Reliability
Part 1: Lifetime Data Analysis (cont'd)

Jean-Yves Dauxois

INSA-Toulouse & IMT, FRANCE

Plan

Plan

Model Checking

Model Checking

Let $\mathfrak{F}$ be a parametric model of distributions (supposed to be continuous in this section) with parameter θ in $\Theta \subset \mathbb{R}^n$, i.e.

$$\mathfrak{F} = \{R_\theta(\cdot) : \theta \in \Theta\}.$$

Model Checking

Let $\mathfrak{F}$ be a parametric model of distributions (supposed to be continuous in this section) with parameter θ in $\Theta \subset \mathbb{R}^n$, i.e.

$$\mathfrak{F} = \{R_\theta(\cdot) : \theta \in \Theta\} .$$

Suppose that we observe a sample (T_i, δ_i), for $i = 1, \ldots, n$.

Question. Does the model $\mathfrak{F}$ fit the data?

Model Checking

Let $\mathfrak{F}$ be a parametric model of distributions (supposed to be continuous in this section) with parameter θ in $\Theta \subset \mathbb{R}^n$, i.e.

$$\mathfrak{F} = \{R_\theta(\cdot) : \theta \in \Theta\}.$$

Suppose that we observe a sample (T_i, δ_i), for $i = 1, \ldots, n$.

Question. Does the model $\mathfrak{F}$ fit the data?

We will consider 2 different types of response.

- $\Rightarrow$ Graphical method.

Model Checking

Let $\mathfrak{F}$ be a parametric model of distributions (supposed to be continuous in this section) with parameter θ in $\Theta \subset \mathbb{R}^n$, i.e.

$$\mathfrak{F} = \{R_\theta(\cdot) : \theta \in \Theta\}.$$

Suppose that we observe a sample (T_i, δ_i), for $i = 1, \ldots, n$.

Question. Does the model $\mathfrak{F}$ fit the data?

We will consider 2 different types of response.

- $\Rightarrow$ Graphical method.
- $\Rightarrow$ Statistical tests.

Model Checking

Graphical method

Graphical method: Hazard-Plotting

As an example, suppose that we want to see if the Weibull distribution is a good model. Remind that:

$$\mathfrak{F} = \{R_\theta(x) = e^{-\left(\frac{x}{\alpha}\right)^\beta}, \text{ where } \theta = (\alpha, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+\}.$$

Graphical method: Hazard-Plotting

As an example, suppose that we want to see if the Weibull distribution is a good model. Remind that:

$$\mathfrak{F} = \{R_\theta(x) = e^{-\left(\frac{x}{\alpha}\right)^\beta}, \text{ where } \theta = (\alpha, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+\}.$$

Recall that we have, for continuous r.v., the relation:

$$R_\theta(x) = \exp(-\Lambda_\theta(x)).$$

Graphical method: Hazard-Plotting

As an example, suppose that we want to see if the Weibull distribution is a good model. Remind that:

$$\mathfrak{F} = \{R_\theta(x) = e^{-\left(\frac{x}{\alpha}\right)^\beta}, \text{ where } \theta = (\alpha, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+\}.$$

Recall that we have, for continuous r.v., the relation:

$$R_\theta(x) = \exp(-\Lambda_\theta(x)).$$

Thus:

$$\begin{aligned} \Lambda_\theta(x) &= \left(\frac{x}{\alpha}\right)^\beta \\ \Longleftrightarrow \ln x &= \frac{1}{\beta}\ln(\Lambda_\theta(x)) + \ln \alpha. \end{aligned}$$

So we have:

$$\ln x = \frac{1}{\beta}\ln(\Lambda_\theta(x)) + \ln\alpha.$$

So we have:

$$\ln x = \frac{1}{\beta} \ln(\Lambda_\theta(x)) + \ln \alpha.$$

Remind that the two parameters α and β and also the cumulative hazard rate function $\Lambda_\theta(\cdot)$ are **unknown**.

So we have:

$$\ln x = \frac{1}{\beta}\ln(\Lambda_\theta(x)) + \ln\alpha.$$

Remind that the two parameters α and β and also the cumulative hazard rate function $\Lambda_\theta(\cdot)$ are **unknown**.

Using the Nelson-Aalen estimator $\widehat{\Lambda}(\cdot)$ of this later function, the idea is to plot the points

$$(\ln(\widehat{\Lambda}(T_i)), \ln(T_i))_{i=1,\ldots,n}$$

and to see if they tend to be aligned. In this case, if n is sufficiently large, one can admit the use the Weibull model on the dataset.

The slope of the line gives us an estimator of $1/\beta$ and its intercept an estimator of $\ln\alpha$, from which we can derive estimators of β and α.

We can follow the same idea for other models. More precisely, one has to plot the points:

- $(\widehat{\Lambda}_n(T_i), T_i)_{i=1,\ldots,n}$ for the exponential distribution;
- $(\varphi^{-1}(1 - \exp(-\widehat{\Lambda}_n(T_i))), \ln(T_i))_{i=1,\ldots,n}$ for the lognormal distribution,

where $\varphi(\cdot)$ is still the c.d.f. of the $N(0, 1)$ distribution.

Examples of application: open hazard_plotting.pdf

⚠

- Disadvantage. This is not a statistical test!

We can follow the same idea for other models. More precisely, one has to plot the points:

- $(\widehat{\Lambda}_n(T_i), T_i)_{i=1,\ldots,n}$ for the exponential distribution;
- $(\varphi^{-1}(1 - \exp(-\widehat{\Lambda}_n(T_i))), \ln(T_i))_{i=1,\ldots,n}$ for the lognormal distribution,

where $\varphi(\cdot)$ is still the c.d.f. of the $N(0,1)$ distribution.

Examples of application: open hazard_plotting.pdf

⚠

- Disadvantage. This is not a statistical test! $\Rightarrow$ **No control of probability of error**.

We can follow the same idea for other models. More precisely, one has to plot the points:

- $(\widehat{\Lambda}_n(T_i), T_i)_{i=1,\ldots,n}$ for the exponential distribution;
- $(\varphi^{-1}(1 - \exp(-\widehat{\Lambda}_n(T_i))), \ln(T_i))_{i=1,\ldots,n}$ for the lognormal distribution,

where $\varphi(\cdot)$ is still the c.d.f. of the $N(0,1)$ distribution.

Examples of application: open hazard_plotting.pdf

⚠

- Disadvantage. This is not a statistical test! $\Rightarrow$ **No control of probability of error**.
- Advantage. In addition of an idea of goodness-of-fit, we have an **estimation of the parameters**.

Model Checking

Goodness-of-fit-tests

Goodness-of-fit-**tests** with uncensored observations

Problem. One wants to test the null hypothesis $H_0 : F = F_0$, where F_0 is a specified c.d.f., against $H_1 : F \neq F_0$.

Goodness-of-fit-**tests** with uncensored observations

Problem. One wants to test the null hypothesis $H_0 : F = F_0$, where F_0 is a specified c.d.f., against $H_1 : F \neq F_0$.

Example:

- One wants to test if a $W(\alpha, \beta)$ distribution could fit the data.

