June 2018

Robust Estimation of Parametric Models for Insurance Loss Data

Chudamani Poudyal
University of Wisconsin-Milwaukee

Robust Estimation of Parametric Models for Insurance Loss Data

by

Chudamani Poudyal

A Dissertation Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
in Mathematics

at

The University of Wisconsin-Milwaukee
August 2018

ABSTRACT

Robust Estimation of Parametric Models for Insurance Loss Data

by

Chudamani Poudyal

The University of Wisconsin-Milwaukee, 2018

Under the Supervision of Professor Vytaras Brazauskas

Parametric statistical models for insurance claims severity are continuous, right-skewed, and frequently heavy-tailed. The data sets that such models are usually fitted to contain outliers that are difficult to identify and separate from genuine data. Moreover, due to commonly used actuarial "loss control strategies," the random variables we observe and wish to model are affected by truncation (due to deductibles), censoring (due to policy limits), scaling (due to coinsurance proportions) and other transformations. In the current practice, statistical inference for loss models is almost exclusively likelihood (MLE) based, which typically results in non-robust parameter estimators, pricing models, and risk measures. To alleviate the lack of robustness of MLE-based inference in risk modeling, two broad classes of parameter estimators – Method of Trimmed Moments (MTM) and Method of Winsorized Moments (MWM) – have been recently developed. MTM and MWM estimators are sufficiently general and flexible, and posses excellent large- and small-sample properties, but they were designed for complete (not transformed) data. In this dissertation, we first redesign MTM estimators to be applicable to claim severity models that are fitted to truncated, censored, and insurance payments data. Asymptotic properties of such estimators are thoroughly investigated and their practical performance is illustrated using Norwegian fire claims data. In addition, we explore several extensions of MTM and MWM estimators for complete data. In particular, we introduce truncated, censored, and insurance payment-type estimators and study their asymptotic properties. Our analysis establishes new connections between data truncation, trimming, and censoring which paves the way for more effective modeling of non-linearly transformed loss data.

To

five ladies in my life

Mother	Bhagiratha Poudyal,
Mother-in-law	Menuka Dahal,
Wife	Bedu Dahal,
Daughters	Gargi and Guras

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

First, I owe a debt of profound gratitude to my supervisor, Professor Vytaras Brazauskas for his invaluable and untiring guidance, encouragement, and suggestions during my five years time at UWM. His guidance, vision, and countless support have made this work possible.

My sincere gratitude goes to Professor Jugal K. Ghorai for his very helpful comments and expert advise on how to improve some of the results in this dissertation.

I am also grateful to Professors Jay H. Beder, Jugal K. Ghorai, Bruce Wade, and Wei Wei for serving on my dissertation committee. I hope to emulate their support and models of academic excellence in my academic career.

Many thanks to Professor Yang Ho who influenced my actuarial career in many dimensions. Thank you all the students in the actuarial science program at UWM, especially Ranadeera Samanthi, Daoping Yu, Qian Zhao, and James Laate for sharing their motivational experience and struggles towards achieving career goals.

Special thanks to Professors Daniel Gervini, Bruce Wade, and Yi Ming Zou for their support and advice on how to complete the program on time. Thanks to the rest of the faculty and staff in the Department of Mathematical Sciences at UWM for every opportunity they have provided in my professional development.

My deep respect to Dr. David R. Finston, Emeritus Professor, New Mexico State University, for his continuous support in my academic career.

Beyond mathematics, I am very much indebted to my family members, including my mother – Bhagiratha Poudyal, wife – Bedu Dahal, and two gorgeous daughters – Gargi and Guras, for sacrificing many things in their lives to help convert my dreams to reality. My special thanks go to all my colleagues, friends and relatives for their suggestions, kind help and motivation.

Chudamani Poudyal

Milwaukee, Wisconsin

May 2018

Chapter 1

Introduction and Preliminaries

1.1 Motivation

Parametric statistical models for insurance claims severity are continuous, right-skewed, and frequently heavy-tailed (see, e.g., Klugman *et al.*, 2012). The data sets that such models are usually fitted to contain outliers that are difficult to identify and separate from genuine data. As a result, there could be a significant difference in statistical inference if the true parametric model is slightly different than the one assumed. Therefore, it is appealing to search for statistical procedures that are insensitive against small perturbations from the assumed models.

In the current practice, statistical inference for loss models is almost exclusively maximum likelihood estimation (MLE) based. The MLE approach typically results in non-robust parameter estimators. The recently proposed estimators based on method of trimmed moments (MTM) (see, e.g., Brazauskas *et al.*, 2009) and method of Winsorized moments (MWM) (see, e.g., Zhao *et al.*, 2018a,b) can address the issue of non-robustness. These estimators are computationally tractable and efficient but were developed for completely observed data. Due to different loss control strategies, insurance loss data are affected by truncation (due to deductibles), censoring (due to policy limits as well as interval censoring), scaling (due to coinsurance proportions), inflation, and other transformations. Therefore, our motivation in this dissertation is to redesign MTM and MWM for such transformed loss data.

1.2 Literature Review

Among many methods of parameter estimation for parametric models the method of moments and the maximum likelihood estimation (MLE) are commonly used in the classical statistical literature (see, e.g., Casella and Berger, 2002, Klugman *et al.*, 2012). MLE is applicable to any form of data sets (i.e., the likelihood function can always be written, Klugman and Parsa, 1993) and involves sophisticated analytical optimization arguments. MLE has a lot of desirable properties such as invariance, asymptotic optimality (in the sense of mean square error) and efficiency, and consistency (see Casella and Berger, 2002, Serfling, 1980, van der Vaart, 1998). On the other hand, MLE is not free from flaws such as non-robustness, possible non-existence, and computational intractability.

Due to the sensitivity of classical statistical estimation procedures to initial model assumptions (see Tukey, 1960), researchers (see, e.g., Hampel, 1968, 1974, Huber, 1964) have become aware and started developing more stable (insensitive) statistical estimation procedures (see Huber and Ronchetti, 2009), which were popularized under the name "robust." The primary focus of the robust procedure is to produce more resistant, stable, and efficient estimators. By design the robust estimators yield a good performance when there are small perturbations from the assumed underlying true distribution (see, e.g., Maronna *et al.*, 2006, for details).

Two broad classes of robust estimators – Method of Trimmed Moments (MTM) and Method of Winsorized Moments (MWM) – have been recently developed in the actuarial and statistical literatures. Both approaches are sufficiently general, belong to the class of L-statistics and thus produce estimators that are robust, computationally efficient and transparent (see Brazauskas, 2009, Brazauskas *et al.*, 2009, Chernoff *et al.*, 1967, Zhao *et al.*, 2018a,b). Fully worked out examples of MTM estimators are available for location-scale families (Brazauskas, 2009, Brazauskas *et al.*, 2009), log-folded-normal, log-folded-Cauchy, and log-folded-t distributions with known degrees of freedom (Brazauskas and Kleefeld, 2011), as well as exponential, single-parameter Pareto, generalized Pareto (Brazauskas and Kleefeld, 2009), and gamma distributions (Kleefeld and Brazauskas, 2012).

As mentioned earlier, in the current actuarial practice, statistical inference for loss models is almost exclusively MLE-based. Besides typical non-robustness of such procedures, MLE implementation on real data is also technically challenging (see discussions by Frees, 2017 and Lee, 2017). This issue is especially evident when one tries to fit complicated multi-parameter models such as mixtures of Erlangs (see Reynkens *et al.*, 2017, Verbelen *et al.*, 2015). Taking this discussion into account, we will redesign the MTM approach for insurance loss data and models with the expectation that it will simplify computation and improve robustness.

1.3 Preliminaries

This section presents the most relevant technical tools and facts which will be used in the rest of the dissertation.

Let $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ be a parameter vector to be estimated. Consider a sequence of estimators $\widehat{\boldsymbol{\theta}}_n$ based on an observed sample $X_1, X_2, \ldots, X_n$ from a population with cumulative distribution function (cdf) $F(\cdot|\boldsymbol{\theta})$, and probability density function (pdf) $f(\cdot|\boldsymbol{\theta})$. The optimality of an estimator is measured in terms of the minimum possible asymptotic variance and is formally defined as follows (see Serfling, 1980).

Definition 1.1. A sequence $\widehat{\boldsymbol{\theta}}_n$ is called *asymptotically efficient estimators* of $\boldsymbol{\theta}$ if

$$\widehat{\boldsymbol{\theta}}_n \sim \mathcal{AN}\left(\boldsymbol{\theta}, \frac{1}{n}\left[\boldsymbol{I}(\boldsymbol{\theta})\right]^{-1}\right), \tag{1.1}$$

where $\mathcal{AN}$ stands for "asymptotically normal" and

$$\boldsymbol{I}(\boldsymbol{\theta}) = \left[\mathbb{E}\left\{\frac{\partial \log\left(f(x|\boldsymbol{\theta})\right)}{\partial\theta_i}\;\frac{\partial \log\left(f(x|\boldsymbol{\theta})\right)}{\partial\theta_j}\right\}\right]_{k\times k} = \left[-\mathbb{E}\left\{\frac{\partial^2 \log\left(f(x|\boldsymbol{\theta})\right)}{\partial\theta_i\partial\theta_j}\right\}\right]_{k\times k},$$

is the Fisher information matrix. Further, since $\left[\boldsymbol{I}(\boldsymbol{\theta})\right]^{-1}$ is finite, relation (1.1) implies that $\widehat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}$. □

Definition 1.1 evaluates the performance of a single estimator. In order to compare the performance of two estimators of the same parameter, the following definition is handy (see Serfling, 1980, for more details).

Definition 1.2. Let $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\boldsymbol{\theta}}_n^*$ be two sequences of estimators of $\boldsymbol{\theta}$ with their respective asymptotic variance-covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^*$. Then the *asymptotic relative efficiency* (ARE) of $\widehat{\boldsymbol{\theta}}_n^*$ with respect to $\widehat{\boldsymbol{\theta}}_n$ is defined as the ratio of the determinants (of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^*$) raised to the power $1/k$:

$$ARE(\widehat{\boldsymbol{\theta}}_n^*, \widehat{\boldsymbol{\theta}}_n) := \left(\frac{det(\boldsymbol{\Sigma})}{det(\boldsymbol{\Sigma}^*)} \right)^{1/k}. \tag{1.2}$$

□

In addition, to evaluate asymptotic properties of functions of asymptotically normal vectors, the delta method is a key tool to use (see Serfling, 1980, van der Vaart, 1998).

Theorem 1.1. *Suppose that* $\widehat{\boldsymbol{\theta}}_n = \left(\widehat{\theta}_{1n}, \widehat{\theta}_{2n}, ..., \widehat{\theta}_{kn}\right) \sim \mathcal{AN}(\boldsymbol{\theta}, n^{-1}\boldsymbol{\Sigma})$ *with* $\boldsymbol{\Sigma}$ *a covariance matrix and neither* $\boldsymbol{\theta}$ *nor* $\boldsymbol{\Sigma}$ *depend on* n. *Let* $g = (g_1, ..., g_m)$, *with each* $g_i : \mathbb{R}^k \to \mathbb{R}$ *for* $1 \leq i \leq m$, *a totally differentiable function with nonzero differential at* $\boldsymbol{\theta}$. *Consider* $\boldsymbol{D} = [d_{ij}]_{m \times k} = \left[\left. \frac{\partial g_i}{\partial x_j} \right|_{\boldsymbol{x}=\boldsymbol{\theta}} \right]_{m \times k}$. *Then* $g(\widehat{\boldsymbol{\theta}}_n) \sim \mathcal{AN}\left(g(\boldsymbol{\theta}), n^{-1}\left(\boldsymbol{D\Sigma D'}\right)\right)$.

1.4 Organization of the Dissertation

The remainder of this dissertation is organized as follows. In Chapter 2, we describe different types of loss data transformations that appear in insurance contract specifications (which are due to the loss control strategies used to construct the contract). In particular, assuming that all observed data satisfy the *i.i.d.* assumption, we study: the complete (not transformed) data scenario; left- and right-truncated data; left- and right-censored data; left-truncated, right-censored, and linearly-transformed data (also known as *payment-per-payment* variable); and interval-censored and linearly-transformed data (also known as *payment-per-loss* variable).

In Chapter 3, we introduce and develop two estimation procedures – maximum likelihood (MLE) and method of trimmed moments (MTM) – for all loss data scenarios discussed in Chapter 2. Taking into account these data transformations, we specify the relevant log-likelihood functions, define sample and population trimmed moments, and describe the procedure for finding MTM estimators. Then asymptotic properties of MLE and MTM estimators are rigorously studied.

In Chapter 4, we use the general formulation of the estimators and specialize them for the exponential and normal distributions. This includes derivation of their computing formulas (or estimating equations) and specification of mean vectors and variance-covariance matrices for their asymptotically normal distributions.

In Chapter 5, MLE and MTM estimators are implemented for the single-parameter Pareto and lognormal models that are fitted to Norwegian fire claims data for the year 1983. The effects of model fitting on insurance contract pricing are then investigated.

In Chapter 6 , we explore several methodological extensions of the newly designed MTM estimators for complete, grouped and exponentially distributed random variables. Specifically, we construct truncated, censored, and insurance payment-type estimators and prove a series of theoretical results about those estimators' existence and asymptotic normality. Our analysis reveals new connections between data truncation, trimming, and censoring.

Finally, in Chapter 7, we summarize the results of this dissertation and briefly discuss our future research plans.

Chapter 2

Loss Data and Models

In this chapter, we review typical transformations of continuous random variables that may be encountered in modeling claim severity. For each type of variable transformation, the resulting probability density function (pdf), cumulative distribution function (cdf) and quantile function (qf) are specified. For some of these descriptions, we closely follow Klugman *et al.* (2012, Sections 12.1 and 13.2).

2.1 Complete Data

Following many standard textbooks on probability and mathematical statistics, we start with the complete data scenario. Suppose the observable random variables

$$X_1, X_2, \ldots, X_n \tag{2.1}$$

are independent and identically distributed ($i.i.d.$) and have the pdf $f(x)$, cdf $F(x)$, sf $S(x) = 1 - F(x)$, and qf $F^{-1}(s)$, $0 < s < 1$. Since loss random variables are nonnegative, the support of $f(x)$ is the set $\{x : x \geq 0\}$. In many practical situations, the $i.i.d.$ assumption seems reasonable, but see Section 7.2.2 for a discussion of other distributional assumptions.

The complete data scenario is not common when claim severities are recorded, but it represents so-called "ground up" losses and thus important to consider. Statistical properties of the ground-up variable are of great interest in risk analysis, product design (for specifying insurance contract parameters), risk transfer considerations, and for other business decisions.

2.2 Truncated Data

Data truncation occurs when sample observations are restricted to some interval, say $(t, T]$ (not necessarily finite, e.g., $T \to \infty$). Measurements and even a count of observations outside the interval are completely unknown. To formalize this discussion, we will say that we observe the *i.i.d.* data

$$X_1^*,\, X_2^*, \ldots, X_n^*, \tag{2.2}$$

where each X^* is equal to the ground-up variable X, if X falls between t and T, and is undefined otherwise. That is, X^* satisfies the following conditional event relationship

$$X^* \overset{d}{=} X \,\big|\, t < X \le T,$$

where $\overset{d}{=}$ denotes "equal in distribution." Due to this relationship, the cdf F_*, pdf f_*, and qf F_*^{-1} of variables X^* are related to F, f, and F^{-1} (see Section 2.1) and given by:

$$F_*(x \,|\, t, T) = \mathbb{P}(X^* \le x | t, T) = \mathbb{P}\left[X \le x \,\big|\, t < X \le T\right] = \begin{cases} 0, & x \le t; \\ \frac{F(x) - F(t)}{F(T) - F(t)}, & t < x \le T; \\ 1, & x > T, \end{cases} \tag{2.3}$$

$$f_*(x \,|\, t, T) = \frac{d}{dx}\Big[F_*(x \,|\, t, T)\Big] = \begin{cases} \frac{f(x)}{F(T) - F(t)}, & t < x \le T; \\ 0, & \text{elsewhere}, \end{cases} \tag{2.4}$$

and

$$F_*^{-1}(s \,|\, t, T) = F^{-1}\big(sF(T) + (1-s)F(t)\big), \quad \text{for } \ 0 \le s \le 1. \tag{2.5}$$

In industry wide databases (such as ORX Loss Data), only losses above some pre-specified threshold, say d, are collected, which results in the *left truncated* data at d. Thus, the observations available to the end-user can be viewed as a realization of random variables (2.2) with $t = d$ and $T \to \infty$. The latter condition slightly simplifies formulas (2.3)–(2.5); one just needs to replace $F(T)$ with 1.

2.3 Censored Data

There are several versions of data censoring that occur in statistical modeling: interval censoring (it includes left and right censoring depending on which end point of the interval is infinite), type I censoring, type II censoring, and random censoring. For actuarial work, the most relevant type is *interval censoring*. It occurs when complete sample observations are available within some interval, say $(t, T]$, but data outside the interval is only partially known. That is, counts are available but actual values are not. To formalize this discussion, we will say that we observe the *i.i.d.* data

$$X_1^{**}, X_2^{**}, \ldots, X_n^{**}, \tag{2.6}$$

where each X^{**} is equal to the ground-up variable X, if X falls between t and T, and is equal to the corresponding end-point of the interval if X is beyond that point. That is, X^{**} is given by

$$X^{**} = \min\big\{\max(t, X),\, T\big\} = \begin{cases} t, & X \le t; \\ X, & t < X \le T; \\ T, & X > T. \end{cases}$$

Due to this relationship, the cdf F_{**}, pdf f_{**}, and qf F_{**}^{-1} of variables X^{**} are related to F, f, and F^{-1} and given by:

$$F_{**}(x \,|\, t, T) = \mathbb{P}\Big[\min\big\{\max(t, X),\, T\big\} \le x\Big] = \begin{cases} 0, & x < t; \\ F(x), & t \le x < T; \\ 1, & x \ge T, \end{cases} \tag{2.7}$$

$$f_{**}(x \,|\, t, T) = \frac{d}{dx}\Big[F_{**}(x \,|\, t, T)\Big] = \begin{cases} F(t), & x = t; \\ f(x), & t < x < T; \\ S(T), & x = T; \\ 0, & \text{elsewhere}, \end{cases} \tag{2.8}$$

and

$$F_{**}^{-1}(s \,|\, t, T) = \begin{cases} t, & s < F(t); \\ F^{-1}(s), & F(t) \le s < F(T); \\ T, & s \ge F(T). \end{cases} \tag{2.9}$$

2.4 Insurance Payments

Insurance contracts have coverage modifications that need to be taken into account when modeling the underlying loss variable. Usually the coverage modifications such as deductibles, policy limits, and coinsurance are introduced as loss control strategies so that unfavorable policyholder behavioral effects (e.g., adverse selection) can be minimized. There are also situations when certain features of the contract emerge naturally (e.g., the value of insured property in general insurance is a natural upper policy limit). Here we describe two common transformations of the loss variable along with the corresponding cdf's, pdf's, and qf's.

Suppose the insurance contract has ordinary deductible d, upper policy limit u, and coinsurance rate c ($0 \leq c \leq 1$). These coverage parameters imply that when a loss X is reported, the insurance company is responsible for a proportion c of X exceeding d, but no more than $c(u-d)$.

Next, if the loss severity X below the deductible d is completely unobservable (even its frequency is unknown), then the observed *i.i.d.* insurance payments $Y_1, \ldots, Y_n$ can be viewed as realizations of *left-truncated*, *right-censored*, and *linearly-transformed* (also known as *per-payment* variable) X:

$$Y \stackrel{d}{=} X | X > d = \begin{cases} c\,(X-d)\,, & d < X \leq u; \\ c\,(u-d)\,, & u < X. \end{cases} \tag{2.10}$$

We can see that the payment variable Y is a linear transformation of a composition of variables X^* and X^{**} (see Sections 2.2 and 2.3). Thus, similar to variables X^* and X^{**}, its cdf G_Y, pdf g_Y, and qf G_Y^{-1} are also related to F, f, and F^{-1} and given by:

$$G_Y(y \,|\, c, d, u) = \mathbb{P}\left[Y \leq y | X > d\right] = \begin{cases} 0, & y \leq 0; \\ \frac{F(y/c+d)-F(d)}{S(d)}, & 0 < y \leq c(u-d); \\ 1, & y > c(u-d), \end{cases} \tag{2.11}$$

$$g_Y(y \,|\, c, d, u) = \frac{d}{dy}\Big[G_Y(y \,|\, c, d, u)\Big] = \begin{cases} \frac{f(y/c+d)}{c[S(d)]}, & 0 < y < c(u-d); \\ \frac{S(u)}{S(d)}, & y = c(u-d); \\ 0, & \text{elsewhere}, \end{cases} \tag{2.12}$$

and

$$G_Y^{-1}(s \,|\, c, d, u) = \begin{cases} c\left[F^{-1}\big(s + (1-s)F(d)\big) - d\right], & 0 \leq s < \frac{F(u)-F(d)}{S(d)}; \\ c(u-d), & \frac{F(u)-F(d)}{S(d)} \leq s \leq 1. \end{cases} \tag{2.13}$$

The scenario that no information is available about X below d is likely to occur when modeling is done based on the data acquired from a third party (e.g., data vendor). For payment data collected in house, the information about the number of policies that did not report claims (equivalently, resulted in a payment of 0) would be available. This minor modification yields different payment variables, say $Z_1, \ldots, Z_n$, which can be treated as *i.i.d.* realizations of *interval-censored* and *linearly-transformed* (also known as *per-loss* variable) X:

$$Z = \begin{cases} 0, & X \leq d; \\ c\,(X-d)\,, & d < X \leq u; \\ c\,(u-d)\,, & u < X. \end{cases} \tag{2.14}$$

Again, its cdf G_Z, pdf g_Z, and qf G_Z^{-1} are related to F, f, and F^{-1} and given by:

$$G_Z(z \,|\, c, d, u) = \mathbb{P}\left[Z \leq z\right] = \begin{cases} 0, & z < 0; \\ F(z/c + d), & 0 \leq z \leq c(u-d); \\ 1, & z > c(u-d), \end{cases} \tag{2.15}$$

$$g_Z(z \,|\, c, d, u) = \frac{d}{dz}\Big[G_Z(z \,|\, c, d, u)\Big] = \begin{cases} F(d), & z = 0; \\ f(z/c + d)/c, & 0 < z < c(u-d); \\ S(u), & z = c(u-d); \\ 0, & \text{elsewhere}, \end{cases} \tag{2.16}$$

and

$$G_Z^{-1}(s \,|\, c, d, u) = \begin{cases} 0, & 0 \leq s \leq F(d); \\ c\,(F^{-1}(s) - d)\,, & F(d) < s < F(u); \\ c(u-d), & F(u) \leq s \leq 1. \end{cases} \tag{2.17}$$

2.5 An Example

For illustrative purposes and to get a better understanding of insurance loss control strategies, let us consider the well-known data set of 30 most damaging hurricanes in the United States from 1925 to 1995 (see, e.g., Pielke and Landsea, 1998). Table 2.1 presents different insurance modifications of the hurricane data in billions of dollars (rounded to two decimal places).

Table 2.1: Top 30 most damaging hurricane losses in the United States from 1925 to 1995 under different data and payment transformations.

Complete losses	Truncated losses $(t, T) = (5, 25)$	Censored losses $(t, T) = (5, 25)$	Pmt.-per-loss, Z $(t, T, c) = (5, 25, .9)$	Pmt.-per-pmt., Y $(t, T, c) = (5, 25, .9)$
2.27	–	5.00	0.00	–
2.40	–	5.00	0.00	–
2.40	–	5.00	0.00	–
2.44	–	5.00	0.00	–
3.00	–	5.00	0.00	–
3.11	–	5.00	0.00	–
3.34	–	5.00	0.00	–
4.06	–	5.00	0.00	–
5.37	5.37	5.37	0.33	0.33
5.84	5.84	5.84	0.75	0.75
6.30	6.30	6.30	1.16	1.16
6.31	6.31	6.31	1.18	1.18
6.54	6.54	6.54	1.38	1.38
7.04	7.04	7.04	1.84	1.84
7.07	7.07	7.07	1.86	1.86
8.31	8.31	8.31	2.98	2.98
9.07	9.07	9.07	3.66	3.66
9.38	9.38	9.38	3.94	3.94
10.23	10.23	10.23	4.71	4.71
10.71	10.71	10.71	5.13	5.13
10.97	10.97	10.97	5.37	5.37
12.05	12.05	12.05	6.34	6.34
12.43	12.43	12.43	6.69	6.69
13.80	13.80	13.80	7.92	7.92
16.63	16.63	16.63	10.47	10.47
16.86	16.86	16.86	10.68	10.68
22.60	22.60	22.60	15.84	15.84
26.62	–	25.00	18.00	18.00
33.09	–	25.00	18.00	18.00
72.30	–	25.00	18.00	18.00

Each column is assumed to be an *i.i.d.* sample from the corresponding distribution. For example, first column is an *i.i.d.* sample of size 30 given by (2.1), second column is an *i.i.d.* sample of size 19 given by (2.2), third column is an *i.i.d.* sample of size 30 given by (2.6), fourth column is an *i.i.d.* sample of size 30 given by (2.14), and the fifth column is an *i.i.d.* sample of size 22 given by (2.10).

Chapter 3

Parameter Estimation

In this chapter we present two estimation procedures – MLE and MTM – along with their asymptotic properties for different loss data scenarios from Chapter 2.

3.1 Maximum Likelihood Estimation

In classical point estimation theory, the method of maximum likelihood estimation (MLE) is the most popular among other methods. MLE, by definition, uses entire sample and is the maximizer of the likelihood function (which is a joint probability density function and/or probability mass function). In the *i.i.d.* case, the likelihood function will be the product of the marginals (pdf's and/or pmf's). Typically, MLE estimators are found by making the logarithmic transformation of the likelihood function, setting its first partial derivatives equal to zero, and solving the resulting system of equations.

3.1.1 Definition

Definition 3.1. The *likelihood function* of the parameter vector $\boldsymbol{\theta}$ for an observed sample $x_1 \in A_1, ..., x_n \in A_n$, where $A_1, \ldots, A_n$ are the events (for example A_j may consist of a single point or an interval), is defined as

$$L(\boldsymbol{\theta}|x_1, \ldots, x_n) = \prod_{i=1}^{n} \mathbb{P}(x_i \in A_i), \tag{3.1}$$

and the corresponding *log-likelihood function* is

$$
\begin{aligned}
l(\boldsymbol{\theta}|x_1,\dots,x_n) &= \log L(\boldsymbol{\theta}|x_1,\dots,x_n) \\
&= \sum_{i=1}^{n} \log\left(\mathbb{P}(x_i \in A_i)\right). \qquad (3.2)
\end{aligned}
$$

The maximizer vector of either the likelihood function (3.1) or the log-likelihood function (3.2) is called the maximum likelihood estimate of the parameter vector $\boldsymbol{\theta}$. □

In the following examples, we specify the likelihood and log-likelihood functions for the data scenarios and models of Chapter 2.

