Statistical methodologies for goodness-of-fit: a comparative analysis of three established approaches

M. Francisco, K. Langohr,M. Besalú, G. Gómez Melis

9^{th} GRBIO retreat

July 11^{th}, 2024

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

GRBIO

Table of Contents

Table of Contents

Goodness-of-fit tests

Goodness-of-fit tests can be applied: available for both complete and censored data.

- Although non-parametric approaches are commonly used for this kind of data, parametric approaches also play an important role in survival analysis.
- Adaptations of classic goodness-of-fit tests to censored data are available.

Tests based on the empirical distribution function

Test the hypothesis:

$H_0 : F(t) = F_0(t)$
$\mathrm{H}_1 : F(t) \neq F_0(t)$, for all $t \geq 0$

Most common ones:

- Kolmogorov-Smirnov
- Cramér-von Mises
- Anderson-Darling

When applying these tests to censored data, the empirical distribution function is replaced by an appropriate estimate of the distribution function.

Kolmogorov-Smirnov

Adaptation proposed by Fleming *et al.* given by:

$$D_n = \sup_t |\hat{F}_n(t) - F_0(t)|$$

Where:

- $\hat{F}_n(t)$ is the estimation of the empirical distribution function of the data.
- n is the sample size.

For **right-censored data** $\hat{F}_n$, is replaced by $\hat{F}_n = 1 - \hat{S}_n = 1 - e^{-\hat{\Lambda}_n}$, where $\hat{\Lambda}_n$ denotes the Nelson-Aalen estimator of the cumulative hazard function.

Kolmogorov-Smirnov

As a result:

$$\hat{D}_n = \sup_{0 \le t \le t_m} |\hat{F}_n(t) - F_0(t)| = \sup_{0 \le t \le t_m} \left| \int_0^t \frac{\hat{S}_n(t) S_0(s)}{\hat{S}_n(s)} d\left[\hat{\Lambda}_n(s) - \Lambda_0(s)\right] \right|,$$

where S_0 and Λ_0 are, respectively, the survival and the cumulative hazard function of the hypothesized distribution and t_m is the largest observed time in the sample.

Cramér-von Mises

The Cramér-von Mises statistic is given by:

$$M_n = n \int_{-\infty}^{+\infty} \left(\hat{F}_n(t) - F_0(t)\right)^2 dF_0(t),$$

where $\hat{F}_n$ is:

- The empirical distribution function when data is complete.
- 1 minus the Kaplan-Meier estimator of the survival function $(\hat{F}_n = 1 - \hat{S}_n)$ if the data has random right-censorship.

Cramér-von Mises

When data has right-censored observations:

$$\hat{M}_n = n_r \sum_{j=1}^{n_r+1} \hat{F}_n(u_{(j-1)})(u_{(j)} - u_{(j-1)}) \left(\hat{F}_n(u_{(j-1)}) - (u_{(j)} + u_{(j-1)}) \right) + \frac{n_r}{3},$$

where, $Y_1, \ldots, Y_{n_r}$ are the n_r observed failure times and $u_{(i)} = F_0(Y_{(i)})$ if we transform the order statistic $Y_{(1)}, \ldots, Y_{(n_r)}$ into Uniform$(0,1)$ random variables $u_{(1)}, \ldots, u_{(n_r)}$.

The asymptotic distribution of $\hat{M}_n$ is not easily implemented.

Anderson-Darling

Anderson-Darling statistics:

$$A_n = n \int_{-\infty}^{+\infty} (\hat{F}_n(t) - F_0(t))^2 \frac{dF_0(t)}{F_0(t)(1 - F_0(t))}.$$

where:

- $\hat{F}_n$ is the empirical distribution function when data are complete.
- $\hat{F}_n$ is 1 minus the Kaplan-Meier estimator of the survival function $(\hat{F}_n = 1 - \hat{S}_n)$ if the data has random right-censorship.