Goodness-of-fit-**tests** with uncensored observations

Problem. One wants to test the null hypothesis $H_0 : F = F_0$, where F_0 is a specified c.d.f., against $H_1 : F \neq F_0$.

Example:

- One wants to test if a $W(\alpha, \beta)$ distribution could fit the data.
- The parameters α and β are **given or estimated** from the data.

Goodness-of-fit-**tests** with uncensored observations

Problem. One wants to test the null hypothesis $H_0 : F = F_0$, where F_0 is a specified c.d.f., against $H_1 : F \neq F_0$.

Example:

- One wants to test if a $W(\alpha, \beta)$ distribution could fit the data.
- The parameters α and β are **given or estimated** from the data. $\Rightarrow$ In the later case, some modifications are necessary.

Goodness-of-fit-**tests** with uncensored observations

Problem. One wants to test the null hypothesis $H_0 : F = F_0$, where F_0 is a specified c.d.f., against $H_1 : F \neq F_0$.

Example:

- One wants to test if a $W(\alpha, \beta)$ distribution could fit the data.
- The parameters α and β are **given or estimated** from the data. $\Rightarrow$ In the later case, some modifications are necessary.

Main idea: Compare the tested distribution F_0 to the estimated cumulative distribution function $\hat{F} = 1 - \hat{R}$.

Goodness-of-fit-**tests** with uncensored observations

Problem. One wants to test the null hypothesis $H_0 : F = F_0$, where F_0 is a specified c.d.f., against $H_1 : F \neq F_0$.

Example:

- One wants to test if a $W(\alpha, \beta)$ distribution could fit the data.
- The parameters α and β are **given or estimated** from the data. $\Rightarrow$ In the later case, some modifications are necessary.

Main idea: Compare the tested distribution F_0 to the estimated cumulative distribution function $\hat{F} = 1 - \hat{R}$. Different measures are available.

Goodness-of-fit-**tests** with uncensored observations

Problem. One wants to test the null hypothesis $H_0 : F = F_0$, where F_0 is a specified c.d.f., against $H_1 : F \neq F_0$.

Example:

- One wants to test if a $W(\alpha, \beta)$ distribution could fit the data.
- The parameters α and β are **given or estimated** from the data. $\Rightarrow$ In the later case, some modifications are necessary.

Main idea: Compare the tested distribution F_0 to the estimated cumulative distribution function $\hat{F} = 1 - \hat{R}$. Different measures are available.$\Rightarrow$ Different tests.

One can use, among others:

1. the χ^2 test;
2. the Kolmogorov-Smirnov test with statistic

$$D_n = \sup_{x \geqslant 0} \left| \widehat{F}_n(x) - F_0(x) \right|;$$

3. the Cramer-von Mises test with statistic

$$W_n^2 = n \int_{\mathbb{R}^+} (\widehat{F}_n(x) - F_0(x))^2 dF_0(x);$$

4. the Anderson-Darling test with statistic

$$A_n^2 = n \int_{\mathbb{R}^+} \frac{(\widehat{F}_n(x) - F_0(x))^2}{F_0(x) R_0(x)} dF_0(x).$$

The previous tests are exact or asymptotic tests.

- **Exact tests**. The result is correct for fixed sample size n.

The previous tests are exact or asymptotic tests.

- **Exact tests**. The result is correct for fixed sample size n.
  $\Rightarrow$ The distribution of the test statistic (or critical value) is known or can be approximated.

The previous tests are exact or asymptotic tests.

- **Exact tests**. The result is correct for fixed sample size n.
  $\Rightarrow$ The distribution of the test statistic (or critical value) is known or can be approximated.
- **Asymptotic tests**. The results are valid when $n \to +\infty$.

The previous tests are exact or asymptotic tests.

- **Exact tests**. The result is correct for fixed sample size n.
  $\Rightarrow$ The distribution of the test statistic (or critical value) is known or can be approximated.
- **Asymptotic tests**. The results are valid when $n \to +\infty$.
  $\Rightarrow$ So we can apply the result for only large values of n.

The previous tests are exact or asymptotic tests.

- **Exact tests**. The result is correct for fixed sample size n. $\Rightarrow$ The distribution of the test statistic (or critical value) is known or can be approximated.
- **Asymptotic tests**. The results are valid when $n \to +\infty$. $\Rightarrow$ So we can apply the result for only large values of n.

⚠ Remind that for the exact test like the χ^2 one, the distribution of the test statistic (or the critical value) has to be modified **if the parameters are** not given but **estimated**.

Goodness-of-fit-tests with censored observations

When there is **censoring** the idea is to adapt the previous nonparametric tests with replacement of the empirical distribution $\hat{F}_n$ by the **Kaplan-Meier estimator** $1 - \hat{R}_n$.

For example the exact or asymptotic distribution of

$$W_n^2 = n \int_0^c (\widehat{F}_n(x) - F_0(x))^2 dF_0(x),$$

where c is fixed, is known under some conditions.

Goodness-of-fit-tests with censored observations

When there is **censoring** the idea is to adapt the previous nonparametric tests with replacement of the empirical distribution $\hat{F}_n$ by the **Kaplan-Meier estimator** $1 - \hat{R}_n$.

For example the exact or asymptotic distribution of

$$W_n^2 = n \int_0^c (\widehat{F}_n(x) - F_0(x))^2 dF_0(x),$$

where c is fixed, is known under some conditions.

Based on the same idea, there exists some tests with idea to compare the cumulative hazard rate function $\Lambda(\cdot)$ to its Nelson-Aalen estimate $\hat{\Lambda}(\cdot)$.

It is also possible to use the test of Hollander and Proschan (1979). Let

$$S_n = \sum_{i=1}^{n} R_0(t_{(i)}) \Delta \widehat{F}_n(t_{(i)}),$$

where $R_0(\cdot) = 1 - F_0(\cdot)$.
One can show that, under H_0, the statistic $S_n^* = \sqrt{n}(S_n - 1/2)/\widehat{\sigma}_n$ is asymptoticallly $N(0,1)$ distributed, where $\widehat{\sigma}_n$ is defined by

$$\widehat{\sigma}_n^2 = \frac{1}{16} \sum_{i=1}^{n} \frac{n}{n-i+1} (R_0^4(t_{(i-1)}) - R_0^4(t_{(i)}))$$

and is an estimator of the standard deviation of S_n. The critical region of the test is:

$$\text{reject } H_0 \Leftrightarrow |S_n^*| > z_{1-\frac{\alpha}{2}}.$$

Another possibility is to use the hazard rate and its Nelson-Aalen estimate.
Let us first rewrite the two hypotheses in term of hazard rate. We want to test

$$H_0 : \lambda(\cdot) = \lambda_0(\cdot) \text{ against } H_1 : \lambda(\cdot) \neq \lambda_0(\cdot).$$

From the asymptotic properties of the Nelson-Aalen estimator, one case consider the test statistic:

$$\frac{\sum_{i=1}^k W(T'_i)\frac{M_i}{Y_i} - \int_0^{T'_k} W(s)\lambda_0(s)ds}{\sqrt{\int_0^{T'_k} W^2(s)\frac{\lambda_0(s)}{Y(s)}ds}},$$

where $W(s)$ is a weight function and $Y(s)$, the number at risk process, is the step function such that $Y(T'_i) = Y_i$. This statistic is asymptotically standard normal distributed.

