Fitting Insurance Claims Severity Data with Coverage Modifications

How to Trick `fitdistrplus::fitdist` to Work with Customized Density Functions

Haoen Cui

May 30, 2020

satRday Chicago

Presentation GitHub Link

Distribution Fitting

General Framework

$$\boxed{\mathcal{P}(\theta) \quad \underset{\text{Parameter Fitting}}{\overset{\text{Data Generating Process}}{\rightleftharpoons}} \quad \mathcal{D} = \{x_i\}_{i=1}^n}$$

- Data generating process
  - Given distribution model $X \sim \mathcal{P}(\theta)$
  - Generate i.i.d. data $\mathcal{D} = \{x_i\}_{i=1}^n$ from $\mathcal{P}$
- Fitting parametric distribution model
  - Given observed data $\mathcal{D} = \{x_i\}_{i=1}^n$, assuming i.i.d.
  - Find parameter $\widehat{\theta}$ such that distribution model $\mathcal{P}(\widehat{\theta})$ "*fits best*"
    - Maximum Likelihood Estimation
    - Generalized Method of Moments
    - Maximum Goodness-of-Fit Estimation

Distribution Fitting Example

Suppose we have an unknown data generating process [1], and we attempt to find a set of parameters (μ^*, σ^*) such that

$$\mathcal{N}(\mu^*, \sigma^{*2})$$

fits the data the best (among normal distribution family).

```r
summary(x)
```

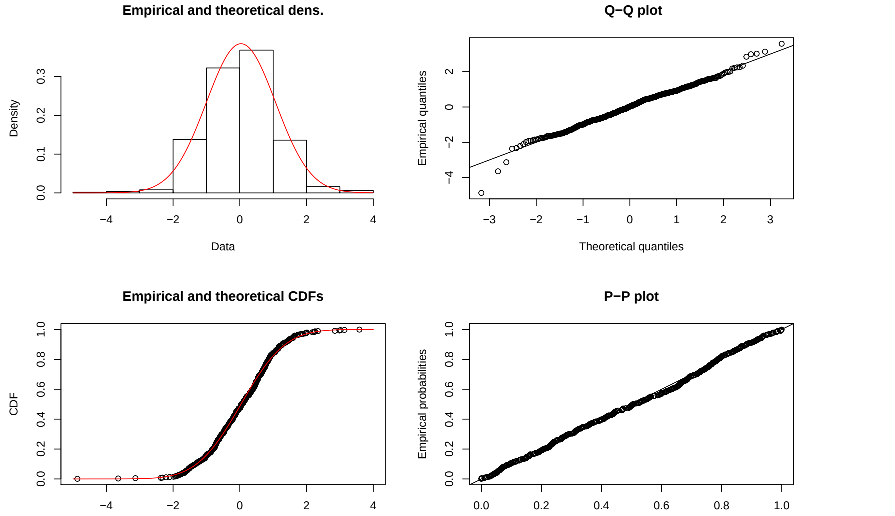
```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -4.86820 -0.67507  0.04759  0.03496  0.75622  3.58477
```

[1] Contaminated normal distribution used for data simulation:

$$(1-p)\cdot\mathcal{N}(\mu, {\sigma_1}^2) + p\cdot\mathcal{N}(\mu, {\sigma_2}^2)$$

Distribution Fitting Example (Cont.)

```r
fit <- fitdistrplus::fitdist(x, distr = dnorm, method = 'mle')
plot(fit)
```



Coverage Modifications

Incurred Raw Loss v.s. Insurer Covered Loss

When an *insured* incurs a loss of amount x, part of the raw amount will be covered by the *insurer* (insurance policy). The process of computing covered loss from raw loss is known as *coverage modification* in actuarial mathematics. Some typical configurations are listed below:

- (Ordinary) Policy Deductible $d \geq 0$

$$x \mapsto (x - d)_+$$

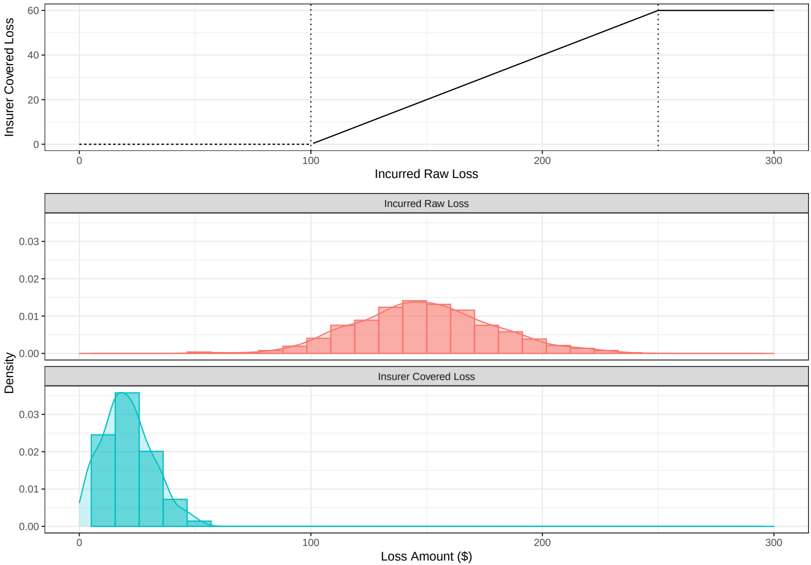
- Policy Limit $u > d$

$$x \mapsto \min\{x, u\}$$

- Coinsurance Factor $\alpha \in [0, 1]$

$$x \mapsto \alpha \cdot x$$

Coverage Modifications at a Glance



Case Study

Policy Specification

For this example, let's consider an insurance policy with

- (Ordinary) Policy Deductible: \$ 5,000
- Policy Limit: \$ 20,000
- Coinsurance Factor: 90 %

```r
# policy params
DEDUCTIBLE <- 5000
LIMIT <- 20000
COINSURANCE <- 0.9

# helper function
coverage_modification <- function(x) modify_loss(
    loss = x,
    deductible = DEDUCTIBLE,
    limit = LIMIT,
    coinsurance = COINSURANCE,
    per.loss = FALSE
)
```

