# Fitting distributions by MME, MGE, QME to non-censored data

*Marie Laure Delignette Muller, Christophe Dutang*

*2016-04-15*

## Contents

# 1 Maximum goodness-of-fit estimation (MGE)

This subsection focuses on alternative estimation methods. One of the alternative for continuous distributions is the maximum goodness-of-fit estimation method also called minimum distance estimation method [@]. In this package this method is proposed with eight different distances: the three classical distances defined in Table~**??**, or one of the variants of the Anderson-Darling distance proposed by (Luceno {2006}) and defined in Table~1. The right-tail AD gives more weight to the right-tail, the left-tail AD gives more weight only to the left tail. Either of the tails, or both of them, can receive even larger weights by using second order Anderson-Darling Statistics.

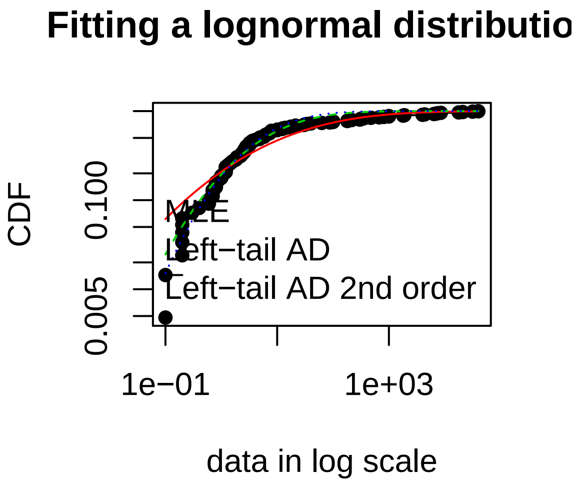
| Statistic | General formula | Computational formula |
|---|---|---|
| Right-tail AD (ADR) | $\int_{-\infty}^{\infty} \frac{(F_n(x)-F(x))^2}{1-F(x)} dx$ | $\frac{n}{2} - 2\sum_{i=1}^{n} F_i - \frac{1}{n}\sum_{i=1}^{n}(2i-1)ln(\overline{F}_{n+1-i})$ |
| Left-tail AD (ADL) | $\int_{-\infty}^{\infty} \frac{(F_n(x)-F(x))^2}{(F(x))} dx$ | $-\frac{3n}{2} + 2\sum_{i=1}^{n} F_i - \frac{1}{n}\sum_{i=1}^{n}(2i-1)ln(F_i)$ |
| Right-tail AD 2nd order (AD2R) | $ad2r = \int_{-\infty}^{\infty} \frac{(F_n(x)-F(x))^2}{(1-F(x))^2} dx$ | $ad2r = 2\sum_{i=1}^{n} ln(\overline{F}_i) + \frac{1}{n}\sum_{i=1}^{n} \frac{2i-1}{\overline{F}_{n+1-i}}$ |
| Left-tail AD 2nd order (AD2L) | $ad2l = \int_{-\infty}^{\infty} \frac{(F_n(x)-F(x))^2}{(F(x))^2} dx$ | $ad2l = 2\sum_{i=1}^{n} ln(F_i) + \frac{1}{n}\sum_{i=1}^{n} \frac{2i-1}{F_i}$ |
| AD 2nd order (AD2) | $ad2r + ad2l$ | $ad2r + ad2l$ |
| where $F_i \overset{\triangle}{=} F(x_i)$; | $\overline{F}_i \overset{\triangle}{=} 1 - F(x_i)$ | |

Table 1: Modified Anderson-Darling statistics as defined by [@Luceno06].

To fit a distribution by maximum goodness-of-fit estimation, one needs to fix the argument `method` to `"mge"` in the call to `fitdist` and to specify the argument `gof` coding for the chosen goodness-of-fit distance. This function is intended to be used only with continuous non-censored data.

Maximum goodness-of-fit estimation may be useful to give more weight to data at one tail of the distribution. In the previous example from ecotoxicology, we used a non classical distribution (the Burr distribution) to correctly fit the empirical distribution especially on its left tail. In order to correctly estimate the 5% percentile, we could also consider the fit of the classical lognormal distribution, but minimizing a goodness-of-fit distance giving more weight to the left tail of the empirical distribution. In what follows, the left tail Anderson-Darling distances of first or second order are used to fit a lognormal to ‘endosulfan} data set (see Figure~**??**).

```
library(fitdistrplus)
data("endosulfan")
ATV <-endosulfan$ATV
fendo.ln <- fitdist(ATV, "lnorm")
fendo.ln.ADL <- fitdist(ATV, "lnorm", method = "mge", gof = "ADL")
fendo.ln.AD2L <- fitdist(ATV, "lnorm", method = "mge", gof = "AD2L")
cdfcomp(list(fendo.ln, fendo.ln.ADL, fendo.ln.AD2L),
  xlogscale = TRUE, ylogscale = TRUE,
  main = "Fitting a lognormal distribution",
  xlegend = "bottomright",
  legendtext = c("MLE","Left-tail AD", "Left-tail AD 2nd order"))
```



