Title

Firstname Lastname
Affiliation

Abstract

—!!!—an abstract is required—!!!—

Keywords: —!!!—at least one keyword is required—!!!—.

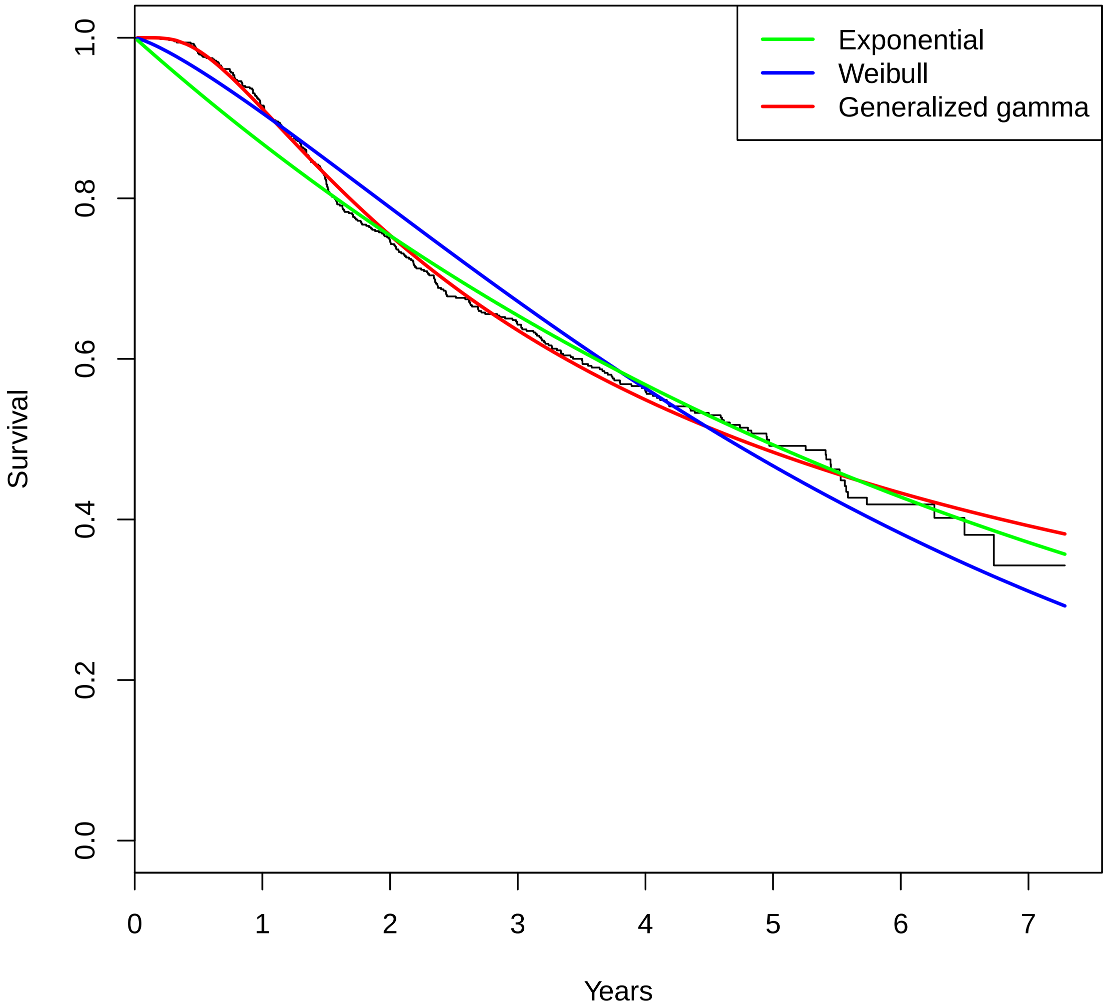
There are two common model choice problems in parametric survival analysis:

1. the selection of covariates, for example in a proportional hazards or accelerated failure time regression model.

2. the selection of the appropriate level of flexibility for a parametric hazard or survival function (given specific "baseline" covariate values).

In this example, focused model comparison is used for the second of these problems. A series of increasingly flexible parametric distributions are fitted to a set of 686 right-censored survival times from patients with primary node positive breast cancer (originally from @sauerbrei1999building, and also provided in the 'flexsurv' package). The exponential, Weibull and generalized gamma models are fitted, which have one, two and three parameters respectively.

```
if (!require("flexsurv")) stop("The `flexsurv` package should be installed to run code in
gnette")
ex <-  flexsurvreg(Surv(recyrs, censrec) ~ 1, data=bc, dist="exponential")
we <- flexsurvreg(Surv(recyrs, censrec) ~ 1, data=bc, dist="weibull")
gg <- flexsurvreg(Surv(recyrs, censrec) ~ 1, data=bc, dist="gengamma")
plot(gg, ci=FALSE, conf.int=FALSE, ylab="Survival", xlab="Years")
lines(we, col="blue", ci=FALSE)
lines(ex, col="green", ci=FALSE)
legend("topright", lty=c(1,1,1), lwd=c(2,2,2), col=c("green", "blue","red"), c("Exponentia
```



Each is a generalisation of the previous one, as described in the **flexsurv** documentation. Therefore we can use focused model comparison, with the "wide" model taken to be the generalized gamma, and test all three models together.

