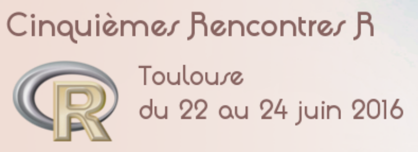
MBBEFD: MODÉLISATION DES TAUX DE DESTRUCTION EN ACTUARIAT NON VIE



Christophe Dutang[1] joint work with
Giorgio Spedicato[2], Markus Gesmann[3]

[1] Institute of Risk and Insurance (IRA) and Mathematics Laboratory of Le Mans (LMM),
Université du Maine, Le Mans, France
[2] ACAS, [3] Chainladder project

June 24, 2016

OUTLINES

1 INTRODUCTION
- Insurance context

2 MIXED-TYPED DISTRIBUTIONS
- One-inflated distributions
- the MBBEFD distribution

3 PACKAGE MBBEFD
- Characterizing functions
- Fitting methods

4 NUMERICAL ILLUSTRATIONS
- Simulation
- Real dataset

5 CONCLUSION
- Perspectives

OUTLINES

1 INTRODUCTION
- Insurance context

2 MIXED-TYPED DISTRIBUTIONS
- One-inflated distributions
- the MBBEFD distribution

3 PACKAGE MBBEFD
- Characterizing functions
- Fitting methods

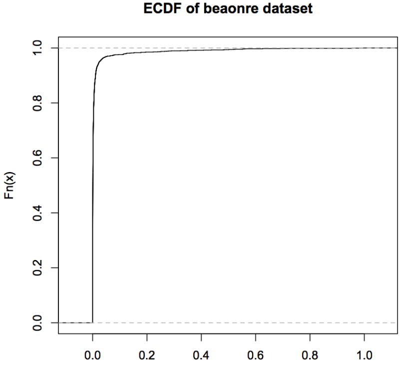
4 NUMERICAL ILLUSTRATIONS
- Simulation
- Real dataset

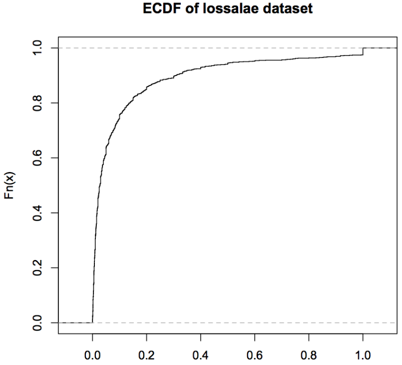
5 CONCLUSION
- Perspectives

Insurance context

Why model destruction rates?

- To make risks comparable, we look at ratio of losses to the underlying exposure.
- Typically, exposure is the sum insured (SI), the total insured value (TIV), or the maximum probable loss (MPL).
- Shifting the metric from loss amounts to damage ratio allows to benchmarks risks (i.e. insurance guarantees).





Reminder: a unified approach with fitdistrplus

Functionalities of the **fitdistrplus** package

- MLE: Extends the `fitdistr` function with fixed arguments, custom optimization algorithms, possible censoring,
- MME: Provides a generic function to perform moment matching estimation with the raw or centered moments,
- QME: Based on the `quantile` function, provides the quantile matching estimation,
- MGE: Maximum goodness-of-fit is now available with the usual statistical distance and their variants.

So we can fit any probability distribution.

OUTLINES

1 INTRODUCTION
- Insurance context

2 MIXED-TYPED DISTRIBUTIONS
- One-inflated distributions
- the MBBEFD distribution

3 PACKAGE MBBEFD
- Characterizing functions
- Fitting methods

4 NUMERICAL ILLUSTRATIONS
- Simulation
- Real dataset

5 CONCLUSION
- Perspectives

One-inflated distributions

Let F_0 be a continuous c.d.f. of a variable X_0. The corresponding c.d.f.[1] of the one-inflated random variable X_1 is

$$F_1(x) = (1-p_1)F_0(x) + p_1 1\!\!1_{[1,+\infty[}(x). \quad (1)$$

The density of X_1 should be considered w.r.t. the measure $\mu(x) = \lambda(x) + \delta_1(x)$ for which

$$f_1(x) = p_1^{1\!\!1_{x=1}} (1-p_1)^{1\!\!1_{x\neq 1}} f_0(x)^{1\!\!1_{x\neq 1}}. \quad (2)$$

The quantile function can be derived[2]

$$Q_1(p) = \begin{cases} Q_0(p/(1-p_1)) & \text{if } p < 1-p_1 \\ 1 & \text{if } p \geq 1-p_1 \end{cases} \quad (3)$$

where $Q_0 = F_0^{-1}$ denotes the quantile function of X_0.