Model Checking

Tests for several samples

Tests for several samples

Suppose that we want to compare the distributions of k different samples, with respective Reliability functions $R_1, \ldots, R_k$.

Tests for several samples

Suppose that we want to compare the distributions of k different samples, with respective Reliability functions $R_1, \ldots, R_k$.

Example: Different technologies of Lightning Arrester (ZNO or the oldest SIC one), different usage groups (uratyp, ur222, ur72, ur360, etc...).

Tests for several samples

Suppose that we want to compare the distributions of k different samples, with respective Reliability functions $R_1, \ldots, R_k$.

Example: Different technologies of Lightning Arrester (ZNO or the oldest SIC one), different usage groups (uratyp, ur222, ur72, ur360, etc...).

$\Rightarrow$ Is the distribution the same among all the different subpopulations?

Tests for several samples

Suppose that we want to compare the distributions of k different samples, with respective Reliability functions $R_1, \ldots, R_k$.

Example: Different technologies of Lightning Arrester (ZNO or the oldest SIC one), different usage groups (uratyp, ur222, ur72, ur360, etc...).

$\Rightarrow$ Is the distribution the same among all the different subpopulations? Do the different types of material have the same behavior along time?

Tests for several samples

Suppose that we want to compare the distributions of k different samples, with respective c.d.f. $F_1, \ldots, F_k$.
Thus, we want to test $H_0 : F_1 = F_2 = \cdots = F_k$ against at least one of the distributions differs from the others.

Let $t_1 < t_2 < \cdots < t_D$ the distinct observed times in the pooled sample. For $j = 1, \ldots, K$ and $i = 1, \ldots, D.$, write:

- d_{ij} the number of failures (i.e. not censored) observed at time t_i in population j,
- y_{ij} the number of individuals at risk just before time t_i in population j,
- $d_i = \sum_{j=1}^{K} d_{ij}$ the number of failures (i.e. not censored) observed at time t_i in the pooled sample,
- $y_i = \sum_{j=1}^{K} y_{ij}$ the number of individuals at risk just before time t_i in the pooled sample.

To compare the estimates of the hazard rate function in the subpopulations with its estimate in the pooled sample, one can consider the statistics

$$Z_j = \sum_{i=1}^{D} W_j(t_i) \left(\frac{d_{ij}}{y_{ij}} - \frac{d_i}{y_i} \right), \text{ pour } j = 1, \ldots, K,$$

where the $W_j(\cdot) \geq 0$ are weight functions.

An usual approach is to consider weight functions of the form

$$W_j(t_i) = y_{ij} W(t_i),$$

where $W(\cdot)$ is a common weight function. In this case, one has:

$$Z_j = \sum_{i=1}^{D} W(t_i) \left(d_{ij} - y_{ij} \frac{d_i}{y_i} \right), \text{ pour } j = 1, \ldots, K.$$

Note that, under H_0, the statistics Z_j are asymptotically close to 0.

Since $\sum_{j=1}^{K} Z_j = 0$, one can consider the statistic

$$S_K = (Z_1, \ldots, Z_{K-1})\hat{\Sigma}^{-1}(Z_1, \ldots, Z_{K-1})'$$

where $\hat{\Sigma}$ is the covariance matrix of the random vector $(Z_1, \ldots, Z_{K-1})$.

One can show that the asymptotical distribution of S_K is a $\chi^2(K-1)$. And the test reject H_0 for large values of S_K.

Many choices are possible for the weight function:

- with $W(\cdot) = 1$, we obtain the Logrank test,
- with $W(t_i) = Y_i$, this is the Gehan-Wilcoxon test.

In the case of $K = 2$ subpopulations, one can also consider the statistic

$$S = \frac{\sum_{i=1}^{D} W(t_i)\left(d_{i1} - y_{i1}\frac{d_i}{y_i}\right)}{\sqrt{\sum_{i=1}^{D} W^2(t_i)\frac{y_{i1}}{y_i}\left(1 - \frac{y_{i1}}{y_i}\right)\left(\frac{y_i - d_i}{y_i - 1}\right) d_i}}$$

which has an asymptotic standard normal distribution.

Plan

The problem of small datasets

Estimators of Mean or Median Ranks

Mean or median ranks estimators

Let $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ be the order statistics of the sample $X_1, \ldots, X_n$ and $x_{(1)}, \ldots, x_{(n)}$ their observation.

Mean or median ranks estimators

Let $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ be the order statistics of the sample $X_1, \ldots, X_n$ and $x_{(1)}, \ldots, x_{(n)}$ their observation.

Remind that we have estimated previously the $F(x_{(i)})$ by

$$\widehat{F}(x_{(i)}) = \frac{i}{n}$$

Mean or median ranks estimators

Let $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ be the order statistics of the sample $X_1, \ldots, X_n$ and $x_{(1)}, \ldots, x_{(n)}$ their observation.

Remind that we have estimated previously the $F(x_{(i)})$ by

$$\widehat{F}(x_{(i)}) = \frac{i}{n} = \frac{\#\{\text{values lower than } x_{(i)}\}}{\text{sample size}},$$

for all $i = 1, \ldots, n$.

Mean or median ranks estimators

Let $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ be the order statistics of the sample $X_1, \ldots, X_n$ and $x_{(1)}, \ldots, x_{(n)}$ their observation.

Remind that we have estimated previously the $F(x_{(i)})$ by

$$\widehat{F}(x_{(i)}) = \frac{i}{n} = \frac{\#\{\text{values lower than } x_{(i)}\}}{\text{sample size}},$$

for all $i = 1, \ldots, n$.

This estimation is called **empirical estimation**.

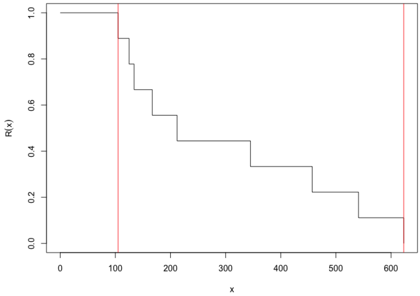

Figure: Competing risks dataset. Empirical estimation of the Reliability function ignoring the type of failure.

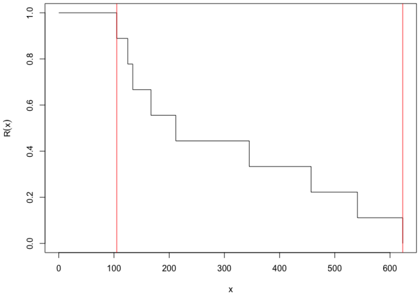

Figure: Competing risks dataset. Empirical estimation of the Reliability function ignoring the type of failure.

Problem. With this empirical estimation, we estimate that the probability of failure outside the range of observations (here $[105, 623]$) is null...