Example 3.1. *Complete Data.*

For an *i.i.d.* sample $x_1,\dots,x_n$ with pdf $f(x|\boldsymbol{\theta})$, the general likelihood (3.1) and log-likelihood (3.2) functions reduce to

$$
L(\boldsymbol{\theta}|x_1,\dots,x_n) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}), \qquad (3.3)
$$

$$
\begin{aligned}
l(\boldsymbol{\theta}|x_1,\dots,x_n) &= \log L(\boldsymbol{\theta}|x_1,\dots,x_n) \\
&= \sum_{i=1}^{n} \log\left(f(x_i|\boldsymbol{\theta})\right), \qquad (3.4)
\end{aligned}
$$

and the MLE of $\boldsymbol{\theta}$ is a maximizer vector of either (3.3) or (3.4). □

Example 3.2. *Truncated Data.*

For an *i.i.d.* sample $x_1^*,\dots,x_n^*$ defined by (2.2) with cdf (2.3), and pdf (2.4), the corresponding likelihood and log-likelihood functions are

$$
\begin{aligned}
L_*(\boldsymbol{\theta}|x_1^*,\dots,x_n^*) &= \prod_{i=1}^{n} f_*(x_i^*) \\
&= \prod_{i=1}^{n} \frac{f(x_i^*|\boldsymbol{\theta})}{F(T|\boldsymbol{\theta}) - F(t|\boldsymbol{\theta})} \\
&= \frac{\prod_{i=1}^{n} f(x_i^*|\boldsymbol{\theta})}{[F(T|\boldsymbol{\theta}) - F(t|\boldsymbol{\theta})]^n}, \qquad (3.5)
\end{aligned}
$$

$$l_*(\boldsymbol{\theta}|x_1^*,\dots,x_n^*) \;=\; \sum_{i=1}^{n} \log\left(f(x_i^*|\boldsymbol{\theta})\right) - n\log\left(F(T|\boldsymbol{\theta}) - F(t|\boldsymbol{\theta})\right). \tag{3.6}$$

$\square$

Example 3.3. *Censored Data.*

For an observed *i.i.d.* sample $x_1^{**},\dots,x_n^{**}$ defined by (2.6) with cdf (2.7), and pdf (2.8), the likelihood and log-likelihood functions are

$$\begin{aligned}
L_{**}(\boldsymbol{\theta}|x_1^{**},\dots,x_n^{**}) &= \prod_{i=1}^{n} f_{**}(x_i^{**}) \\
&= \left(\prod_{x_i^{**}=t} F(t|\boldsymbol{\theta})\right)\left(\prod_{t<x_i^{**}<T} f(x_i^{**}|\boldsymbol{\theta})\right)\left(\prod_{x_i^{**}=T}\left[1-F(T|\boldsymbol{\theta})\right]\right) \\
&= \left(F(t|\boldsymbol{\theta})\right)^{\sum_{i=1}^{n}\mathbb{1}\{x_i^{**}=t\}}\left(\prod_{t<x_i^{**}<T} f(x_i^{**}|\boldsymbol{\theta})\right)\left(S(T|\boldsymbol{\theta})\right)^{\sum_{i=1}^{n}\mathbb{1}\{x_i^{**}=T\}},
\end{aligned} \tag{3.7}$$

$$\begin{aligned}
l_{**}(\boldsymbol{\theta}|x_1^{**},\dots,x_n^{**}) &= \log\left(F(t|\boldsymbol{\theta})\right)\sum_{i=1}^{n}\mathbb{1}\{x_i^{**}=t\} + \sum_{t<x_i^{**}<T}\log\left(f(x_i^{**}|\boldsymbol{\theta})\right) \\
&\quad + \log\left(S(T|\boldsymbol{\theta})\right)\sum_{i=1}^{n}\mathbb{1}\{x_i^{**}=T\}.
\end{aligned} \tag{3.8}$$

$\square$

Example 3.4. *Payment-per-payment Data.*

For an observed *i.i.d.* sample $y_1,\dots,y_n$ defined by (2.10) with cdf (2.11), and pdf (2.12), the likelihood and log-likelihood functions are

$$\begin{aligned}
L_{\otimes}(\boldsymbol{\theta}|y_1,\dots,y_n) &= \left(\prod_{0<y_i<c(u-d)} \frac{f\left(\frac{y_i}{c}+d|\boldsymbol{\theta}\right)}{cS(d|\boldsymbol{\theta})}\right)\left(\prod_{y_i=c(u-d)} \frac{1-F(u|\boldsymbol{\theta})}{1-F(d|\boldsymbol{\theta})}\right) \\
&= \frac{\left(S(u|\boldsymbol{\theta})\right)^{\sum_{i=1}^{n}\mathbb{1}\{y_i=c(u-d)\}}\, c^{-\sum_{i=1}^{n}\mathbb{1}\{0<y_i<c(u-d)\}}}{\left(S(d|\boldsymbol{\theta})\right)^{n}} \\
&\quad \times \left(\prod_{0<y_i<c(u-d)} f\left(\frac{y_i}{c}+d|\boldsymbol{\theta}\right)\right),
\end{aligned} \tag{3.9}$$

$$l_{\otimes}(\boldsymbol{\theta}|y_1,\ldots,y_n) = \log\left(S(u|\boldsymbol{\theta})\right)\sum_{i=1}^{n}\mathbb{1}\{y_i = c(u-d)\} - n\log\left(S(d|\boldsymbol{\theta})\right)$$
$$- \sum_{i=1}^{n}\mathbb{1}\{0 < y_i < c(u-d)\}\left(\log(c) - \log\left(f\left(\frac{y_i}{c} + d|\boldsymbol{\theta}\right)\right)\right). \tag{3.10}$$

□

Example 3.5. *Payment-per-loss Data.*

For an *i.i.d.* sample $z_1,\ldots,z_n$ defined by (2.14) with cdf (2.15), and pdf (2.16), the corresponding likelihood and log-likelihood functions are

$$L_{\otimes\otimes}(\boldsymbol{\theta}|\boldsymbol{z}) = \left[F(d|\boldsymbol{\theta})\right]^{\sum_{i=1}^{n}\mathbb{1}\{z_i=0\}}\left(\prod_{0<z_j<c(u-d)}\frac{f\left(\frac{z_i}{c}+d|\boldsymbol{\theta}\right)}{c}\right)$$
$$\times\left[S(u|\boldsymbol{\theta})\right]^{\sum_{i=1}^{n}\mathbb{1}\{z_i=c(u-d)\}}, \tag{3.11}$$

$$l_{\otimes\otimes}(\boldsymbol{\theta}|\boldsymbol{z}) = \log\left(F(d|\boldsymbol{\theta})\right)\sum_{i=1}^{n}\mathbb{1}\{z_i = 0\} + \log\left(S(u|\boldsymbol{\theta})\right)\sum_{i=1}^{n}\mathbb{1}\{z_i = c(u-d)\}$$
$$- \sum_{i=1}^{n}\mathbb{1}\{0 < z_i < c(u-d)\}\left(\log(c) - \log\left(f\left(\frac{z_i}{c} + d|\boldsymbol{\theta}\right)\right)\right), \tag{3.12}$$

where $\boldsymbol{z} := (z_1,\ldots,z_n)$. □

3.1.2 Asymptotic Properties

Under certain regularity conditions (see, e.g., Serfling, 1980) on pdf and the likelihood function, MLEs are consistent, efficient, and asymptotically normal. In the following sequence of examples, we summarize the asymptotic properties of MLEs for different loss data scenarios from Chapter 2.

Example 3.6. *Complete Data.*

For an *i.i.d.* sample $x_1,\ldots,x_n$ with pdf $f(x|\boldsymbol{\theta})$, MLE of $\boldsymbol{\theta}$ is found by maximizing the likelihood function (3.3); let us denote it as $\widehat{\boldsymbol{\theta}}_n$. Then, according to Definition 1.1,

$$\widehat{\boldsymbol{\theta}}_n \sim \mathcal{AN}\left(\boldsymbol{\theta}, \frac{1}{n}\left[\boldsymbol{I}(\boldsymbol{\theta})\right]^{-1}\right),$$

which also implies that $\widehat{\boldsymbol{\theta}}_n$ is consistent and efficient. $\square$

Example 3.7. *Truncated Data.*

For an *i.i.d.* sample $x_1^*, \ldots, x_n^*$ defined by (2.2) with pdf (2.4) and cdf (2.3), MLE of $\boldsymbol{\theta}$ is found by maximizing the likelihood function (3.5) and/or the log-likelihood function (3.6); let us denote it as $\widehat{\boldsymbol{\theta}}_n$. Then $\widehat{\boldsymbol{\theta}}_n \sim \mathcal{AN}\left(\boldsymbol{\theta}, \frac{1}{n}\left[\boldsymbol{I}_*(\boldsymbol{\theta})\right]^{-1}\right)$, where the Fisher information matrix $\boldsymbol{I}_*(\boldsymbol{\theta})$ is given by

$$\Bigg[\mathbb{E}\Bigg\{\frac{\partial\left[\log\left(f(x|\boldsymbol{\theta})\right) - \log\left(F(T|\boldsymbol{\theta}) - F(t|\boldsymbol{\theta})\right)\right]}{\partial\theta_i} \times \frac{\partial\left[\log\left(f(x|\boldsymbol{\theta})\right) - \log\left(F(T|\boldsymbol{\theta}) - F(t|\boldsymbol{\theta})\right)\right]}{\partial\theta_j}\Bigg\}\Bigg]_{k\times k}. \tag{3.13}$$

$\square$

Example 3.8. *Censored Data.*

For an *i.i.d.* sample $x_1^{**}, \ldots, x_n^{**}$ defined by (2.6) with pdf (2.8) and cdf (2.7), MLE of $\boldsymbol{\theta}$ is found by maximizing the likelihood function (3.7) and/or the log-likelihood function (3.8); let us denote it as $\widehat{\boldsymbol{\theta}}_n$. Then $\widehat{\boldsymbol{\theta}}_n \sim \mathcal{AN}\left(\boldsymbol{\theta}, \frac{1}{n}\left[\boldsymbol{I}_{**}(\boldsymbol{\theta})\right]^{-1}\right)$, where the (i, j)th entry of $\boldsymbol{I}_{**}(\boldsymbol{\theta})$ is given by

$$\mathbb{E}\Bigg\{\frac{\partial\left[\log\left(F(t|\boldsymbol{\theta})\right)\mathbb{1}\{x=t\} + \log\left(f(x|\boldsymbol{\theta})\right)\mathbb{1}\{t<x<T\} + \log\left(S(T|\boldsymbol{\theta})\right)\mathbb{1}\{x=T\}\right]}{\partial\theta_j} \times \frac{\partial\left[\log\left(F(t|\boldsymbol{\theta})\right)\mathbb{1}\{x=t\} + \log\left(f(x|\boldsymbol{\theta})\right)\mathbb{1}\{t<x<T\} + \log\left(S(T|\boldsymbol{\theta})\right)\mathbb{1}\{x=T\}\right]}{\partial\theta_j}\Bigg\}. \tag{3.14}$$

$\square$

Example 3.9. *Payment-per-payment Data.*

For an *i.i.d.* sample $y_1, \ldots, y_n$ defined by (2.10) with pdf (2.12) and cdf (2.11), MLE of $\boldsymbol{\theta}$ is found by maximizing the likelihood function (3.9) and/or the log-likelihood function (3.10); let us denote it as $\widehat{\boldsymbol{\theta}}_n$. Then $\widehat{\boldsymbol{\theta}}_n \sim \mathcal{AN}\left(\boldsymbol{\theta}, \frac{1}{n}\left[\boldsymbol{I}_{\otimes}(\boldsymbol{\theta})\right]^{-1}\right)$, where the (i, j)th entry of $\boldsymbol{I}_{\otimes}(\boldsymbol{\theta})$ is given by

$$\mathbb{E}\Bigg\{\frac{\partial\left[\log\left(S(u|\boldsymbol{\theta})\right)\mathbb{1}\{y=c^*\} - \log\left(S(d|\boldsymbol{\theta})\right) - \xi_{\otimes}(c, y^*)\mathbb{1}\{0<y<c^*\}\right]}{\partial\theta_i}$$

$$\times \frac{\partial \left[\log\left(S(u|\boldsymbol{\theta})\right)\mathbb{1}\{y=c^*\} - \log\left(S(d|\boldsymbol{\theta})\right) - \xi_{\otimes}(c,y^*)\mathbb{1}\{0<y<c^*\}\right)\right]}{\partial\theta_j}\Bigg\}, \tag{3.15}$$

where $\xi_{\otimes}(c,y^*) := \log(c) - \log\left(f\left(y^*|\boldsymbol{\theta}\right)\right)$, $y^* := \frac{y}{c} + d$, and $c^* := c(u-d)$. $\square$

Example 3.10. *Payment-per-loss Data.*

For an *i.i.d.* sample $z_1, \ldots, z_n$ defined by (2.14) with pdf (2.16) and cdf (2.15), MLE of $\boldsymbol{\theta}$ is found by maximizing the likelihood function (3.11) and/or the log-likelihood function (3.12); let us denote it as $\widehat{\boldsymbol{\theta}}_n$. Then $\widehat{\boldsymbol{\theta}}_n \sim \mathcal{AN}\left(\boldsymbol{\theta}, \frac{1}{n}\left[\boldsymbol{I}_{\otimes\otimes}(\boldsymbol{\theta})\right]^{-1}\right)$, where the (i,j)th entry of $\boldsymbol{I}_{\otimes\otimes}(\boldsymbol{\theta})$ is given by

$$\mathbb{E}\Bigg\{\frac{\partial\left[\log\left(F(d|\boldsymbol{\theta})\right)\mathbb{1}\{z=0\} + \log\left(S(u|\boldsymbol{\theta})\right)\mathbb{1}\{z=c^*\} - \xi_{\otimes\otimes}(c,z^*)\mathbb{1}\{0<z<c^*\}\right]}{\partial\theta_i}$$

$$\times \frac{\partial\left[\log\left(F(d|\boldsymbol{\theta})\right)\mathbb{1}\{z=0\} + \log\left(S(u|\boldsymbol{\theta})\right)\mathbb{1}\{z=c^*\} - \xi_{\otimes\otimes}(c,z^*)\mathbb{1}\{0<z<c^*\}\right]}{\partial\theta_j}\Bigg\},$$

where $\xi_{\otimes\otimes}(c,z^*) := \log(c) - \log\left(f\left(z^*|\boldsymbol{\theta}\right)\right)$, $z^* := \frac{z}{c} + d$, and $c^* := c(u-d)$. $\square$

3.2 Method of Trimmed Moments

MTM works like the method of moments but is designed to reduce the effect of possible spurious outliers. To control the influence of extremes, a general strategy is to trim certain proportion of the ordered sample data on both tails (for example, 5% of lower statistics and 10% of upper) and then apply the method of moments on the remaining data. The choice of trimming proportions allows the user to balance robustness and efficiency trade-offs (see Brazauskas *et al.*, 2009, for details).

3.2.1 Definition

Let $X_1, X_2, ..., X_n \overset{i.i.d.}{\sim} X$, random variables, where $X \sim F(x|\boldsymbol{\theta})$ with k unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. Denote the order statistics of $X_1, \ldots, X_n$ by $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$. Then the MTM estimators of $\theta_1, \theta_2, ..., \theta_k$ are found as follows:

- Compute the sample trimmed moments

$$\widehat{\mu}_j = \frac{1}{n - m_n(j) - m_n^*(j)} \sum_{i=m_n(j)+1}^{n-m_n^*(j)} h_j(X_{i:n}), \qquad 1 \leq j \leq k. \tag{3.16}$$

The $h_j's$ in (3.16) are specially chosen functions and $m_n(j)$ and $m_n^*(j)$ are integers such that $0 \leq m_n(j) < n - m_n^*(j) \leq n$ with $\frac{m_n(j)}{n} \to a_j$ and $\frac{m_n^*(j)}{n} \to b_j$ as $n \to \infty$, where the proportions a_j and b_j are chosen by researcher.

- Compute the corresponding population trimmed moments

$$\mu_j = \frac{1}{1 - a_j - b_j} \int_{a_j}^{1-b_j} h_j(F^{-1}(u|\boldsymbol{\theta}))\, du, \quad 1 \leq j \leq k. \tag{3.17}$$

  In (3.17), $F^{-1}(u|\boldsymbol{\theta}) = \inf\{x : F(x|\boldsymbol{\theta}) \geq u\}$ is the quantile function.

- Now, match the sample and population trimmed moments from (3.16) and (3.17) to get the following system of equations for $\theta_1, \theta_2, ..., \theta_k$

$$\begin{cases} \mu_1(\theta_1, \ldots, \theta_k) & = & \widehat{\mu}_1 \\ & \vdots & \\ \mu_k(\theta_1, \ldots, \theta_k) & = & \widehat{\mu}_k \end{cases} \tag{3.18}$$

Definition 3.2. A solution, say $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_{1n}, \widehat{\theta}_{2n}, ..., \widehat{\theta}_{kn})$, if it exists, to the system of equations (3.18) is called the *method of trimmed moments (MTM)* estimator of $\boldsymbol{\theta}$. Thus, $\widehat{\theta}_{jn} =: g_j(\widehat{\mu}_1, \widehat{\mu}_2, ..., \widehat{\mu}_k)$, $1 \leq j \leq k$, are the MTM estimators of $\theta_1, \theta_2, ..., \theta_k$.

The following examples customize the MTM estimators for data scenarios of Chapter 2.

Example 3.11. *Complete Data.*

For this scenario, the MTM estimators are found according to equations (3.16) - (3.18). □

Example 3.12. *Truncated Data.*

For an *i.i.d.* sample $x_1^*, x_2^*, \ldots, x_n^*$ defined by (2.2) with cdf (2.3) and qf (2.5), the sample and population trimmed moments in equations (3.16) and (3.17), respectively, are given by

$$\widehat{\mu}_{*,j} = \frac{1}{n - m_n(j) - m_n^*(j)} \sum_{i=m_n(j)+1}^{n-m_n^*(j)} h_j(x_{i:n}^*), \qquad 1 \leq j \leq k, \tag{3.19}$$

and

$$
\begin{aligned}
\mu_{*,j} &= \frac{1}{1-a_j-b_j}\int_{a_j}^{1-b_j} h_j(F_*^{-1}(u))\,du \\
&= \frac{1}{1-a_j-b_j}\int_{a_j}^{1-b_j} h_j\left(F^{-1}\big(uF(T|\boldsymbol{\theta})+(1-u)F(t|\boldsymbol{\theta})|\boldsymbol{\theta}\big)\right)\,du, \quad 1\le j\le k.
\end{aligned}
\tag{3.20}
$$

MTM of $\boldsymbol{\theta}$ is found by solving (3.18) if a solution exists. $\square$

It is well-known in the operational risk literature that the standard method-of-moments and maximum likelihood estimators present significant technical challenges in practice (see Ergashev *et al.*, 2016). In view of this and since operational risk data is a special case of the truncated data (i.e., $T\to\infty$), the MTM estimators of Example 3.12 offer an attractive model estimation alternative.

Example 3.13. *Censored Data.*

For an *i.i.d.* sample $x_1^{**}, x_2^{**}, \ldots, x_n^{**}$ defined by (2.6) with cdf (2.7) and qf (2.9), the sample and population trimmed moments in equations (3.16) and (3.17), respectively, are given by

$$
\widehat{\mu}_{**,j} = \frac{1}{n-m_n(j)-m_n^*(j)}\sum_{i=m_n(j)+1}^{n-m_n^*(j)} h_j(x_{i:n}^{**}), \qquad 1\le j\le k, \tag{3.21}
$$

and

$$
\mu_{**,j} = \frac{1}{1-a_j-b_j}\int_{a_j}^{1-b_j} h_j(F_{**}^{-1}(u))\,du, \quad 1\le j\le k. \tag{3.22}
$$

MTM of $\boldsymbol{\theta}$ is found by solving (3.18) if a solution exists. $\square$

Example 3.14. *Payment-per-payment Data.*

For an *i.i.d.* sample $y_1, y_2, \ldots, y_n$ defined by (2.10) with cdf (2.11) and qf (2.13), the sample and population trimmed moments in equations (3.16) and (3.17), respectively, are given by

$$
\widehat{\mu}_{\otimes,j} = \frac{1}{n-m_n(j)-m_n^*(j)}\sum_{i=m_n(j)+1}^{n-m_n^*(j)} h_j(y_{i:n}), \qquad 1\le j\le k, \tag{3.23}
$$

and

$$\mu_{\otimes,j} = \frac{1}{1-a_j-b_j} \int_{a_j}^{1-b_j} h_j(G_Y^{-1}(s))\, ds, \quad 1 \leq j \leq k. \tag{3.24}$$

MTM of $\boldsymbol{\theta}$ is found by solving (3.18) if a solution exists. □

Example 3.15. *Payment-per-loss Data.*

For an *i.i.d.* sample $z_1, z_2, \ldots, z_n$ defined by (2.14) with cdf (2.15) and qf (2.17), the sample and population trimmed moments in equations (3.16) and (3.17), respectively, are given by

$$\widehat{\mu}_{\otimes\otimes,j} = \frac{1}{n-m_n(j)-m_n^*(j)} \sum_{i=m_n(j)+1}^{n-m_n^*(j)} h_j(z_{i:n}), \qquad 1 \leq j \leq k, \tag{3.25}$$

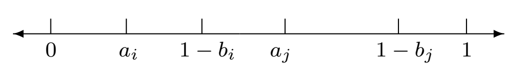
and

$$\mu_{\otimes\otimes,j} = \frac{1}{1-a_j-b_j} \int_{a_j}^{1-b_j} h_j(G_Z^{-1}(s))\, ds, \quad 1 \leq j \leq k. \tag{3.26}$$

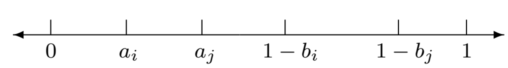
MTM of $\boldsymbol{\theta}$ is found by solving (3.18) if a solution exists. □

Note 3.1. In the procedure (3.16) - (3.18), and depending on the data scenario, there are quite a few arrangements of the proportions (a_i, b_i) and (a_j, b_j) and their positioning in the sample with respect to $F(t)$ and $F(T)$. For example, in the complete data case (i.e., $F(t) = 0$ and $F(T) = 1$), the entries of the variance-covariance matrix $\boldsymbol{\Sigma}$ (see equations 3.28 and 3.29) of the random vector $(\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_k)$ actually depend on the proportions (a_i, b_i) and (a_j, b_j) and there are six possible combinations of these proportions:

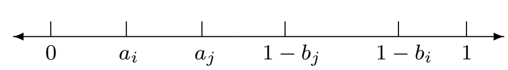
1. $0 \leq a_i \leq 1-b_i \leq a_j \leq 1-b_j \leq 1,$



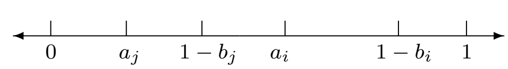
2. $0 \leq a_i \leq a_j \leq 1-b_i \leq 1-b_j \leq 1,$



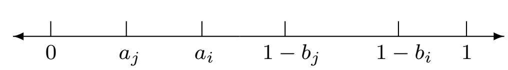
3. $0 \leq a_i \leq a_j \leq 1-b_j \leq 1-b_i \leq 1,$



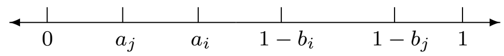
4. $0 \leq a_j \leq 1-b_j \leq a_i \leq 1-b_i \leq 1,$



5. $0 \leq a_j \leq a_i \leq 1-b_j \leq 1-b_i \leq 1,$



6. $0 \leq a_j \leq a_i \leq 1 - b_i \leq 1 - b_j \leq 1$.



Each choice results in an MTM estimator with different robustness and efficiency properties. Following the existing literature, we will consider the case when the respective lower and upper proportions of two sample trimmed moments are identical, i.e., $0 \leq a = a_i = a_j < 1-b_i = 1-b_j = 1-b \leq 1$. Also, for the other data scenarios, we will choose a and b so that they are inside the interval $[F(t), F(T)]$. Such a choice results in MTM estimators that will be resistant against outliers, i.e., observations that are inconsistent with the model and most likely appearing at the boundaries t and T (see also more detailed discussion in Notes 3.4 - 3.5). □

Note 3.2. In view of Note 3.1, the MTM estimators with $a > 0$ and $b > 0$ ($0 \leq F(t) \leq a < 1 - b \leq F(T) \leq 1$) are globally robust with the lower and upper breakdown points given by $LBP = a$ and $UBP = b$, respectively. The robustness of such estimators against small or large outliers comes from the fact that in the computation of estimates the influence of the order statistics with the index less than $n \times LBP$ or higher than $n \times (1 - UBP)$ is controlled in some way. For more details on LBP and UBP, see Brazauskas and Serfling (2000) and Serfling (2002). □

Note 3.3. For truncated data, the choice of $0 \leq F(t) \leq a < 1 - b \leq F(T) \leq 1$ yields the following expressions of the sample and population trimmed moments, respectively (see equations 3.19 and 3.20):

$$\widehat{\mu}_{*,j} = \frac{1}{n - m_n(j) - m_n^*(j)} \sum_{i=m_n(j)+1}^{n-m_n^*(j)} h_j(x_{i:n}^*), \qquad 1 \leq j \leq k,$$

and

$$\begin{aligned} \mu_{*,j} &= \frac{1}{1-a-b} \int_a^{1-b} h_j(F_*^{-1}(u))\, du \\ &= \frac{1}{1-a-b} \int_a^{1-b} h_j\left(F^{-1}\big(uF(T|\boldsymbol{\theta}) + (1-u)F(t|\boldsymbol{\theta})|\boldsymbol{\theta}\big)\right) du, \quad 1 \leq j \leq k. \end{aligned}$$

The remaining steps are the same as in (3.16) - (3.18). □

Note 3.4. For censored data, the choice of $0 \leq F(t) \leq a < 1-b \leq F(T) \leq 1$ makes the procedure (3.16) - (3.18) equivalent to the one for complete data. This is because

$$\int_a^{1-b} h_j\left(F_{**}^{-1}(u)\right)\, du = \int_a^{1-b} h_j\left(F^{-1}(u|\boldsymbol{\theta})\right)\, du.$$

The remaining combinations of $a,\ b,\ F(t)$, and $F(T)$ are listed in Appendix A. It is also important to note that for this procedure to work, one needs to first estimate $F(t)$ and $F(T)$, which can be done as follows

$$\widehat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i^{**} = t\}, \quad \text{and} \quad \widehat{F}(T) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i^{**} = T\}.$$

□

Note 3.5. For payment-per-payment data, there are three different cases to consider. After defining $s^* := \frac{F(u|\boldsymbol{\theta})-F(d|\boldsymbol{\theta})}{S(d|\boldsymbol{\theta})}$, we have

1. $0 \leq a_j \leq s^* \leq 1-b_j \leq 1$:

$$\begin{aligned}
\mu_{\otimes,j} &= \frac{1}{1-a_j-b_j}\left[\int_{a_j}^{s^*} h_j\left(c\left[F^{-1}\big(s+(1-s)F(d|\boldsymbol{\theta})|\boldsymbol{\theta}\big)-d\right]\right)\, ds \right.\\
&\qquad \left. + \int_{s^*}^{1-b_j} h_j(c(u-d))\, ds\right] \\
&= \frac{1}{1-a_j-b_j}\int_{a_j}^{s^*} h_j\left(c\left[F^{-1}\big(s+(1-s)F(d|\boldsymbol{\theta})|\boldsymbol{\theta}\big)-d\right]\right)\, ds \\
&\qquad + \frac{1-b_j-s^*}{1-a_j-b_j} h_j(c(u-d)). \qquad (3.27a)
\end{aligned}$$

2. $0 \leq a_j < 1-b_j \leq s^* \leq 1$:

$$\mu_{\otimes,j} = \frac{1}{1-a_j-b_j}\int_{a_j}^{1-b_j} h_j\left(c\left[F^{-1}\big(s+(1-s)F(d|\boldsymbol{\theta})|\boldsymbol{\theta}\big)-d\right]\right)\, ds. \qquad (3.27b)$$

3. $0 \leq s^* \leq a_j < 1-b_j \leq 1$:

$$\mu_{\otimes,j} = \frac{1}{1-a_j-b_j}\int_{a_j}^{1-b_j} h_j(c(u-d))\, ds = h_j(c(u-d)). \qquad (3.27c)$$

In order to implement the procedure, one needs to estimate s^* directly from available data, which can be done as follows

$$\widehat{s^*} = \frac{\widehat{\mathbb{P}}\,[0 < Y < c(u-d)]}{\widehat{\mathbb{P}}\,[Y > 0]} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{0 < y_i < c(u-d)\}.$$

More specifically, if both uncensored and censored sample observations participate in $\widehat{\mu}_{\otimes,j}$, then we end up with the first case. If the censored observations, i.e., $Y_i = c(u-d)$, $1 \le j \le k$ are not involved in computing $\widehat{\mu}_{\otimes,j}$, then we end up with the second case. And, finally if $\widehat{\mu}_{\otimes,j}$ is computed only with censored observations, then we are in the third case, but in this case (equation 3.27c) the population trimmed moment $\mu_{\otimes,j}$ is no longer a function of the parameter to be estimated. Thus, in that case MTM is not recommended approach of estimation. We may also rule out the third case by choosing $a_j = 0$, i.e., no trimming on the left. □

Note 3.6. The payment-per-loss scenario is a special case of censored data. MTM estimators follow from Note 3.4 with obvious adjustments of notation. □

3.2.2 Asymptotic Properties

MTM estimators belong to the class of L-statistics whose general asymptotic properties have been established by Chernoff *et al.* (1967). Specifically, asymptotically equation (3.16) is equivalent to

$$\widehat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} J_j\left(\frac{i}{n+1}\right) h_j(X_{i:n}), \qquad 1 \le j \le k,$$

where

$$J_j(s) = \begin{cases} (1-a_j-b_j)^{-1}, & a_j < s < 1-b_j; \\ 0, & \text{otherwise,} \end{cases}$$

is the weight generating function. The population trimmed moments take the form

$$\mu_j = \int_0^1 J_j(u) H_j(u)\, du, \quad 1 \le j \le k,$$

where $H_j = h_j \circ F^{-1}$. Also, consider

$$\alpha_j(u) = \frac{1}{1-u}\int_u^1 J_j(v) H_j'(v)(1-v)\, dv, \quad 1 \le j \le k.$$

Then the k-variate vector $\sqrt{n}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ converges in distribution to the k-variate normal random vector with mean $\mathbf{0}$ and the variance-covariance matrix $\boldsymbol{\Sigma} := \left[\sigma^2_{ij}\right]_{i,j=1}^{k}$ with the entries (see Chernoff *et al.*, 1967, Remark 9)

$$\begin{aligned}\sigma^2_{ij} &= \int_0^1 \alpha_i(u)\alpha_j(u)\,du \\ &= \int_0^1 \left(\frac{1}{1-u}\right)^2 \left[\int_u^1 J_i(v)H_i'(v)(1-v)\,dv \int_u^1 J_j(w)H_j'(w)(1-w)\,dw\right] du.\end{aligned} \tag{3.28}$$

Further, Brazauskas *et al.* (2007) independently established an equivalent simplified expression for σ^2_{ij}, which can be written as

$$\sigma^2_{ij} = \frac{\int_{a_i}^{1-b_i}\int_{a_j}^{1-b_j} \left(\min\{u,v\} - uv\right)\, dh_j(F^{-1}(v|\boldsymbol{\theta}))\; dh_i(F^{-1}(u|\boldsymbol{\theta}))}{(1-a_i-b_i)(1-a_j-b_j)}. \tag{3.29}$$

Thus,

$$(\widehat{\mu}_1, \widehat{\mu}_2, ..., \widehat{\mu}_k) \sim \mathcal{AN}((\mu_1, \mu_2, ..., \mu_k), n^{-1}\boldsymbol{\Sigma}). \tag{3.30}$$

The relation (3.30) along with the delta method (see Theorem 1.1) can be used to derive the asymptotic distribution of the MTM estimators. Consider $g = (g_1, ..., g_k)$ such that $\widehat{\boldsymbol{\theta}}_n = g(\widehat{\boldsymbol{\mu}})$. Then

$$\widehat{\boldsymbol{\theta}}_n \sim \mathcal{AN}\left(g(\boldsymbol{\mu}), n^{-1}\left(\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}'\right)\right), \tag{3.31}$$

with the usual meaning of $\boldsymbol{D}$ as in Theorem 1.1. Relation (3.31) can be used to perform statistical inference based on the MTM estimators.

