

Parametric Survival Analysis

Zachary R. McCaw

2021-07-19

Contents

- Introduction
- Simulation
- Estimation
- Group Contrasts

Introduction

Overview

This package performs estimation and inference on parametric survival curves. Supported distributions include the exponential, gamma, generalized gamma, log-normal, and Weibull. Data are expected in time-to-event format, with a status indicator to accommodate non-informative right censoring. The function `FitParaSurv` provides maximum likelihood estimates (MLEs) and confidence intervals (CIs) for the distribution of interest. Estimates are presented for model parameters, and for characteristics of the event time distribution (the mean, median, variance, and restricted mean). The function `CompParaSurv` compares the fitted survival distributions of two treatment arms with respect to the difference and ratio of means, medians, and restricted means. For each contrast, point estimates, CIs, and p-values assessing the null hypothesis of no between-group difference are presented. While by default the between-group difference is evaluated using p-values, an option is provided for obtaining permutation p-values.

Setting

Suppose the data consist of pairs (U_i, δ_i) , where the observation time $U_i = \min(T_i, C_i)$ is the minimum of an event time T_i and a censoring time C_i ; and $\delta_i = I(T_i \leq C_i)$ is an indicator that an event was observed. The event times T_i are assumed to follow a survival distribution S_T parameterized by θ . The distribution of censoring times C_i is left unspecified. Estimation of θ proceeds by maximizing the right-censored log likelihood $\ell(\theta) = \sum_{i=1}^n \delta_i \ln h_i + \ln S_i$, where S_i and h_i denote the survival and hazard contributions of the i th subject. The asymptotic covariance of the MLE $\hat{\theta}$ is estimated using the (inverse) observed information, and standard errors for functions of the MLE (e.g. means and medians) are obtained from the Δ -method.

Estimation

Overview

The function `FitParaSurv` requires a `data.frame` containing observation times `time` and status indicators `status` as input. `status` is coded as 1 if an event was observed, and as 0 if censored. The distribution of interest is specified using `dist`, which defaults to Weibull. The output is an object of class `fit`. The slot `fit@Parameters` contains the parameter estimates and CIs. The slot `fit@Information` contains the observed sample information matrix. The slot `fit@Outcome` contains characteristics of the fitted event time distribution. If truncation times `tau` are provided, the corresponding RMSTs are stored in `fit@RMST`. Functions are presented below for simulating from and estimating the exponential, gamma, generalized gamma, log-normal, and Weibull distributions. If an expected censoring proportion `p` is provided, then the event time are subject to non-informative random right censoring.

Exponential

The exponential distribution is parameterized in terms of the rate λ . The density is:

$$f(t) = \lambda e^{-\lambda t}, t > 0$$

In the following, $n = 10^3$ exponential event times are simulated, with rate $\lambda = 2$ and expected censoring proportion 20%. Generative parameters are recovered using `FitParaSurv` with `dist = "exp"`.

```
# Generate exponential event time data.
data <- GenData(n = 1e3, dist = "exp", theta = c(2), p = 0.2)

# Estimate parameters.
fit <- FitParaSurv(data, dist = "exp")
show(fit)
```

```
## Fitted Exponential Distribution.
## Estimated Parameters:
##   Aspect Estimate   SE     L     U
## 1   Rate      1.949 0.068 1.815 2.083
##
## Distribution Properties:
##   Aspect Estimate   SE     L     U
## 1   Mean      0.513 0.018 0.478 0.548
## 2   Median    0.356 0.012 0.331 0.380
## 3 Variance    0.263 0.018 0.227 0.300
```

Gamma

The gamma distribution is parameterized in terms of the shape α and rate λ . The density is:

$$f(t) = \frac{\lambda}{\Gamma(\alpha)} (\lambda t)^{\alpha-1} e^{-\lambda t}, t > 0$$