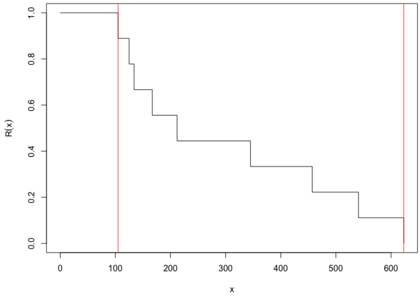

Figure: Competing risks dataset. Empirical estimation of the Reliability function ignoring the type of failure.

Problem. With this empirical estimation, we estimate that the probability of failure outside the range of observations (here $[105, 623]$) is null... $\Rightarrow$ Might appear unreasonable.

Aim. Try to find alternative estimates of our previous estimation of $F(x_{(i)})$ by

$$\widehat{F}(x_{(i)}) = \frac{i}{n}.$$

Aim. Try to find alternative estimates of our previous estimation of $F(x_{(i)})$ by

$$\widehat{F}(x_{(i)}) = \frac{i}{n}.$$

Note that the $F(x_{(i)})$, for $i = 1, \ldots, n$, are realizations of the r.v. $F(X_{(i)})$, for $i = 1, \ldots, n$.

Aim. Try to find alternative estimates of our previous estimation of $F(x_{(i)})$ by

$$\widehat{F}(x_{(i)}) = \frac{i}{n}.$$

Note that the $F(x_{(i)})$, for $i = 1, \ldots, n$, are realizations of the r.v. $F(X_{(i)})$, for $i = 1, \ldots, n$.

Deriving the p.d.f. of the r.v. $F(X_{(i)})$ will help us to introduce other estimators of the $F(x_{(i)})$, for $i = 1, \ldots, n$.

One knows that the p.d.f. of the ith order statistics is:

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!}(F(x))^{i-1}f(x)(1-F(x))^{n-i}.$$

One knows that the p.d.f. of the ith order statistics is:

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!}(F(x))^{i-1}f(x)(1-F(x))^{n-i}.$$

Since $F(\cdot)$ is increasing, we have:

$$F(X_{(i)}) = (F(X))_{(i)}, \text{ pour } i = 1, \ldots, n.$$

One knows that the p.d.f. of the ith order statistics is:

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!}(F(x))^{i-1}f(x)(1-F(x))^{n-i}.$$

Since $F(\cdot)$ is increasing, we have:

$$F(X_{(i)}) = (F(X))_{(i)}, \text{ pour } i = 1, \ldots, n.$$

Since $(F(X_i))_{i=1,\ldots,n}$ are iid r.v. with uniform distribution on $[0,1]$, the p.d.f. of the r.v. $F(X_{(i)})$ is

$$f_{F(X)_{(i)}}(y) = \frac{n!}{(i-1)!(n-i)!}y^{i-1}(1-y)^{n-i}.$$

This shows that the r.v. $F(X_{(i)})$ has a **beta** distribution with parameters i and $n-i+1$.

Using the mean of the distribution of $F(X_{(i)})$ as an estimator of $F(x_{(i)})$ we obtain the **Mean ranks estimates** (also known as Johnson's method):

$$\widehat{F(x_{(i)})} = \frac{i}{n+1},$$

for $i = 1, \ldots, n$.

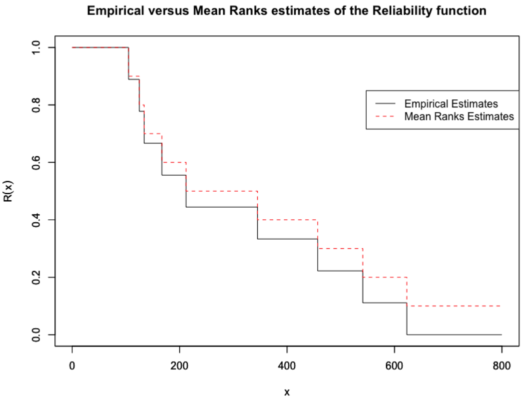

Figure: Competing risks dataset. Empirical estimates and Mean Ranks estimates of the Reliability function ignoring the type of failure.

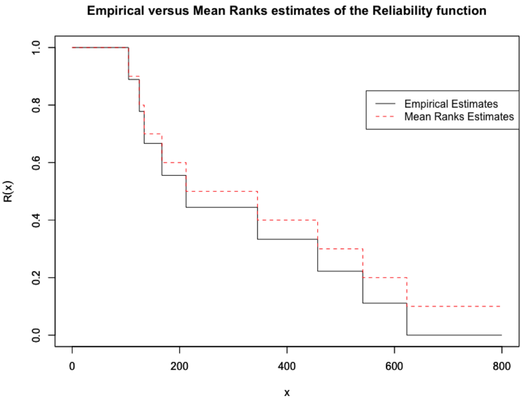

Figure: Competing risks dataset. Empirical estimates and Mean Ranks estimates of the Reliability function ignoring the type of failure.

Remark. Still a problem for small values. $\Rightarrow$ The probability to have failures before the smallest observed time (here 105) is estimated to be null.

Using the median yields to the **Median Rank estimates**:

$$\widehat{F(x_{(i)})} = \widetilde{y}_i$$

where $\widetilde{y}_i$ is the median of the distribution of $F_{F(X_{(i)})}$, i.e. such that

$$F_{F(X_{(i)})}(\widetilde{y}_i) = 1/2.$$

Using the median yields to the **Median Rank estimates**:

$$\widehat{F(x_{(i)})} = \widetilde{y}_i$$

where $\widetilde{y}_i$ is the median of the distribution of $F_{F(X_{(i)})}$, i.e. such that

$$F_{F(X_{(i)})}(\widetilde{y}_i) = 1/2.$$

An approximation of $\widetilde{y}_i$ is given by:

$$\widehat{F(x_{(i)})} \cong \frac{i - 0,3}{n + 0,4}.$$

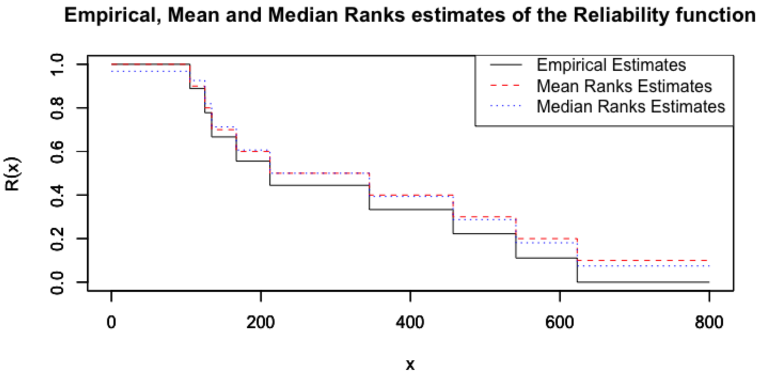

Figure: Competing risks dataset. Empirical estimates, Mean and Median Ranks estimates of the Reliability function ignoring the type of failure.

The problem of small datasets

Bayesian inference

(Parametric) Bayesian inference

When we have in hands a small dataset, it could be of interest to improve the (poor) estimates obtained from the data using **expert advises**. The bayesian approach brings us a nice way to do it.