Anderson-Darling

When data has right-censored observations:

$$\hat{A}_n = -n_r + n_r \sum_{j=1}^{n_r} (\hat{F}_n(u_{(j-1)}) - 1)^2 \left[\log|1 - u_{(j-1)}| - \log|1 - u_{(j)}|\right] + n_r \sum_{j=1}^{n_r - 1} \hat{F}_n^2(u_{(j)}) \left[\log|u_{(j+1)}| - \log|u_{(j)}|\right] - n_r \log|u_{(n)}|.$$

The asymptotic distribution of $\hat{A}_n$ is not easily implemented.

Table of Contents

GofCens package

Goodness-of-Fit Methods for Complete and Right-Censored Data

```
install.packages("Gofcens")
library(GofCens)
```

Graphical tools and goodness-of-fit tests for complete and right-censored data.

Goodness-of-fit test:

- *KScens*
- *CvMcens*
- *ADcens*
- *chisqcens*

Graphical tools:

- *kmPlot*
- *probPlot*
- *cumhazPlot*

Example

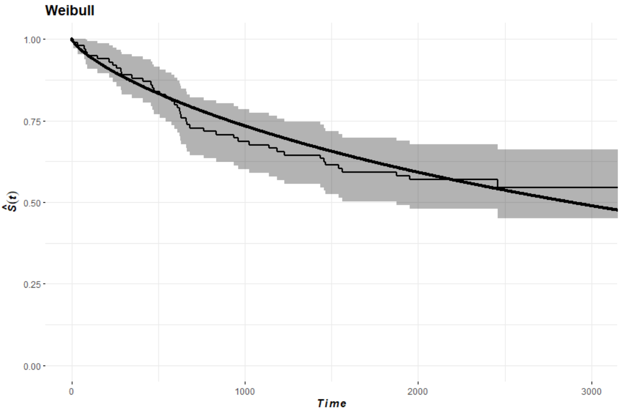
```
> library(survival)
> colonsamp <- colon[sample(nrow(colon), 100), ]
> head(colonsamp)
```

id	time	status
293	2288	1
104	614	1
446	1106	1
863	751	1
124	2862	0
42	3030	0

- $H_0 : X \sim Weibull$

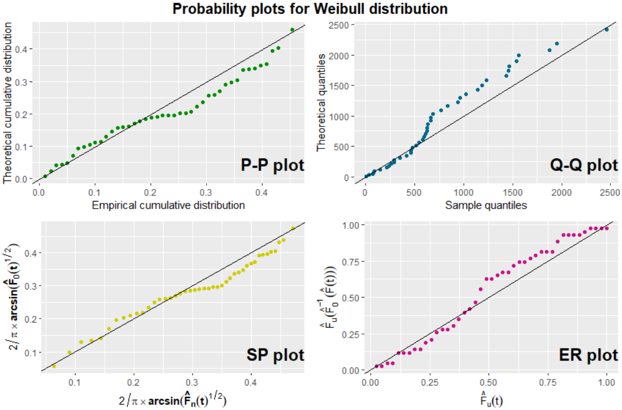
GofCens: kmPlot

- Generates a plot that combines a **Kaplan-Meier survival curve** and a **parametric survival curve** in the same graph.



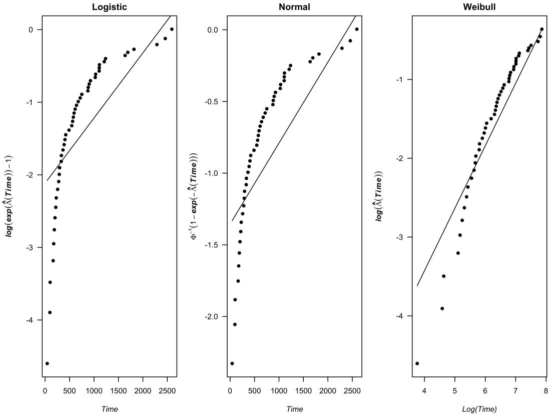
GofCens: probPlot

- Provides four types of probability plots: **P-P plot**, **Q-Q plot**, **Stabilised probability plot** and **Empirically Rescaled plot**.



GofCens: cumhazPlot

- Cumulative hazard plot to check if a certain distribution is an appropriate choice for the data.