[1] no density w.r.t. the Lebesgues measure but an improper density $(1-p)F_0'(x)$ and a probability mass p_1 at $x=1$.

[2] Assuming X_0 is valued on the unit-interval (i.e. $F_0(1)=1$).

Example of one-inflated distributions and estimation

- one-inflated beta:

$$F_1(x) = \begin{cases} 0 & \text{if } x < 0 \\ I(x; a, b)(1 - p_1) & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

where $I(x; a, b)$ denotes the incomplete beta ratio function. The improper density function is $f_1(x) = (1 - p_1)\frac{x^{a-1}(1-x)^{b-1}}{\beta(a,b)}$. The expectation is

$$E(X_1) = p_1 + (1 - p_1)\frac{a}{a+b}.$$

Proposition

- *The maximum likelihood estimator of p_1 is the empirical proportion of total loss $\hat{p}_1 = \frac{1}{n}\sum_{i=1}^{n} 1\!\!1_{X_i=1}$, whereas the estimator $\hat{\theta}$ is obtained separately.*
- *$(\hat{p}_1, \hat{\theta})$ converges in distribution to (p_1, θ) as*

$$\sqrt{n}\left(\begin{pmatrix}\hat{p}_1 \\ \hat{\theta}\end{pmatrix} - \begin{pmatrix}p_1 \\ \theta\end{pmatrix}\right) \xrightarrow[n\rightarrow+\infty]{\mathcal{D}} \mathcal{N}_{s+1}(0, \tilde{I}(p_1, \theta)^{-1}),\ \tilde{I}(p_1, \theta) = \begin{pmatrix}\frac{1}{p_1(1-p_1)} & 0 \\ 0 & I_0(\theta)\end{pmatrix}.$$

where $\tilde{I}(p_1, \theta)$ is the Fisher information matrix.

Exposure curves

The exposure curve function of X is defined as the ratio of the limited expected value and the expectation

$$G_X(d) = \frac{E(\min(X,d))}{E(X)}, \; d \in [0,1].$$

- exposure curves are concave increasing.
- direct link between the c.d.f. and the exposure curve :

$$G_X(d) = \frac{\int_0^d (1 - F_X(x))dx}{\int_0^1 (1 - F_X(x))dx} \Leftrightarrow F_X(x) = \left(1 - \frac{G'_X(x)}{G'_X(0)}\right) 1\!\!1_{[0,1[}(x) + 1\!\!1_{[1,+\infty[}(x).$$

Examples

- uniform sur $[0,1]$: $G_X(d) = d(2-d)$.
- Dirac en 1 : $G_X(d) = d$.

MBBEFD(A,B)

Introduced by [B97], the MBBEFD distribution is defined by the following exposure curve for $(a, b) \in \mathcal{D}_{a,b}$

$$\forall x \in I,\ G_X(x) = \begin{cases} \frac{\ln(\frac{a+b^x}{a+1})}{\ln(\frac{a+b}{a+1})} & \text{if } a(1-b) > 0 \\ \frac{1-b^x}{1-b} & \text{if } a = +\infty \text{ and } b < 1 \\ x & \text{if } a = 0 \text{ or } b = 1. \end{cases} \quad (4)$$

where the parameter domain is $\mathcal{D}_{a,b} = \{(a, b), a + 1 > 0, a(1-b) \geq 0, b > 0\}$.

NB: *MBBEFD* stands for Maxwell Boltzmann Bose Einstein Fermi Dirac

D,P,Q FUNCTIONS

The distribution function is

$$\forall x \in I,\ F_X(x) = \begin{cases} \left(1 - \frac{(a+1)b^x}{a+b^x}\right) 1\!\!1_{[0,1[}(x) + 1\!\!1_{[1,+\infty[}(x) & \text{if } a(1-b) > 0 \\ (1-b^x) 1\!\!1_{[0,1[}(x) + 1\!\!1_{[1,+\infty[}(x) & \text{if } a = +\infty \text{ and } b < 1 \\ 1\!\!1_{[1,+\infty[}(x) & \text{if } a = 0 \text{ or } b = 1. \end{cases}$$

A mixed-type distribution with

- a probability mass $P(X=1) = \frac{(a+1)b}{a+b}$,
- an improper density

$$\tilde{f}_X(x) \begin{cases} -\frac{a(a+1)b^x \ln(b)}{(a+b^x)^2} 1\!\!1_{[0,1[}(x) & \text{if } a(1-b) > 0 \\ -\ln(b) b^x 1\!\!1_{[0,1[}(x) & \text{if } a = +\infty \text{ and } b < 1 \\ 0 & \text{if } a = 0 \text{ or } b = 1. \end{cases}$$

PROPOSITION

The MBBEFD(a, b) verifies the regularity and differentiability conditions w.r.t. μ of Theorem 6.5.1 of [CL98], so the MLE converges (in distrib.) to the true value.