(Parametric) Bayesian inference

When we have in hands a small dataset, it could be of interest to improve the (poor) estimates obtained from the data using **expert advises**. The bayesian approach brings us a nice way to do it.

Parametric Model

Let X still be the lifetime r.v. with p.d.f. $f_\theta(x)$ where θ is the unknown (multidimensional) parameter.

Example. $X \sim \mathcal{E}(\lambda)$, where λ is unknown.

Bayesian Framework

- **Main hypothesis**: the parameter θ is supposed to be random;

Bayesian Framework

- **Main hypothesis**: the parameter θ is supposed to be random;
- Its p.d.f. $h(\theta)$ is called the **prior distribution**;

Bayesian Framework

- **Main hypothesis**: the parameter θ is supposed to be random;
- Its p.d.f. $h(\theta)$ is called the **prior distribution**;
- The p.d.f. of (X, θ) is: $g(x, \theta) = f_\theta(x)h(\theta)$;

Bayesian Framework

- **Main hypothesis**: the parameter θ is supposed to be random;
- Its p.d.f. $h(\theta)$ is called the **prior distribution**;
- The p.d.f. of (X, θ) is: $g(x, \theta) = f_\theta(x)h(\theta)$;
- The marginal distribution of X,

$$f^*(x) = \int_\Theta g(x, \theta)d\theta,$$

is called the **predictive distribution**.

Suppose that **one observation** x of X is available, then

- the conditional distribution of θ given $X = x$ is $h(\theta|x) = g(x, \theta)/f^*(x)$;

Suppose that **one observation** x of X is available, then

- the conditional distribution of θ given $X = x$ is $h(\theta|x) = g(x, \theta)/f^*(x)$;
- this conditional distribution is called the **posterior distribution.**

Suppose that **one observation** x of X is available, then

- the conditional distribution of θ given $X = x$ is $h(\theta|x) = g(x, \theta)/f^*(x)$;
- this conditional distribution is called the **posterior distribution.**

If a sample sample $X_1, \ldots, X_n$ is observed, with values $x_1, \ldots, x_n$

- the conditional distribution of θ given $X_1 = x_1, \ldots, X_n = x_n$ is

$$h(\theta|x_1, \ldots, x_n) = \frac{\prod_{i=1}^n f_\theta(x_i)h(\theta)}{\int \prod_{i=1}^n f_\theta(x_i)h(\theta)d\theta}.$$

Bayesian estimator

Definition

The posterior quadratic risk is defined by

$$R(\hat{\theta}, x_1, \ldots, x_n) = \int (\hat{\theta} - \theta)^2 h(\theta | x_1, \ldots, x_n) d\theta.$$

The bayes estimate is the value $\hat{\theta}$ which minimizes the posterior risk given $x_1, \ldots, x_n$.

Bayesian estimator

Definition

The posterior quadratic risk is defined by

$$R(\hat{\theta}, x_1, \ldots, x_n) = \int (\hat{\theta} - \theta)^2 h(\theta | x_1, \ldots, x_n) d\theta.$$

The bayes estimate is the value $\hat{\theta}$ which minimizes the posterior risk given $x_1, \ldots, x_n$.

Example. Suppose that the distribution of X depends of an unidimensional parameter θ. The bayesian estimate is:

$$\widehat{\theta} = \mathbb{E}(\theta | x_1, \ldots, x_n) = \int \theta h(\theta | x_1, \ldots, x_n) d\theta.$$

Choice of a prior distribution for θ

The choice of the prior distribution $h(\theta)$ can be driven by two considerations.

- Choose a prior distribution which easily represents the expert's knowledge on the behavior of the lifetime X.

Choice of a prior distribution for θ

The choice of the prior distribution $h(\theta)$ can be driven by two considerations.

- Choose a prior distribution which easily represents the expert's knowledge on the behavior of the lifetime X.
- Choose a prior distribution which facilitates the mathematical developments.

Choice of a prior distribution for θ

The choice of the prior distribution $h(\theta)$ can be driven by two considerations.

- Choose a prior distribution which easily represents the expert's knowledge on the behavior of the lifetime X.
- Choose a prior distribution which facilitates the mathematical developments.

⚠ Different choices for the prior distribution

Choice of a prior distribution for θ

The choice of the prior distribution $h(\theta)$ can be driven by two considerations.

- Choose a prior distribution which easily represents the expert's knowledge on the behavior of the lifetime X.
- Choose a prior distribution which facilitates the mathematical developments.

⚠ Different choices for the prior distribution $\Rightarrow$ Different posterior distributions

Choice of a prior distribution for θ

The choice of the prior distribution $h(\theta)$ can be driven by two considerations.

- Choose a prior distribution which easily represents the expert's knowledge on the behavior of the lifetime X.
- Choose a prior distribution which facilitates the mathematical developments.

⚠ Different choices for the prior distribution$\Rightarrow$ Different posterior distributions$\Rightarrow$ **Different estimates!**

Modeling of the Expert's knowledge

Example.

- Suppose $X \sim \mathcal{E}(\lambda)$, where λ is unknown.

Modeling of the Expert's knowledge

Example.

- Suppose $X \sim \mathcal{E}(\lambda)$, where λ is unknown.
- Suppose that the expert is able to say the mean value for λ is 10 and that with probability more than 99% this value of λ is inside of the interval $[4, 16]$.

Modeling of the Expert's knowledge

Example.

- Suppose $X \sim \mathcal{E}(\lambda)$, where λ is unknown.
- Suppose that the expert is able to say the mean value for λ is 10 and that with probability more than 99% this value of λ is inside of the interval $[4, 16]$.

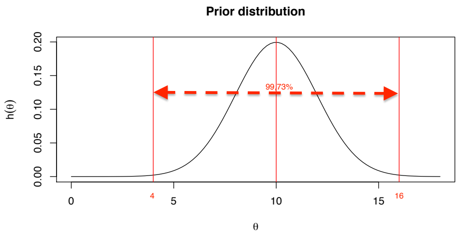

Figure: Modeling the Expert's knowledge. Gaussian prior according to a $N(10, 4)$.

Modeling of the Expert's knowledge (Cont'd)

If the expert has not a precise idea on the distribution of λ but only knows the bounds of an interval for λ, one could use uninformative prior through an uniform distribution.

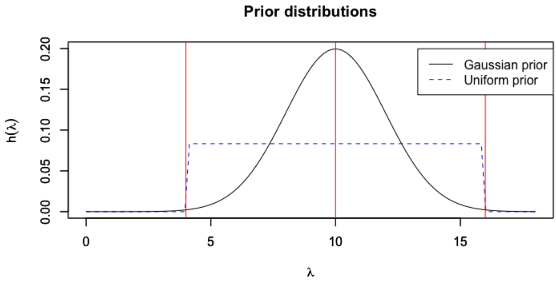

Figure: Modeling the Expert's knowledge. $N(10, 4)$ and $\mathcal{U}_{[4,16]}$ prior distributions.