Comparing the 5% percentiles (HC5) calculated using these three fits to the one calculated from the MLE fit of the Burr distribution, we can observe, on this example, that fitting the lognormal distribution by maximizing left tail Anderson-Darling distances of first or second order enables to approach the value obtained by fitting the Burr distribution by MLE.

```
library(actuar)
fendo.B <- fitdist(ATV, "burr", start = list(shape1 = 0.3, shape2 = 1,
  rate = 1))

(HC5.estimates <- c(
  empirical = as.numeric(quantile(ATV, probs = 0.05)),
  Burr = as.numeric(quantile(fendo.B, probs = 0.05)$quantiles),
  lognormal_MLE = as.numeric(quantile(fendo.ln, probs = 0.05)$quantiles),
  lognormal_AD2 = as.numeric(quantile(fendo.ln.ADL,
    probs = 0.05)$quantiles),
  lognormal_AD2L = as.numeric(quantile(fendo.ln.AD2L,
    probs = 0.05)$quantiles)))
```

```
##      empirical           Burr  lognormal_MLE  lognormal_AD2 lognormal_AD2L
##     0.20000000     0.29392593     0.07258961     0.19590686     0.25877232
```

# 2 Moment matching estimation (MME)

The moment matching estimation (MME) is another method commonly used to fit parametric distributions (Vose 2010). MME consists in finding the value of the parameter $\theta$ that equalizes the first theoretical raw moments of the parametric distribution to the corresponding empirical raw moments as in Equation~(1):

$$E(X^k|\theta) = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \tag{1}$$

for $k = 1, \dots, d$, with $d$ the number of parameters to estimate and $x_i$ the $n$ observations of variable $X$. For moments of order greater than or equal to 2, it may also be relevant to match centered moments. Therefore, we match the moments given in Equation~(2):

$$E(X|\theta) = \overline{x} \ , \ E\left((X - E(X))^k|\theta\right) = m_k, \text{ for } k = 2, \dots, d, \tag{2}$$

where $m_k$ denotes the empirical centered moments. This method can be performed by setting the argument `method` to `"mme"` in the call to `fitdist`. The estimate is computed by a closed-form formula for the following distributions: normal, lognormal, exponential, Poisson, gamma, logistic, negative binomial, geometric, beta and uniform distributions. In this case, for distributions characterized by one parameter (geometric, Poisson and exponential), this parameter is simply estimated by matching theoretical and observed means, and for distributions characterized by two parameters, these parameters are estimated by matching theoretical and observed means and variances (Vose 2010). For other distributions, the equation of moments is solved numerically using the `optim` function by minimizing the sum of squared differences between observed and theoretical moments (see the **fitdistrplus** reference manual for technical details (Delignette-Muller et al. 2014).

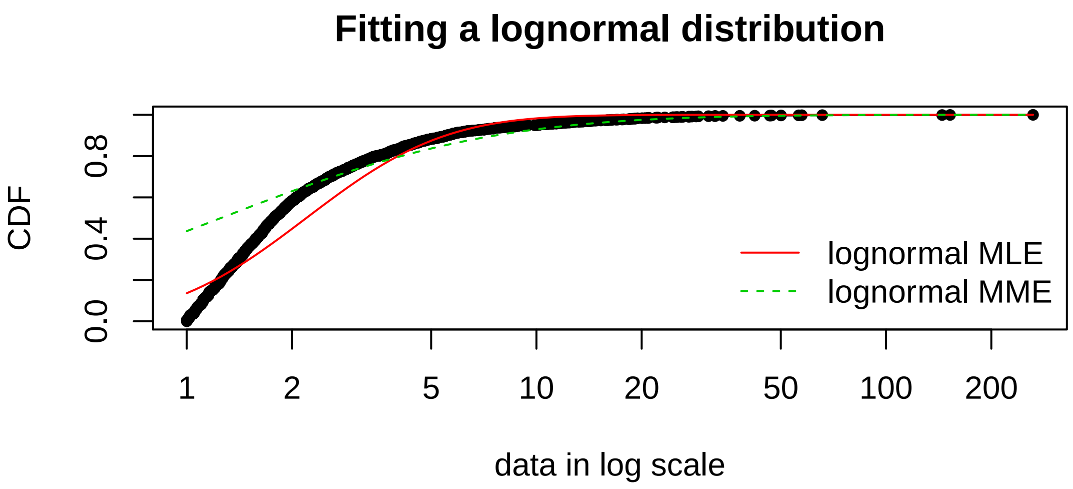
A classical data set from the Danish insurance industry published in (McNeil 1997) will be used to illustrate this method. In **fitdistrplus**, the data set is stored in `danishuni` for the univariate version and contains the loss amounts collected at Copenhagen Reinsurance between 1980 and 1990. In actuarial science, it is standard to consider positive heavy-tailed distributions and have a special focus on the right-tail of the distributions. In this numerical experiment, we choose classic actuarial distributions for loss modelling: the lognormal distribution and the Pareto type II distribution (Klugman, Panjer, and Willmot 2009).