MBBEFD(G,B)

Let h be the function $h : (a, b) \mapsto \begin{pmatrix} \frac{a+b}{b(a+1)} \\ b \end{pmatrix}$. h is a bijection from $\mathcal{D}^i_{a,b}$ to $\mathcal{D}^i_{g,b}$ where

$$\mathcal{D}^1_{g,b} = \{(g, b) \in (1, +\infty)^2, bg > 1\}, \ \mathcal{D}^2_{g,b} = \{(g, b) \in (1, +\infty) \times (0, 1), bg < 1\}.$$

Hence, the new parametrization *MBBEFD*(g, b) is defined as

$$G_X(x) = \begin{cases} \frac{\ln\left(\frac{(g-1)b}{1-b} + \frac{1-gb}{1-b} b^x\right)}{\ln(gb)} & \text{if } g > 1, b \neq 1, b \neq 1/g \\ \frac{\ln(1+(g-1)x)}{\ln(g)} & \text{if } g > 1, b = 1 \\ \frac{1-b^x}{1-b} & \text{if } g > 1, bg = 1 \\ x & \text{if } g = 1 \text{ or } b = 0 \end{cases} \quad (5)$$

The distribution function (and other functions) can be derived

$$F_X(x) = \begin{cases} \left(1 - \frac{1-b}{(g-1)b^{1-x}+1-gb}\right) 1\!\!1_{[0,1[}(x) + 1\!\!1_{[1,+\infty[}(x) & \text{if } g > 1, b \neq 1, b \neq 1/g \\ (1 - b^x) 1\!\!1_{[0,1[}(x) + 1\!\!1_{[1,+\infty[}(x) & \text{if } g > 1, bg = 1 \\ 1\!\!1_{[1,+\infty[}(x) & \text{if } g = 1 \text{ or } b = 0 \end{cases}$$

OUTLINES

1 INTRODUCTION
- Insurance context

2 MIXED-TYPED DISTRIBUTIONS
- One-inflated distributions
- the MBBEFD distribution

3 PACKAGE MBBEFD
- Characterizing functions
- Fitting methods

4 NUMERICAL ILLUSTRATIONS
- Simulation
- Real dataset

5 CONCLUSION
- Perspectives

D, P, Q, R FUNCTIONS FOR ONE-INFLATED DISTRIBUTIONS

- generic one-inflated distributions:

```
doifun(x, dfun, p1, log=FALSE, ...)
poifun(q, pfun, p1, lower.tail = TRUE, log.p = FALSE, ...)
qoifun(p, qfun, p1, lower.tail = TRUE, log.p = FALSE, ...)
roifun(n, rfun, p1, ...)
```

- specific one-inflated distribution : `<d,p,q,r>d` for `d=unif`, `stpareto`, `beta`, `gbeta`.
- exposure curves via `ec` function, e.g. `ecoibeta`
- moments via `m` function, e.g. `moibeta`
- total loss via `tl` function, e.g. `tloibeta`

D, P, Q, R FUNCTIONS FOR THE MBBEFD DISTRIBUTION

- 1st parametrization:

```
dmbbefd(x, a, b, log=FALSE, g)
pmbbefd(q, a, b, lower.tail = TRUE, log.p = FALSE, g)
qmbbefd(p, a, b, lower.tail = TRUE, log.p = FALSE, g)
rmbbefd(n, a, b)
```

- 2nd parametrization: `<d,p,q,r>MBBEFD`
- exposure curves via `ec` function, e.g. `ecmbbefd`
- moments via `m` function, e.g. `mmbbefd`
- total loss via `tl` function, e.g. `tlmbbefd`

(MAXIMUM LIKELIHOOD) ESTIMATION

`fitDR()`

- provides MLE for the following list of distributions: `unif`, `stpareto`, `beta`, `gbeta`, `mbbefd`, `MBBEFD`.
- generates an object of class `"fitDR"` inheriting from the class `"fitdist"`.
- has access to all summarizing functions from **fitdistrplus**: `print`, `summary`, `logLik`, `coef`, `vcov`, `gofstat`,
- has access to all plotting functions from **fitdistrplus**: `cdfcomp, qqcomp, ppcomp, denscomp`.
- has access to bootstrap functions from **fitdistrplus**: `bootdist` and its generic functions,
- provides also total-loss-moment matching estimation.