GofCens

Null Distribution
Test statistic
p-value
Maximum likelihood estimates of the parameters of the distribution under study

```
> KScens(colonsamp$time, colonsamp$status, distr = "weibull")
Distribution: weibull

KS Test results:
      A  p-value    F(ym)       ym
  0.637    0.627    0.491 3238.000

Parameter estimates:
  shape    scale
  0.762 4659.238
```

GofCens

```
> ADcens(colonsamp$time, colonsamp$status, distr = "weibull")
Distribution: weibull

AD Test results:
      AD p-value
 10.343   0.454

Parameter estimates:
   shape    scale
   0.762 4659.238
```

```
> CvMcens(colonsamp$time, colonsamp$status, distr = "weibull")
Distribution: weibull

CvM Test results:
     CvM p-value
   2.349   0.206

Parameter estimates:
   shape    scale
   0.762 4659.238
```

GofCens

```
> gofcens(colonsamp$time, colonsamp$status, distr = "weibull")
Distribution: weibull

Test statistics
    KS     CvM      AD
 0.637   2.349 10.343

p-values
   KS   CvM    AD
0.627 0.174 0.457

Parameter estimates:
   shape     scale
   0.762 4659.238
```

- For the tests Cramér-von Mises, Anderson-Darling and Chi-square, the asymptotic distributions of goodness-of-fit statistics are difficult to obtain/implement in the presence of censored data.
- The p-values associated with these tests are obtained via **bootstrap** methods.

Bootstrap methods for right-censored data

$H_0 : F(t) = F_0(t; \theta)$

1. Observed data is utilized to estimate the parameter θ, denoted as $\hat{\theta}_n$, using maximum likelihood estimation
2. Generation of B independent bootstrap samples of the same size (n) as the original data set:
   - Generation of survival times $T_1^b, \dots, T_n^b$ from the fitted distribution $F_0(t; \hat{\theta}_n)$.
   - Generation of censoring times $C_1^b, \dots, C_n^b$ from the nonparametric estimation of H obtained with the Kaplan-Meier estimator.
   - Generation of observed survival times $Y_i^b = \min(T_i^b, C_i^b)$, and event indicators $\delta_i^b = \mathbf{1}\{T_i^b \leq C_i^b\}, i = 1, \dots n$

Bootstrap methods for right-censored data

- Maximum likelihood estimation of the parameter, $\hat{\theta}_n^b$, given $(Y_i^b, \delta_i^b), i = 1, \ldots n$.
- Computation of the test statistic, $(\hat{G}_n^{\hat{\theta}_n})_b$.

3. Repetition of this process for many bootstrap samples (default B=999)
4. Sequence of bootstrap statistics, $(\hat{G}_n^{\hat{\theta}_n})_b$, $b = 1, \cdots, B$, represents the empirical distribution of the statistic under the null hypothesis
5. p value is the proportion of bootstrap statistic values $(\hat{G}_n^{\hat{\theta}_n})_b$ greater than or equal to the observed statistic $\hat{G}_n$

Motivation

What is the most correct test to apply?
What does it depend on?

Table of Contents

Simulation study

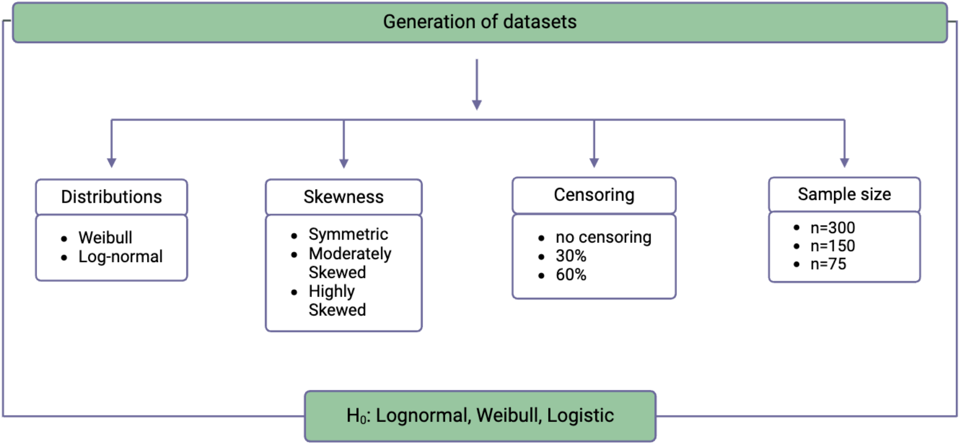
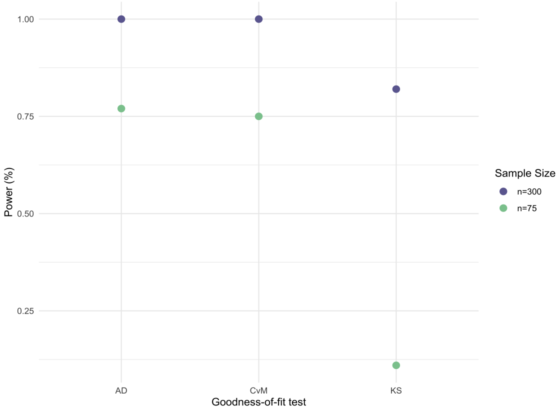