The lognormal distribution is fitted to ‘danishuni} data set by matching moments implemented as a closed-form formula. On the left-hand graph of Figure~**??**, the fitted distribution functions obtained using the moment matching estimation (MME) and maximum likelihood estimation (MLE) methods are compared. The MME method provides a more cautious estimation of the insurance risk as the MME-fitted distribution function (resp. MLE-fitted) underestimates (overestimates) the empirical distribution function for large values of claim amounts.

```r
data("danishuni")
str(danishuni)
```

```
## 'data.frame':    2167 obs. of  2 variables:
##  $ Date: Date, format: "1980-01-03" "1980-01-04" ...
##  $ Loss: num  1.68 2.09 1.73 1.78 4.61 ...
```

```r
fdanish.ln.MLE <- fitdist(danishuni$Loss, "lnorm")
fdanish.ln.MME <- fitdist(danishuni$Loss, "lnorm", method = "mme",
  order = 1:2)
cdfcomp(list(fdanish.ln.MLE, fdanish.ln.MME),
  legend = c("lognormal MLE", "lognormal MME"),
  main = "Fitting a lognormal distribution",
  xlogscale = TRUE, datapch = 20)
```



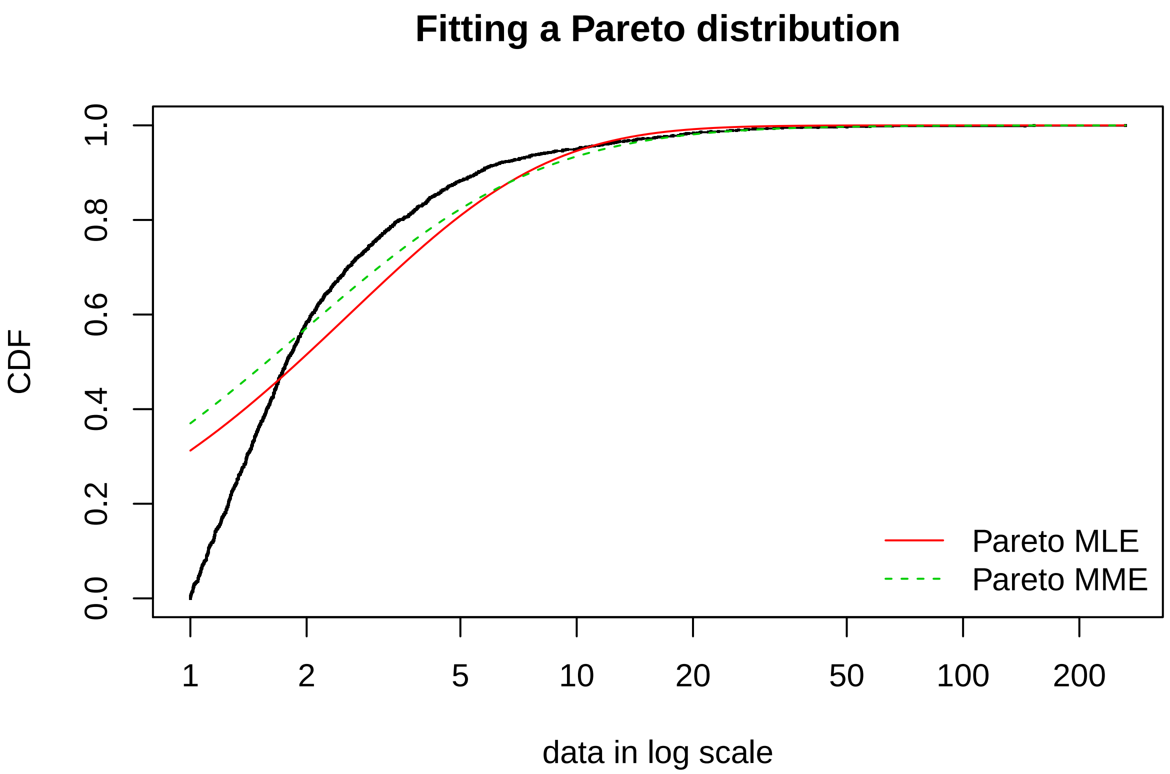
In a second time, a Pareto distribution, which gives more weight to the right-tail of the distribution, is fitted. As the lognormal distribution, the Pareto has two parameters, which allows a fair comparison. The Burr distribution (with its three parameters) would lead to a better fit.