Example 3.16. *Complete Data.*

For this data scenario, with $a = a_i = a_j$ and $b = b_i = b_j$, the entries σ^2_{ij} of the variance-covariance matrix $\boldsymbol{\Sigma}$ are given by equation (3.29). □

Example 3.17. *Truncated Data.*

For this data scenario, with $0 \le F(t) \le a < 1-b \le F(T) \le 1$, the vector

$$\widehat{\boldsymbol{\mu}}_* \sim \mathcal{AN}(\boldsymbol{\mu}_*, n^{-1}\boldsymbol{\Sigma}_*),$$

where the entries σ_{ij}^2 of $\mathbf{\Sigma}_*$ are given by

$$\sigma_{ij}^2 = (1-a-b)^{-2} \int_a^{1-b} \int_a^{1-b} \left(\min\{u,v\} - uv\right) \, dh_j(F_*^{-1}(v)) \, dh_i(F_*^{-1}(u)), \quad (3.32)$$

with $F_*^{-1}(u) = F^{-1}\left(uF(T|\boldsymbol{\theta}) + (1-u)F(t|\boldsymbol{\theta})|\boldsymbol{\theta}\right)$. $\square$

Example 3.18. *Censored Data.*

For this data scenario, with $0 \le F(t) \le a < 1-b \le F(t) \le 1$, the vector

$$\widehat{\boldsymbol{\mu}}_{**} \sim \mathcal{AN}(\boldsymbol{\mu}_{**}, n^{-1}\mathbf{\Sigma}_{**}),$$

where the entries σ_{ij}^2 of $\mathbf{\Sigma}_{**}$ are given by

$$\sigma_{ij}^2 = (1-a-b)^{-2} \int_a^{1-b} \int_a^{1-b} \left(\min\{u,v\} - uv\right) \, dh_j(F_{**}^{-1}(v)) \, dh_i(F_{**}^{-1}(u)),$$

with the qf F_{**}^{-1} given by equation (2.9). $\square$

Example 3.19. *Payment-per-payment Data.*

For this data scenario, with $0 \le a < 1-b \le \frac{F(u)-F(d)}{S(d)} \le 1$, the vector

$$\widehat{\boldsymbol{\mu}}_{\otimes} \sim \mathcal{AN}(\boldsymbol{\mu}_{\otimes}, n^{-1}\mathbf{\Sigma}_{\otimes}),$$

where the entries σ_{ij}^2 of $\mathbf{\Sigma}_{\otimes}$ are given by

$$\sigma_{ij}^2 = (1-a-b)^{-2} \int_a^{1-b} \int_a^{1-b} \left(\min\{u,v\} - uv\right) \, dh_j(G_Y^{-1}(v)) \, dh_i(G_Y^{-1}(u)), \quad (3.33)$$

with the qf G_Y^{-1} defined by equation (2.13). $\square$

Example 3.20. *Payment-per-loss Data.*

For this data scenario, with $0 \le F(t) \le a < 1-b \le F(T) \le 1$, the vector

$$\widehat{\boldsymbol{\mu}}_{\otimes\otimes} \sim \mathcal{AN}(\boldsymbol{\mu}_{\otimes\otimes}, n^{-1}\mathbf{\Sigma}_{\otimes\otimes}),$$

where the entries σ_{ij}^2 of $\mathbf{\Sigma}_{\otimes\otimes}$ are given by

$$\sigma_{ij}^2 = (1-a-b)^{-2} \int_a^{1-b} \int_a^{1-b} \left(\min\{u,v\} - uv\right) \, dh_j(G_Z^{-1}(v)) \, dh_i(G_Z^{-1}(u)),$$

with the qf G_Y^{-1} defined by equation (2.17). $\square$

Chapter 4

Analytic Examples

In this chapter, we derive MLE and MTM estimators for the parameters of exponential and normal distributions, under the data scenarios of Chapter 2. Note that for insurance losses the equivalent (after the logarithmic transformation) models are Pareto and lognormal. Thus, the estimators derived in Sections 4.1 and 4.2 can easily be adjusted for Pareto and lognormal models. Their asymptotic properties will remain valid as well.

4.1 Exponential and Pareto Models

Let $X \sim Exp(\theta)$ with the ground-up loss distribution function $F(x|\theta) = 1 - e^{-\frac{x}{\theta}}$, density function $f(x|\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, \quad x > 0$, and the quantile function $F^{-1}(u|\theta) = -\theta \log(1-u)$.

4.1.1 Maximum Likelihood Estimation

In the following examples, we specify the maximum likelihood estimators for the data scenarios and models of Chapter 2 when $X \sim Exp(\theta)$.

Example 4.1. *Complete Data.*

For an *i.i.d.* sample $x_1, \ldots, x_n$ with pdf $f(x|\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$, the log-likelihood function (equation 3.4) becomes

$$l(\theta|x_1, \ldots, x_n) = \sum_{i=1}^{n} \log(f(x_i|\theta)) = -n \log\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i. \tag{4.1}$$

Then, setting $l'(\theta|x_1,\ldots,x_n)=0$ and solving for θ yields $\widehat{\theta}_n=\widehat{\mu}$, where $\widehat{\mu}$ is the sample mean. Also, it readily follows that $\widehat{\theta}_n\sim\mathcal{AN}\left(\theta,n^{-1}\theta^2\right)$. □

Example 4.2. *Truncated Data.*

For an *i.i.d.* sample $x_1^*,\ldots,x_n^*$, of truncated data in the interval $(t,T]$, defined by (2.2) with pdf (2.4), the log-likelihood function (equation 3.6) becomes

$$
\begin{aligned}
l_*(\theta|x_1^*,\ldots,x_n^*) &= \sum_{i=1}^{n}\log\left(f(x_i^*|\theta)\right)-n\log\left(F(T|\theta)-F(t|\theta)\right)\\
&= -n\log\theta-n\log\left(e^{-\frac{t}{\theta}}-e^{-\frac{T}{\theta}}\right)-\frac{1}{\theta}\sum_{i=1}^{n}x_i^* \qquad (4.2)
\end{aligned}
$$

Setting $l_*'(\theta|x_1^*,\ldots,x_n^*)=0$ yields the equation

$$
\widehat{\mu}=\theta+\frac{te^{-\frac{t}{\theta}}-Te^{-\frac{T}{\theta}}}{e^{-\frac{t}{\theta}}-e^{-\frac{T}{\theta}}}, \qquad (4.3)
$$

where $\widehat{\mu}=n^{-1}\sum_{i=1}^{n}x_i^*$ is the sample mean. Solving (4.3) for θ leads to the MLE $\widehat{\theta}_n$.

Proposition 4.1. *If $\widehat{\mu}\geq\frac{t+T}{2}$, then the MLE estimate $\widehat{\theta}_n$ of θ does not exist.*

Proof. It follows directly from the proof of Theorem 6.5. □

In this case the Fisher information matrix $\boldsymbol{I}_*(\theta)$ given by (3.13) is a scalar and can be computed as

$$
\boldsymbol{I}_*(\theta)=\frac{1}{\theta^2}-\left(\frac{T-t}{\theta^2}\right)^2\frac{e^{-\frac{t+T}{\theta}}}{\left(e^{-\frac{t}{\theta}}-e^{-\frac{T}{\theta}}\right)^2}. \qquad (4.4)
$$

Therefore,

$$
\widehat{\theta}_n\sim\mathcal{AN}\left(\theta,\frac{1}{n}\left[\boldsymbol{I}_*(\theta)\right]^{-1}\right),
$$

which also implies that $\widehat{\theta}_n$ is consistent and efficient. □

Example 4.3. *Censored Data.*

If an *i.i.d.* sample $x_1^{**},\ldots,x_n^{**}$ of censored data, defined by (2.6), is observed, then the log-likelihood function (equation 3.8) becomes

$$
\begin{aligned}
l_{**}(\theta|x_1^{**},\ldots,x_n^{**}) &= \log\left(F(t|\theta)\right)\sum_{i=1}^{n}\mathbb{1}\{x_i^{**}=t\} + \sum_{t<x_i^{**}<T}\log\left(f(x_i^{**}|\theta)\right) \\
&\quad + \log\left(S(T|\theta)\right)\sum_{i=1}^{n}\mathbb{1}\{x_i^{**}=T\} \\
&= \log\left(1-e^{-\frac{t}{\theta}}\right)\sum_{i=1}^{n}\mathbb{1}\{x_i^{**}=t\} - \sum_{t<x_i^{**}<T}\left(\log(\theta)+\frac{x_i^{**}}{\theta}\right) \\
&\quad - \frac{T}{\theta}\sum_{i=1}^{n}\mathbb{1}\{x_i^{**}=T\}. \qquad (4.5)
\end{aligned}
$$

Setting $l'_{**}(\theta|x_1^{**},\ldots,x_n^{**})=0$ gives

$$
-\frac{te^{-\frac{t}{\theta}}}{\theta^2(1-e^{-\frac{t}{\theta}})}\sum_{i=1}^{n}\mathbb{1}\{x_i^{**}=t\} - \sum_{t<x_i^{**}<T}\left(\frac{1}{\theta}-\frac{x_i^{**}}{\theta^2}\right) + \frac{T}{\theta^2}\sum_{i=1}^{n}\mathbb{1}\{x_i^{**}=T\} = 0, \quad (4.6)
$$

which should be solved numerically to obtain $\widehat{\theta}_n$.

The information matrix $\boldsymbol{I}_{**}(\theta)$ given by equation (3.14) is a scalar. With straightforward calculation, we get

$$
\boldsymbol{I}_{**}(\theta) = \frac{t^2 S(t|\theta)}{\theta^4 F(t|\theta)} + \frac{F(T|\theta)-F(t|\theta)}{\theta^2}, \qquad (4.7)
$$

and hence,

$$
\widehat{\theta}_n \sim \mathcal{AN}\left(\theta, \frac{1}{n}\left[\boldsymbol{I}_{**}(\theta)\right]^{-1}\right),
$$

which also implies that $\widehat{\theta}_n$ is consistent and efficient. □

Example 4.4. *Payment-per-payment Data.*

For an observed *i.i.d.* sample $y_1,\ldots,y_n$ defined by (2.10), the log-likelihood function (3.10) becomes

$$
\begin{aligned}
l_{\otimes}(\theta|y_1,\ldots,y_n) &= \log\left(S(u|\boldsymbol{\theta})\right)\sum_{i=1}^{n}\mathbb{1}\{y_i=c(u-d)\} - n\log\left(S(d|\boldsymbol{\theta})\right) \\
&\quad - \sum_{i=1}^{n}\mathbb{1}\{0<y_i<c(u-d)\}\left(\log(c)-\log\left(f\left(\frac{y_i}{c}+d|\boldsymbol{\theta}\right)\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \log\left(e^{-\frac{u}{\theta}}\right)\sum_{i=1}^{n} \mathbb{1}\{y_i = c(u-d)\} - n\log\left(e^{-\frac{d}{\theta}}\right) \\
&\quad - \sum_{i=1}^{n} \mathbb{1}\{0 < y_i < c(u-d)\}\left(\log(c) - \log\left(\frac{1}{\theta}e^{-\frac{\frac{y_i}{c}+d}{\theta}}\right)\right) \\
&= -\frac{u}{\theta}\sum_{i=1}^{n} \mathbb{1}\{y_i = c(u-d)\} + \frac{nd}{\theta} \\
&\quad - \sum_{i=1}^{n} \mathbb{1}\{0 < y_i < c(u-d)\}\left(\log(c) + \log(\theta) + \frac{\frac{y_i}{c}+d}{\theta}\right), \qquad (4.8)
\end{aligned}
$$

Setting $l'_{\otimes}(\theta|y_1,\ldots,y_n) = 0$ yields the explicit formula

$$
\widehat{\theta}_n = (u-d)\frac{\sum_{i=1}^{n} \mathbb{1}\{y_i = c(u-d)\}}{\sum_{i=1}^{n} \mathbb{1}\{0 < y_i < c(u-d)\}} + \frac{1}{c}\frac{\sum_{i=1}^{n} y_i \mathbb{1}\{0 < y_i < c(u-d)\}}{\sum_{i=1}^{n} \mathbb{1}\{0 < y_i < c(u-d)\}}. \qquad (4.9)
$$

The information matrix $\boldsymbol{I}_{\otimes}(\theta)$, a scalar, given by the expression (3.15) can be computed as

$$
\boldsymbol{I}_{\otimes}(\theta) = \frac{F(u|\theta) - F(d|\theta)}{\theta^2 S(d|\theta)}, \qquad (4.10)
$$

and hence,

$$
\widehat{\theta}_n \sim \mathcal{AN}\left(\theta, \frac{1}{n}\left[\boldsymbol{I}_{\otimes}(\theta)\right]^{-1}\right),
$$

which also implies that $\widehat{\theta}_n$ is consistent and efficient. □

Example 4.5. *Payment-per-loss Data.*

Consider an observed *i.i.d.* sample $z_1, \ldots, z_n$ defined by (2.14). Then the linearly transformed *i.i.d.* sample $\frac{z_1}{c} + d, \ldots, \frac{z_n}{c} + d$ is exactly the sample $x_1^{**}, \ldots, x_n^{**}$ treated in Example 4.3 with $d \equiv t$ and $u \equiv T$. □

4.1.2 Method of Trimmed Moments

In the following examples, we specify MTM estimators for the data scenarios and models of Chapter 2 when $X \sim Exp(\theta)$. Since $\boldsymbol{\theta}$ is a scalar, only one function h is needed. The most convenient choice is $h(x) = x$.

Example 4.6. *Complete Data.*

This scenario has been fully investigated by Brazauskas *et al.* (2009). The sample trimmed moment (3.16) takes the form

$$\widehat{\mu} = \frac{1}{n - m_n - m_n^*} \sum_{i=m_n+1}^{n-m_n^*} x_{i:n},$$

with $m_n/n \to a$ and $m_n^*/n \to b$. The corresponding population trimmed moment (3.17) is given by

$$\begin{aligned}
\mu &= \frac{1}{1-a-b} \int_a^{1-b} F^{-1}(u|\theta)\, du \\
&= -\frac{\theta}{1-a-b} \int_a^{1-b} \log(1-u)\, du \\
&=: -\frac{\theta I(a, 1-b)}{1-a-b}.
\end{aligned}$$

The MTM estimator of θ is then

$$\widehat{\widehat{\theta}}_n = -\frac{\widehat{\mu}(1-a-b)}{I(a,1-b)}. \tag{4.11}$$

Its asymptotic distribution is

$$\widehat{\widehat{\theta}}_n \sim \mathcal{AN}\left(\theta, \frac{\theta^2}{n}\Delta\right), \quad \text{with} \quad \Delta = \frac{J(a,1-b)}{[I(a,1-b)]^2},$$

where

$$J(a,1-b) := \int_a^{1-b} \int_a^{1-b} \frac{\min\{u,v\} - uv}{(1-u)(1-v)}\, dv\, du.$$

This also implies that $\widehat{\widehat{\theta}}_n$ is consistent. □

Example 4.7. *Truncated Data.*

For an *i.i.d.* sample $x_1^*, \ldots, x_n^*$ defined by (2.2) the sample trimmed moment (3.16) is given by

$$\widehat{\mu}_* = \frac{1}{n - m_n - m_n^*} \sum_{i=m_n+1}^{n-m_n^*} x_{i:n}^*,$$

with $m_n/n \to a$ and $m_n^*/n \to b$. The corresponding population trimmed moment (3.20) with the qf defined by (2.5) is given by

$$\begin{aligned}\mu_* &= \frac{1}{1-a-b}\int_a^{1-b} F_*^{-1}(u)\,du \\ &= \frac{1}{1-a-b}\int_a^{1-b} F^{-1}\big(uF(T|\theta)+(1-u)F(t|\theta)|\theta\big)\,du, \\ &= -\frac{\theta}{1-a-b}\int_a^{1-b}\log\left(e^{-\frac{t}{\theta}}-up(\theta)\right)du, \end{aligned} \tag{4.12}$$

where $p(\theta) := e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}$. Clearly, the equation $\mu_* = \widehat{\mu}_*$ needs to be solved numerically for θ. Let us denote the solution by $\widehat{\widehat{\theta}}_n$, if it exists. Its asymptotic distribution is

$$\widehat{\widehat{\theta}}_n \sim \mathcal{AN}\left(\theta, n^{-1}\boldsymbol{D}_*^2\boldsymbol{\Sigma}_*\right), \tag{4.13}$$

where the single entry of $\boldsymbol{\Sigma}_*$ is given by (3.32):

$$\begin{aligned}\sigma_{11}^2 &= (1-a-b)^{-2}\int_a^{1-b}\int_a^{1-b}\left(\min\{u,v\}-uv\right)\,dF_*^{-1}(v)\,dF_*^{-1}(u) \\ &= \frac{(\theta p(\theta))^2}{(1-a-b)^2}\int_a^{1-b}\int_a^{1-b}\frac{\min\{u,v\}-uv}{\left(e^{-\frac{t}{\theta}}-vp(\theta)\right)\left(e^{-\frac{t}{\theta}}-up(\theta)\right)}\,dv\,du,\end{aligned}$$

and the Jacobian $\boldsymbol{D}_*$ entry is found with implicit differentiation of θ with respect to μ_* from equation (4.12) as follows:

$$\boldsymbol{D}_* = \frac{d\theta}{d\mu_*} = -(1-a-b)\left[\int_a^{1-b}\left\{\log\left(e^{-\frac{t}{\theta}}-up(\theta)\right)+\frac{te^{-\frac{t}{\theta}}-up_{tT}(\theta)}{\theta\left(e^{-\frac{t}{\theta}}-up(\theta)\right)}\right\}du\right]^{-1},$$

where $p_{tT}(\theta) := te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}}$. □

Example 4.8. *Censored Data.*

For an *i.i.d.* sample $x_1^{**},\ldots,x_n^{**}$ defined by (2.6) the sample trimmed moment (3.16) is given by

$$\widehat{\mu}_* = \frac{1}{n-m_n-m_n^*}\sum_{i=m_n+1}^{n-m_n^*} x_{i:n}^{**},$$

with $m_n/n \to a$ and $m_n^*/n \to b$. The corresponding population trimmed moment (3.20) with the qf defined by (2.9) is given by

$$\begin{aligned}
\mu_{**} &= \frac{1}{1-a-b}\int_a^{1-b} F_{**}^{-1}(u)\,du \\
&= \frac{1}{1-a-b}\int_a^{1-b} F^{-1}(u|\theta)\,du, \\
&= -\frac{\theta I(a,1-b)}{1-a-b},
\end{aligned} \tag{4.14}$$

by assuming the most general case that $0 \le F(t) \le a < 1-b \le F(T) \le 1$ as in Example 3.13. Thus, with the assumption $0 \le F(t) \le a < 1-b \le F(T) \le 1$, this case translates to the complete case as in Example 4.6. $\square$

Example 4.9. *Payment-per-payment Data.*

For an *i.i.d.* sample $y_1, \ldots, y_n$ defined by (2.10) the sample trimmed moment (3.16) is given by

$$\widehat{\mu}_{\otimes} = \frac{1}{n-m_n-m_n^*}\sum_{i=m_n+1}^{n-m_n^*} y_{i:n},$$

with $m_n/n \to a$ and $m_n^*/n \to b$. Assume that $m_n^* \ge \sum_{i=1}^n \mathbb{1}\{y_i = c(u-d)\}$, then we end up with the most general second case of Example 3.14. That, is $0 \le a < 1-b \le s^* \le 1$ with $s^* = \frac{F(u)-F(d)}{S(d)}$. Then, the corresponding population trimmed moment (3.27b) with the qf defined by (2.13) is given by

$$\begin{aligned}
\mu_{\otimes} &= \frac{1}{1-a-b}\int_a^{1-b} G_Y^{-1}(u)\,du \\
&= \frac{1}{1-a-b}\int_a^{1-b} c\left[F^{-1}\big(s+(1-s)F(d|\theta)|\theta\big) - d\right]\,ds \\
&= \frac{1}{1-a-b}\int_a^{1-b} c\left[-\theta\log\left(1-s-(1-s)F(d|\theta)\right) - d\right] ds \\
&= -\frac{c}{1-a-b}\left[\theta I(a,1-b) + \theta\log\left(1-F(d|\theta)\right)(1-a-b) + d(1-a-b)\right] \\
&= -\frac{c}{1-a-b}\left[\theta I(a,1-b) + \theta\frac{d}{\theta}(1-a-b) - d(1-a-b)\right]
\end{aligned}$$

$$= -\frac{c\theta I(a,1-b)}{1-a-b}.$$

Setting $\mu_\otimes = \widehat{\mu}$ yields

$$\widehat{\widehat{\theta}}_n = -\frac{\widehat{\mu}(1-a-b)}{cI(a,1-b)} =: g(\widehat{\mu}). \tag{4.15}$$

The entries of the matrix $\mathbf{\Sigma}_\otimes$, which is one-dimensional, follow from equation (3.33):

$$\begin{aligned}
\sigma_{11}^2 &= \frac{\int_a^{1-b}\int_a^{1-b} (\min\{w,v\} - wv)\; dh(G_Y^{-1}(v))\; dh(G_Y^{-1}(w))}{(1-a-b)(1-a-b)} \\
&= \frac{c^2}{(1-a-b)^2}\int_a^{1-b}\int_a^{1-b}\left[\frac{(\min\{w,v\}-wv)\,(1-F(d|\theta))}{f\left[F^{-1}(v(1-F(d|\theta))+F(d|\theta)|\theta)|\theta\right]}\right. \\
&\qquad\qquad \left.\times\frac{1-F(d|\theta)}{f\left[F^{-1}(w(1-F(d|\theta))+F(d|\theta)|\theta)|\theta\right]}\right] dv\,dw \\
&= c^2\left(\frac{1-F(d|\theta)}{1-a-b}\right)^2\int_a^{1-b}\int_a^{1-b}\frac{(\min\{w,v\}-wv)\; dv\,dw}{f(d-\theta\log(1-v)|\theta)f(d-\theta\log(1-w)|\theta)} \\
&= \frac{c^2\theta^2}{(1-a-b)^2}J(a,1-b).
\end{aligned}$$

Also, $\left.\frac{\partial g}{\partial\widehat{\mu}}\right|_\mu = -\frac{1-a-b}{cI(a,1-b)}$. Then, $\boldsymbol{D}_\otimes\mathbf{\Sigma}_\otimes\boldsymbol{D}_\otimes' = \theta^2\frac{J(a,1-b)}{[I(a,1-b)]^2}$. Therefore

$$\widehat{\widehat{\theta}}_n \sim \mathcal{AN}\left(\theta, \frac{\theta^2}{n}\Delta\right) \text{ with } \Delta = \frac{J(a,1-b)}{[I(a,1-b)]^2},$$

which implies that $\widehat{\widehat{\theta}}_n$ is consistent. □

Example 4.10. *Payment-per-loss Data.*

Consider an observed *i.i.d.* sample $z_1,\ldots,z_n$ defined by (2.14). Then, as in Example 4.5, the linearly transformed *i.i.d.* sample $\frac{z_1}{c}+d,\ldots,\frac{z_n}{c}+d$ is exactly the sample $x_1^{**},\ldots,x_n^{**}$ treated in Example 4.8 with $d\equiv t$ and $u\equiv T$. □

4.2 Normal and Lognormal Models

Let $X \sim \text{Normal}(\theta,\sigma^2)$ with ground-up loss distribution function

$$F(x) = \Phi\left(\frac{x-\theta}{\sigma}\right), \qquad -\infty < x < \infty, \tag{4.16}$$

where $-\infty < \theta < \infty$, the location parameter, $0 < \sigma < \infty$, scale parameter, and Φ is the cdf of the standard normal distribution. The corresponding ground-up density function is

$$f(x) = \frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right), \qquad -\infty < x < \infty, \tag{4.17}$$

where ϕ is the pdf of the standard normal distribution given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \qquad -\infty < x < \infty. \tag{4.18}$$

The quantile function F^{-1} is given by

$$F^{-1}(u|\boldsymbol{\theta}) = \theta + \sigma\Phi^{-1}(u). \tag{4.19}$$

The parameter vector to be estimated is $\boldsymbol{\theta} = (\theta, \sigma)$.