Data Simulation

Suppose the incurred raw loss amount follows an unknown data generating process

$$L \sim \ln \mathcal{N}(\mu, \sigma)$$

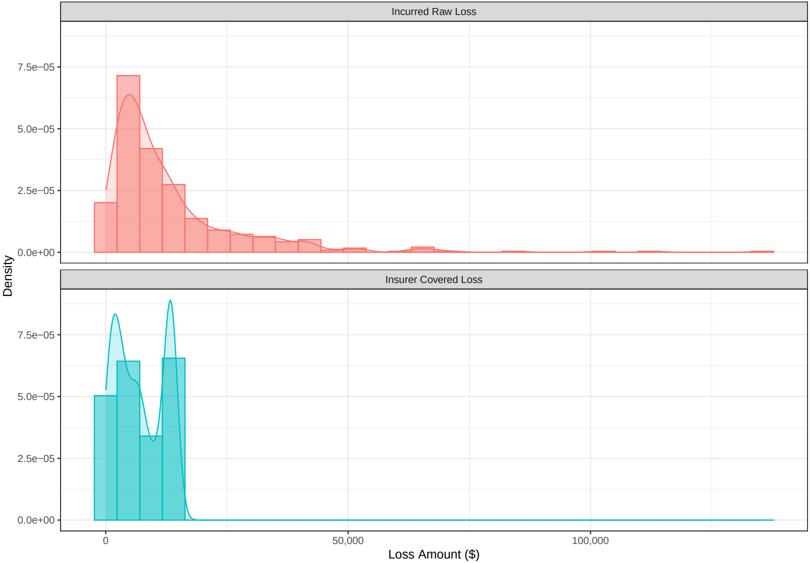
with the following parameters

```
# lognormal distribution param
MU <- 9
SIGMA <- 1
```

Let's simulate some data for this case study

```
# data simulation
loss_raw <- rlnorm(n = n_sample, meanlog = MU, sdlog = SIGMA)
loss_mod <- coverage_modification(loss_raw)
```

Data Simulation (Cont.)



Parameter Estimation

Affine transformation of modified loss data alone won't be sufficient. (Statistics omitted here.) Instead, we should transform the probability density function according to the contract parameters.

For modified loss $y \in (0, \alpha(u-d))$,

$$F_Y(y) = \frac{F_X\left(\frac{y+\alpha d}{\alpha(1+r)}\right) - F_X\left(\frac{d}{1+r}\right)}{1 - F_X\left(\frac{d}{1+r}\right)}$$

```r
# specify only CDF of loss_raw to get CDF of loss_mod
ploss_mod <- actuar::coverage(
    cdf = plnorm,
    deductible = DEDUCTIBLE,
    limit = LIMIT,
    coinsurance = COINSURANCE,
    franchise = FALSE,
    inflation = 0,
    per.loss = FALSE
)
```

Parameter Estimation (Cont.)

For modified loss $y \in (0, \alpha(u-d))$,

$$f_Y(y) = \frac{1}{\alpha(1+r)} \cdot \frac{f_X\left(\frac{y+\alpha d}{\alpha(1+r)}\right)}{1 - F_X\left(\frac{d}{1+r}\right)}$$

with a point probability mass at $y = \alpha(u-d)$,

$$\frac{1 - F_X\left(\frac{u}{1+r}\right)}{1 - F_X\left(\frac{d}{1+r}\right)}$$

```r
# specify both PDF and CDF of loss_raw to get PDF of loss_mod
dloss_mod <- actuar::coverage(
    pdf = dlnorm, cdf = plnorm,
    deductible = DEDUCTIBLE,
    limit = LIMIT,
    coinsurance = COINSURANCE,
    franchise = FALSE,
    inflation = 0,
    per.loss = FALSE
)
```

Parameter Estimation (Cont.)

By supplying our customized density function `dloss_mod` (which captures coverage modifications), we can estimate the parameters of raw loss using `fitdistrplus::fitdist` as before.

```
# parameter fit
loss_mod_fit <- fitdistrplus::fitdist(
    data = loss_mod,
    distr = dloss_mod,
    start = start_est,
    method = 'mle'
)
loss_mod_fit
```

```
## Fitting of the distribution ' loss_mod ' by maximum likelihood
## Parameters:
##         estimate Std. Error
## meanlog 8.960229  0.2261148
## sdlog   1.027789  0.1496204
```

Concluding Remarks

We can fit transformed data using existing approaches if we can construct the corresponding distribution functions.

- `actuar`: Actuarial Functions and Heavy Tailed Distributions
  - `actuar::coverage`: Density and Cumulative Distribution Function for Modified Data
- `fitdistrplus`: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data
  - `fitdistrplus::fitdist`: Fit of Univariate Distributions to Non-Censored Data