In the following, $n = 10^3$ gamma event times are simulated, with shape $\alpha = 2$, rate $\lambda = 2$, and expected censoring proportion 25%. Generative parameters are recovered using `FitParaSurv` with `dist = "gamma"`. RMST at $\tau = 0.5$ is requested.

```

# Generate gamma event time data.
data <- GenData(n = 1e3, dist = "gamma", theta = c(2, 2), p = 0.25)

# Estimate parameters.
fit <- FitParaSurv(data, dist = "gamma", tau = 0.5)
show(fit)

## Fitted Gamma Distribution.
## Estimated Parameters:
##   Aspect Estimate   SE     L     U
## 1  Shape     2.236 0.103 2.034 2.438
## 2  Rate     2.292 0.126 2.044 2.539
##
## Distribution Properties:
##   Aspect Estimate   SE     L     U
## 1   Mean     0.976 0.024 0.929 1.022
## 2  Median     0.835 0.020 0.795 0.874
## 3 Variance     0.426 0.030 0.366 0.485
##
## Restricted Mean Survival Times:
##   Tau Estimate   SE     L     U
## 1 0.5     0.453 0.003 0.447 0.459

```

Generalized Gamma

The generalized gamma distribution is parameterized in terms of shapes α and β , and rate λ . The density is:

$$f(t) = \frac{\beta\lambda}{\Gamma(\alpha)} (\lambda t)^{\alpha\beta-1} e^{-(\lambda t)^\beta}, t > 0$$

The standard gamma and Weibull distributions are nested within the generalized gamma. Setting $\beta = 1$ recovers the standard gamma, while setting $\alpha = 1$ recovers the Weibull. In the following, $n = 10^4$ generalized gamma event times are simulated, with shapes $\alpha = 2$ and $\beta = 2$, rate $\lambda = 2$, and expected censoring proportion 10%. Generative parameters are recovered using `FitParaSurv` with `dist="gen-gamma"`. Accurate estimation of the generalized gamma generally requires larger sample sizes. Fitting progress is monitored by setting `report = TRUE`.

```

set.seed(102)

# Generate generalized gamma event time data.
data <- GenData(n = 1e4, dist = "gen-gamma", theta = c(2, 2, 2), p = 0.1)

# Estimate parameters.
fit <- FitParaSurv(data, dist = "gen-gamma", report = TRUE)
show(fit)

## Objective increment: 29.5
## Objective increment: 0.103
## Objective increment: 6.59e-06
## Objective increment: 2.18e-11
## 3 update(s) performed before tolerance limit.
##
## Fitted Generalized Gamma Distribution.
## Estimated Parameters:

```

```
## Aspect Estimate SE L U
## 1 Alpha 2.020 0.146 1.734 2.306
## 2 Beta 1.996 0.083 1.833 2.160
## 3 Lambda 2.023 0.112 1.803 2.242
##
## Distribution Properties:
## Aspect Estimate SE L U
## 1 Mean 0.661 0.002 0.656 0.666
## 2 Median 0.840 0.026 0.789 0.890
## 3 Variance 0.057 0.001 0.055 0.059
```

The final parameter estimates for the generalized gamma distribution may be sensitive to the initial values. If the Newton-Raphson iteration is initialized too far from the optimum of the log-likelihood, the search may not reach the maximum. If the search halts prematurely, the fitting procedure will indicate that the information matrix was not positive definite. Supplying initial parameter values, as in the following, may help to improve the final estimates.

```
# Initialization.
fit <- FitParaSurv(
  data,
  dist = "gen-gamma",
  init = list(alpha = 2, beta = 2, lambda = 2)
)
show(fit)
```

Log-Normal

The log-normal distribution is parameterized in terms of the location μ and scale σ . The density is:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}, t > 0$$