The following additional notation will be used in this section.

<u>NOTATION:</u> Let t and T be the left and right truncation thresholds, respectively, as defined in Section 2.2. Define:

$$R := T - t, \qquad t_z := \frac{t-\theta}{\sigma}, \qquad T_z := \frac{T-\theta}{\sigma}.$$

Then, obviously

$$\theta = t - \sigma t_z, \qquad T_z = t_z + \frac{R}{\sigma}. \tag{4.20}$$

Additionally,

$$\begin{cases} K_{0,t}(t_z) & := & 1 - \Phi(t_z), \\ K_{0,T}(T_z) & := & 1 - \Phi(T_z). \end{cases} \tag{4.21}$$

Since $K_{0,\cdot}$ is defined by equation (4.21), we recursively define

$$K_{n,\cdot}(z) = \int_z^\infty K_{n-1,\cdot}(s)\,ds, \tag{4.22}$$

then consequently

$$\frac{dK_{n,\cdot}}{dz} = -K_{n,\cdot}. \tag{4.23}$$

Lemma 4.1. *For all* $n \geq -1$ *(Cohen, 1950),*

$$(n+1)K_{n+1,.}(z) + zK_{n,.}(z) - K_{n-1,.}(z) = 0. \tag{4.24}$$

Note 4.1. The variable t_z is considered to be the independent parameter of location in Examples 4.12, 4.13, and 4.14. The mean, θ, is a linear function of t_z given by (4.20). □

4.2.1 Maximum Likelihood Estimation

We now derive the maximum likelihood estimators for all the data scenarios of Chapter 2 when $X \sim \text{Normal}(\theta, \sigma^2)$.

Example 4.11. *Complete Data.*

It can be found in any standard statistics textbook (see, e.g., Wasserman, 2004) that for a completely observed *i.i.d.* normal sample $x_1, \ldots, x_n$ of size n, the MLE estimators of $\boldsymbol{\theta} = (\theta, \sigma)$ are:

$$\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_n, \widehat{\sigma}_n) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i, \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \widehat{\theta}_n\right)^2} \right),$$

and

$$\widehat{\boldsymbol{\theta}}_n \sim \mathcal{AN}\left((\theta,\sigma), \frac{\sigma^2}{n}\boldsymbol{S}_0\right) \quad \text{with} \quad \boldsymbol{S}_0 = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \tag{4.25}$$

□

Example 4.12. *Truncated Data.*

Let $x_1^*, \ldots, x_n^*$ be an *i.i.d.* sample of truncated normal data in the interval $(t, T]$ with cdf (2.3) and pdf (2.4) . Let us linearly transform the observed data via $y := x^* - t$. Clearly, $0 < y \leq R$ and the density function of the random variable Y is given by

$$\begin{aligned} f_*(y) &= \frac{f(y+t|\boldsymbol{\theta})}{F(T|\boldsymbol{\theta}) - F(t|\boldsymbol{\theta})} \\ &= \frac{e^{-\frac{1}{2}\left(t_z + \frac{y}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}(K_{0,t} - K_{0,T})}, \end{aligned}$$

The log-likelihood function given by equation (3.6) for an $i.i.d.$ sample of size n of the random variable $Y := X^* - t$ takes the form

$$l_*(t_z, \sigma) = -n \log\left((K_{0,t} - K_{0,T})\sigma\sqrt{2\pi}\right) - \frac{1}{2}\sum_{i=1}^{n}\left(t_z + \frac{y_i}{\sigma}\right)^2. \tag{4.26}$$

Using equations (4.18), (4.20), (4.21), and (4.22), we summarize the following derivatives:

$$\begin{cases} \frac{\partial t_z}{\partial t_z} = 1, & \frac{\partial t_z}{\partial \sigma} = 0, \\ \frac{\partial T_z}{\partial t_z} = 1, & \frac{\partial T_z}{\partial \sigma} = -\frac{R}{\sigma^2} \\ \frac{\partial \phi(t_z)}{\partial t_z} = -t_z\phi(t_z), & \frac{\partial \phi(t_z)}{\partial \sigma} = 0, \\ \frac{\partial \phi(T_z)}{\partial t_z} = -T_z\phi(T_z), & \frac{\partial \phi(T_z)}{\partial \sigma} = \frac{RT_z\phi(T_z)}{\sigma^2}, \\ \frac{\partial K_{0,t}}{\partial t_z} = -\phi(t_z), & \frac{\partial K_{0,t}}{\partial \sigma} = 0, \\ \frac{\partial K_{0,T}}{\partial t_z} = -\phi(T_z), & \frac{\partial K_{0,T}}{\partial \sigma} = \phi(T_z)\frac{R}{\sigma^2}. \end{cases} \tag{4.27}$$

Therefore, setting

$$\begin{cases} \frac{\partial l_*}{\partial t_z} &= 0, \\ \frac{\partial l_*}{\partial \sigma} &= 0, \end{cases}$$

yields the system of MLE equations

$$\begin{cases} \frac{\partial l_*}{\partial t_z} = \dfrac{n(\phi(t_z) - \phi(T_z))}{K_{0,t} - K_{0,T}} - \sum_{i=1}^{n}\left(t_z + \dfrac{y_i}{\sigma}\right) = 0, \\ \frac{\partial l_*}{\partial \sigma} = \left(\dfrac{n\phi(T_z)}{K_{0,t} - K_{0,T}}\right)\dfrac{R}{\sigma^2} - \dfrac{n}{\sigma} + \dfrac{1}{\sigma^2}\sum_{i=1}^{n}\left[y_i\left(t_z + \dfrac{y_i}{\sigma}\right)\right] = 0. \end{cases} \tag{4.28}$$

Consider,

$$Z_1 := \frac{\phi(t_z)}{K_{0,t} - K_{0,T}}, \qquad Z_2 := \frac{\phi(T_z)}{K_{0,t} - K_{0,T}}, \tag{4.29}$$

then the MLE system of equations (4.28) becomes

$$\begin{cases} \sigma\left(Z_1 - Z_2 - t_z\right) - \widehat{\mu}_1 = 0, \\ \sigma^2\left(1 - t_z(Z_1 - Z_2 - t_z) - \frac{Z_2 R}{\sigma}\right) - \widehat{\mu}_2 = 0, \end{cases} \tag{4.30}$$

where $\widehat{\mu}_1$ and $\widehat{\mu}_2$ are the first and second sample moments given by $\widehat{\mu}_j = n^{-1}\sum_{i=1}^{n} y_i^j$,

$j = 1, 2$.

The system of equations (4.30) can be solved for $\widehat{\sigma}_n$ and $\widehat{t}_{z,n}$ by using a modified Newton-Raphson method (see, e.g., Cohen, 1950) with the initializing values

$$\sigma_{\text{start}} = \sqrt{\widehat{\mu}_2 - \widehat{\mu}_1^2}, \quad \text{and} \quad t_{z,\text{start}} = -\frac{\widehat{\mu}_1}{\sqrt{\widehat{\mu}_2 - \widehat{\mu}_1^2}}.$$

To establish the asymptotic distribution of $(\widehat{t}_{z,n}, \widehat{\sigma}_n)$, define (Cohen, 1950)

$$\begin{cases} f_1(t_z, T_z) & := & -\left[1 + t_z Z_1 - T_z Z_2 - (Z_1 - Z_2)^2\right], \\ f_2(t_z, T_z) & := & \frac{RZ_2}{\sigma}\left[(Z_1 - Z_2) - T_z\right] + \left[Z_1 - Z_2 - t_z\right], \\ f_3(t_z, T_z) & := & \left(\frac{R}{\sigma}\right)^2 Z_2(Z_2 + T_z) - \left[2 - t_z(Z_1 - Z_2 - t_z) - \frac{Z_2 R}{\sigma}\right]. \end{cases} \tag{4.31}$$

Let

$$\boldsymbol{I}_*(t_z, \sigma) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

be the Fisher information matrix given by equation (3.13), then

$$a_{11} = -f_1(t_z, T_z),$$

$$a_{12} = a_{21} = -\sigma^{-1} f_2(t_z, T_z),$$

$$a_{22} = -\sigma^{-2} f_3(t_z, T_z).$$

Therefore,

$$(\widehat{t}_{z,n}, \widehat{\sigma}_n) \sim \mathcal{AN}\left((t_z, \sigma), \frac{1}{n}\begin{bmatrix} -\frac{f_3}{f_1 f_3 - f_2^2} & \frac{\sigma f_2}{f_1 f_3 - f_2^2} \\ \frac{\sigma f_2}{f_1 f_3 - f_2^2} & -\frac{\sigma^2 f_1}{f_1 f_3 - f_2^2} \end{bmatrix}\right). \qquad \square$$

Example 4.13. *Censored Data*

Let $x_1^{**}, \ldots, x_n^{**}$ be an *i.i.d.* sample with cdf (2.7) and pdf (2.8). Before going in further details, let us borrow all the symbols from Example 4.12. In addition, define

$$n_0 := \sum_{i=1}^{n} \mathbb{1}\{y_i = 0\}, \qquad n_1 := \sum_{i=1}^{n} \mathbb{1}\{0 < y_i < R\}, \qquad n_2 := \sum_{i=1}^{n} \mathbb{1}\{y_i = R\}.$$

Note that $n = n_0 + n_1 + n_2$. In this case the log-likelihood function given by equation (3.8) becomes

$$l_{**}(t_z, \sigma) = \text{constant} + n_0 \log\left(1 - K_{0,t}\right) - n_1 \log\left(\sigma\right)$$

$$- \frac{1}{2} \sum_{0<y_i<R} \left(t_z + \frac{y_i}{\sigma} \right)^2 + n_2 \log \left(K_{0,T} \right). \tag{4.32}$$

For the censored case the system of equations (4.28) takes the form

$$\begin{cases} \frac{\partial l_{**}}{\partial t_z} &= n_0 \dfrac{\phi(t_z)}{1 - K_{0,t}} - n_2 \dfrac{\phi(T_z)}{K_{0,T}} - \displaystyle\sum_{0<y_i<R} \left(t_z + \frac{y_i}{\sigma} \right) &= 0, \\ \frac{\partial l_{**}}{\partial \sigma} &= -n_1 \dfrac{1}{\sigma} + \dfrac{1}{\sigma^2} \displaystyle\sum_{0<y_i<R} \left[y_i \left(t_z + \frac{y_i}{\sigma} \right) \right] + n_2 \dfrac{\phi(T_z)}{K_{0,T}} &= 0. \end{cases} \tag{4.33}$$

Similarly with (4.29), define

$$\begin{cases} Y_1 &:= \frac{n_0}{n_1} \frac{\phi(t_z)}{1 - K_{0,t}}, \\ Y_2 &:= \frac{n_2}{n_1} \frac{\phi(T_z)}{K_{0,T}}, \end{cases} \tag{4.34}$$

then the MLE system of equations (4.33) becomes

$$\begin{cases} \sigma \left(Y_1 - Y_2 - t_z \right) - \widehat{\mu}_1 &= 0, \\ \sigma^2 \left(1 - t_z (Y_1 - Y_2 - t_z) - \frac{Y_2 R}{\sigma} \right) - \widehat{\mu}_2 &= 0, \end{cases} \tag{4.35}$$

where $\widehat{\mu}_1$ and $\widehat{\mu}_2$ are the first and second sample moments, $\widehat{\mu}_j = n_1^{-1} \sum_{i=1}^{n} \mathbb{1}\{0 < y_i < R\} y_i^j$, $j = 1, 2$. To solve the system (4.35) for $\widehat{t}_{z,n}$ and $\widehat{\sigma}_n$ we initialize the system as below (Cohen, 1950). The initial values $t_{z,\text{start}}$ and $T_{z,\text{start}}$ are, respectively, the solution of the following equations.

$$\frac{n_0}{n} = 1 - K_{0,t} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_z} e^{-\frac{s^2}{2}} \, ds,$$

$$\frac{n_2}{n} = K_{0,T} = \frac{1}{\sqrt{2\pi}} \int_{T_z}^{\infty} e^{-\frac{s^2}{2}} \, ds,$$

then from (4.20), we initialize σ as

$$\sigma_{\text{start}} = \frac{R}{T_{z,\text{start}} - t_{z,\text{start}}}.$$

Let

$$\boldsymbol{I}_{**}(t_z, \sigma) = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

be the Fisher information matrix given by equation (3.14), then

$$b_{11} = -\left(K_{0,t} - K_{0,T}\right) g_1(t_z, T_z),$$

$$b_{12} = b_{21} = -\left(K_{0,t} - K_{0,T}\right) \sigma^{-1} g_2(t_z, T_z),$$

$$b_{22} = -\left(K_{0,t} - K_{0,T}\right) \sigma^{-2} g_3(t_z, T_z),$$

where

$$\left\{ \begin{array}{lcl} g_1(t_z, T_z) & := & -\left[1 + t_z Z_1 - T_z Z_2 + \frac{\phi(t_z)}{1-K_{0,t}} Z_1 + \frac{\phi(T_z)}{K_{0,T}} Z_2\right], \\ g_2(t_z, T_z) & := & \frac{R Z_2}{\sigma}\left[\frac{\phi(T_z)}{K_{0,T}} - T_z\right] + \left[Z_1 - Z_2 - t_z\right], \\ g_3(t_z, T_z) & := & \left(\frac{R}{\sigma}\right)^2 Z_2 \left(T_z - \frac{\phi(T_z)}{K_{0,T}}\right) - \left[2 - t_z(Z_1 - Z_2 - t_z) - \frac{Z_2 R}{\sigma}\right]. \end{array} \right. \tag{4.36}$$

Then it follows that

$$\left(\widehat{t}_{z,n}, \widehat{\sigma}_n\right) \sim \mathcal{AN}\left((t_z, \sigma), \frac{1}{n\left(K_{0,t} - K_{0,T}\right)\left(g_1 g_3 - g_2^2\right)} \begin{bmatrix} -g_3 & \sigma g_2 \\ \sigma g_2 & -\sigma^2 g_1 \end{bmatrix}\right). \qquad \square$$

Example 4.14. *Payment-per-payment Data.*

Let $y_1, \ldots, y_n$ be an *i.i.d.* sample with cdf (2.11) and pdf (2.12) with policy limit u, deductible d, and coinsurance factor c. For notational simplicity and to borrow the symbols from previous examples, we assume that $d \equiv t$ and $u \equiv T$. Then, the log-likelihood function given by equation (3.10) becomes

$$\begin{aligned} l_{\otimes}(t_z, \sigma) = \text{constant} + n_2 \log\left(K_{0,T}\right) - n \log\left(K_{0,t}\right) - n_1 \log\left(\sigma\right) \\ - \frac{1}{2} \sum_{0<y_i<R} \left(t_z + \frac{y_i}{c\sigma}\right)^2. \end{aligned} \tag{4.37}$$

Thus, setting

$$\left\{ \begin{array}{lcl} \frac{\partial l_{\otimes}}{\partial t_z} & = & 0, \\ \frac{\partial l_{\otimes}}{\partial \sigma} & = & 0, \end{array} \right.$$

yields the system of MLE equations

$$
\begin{cases} n\dfrac{\phi(t_z)}{K_{0,t}} - n_2\dfrac{\phi(T_z)}{K_{0,T}} - \displaystyle\sum_{0<y_i<R}\left(t_z + \dfrac{y_i}{c\sigma}\right) & = & 0, \\ n_2\dfrac{R\phi(T_z)}{\sigma^2 K_{0,T}} - \dfrac{n_1}{\sigma} + \dfrac{1}{c\sigma^2}\displaystyle\sum_{0<y_i<R} y_i\left(t_z + \dfrac{y_i}{c\sigma}\right) & = & 0. \end{cases} \tag{4.38}
$$

Define

$$
\begin{cases} Q_1 & := \frac{n}{n_1}\frac{\phi(t_z)}{K_{0,t}}, \\ Q_2 & := \frac{n_2}{n_1}\frac{\phi(T_z)}{K_{0,T}}, \end{cases} \tag{4.39}
$$

then the system of MLE equations (4.38) takes the form

$$
\begin{cases} \sigma\left(Q_1 - Q_2 - t_z\right) - c^{-1}\widehat{\mu}_1 & = & 0, \\ \sigma^2\left(1 - t_z(Q_1 - Q_2 - t_z) - \frac{Q_2 R}{\sigma}\right) - c^{-2}\widehat{\mu}_2 & = & 0, \end{cases} \tag{4.40}
$$

where $\widehat{\mu}_1$ and $\widehat{\mu}_2$ are the first and second sample moments, $\widehat{\mu}_j = n_1^{-1}\sum_{i=1}^{n}\mathbb{1}\{0 < y_i < R\}y_i^j$, $j = 1, 2$. To solve the system (4.40) for $\widehat{t}_{z,n}$ and $\widehat{\sigma}_n$ we initialize the system as below:

$$
\sigma_{\text{start}} = \sqrt{c^{-2}\widehat{\mu}_2 - (c^{-1}\widehat{\mu}_1)^2} \qquad \text{and} \qquad t_{z,\text{start}} = -\frac{c^{-1}\widehat{\mu}_1}{\sigma_{\text{start}}}.
$$

Let

$$
\boldsymbol{I}_{\otimes}(t_z, \sigma) = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}
$$

be the Fisher information matrix given by equation (3.15), then

$$
c_{11} = -\frac{(K_{0,t} - K_{0,T})}{K_{0,t}} r_1(t_z, T_z),
$$

$$
c_{12} = c_{21} = -\frac{(K_{0,t} - K_{0,T})}{K_{0,t}} \sigma^{-1} r_2(t_z, T_z),
$$

$$
c_{22} = -\frac{(K_{0,t} - K_{0,T})}{K_{0,t}} \sigma^{-2} r_3(t_z, T_z),
$$

where

$$\begin{cases} r_1(t_z, T_z) & := & -\left[1 + t_z Z_1 - T_z Z_2 - \frac{\phi(t_z)}{K_{0,t}} Z_1 + \frac{\phi(T_z)}{K_{0,T}} Z_2\right], \\ r_2(t_z, T_z) & := & \frac{R Z_2}{\sigma}\left[\frac{\phi(T_z)}{K_{0,T}} - T_z\right] + [Z_1 - Z_2 - t_z], \\ r_3(t_z, T_z) & := & \left(\frac{R}{\sigma}\right)^2 Z_2 \left(T_z - \frac{\phi(T_z)}{K_{0,T}}\right) - \left[2 - t_z(Z_1 - Z_2 - t_z) - \frac{Z_2 R}{\sigma}\right]. \end{cases} \tag{4.41}$$

Therefore, it follows that

$$(\widehat{t}_{z,n}, \widehat{\sigma}_n) \sim \mathcal{AN}\left((t_z, \sigma), \frac{K_{0,t}}{n\,(K_{0,t} - K_{0,T})\,(r_1 r_3 - r_2^2)} \begin{bmatrix} -r_3 & \sigma r_2 \\ \sigma r_2 & -\sigma^2 r_1 \end{bmatrix}\right). \qquad \square$$

Example 4.15. *Payment-per-Loss Data.*

Consider an observed *i.i.d.* sample $z_1, \ldots, z_n$ defined by (2.14). Then the linearly transformed *i.i.d.* sample $\frac{z_1}{c} + d, \ldots, \frac{z_n}{c} + d$ is exactly the sample $x_1^{**}, \ldots, x_n^{**}$ treated in Example 4.13 with $d \equiv t$ and $u \equiv T$. $\square$

4.2.2 Method of Trimmed Moments

In the following examples, we specify MTM estimators for the data scenarios and models of Chapter 2 when $X \sim \text{Normal}(\theta, \sigma^2)$. Since the location parameter θ can be any real number, we choose $h_1(x) = x$, but to ensure that the estimator of σ is positive, we choose $h_2(x) = x^2$.

Example 4.16. *Complete Data.*

This scenario has been fully investigated by Brazauskas *et al.* (2009). Using equation (3.16), the sample trimmed moments are:

$$\widehat{\mu}_1 = \frac{1}{n - m_n(1) - m_n^*(1)} \sum_{i=m_n(1)+1}^{n-m_n^*(1)} x_{i:n},$$

$$\widehat{\mu}_2 = \frac{1}{n - m_n(2) - m_n^*(2)} \sum_{i=m_n(2)+1}^{n-m_n^*(2)} x_{i:n}^2,$$

with $m_n(1)/n = m_n(2)/n \to a$ and $m_n^*(1)/n = m_n^*(2)/n \to b$.

The corresponding population trimmed moments by using equation (3.17) are computed as

$$\mu_1 := \frac{1}{1-a-b}\int_a^{1-b} F^{-1}(u|\boldsymbol{\theta})\,du = \theta + \sigma c_1,$$

$$\mu_2 := \frac{1}{1-a-b}\int_a^{1-b} \left[F^{-1}(u|\boldsymbol{\theta})\right]^2\,du = \theta^2 + 2\theta\sigma c_1 + \sigma^2 c_2,$$

where

$$c_k \equiv c_k(\Phi, a, b) := \frac{1}{1-a-b}\int_a^{1-b} \left[\Phi^{-1}(u)\right]^k\,du. \tag{4.42}$$

The MTM estimators of θ and σ are then

$$\begin{cases} \widehat{\widehat{\theta}}_n &= \widehat{\mu}_1 - c_1\widehat{\widehat{\sigma}}_n =: g_1(\widehat{\mu}_1, \widehat{\mu}_2), \\ \widehat{\widehat{\sigma}}_n &= \sqrt{(\widehat{\mu}_2 - \widehat{\mu}_1^2/(c_2 - c_1^2))} =: g_2(\widehat{\mu}_1, \widehat{\mu}_2). \end{cases} \tag{4.43}$$

The asymptotic distribution is

$$(\widehat{\widehat{\theta}}_n, \widehat{\widehat{\sigma}}_n) \sim \mathcal{AN}\left((\theta,\sigma), \frac{\sigma^2}{n}\boldsymbol{S}\right) \quad \text{with} \quad \boldsymbol{S} := \sigma^{-2}\boldsymbol{D\Sigma D'}, \tag{4.44}$$

where the variance-covariance matrix $\boldsymbol{\Sigma}$ is computed by using equation (3.29). The entries of the matrix $\boldsymbol{D}$ as in Theorem 1.1 can be computed by using the functions $g_i,\ 1 \le i \le 2$ given by (4.43).

$$\begin{aligned} d_{11} &= \left.\frac{\partial g_1}{\partial \widehat{\mu}_1}\right|_{(\mu_1,\mu_2)} = 1 - c_1\left.\frac{\partial g_2}{\partial \widehat{\mu}_1}\right|_{(\mu_1,\mu_2)} = \frac{c_1\theta + c_2\sigma}{\sigma(c_2 - c_1^2)}, \\ d_{12} &= \left.\frac{\partial g_1}{\partial \widehat{\mu}_2}\right|_{(\mu_1,\mu_2)} = -c_1\left.\frac{\partial g_2}{\partial \widehat{\mu}_1}\right|_{(\mu_1,\mu_2)} = -\frac{c_1}{2\sigma(c_2 - c_1^2)}, \\ d_{21} &= \left.\frac{\partial g_2}{\partial \widehat{\mu}_1}\right|_{(\mu_1,\mu_2)} = \left.\frac{-\widehat{\mu}_1}{\sqrt{(\widehat{\mu}_2 - \widehat{\mu}_1^2)/(c_2 - c_1^2)}}\right|_{(\mu_1,\mu_2)} = -\frac{\theta + c_1\sigma}{\sigma(c_2 - c_1^2)}, \\ d_{22} &= \left.\frac{\partial g_2}{\partial \widehat{\mu}_1}\right|_{(\mu_2,\mu_2)} = \left.\frac{0.5}{\sqrt{(\widehat{\mu}_2 - \widehat{\mu}_1^2)/(c_2 - c_1^2)}}\right|_{(\mu_1,\mu_2)} = \frac{1}{2\sigma(c_2 - c_1^2)}, \end{aligned} \tag{4.45}$$

The matrix $\boldsymbol{S}$ does not depend on any unknown parameters which makes the estimator vector $\widehat{\widehat{\boldsymbol{\theta}}}_n = (\widehat{\widehat{\theta}}_n, \widehat{\widehat{\sigma}}_n)$ asymptotically consistent. □

Example 4.17. *Truncated Data.*

Let $x_1^*, \ldots, x_n^*$ be an *i.i.d.* sample of truncated normal data in the interval $(t, T]$

with qf given by (2.5). Then, following the procedure of Example 3.12, we have

$$\widehat{\mu}_{*,1} = \frac{1}{n - m_n(1) - m_n^*(1)} \sum_{i=m_n(1)+1}^{n-m_n^*(1)} x_{i:n}^*,$$

$$\widehat{\mu}_{*,2} = \frac{1}{n - m_n(2) - m_n^*(2)} \sum_{i=m_n(2)+1}^{n-m_n^*(2)} x_{i:n}^{*2},$$

with $m_n(1)/n = m_n(2)/n \to a$ and $m_n^*(1)/n = m_n^*(2)/n \to b$.

The corresponding population trimmed moments by using equation (3.20) with qf (2.5) are computed as

$$\mu_{*,1} := \frac{1}{1-a-b} \int_a^{1-b} F_*^{-1}(u)\, du = \theta + \sigma c_{*,1},$$

$$\mu_{*,2} := \frac{1}{1-a-b} \int_a^{1-b} \left[F_*^{-1}(u)\right]^2\, du = \theta^2 + 2\theta\sigma c_{*,1} + \sigma^2 c_{*,2},$$

where

$$c_{*,k} \equiv c_{*,k}(\Phi, a, b, t, T) := \frac{1}{1-a-b} \int_a^{1-b} \left[\Phi^{-1}\left(u\Phi(T_z) + (1-u)\Phi(t_z)\right)\right]^k\, du, \tag{4.46}$$

and

$$t_z := \frac{t-\theta}{\sigma}, \qquad T_z := \frac{T-\theta}{\sigma}.$$

as defined at the beginning of Section 4.2. Note that, $c_{*,k}$ do depend on the unknown parameters θ and σ. Equating $\mu_{*,1} = \widehat{\mu}_{*,1}$, and $\mu_{*,2} = \widehat{\mu}_{*,2}$ yields the implicit (the system 4.43 is explicit) system of equations to be solved for θ and σ:

$$\begin{cases} \theta &= \widehat{\mu}_{*,1} - c_{*,1}\sigma, \\ \sigma &= \sqrt{(\widehat{\mu}_{*,2} - \widehat{\mu}_{*,1}^2)/(c_{*,2} - c_{*,1}^2)} \end{cases}. \tag{4.47}$$

The system of equations (4.47) can be solved for $\widehat{\widehat{\sigma}}_n$ and $\widehat{\widehat{\theta}}_n$ by using an iterative numerical method with the initializing values

$$\sigma_{\text{start}} = \sqrt{\widehat{\mu}_{*,2} - \widehat{\mu}_{*,1}^2}, \quad \text{and} \quad \theta_{\text{start}} = \widehat{\mu}_{*,1}.$$

A special case with $a = 0 = b$ (which translates to classical method of moments), can be handled in the following way. Instead of the first two moments, i.e., $h_1(x) = x$, $h_2(x) = x^2$, we take the first four sample moments and match them with the corresponding population moments. That is

$$\widehat{\mu}_{*,j} = \frac{\sum_{i=1}^{n} x_i^{*j}}{n}, \qquad \text{for} \qquad j = 1, 2, 3, 4.$$

The expected value and the variance of the doubly truncated normal random variable can be found in Ergashev *et al.* (2016), Johnson *et al.* (1994). However, with the help of mathematical induction, all moments can be computed as follows.

