{t}\right)^{\beta_1/\beta_1^*} = \lambda_0(1)\frac{1}{t},$$

or, moving the term on $1/t$ to the right-hand side,

$$\lambda_0(t) = \lambda_0(1)t^{\beta_1/\beta_1^*-1}.$$

Repeat this exercise with a covariate vector x that has $\log t/\beta_i^*$ in the i-th slot and 0 everywhere else, so that $x'\beta^* = -\log t$ and $x'\beta = -\log t\beta_i/\beta_i^*$.

We find the same result but with $\beta_i/\beta_i^* - 1$ as the exponent of t. If the condition is to be true for all x, then the ratio of the coefficients must be constant,

$$\frac{\beta_i}{\beta_i^*} = \frac{\beta}{\beta^*} = p,$$

say. This leads to the solution

$$\lambda_0(t) = \lambda_0(1)t^{p-1},$$

which can be recognized as a Weibull hazard. To see this last point write the result in the more familiar form

$$\lambda_0(t) = \lambda^p p t^{p-1} = \lambda p(\lambda t)^{p-1},$$

where I have taken the constant to be $\lambda_0(1) = \lambda^p p$, which is the same as defining $\lambda = (\lambda_0(1)/p)^{1/p}$.

This result shows that the Weibull is the *only* distribution that is closed under both the accelerated life and proportional hazards families.

Note that the accelerated life and proportional hazards parameters β^* and β are proportional to each other, with proportionality constant p. In particular, they are equal for $p = 1$.

Thus, doubling the risk in an exponential model makes time go twice as fast. But doubling the risk in a Weibull model with $p = 2$ makes time go only about 40% faster. Can you see why?

2.5 Proportional Odds

An alternative approach to survival modelling is to assume that the effect of the covariates is to increase or decrease the *odds* of dying by a given duration by a proportionate amount:

$$\frac{1 - S(t,x)}{S(t,x)} = \frac{1 - S_0(t)}{S_0(t)} e^{x'\beta},$$

where $S_0(t)$ is a baseline survivor function, taken from a suitable distribution, and $\exp\{x'\beta\}$ is a multiplier reflecting the proportionate increase in the odds associated with covariate values x.

Taking logs, we find that

$$\text{logit}(1 - S(t,x)) = \text{logit}(1 - S_0(t)) + x'\beta,$$

so the covariate effects are linear in the logit scale.

A somewhat more general version of the proportional odds model (but without covariates) is known as the *relational logit model* in demography. The idea is to allow the log-odds of dying in a given population to be a linear function of the log-odds in a reference or baseline population, so that

$$\text{logit}(1 - S(t)) = \alpha + \theta \text{logit}(1 - S_0(t)).$$

These models were popularized by Brass. The proportional odds model is the special case where $\theta = 1$ (but we let the constant α depend on covariates).

These models could be defined in terms of the odds of *surviving* to duration t, but this merely changes the sign of β. I prefer the definition in terms of the odds of *dying* because it preserves the interpretation of the β coefficients as increasing the risk, which is consistent with hazard models. This is also the reason why I used a minus sign when defining the coefficients for accelerated life models.

As an example consider a proportional odds model with a log-logistic baseline. The corresponding survival function, its complement, and the odds of dying are

$$S_0(t) = \frac{1}{1 + (\lambda t)^p}, \qquad 1 - S_0(t) = \frac{(\lambda t)^p}{1 + (\lambda t)^p}, \qquad \frac{1 - S_0(t)}{S_0(t)} = (\lambda t)^p.$$

Multiplying the odds by $\exp\{x'\beta\}$ yields another log-logistic model, this time with $\lambda^* = \lambda e^{x'\beta/p}$ and $p^* = p$. Thus, the log-logistic family is closed under proportionality of odds.

This is not true of other distributions. For example if we start with a Weibull baseline and multiply the odds of dying by a constant, the resulting distribution is not Weibull.

2.6 Proportional Odds And Accelerated Life

Do the proportional odds and accelerated life models ever coincide? The answer is yes, when (and only when) the baseline is log-logistic.

The proof follows essentially the same steps as the proof for the intersection of the proportional hazards and accelerated life models.

3 Maximum Likelihood Estimation

All parametric models may be fit by maximizing the appropriate likelihood function.

The data consist of pairs $\{t_i, d_i\}$ where

- t_i is the survival or censoring time, and
- d_i is a death indicator, taking the value 1 for deaths and 0 for censored cases

The likelihood function under general non-informative censoring has the form

$$L(\theta) = \prod_{i=1}^{n} \lambda(t_i|x_i)^{d_i} S(t_i|x_i),$$

and in general must be maximized numerically using a procedure such as Newton-Raphson.

Kalbfleisch and Prentice have a nice discussion of the procedures that need to be followed in fitting parametric models, including first and second derivatives for accelerated life models using the parametric distributions discussed here.

Stata's `streg` can fit a number of parametric models, including exponential, Weibull and Gompertz in the proportional hazards framework, and log-normal, log-logistic, and generalized gamma (as well as exponential and Weibull) in the accelerated failure-time framework. Now you know why the Weibull is included in both the PH and AFT metrics.