flexsurv: Distributions reference

Christopher H. Jackson
MRC Biostatistics Unit, Cambridge, UK
chris.jackson@mrc-bsu.cam.ac.uk

Abstract

A reference guide for the distributions built into flexsurv

Keywords: survival.

This document lists the following information for each of the built-in distributions in `flexsurvreg`.

- How the distribution is identified in the `dist` argument to `flexsurvreg`.

- Name of the `d` function in R for the probability density, showing how names of the arguments in the R function correspond to the symbols in the formula for the survivor function. In each case, there are corresponding R functions with names beginning with `p,q` and `r` instead of `d`, for the cumulative distribution, quantiles and random number generation.

  The R `help` page for the function named here will contain more information about the distribution, e.g. the probability density function, and how particular distributions are defined as special cases of other distributions.

- The survivor function $S(t|\boldsymbol{\theta})$ as a function of time t and the parameters of the distribution $\boldsymbol{\theta}$. Any restrictions on the allowed values of the parameters are noted.

- The location parameter of the distribution, now indexed by j, and a function describing how the location parameter depends on covariate values $\mathbf{z}_j$ and covariate effects $\boldsymbol{\beta}$.

  $\boldsymbol{\beta}$ are the covariate effects indicated when printing a `flexsurvreg` object, which can take any real value.

- A more interpretable covariate effect measure, defined as a function of $\boldsymbol{\beta}$.

In a *proportional hazards* model, the "more interpretable" effect measure presented is the hazard ratio (HR) for one unit of the covariate.

In an *accelerated failure time model*, the time acceleration factor (TAF) for one unit of the covariate is presented. The survivor function is of the form $S(t) = S^*(ct)$ where c is some constant that depends on the parameters. Multiplying c by 2 has the same effect on survival as multiplying t by 2, doubling the speed of time and halving the expected survival time. An alternative way of presenting effects from an accelerated failure time model is the ratio of expected times to the event between covariate values of 1 and 0. This equals 1/TAF, and may be easier to interpret if the time to the event is of direct interest.

A HR below 1 and a TAF above 1 both indicate that higher covariate values are associated with a higher risk of the event, or shorter times to the event.

Weibull (accelerated failure time)

```
flexsurvreg(..., dist="weibull")
dweibull(..., shape=a, scale=mu)
```

$$S(t|\mu, a) = \exp(-(t/\mu)^a), \qquad \mu > 0, a > 0$$

$$\mu_j = \exp(\mathbf{z}_j \boldsymbol{\beta}), \quad TAF = \exp(-\boldsymbol{\beta})$$

Weibull (proportional hazards)

```
flexsurvreg(..., dist="weibullPH")
dweibullPH(..., shape=a, scale=lambda)
```

$$S(t|\lambda, a) = \exp(-\lambda t^a), \qquad \lambda > 0, a > 0$$

$$\lambda_j = \exp(\mathbf{z}_j \boldsymbol{\beta}), \quad HR = \exp(\boldsymbol{\beta})$$

(Note the argument "scale" to `dweibullPH` would perhaps better have been called "rate", given the analogy with the rate of the exponential model).

Gamma

```
flexsurvreg(..., dist="gamma")
dgamma(..., shape=a, rate=mu)
```

$$S(t|a, \mu) = 1 - \int_0^t \frac{x^{a-1}\mu^a \exp-(\mu x)}{\Gamma(a)} dx, \qquad \mu > 0, a > 0$$

$$\mu_j = \exp(\mathbf{z}_j \boldsymbol{\beta}), \quad TAF = \exp(\boldsymbol{\beta})$$

Exponential

```
flexsurvreg(..., dist="exp")
dexp(..., rate=lambda)
```

$$S(t|\lambda) = \exp(-\lambda t), \qquad \lambda > 0$$

$$\lambda_j = \exp(\mathbf{z}_j \boldsymbol{\beta}), \quad HR = \exp(\boldsymbol{\beta}), \quad TAF = \exp(\boldsymbol{\beta})$$

Log-logistic

```
flexsurvreg(..., dist="llogis")
dllogis(..., shape=a, scale=b)
```

$$S(t|a,b) = 1/(1 + (t/b)^a), \qquad a > 0, b > 0$$

$$b_j = \exp(\mathbf{z}_j\boldsymbol{\beta}), \quad TAF = \exp(-\boldsymbol{\beta})$$

Log-normal

```
flexsurvreg(..., dist="lnorm")
dlnorm(..., meanlog=mu, sdlog=sigma)
```

$$S(t|\mu,\sigma) = 1 - \int_0^t \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\} dx, \qquad \sigma > 0$$

$$\mu_j = z_j\boldsymbol{\beta}, \quad TAF = \exp(-\boldsymbol{\beta})$$

Gompertz

```
flexsurvreg(..., dist="gompertz")
dgompertz(..., shape=a, rate=b)
```

$$S(t|a,b) = \exp(-(b/a)(\exp(at) - 1)), b > 0$$

Note that $a < 0$ is permitted, in which case $S(t|)$ tends to a non-zero probability as t increases, i.e. a probability of living forever.

$$b_j = \exp(\mathbf{z}_j\boldsymbol{\beta}), \quad HR = \exp(\boldsymbol{\beta})$$