In the following, $n = 10^3$ log-normal event times are simulated, with location $\mu = 1$, scale $\sigma = 2$, and expected censoring proportion 15%. Generative parameters are recovered using `FitParaSurv` with `dist = "log-normal"`. RMSTs at $\tau = c(5, 10, 25)$ are requested.

```
# Generate log-normal event time data.
data <- GenData(n = 1e3, dist = "log-normal", theta = c(1, 2), p = 0.15)

# Estimate parameters.
fit <- FitParaSurv(data, dist = "log-normal", tau = c(5, 10, 25))
show(fit)
```

```
## Fitted Log-Normal Distribution.
## Estimated Parameters:
## Aspect Estimate SE L U
## 1 Location 1.039 0.068 0.907 1.172
## 2 Scale 2.081 0.051 1.980 2.182
##
## Distribution Properties:
## Aspect Estimate SE L U
## 1 Mean 24.627 3.242 18.274 30.981
## 2 Median 2.827 0.191 2.452 3.201
## 3 Variance 45430.384 21041.346 4190.104 86670.665
##
```

```
## Restricted Mean Survival Times:
##   Tau Estimate   SE     L     U
## 1   5     2.832 0.059 2.717 2.947
## 2  10     4.450 0.119 4.216 4.684
## 3  25     7.398 0.279 6.852 7.945
```

Weibull

The Weibull distribution is parameterized in terms of the shape α and rate λ . The density is:

$$f(t) = \alpha\lambda(\lambda t)^{\alpha-1}e^{-(\lambda t)^\alpha}, t > 0$$

In the following, $n = 10^3$ Weibull event times are simulated, with shape $\alpha = 2$, rate $\lambda = 2$, and expected censoring proportion 30%. Generative parameters are recovered using `FitParaSurv` with `dist = "weibull"`.

```
# Generate Weibull event time data.
data <- GenData(n = 1e3, dist = "weibull", theta = c(2, 2), p = 0.3)

# Estimate parameters.
fit <- FitParaSurv(data, dist = "weibull")
show(fit)
```

```
## Fitted Weibull Distribution.
## Estimated Parameters:
##   Aspect Estimate   SE     L     U
## 1  Shape     1.994 0.060 1.877 2.111
## 2  Rate     1.961 0.037 1.888 2.035
##
## Distribution Properties:
##   Aspect Estimate   SE     L     U
## 1   Mean     0.452 0.009 0.435 0.469
## 2  Median     0.424 0.008 0.408 0.441
## 3 Variance     0.056 0.004 0.049 0.063
```

Group Contrasts

Overview

The function `CompParaSurv` takes the observation times `time`, status indicators `status`, and the treatment arms `arm` as inputs. `arm` is coded as 1 for the target group, and 0 for the reference group. The distributions for the target and reference groups are selected using `dist1` and `dist0`, respectively. If unspecified, the distribution for the reference group defaults to the distribution for the target group, and the distribution for the target group defaults to Weibull. The output is an object of class `contrast`. The slots `contrast@Model1` and `contrast@Model2` contain the fitted models for the target and reference groups. See `fit` for a description of class `fit` objects. The slot `contrast@Location` contains the location parameter contrasts, CIs, and p-values. If truncation times `tau` are specified, the slot `contrast@RMST` will contain these contrasts.

By default, all CIs and p-values are asymptotic. If the `reps` argument is supplied, then permutation p-values are calculated using that many reshuffles of the treatment assignments.