We use the implementation of the **actuar** package providing raw and centered moments for that distribution (in addition to `d`, `p`, `q` and `r` functions (Goulet 2012). Fitting a heavy-tailed distribution for which the first and the second moments do not exist for certain values of the shape parameter requires some cautiousness. This is carried out by providing, for the optimization process, a lower and an upper bound for each parameter. The code below calls the L-BFGS-B optimization method in `optim`, since this quasi-Newton allows box constraints[1]. We choose match moments defined in Equation~(1), and so a function for computing the empirical raw moment (called `memp` in our example) is passed to `fitdist`. For two-parameter distributions (i.e., $d = 2$),
Equations~(1) and (2) are equivalent.

```
library("actuar")
fdanish.P.MLE <- fitdist(danishuni$Loss, "pareto",
  start = list(shape = 10, scale = 10), lower = 2+1e-6, upper = Inf)
memp <- function(x, order) sum(x^order)/length(x)
fdanish.P.MME <- fitdist(danishuni$Loss, "pareto", method = "mme",
  order = 1:2, memp = "memp", start = list(shape = 10, scale = 10),
  lower = c(2+1e-6, 2+1e-6), upper = c(Inf, Inf))
cdfcomp(list(fdanish.P.MLE, fdanish.P.MME),
  legend = c("Pareto MLE", "Pareto MME"),
  main = "Fitting a Pareto distribution",
  xlogscale = TRUE, datapch = ".")
```

---

[1] That is what the B stands for.



```r
gofstat(list(fdanish.ln.MLE, fdanish.P.MLE,
  fdanish.ln.MME, fdanish.P.MME),
  fitnames = c("lnorm.mle", "Pareto.mle", "lnorm.mme", "Pareto.mme"))
```

```
## Goodness-of-fit statistics
##                                lnorm.mle  Pareto.mle   lnorm.mme
## Kolmogorov-Smirnov statistic  0.1374619   0.3124174   0.4367645
## Cramer-von Mises statistic   14.7911467  37.7226866  88.9503123
## Anderson-Darling statistic   87.1933309 208.3387883 416.2567475
##                                Pareto.mme
## Kolmogorov-Smirnov statistic   0.3700154
## Cramer-von Mises statistic    55.4266567
## Anderson-Darling statistic   281.5838880
##
## Goodness-of-fit criteria
##                                 lnorm.mle Pareto.mle lnorm.mme Pareto.mme
## Aikake's Information Criterion  8119.795   9249.666  9791.887   9408.535
## Bayesian Information Criterion  8131.157   9261.029  9803.249   9419.897
```

As shown on Figure~**??**, MME and MLE fits are far less distant (when looking at the right-tail) for the Pareto distribution than for the lognormal distribution on this data set. Furthermore, for these two distributions, the MME method better fits the right-tail of the distribution from a visual point of view. This seems logical since empirical moments are influenced by large observed values. In the previous traces, we gave the values of goodness-of-fit statistics. Whatever the statistic considered, the MLE-fitted lognormal always provides the best fit to the observed data.

Maximum likelihood and moment matching estimations are certainly the most commonly used method for fitting distributions (Cullen and Frey 1999). Keeping in mind that these two methods may produce very different results, the user should be aware of its great sensitivity to outliers when choosing the moment matching estimation. This may be seen as an advantage in our example if the objective is to better describe the right tail of the distribution, but it may be seen as a drawback if the objective is different.