Lemma 4.2. *Define $\mu_{*,-1} = 0$ and $\mu_{*,0} = 1$. Then the moments of doubly truncated, in the interval $(t, T]$, normally distributed random variable X^* with parameters (θ, σ) are given by,*

$$\mu_{*,j} = (j-1)\sigma^2 \mu_{*,j-2} + \theta \mu_{*,j-1} + \sigma p^{-1} \left[t^{j-1} \phi\left(t_z\right) - T^{j-1} \phi\left(T_z\right) \right], j = 1, 2, \ldots$$

where $p = \Phi\left(T_z\right) - \Phi\left(t_z\right)$.

Then, by Central Limit Theorem (see, e.g., Serfling, 1980), we have

$$\widehat{\boldsymbol{\mu}}_* \sim \mathcal{AN}\left(\boldsymbol{\mu}_*, \frac{1}{n}\boldsymbol{\Sigma}_*\right),$$

where the (i, j)th entry of $\boldsymbol{\Sigma}_*$ is equal to $\mu_{*,i+j} - \mu_{*,i}\mu_{*,j}$ for $j = 1, 2, 3, 4$. Following the work of Shah and Jaiswal (1966), θ and σ^2 can be expressed as

$$\theta = t - \frac{N_\theta}{D_\theta} =: g_\theta(\mu_{*,1}, \mu_{*,2}, \mu_{*,3}, \mu_{*,4}),$$

$$\sigma^2 = \frac{N_\sigma}{D_\sigma} =: g_\sigma(\mu_{*,1}, \mu_{*,2}, \mu_{*,3}, \mu_{*,4}),$$

where $D_\theta = D_\sigma$ and

$$
\begin{aligned}
R &:= T - t, \quad \text{as defined above,}\\
N_\theta &:= -2t(2R^2+6tR+5t^2)\mu_{*,1}^2 + (12tR+15t^2+2R^2)\mu_{*,1}\mu_{*,2} + 2t\mu_{*,1}\mu_{*,3}\\
&\quad + t^2(3R^2+8tR+5t^2)\mu_{*,1} - 3(T+2t)\mu_{*,2} + 3\mu_2\mu_{*,3} + tR^2\mu_{*,2}\\
&\quad - (4tR+5t^2+R^2)\mu_{*,3} + (T+t)\mu_{*,4} - t^3R^2 - 2t^4R - t^5,\\
N_\sigma &:= t^2T^2\mu_{*,1}^2 - t(R^2+3tR+2t^2)\mu_{*,1}\mu_{*,2} - (2tR+2t^2+R^2)\mu_{*,1}\mu_{*,3}\\
&\quad + (3tR+3t^2+R^2)\mu_{*,2}^2 - (T+t)\mu_{*,2}\mu_{*,3} - \mu_{*,2}\mu_{*,4} - t^2T^2\mu_{*,2}\\
&\quad + \mu_{*,3}^2 + t(R^2+3tR+2t^2)\mu_{*,3} - tT\mu_{*,4} + (T+t)\mu_{*,1}\mu_{*,4},\\
D_\theta &:= -2(tR+3t+R^2)\mu_{*,1}^2 + 3(T+t)\mu_{*,1}\mu_{*,2} + 2t(R^2+3tR+2t^2)\mu_{*,1}\\
&\quad + 2\mu_{*,1}\mu_{*,3} - 3\mu_{*,2}^2 + R^2\mu_{*,2} - (T+t)\mu_{*,3} - t^2R^2 - 2t^3R - t^4.
\end{aligned}
$$

Now, let the Jacobian matrix be

$$
\boldsymbol{D}_* := \begin{bmatrix} \frac{\partial g_\theta}{\partial \mu_{*,1}} & \frac{\partial g_\theta}{\partial \mu_{*,2}} & \frac{\partial g_\theta}{\partial \mu_{*,3}} & \frac{\partial g_\theta}{\partial \mu_{*,4}} \\ \frac{\partial g_\sigma}{\partial \mu_{*,1}} & \frac{\partial g_\sigma}{\partial \mu_{*,2}} & \frac{\partial g_\sigma}{\partial \mu_{*,3}} & \frac{\partial g_\sigma}{\partial \mu_{*,4}} \end{bmatrix}.
$$

Then by the delta method (Theorem 1.1), we have

$$
(\widehat{\widehat{\theta}}_n, \widehat{\widehat{\sigma}}^2_n) \sim \mathcal{AN}\left((\theta,\sigma^2), \frac{1}{n}\boldsymbol{D}_*\boldsymbol{\Sigma}_*\boldsymbol{D}_*'\right).
$$

□

Example 4.18. *Censored Data.*

Let $x_1^{**}, \ldots, x_n^{**}$ be an *i.i.d.* sample of censored normal data in the interval $(t, T]$ with qf given by (2.9). Then, following the procedure of Example 3.13, we have

$$
\widehat{\mu}_{**,1} = \frac{1}{n - m_n - m_n^*} \sum_{i=m_n+1}^{n-m_n^*} x_{i:n}^{**},
$$

$$
\widehat{\mu}_{**,2} = \frac{1}{n - m_n - m_n^*} \sum_{i=m_n+1}^{n-m_n^*} x_{i:n}^{**2},
$$

with $m_n/n \to a$ and $m_n^*/n \to b$. The corresponding population trimmed moments (3.22) with the qf defined by (2.9) are given by

$$\begin{aligned}
\mu_{**,1} &:= \frac{1}{1-a-b}\int_a^{1-b} F_{**}^{-1}(u)\, du \\
&= \frac{1}{1-a-b}\int_a^{1-b} F^{-1}(u|\boldsymbol{\theta})\, du \\
&= \theta + \sigma c_1, \\
\mu_{**,2} &:= \frac{1}{1-a-b}\int_a^{1-b} \left[F_{**}^{-1}(u)\right]^2\, du \\
&= \frac{1}{1-a-b}\int_a^{1-b} \left[F^{-1}(u|\boldsymbol{\theta})\right]^2\, du \\
&= \theta^2 + 2\theta\sigma c_1 + \sigma^2 c_2,
\end{aligned}$$

by assuming the most general case that $0 \leq F(t) \leq a < 1-b \leq F(T) \leq 1$ as in Example 3.13. Thus, with the assumption $0 \leq F(t) \leq a < 1-b \leq F(T) \leq 1$, this case translates to the complete case as in Example 4.16. $\square$

Example 4.19. *Payment-per-payment Data.*

Let $y_1, \ldots, y_n$ be an *i.i.d.* sample of payment-per-payment data defined by (2.10) with qf (2.13). Since c, d, and u are assumed to be known constants, then we linearly transform the sample as $\frac{y_1}{c} + d, \ldots, \frac{y_n}{c} + d$. Then, following the procedure of Example 3.14, the sample trimmed moments given by (3.16) are

$$\widehat{\mu}_{\otimes,1} = \frac{1}{n - m_n - m_n^*}\sum_{i=m_n+1}^{n-m_n^*}\left(\frac{y_{i:n}}{c} + d\right),$$

$$\widehat{\mu}_{\otimes,2} = \frac{1}{n - m_n - m_n^*}\sum_{i=m_n+1}^{n-m_n^*}\left(\frac{y_{i:n}}{c} + d\right)^2,$$

with $m_n/n \to a$ and $m_n^*/n \to b$. As in Example 4.9, assume that $m_n^* \geq \sum_{i=1}^{n} \mathbb{1}\{y_i = c(u-d)\}$, then we end up with the most general second case of Example 3.14. That, is $0 \leq a < 1-b \leq s^* \leq 1$ with $s^* = \frac{F(u)-F(d)}{S(d)}$. Then, the corresponding population trimmed moments (3.27b) with the qf defined by (2.13) are given by

$$\begin{aligned}\mu_{\otimes,1} &= \frac{1}{1-a-b}\int_a^{1-b}\big[F^{-1}\big(s+(1-s)F(d|\boldsymbol{\theta})|\boldsymbol{\theta}\big)\big]\ ds,\\ &= \theta+\sigma c_{\otimes,1},\\ \mu_{\otimes,2} &= \frac{1}{1-a-b}\int_a^{1-b}\big[F^{-1}\big(s+(1-s)F(d|\boldsymbol{\theta})|\boldsymbol{\theta}\big)\big]^2\ ds\\ &= \theta^2+2\theta\sigma c_{\otimes,1}+\sigma^2 c_{\otimes,2},\end{aligned}$$

where

$$c_{\otimes,k}\equiv c_{\otimes,k}(\Phi,a,b,d,u):=\frac{1}{1-a-b}\int_a^{1-b}\left[\Phi^{-1}\left(s+(1-s)\Phi(d_z)\right)\right]^k\ ds, \tag{4.48}$$

and $d_z := (d-\theta)/\sigma$. Again, $c_{\otimes,k}$ depend on the unknown parameters. Equating $\mu_{\otimes,1}=\widehat{\mu}_{\otimes,1}$, and $\mu_{\otimes,2}=\widehat{\mu}_{\otimes,2}$ yields the implicit system of equations to be solved for θ and σ:

$$\left\{\begin{array}{lcl}\theta &=& \widehat{\mu}_{\otimes,1}-c_{\otimes,1}\sigma =: g_1(\widehat{\mu}_{\otimes,1},\widehat{\mu}_{\otimes,2}),\\ \sigma &=& \sqrt{(\widehat{\mu}_{\otimes,2}-\widehat{\mu}^2_{\otimes,1})/(c_{\otimes,2}-c^2_{\otimes,1})} =: g_2(\widehat{\mu}_{\otimes,1},\widehat{\mu}_{\otimes,2}).\end{array}\right. \tag{4.49}$$

The system of equations (4.49) can be solved for $\widehat{\widehat{\sigma}}_n$ and $\widehat{\widehat{\theta}}_n$ by using an iterative numerical method with the initializing values

$$\sigma_{\text{start}}=\sqrt{\widehat{\mu}_{\otimes,2}-\widehat{\mu}^2_{\otimes,1}},\quad \text{and}\quad \theta_{\text{start}}=\widehat{\mu}_{\otimes,1}.$$

The entries of the variance-covariance matrix $\mathbf{\Sigma}_\otimes$ calculated using (3.33) are

$$\begin{aligned}\sigma^2_{11} &= \sigma^2\frac{(1-\Phi(d_z))^2}{(1-a-b)^2}\\ &\quad\times\int_a^{1-b}\int_a^{1-b}\frac{\min\{u,v\}-uv}{\phi\left[\Phi^{-1}\left(v+(1-v)\Phi(d_z)\right)\right]\phi\left[\Phi^{-1}\left(u+(1-u)\Phi(d_z)\right)\right]}\,dv\,du,\\ \sigma^2_{12} &= \frac{2\theta\sigma^2(1-\Phi(d_z))^2}{(1-a-b)^2}\\ &\quad\times\int_a^{1-b}\int_a^{1-b}\frac{\min\{u,v\}-uv}{\phi\left[\Phi^{-1}\left(v+(1-v)\Phi(d_z)\right)\right]\phi\left[\Phi^{-1}\left(u+(1-u)\Phi(d_z)\right)\right]}\,dv\,du\end{aligned}$$

$$
\begin{aligned}
&+\frac{2\sigma^3(1-\Phi(d_z))^2}{(1-a-b)^2}\\
&\times\int_a^{1-b}\int_a^{1-b}\frac{[\min\{u,v\}-uv]\,\Phi^{-1}(u+(1-u)\Phi(d_z))}{\phi\left[\Phi^{-1}(v+(1-v)\Phi(d_z))\right]\phi\left[\Phi^{-1}(u+(1-u)\Phi(d_z))\right]}\,dv\,du,\\
\sigma_{22}^2=&\ \frac{4\theta^2\sigma^2(1-\Phi(d_z))^2}{(1-a-b)^2}\\
&\times\int_a^{1-b}\int_a^{1-b}\frac{\min\{u,v\}-uv}{\phi\left[\Phi^{-1}(v+(1-v)\Phi(d_z))\right]\phi\left[\Phi^{-1}(u+(1-u)\Phi(d_z))\right]}\,dv\,du\\
&+\frac{8\theta\sigma^3(1-\Phi(d_z))^2}{(1-a-b)^2}\\
&\times\int_a^{1-b}\int_a^{1-b}\frac{[\min\{u,v\}-uv]\,\Phi^{-1}(u+(1-u)\Phi(d_z))}{\phi\left[\Phi^{-1}(v+(1-v)\Phi(d_z))\right]\phi\left[\Phi^{-1}(u+(1-u)\Phi(d_z))\right]}\,dv\,du\\
&+\frac{4\sigma^4(1-\Phi(d_z))^2}{(1-a-b)^2}\int_a^{1-b}\int_a^{1-b}\left[\frac{[\min\{u,v\}-uv]\,\Phi^{-1}(u+(1-u)\Phi(d_z))}{\phi\left[\Phi^{-1}(v+(1-v)\Phi(d_z))\right]}\right.\\
&\qquad\qquad\left.\times\frac{\Phi^{-1}(v+(1-v)\Phi(d_z))}{\phi\left[\Phi^{-1}(u+(1-u)\Phi(d_z))\right]}\right]dv\,du.
\end{aligned}
$$

For $k=1,2$; it follows evidently that

$$
\begin{cases}
\frac{\partial c_{\otimes,k}}{\partial\theta} = -\frac{2^{k-1}\phi(d_z)}{\sigma(1-a-b)}\displaystyle\int_a^{1-b}\frac{(1-s)\left[\Phi^{-1}(s+(1-s)\Phi(d_z))\right]^{k-1}}{\phi\left[\Phi^{-1}(s+(1-s)\Phi(d_z))\right]}\,ds,\\
\frac{\partial c_{\otimes,k}}{\partial\sigma} = -\frac{2^{k-1}(d-\theta)\phi(d_z)}{\sigma^2(1-a-b)}\displaystyle\int_a^{1-b}\frac{(1-s)\left[\Phi^{-1}(s+(1-s)\Phi(d_z))\right]^{k-1}}{\phi\left[\Phi^{-1}(s+(1-s)\Phi(d_z))\right]}\,ds.
\end{cases}
\tag{4.50}
$$

For $k=1,2$; let us denote

$$
\theta_{\mu_{\otimes,k}} := \left.\frac{\partial g_1}{\partial\widehat{\mu}_{\otimes,k}}\right|_{(\mu_{\otimes,1},\mu_{\otimes,2})}\quad\text{and}\quad \sigma_{\mu_{\otimes,k}} := \left.\frac{\partial g_2}{\partial\widehat{\mu}_{\otimes,k}}\right|_{(\mu_{\otimes,1},\mu_{\otimes,2})}.
$$

Consider the following more notations

$$
\begin{aligned}
f_{11}(\theta,\sigma) &:= 1+\sigma\frac{\partial c_{\otimes,1}}{\partial\theta}, & f_{12}(\theta,\sigma) &:= c_{\otimes,1}+\sigma\frac{\partial c_{\otimes,1}}{\partial\sigma},\\
f_{21}(\theta,\sigma) &:= \frac{\partial c_{\otimes,2}}{\partial\theta}-2c_{\otimes,1}\frac{\partial c_{\otimes,1}}{\partial\theta}, & f_{22}(\theta,\sigma) &:= \frac{\partial c_{\otimes,2}}{\partial\sigma}-2c_{\otimes,1}\frac{\partial c_{\otimes,1}}{\partial\sigma}.
\end{aligned}
$$

The entries of the matrix $\boldsymbol{D}_\otimes$ are found by implicitly differentiating the functions g_i (with multivariate chain rule) from equations (4.49) with the help of equations

(4.50).

$$d_{11} = \theta_{\mu_{\otimes,1}} = \frac{1 - f_{12}\sigma_{\mu_{\otimes,1}}}{f_{11}} = \frac{1 - f_{12}d_{21}}{f_{11}},$$

$$d_{12} = \theta_{\mu_{\otimes,2}} = -\frac{f_{12}\sigma_{\mu_{\otimes,2}}}{f_{11}} = -\frac{f_{12}d_{22}}{f_{11}},$$

$$d_{21} = \sigma_{\mu_{\otimes,1}} = -\frac{K\left[2f_{11}\mu_{\otimes,1}(c_{\otimes,2} - c_{\otimes,1}^2) + f_{21}(\mu_{\otimes,2} - \mu_{\otimes,1}^2)\right]}{f_{11}(c_{\otimes,2} - c_{\otimes,1}^2)^2 + K(\mu_{\otimes,2} - \mu_{\otimes,1}^2)(f_{11}f_{22} - f_{12}f_{21})},$$

$$d_{22} = \sigma_{\mu_{\otimes,2}} = \frac{Kf_{11}(c_{\otimes,2} - c_{\otimes,1}^2)}{f_{11}(c_{\otimes,2} - c_{\otimes,1}^2)^2 + K(\mu_{\otimes,2} - \mu_{\otimes,1}^2)(f_{11}f_{22} - f_{12}f_{21})},$$

where $K := \frac{1}{2}\sqrt{\frac{c_{\otimes,2} - c_{\otimes,1}^2}{\mu_{\otimes,2} - \mu_{\otimes,1}^2}}$. Hence the asymptotic result (3.31) becomes

$$(\widehat{\widehat{\theta}}_n, \widehat{\widehat{\sigma}}_n) \sim \mathcal{AN}\left((\theta, \sigma), n^{-1}\boldsymbol{D}_{\otimes}\boldsymbol{\Sigma}_{\otimes}\boldsymbol{D}_{\otimes}'\right). \tag{4.51}$$

$\square$

Example 4.20. *Payment-per-loss Data.*

Consider an observed *i.i.d.* sample $z_1, \ldots, z_n$ defined by (2.14). Then, as in Example 4.15, the linearly transformed *i.i.d.* sample $\frac{z_1}{c} + d, \ldots, \frac{z_n}{c} + d$ is exactly the sample $x_1^{**}, \ldots, x_n^{**}$ treated in Example 4.18 with $d \equiv t$ and $u \equiv T$. $\square$

Chapter 5

Real Data Illustrations

In this chapter, we study the practical performance of the estimators developed in Chapter 4. Specifically, we fit Pareto I and lognormal models using MLE and MTM approaches to the Norwegian fire claims data for the year 1983. After validating the models, we use them to price an insurance contract and investigate the effect of model estimation on the actuarial premium.

Let us start by introducing the data set.

- Data set is available at

  http://lstat.kuleuven.be/Wiley/ (in Chapter 1, file NORWEGIANFIRE.TXT).

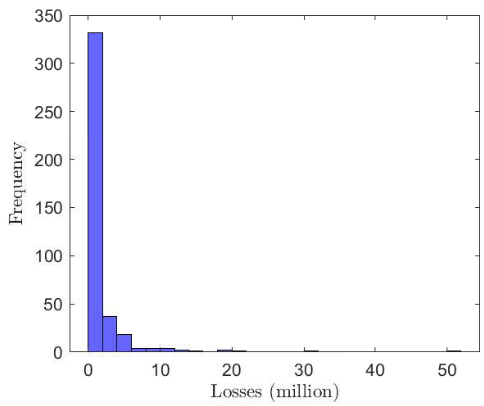
- It represents total damage done by fires in Norway for the year 1983.

- Losses are measured in thousands of Norwegian kroner. It is unknown if claims were inflation adjusted.

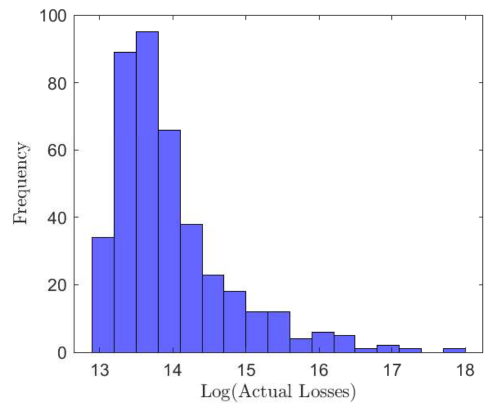
- Only damages above $500,000$ are reported (i.e., data is left-truncated at $d := 500,000$). The sample size is $n = 407$.

5.1 Modeling Severity

As can be seen from Figure 5.1, the histograms of data and $\log(\text{data})$ are similar to many insurance loss distributions. Moreover, there are few observations far in the right tail, which suggests that a right-skewed and heavy-tailed distribution might be a reasonable choice. In view of this, we will consider Pareto I and lognormal distributions.



(a) Original Data



(b) Log transformation of the data

Figure 5.1: The histograms of Norwegian fire claims and log-transformed claims for the year 1983.

The cumulative distribution function (cdf) of the single parameter Pareto and lognormal models are, respectively, given by

$$\text{Pareto I}(\alpha, x_0) : F(x) = 1 - \left(\frac{x_0}{x}\right)^{\alpha}, \; x > x_0, \quad \text{and}$$

$$\text{LN}(\theta, \sigma) : F(x) = \Phi\left(\frac{\log(x) - \theta}{\sigma}\right), \; x > 0,$$

where $\alpha > 0$, $-\infty < \theta < \infty$, and $\sigma > 0$ are unknown parameters with Φ denoting the standard normal cdf. The parameter $x_0 > 0$ is assumed to be known in advance and we set it to be $x_0 = 1$ (i.e., one Norwegian krone).

Since the data is left truncated, we will use MLE and MTM estimators developed in Examples 4.9 and 4.19, with $u = \infty$. The trimming proportion pairs given by Table 5.1 will be used:

Table 5.1: Trimming proportion pairs for real data illustrations.

	Asymmetric Trimming			Symmetric Trimming			
Name	MTM2	MTM3	MTM4	MTM1	MTM5	MTM6	MTM7
a	.00	.00	.00	.00	.05	.10	.25
b	.05	.10	.25	.00	.05	.10	.25

Note that as follows from the results of Chapter 6, MTM1 coincides with MLE for Pareto I. Similar result for the lognormal distribution has been established by Ergashev *et al.* (2016). The estimated parameters are listed in Table 5.2 (at the end of this chapter).

5.2 Model Validation

To assess the quality of fits, we first present the quantile-quantile plots and then compute two goodness-of-fit statistics and their p-values for the observed data.

5.2.1 Quantile-Quantile Plots

In Figures 5.2 and 5.3, we present plots of the MTM fitted-versus-observed quantiles for Pareto I and lognormal models. That is, the points plotted in those graphs are the following pairs:

$$\left(\log\left(\widehat{F}_*^{-1}\left[s_i + (1 - s_i)\widehat{F}_*(d)\right]\right), \log\left(x_{i:n}^*\right)\right),$$

where $\widehat{F}_*$ is the estimated parametric cdf, $\widehat{F}_*^{-1}$ is the estimated parametric qf (see Section 2.2 for the corresponding definitions with $T \to \infty$, and $d \equiv t$), $x_{1:n}^* < \cdots < x_{n:n}^*$ denote the ordered observed claim severities, and $s_i = \frac{2i-1}{2n}$, $i = 1, 2, \ldots, n$, is the quantile level. The qq-plot pairs can easily be adjusted for complete loss data via $d \to 0$. Note that there are nine observations which are exactly equal to the priority $d = 500,000$. For the purpose of parameter estimation, construction of quantile-quantile plots, and for computation of other model validation measures, such data clusters were de-grouped using the method described in Brazauskas and Serfling (2003).

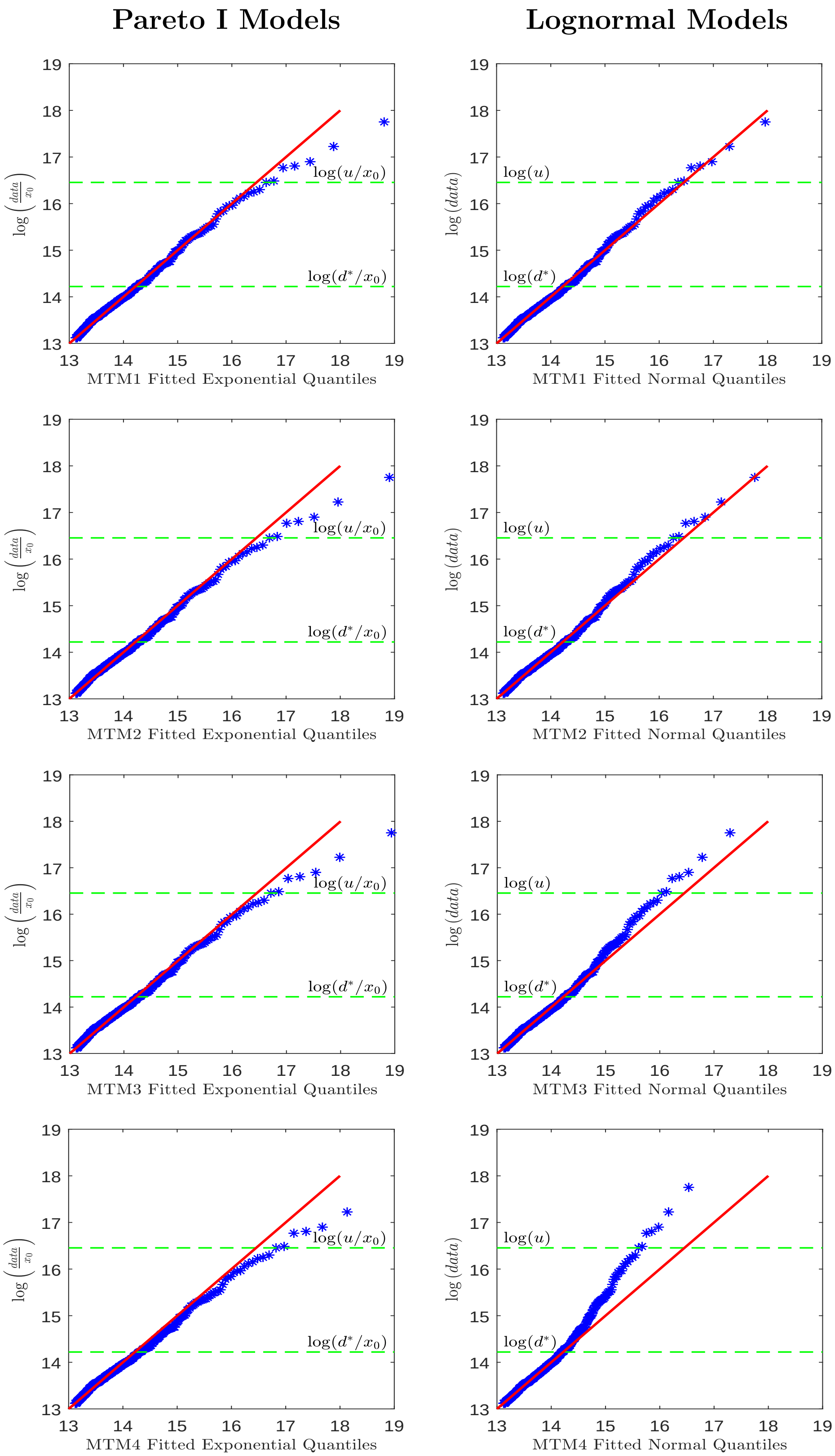


Figure 5.2: Log-fitted versus log-observed quantiles of Norwegian fire claims, using MLE (MTM1) and asymmetrically trimmed MTM estimators. Left panels: Pareto I models. Right panels: Lognormal models. The solid $45°$ red line represents the perfect fit. The dashed lines reflect contract specifications (to be used in Section 5.3).

