Statistical Modeling of Loss Distributions Using **actuar**

Vincent Goulet

École d'actuariat, Université Laval
Québec, Canada

actuar

- Provides additional Actuarial Science functionality to R
- Current version covers
  - Loss distribution modeling
  - Risk theory (including ruin theory)
  - Simulation of compound hierarchical models
  - Credibility theory

Summary

Statistical
Modeling of
Loss
Distributions
Using
actuar

Vincent
Goulet

Probability
Laws

Grouped
Data

Minimum
Distance
Estimation

Censored
Data

1 Probability Laws

2 Grouped Data

3 Minimum Distance Estimation

4 Censored Data

Summary

Statistical
Modeling of
Loss
Distributions
Using
actuar

Vincent
Goulet

Probability
Laws

Grouped
Data

Minimum
Distance
Estimation

Censored
Data

- Support for 18 probability laws not in base R
- Mostly positive, heavy tail distributions
- New utility functions in addition to d*foo*, p*foo*, q*foo*, r*foo*

- Transformed Beta Family
  - 9 special cases (including Burr and Pareto)
- Transformed Gamma Family
  - 5 special cases (including inverse distributions)
- Loggamma
- Single parameter Pareto
- Generalized Beta
- Phase-type distributions

New Utility Functions

- m*foo* to compute theoretical raw moments

$$m_k = E[X^k]$$

- lev*foo* to compute theoretical limited moments

$$E[(X \wedge x)^k] = E[\min(X, x)^k]$$

- mgf*foo* to compute the moment generating function

$$M_X(t) = E[e^{tX}]$$

  when it exists

- Also support for: beta, exponential, chi-square, gamma, lognormal, normal (no lev), uniform, Weibull, inverse Gaussian

New Utility Functions

- m*foo* to compute theoretical raw moments

$$m_k = E[X^k]$$

- lev*foo* to compute theoretical limited moments

$$E[(X \wedge x)^k] = E[\min(X, x)^k]$$

- mgf*foo* to compute the moment generating function

$$M_X(t) = E[e^{tX}]$$

  when it exists

- Also support for: beta, exponential, chi-square, gamma, lognormal, normal (no lev), uniform, Weibull, inverse Gaussian

Summary

- Data presented in an interval-frequency manner:

Group	Line 1	Line 2
$(0, 25]$	30	26
$(25, 50]$	31	33
$(50, 100]$	57	31

- Need for a "standard" storage method
- Useful for minimum distance estimation

Creation and Manipulation of Objects

Statistical
Modeling of
Loss
Distributions
Using
actuar

Vincent
Goulet

Probability
Laws

Grouped
Data

Minimum
Distance
Estimation

Censored
Data

```
> x <- grouped.data(Group = c(0, 25,
+     50, 100), Line.1 = c(30, 31, 57),
+     Line.2 = c(26, 33, 31))
> x

      Group Line.1 Line.2
1  (0,  25]     30     26
2 (25,  50]     31     33
3 (50, 100]     57     31
```

Calculation of Empirical Moments

```
> mean(x)

  Line.1   Line.2
49.25847 43.19444

> emm(x, 2)

  Line.1   Line.2
3253.884 2604.167

> E <- elev(x[, -3])
> E(c(25, 50))

[1] 21.82203 37.18220
```

Plot of the Histogram and Ogive

```
> hist(x[, -3])
```

```
> plot(ogive(x[, -3]))
```


Histogram of x[, −3]
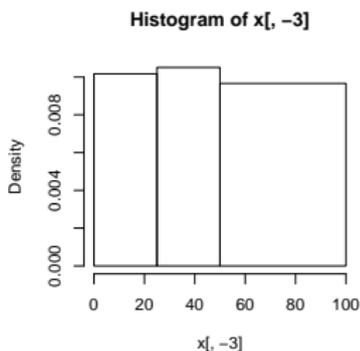
ogive(x[, −3])

Summary

Statistical
Modeling of
Loss
Distributions
Using
actuar

Vincent
Goulet

Probability
Laws

Grouped
Data

Minimum
Distance
Estimation

Censored
Data

1 Cramér-von Mises

$$d(\theta) = \sum_{j=1}^{n} w_j [F(x_j; \theta) - F_n(x_j; \theta)]^2$$

2 Modified chi-square

$$d(\theta) = \sum_{j=1}^{r} w_j [n(F(c_j; \theta) - F(c_{j-1}; \theta)) - n_j]^2,$$

3 Layer average severity

$$d(\theta) = \sum_{j=1}^{r} w_j [\text{LAS}(c_{j-1}, c_j; \theta) - \tilde{\text{LAS}}_n(c_{j-1}, c_j; \theta)]^2,$$

where $\text{LAS}(x, y) = E[\min(X, y)] - E[\min(X, x)]$

1 Cramér-von Mises

$$d(\theta) = \sum_{j=1}^{n} w_j [F(x_j; \theta) - F_n(x_j; \theta)]^2$$

2 Modified chi-square

$$d(\theta) = \sum_{j=1}^{r} w_j [n(F(c_j; \theta) - F(c_{j-1}; \theta)) - n_j]^2,$$