Conjugate distributions

- Another criteria in the choice of the prior distribution could be the ease of calculus.

Conjugate distributions

- Another criteria in the choice of the prior distribution could be the ease of calculus.
- Some specific choices of prior distribution can simplify the mathematical expression of the bayesian estimator.

Conjugate distributions

- Another criteria in the choice of the prior distribution could be the ease of calculus.
- Some specific choices of prior distribution can simplify the mathematical expression of the bayesian estimator.
- It is the case when we have **conjugate priors**, that is to say when the prior and posterior distributions belong to the same parametric family.

Conjugate distributions

- Another criteria in the choice of the prior distribution could be the ease of calculus.
- Some specific choices of prior distribution can simplify the mathematical expression of the bayesian estimator.
- It is the case when we have **conjugate priors**, that is to say when the prior and posterior distributions belong to the same parametric family.

Example: $X \sim \mathcal{P}(\lambda)$ and $\lambda \sim \gamma(\alpha, \beta)$ with p.d.f.

$$h(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \mathbb{1}_{\mathbb{R}^+}(\lambda).$$

Conjugate distributions

- Another criteria in the choice of the prior distribution could be the ease of calculus.
- Some specific choices of prior distribution can simplify the mathematical expression of the bayesian estimator.
- It is the case when we have **conjugate priors**, that is to say when the prior and posterior distributions belong to the same parametric family.

Example: $X \sim \mathcal{P}(\lambda)$ and $\lambda \sim \gamma(\alpha, \beta)$ with p.d.f.

$$h(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \mathbb{1}_{\mathbb{R}^+}(\lambda).$$

Then, the posterior distribution of λ given $X = x$ is $\gamma(\alpha + x, \beta + 1)$.

Example: Exponential lifetimes

Conjugate priors property. If X is exponentially distributed with parameter λ and if λ has a $\gamma(\alpha, \beta)$ prior distribution, then the posterior distribution of λ, given $X = x$, is $\gamma(\alpha + 1, \beta + x)$.

Example: Exponential lifetimes

Conjugate priors property. If X is exponentially distributed with parameter λ and if λ has a $\gamma(\alpha, \beta)$ prior distribution, then the posterior distribution of λ, given $X = x$, is $\gamma(\alpha + 1, \beta + x)$.

Under Type I censoring. Let $(t_1, \delta_1), \ldots, (t_n, \delta_n)$ be the observed data and

$$
\begin{aligned}
Nof &= \sum_{i=1}^{n} \delta_i \\
TTT &= \sum_{i=1}^{n} t_i
\end{aligned}
$$

respectively the **number of failures** and the **total time on test**.

Example: Exponential lifetimes

Conjugate priors property. If X is exponentially distributed with parameter λ and if λ has a $\gamma(\alpha, \beta)$ prior distribution, then the posterior distribution of λ, given $X = x$, is $\gamma(\alpha + 1, \beta + x)$.

Under Type I censoring. Let $(t_1, \delta_1), \ldots, (t_n, \delta_n)$ be the observed data and

$$\begin{aligned} Nof &= \sum_{i=1}^{n} \delta_i \\ TTT &= \sum_{i=1}^{n} t_i \end{aligned}$$

respectively the **number of failures** and the **total time on test**.

Then, the posterior distribution of λ given the observed $(t_1, \delta_1), \ldots, (t_n, \delta_n)$ is $\gamma(\alpha + Nof, \beta + TTT)$.

Advantage: We know that the mean of a $\gamma(a, b)$ distribution is a/b.

Advantage: We know that the mean of a $\gamma(a, b)$ distribution is a/b.

Thus, knowing that the posterior distribution of λ

$$h(\lambda|(t_1, \delta_1), \ldots, (t_n, \delta_n)) = \gamma(\alpha + Nof, \beta + TTT),$$

the bayesian estimate of λ is:

$$\widehat{\lambda}_n = \mathbb{E}(\lambda|(t_1, \delta_1), \ldots, (t_n, \delta_n)) = \frac{\alpha + Nof}{\beta + TTT}.$$

Recall that the MLE estimate of λ is

$$MLE(\lambda) = \frac{Nof}{TTT}.$$

Recall that the MLE estimate of λ is

$$MLE(\lambda) = \frac{Nof}{TTT}.$$

Then, we have:

$$\widehat{\lambda}_n = \frac{\beta}{\beta + TTT}\mathbb{E}_{\gamma(\alpha,\beta)}(\lambda) + \frac{TTT}{\beta + TTT}MLE(\lambda).$$

Recall that the MLE estimate of λ is

$$MLE(\lambda) = \frac{Nof}{TTT}.$$

Then, we have:

$$\widehat{\lambda}_n = \frac{\beta}{\beta + TTT}\mathbb{E}_{\gamma(\alpha,\beta)}(\lambda) + \frac{TTT}{\beta + TTT}MLE(\lambda).$$

This shows that the bayesian estimate is a mix of the prior knowledge, which doesn't use the data, and the frequentist estimation, which is entirely based on the data.

The bayesian estimate of the Reliability function is:

$$\widehat{R}_n(x) = \mathbb{E}(e^{-\lambda x} | (t_1, \delta_1), \ldots, (t_n, \delta_n))$$

The bayesian estimate of the Reliability function is:

$$\widehat{R}_n(x) = \mathbb{E}(e^{-\lambda x} | (t_1, \delta_1), \ldots, (t_n, \delta_n)) = \left(\frac{\beta + TTT}{\beta + TTT + x} \right)^{Nof + \alpha}.$$

The bayesian estimate of the Reliability function is:

$$\widehat{R}_n(x) = \mathbb{E}(e^{-\lambda x}|(t_1, \delta_1), \ldots, (t_n, \delta_n)) = \left(\frac{\beta + TTT}{\beta + TTT + x}\right)^{Nof + \alpha}.$$

Credibility interval

Definition

An interval $C_\xi(x)$ is called bayesian interval with credibility ξ (or with level $1 - \xi$) if the posterior probability that θ lies in $C_\xi(x)$ is greater than ξ, i.e.

$$P(\theta \in C_\xi(x)|x) \geq \xi.$$

Plan

Plan

Regression models

Parametric regression models

Parametric regression model

Idea.

- The lifetime distribution is known up to a parameter θ.

Parametric regression model

Idea.

- The lifetime distribution is known up to a parameter θ.
- The parameter θ is not the same for each item (system).

Parametric regression model

Idea.

- The lifetime distribution is known up to a parameter θ.
- The parameter θ is not the same for each item (system).
- This parameter depends on some observable covariates z.

Parametric regression model

Idea.

- The lifetime distribution is known up to a parameter θ.
- The parameter θ is not the same for each item (system).
- This parameter depends on some observable covariates z.

Example.

- X is the lifetime of a Lightning arrester.

Parametric regression model

Idea.

- The lifetime distribution is known up to a parameter θ.
- The parameter θ is not the same for each item (system).
- This parameter depends on some observable covariates z.

Example.

- X is the lifetime of a Lightning arrester.
- Available covariates in the file: z_1=techno, z_2=group.

Parametric regression model

Idea.