## 3 Quantile matching estimation (QME)

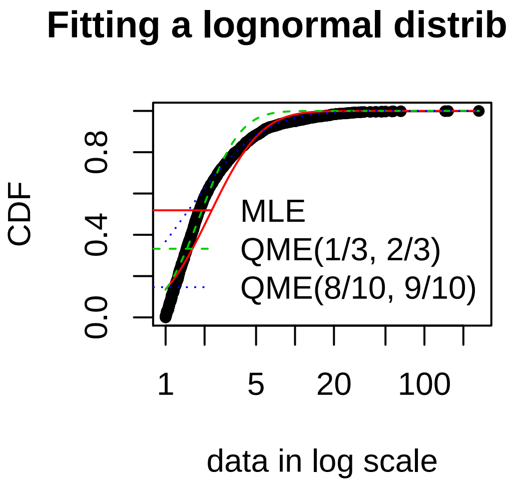
Fitting of a parametric distribution may also be done by matching theoretical quantiles of the parametric distributions (for specified probabilities) against the empirical quantiles ((Tse 2009)). The equality of theoretical and empirical qunatiles is expressed by Equation~(3) below, which is very similar to Equations~(1) and (2):

$$F^{-1}(p_k|\theta) = Q_{n,p_k} \tag{3}$$

for $k = 1, \dots, d$, with $d$ the number of parameters to estimate (dimension of $\theta$ if there is no fixed parameters) and $Q_{n,p_k}$ the empirical quantiles calculated from data for specified probabilities $p_k$. The choice $p_k$'s is linked to the application. For instance, the World Health Organisation uses 3%, 15%, 50%, 85%, 97% in its charts, therefore one could choose $p_1 = 50\%$, $p_2 = 15\%$, $p_3 = 85\%$ for a three-parameter distribution. Another relevant example is $p_1 = 99\%$ when assessing the right-tail of the distribution, which is typical in insurance/finance.

Quantile matching estimation (QME) is performed by setting the argument `method` to `"qme"` in the call to `fitdist` and adding an argument `probs` defining the probabilities for which the quantile matching is performed. The length of this vector must be equal to the number of parameters to estimate (as the vector of moment orders for MME). Empirical quantiles are computed using the `quantile` function of the **stats** package using `type=7` by default (see `?quantile` and (Hyndman and Fan 1996)). But the type of quantile can be easily changed by using the `qty` argument in the call to the `qme` function.
The quantile matching is carried out numerically, by minimizing the sum of squared differences between observed and theoretical quantiles.

```r
fdanish.ln.QME1 <- fitdist(danishuni$Loss, "lnorm", method = "qme",
  probs = c(1/3, 2/3))
fdanish.ln.QME2 <- fitdist(danishuni$Loss, "lnorm", method = "qme",
  probs = c(8/10, 9/10))
cdfcomp(list(fdanish.ln.MLE, fdanish.ln.QME1, fdanish.ln.QME2),
  legend = c("MLE", "QME(1/3, 2/3)", "QME(8/10, 9/10)"),
  main = "Fitting a lognormal distribution",
  xlogscale = TRUE, datapch = 20)
```



Above is an example of fitting of a lognormal distribution to `danishuni` data set by matching probabilities $(p_1 = 1/3, p_2 = 2/3)$ and $(p_1 = 8/10, p_2 = 9/10)$. As expected, the second QME fit gives more weight to the right-tail of the distribution. , despite we do not choose the Pareto type-II distribution. Compared to the maximum likelihood estimation, the second QME fit best suits the right-tail of the distribution, whereas the first QME fit best models the body of the distribution. The quantile matching estimation is of particular interest when we need to focus around particular quantiles, e.g., $p = 99.5\%$ in the Solvency II insurance context or $p = 5\%$ for the HC5 estimation in the ecotoxicology context.

# References

Cullen, A.C., and H.C. Frey. 1999. *Probabilistic Techniques in Exposure Assessment*. 1st ed. Plenum Publishing Co.

Delignette-Muller, M.L., R. Pouillot, J.B. Denis, and C. Dutang. 2014. *Fitdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data*. http://CRAN.R-project.org/package=fitdistrplus.

Goulet, V. 2012. *Actuar: An R Package for Actuarial Science*. http://CRAN.R-project.org/package=actuar.

Hyndman, R.J., and Y. Fan. 1996. "Sample Quantiles in Statistical Packages." *The American Statistician* 50: 361–65.

Klugman, S.A., H.H. Panjer, and G.E. Willmot. 2009. *Loss Models: from Data to Decisions*. 3rd ed. John Wiley & Sons.

Luceno, A. {2006}. "Fitting the Generalized Pareto Distribution to Data Using Maximum Goodness-of-fit Estimators." *Computational Statistics and Data Analysis* 51 (2): 904–17.

McNeil, A.J. 1997. "Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory." *ASTIN Bulletin* 27 (1): 117–37.

Tse, Y.K. 2009. *Nonlife Actuarial Models: Theory, Methods and Evaluation*. 1st ed. International Series on Actuarial Science. Cambridge University Press.

Vose, D. 2010. *Quantitative Risk Analysis. a Guide to Monte Carlo Simulation Modelling*. 1st ed. John Wiley & Sons.