`eecf()`

- produces an object of class `"eecf"`, `"function"`
- has generic functions `print`, `summary`, `plot`
- `eccomp()` plot exposure curves of multiple fits.

EXAMPLE

```
> x <- roibeta(1e3, 3, 2, 1/6)
> f1 <- fitDR(x, "oibeta", method="mle")
  shape1   shape2
3.032457 2.029166
  shape1   shape2       p1
3.032457 2.029166 0.167000
           shape1      shape2
shape1 0.02115524 0.011131107
shape2 0.01113111 0.008749516
> summary(f1)
Fitting of the distribution ' oibeta ' by maximum likelihood
Parameters :
       estimate Std. Error
shape1 3.032457 0.14544840
shape2 2.029166 0.09353885
p1     0.167000 0.37297587
Loglikelihood:  -252.4742   AIC:  510.9485   BIC:  525.6718
Correlation matrix:
          shape1    shape2 p1
shape1 1.0000000 0.8181583  0
shape2 0.8181583 1.0000000  0
p1     0.0000000 0.0000000  1
```

OUTLINES

1 INTRODUCTION
- Insurance context

2 MIXED-TYPED DISTRIBUTIONS
- One-inflated distributions
- the MBBEFD distribution

3 PACKAGE MBBEFD
- Characterizing functions
- Fitting methods

4 NUMERICAL ILLUSTRATIONS
- Simulation
- Real dataset

5 CONCLUSION
- Perspectives

ASSESSING BIAS AND VARIANCE – ONE-INFLATED DISTR.

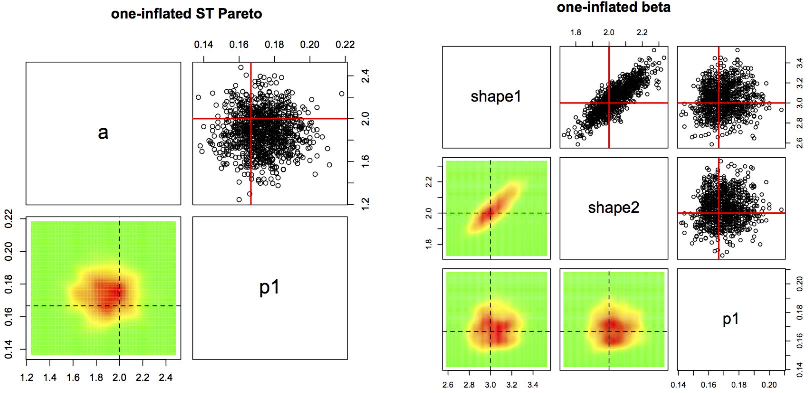


FIGURE: Bootstrap estimate of ML Estimators for oiPareto and oibeta, sample size $n = 1000$, bootstrap size $b = 1000$: function `bootdist` on `fitDR` outputs

ASSESSING BIAS AND VARIANCE – MBBEFD(A,B)

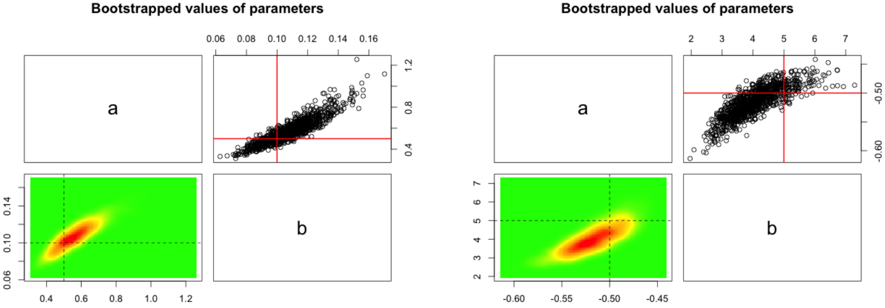


FIGURE: Bootstrap estimate of ML Estimators for MBBEFD(a,b), sample size $n = 1000$, bootstrap size $b = 1000$ for domain $\mathcal{D}^1_{a,b}$ (left) and $\mathcal{D}^2_{a,b}$ (right)