Table of Contents

Preliminary results

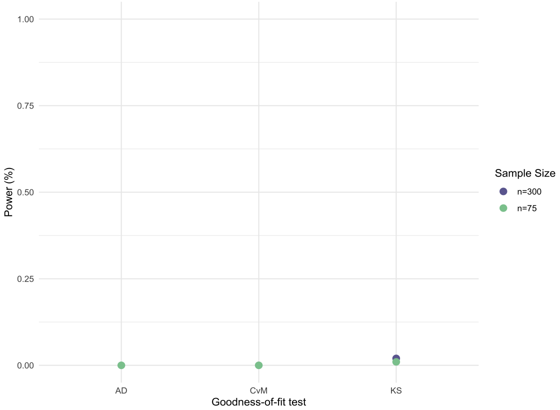
- Highly skewed Log-normal | Complete data
- $H_0 : X \sim Weibull$



- Anderson-Darling and Cramér-von Mises have good power regardless of sample size.
- Kolmogorov-Smirnov has a good power for big sample size.
- Increase of sample size leads to increase of power.

Preliminary results

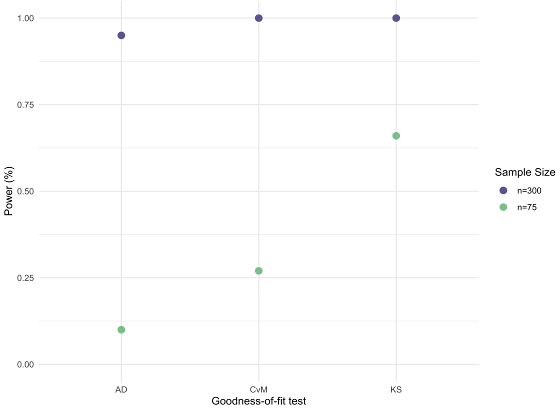
- Highly skewed Log-normal | 60 % censored data
- $H_0 : X \sim Weibull$



- Power results are not adequate regardless of the test applied.

Preliminary results

- Highly skewed Log-normal | 60 % censored data
- $H_0 : X \sim Logistic$



- For large sample sizes, the tests produce good results.
- For a small sample size, Kolmogorov-Smirnov is the one that behaves better.

References

K. Langohr, M. Besalú, M. Francisco, G. Gómez, GofCens: Goodness-of-Fit Methods for Complete and Right-Censored Data, R package version 0.98 (2024).

URL https://CRAN.R-project.org/package=GofCens T. R. Fleming, J. R. O'Fallon, P. C. O'Brien, D. P. Harrington, Modified kolmogorov-smirnov test procedures with application to arbitrarily

right-censored data, Biometrics (1980) 607–625. J. A. Koziol, S. B. Green, A cramér-von mises statistic for randomly censored data, Biometrika 63 (3) (1976) 465–474.

A. N. Pettitt, M. A. Stephens, Modified cramér-von mises statistics for censored data, Biometrika 63 (2) (1976) 291–298

Thank you for the attention!

Gracias por la atención!

Gràcies per l'atenció!