3 Layer average severity

$$d(\theta) = \sum_{j=1}^{r} w_j [\text{LAS}(c_{j-1}, c_j; \theta) - \tilde{\text{LAS}}_n(c_{j-1}, c_j; \theta)]^2,$$

where $\text{LAS}(x, y) = E[\min(X, y)] - E[\min(X, x)]$
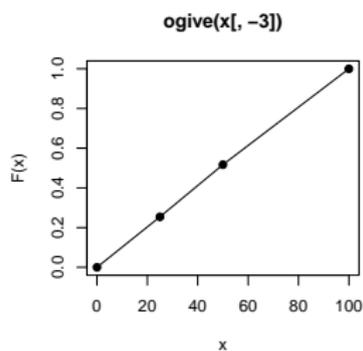
mde() Supports Three Distance Measures

1. Cramér-von Mises

$$d(\theta) = \sum_{j=1}^{n} w_j [F(x_j; \theta) - F_n(x_j; \theta)]^2$$

2. Modified chi-square

$$d(\theta) = \sum_{j=1}^{r} w_j [n(F(c_j; \theta) - F(c_{j-1}; \theta)) - n_j]^2,$$

3. Layer average severity

$$d(\theta) = \sum_{j=1}^{r} w_j [\text{LAS}(c_{j-1}, c_j; \theta) - \tilde{\text{LAS}}_n(c_{j-1}, c_j; \theta)]^2,$$

where $\text{LAS}(x, y) = E[\min(X, y)] - E[\min(X, x)]$

Summary

Statistical
Modeling of
Loss
Distributions
Using
actuar

Vincent
Goulet

Probability
Laws

Grouped
Data

Minimum
Distance
Estimation

Censored
Data

1 Probability Laws

2 Grouped Data

3 Minimum Distance Estimation

4 Censored Data

Context

- Common in statistical and actuarial applications to work with censored data
- Actuarial terminology:

left censoring	$\Leftrightarrow$	(ordinary) deductible
right censoring	$\Leftrightarrow$	policy limit
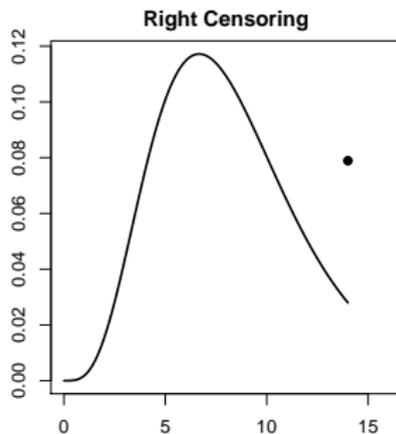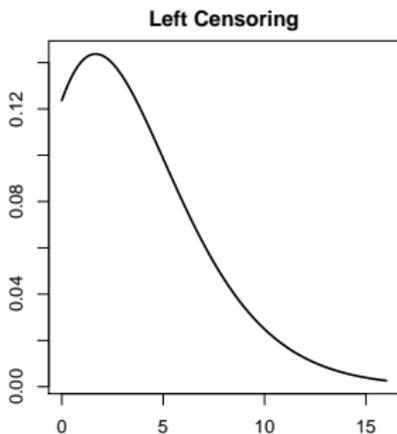
A Different Approach

- Package **survival** has extensive support for censored distributions
- Our approach is different
- `coverage()` returns pdf or cdf of censored distribution (with many options)
- function can be used in fitting as usual (`fitdistr()`, `mde()`, ...)

```
> f <- coverage(pdf = dgamma, cdf = pgamma,
+     deductible = 1, limit = 10)

> fitdistr(y, f, start = list(shape = 2,
+     rate = 0.5))

    shape         rate
  4.5822202    0.8634705
 (0.7672822) (0.1518537)
```

```
> f <- coverage(pdf = dgamma, cdf = pgamma,
+     deductible = 1, limit = 10)

> fitdistr(y, f, start = list(shape = 2,
+     rate = 0.5))

     shape        rate
  4.5822202    0.8634705
 (0.7672822) (0.1518537)
```

More Information

Statistical
Modeling of
Loss
Distributions
Using
actuar

Vincent
Goulet

Probability
Laws

Grouped
Data

Minimum
Distance
Estimation

Censored
Data

- Project's web site

    `http://www.actuar-project.org`

- Package vignettes

    ```
    actuar          Introduction to actuar
    coverage        Complete formulas used by
                    coverage
    credibility     Risk theory features
    lossdist        Loss distributions modeling
                    features
    risk            Risk theory features
    ```

- Demo files