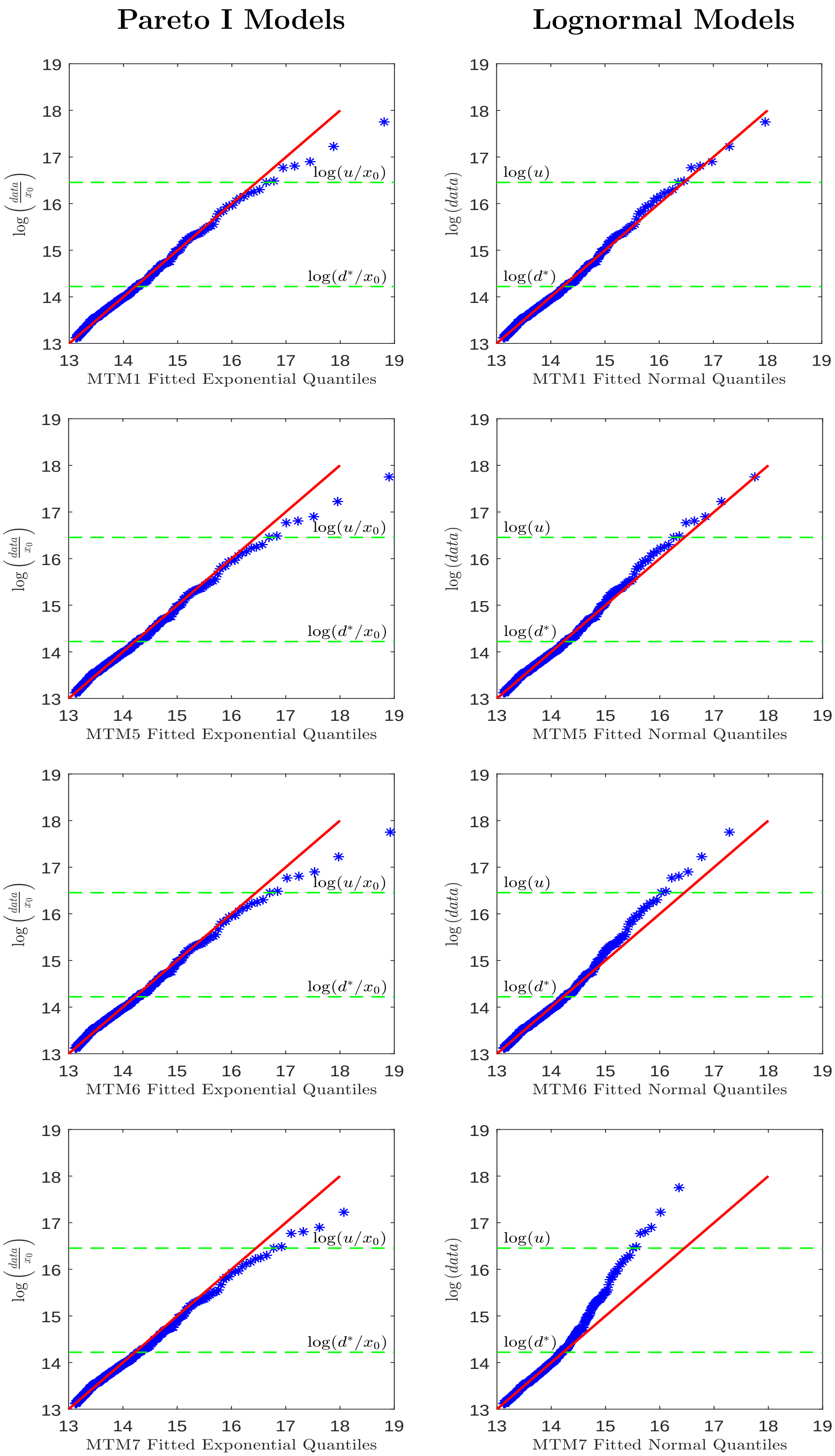


Figure 5.3: Log-fitted versus log-observed quantiles of Norwegian fire claims, using MLE (MTM1) and symmetrically trimmed MTM estimators. Left panels: Pareto I models. Right panels: Lognormal models. The solid 45° red line represents the perfect fit. The dashed lines reflect contract specifications (to be used in Section 5.3).

Discussion of Figures 5.2 and 5.3: All qq-plots seem to be alike for the Pareto I model regardless of the trimming proportions. Since 45° line is above the data points, Pareto I models seem to overestimate the right tail of the data. In contrast, all lognormal models underestimate the right tail of the claims. Also, this model is quite sensitive to the choice of the trimming proportions. Overall, we can infer from the qq-plots that the single parameter Pareto models capture the pattern of the data better than the lognormal counterparts. □

5.2.2 Goodness-of-Fit Statistics

To formally assess the "closeness" of the fitted model to the observed data, we will measure the distance (according to a selected measure) between the truncated empirical distribution function $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{d < x_i \leq x\}$ and the parametrically estimated left truncated (at deductible $d = 500,000$) distribution function:

$$\widehat{F}_*(x) = \frac{\widehat{F}(x) - \widehat{F}(d)}{1 - \widehat{F}(d)}, \quad x > d.$$

There are multiple options available to accomplish this task, for example, the mean absolute deviation is used both in Brazauskas *et al.* (2009) and Zhao *et al.* (2018a), but in this dissertation we choose to work with two popular discrepancy measures for individual data. They are (see, e.g., Klugman *et al.*, 2012, Chapter 16):

i. *maximum absolute* distance: Kolmogorov-Smirnov (KS) statistic, and

ii. *cumulative weighted quadratic* distance: Anderson-Darling (AD) statistic.

The Kolmogorov-Smirnov test statistic for truncated data is defined as

$$D = \max_{d < x \leq u} \left| F_n(x) - \widehat{F}_*(x) \right|,$$

where u is the right censoring point ($u = \infty$ if there is no censoring, which is the case for the data considered in this chapter). For computational purposes, the following form is more convenient:

$$D = \max_{1 \leq i \leq n} \left\{ \left| \widehat{F}_*(x^*_{i:n}) - \frac{i-1}{n} \right|, \left| \widehat{F}_*(x^*_{i:n}) - \frac{i}{n} \right| \right\},$$

where $x^*_{1:n} \leq x^*_{2:n} \leq \cdots \leq x^*_{n:n}$ denote the ordered claim severities. The Anderson-Darling test statistic for truncated data type is defined as

$$A^2 = \int_d^u \frac{\left[F_n(x) - \widehat{F}_*(x)\right]^2}{\widehat{F}_*(x)\left[1 - \widehat{F}_*(x)\right]} \widehat{f}_*(x)\, dx.$$

And the corresponding computationally convenient form is given by

$$\begin{aligned} A^2 &= -n\widehat{F}_*(u) + n\sum_{i=0}^{k} [1 - F_n(y_i)]^2 \log\left(\frac{1 - \widehat{F}_*(y_i)}{1 - \widehat{F}_*(y_{i+1})}\right) \\ &\quad + n\sum_{i=1}^{k} [F_n(y_i)]^2 \log\left(\frac{\widehat{F}_*(y_{i+1})}{\widehat{F}_*(y_i)}\right), \end{aligned}$$

where the unique non-censored data points are $d = y_0 < y_1 < \cdots < y_k < y_{k+1} = u < \infty$. If $u = \infty$, i.e., the observed data is only left truncated (no censoring), then

$$\begin{aligned} A^2 &= -n + n\sum_{i=0}^{k-1} [1 - F_n(y_i)]^2 \log\left(\frac{1 - \widehat{F}_*(y_i)}{1 - \widehat{F}_*(y_{i+1})}\right) \\ &\quad + n\sum_{i=1}^{k} [F_n(y_i)]^2 \log\left(\frac{\widehat{F}_*(y_{i+1})}{\widehat{F}_*(y_i)}\right). \end{aligned}$$

Both KS and AD test statistics can be reduced to the complete data case by letting $u \to \infty$ and $d \to 0$. The formal hypothesis testing results are summarized in Table 5.2. There we arrive at similar conclusions to those based on the qq-plots, i.e., Figures 5.2 and 5.3.

Discussion of Table 5.2: The p-values of both KS and AD test statistics for single parameter Pareto are larger than the corresponding values for lognormal models. For any trimming proportion, both KS and AD p-values for Pareto models are slightly higher than the corresponding values for lognormal models. Further, AD p-values for Pareto models fitted with MTM1, MTM3, and MTM6 exceed 0.10, which indicates that the single parameter Pareto is a plausible model for the data fitted with those trimming proportions. Neither KS nor AD p-values are higher than 0.10 for any trimming proportion used to fit lognormal models, which means that lognormal distribution may be inadequate for this data set. $\square$

5.3 Actuarial Premiums

We now consider estimation of the severity component of the pure premium for an insurance benefit equal to the amount by which a fire claim exceeds 1.5 million with co-insurance factor $c = 0.8$ and the maximum paid benefit of 10 million. That is,

$$\begin{aligned} \text{Deductible: } & d^* = 1,500,000; \\ \text{Policy Limit: } & u = 14,000,000; \quad \text{and} \\ \text{Co-insurance factor: } & c = 0.8. \end{aligned}$$

Note that there are five claims larger than u. Now, consider the random variable Y defined by equation (2.10) with the above specifications. That is,

$$Y \stackrel{d}{=} X|X > d^* = \begin{cases} 0.8\,(X - 1.5)\,, & 1.5 < X \leq 14; \\ 10, & 14 < X. \end{cases} \tag{5.1}$$

Then, we seek

$$\begin{aligned} \Pi[F] = \mathbb{E}[Y] &= \frac{c\,\{\mathbb{E}\,(X \wedge u) - \mathbb{E}\,(X \wedge d^*)\}}{\mathbb{P}(X > d^*)} \qquad (5.2) \\ &= \frac{c\left\{\displaystyle\int_{d^*}^{u} (x - d^*)\,dF(x|\boldsymbol{\theta}) + (u - d^*)(1 - F(u|\boldsymbol{\theta}))\right\}}{\mathbb{P}(X > d^*)} \\ &= \frac{0.8\left\{\displaystyle\int_{1.5}^{14} (x - 1.5)\,dF(x|\boldsymbol{\theta}) + 12.5(1 - F(14))\right\}}{\mathbb{P}(X > 1.5)}. \end{aligned}$$

The premium values including their corresponding 95% confidence intervals (CIs) with different fitted models are summarized in Table 5.2. We can compare the estimated parametric premiums, $\Pi[\widehat{F}]$, with the empirical premium $\Pi[F_m]$, where F_m denotes the empirical distribution function and $m = \sum_{i=1}^{n} \mathbb{1}\{x_i > d^*\}$. By Central Limit Theorem,

$$\Pi[F_m] \sim \mathcal{AN}\left(\Pi[F], m^{-1} Var(F)\right),$$

where $Var(F) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$. Thus, $\Pi[F_m]$ is simply the sample mean of the random variable, Y, and the sample variance of $\Pi[F_m]$ is the sample variance of Y divided by the number of observations above 1.5 (million).

On the other hand, the asymptotic distribution of $\Pi[\widehat{F}]$ can be established by using the asymptotic normality results for parameter estimators in conjunction with the delta method. That is, by Central Limit Theorem,

$$\Pi[\widehat{F}] \sim \mathcal{AN}\left(\Pi[F], n^{-1}(\nabla\Pi[F])\boldsymbol{\Sigma}_1(\nabla\Pi[F])'\right),$$

where $\nabla\Pi[F]$ is the gradient vector evaluated at the parameter vector $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}_1$ is the variance-covariance matrix of $\widehat{\boldsymbol{\theta}}_n$. In particular, for Pareto I$(\alpha, x_0 = 1)$ model,

$$\begin{aligned}
\nabla\Pi[F] &= \left.\frac{d\Pi[F]}{d\alpha}\right|_{\alpha} \\
&= \frac{cd^*}{\alpha - 1}\left(\frac{d^*}{u}\right)^{\alpha-1}\left[\frac{1}{1-\alpha} - \log\left(\frac{d^*}{u}\right)\right] - \frac{cd^*}{(\alpha-1)^2},
\end{aligned}$$

and $\boldsymbol{\Sigma}_1 = \boldsymbol{D}_{\otimes}\boldsymbol{\Sigma}_{\otimes}\boldsymbol{D}'_{\otimes}$ with $\boldsymbol{D}_{\otimes}$ and $\boldsymbol{\Sigma}_{\otimes}$ as given by Example 4.9. Similarly for Lognormal(θ, σ) model,

$$\nabla\Pi[F] = \left.\left(\frac{\partial\Pi[F]}{\partial\theta}, \frac{\partial\Pi[F]}{\partial\sigma}\right)\right|_{(\theta,\sigma)},$$

with

$$\begin{aligned}
\frac{\partial\Pi[F]}{\partial\theta} &= cK_{0,t}^{-2}\left[\left\{e^{\theta+\frac{\sigma^2}{2}}\left(H(\theta,\sigma) + \frac{\partial H(\theta,\sigma)}{\partial\theta}\right) + u\frac{\partial K_{0,T}}{\partial\theta} - d^*\frac{\partial K_{0,t}}{\partial\theta}\right\}K_{0,t}\right. \\
&\qquad \left. - \left(e^{\theta+\frac{\sigma^2}{2}}H(\theta,\sigma) + uK_{0,T} - d^*K_{0,t}\right)\frac{\partial K_{0,t}}{\partial\theta}\right], \\
\frac{\partial\Pi[F]}{\partial\sigma} &= cK_{0,t}^{-2}\left[\left\{e^{\theta+\frac{\sigma^2}{2}}\left(\sigma H(\theta,\sigma) + \frac{\partial H(\theta,\sigma)}{\partial\sigma}\right) + u\frac{\partial K_{0,T}}{\partial\sigma} - d^*\frac{\partial K_{0,t}}{\partial\sigma}\right\}K_{0,t}\right. \\
&\qquad \left. - \left(e^{\theta+\frac{\sigma^2}{2}}H(\theta,\sigma) + uK_{0,T} - d^*K_{0,t}\right)\frac{\partial K_{0,t}}{\partial\sigma}\right],
\end{aligned}$$

where

$$
\begin{cases}
T_z & := \frac{\log(u)-\theta}{\sigma}, \quad t_z := \frac{\log(d^*)-\theta}{\sigma}, \\
K_{0,T} & := 1-\Phi(T_z), \quad K_{0,t} := 1-\Phi(t_z), \\
H(\theta,\sigma) & := \Phi(T_z-\sigma)-\Phi(t_z-\sigma), \\
\frac{\partial H(\theta,\sigma)}{\partial\theta} & = \frac{1}{\sigma}\left[\phi(t_z-\sigma)-\phi(T_z-\sigma)\right], \\
\frac{\partial H(\theta,\sigma)}{\partial\sigma} & = \phi(t_z-\sigma)\left(\frac{t_z}{\sigma}+1\right)-\phi(T_z-\sigma)\left(\frac{T_z}{\sigma}+1\right), \\
\frac{\partial K_{0,t}}{\partial\theta} & = \frac{1}{\sigma}\phi(t_z), \quad \frac{\partial K_{0,t}}{\partial\sigma} = \frac{t_z}{\sigma}\phi(t_z), \\
\frac{\partial K_{0,T}}{\partial\theta} & = \frac{1}{\sigma}\phi(T_z), \quad \frac{\partial K_{0,T}}{\partial\sigma} = \frac{T_z}{\sigma}\phi(T_z),
\end{cases}
$$

and $\boldsymbol{\Sigma}_1 = \boldsymbol{D}_\otimes \boldsymbol{\Sigma}_\otimes \boldsymbol{D}'_\otimes$ with $\boldsymbol{D}_\otimes$ and $\boldsymbol{\Sigma}_\otimes$ as given by Example 4.19.

Discussion of Table 5.2 (continued): First, note that $1 < \widehat{\alpha} < 2$, which implies that the claims distribution is heavy-tailed (because for $\alpha < 2$, the variance of Pareto I is infinite). Second, as suggested by model validation, actuarial premiums based on Pareto I models exceed the empirical premium, and those based on seemingly inappropriate lognormal models are significantly below. Although the latter observation may not be accurate as the MLE, MTM2, and MTM5 based fits look good. Third, it is quite surprising and counter-intuitive to see that the premium estimates change less for MLE than MTM estimators, when the distributional assumption is changed. Further, for Pareto I model, CIs via MTM are very close to CI via MLE. For lognormal model, CIs via MTM are mush shorter than the corresponding CI via MLE. Finally, the main advantage of parametric procedures (both MTM and MLE) over the empirical approach is that, in general, all the parametric intervals are shorter than the empirical one. It is also evident from the model validation and premium calculation that the MTM approach with appropriate trimming proportions lead to premium point estimate which are closer to empirical counter part than those with over-cutting (i.e., MTM4 and MTM7). □

Table 5.2: Goodness-of-fit measurements, parameter estimates, actuarial premiums and 95% confidence intervals of the premium for the layer $[1,500,000; 14,000,000]$ with co-insurance factor $c = 0.8$.

Estimators	Pareto I Model				Lognormal Model			
	x_0	$\widehat{\alpha}$	AP ($\times 10^6$)	CIs ($\times 10^6$)	$\widehat{\theta}$	$\widehat{\sigma}$	APs ($\times 10^6$)	CIs ($\times 10^6$)
MLE = MTM1($a = .00, b = .00$)	1	1.18	2.21	$(1.95, 2.47)$	6.04	2.71	1.94	$(1.05, 2.84)$
MTM2($a = .00, b = .05$)	1	1.16	2.26	$(1.98, 2.54)$	8.02	2.37	1.85	$(1.36, 2.35)$
MTM3($a = .00, b = .10$)	1	1.15	2.27	$(1.98, 2.56)$	10.74	1.77	1.61	$(1.37, 1.85)$
MTM4($a = .00, b = .25$)	1	1.12	2.36	$(2.02, 2.69)$	12.63	1.17	1.13	$(1.10, 1.16)$
MTM5($a = .05, b = .05$)	1	1.16	2.27	$(1.98, 2.53)$	8.12	2.35	1.85	$(1.24, 2.46)$
MTM6($a = .10, b = .10$)	1	1.15	2.26	$(1.98, 2.56)$	10.78	1.76	1.60	$(1.29, 1.91)$
MTM7($a = .25, b = .25$)	1	1.13	2.32	$(1.99, 2.65)$	12.89	1.05	1.02	$(0.61, 1.42)$
Empirical	-	-	2.20	$(1.54, 2.89)$	-	-	2.20	$(1.54, 2.89)$

Goodness-of-fit Measures	KS		AD		KS		AD	
	TS	p-val	TS	p-val	TS	p-val	TS	p-val
MLE = MTM1($a = .00, b = .00$)	0.062	0.02	1.822	0.12	0.044	0.02	1.242	0.03
MTM2($a = .00, b = .05$)	0.057	0.02	1.681	0.08	0.041	0.01	1.209	0.01
MTM3($a = .00, b = .10$)	0.056	0.02	1.658	0.16	0.035	0.01	1.323	0.01
MTM4($a = .00, b = .25$)	0.048	0.05	1.700	0.05	0.053	0.01	3.246	0.01
MTM5($a = .05, b = .05$)	0.057	0.02	1.683	0.08	0.041	0.01	1.208	0.01
MTM6($a = .10, b = .10$)	0.056	0.02	1.665	0.16	0.035	0.01	1.334	0.01
MTM7($a = .25, b = .25$)	0.051	0.03	1.646	0.02	0.060	0.02	4.367	0.00

KS = Kolmogorov-Smirnov; AD = Anderson-Darling; TS = Test Statistics; p-val = p-values; AP = Actuarial Premiums; CIs = Confidence Intervals. The p-values are computed using parametric bootstrap (see, e.g., Wasserman, 2006, Chapter 3) with 1000 simulation runs.

Chapter 6

Methodological Extensions

In this chapter, we use the ideas of loss data scenarios of Chapter 2 to construct novel estimation procedures for complete data given by (2.1). We develop and study asymptotic properties of the newly proposed estimators. Several connections between data truncation, trimming, and censoring will also be established.

6.1 Method of Truncated Moments

Instead of trimming fixed proportion from both tails as investigated by Brazauskas *et al.* (2009), in this approach of parametric estimation we truncate the data from below at lower threshold and from above at upper threshold and then apply the method of moments on the remaining data. We call such an approach method-of-truncated-moments (MTuM).

6.1.1 Definition

Let $X_1, X_2, ..., X_n$ be *i.i.d.* random variables with common cdf $F(\cdot|\boldsymbol{\theta})$. The truncated moments estimators of $\theta_1, \theta_2, ..., \theta_k$ are computed according to the following procedure.

- The sample truncated moments are computed as

$$\widehat{\mu}_j = \frac{\sum_{i=1}^{n} h_j(X_i)\mathbb{1}\{t_j < X_i \leq T_j\}}{\sum_{i=1}^{n} \mathbb{1}\{t_j < X_i \leq T_j\}}, \qquad 1 \leq j \leq k. \tag{6.1}$$

The $h_j's$ in (6.1) are specially chosen functions as well as the thresholds t_j and

T_j are chosen by the researcher. In general, it is reasonable to assume that $X_{1:n} \leq t_j < T_j \leq X_{n:n}$, for all $1 \leq j \leq k$.

- Derive the corresponding population truncated moments as

$$\begin{aligned}\mu_j(\theta_1, \theta_2, ..., \theta_k) &= \mathbb{E}\left[h_j(X)|t_j < X \leq T_j\right] \\ &= \frac{\mathbb{E}\left[h_j(X)\mathbb{1}\{t_j < X \leq T_j\}\right]}{\mathbb{P}(t_j < X \leq T_j)} \\ &= \frac{\int_{t_j}^{T_j} h_j(x) f(x|\boldsymbol{\theta})\, dx}{F(T_j|\boldsymbol{\theta}) - F(t_j|\boldsymbol{\theta})}, \quad 1 \leq j \leq k. \end{aligned} \tag{6.2}$$

- Now, match the sample and population truncated moments from (6.1) and (6.2) to get the following system of equations for $\theta_1, \theta_2, ..., \theta_k$:

$$\begin{cases} \mu_1(\theta_1, \ldots, \theta_k) & = & \widehat{\mu}_1 \\ & \vdots & \\ \mu_k(\theta_1, \ldots, \theta_k) & = & \widehat{\mu}_k \end{cases} \tag{6.3}$$

Definition 6.1. A solution to the system of equations (6.3), say $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_k)$, if it exists, is called the *method of truncated moments (MTuM)* estimator of $\boldsymbol{\theta}$. Thus, $\widehat{\theta}_j =: g_j(\widehat{\mu}_1, \widehat{\mu}_2, ..., \widehat{\mu}_k)$, $1 \leq j \leq k$ are the MTuM estimators of $\theta_1, \theta_2, ..., \theta_k$.

6.1.2 Asymptotic Properties

For $1 \leq j, j' \leq k$ and for any positive integer n, define $\mathbb{1}\{t_{jj'} < X \leq T_{jj'}\} := \mathbb{1}\{t_j < X \leq T_j\}\mathbb{1}\{t_{j'} < X \leq T_{j'}\}$ and consider the following additional notations:

$$\begin{aligned} &Z_j := h_j(X), \quad h_{jj'}(x) := h_j(x)h_{j'}(x), && p_j := F(T_j|\boldsymbol{\theta}) - F(t_j|\boldsymbol{\theta}), \\ &Y_{jj'} := Y_j Y_{j'}, \qquad Y_j := Z_j \mathbb{1}\{t_j < X \leq T_j\}, && p_{jj'} := F(T_{jj'}|\boldsymbol{\theta}) - F(t_{jj'}|\boldsymbol{\theta}), \\ &r_j := h_j(t_j), \qquad W_{jj'} := Z_j \mathbb{1}\{t_{jj'} < X \leq T_{jj'}\}, && p_{j,n} := F_n(T_j) - F_n(t_j), \\ &R_j := h_j(T_j), \end{aligned}$$

where $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}$ is the empirical distribution function. Note that $Y_{jj'} = Y_{j'j}$ but $W_{jj'} \neq W_{j'j}$ for $j \neq j'$, in general. With those notations, the density of $Y_j (1 \leq j \leq k)$ can be expressed as

$$f_{Y_j}(x) = \begin{cases} 1 - F_{Z_j}(R_j|\boldsymbol{\theta}) + F_{Z_j}(r_j|\boldsymbol{\theta}), & x = 0; \\ f_{Z_j}(x|\boldsymbol{\theta}), & r_j < x < R_j; \\ 0, & \text{otherwise.} \end{cases}$$

The density of the random variables $Y_{jj'} = Y_{j'j}$ and $W_{jj'}$ can be constructed with the four possible scenarios which are listed in Appendix A. To establish the asymptotic distribution of $\widehat{\boldsymbol{\mu}}$, we need the following lemma (its proof can be found in Serfling, 1980).

Lemma 6.1. *For* $1 \leq j, j' \leq k$,

$$\mathbb{C}ov(Y_j, Y_{j'}) = \mu_{Y_{jj'}} - \mu_{Y_j}\mu_{Y_{j'}},$$

$$\mathbb{C}ov(Y_j; p_{j',1}) = \mu_{W_{jj'}} - \mu_{Y_j} p_{j'},$$

$$\mathbb{C}ov(p_{j,1}; p_{j',1}) = p_{jj'} - p_j p_{j'}.$$

Consider a $2k$ - dimensional random vector $\boldsymbol{V} := (Y_1, \ldots, Y_k, p_{1,1}, \ldots, p_{k,1})$. Clearly the mean vector of $\boldsymbol{V}$ is $\boldsymbol{\mu_V} = (\mu_{Y_1}, \ldots, \mu_{Y_k}, p_1, \ldots, p_k)$ and with Lemma 6.1, the variance-covariance matrix is $\boldsymbol{\Sigma_V} = \left[\sigma^2_{\boldsymbol{V},jj'}\right]_{j,j'=1}^{2k}$, where

$$\sigma^2_{\boldsymbol{V},jj'} = \begin{cases} \mu_{Y_{jj'}} - \mu_{Y_j}\mu_{Y_{j'}}, & 1 \leq j, j' \leq k; \\ \mu_{W_{j(j'-k)}} - \mu_{Y_j} p_{j'-k}, & 1 \leq j \leq k; k+1 \leq j' \leq 2k; \\ \mu_{W_{(j-k)j'}} - \mu_{Y_{j'}} p_{j-k}, & 1 \leq j' \leq k; k+1 \leq j \leq 2k; \\ p_{(j-k)(j'-k)} - p_{j-k} p_{j'-k}, & k+1 \leq j, j' \leq 2k. \end{cases}$$

Theorem 6.1. *The empirical estimator*

$$\begin{aligned} \widehat{\boldsymbol{\mu}}_V &:= \frac{1}{n}\left(\sum_{i=1}^n Y_{1,i}, \ldots, \sum_{i=1}^n Y_{k,i}, \sum_{i=1}^n p_{1,i}, \ldots, \sum_{i=1}^n p_{k,i}\right) \\ &= \left(\overline{Y}_{1,n}, \ldots, \overline{Y}_{k,n}, p_{1,n}, \ldots, p_{k,n}\right) \end{aligned}$$

of the mean vector $\boldsymbol{\mu}_V$ *has the following asymptotic distribution:*

$$\widehat{\boldsymbol{\mu}}_V \sim \mathcal{AN}\left(\boldsymbol{\mu_V}, \frac{1}{n}\boldsymbol{\Sigma_V}\right).$$

Proof. Let $\{\boldsymbol{V}_n\}$ be a sequence of *i.i.d.* $\boldsymbol{V}$ random vectors, then by multivariate Central Limit Theorem (see, e.g., Serfling, 1980), we have:

$$\left(\overline{Y}_{1,n}, \ldots, \overline{Y}_{k,n}, p_{1,n}, \ldots, p_{k,n}\right) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{V_i} \sim \mathcal{AN}\left(\boldsymbol{\mu_V}, \frac{1}{n}\boldsymbol{\Sigma_V}\right). \qquad \square$$

The system of MTuM equations (6.3) can be written as:

$$\begin{cases} \mu_1(\theta_1, \ldots, \theta_k) & = & \widehat{\mu}_1 & = & \frac{\overline{Y}_{1,n}}{p_{1,n}} \\ & \vdots & & \vdots & \\ \mu_k(\theta_1, \ldots, \theta_k) & = & \widehat{\mu}_k & = & \frac{\overline{Y}_{k,n}}{p_{k,n}} \end{cases} \tag{6.4}$$

Lemma 6.2. *Consider a function* $g_{\boldsymbol{V}} : \mathbb{R}^{2k} \to \mathbb{R}^k$ *for* $\boldsymbol{x} = (x_1, x_2, \ldots, x_{2k})$ *defined by*

$$g_{\boldsymbol{V}}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \ldots, g_k(\boldsymbol{x})) = \left(\frac{x_1}{x_{k+1}}, \ldots, \frac{x_k}{x_{2k}}\right),$$

where $x_i \neq 0,\ i = k+1, \ldots, 2k$. *Then* $g_{\boldsymbol{V}}$ *is totally differentiable at any point* $\boldsymbol{x}_0 \in \mathbb{R}^{2k}$.