- The lifetime distribution is known up to a parameter θ.
- The parameter θ is not the same for each item (system).
- This parameter depends on some observable covariates z.

Example.

- X is the lifetime of a Lightning arrester.
- Available covariates in the file: z_1=techno, z_2=group.
- If we assume (or can check...) that the lifetime distribution of the Lightning arrester depends on these covariates, then we can consider a parametric regression model.

Exponential regression model

The exponential regression model assumes that:

- for all values z of the (possibly multidimensional) covariate Z, the lifetime X has an exponential distribution,

Exponential regression model

The exponential regression model assumes that:

- for all values z of the (possibly multidimensional) covariate Z, the lifetime X has an exponential distribution,
- but the parameter $\lambda(z)$ of the distribution is function of z
- i.e the Reliability function is:

$$R(x|Z=z)=\exp(-\lambda(z)x), \quad \text{for all} \quad x\geq 0.$$

It is rather common to use the function $\lambda(z) = \exp(z^T\beta)$ where $z = (z_1, \ldots, z_p)$ is the vector of covariates and $\beta = (\beta_1, \ldots, \beta_p)$ is the vector of regression parameters.

It is rather common to use the function $\lambda(z) = \exp(z^T \beta)$ where $z = (z_1, \ldots, z_p)$ is the vector of covariates and $\beta = (\beta_1, \ldots, \beta_p)$ is the vector of regression parameters.

But other choices are possible.

It is rather common to use the function $\lambda(z) = \exp(z^T\beta)$ where $z = (z_1, \ldots, z_p)$ is the vector of covariates and $\beta = (\beta_1, \ldots, \beta_p)$ is the vector of regression parameters.

But other choices are possible.

Example. Lifetime of a lightning arrester has an exponential distribution with parameter $\lambda(z_1, z_2) = \exp(\beta_1 z_1 + \beta_2 z_2)$.

It is rather common to use the function $\lambda(z) = \exp(z^T \beta)$ where $z = (z_1, \ldots, z_p)$ is the vector of covariates and $\beta = (\beta_1, \ldots, \beta_p)$ is the vector of regression parameters.

But other choices are possible.

Example. Lifetime of a lightning arrester has an exponential distribution with parameter $\lambda(z_1, z_2) = \exp(\beta_1 z_1 + \beta_2 z_2)$. And the problem is to find how the covariates act on the lifetime.

It is rather common to use the function $\lambda(z) = \exp(z^T\beta)$ where $z = (z_1, \ldots, z_p)$ is the vector of covariates and $\beta = (\beta_1, \ldots, \beta_p)$ is the vector of regression parameters.

But other choices are possible.

Example. Lifetime of a lightning arrester has an exponential distribution with parameter $\lambda(z_1, z_2) = \exp(\beta_1 z_1 + \beta_2 z_2)$. And the problem is to find how the covariates act on the lifetime.

The regression parameters can be estimated by maximizing the likelihood (with or without censoring).

Weibull regression model

A Weibull regression model can be obtained if one assumes *e.g.* that, given the value z of the (possibly multidimensional) covariate Z, the Reliability function of X is given by:

$$R(x|Z=z) = \exp\left(-\left(\frac{x}{\alpha(z)}\right)^{\delta}\right), \text{ for all } x \geq 0,$$

where $\alpha(\cdot)$ is generally taken like $\alpha(z) = \exp(z^T\beta)$.

Weibull regression model

A Weibull regression model can be obtained if one assumes *e.g.* that, given the value z of the (possibly multidimensional) covariate Z, the Reliability function of X is given by:

$$R(x|Z=z) = \exp\left(-\left(\frac{x}{\alpha(z)}\right)^{\delta}\right), \text{ for all } x \geq 0,$$

where $\alpha(\cdot)$ is generally taken like $\alpha(z) = \exp(z^T\beta)$. But we could also assume that the parameter δ also depends on z...

Weibull regression model

A Weibull regression model can be obtained if one assumes *e.g.* that, given the value z of the (possibly multidimensional) covariate Z, the Reliability function of X is given by:

$$R(x|Z=z) = \exp\left(-\left(\frac{x}{\alpha(z)}\right)^{\delta}\right), \text{ for all } x \geq 0,$$

where $\alpha(\cdot)$ is generally taken like $\alpha(z) = \exp(z^T\beta)$. But we could also assume that the parameter δ also depends on z...

Remark. Of course, one can introduce other parametric regression models using the log-normal distribution, the gamma distribution, etc...

It's interesting to note that the r.v. $U = \delta(\log X - \log \alpha(z))$ has its survival function given by

$$R_U(u|Z = z) = \exp(-\exp(u)).$$

This is the Gumbel distribution (also known as the generalized extreme value distribution).

It's interesting to note that the r.v. $U = \delta(\log X - \log \alpha(z))$ has its survival function given by

$$R_U(u|Z = z) = \exp(-\exp(u)).$$

This is the Gumbel distribution (also known as the generalized extreme value distribution).

Writing $\sigma = 1/\delta$ and $\mu(z) = \log \alpha(z)$ we have:

$$Y = \log(X) = \mu(z) + \sigma U.$$

It's interesting to note that the r.v. $U = \delta(\log X - \log \alpha(z))$ has its survival function given by

$$R_U(u|Z = z) = \exp(-\exp(u)).$$

This is the Gumbel distribution (also known as the generalized extreme value distribution).

Writing $\sigma = 1/\delta$ and $\mu(z) = \log \alpha(z)$ we have:

$$Y = \log(X) = \mu(z) + \sigma U.$$

This is a **position-scale model**.

With the choice $\alpha(z) = \exp(z^t\beta)$ in the initial Weibull regression model, we get:

$$Y = z^t\beta + \sigma U.$$

With the choice $\alpha(z) = \exp(z^t \beta)$ in the initial Weibull regression model, we get:

$$Y = z^t \beta + \sigma U.$$

This is nothing but the **linear model** (but without the gaussian assumption!).

With the choice $\alpha(z) = \exp(z^t\beta)$ in the initial Weibull regression model, we get:

$$Y = z^t\beta + \sigma U.$$

This is nothing but the **linear model** (but without the gaussian assumption!).

- The Ordinary Least Square method is not adapted to handle censored data.
- So estimation of β and σ are obtained through maximum likelihood estimation.

With the choice $\alpha(z) = \exp(z^t\beta)$ in the initial Weibull regression model, we get:

$$Y = z^t\beta + \sigma U.$$

This is nothing but the **linear model** (but without the gaussian assumption!).

- The Ordinary Least Square method is not adapted to handle censored data.
- So estimation of β and σ are obtained through maximum likelihood estimation.

This equation

$$Y = z^t\beta + \sigma U.$$

allows us to introduce new models by changing the distribution of U.

- If U has a normal distribution

This equation

$$Y = z^t \beta + \sigma U.$$

allows us to introduce new models by changing the distribution of U.

- If U has a normal distribution$\Rightarrow$ gaussian linear model.

This equation

$$Y = z^t \beta + \sigma U.$$

allows us to introduce new models by changing the distribution of U.