However this requires the Weibull and exponential models to be reparameterised as generalized gamma models with parameters μ, σ, Q fixed to special values: the exponential has $\sigma = 1, Q = 1$ and its rate parameters is $1/exp(\mu)$, while the Weibull model has $Q = 1$, shape $1/\sigma$ and scale $exp(\mu)$. This is done by supplying 'inits' and 'fixedpars' to 'flexsurvreg' as follows.

```
we2 <- flexsurvreg(Surv(recyrs, censrec) ~ 1, data=bc, dist="gengamma", inits=c(1,1,1), fi
pars=3)  # weibull with shape 1/sigma, scale exp(mu)
ex2 <- flexsurvreg(Surv(recyrs, censrec) ~ 1, data=bc, dist="gengamma", inits=c(1,1,1), fi
pars=c(2,3)) # exponential model with rate 1/exp(mu)
```

The focus parameter is taken to be the expected survival over 8 years (the "restricted mean" survival), and the FIC statistics are calculated for the three models. The matrix ‘indmat‘ indicates which of the three parameters μ, σ, Q are included in each model.

In the ‘fic‘ method for ‘flexsurvreg‘ model objects, the parameters need to be transformed to a real-line, since ‘flexsurvreg‘ stores the covariance matrix of the parameter estimates on this scale. Therefore, since σ is defined to be positive, the focus is specified as a function of $(\boldsymbol{\theta}, \boldsymbol{\gamma}) = (\mu, \log(\sigma), Q)$ instead of (μ, σ, Q). The special values $\boldsymbol{\gamma}_0$ of the parameters $\boldsymbol{\gamma} = (\log(\sigma), Q)$ which define the ”narrow” exponential model as a special case of the ”wide” generalized gamma are also defined here as $\boldsymbol{\gamma}_0 = (0, 1)$.

```
library(fic)
focus <- function(par){
    rmst_gengamma(8, par[1], exp(par[2]), par[3])
}
indmat <- rbind(exp    = c(1,0,0),
                weib   = c(1,1,0),
                ggamma = c(1,1,1))
gamma0 <- c(0,1)
fic(gg, inds=indmat, gamma0=gamma0, focus=focus, sub=list(ex2, we2, gg))

##   vals   mods    FIC  rmse rmse.adj   bias bias.adj     se focus
## 1    A    exp 125.51 0.431    0.439 -0.428   -0.428 0.0985  4.79
## 2    A   weib  44.57 0.261    0.261 -0.233   -0.230 0.1230  4.71
## 3    A ggamma   8.99 0.128    0.128  0.000    0.000 0.1275  4.84
```

The generalized gamma model gives the estimate of 8-year survival with the lowest mean square error.

We might expect a simpler model to give more precise estimates in situations with less data. In the following example, 50 uncensored survival times are simulated from a standard exponential distribution. A generalized gamma model is fitted to it, and treated as the ”wide” model in a FIC comparison of an exponential, Weibull and gamma.

```
set.seed(1)
y <- rexp(50); cen <- rep(1,50)
gge <- flexsurvreg(Surv(y, cen) ~ 1, dist="gengamma")
fic(gge, inds=indmat, gamma0=gamma0, focus=focus)

##   vals   mods    FIC  rmse rmse.adj      bias bias.adj    se
## 1    A    exp 0.0157 0.122    0.123  1.77e-02 1.77e-02 0.121
## 2    A   weib 0.0146 0.122    0.122 -2.89e-03 0.00e+00 0.122
## 3    A ggamma 0.0152 0.122    0.122  3.93e-18 8.96e-19 0.122
```

In this case, the Weibull, rather than the generalized gamma, has the lowest MSE for the survival estimate, though the differences are tiny. Note that the FIC analysis assumes that the biggest model is true, so it will not necessarily select the true model in situations where one exists.

There is also a generalized F distribution, which generalizes these models further by including a fourth parameter P. However the FIC method will not be able to include this model in the comparison, as the special value $P = 0$, which defines the generalized gamma as a special case of the generalized F, is on the boundary of the parameter space, violating the asymptotic theory required by the method.

Spline-based models are an alternative way of defining very flexible parametric survival models, see the 'flexsurvspline' function. FIC may be possible for these models, however they would need to be set up carefully so that that smaller models are all nested within a single wide model, for example by choosing knot locations manually, and this has not been investigated.

0.1. alternative focuses

e.g. hazard ratio, cumulative hazard.

Refer to Cox regression

0.2. fic method for "survreg" models

`fic` also has a built-in method for comparing models fitted using the `survreg` function of the `survival` package. The exponential and Weibull models above can also be compared in the same way, but this time using the Weibull as the "wide" model. The focus function is again specified as the restricted mean survival over 8 years, using the `rmst_weibull` function in the **flexsurv** package. An awkward reparametrisation is necessary here. The first parameter of the Weibull distribution in `flexsurv` equals `1/exp(par[2])` where `par[2]` is the `Log(scale)` reported by `survreg`, as can be seen in `summary(we)`. The second parameter of the Weibull distribution in `flexsurv` equals `exp(par[1])`, the exponential of the `(Intercept)` parameter reported in the `survreg` summary output.

```
if (!require("survival")) stop("The `survival` package should be installed to run code in
ex <-  survreg(Surv(recyrs, censrec) ~ 1, data=bc, dist="exponential")
we <-  survreg(Surv(recyrs, censrec) ~ 1, data=bc, dist="weibull")
focus <- function(par){
    rmst_weibull(8, 1/exp(par[2]), exp(par[1]))
}
indmat <- rbind(exp    = c(1,0),
                weib   = c(1,1))
fic(we, inds=indmat, focus=focus, sub=list(ex, we))

##   vals mods   FIC  rmse rmse.adj   bias bias.adj    se focus
## 1    A  exp 3.797 0.143    0.144 0.0744   0.0744 0.123  4.79
## 2    A weib 0.326 0.124    0.124 0.0000   0.0000 0.124  4.71
```

The more flexible Weibull gives a better estimate, as we found before.