Proof. A proof directly follows from Serfling (1980, Lemma 1.12.2). $\square$

With the help of Theorem 6.1 and Lemma 6.2, we are now ready to state the asymptotic distribution of the truncated sample moment vector $\widehat{\boldsymbol{\mu}}$.

Theorem 6.2. *The asymptotic joint distribution of the truncated sample moment vector* $(\widehat{\mu}_1, \ldots, \widehat{\mu}_k)$ *is given by* $N\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right)$ *with* $\boldsymbol{\Sigma} = \boldsymbol{D_V}\Sigma_{\boldsymbol{V}}D'_{\boldsymbol{V}} =: \left[\sigma^2_{jj'}\right]_{k\times k}$, *where*

$$\sigma^2_{jj'} = \frac{1}{p_{j'}}\left(\frac{\mu_{Y_{jj'}} - \mu_{Y_j}\mu_{Y_{j'}}}{p_j} - \frac{\mu_{Y_j}(\mu_{W_{jj'}} - \mu_{Y_{j'}}p_j)}{p_j^2}\right) - \frac{\mu_{Y_{j'}}}{p_{j'}^2}\left(\frac{\mu_{W_{jj'}} - \mu_{Y_j}p_{j'}}{p_j} - \frac{\mu_{Y_j}(p_{jj'} - p_jp_{j'})}{p_j^2}\right).$$

Proof. See Appendix A. $\square$

Now, with $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_k)$ and $g_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\mu}}) = (g_{1,\boldsymbol{\theta}}(\widehat{\boldsymbol{\mu}}), \ldots, g_{k,\boldsymbol{\theta}}(\widehat{\boldsymbol{\mu}})) = \widehat{\boldsymbol{\theta}}$, then again by the delta method, we have the following main result of this section.

Theorem 6.3. *The MTuM estimator of $\boldsymbol{\theta}$, denoted by $\widehat{\boldsymbol{\theta}}$, has the following asymptotic distribution:*

$$\widehat{\boldsymbol{\theta}} = \left(\widehat{\theta}_1, \ldots, \widehat{\theta}_k\right) \sim \mathcal{AN}\left(\boldsymbol{\theta}, \frac{1}{n}\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}'\right),$$

where the Jacobian $\boldsymbol{D}$ is given by (see Theorem 1.1) $\boldsymbol{D} = \left[\left.\frac{\partial g_{j,\boldsymbol{\theta}}}{\partial \widehat{\mu}_{j'}}\right|_{\widehat{\boldsymbol{\mu}}=\boldsymbol{\mu}}\right]_{k\times k} =: [d_{jj'}]_{k\times k}$ and the variance-covariance matrix $\boldsymbol{\Sigma}$ has the same form as in Theorem 6.2.

Note 6.1. In view of the above derivations, we notice that data trimming and thus MTM can be interpreted as special cases of data truncation and thus MTuM, respectively. To see that, let F be the distribution function of X. For $1 \leq j \leq k$, consider $F(t_j|\boldsymbol{\theta}) = a_j$ and $F(T_j|\boldsymbol{\theta}) = 1 - b_j$. Then, using integration by substitution with $U = F(X)$, the equation (6.2) becomes

$$\begin{aligned}
\mu_j(\theta_1, \theta_2, ..., \theta_k) &= \frac{\int_{t_j}^{T_j} h_j(x) f(x|\boldsymbol{\theta})\, dx}{F(T_j|\boldsymbol{\theta}) - F(t_j|\boldsymbol{\theta})} \\
&= \frac{\int_{F(t_j|\boldsymbol{\theta})}^{F(T_j|\boldsymbol{\theta})} h_j(F^{-1}(u|\boldsymbol{\theta}))\, du}{F(T_j|\boldsymbol{\theta}) - F(t_j|\boldsymbol{\theta})} \qquad (6.5a) \\
&= \frac{\int_{a_j}^{1-b_j} h_j(F^{-1}(u|\boldsymbol{\theta}))\, du}{1 - a_j - b_j}, \qquad (6.5b)
\end{aligned}$$

which is equivalent to (3.17). □

Note 6.2. For estimation purposes these two approaches (i.e., MTM and MTuM) are very different. With the MTuM approach, the limits of integration as well as the denominator in equation (6.5a) are unknowns, which creates technical complications when we want to assess the asymptotic properties of MTuM estimators. On the other hand, with the MTM approach, both the limits of integration and the denominator in equation (6.5b) are constants, which simplifies the matters significantly. Indeed, as is evident from complete data examples in Chapter 4 (as well as those presented by Brazauskas *et al.*, 2009 and Zhao *et al.*, 2018a), MTM leads to explicit formulas for all location-scale families and their variants, but that is not the case with MTuM. In view of this, we will consider the MTuM approach further only for some data scenarios, but not all. $\square$

6.2 Exponential and Pareto Models

In this section, we derive MTuM and related estimators for the parameter of exponential distribution for completely observed data. For this particular distribution, we also explore two additional methods: method of censored moments and insurance payment-type estimators. Several connections between different approaches are established. For insurance losses the equivalent (after the logarithmic transformation) model is Pareto. Thus, the estimators derived in this section can easily be adjusted for Pareto model. Their asymptotic properties will remain valid as well.

As in Section 4.1, let $X \sim Exp(\theta)$ with cdf $F(x|\theta) = 1 - e^{-\frac{x}{\theta}}$ and pdf $f(x|\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$, $x > 0$. Since there is a single parameter, θ, to be estimated, as in Section 4.1.2, we consider the function $h(x) = x$.

6.2.1 Method of Truncated Moments

Let $X_1, \ldots, X_n$ be *i.i.d.* random variables defined by (2.1). Consider t and T be the left and right truncation points, respectively. Then the sample truncated moment is given by

$$\begin{aligned}
\widehat{\mu} &= \frac{\sum_{i=1}^{n} X_i \mathbb{1}\{t < X_i \leq T\}}{\sum_{i=1}^{n} \mathbb{1}\{t < X_i \leq T\}} \\
&= \frac{\sum_{i=1}^{n} X_i \mathbb{1}\{t < X_i \leq T\}}{n} \frac{n}{\sum_{i=1}^{n} \mathbb{1}\{t < X_i \leq T\}} \\
&= \frac{\left(\sum_{i=1}^{n} Y_i\right)/n}{F_n(T) - F_n(t)} \\
&= \frac{\overline{Y}_n}{F_n(T) - F_n(t)} \\
&= \frac{\overline{Y}_n}{p_n},
\end{aligned}$$

where $Y_1, Y_2, \ldots, Y_n \overset{i.i.d.}{\sim} Y := X\mathbb{1}\{t < X \leq T\}$ and $p_n := F_n(T) - F_n(t)$ with $p(\theta) = F(T|\theta) - F(t|\theta) = e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}$.

Theorem 6.4. *The mean and the variance of the random variable Y are respectively given by*

$$\mu_Y = \theta\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right) + te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}},$$

$$\sigma_Y^2 = 2\theta^2\left(\Gamma\left(3; \frac{T}{\theta}\right) - \Gamma\left(3; \frac{t}{\theta}\right)\right) - \mu_Y^2,$$

where $\Gamma(\alpha; x)$ with $\alpha > 0, x > 0$ is the incomplete gamma function defined as

$$\Gamma(\alpha; x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t}\, dt,$$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt.$$

Proof. See Appendix A. □

From Theorem 6.2, $\widehat{\mu} \sim \mathcal{AN}\left(\frac{\mu_Y}{p}, \frac{1}{n}\left(\frac{\sigma_Y^2}{p^2} - \frac{(1-p)\mu_Y^2}{p^3}\right)\right)$. Note that the asymptotic variance of $\widehat{\mu}$ is exactly equal to the approximation through the second order Taylor series expansion of the ratio of the asymptotic distribution of $\overline{Y}_n$ and p_n as mentioned in Hayya *et al.* (1975).

The population version of $\widehat{\mu}$ is given by

$$\begin{aligned}
\mu_* &= \mathbb{E}[X|t < X \leq T] \\
&= \frac{\mathbb{E}[Y]}{F(T|\theta) - F(t|\theta)} \\
&= \frac{\theta\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right) + te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}}}{F(T|\theta) - F(t|\theta)} \\
&= \frac{e^{-\frac{t}{\theta}}(\theta + t) - e^{-\frac{T}{\theta}}(\theta + T)}{e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}} \\
&= \frac{\mu_Y}{e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}},
\end{aligned}$$

where $\mu_Y = e^{-\frac{t}{\theta}}(\theta + t) - e^{-\frac{T}{\theta}}(\theta + T)$.

Theorem 6.5. *The equation $\mu_* = \widehat{\mu}$ has a unique solution $\widehat{\theta}$ provided that $t < \widehat{\mu} < \frac{t+T}{2}$. Otherwise, the solution does not exist.*

Proof. It is clear that $t < \widehat{\mu} < T$. Also, $\mu_*(\theta) = \frac{e^{-\frac{t}{\theta}}(t+\theta) - e^{-\frac{T}{\theta}}(T+\theta)}{e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}}$. Then, in order to establish the result, it is enough to prove the following statements:

1. $\mu_*(\theta)$ is strictly increasing,

2. $\lim_{\theta \to 0+} \mu_*(\theta) = t$, and

3. $\lim_{\theta \to \infty} \mu_*(\theta) = \frac{t+T}{2}$.

First of all, let us establish that $\mu_*(\theta)$ is strictly increasing.

$$
\begin{aligned}
\mu_*(\theta) &= \frac{te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}}}{e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}} + \theta. \\
\therefore\ \mu_*'(\theta) &= \frac{\left(\left(\frac{t}{\theta}\right)^2 e^{-\frac{t}{\theta}} - \left(\frac{T}{\theta}\right)^2 e^{-\frac{T}{\theta}}\right)\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right) + \left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right)^2}{\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right)^2} \\
&= \frac{-e^{-\frac{t+T}{\theta}}\left(\left(\frac{T}{\theta}\right)^2 + \left(\frac{t}{\theta}\right)^2\right) + \frac{2tT}{\theta^2} e^{-\frac{t+T}{\theta}} + \left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right)^2}{\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right)^2} \\
&= \frac{-4e^{-\frac{t+T}{\theta}}(t^2+T^2) + 8tTe^{-\frac{t+T}{\theta}} + 4\theta^2\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right)^2}{4\theta^2\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right)^2} \\
&= \frac{e^{-\frac{t+T}{\theta}}\left[-4(t^2+T^2) + 8tT + 4\theta^2\left(e^{\frac{t-T}{2\theta}} - e^{-\frac{t-T}{2\theta}}\right)^2\right]}{e^{-\frac{t+T}{\theta}}4\theta^2\left(e^{\frac{t-T}{2\theta}} - e^{-\frac{t-T}{2\theta}}\right)^2} \\
&= 1 - \frac{4(t-T)^2}{4\theta^2\left(e^{\frac{t-T}{2\theta}} - e^{-\frac{t-T}{2\theta}}\right)^2} \\
&= 1 - \left(\frac{t-T}{2\theta}\right)^2\left(\frac{2}{e^{\frac{t-T}{2\theta}} - e^{-\frac{t-T}{2\theta}}}\right)^2 \\
&= 1 - \left(\frac{t-T}{2\theta}\right)^2 \operatorname{csch}^2\left(\frac{t-T}{2\theta}\right).
\end{aligned}
$$

Therefore, $\mu_*'(\theta) > 0$ if and only if $\left(\frac{t-T}{2\theta}\right)^2 < \sinh^2\left(\frac{t-T}{2\theta}\right)$, which is true since $x < \sinh x$ for all $x > 0$ and $x > \sinh x$ for all $x < 0$.

Further,

$$
\begin{aligned}
\lim_{\theta\to 0+} \mu_*(\theta) &= \lim_{\theta\to 0+} \left[\theta + \frac{te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}}}{e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}}\right] \\
&= \lim_{\theta\to 0+} \frac{e^{-\frac{t}{\theta}}\left(t - Te^{\frac{t-T}{\theta}}\right)}{e^{-\frac{t}{\theta}}\left(1 - e^{\frac{t-T}{\theta}}\right)} \\
&= t. \\
\lim_{\theta\to \infty} \mu_*(\theta) &= \lim_{\theta\to \infty} \left[\theta + \frac{te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}}}{e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}}\right] \\
&\overset{y:=\frac{1}{\theta}}{=} \lim_{y\to 0+} \left[\frac{1}{y} + \frac{te^{-ty} - Te^{-Ty}}{e^{-ty} - e^{-Ty}}\right] \\
&= \lim_{y\to 0+} \left[\frac{1}{y} + \frac{te^{(T-t)y} - T}{e^{(T-t)y} - 1}\right] \\
&= \lim_{y\to 0+} \left[\frac{e^{(T-t)y} - 1 + tye^{(T-t)y} - Ty}{y\left(e^{(T-t)y} - 1\right)}\right] \\
&= \lim_{y\to 0+} \left[\frac{e^{(T-t)y} - 1 - ty - (T-t)y + tye^{(T-t)y}}{y\left(e^{(T-t)y} - 1\right)}\right] \\
&= \lim_{y\to 0+} \left[\frac{ty\left(e^{(T-t)y} - 1\right)}{y\left(e^{(T-t)y} - 1\right)} - \frac{(T-t)y - e^{(T-t)y} + 1}{y\left(e^{(T-t)y} - 1\right)}\right] \\
&= t - \lim_{y\to 0+} \frac{(T-t)y - e^{(T-t)y} + 1}{y\left(e^{(T-t)y} - 1\right)} \\
&= t - \lim_{y\to 0+} \frac{(T-t) - (T-t)e^{(T-t)y}}{e^{(T-t)y} - 1 + y(T-t)e^{(T-t)y}} \\
&= t - \lim_{y\to 0+} \frac{-(T-t)^2 e^{(T-t)y}}{(T-t)e^{(T-t)y} + y(T-t)e^{(T-t)y} + (T-t)e^{(T-t)y}} \\
&= t + \frac{(T-t)^2}{T-t+T-t} \\
&= \frac{t+T}{2}. \qquad \square
\end{aligned}
$$

Let $\theta' := \frac{d\theta}{d\mu_*}$, then

$$\mu_* = \frac{\mu_Y}{e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}}$$

$$\Rightarrow \quad \mu_* \left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right) = \mu_Y$$

$$\Rightarrow \quad \theta' = \frac{p\theta^2}{te^{-\frac{t}{\theta}}(\theta + t) - Te^{-\frac{T}{\theta}}(\theta + T) + p\theta^2 - \mu_* \left(te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}}\right)}.$$

Therefore, by the delta method (see Theorem 1.1), we have that

$$\widehat{\theta} \sim \mathcal{AN}\left(\theta, (\theta')^2 \left(\frac{\sigma_Y^2}{np^2} - \frac{(1-p)\mu_Y^2}{np^3}\right)\right),$$

and hence from Example 4.1, we have

$$ARE(\widehat{\theta}_{MTuM}, \widehat{\theta}_{MLE}) = \frac{\theta^2 p^3}{(\theta')^2 \left(\sigma_Y^2 p - (1-p)\mu_Y^2\right)}.$$

Table 6.1 provides numerical illustrations of ARE computation.

Table 6.1: $ARE(\widehat{\theta}_{MTuM}, \widehat{\theta}_{MLE})$ for selected t and T for $Exp(\theta = 10)$.

	$T_{(1\,-\,F(T\mid\theta))}$							
$t_{(F(t\mid\theta))}$	∞(.00)	29.96(.05)	23.03(.10)	18.97(.15)	13.86(.25)	7.13(.49)	3.57(.70)	1.63 (.85)
0(.00)	1	.478	.311	.216	.109	.021	.003	.000
0.51(.05)	.950	.443	.284	.193	.095	.016	.002	.000
1.05(.10)	.900	.408	.257	.172	.082	.012	.001	.000
1.63(.15)	.850	.373	.231	.152	.069	.009	.000	-
2.88(.25)	.750	.307	.182	.114	.047	.004	.000	-
6.73(.49)	.510	.161	.080	.042	.011	.000	-	-
12.04(.70)	.300	.057	.019	.006	.000	-	-	-
18.97(.85)	.150	.009	.001	-	-	-	-	-

Discussion of Table 6.1: The truncation thresholds t and T are rounded to two decimal places; for example, $0.51 \approx F^{-1}(0.05)$, $18.97 \approx F^{-1}(0.85)$, etc. The entries are all smaller than the corresponding entries for MTM. For example, if the lower and upper truncation thresholds are, respectively, $t = F^{-1}(0.05)$ and $T = F^{-1}(0.95)$ then $ARE(\widehat{\theta}_{MTuM}, \widehat{\theta}_{MLE}) = 0.443$ but with similar trimming proportion (i.e., $a = 0.05 = b$), $ARE(\widehat{\theta}_{MTM}, \widehat{\theta}_{MLE}) = 0.918$. That is, we loose approximately 52% efficiency by going from MTM to MTuM. The reason that MTuM relative efficiency is much lower than the corresponding MTM is that the trimmed sample size is always constant given that the trimming proportions are fixed. On the other hand, even if we fix the truncation thresholds, the truncated sample size is random. $\square$

Note that if $X \sim$ Pareto I (α, x_0) with x_0 known then $Y := \log\left(\frac{X}{x_0}\right) \sim \mathrm{Exp}\left(\frac{1}{\alpha} =: \theta\right)$. So, estimators of α of the single-parameter Pareto distribution will share the same AREs with estimators of $Exp(\theta)$, given that $h(x) = \log\left(\frac{x}{x_0}\right)$. The following result for single-parameter Pareto has been partially derived by Clark (2013), but can easily be extended using the tools of this section.

Theorem 6.6. *Let t and T be the left and right truncation point, respectively, for $X \sim$ Pareto I(α, x_0). Also, define $A_{tT} := T^{\alpha}\left(1 - \alpha \log\left(\frac{x_0}{t}\right)\right) - t^{\alpha}\left(1 - \alpha \log\left(\frac{x_0}{T}\right)\right)$ and $g_{tT}(\alpha) := \frac{A_{tT}}{\alpha(T^{\alpha} - t^{\alpha})}$. Then the equation $\widehat{\mu} = \mu_*$ has a unique solution provided that $\lim_{\alpha \to \infty} g_{tT}(\alpha) < \widehat{\mu} < \lim_{\alpha \to 0+} g_{tT}(\alpha)$.*

Proof. See Appendix A. $\square$

6.2.2 Method of Fixed Censored Moments

Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} \mathrm{Exp}(\theta)$ random variables. Also, let t and T be the left- and right-hand side censored points (see Section 2.3), respectively. The sample censored mean is given by

$$\widehat{\mu} = \frac{t \sum_{i=1}^{n} \mathbb{1}\{X_i \leq t\} + \sum_{i=1}^{n} X_i \mathbb{1}\{t < X_i \leq T\} + T \sum_{i=1}^{n} \mathbb{1}\{X_i > T\}}{n}.$$

Define $X^{**} = t\mathbb{1}\{X \leq t\} + X\mathbb{1}\{t < X \leq T\} + T\mathbb{1}\{X > T\}$ as in Section 2.3. Then the corresponding population censored moments are:

$$\mu_{**} = \mathbb{E}[X^{**}] = t(1 - e^{-\frac{t}{\theta}}) + \mu_Y + Te^{-\frac{T}{\theta}},$$

$$\mu_{(X^{**})^2} = \mathbb{E}[(X^{**})^2] = t^2(1 - e^{-\frac{t}{\theta}}) + \mathbb{E}[Y^2] + T^2 e^{-\frac{T}{\theta}}.$$

Thus, $\sigma^2_{X^{**}} = \mu_{(X^{**})^2} - \mu^2_{**}$. Moreover, setting $\mu_{**} = \widehat{\mu}$ implies

$$t + \theta\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right) = \widehat{\mu}.$$

Theorem 6.7. *The equation $\widehat{\mu} = \mu_{**}$ has a unique solution $\widehat{\theta}$ provided that $t < \widehat{\mu} < T$. Otherwise, the solution does not exist.*

Proof. Assume $t < \widehat{\mu} < T$. Let $\mu_{**}(\theta) = t + \theta\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right)$. Now, in order to have unique solution it is enough to show that $\mu_{**}(\theta)$ is strictly increasing with $\lim_{\theta\to 0+}\mu_{**}(\theta) = t$ and $\lim_{\theta\to\infty}\mu_{**}(\theta) = T$. Clearly,

$$\mu_{**}^{'}(\theta) = \frac{1}{\theta}\left(te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}}\right) + e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}.$$

So,

$$\begin{aligned}
& \mu_{**}^{'}(\theta) > 0 \\
\Longleftrightarrow\quad & e^{-\frac{T}{\theta}}(T+\theta) < e^{-\frac{t}{\theta}}(t+\theta) \\
\Longleftrightarrow\quad & \frac{T+\theta}{t+\theta} < e^{\frac{T-t}{\theta}} \\
\Longleftrightarrow\quad & T+\theta < (t+\theta)\left(1 + \frac{T-t}{\theta} + \frac{\left(\frac{T-t}{\theta}\right)^2}{2!} + \cdots\right) \\
\Longleftrightarrow\quad & T+\theta < T+\theta + \text{a positive term.}
\end{aligned}$$

Thus, $\mu_{**}^{'}(\theta) > 0$ and hence $\mu_{**}(\theta)$ is strictly increasing. Moreover,

$$\begin{aligned}
\lim_{\theta\to 0+}\mu_{**}(\theta) &= t. \\
\lim_{\theta\to\infty}\mu_{**}(\theta) &= \lim_{\theta\to\infty}\frac{e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}}{\frac{1}{\theta}} + t \\
&= T - t + t = T.
\end{aligned}$$

□

Further,

$$\begin{aligned}
& \mu_{**} = \mathbb{E}[X^{**}] \\
\Rightarrow\quad & \mu_{**} = t - te^{-\frac{t}{\theta}} + \theta\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right) + te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}} + Te^{-\frac{T}{\theta}} \\
\Rightarrow\quad & \mu_{**} = t + \theta\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right)
\end{aligned}$$

$$\Rightarrow \quad \theta' = \frac{\theta}{p\theta + te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}}}.$$

Theorem 6.8. *For method of censored moments (MCM) with $a = F(t|\theta)$ and $b = 1 - F(T|\theta)$, then the following result holds:*

$$ARE(\widehat{\theta}_{MCM}, \widehat{\theta}_{MLE}) = ARE(\widehat{\theta}_{MTM}, \widehat{\theta}_{MLE}).$$

Proof. From Example 4.6, we know that

$$\widehat{\theta}_{MTM} \sim \mathcal{AN}\left(\theta, \frac{\theta^2}{n}\Delta\right), \quad \text{with} \quad \Delta = \frac{J(a, 1-b)}{[I(a, 1-b)]^2},$$

where

$$J(a, 1-b) = \int_a^{1-b} \int_a^{1-b} \frac{\min\{u, v\} - uv}{(1-u)(1-v)}\, dv\, du.$$

Therefore,

$$ARE(\widehat{\theta}_{MTM}, \widehat{\theta}_{MLE}) = \frac{1}{\Delta} = \frac{[I(a, 1-b)]^2}{J(a, 1-b)}.$$

On the other hand,

$$ARE(\widehat{\theta}_{MCM}, \widehat{\theta}_{MLE}) = \frac{(p\theta + te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}})^2}{\sigma^2_{X^{**}}}.$$

So, we need to show that,

$$\begin{aligned} ARE(\widehat{\theta}_{MCM}, \widehat{\theta}_{MLE}) &= ARE(\widehat{\theta}_{MTM}, \widehat{\theta}_{MLE}) \\ &= \frac{\theta^2 [I(a, 1-b)]^2}{\sigma^2_{X^{**}}}. \end{aligned}$$

That is, $J(a, 1-b) = \frac{\sigma^2_{X^{**}}}{\theta^2}$. For that, we have:

$$\begin{aligned} \sigma^2_{X^{**}} &= t^2(1 - e^{-\frac{t}{\theta}}) + 2\theta^2 \left[\Gamma\left(3; \frac{t}{\theta}\right) - \Gamma\left(3; \frac{T}{\theta}\right)\right] + T^2 e^{-\frac{T}{\theta}} - t^2 - 2t\theta p - \theta^2 p^2 \\ &= 2\theta^2(1 - a - b) + 2\theta(-\theta(1-a)\log(1-a) + \theta b \log(b)) \\ &\quad - \theta(1 - a - b)(-2\theta \log(1-a) + \theta(1 - a - b)). \\ \frac{\sigma^2_{X^{**}}}{\theta^2} &= 2(1 - a - b) + 2(b \log(b) - (1-a)\log(1-a)) \\ &\quad - (1 - a - b)(1 - a - b - 2\log(1-a)) \end{aligned}$$

$$
\begin{aligned}
&= 2(1-a-b) + 2b\log(b) - 2(1-a)\log(1-a) \\
&\quad -(1-a-b) + 2\log(1-a) + 2\log(1-a) \\
&\quad + a(1-a-b) - 2a\log(1-a) \\
&\quad + b(1-a-b) - 2b\log(1-a) \\
&= (1-a-b) + 2b\log(b) - 2(1-a)\log(1-a) \\
&\quad + 2(1-a-b)\log(1-a) + a(1-a-b) + b(1-a-b) \\
&= a(1-a-b) + (1-a-b)\log(1-a) \\
&\quad + (1-a-b) - 2(1-a)\log(1-a) + (1-a-b)\log(1-a) + b\log(b) \\
&\quad + b(1-a-b) + b\log(b) \\
&= a(1-a-b) + (1-a-b)\log(1-a) \\
&\quad + (1-a-b) - 2(1-a)\log(1-a) + (1-a)\log(1-a) - b\log(1-a) \\
&\quad + b\log(b) + b(1-a-b) + b\log(b) \\
&= a(1-a-b) + (1-a-b)\log(1-a) \\
&\quad - \left[-(1-a-b) + (1-a)\log(1-a) - b\log(b)\right] \\
&\quad - b\left[(a-1+b) + \log(1-a) - \log(b)\right] \\
&= (1-a-b)\left[a + \log(1-a)\right] - I(a, 1-b) \\
&\quad + (1-b-1)\left[a - 1 + b + \log\left(\frac{1-a}{b}\right)\right] \\
&= (1-a-b)\left[a + \log(1-a)\right] - I(a, 1-b) + (1-b-1)I_1(a, 1-b) \\
&= J(a, 1-b).
\end{aligned}
$$

where $I_1(a, 1-b) := \int_a^{1-b} \frac{u}{1-u}\, du = (a-1+b) + \log\left(\frac{1-a}{b}\right)$. □

Table 6.2 provides numerical illustrations of ARE computation.

Table 6.2: $ARE(\widehat{\theta}_{MCM}, \widehat{\theta}_{MLE})$ for selected t and T for $Exp(\theta = 10)$.