LOSSALAE DATASET – DISTRIBUTION FUNCTION

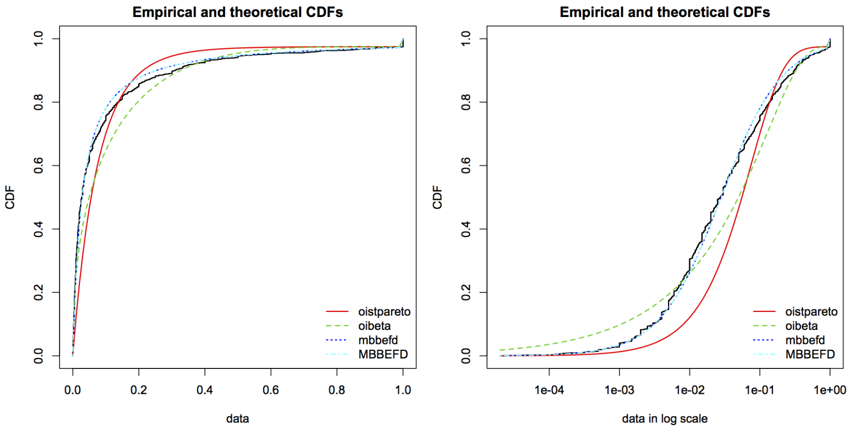


FIGURE: Fitted cdf on `lossalae`: function `cdfcomp` on `fitDR` outputs

LOSSALAE DATASET – PPPLOT AND ECPLOT

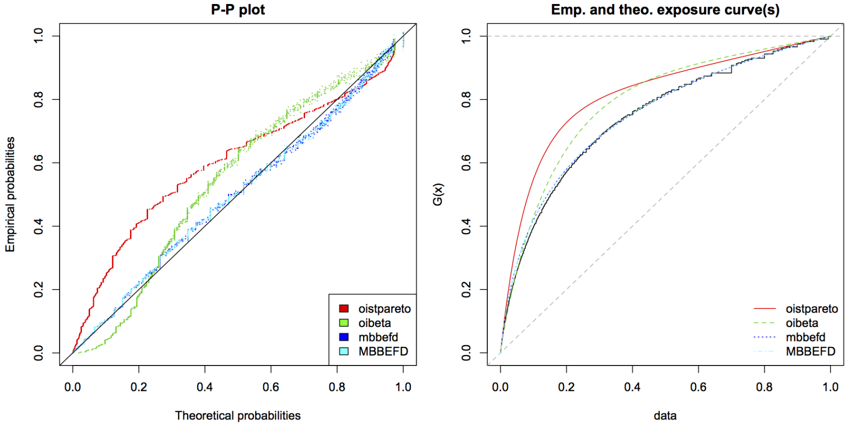


FIGURE: PP-plot and fitted densities on `lossalae`: functions `ppcomp`, `eccomp` on `fitDR` outputs

OUTLINES

1 INTRODUCTION
- Insurance context

2 MIXED-TYPED DISTRIBUTIONS
- One-inflated distributions
- the MBBEFD distribution

3 PACKAGE MBBEFD
- Characterizing functions
- Fitting methods

4 NUMERICAL ILLUSTRATIONS
- Simulation
- Real dataset

5 CONCLUSION
- Perspectives

Conclusion and Perspectives

- Fitting one-inflated distribution is carried out in a two-step procedure.
  1. estimate p_1 and compute the set of observations < 1,
  2. estimate other parameters on the other set.

- Fitting MBBEFD distribution is rather hard: use a three-step procedure.
  1. compute prefitting values based on parameter transformation,
  2. estimate parameters on subsets $\mathcal{D}^1$, $\mathcal{D}^2$
  3. take the most likely parameters.

- R package development:
  1. **mbbefd** [SDG16] for `d`, `p`, `q`, `r` functions of new distributions; computing exposure curve (theo. and emp.), fitting function `fitDR` inheriting from the `fitdist` class.
  2. use **fitdistrplus** [DMD16] for the fitting process.

- Regression models : why not consider explanatory variables?
  - dataset `asiacomrisk` contains large commercial losses caused by man-made risks with variables (usage, country,...).
  - how to use explanatory variables? in estimating the total loss probability?

REFERENCES

Bernegger, S. (1997),
The Swiss Re Exposure Curves and the MBBEFD Distribution Class
ASTIN Bulletin 27(1), 99–111.

Casella, G. & Lehmann, E. (1998),
Theory of Point Estimation,
Springer-Verlag.

Spedicato, G., Dutang, C. & Gesmann, M. (2016).
mbbefd: Maxwell Boltzmann Bose Einstein Fermi Dirac Distribution and Destruction Rate Modelling.
http://github.com/spedygiorgio/mbbefd

Delignette-Muller, M.L. & Dutang, C. (2016).
fitdistrplus: An R Package for Fitting Distributions.
http://r-forge.r-project.org/projects/riskassessment/

Dutang, C. (2016).
CASdatasets: Insurance datasets.
http://cas.uqam.ca/