Examples

Comparison of Distinct Gammas

The target group consists of 10^3 observations from a gamma distribution with shape $\alpha = 2$, rate $\lambda = 1$, and 25% censoring. The reference group consists of 10^3 observations from a gamma distribution with the same shape ($\alpha = 2$) but different rate ($\lambda = 2$) and censoring frequency (15%). The true difference of mean survival times is 1.0, and the true ratio is 2.0. The true difference of median survival times is 0.84, and the true ratio is 2.

```
set.seed(101)

# Target group.
df1 <- GenData(n = 1e3, dist = "gamma", theta = c(2, 1), p = 0.25)
df1$arm <- 1

# Reference group.
df0 <- GenData(n = 1e3, dist = "gamma", theta = c(2, 2), p = 0.15)
df0$arm <- 0

# Overall data set.
data <- rbind(df1, df0)

# Compare fitted distributions.
comp <- CompParaSurv(data, dist1 = "gamma", dist0 = "gamma")
cat("\n")
show(comp)
```

```
##
## Contrast of Fitted Gamma Distributions.
##
## Fitted Characteristics for Group 1:
##   Aspect Estimate   SE    L    U
## 1   Mean      2.046 0.055 1.937 2.154
## 2   Median    1.685 0.046 1.595 1.775
## 3  Variance    2.306 0.175 1.963 2.649
##
## Fitted Characteristics for Group 0:
##   Aspect Estimate   SE    L    U
## 1   Mean      1.017 0.024 0.970 1.064
## 2   Median    0.859 0.021 0.818 0.899
## 3  Variance    0.502 0.034 0.436 0.567
##
## Location:
##   Contrast Point   SE    L    U P
## 1 Mean1-Mean0 1.029 0.060 0.910 1.147 0
## 2 Mean1/Mean0 2.011 0.072 1.874 2.158 0
## 3  Med1-Med0 0.826 0.050 0.728 0.925 0
## 4  Med1/Med0 1.962 0.072 1.827 2.108 0
```

Comparison of Equivalent Weibulls

Both the target and the reference groups consist of 10^3 observations from the Weibull distribution with shape $\alpha = 2$ and rate $\lambda = 2$. However, the target group is subject to 50% censoring, while the reference group is

uncensored. The true difference in means and medians is 0.0. The true ratio of means and medians is 1.0.

```
# Target group.
df1 <- GenData(n = 1e3, dist = "weibull", theta = c(2, 2), p = 0.5)
df1$arm <- 1

# Reference group.
d0 <- GenData(n = 1e3, dist = "weibull", theta = c(2, 2), p = 0.0)
d0$arm <- 0

# Overall data set.
data <- rbind(df1, df0)

# Compare fitted distributions.
comp <- CompParaSurv(data, dist1 = "weibull", dist0 = "weibull")
cat("\n")
show(comp)
```

```
##
## Contrast of Fitted Weibull Distributions.
##
## Fitted Characteristics for Group 1:
##   Aspect Estimate   SE    L    U
## 1   Mean      0.428 0.010 0.410 0.447
## 2   Median    0.403 0.009 0.385 0.420
## 3 Variance    0.050 0.004 0.041 0.058
##
## Fitted Characteristics for Group 0:
##   Aspect Estimate   SE    L    U
## 1   Mean      1.018 0.023 0.973 1.063
## 2   Median    0.889 0.022 0.845 0.933
## 3 Variance    0.464 0.028 0.408 0.519
##
## Location:
##   Contrast Point   SE    L    U P
## 1 Mean1-Mean0 -0.590 0.025 -0.639 -0.541 0
## 2 Mean1/Mean0  0.421 0.013  0.395  0.448 0
## 3 Med1-Med0   -0.486 0.024 -0.533 -0.438 0
## 4 Med1/Med0   0.453 0.015  0.425  0.484 0
```

Log-Normal v. Weibull, Different Means, Same Medians

The target group consists of 10^3 observations from a log-normal distribution with location $\mu = 0$ and scale $\sigma = \sqrt{2 \ln 2}$. The reference group consists of 10^3 observations from a Weibull distribution with shape $\alpha = 2$ and rate $\lambda = \sqrt{\ln(2)}$. In each case the expected censoring proportion is 10%. The true difference of mean survival times is 0.94, and the true ratio is 1.88. The true difference of median survival times is 0.0, and the true ratio is 1.0.

```
set.seed(105)

# Target group.
df1 <- GenData(n = 1e3, dist = "log-normal", theta = c(0, sqrt(2 * log(2))), p = 0.1)
df1$arm <- 1
```

```

# Reference group.
d0 <- GenData(n = 1e3, dist = "weibull", theta = c(2, sqrt(log(2))), p = 0.1)
d0$arm <- 0

# Overall data set.
data <- rbind(df1, df0)

# Compare fitted distributions.
comp <- CompParaSurv(data, dist1 = "log-normal", dist0 = "weibull")
cat("\n")
show(comp)