Generalised gamma (Prentice)

```
flexsurvreg(..., dist="gengamma")
dgengamma(..., mu, sigma, Q)
```

$$S(t|\mu,\sigma,Q) = \begin{array}{ll} S_G(\frac{\exp(Qw)}{Q^2} \mid \frac{1}{Q^2}, 1) & (Q > 0) \\ 1 - S_G(\frac{\exp(Qw)}{Q^2} \mid \frac{1}{Q^2}, 1) & (Q < 0) \\ S_L(t \mid \mu, \sigma) & (Q = 0) \end{array}$$

where $w = (\log(t) - \mu)/\sigma$, $S_G(t|a,1)$ is the survivor function of the gamma distribution with shape a and scale 1, $S_L(t|\mu,\sigma)$ is the survivor function of the log-normal distribution with log-scale mean μ and standard deviation σ, μ, Q are unrestricted, and σ is positive.

$$\mu_j = z_j \boldsymbol{\beta}, \quad TAF = \exp(-\boldsymbol{\beta})$$

Generalised gamma (Stacy)

```
flexsurvreg(..., dist="gengamma.orig")
dgengamma.orig(..., shape=b, scale=a, k=k)
```

$$S(t|a, b, k) = 1 - \int_0^t \frac{bx^{bk-1}}{\Gamma(k)a^{bk}} \exp(-(x/a)^b)dx$$

$$a_j = \exp(z_j \boldsymbol{\beta}), \quad TAF = \exp(-\boldsymbol{\beta})$$

Generalised F (Prentice)

```
flexsurvreg(..., dist="genf")
dgenf(..., mu, sigma, Q, P)
```

$$S(t|\mu, \sigma, Q, P) = 1 - \int_0^t \frac{\delta(s_1/s_2)^{s_1} e^{s_1 w}}{\sigma x (1 + s_1 e^w/s_2)^{(s_1+s_2)} B(s_1, s_2)}, \qquad \sigma > 0, P > 0$$

where $s_1 = 2(Q^2 + 2P + Q\delta)^{-1}$, $s_2 = 2(Q^2 + 2P - Q\delta)^{-1}$, $\delta = (Q^2 + 2P)^{1/2}$ and $w = (\log(x) - \mu)\delta/\sigma$

$$\mu_j = x_j \boldsymbol{\beta}, \quad TAF = \exp(-\boldsymbol{\beta})$$

Generalised F (original)

```
flexsurvreg(..., dist="genf.orig")
dgenf.orig(..., mu, sigma, s1, s2)
```

$$S(t|\mu, \sigma, s_1, s_2) = 1 - \int_0^t \frac{(s_1/s_2)^{s_1} e^{s_1 w}}{\sigma x (1 + s_1 e^w/s_2)^{s_1+s_2} B(s_1, s_2)} dx, \qquad \sigma > 0, s_1 > 0, s_2 > 0$$

where $w = (\log(x) - \mu)/\sigma$, and $B(s_1, s_2) = \frac{\Gamma(s_1)\Gamma(s_2)}{\Gamma(s_1+s_2)}$ is the beta function.

$$\mu_j = z_j \boldsymbol{\beta}, \quad TAF = \exp(-\boldsymbol{\beta})$$

Spline (Royston/Parmar)

```
flexsurvspline(...,scale="hazard")
flexsurvspline(...,scale="odds")
flexsurvspline(...,scale="normal")
```
`dsurvspline(..., gamma)` where the argument `gamma` collects together all parameters,

or the alternative forms for the density/distribution functions (flexsurv versions 2.0 and later) where different parameters are in separate arguments:

dsurvspline0(..., gamma0, gamma1) (no internal knots)

dsurvspline1(..., gamma0, gamma1, gamma2) (1 internal knot)

dsurvspline2(..., gamma0, gamma1, gamma2, gamma3) (2 internal knots)

etc., all the way up to dsurvspline7.

$$g(S(t)) = s(x, \boldsymbol{\gamma})$$

where $x = \log(t)$ and $s()$ is a natural cubic spline function with m internal knots:

$$s(x, \boldsymbol{\gamma}) = \gamma_0 + \gamma_1 x + \gamma_2 v_1(x) + \ldots + \gamma_{m+1} v_m(x)$$

where $v_j(x)$ is the jth *basis* function,

$$v_j(x) = (x - k_j)^3_+ - \lambda_j(x - k_{min})^3_+ - (1 - \lambda_j)(x - k_{max})^3_+, \qquad \lambda_j = \frac{k_{max} - k_j}{k_{max} - k_{min}}$$

and $(x - a)_+ = max(0, x - a)$.

The link function relating the survivor function to the spline is:

$$g(S(t)) = \begin{matrix} \log(-\log(S(t))) & (\texttt{scale="hazard"}) \\ \log(S(t)^{-1} - 1) & (\texttt{scale="odds"}) \\ -\Phi^{-1}(S(t)) & (\texttt{scale="normal"}) \end{matrix}$$

For more details, see the main flexsurv vignette.

The location parameter is

$$\gamma_{0j} = z_j \boldsymbol{\beta}$$

so that higher values of the location parameter γ_0 lead to higher hazards, or lower survival, for all three choices of link function. For scale="hazard", $\exp(\boldsymbol{\beta})$ is the hazard ratio, and for scale="odds", $\exp(\boldsymbol{\beta})$ is an odds ratio.