- If U has a normal distribution$\Rightarrow$ gaussian linear model. This is the log-normal regression model for X since $X|Z = z \sim LN(\beta^t z, \sigma)$.

This equation

$$Y = z^t\beta + \sigma U.$$

allows us to introduce new models by changing the distribution of U.

- If U has a normal distribution$\Rightarrow$ gaussian linear model. This is the log-normal regression model for X since $X|Z = z \sim LN(\beta^t z, \sigma)$.
- If U has a logistic distribution

This equation

$$Y = z^t\beta + \sigma U.$$

allows us to introduce new models by changing the distribution of U.

- If U has a normal distribution$\Rightarrow$ gaussian linear model. This is the log-normal regression model for X since $X|Z = z \sim LN(\beta^t z, \sigma)$.
- If U has a logistic distribution$\Rightarrow$ log-logistic regression model for X with parameters $\mu(z)$ and σ.

Proportional Hazards Model

Let us come back to the Weibull regression model where $X|Z = z \sim W(\alpha(z), \delta)$. The hazard rate function of X is:

$$\lambda(x|z) = \frac{\delta}{\alpha(z)} \left(\frac{x}{\alpha(z)} \right)^{\delta - 1}.$$

Thus, for two different covariates z_1 and z_2 one can write

$$\frac{\lambda(x|z_1)}{\lambda(x|z_2)} = \left(\frac{\alpha(z_2)}{\alpha(z_1)} \right)^{\delta},$$

which doesn't depend on x. This is a **proportional hazards model**.

Definition

When the hazard rate function can be written like

$$\lambda(x|z) = \lambda_0(x)g(z),$$

where $\lambda_0(\cdot)$ *and* $g(\cdot)$ *are known or unknown functions, we say that we have a* **proportional hazards model**.

The function $\lambda_0(\cdot)$ is called the **baseline** hazard rate function and corresponds to the hazard rate of an individual with covariate z such that $g(z) = 1$.

The proportional hazards model has been introduced by **Cox** in 1972 with function

$$g(z) = \exp(z^T \beta),$$

where β is an unknown regression parameter. In this case, we say that we have a Cox model.

Graphical checking

Let us come back to the general proportional hazards model. We can write:

$$R(x|z) = (R_0(x))^{g(z)},$$

where $R_0(x) = \exp(-\Lambda_0(x))$ is the baseline survival function.

The relation

$$\log(-\log R(x|z)) = \log g(z) + \log(-\log R_0(x))$$

allows us to obtain a graphical method in order to check if a proportional hazards model could fit the data, since in this case the difference $\log(-\log R(x|z_1)) - \log(-\log R(x|z_2))$ is constant in x.

Plan

Regression models

Semiparametric regression model

Cox's model

The **Cox semiparametric model** assumes that the hazard rate function depends on the covariates through the relation:

$$\lambda(x|z) = \lambda_0(x)\exp(z^T\beta).$$

Cox's model

The **Cox semiparametric model** assumes that the hazard rate function depends on the covariates through the relation:

$$\lambda(x|z) = \lambda_0(x)\exp(z^T\beta).$$

The baseline hazard rate function $\lambda_0(\cdot)$ and the regression parameter β are unknown.

Cox's model

The **Cox semiparametric model** assumes that the hazard rate function depends on the covariates through the relation:

$$\lambda(x|z) = \lambda_0(x) \exp(z^T \beta).$$

The baseline hazard rate function $\lambda_0(\cdot)$ and the regression parameter β are unknown.

- **Advantage**. Thanks to the unknown function $\lambda_0(\cdot)$, this model is more flexible than the parametric regression model to fit different types of distributions.

Cox's model

The **Cox semiparametric model** assumes that the hazard rate function depends on the covariates through the relation:

$$\lambda(x|z) = \lambda_0(x) \exp(z^T \beta).$$

The baseline hazard rate function $\lambda_0(\cdot)$ and the regression parameter β are unknown.

- **Advantage**. Thanks to the unknown function $\lambda_0(\cdot)$, this model is more flexible than the parametric regression model to fit different types of distributions.
- **Disadvantage**. There is more than the regression parameter to estimate, there is also the baseline function.

Cox's model

The **Cox semiparametric model** assumes that the hazard rate function depends on the covariates through the relation:

$$\lambda(x|z) = \lambda_0(x) \exp(z^T \beta).$$

The baseline hazard rate function $\lambda_0(\cdot)$ and the regression parameter β are unknown.

- **Advantage**. Thanks to the unknown function $\lambda_0(\cdot)$, this model is more flexible than the parametric regression model to fit different types of distributions.
- **Disadvantage**. There is more than the regression parameter to estimate, there is also the baseline function.

Problem. How to estimate $\lambda_0(\cdot)$ and β?

Inference in Cox's model

The likelihood of the (censored) observations can be written like:

$$
\begin{aligned}
& L((t_1,\delta_1),\ldots,(t_n,\delta_n);\beta,\lambda_0(\cdot)) \\
= & \prod_{i=1}^{n}\left(\frac{e^{z_i^T\beta}}{\sum_{l\in Y_i} e^{z_l^T\beta}}\right)^{\delta_i}\prod_{i=1}^{n}\left(\lambda_0(t_i)\times\sum_{l\in Y_i} e^{z_l^T\beta}\right)^{\delta_i} \\
& \times\prod_{i=1}^{n}(R_0(t_i))^{\exp z_i^T\beta},
\end{aligned}
$$

where $Y(x)$ is the number of subjects at risk at time x and Y_i is its value at time T_i.

Partial Likelihood

The **partial likelihood** introduced by Cox is given by the first term of the previous expression of the likelihood, that is:

$$L_1(\beta) = \prod_{i:\delta_i=1} \left(\frac{e^{z_i^T \beta}}{\sum_{l \in Y(t_i)} e^{z_l^T \beta}} \right).$$

Partial Likelihood

The **partial likelihood** introduced by Cox is given by the first term of the previous expression of the likelihood, that is:

$$L_1(\beta) = \prod_{i:\delta_i=1} \left(\frac{e^{z_i^T \beta}}{\sum_{l \in Y(t_i)} e^{z_l^T \beta}} \right).$$

The term

$$\frac{e^{z_i^T \beta}}{\sum_{l \in Y(t_i)} e^{z_l^T \beta}}$$

is the probability that the failure is due to individual i with covariate z_i, given that a failure has been observed at time t_i.

Partial Likelihood

The **partial likelihood** introduced by Cox is given by the first term of the previous expression of the likelihood, that is:

$$L_1(\beta) = \prod_{i:\delta_i=1} \left(\frac{e^{z_i^T \beta}}{\sum_{l \in Y(t_i)} e^{z_l^T \beta}} \right).$$

The term

$$\frac{e^{z_i^T \beta}}{\sum_{l \in Y(t_i)} e^{z_l^T \beta}}$$

is the probability that the failure is due to individual i with covariate z_i, given that a failure has been observed at time t_i.

The estimator of β obtained by maximizing this partial likelihood has the same kind of properties than the usual maximum likelihood estimator: consistency and asymptotic normality are fulfilled.