```

```

##
## Contrast of Fitted Log-Normal v. Weibull
##
## Fitted Characteristics for Group 1:
##   Aspect Estimate   SE    L    U
## 1   Mean      1.963 0.097 1.773 2.154
## 2   Median    1.024 0.038 0.950 1.098
## 3 Variance   10.308 1.741 6.895 13.721
##
## Fitted Characteristics for Group 0:
##   Aspect Estimate   SE    L    U
## 1   Mean      1.018 0.023 0.973 1.063
## 2   Median    0.889 0.022 0.845 0.933
## 3 Variance    0.464 0.028 0.408 0.519
##
## Location:
##   Contrast Point   SE    L    U    P
## 1 Mean1-Mean0 0.945 0.100 0.749 1.141 0.000
## 2 Mean1/Mean0 1.928 0.105 1.733 2.145 0.000
## 3 Med1-Med0 0.136 0.044 0.050 0.221 0.002
## 4 Med1/Med0 1.152 0.051 1.056 1.258 0.001

```

Gamma v. Exponential, Same Means, Different Medians

The target group consists of 10^3 observations from a gamma distribution with shape $\alpha = 4$ and scale $\lambda = 4$. The reference group consists of 10^3 observations from an exponential distribution with rate $\lambda = 1$. In each case the expected censoring proportion is 20%. The true difference of mean survival times is 0.0, and the true ratio is 1.0. The true difference of median survival times is 0.22, and the true ratio is 1.32. RMSTs at $\tau = c(0.5, 1.0, 1.5)$ are requested.

```

set.seed(106)

# Target group.
df1 <- GenData(n = 1e3, dist = "gamma", theta = c(4, 4), p = 0.2)
df1$arm <- 1

# Reference group.
df0 <- GenData(n = 1e3, dist = "exp", theta = c(1), p = 0.2)
df0$arm <- 0

# Overall data set.

```



```

data <- rbind(df1, df0)

# Compare fitted distributions.
comp <- CompParaSurv(data, dist1 = "gamma", dist0 = "exp", tau = c(0.5, 1.0, 1.5))
cat("\n")
show(comp)

```

```

##
## Contrast of Fitted Gamma v. Exponential
##
## Fitted Characteristics for Group 1:
##   Aspect Estimate   SE     L     U
## 1   Mean      0.993 0.017 0.960 1.027
## 2   Median    0.913 0.016 0.882 0.944
## 3  Variance   0.242 0.014 0.214 0.271
##
## Fitted Characteristics for Group 0:
##   Aspect Estimate   SE     L     U
## 1   Mean      0.984 0.035 0.915 1.053
## 2   Median    0.682 0.024 0.634 0.730
## 3  Variance   0.968 0.069 0.833 1.103
##
## Location:
##   Contrast Point   SE     L     U     P
## 1 Mean1-Mean0 0.009 0.039 -0.067 0.086 0.812
## 2 Mean1/Mean0 1.009 0.040 0.934 1.091 0.812
## 3  Med1-Med0 0.231 0.029 0.174 0.288 0.000
## 4  Med1/Med0 1.339 0.053 1.239 1.447 0.000
##
## RMST:
##   Tau   Contrast Point   SE     L     U P
## 1.1 0.5 RMST1-RMST0 0.089 0.004 0.082 0.097 0
## 1.2 0.5 RMST1/RMST0 1.228 0.011 1.207 1.250 0
## 2.1 1.0 RMST1-RMST0 0.176 0.012 0.152 0.199 0
## 2.2 1.0 RMST1/RMST0 1.280 0.023 1.236 1.325 0
## 3.1 1.5 RMST1-RMST0 0.168 0.020 0.129 0.208 0
## 3.2 1.5 RMST1/RMST0 1.219 0.030 1.162 1.279 0

```