$t_{(F(t\|\theta))}$	$T_{(1-F(T\|\theta))}$							
	∞(.00)	29.96(.05)	23.03(.10)	18.97(.15)	13.86(.25)	7.13(.49)	3.57(.70)	1.63 (.85)
0(.00)	1	.918	.847	.783	.666	.423	.238	.116
0.51(.05)	1	.918	.884	.783	.667	.425	.242	.122
1.05(.10)	1	.918	.848	.785	.669	.430	.250	.135
1.63(.15)	.999	.918	.850	.787	.672	.437	.260	-
2.88(.25)	.995	.918	.851	.790	.679	.452	.284	-
6.73(.49)	.958	.897	.839	.786	.688	.487	-	-
12.04(.70)	.857	.824	.781	.738	.659	-	-	-
18.97(.85)	.681	.688	.663	-	-	-	-	-

Discussion of Table 6.2: The truncation thresholds t and T are rounded to two decimal places; for example, $0.51 \approx F^{-1}(0.05)$, $18.97 \approx F^{-1}(0.85)$, etc. Due to Theorem 6.8, this table is identical to $ARE(\widehat{\theta}_{MTM}, \widehat{\theta}_{MLE})$ table which can be found in Brazauskas *et al.* (2009). $\square$

6.2.3 Insurance Payment Estimators

Let $X_1, \ldots, X_n$ be *i.i.d.* random variables with common exponential cdf $F(\cdot|\theta)$. Define the left- truncated (at t) and right-censored (at T) sample moment

$$\widehat{\mu} = \frac{\sum_{i=1}^n X_i \mathbb{1}\{t < X_i \le T\} + T\sum_{i=1}^n \mathbb{1}\{X_i > T\}}{\sum_{i=1}^n \mathbb{1}\{X_i > t\}}$$

$$= \frac{\overline{X^{\otimes}}_n}{p_n}.$$

where $X^{\otimes} = X\mathbb{1}\{t < X \le T\} + T\mathbb{1}\{X > T\}$, $p_n = 1 - F_n(t)$, and $p = 1 - F(t|\theta)$. The covariance of $X^{\otimes}$ and p_1 is given as

$$\sigma^2_{X^{\otimes}p_1} = \mathbb{C}ov(X^{\otimes}, p_1) = \mu_{X^{\otimes}}(1-p),$$

with

$$\mathbb{E}[X^{\otimes}] = \mu_Y + T(1 - F(T|\theta)),$$

$$\mathbb{E}[X^{\otimes 2}] = \mathbb{E}[Y^2] + T^2(1 - F(T|\theta)).$$

Then, by multivariate Central Limit Theorem, we have

$$\left(\overline{X^{\otimes}}_n, p_n\right) \sim \mathcal{AN}\left((\mu_{X^{\otimes}}, p), \frac{1}{n}\begin{bmatrix} \sigma^2_{X^{\otimes}} & \sigma^2_{X^{\otimes}p_1} \\ \sigma^2_{X^{\otimes}p_1} & p(1-p)\end{bmatrix}\right).$$

Then, by the delta method (Theorem 1.1) with a function $g(x_1, x_2) = \frac{x_1}{x_2}$, we have

$$\widehat{\mu} = \frac{\overline{X^{\otimes}}_n}{p_n} \sim \mathcal{AN}\left(\frac{\mu_{X^{\otimes}}}{p}, \frac{1}{n}\left(\frac{\sigma^2_{X^{\otimes}}}{p^2} - \frac{\mu^2_{X^{\otimes}}(1-p)}{p^3}\right)\right).$$

The population version of $\widehat{\mu}$ is given by

$$\mu_{\otimes} = \frac{\mathbb{E}[X^{\otimes}]}{1 - F(t|\theta)}$$

$$\Rightarrow \quad \mu_{\otimes} = \frac{\theta\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right) + te^{-\frac{t}{\theta}}}{e^{-\frac{t}{\theta}}}$$

$$\Rightarrow \quad \theta' = \frac{p\theta^2}{\theta^2\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right) + \theta\left(te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}}\right) + t^2p - tp\mu_{\otimes}}.$$

Therefore, again by the delta method, we get:

$$ARE(\widehat{\theta}_{MTCM}, \widehat{\theta}_{MLE}) = \frac{\theta^2 p^3}{(\theta')^2\left(p\sigma^2_{X^{\otimes}} - (1-p)\mu^2_{X^{\otimes}}\right)}.$$

Table 6.3 provides numerical illustrations of ARE computation.

Table 6.3: $ARE(\widehat{\theta}_{MTCM}, \widehat{\theta}_{MLE})$ for selected t and T for $Exp(\theta = 10)$.

	$T_{(1-F(T\|\theta))}$							
$t_{(F(t\|\theta))}$	∞(.00)	29.96(.05)	23.03(.10)	18.97(.15)	13.86(.25)	7.13(.49)	3.57(.70)	1.63 (.85)
0(.00)	1	.918	.847	.783	.666	.423	.238	.116
0.51(.05)	.950	.868	.798	.753	.619	.379	.197	.076
1.05(.10)	.900	.818	.749	.686	.572	.336	.156	.038
1.63(.15)	.850	.769	.700	.638	.525	.293	.116	-
2.88(.25)	.750	.670	.603	.542	.433	.208	.038	-
6.73(.49)	.510	.433	.371	.315	.216	.015	-	-
12.04(.70)	.300	.229	.173	.124	.039	-	-	-
18.97(.85)	.150	.087	.040	-	-	-	-	-

Discussion of Table 6.3: The truncation thresholds t and T are rounded to two decimal places, for example, $0.51 \approx F^{-1}(0.05)$, $18.97 \approx F^{-1}(0.85)$, etc. Comparing the corresponding entries among Tables 6.1, 6.2, and 6.3; the entries in 6.1 are the lowest and in 6.2 are the highest. The reason that the entries in Table 6.1 are the highest is that the observations beyond the truncation thresholds are disregarded in order to control the influence of extremes in the statistical inference. The MCM controls such influence of extremes differently, i.e., those observations which are beyond the thresholds are adjusted to be equal to the corresponding thresholds and hence increase the efficiency significantly (Table 6.2). MTCM controls the influence of extremes by disregarding the observations below lower threshold and adjusting the observations above upper threshold to be equal to the upper threshold which makes the entries in Table 6.3 in between Table 6.1 and Table 6.2. ∎

Theorem 6.9. *The equation $\widehat{\mu} = \mu_{\otimes}$ has a unique solution $\widehat{\theta}$ provided that $t < \widehat{\mu} < T$. Otherwise, the solution does not exist.*

Proof. Assume $t < \widehat{\mu} < T$. Again, let $\mu_{\otimes}(\theta) = \frac{\theta\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right) + te^{-\frac{t}{\theta}}}{e^{-\frac{t}{\theta}}}$. Then in order to have unique solution, it is enough to show that $\mu_{\otimes}(\theta)$ is strictly increasing with $\lim_{\theta\to 0+}\mu_{\otimes}(\theta) = t$ and $\lim_{\theta\to\infty}\mu_{\otimes}(\theta) = T$. Clearly,

$$\mu_{\otimes}^{'}(\theta) = \frac{\theta - e^{\frac{t-T}{\theta}}(T - t + \theta)}{\theta}.$$

So,

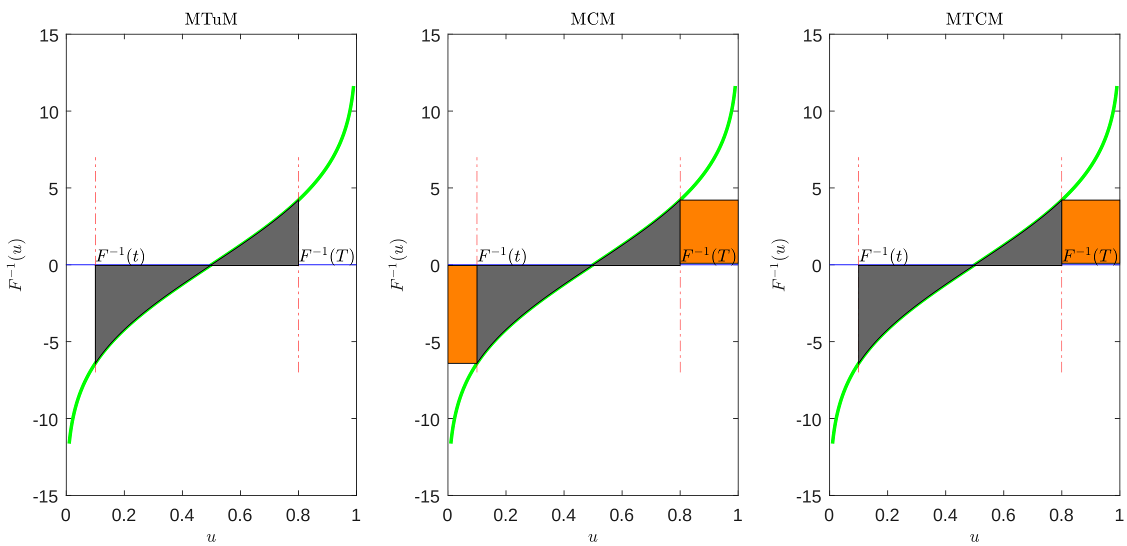
$$\begin{aligned}
& \mu_{\otimes}^{'}(\theta) > 0 \\
\iff\quad & e^{\frac{t-T}{\theta}}(T - t + \theta) < \theta \\
\iff\quad & T - t + \theta < \theta e^{\frac{T-t}{\theta}} \\
\iff\quad & T - t + \theta < \theta\left(1 + \frac{T-t}{\theta} + \frac{\left(\frac{T-t}{\theta}\right)^2}{2!} + \cdots\right) \\
\iff\quad & T - t + \theta < T - t + \theta + \text{a positive term.}
\end{aligned}$$

Thus, $\mu_{\otimes}^{'}(\theta) > 0$ and hence $\mu_{\otimes}(\theta)$ is strictly increasing. Moreover,

$$\lim_{\theta\to 0+}\mu_{\otimes}(\theta) = t.$$

$$
\begin{aligned}
\lim_{\theta\to\infty} \mu_{\otimes}(\theta) &= \lim_{\theta\to\infty} \left[\theta - \theta e^{\frac{t-T}{\theta}} + t\right] \\
&= \lim_{\theta\to\infty} \left[\theta - \theta\left(1 + \frac{t-T}{\theta} + \frac{\left(\frac{t-T}{\theta}\right)^2}{2!} + \cdots\right)\right] + t \\
&= -t + T + t = T. \qquad \square
\end{aligned}
$$

In Figure 6.1, we illustrate how these three approaches - MTuM, MCM, and MTCM - act on the underlying quantile function and thus data.



MTuM - Method of Truncated Moments
MCM - Method of Censored Moments
MTCM - Method of Left Truncated and Right Censored Moments

Figure 6.1: MTuM (left panel), MCM (middle panel), and MTCM (right panel).

6.2.4 Grouped Data

To protect the privacy of policyholders (e.g., individuals, small businesses, privately owned companies, local government funds), data vendors and publicly available databases provide summarized data, in a grouped format. For statistical inference, we view such data as *i.i.d.* realizations of a random variable that was subjected to interval censoring by multiple, say m, contiguous intervals. That is, in the *complete* data case, we observe the following *empirical* frequencies of X:

$$\widehat{\mathbb{P}}\big[c_{j-1} < X \leq c_j\big] = F_n(c_j) - F_n(c_{j-1}) = \frac{n_j}{n}, \qquad j = 1, \ldots, m+1,$$

where F_n denotes the empirical distribution function, $n = \sum_{j=1}^{m+1} n_j$ is the sample size, and the group boundaries satisfy $0 = c_0 < c_1 < \cdots < c_{m-1} < c_m < c_{m+1} = \infty$.

Computation of the empirical distribution function at the group boundaries is clear, but inside the intervals we consider the linearly interpolated empirical cdf as defined in Klugman *et al.* (2012). The linearly interpolated empirical cdf, called "ogive" and denoted by F_n, is defined as

$$F_n(x) = \frac{c_j - x}{c_j - c_{j-1}} F_n(c_{j-1}) + \frac{x - c_{j-1}}{c_j - c_{j-1}} F_n(c_j), \qquad c_{j-1} < x \leq c_j, \quad j \leq m. \tag{6.6}$$

The corresponding linearized population cdf F_1 is defined by

$$F_1(x) = \begin{cases} \frac{c_j - x}{c_j - c_{j-1}} F(c_{j-1}|\theta) + \frac{x - c_{j-1}}{c_j - c_{j-1}} F(c_j|\theta), & c_{j-1} < x \leq c_j, \quad j \leq m; \\ F(x|\theta), & x > c_m. \end{cases} \tag{6.7}$$

It is important to note that the empirical distribution F_n is not defined in the interval $(c_m, c_{m+1} = \infty)$ as it is impossible to draw a straight line joining two points $(c_m, F_n(c_m))$ and $(\infty, 1)$ unless $F_n(c_m) = 1$. The corresponding density function f_n, called the histogram, is defined as

$$f_n(x) = \frac{F_n(c_j) - F_n(c_{j-1})}{c_j - c_{j-1}} = \frac{n_j}{n(c_j - c_{j-1})}, \qquad c_{j-1} < x < c_j, \quad j \leq m, \tag{6.8}$$

where n_j is the frequency of the interval $(c_{j-1}, c_j]$.

The empirical quantile function (the inverse of F_n) is then computed as

$$F_n^{-1}(s) = c_{j-1} + \frac{(c_j - c_{j-1})(s - F_n(c_{j-1}))}{F_n(c_j) - F_n(c_{j-1})}, \quad F_n(c_{j-1}) < s \leq F_n(c_j), \; j \leq m. \tag{6.9}$$

Similarly,

$$F_1^{-1}(s|\theta) = \begin{cases} c_{j-1} + \frac{(c_j - c_{j-1})(s - F(c_{j-1}|\theta))}{F(c_j|\theta) - F(c_{j-1}|\theta)}, & F(c_{j-1}|\theta) < s \leq F(c_j|\theta), \; j \leq m; \\ F^{-1}(s|\theta), & s > F(c_m|\theta). \end{cases} \tag{6.10}$$

As was the case with individual data, defined by (2.1), the loss variable X

observed in a grouped format may also be affected by additional transformations: truncation, interval censoring, coverage modifications. In those cases, the underlying distribution function would have to be modified accordingly. For example, if m groups (n observations in total) are provided and it is known that only data above deductible d appeared, then the distributional assumption is that we observe

$$\widehat{\mathbb{P}}\left[c_{j-1} < X \leq c_j \,\middle|\, X > d\right] = \frac{n_j}{n}, \qquad j = 1, \ldots, m+1,$$

with the group boundaries satisfying $d = c_0 < c_1 < \cdots < c_m < c_{m+1} = \infty$.

By using the empirical cdf (equation 6.6) and pdf (equation 6.8), the sample truncated moments for a grouped data corresponding to the equation (6.1) is given by

$$\widehat{\mu} = \frac{1}{F_n(T_j) - F_n(t_j)} \int_{t_j}^{T_j} h_j(x) f_n(x)\, dx, \qquad 1 \leq j \leq k. \tag{6.11}$$

Note that F_n is not defined on the interval (c_m, c_{m+1}) as it is impossible to linearly interpolate a finite point and infinity. As a consequence, and F_n^{-1} is not defined on the interval $(F_n(c_m), 1]$. Therefore, in order to apply the MTM approach for grouped sample then we need to make sure that $F_n^{-1}(1-b) \leq c_m$, that is, $1 - b \leq F_n(c_m)$. Similarly, it is required to have the condition $T \leq c_m$ in order to apply MTuM for grouped data. Here, we analyze MTuM approach for exponential random variable. Let t and T be the left and right truncation points, respectively. Let us introduce the following notations:

$$\left.\begin{aligned}
p_j = p_j(\theta) &:= F(c_j|\theta) \\
P_j = P_j(\theta) &:= F(c_j|\theta) - F(c_{j-1}|\theta) \\
p_{j,n} &:= F_n(c_j) \\
\sigma^2_{j,j'} &:= \mathbb{C}ov(F_n(c_j), (F_n(c_{j'})) \\
&= \mathbb{C}ov(p_j, p_{j'}) \\
I_{i,j} &:= \mathbb{1}\{X_i \leq c_j\} \\
J_{i,j} &:= \mathbb{1}\{X_i > c_j\}
\end{aligned}\right\} \text{ for } 0 \leq j, j' \leq m+1; 0 \leq i \leq n.$$

Proposition 6.1. *Suppose* $1 \leq j \leq j' \leq m$. *Then* $\mathbb{C}ov(p_{j,n}, 1 - p_{j',n}) = -\frac{p_j(1-p_{j'})}{n}$.

Proof. Clearly,

$$p_{j,n} = \frac{\sum_{i=1}^{n} I_{i,j}}{n},$$
$$1 - p_{j',n} = \frac{\sum_{i=1}^{n} J_{i,j'}}{n}.$$

Therefore,

$$\begin{aligned}
\mathbb{C}ov(p_{j,n}, 1 - p_{j',n}) &= \mathbb{C}ov\left(\frac{\sum_{i=1}^{n} I_{i,j}}{n}, \frac{\sum_{i=1}^{n} J_{i,j'}}{n}\right) \\
&= \frac{1}{n^2}\mathbb{C}ov\left(\sum_{i=1}^{n} I_{i,j}, \sum_{i=1}^{n} J_{i,j'}\right) \\
&= \frac{1}{n^2}\sum_{k=1}^{n}\sum_{i=1}^{n} \mathbb{C}ov\left(I_{k,j}, J_{i,j'}\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{C}ov\left(I_{i,j}, J_{i,j'}\right) \\
&= \frac{1}{n^2} n \mathbb{C}ov\left(I_{1,j}, J_{1,j'}\right) \\
&= \frac{1}{n}\left[\mathbb{E}(I_{1,j}J_{1,j'}) - \mathbb{E}(I_{1,j})\mathbb{E}(J_{1,j'})\right] \\
&= \frac{1}{n}\left[0 - p_j(1 - p_{j'})\right] \\
&= -\frac{p_j(1 - p_{j'})}{n}.
\end{aligned}$$
□

The following corollary is an immediate consequence of Proposition 6.1.

Corollary 6.1. *Let* $(F_n(c_1), \ldots, F_n(c_m))$ *be a vector of empirical distribution function evaluated at the group boundaries vector* $(c_1, \ldots, c_m)$. *Then,* $(F_n(c_1), \ldots, F_n(c_m))$ *is* $\mathcal{AN}\left(\boldsymbol{F}, n^{-1}\boldsymbol{\Sigma}\right)$, *where* $\boldsymbol{F} = (F(c_1|\theta), \ldots, F(c_m|\theta))$, $\boldsymbol{\Sigma} = \left[\sigma^2_{jj'}\right]_{j,j'=1}^{m}$, *with* $\sigma^2_{jj'} = \sigma^2_{j'j} = F(c_j|\theta)(1 - F(c_{j'}|\theta))$ *for all* $j \leq j'$.

Assume that $c_0 \le c_{l-1} < t \le c_l \le c_r < T \le c_{r+1} \le c_m$. Then,

$$F_n(t) = A_1 F_n(c_{l-1}) + B_1 F_n(c_l),$$

$$F_n(T) = A_2 F_n(c_r) + B_2 F_n(c_{r+1}).$$

where $A_1 := \frac{c_l - t}{c_l - c_{l-1}}$, $A_2 := \frac{c_{r+1} - T}{c_{r+1} - c_r}$, $B_1 := \frac{t - c_{l-1}}{c_l - c_{l-1}}$, $B_2 := \frac{T - c_r}{c_{r+1} - c_r}$.

Also, consider $u_l := \frac{c_l^2 - t^2}{2(c_l - c_{l-1})}$, $v_i := \frac{c_i + c_{i-1}}{2}$, $z_r := \frac{T^2 - c_r^2}{2(c_{r+1} - c_r)}$.

Then, after some computation, we get

$$\begin{aligned} g_\mu(p_{1,n}, \ldots, p_{m,n}) &:= \widehat{\mu} \\ &= \frac{u_l(p_{l,n} - p_{l-1,n}) + \sum_{i=l+1}^{r} v_i(p_{i,n} - p_{i-1,n}) + z_r(p_{r+1,n} - p_{r,n})}{A_2 p_{r,n} + B_2 p_{r+1,n} - A_1 p_{l-1,n} - B_1 p_{l,n}} \\ &=: \frac{N}{H}. \end{aligned}$$

Note that $p_{0,n}, = 0$. Thus, by the delta method (see Theorem 1.1),

$$\widehat{\mu} \sim \mathcal{AN}\left(\mu = g_\mu(\boldsymbol{F}), n^{-1} \boldsymbol{D}_\mu \boldsymbol{\Sigma} \boldsymbol{D}_\mu'\right),$$

where $\boldsymbol{D}_\mu := \left(\left(\frac{\partial g_\mu}{\partial p_{1,n}}, \ldots, \frac{\partial g_\mu}{\partial p_{m,n}}\right)_{\boldsymbol{p}=\boldsymbol{F}}\right)$ and $\boldsymbol{p} := (p_{1,n}, \ldots, p_{m,n})'$. Consider $\boldsymbol{\Sigma}_\mu := \boldsymbol{D}_\mu \boldsymbol{\Sigma} \boldsymbol{D}_\mu'$. Clearly, if $2 \le l < r$ then

$$\frac{\partial g_\mu}{\partial p_{j,n}} = \begin{cases} 0, & \text{for } 1 \le j \le l-2 \text{ or } j \ge r+2; \\ \frac{-u_l H + A_1 N}{H^2}, & \text{for } j = l-1; \\ \frac{(u_l - v_{l+1})H + B_1 N}{H^2}, & \text{for } j = l; \\ \frac{c_{j-1} - c_{j+1}}{2H}, & \text{for } l+1 \le j \le r-1; \\ \frac{(v_r - z_r)H - A_2 N}{H^2}, & \text{for } j = r; \\ \frac{z_r H - B_2 N}{H^2}, & \text{for } j = r+1. \end{cases}$$

And if $l = r$,

$$\frac{\partial g_\mu}{\partial p_{j,n}} = \begin{cases} 0, & \text{for } 1 \le j \le l-2 \text{ or } j \ge l+2; \\ \frac{-u_l H + A_1 N}{H^2}, & \text{for } j = l-1; \\ \frac{(u_l - z_r)H - (A_2 - B_1)N}{H^2}, & \text{for } j = l; \\ \frac{z_r H - B_2 N}{H^2}, & \text{for } j = l+1. \end{cases}$$

By using equation (6.7), the corresponding population mean is

$$g_{tT}(\theta) := \mu = \frac{u_l P_l(\theta) + \sum_{i=l+1}^{r} v_i P_i(\theta) + z_r P_{r+1}(\theta)}{A_2 p_r(\theta) + B_2 p_{r+1}(\theta) - A_1 p_{l-1}(\theta) - B_1 p_l(\theta)} = \frac{N^*}{H^*}.$$

Conjecture 6.1. The function $g_{tT}(\theta)$ is strictly increasing.

Proposition 6.2.

$$\lim_{\theta \to 0+} g_{tT}(\theta) = \frac{u_l}{A_1},$$

$$\lim_{\theta \to \infty} g_{tT}(\theta) = \frac{u_l(c_{l-1} - c_l) + \sum_{i=l+1}^{r} v_i(c_{i-1} - c_i) + z_r(c_r - c_{r+1})}{-A_2 c_r - B_2 c_{r+1} + A_1 c_{l-1} + B_1 c_l}.$$

Proof. These limits can be established by using L'Hôpital's rule. □

Now, assuming the Conjecture 6.1 is true then with Proposition 6.2, we have

Theorem 6.10. *The equation $\widehat{\mu} = g_{tT}(\theta)$ has a unique solution $\widehat{\theta}$ provided that* $\frac{u_l}{A_1} < \widehat{\mu} < \frac{u_l(c_{l-1}-c_l)+\sum_{i=l+1}^{r} v_i(c_{i-1}-c_i)+z_r(c_r-c_{r+1})}{-A_2c_r-B_2c_{r+1}+A_1c_{l-1}+B_1c_l}$.

Solve the equation $\widehat{\mu} = \mu$ for θ, say $\widehat{\theta} =: g_\theta(\widehat{\mu})$. Then, again by the delta method, we conclude that $\widehat{\theta} \sim \mathcal{AN}\left(g_\theta(\mu), n^{-1}\left(g'_\theta(\mu)\right)^2 \mathbf{\Sigma}_\mu\right)$. Note that if both the left- and right-truncation points lie on the same interval, then $\widehat{\mu} = \frac{t+T}{2} = \mu$. So the parameter to be estimated disappears from the equation and hence we do not consider this case for further investigation. Let

$$P := u_l(e^{-\frac{c_{l-1}}{\theta}} - e^{-\frac{c_l}{\theta}}) + \sum_{i=l+1}^{r} v_i(e^{-\frac{c_{i-1}}{\theta}} - e^{-\frac{c_i}{\theta}}) + z_r(e^{-\frac{c_r}{\theta}} - e^{-\frac{c_{r+1}}{\theta}}),$$

$$Q := B_2(1 - e^{-\frac{c_{r+1}}{\theta}}) - A_1(1 - e^{-\frac{c_{l-1}}{\theta}}) - B_1(1 - e^{-\frac{c_l}{\theta}}).$$

Then, we get a fixed point function as $\theta = G(\theta)$, where

$$G(\theta) = -\frac{c_r}{\log\left(\frac{\widehat{\mu}A_2 - P + \widehat{\mu}Q}{\widehat{\mu}A_2}\right)}.$$

However, we need to consider the condition $\widehat{\mu}(A_2 + Q) > P$. Therefore, we need to be careful about the initialization of θ as the right truncation point T cannot be a boundary point. Because if it was, then $A_2 = 0$ and we would not able to divide by A_2 in the fixed point function $\theta = G(\theta)$.

Now, let us compute the derivative of g_θ with respect to μ, using implicit differentiation.

Case 1: Assume that the two truncation points are in two consecutive intervals, i.e., assume that $l = r$. Then $\theta' = g'_\theta(\widehat{\mu}) = \frac{A-B}{\Lambda+\Delta}$, where

$$\begin{aligned}
A &:= A_2 + B_2 - A_1 - B_1, \\
B &:= A_2 e^{-\frac{c_r}{\theta}} + B_2 e^{-\frac{c_{r+1}}{\theta}} - A_1 e^{-\frac{c_{l-1}}{\theta}} - B_1 e^{-\frac{c_l}{\theta}}, \\
\Lambda &:= \frac{u_l}{\theta^2}\left(c_{l-1}e^{-\frac{c_{l-1}}{\theta}} - c_l e^{-\frac{c_l}{\theta}}\right) + \frac{z_r}{\theta^2}\left(c_r e^{-\frac{c_r}{\theta}} - c_{r+1}e^{-\frac{c_{r+1}}{\theta}}\right), \\
\Delta &:= \frac{\widehat{\mu}}{\theta^2}\left(A_2 c_r e^{-\frac{c_r}{\theta}} + B_2 c_{r+1} e^{-\frac{c_{r+1}}{\theta}} - A_1 c_{l-1} e^{-\frac{c_{l-1}}{\theta}} - B_1 c_l e^{-\frac{c_l}{\theta}}\right).
\end{aligned}$$

Case 2: The other case is that the two truncation points are not in two consecutive intervals, i.e., assume that $l < r$. Then $\theta' = g'_\theta(\widehat{\mu}) = \frac{A-B}{\Gamma+\Delta}$, where $\Gamma := \Lambda + \sum_{i=l+1}^{r} \frac{v_i}{\theta^2}\left(c_{i-1}e^{-\frac{c_{i-1}}{\theta}} - c_i e^{-\frac{c_i}{\theta}}\right)$ and A, B, Λ, and Δ are defined above.

To get Exponential Grouped MLE, consider $P_j(\theta) := e^{-\frac{c_{j-1}}{\theta}} - e^{-\frac{c_j}{\theta}}$. Then, following Hongqi and Lixin (2002), we have, $\widehat{\theta}_{MLE} \sim \mathcal{AN}\left(\theta, \frac{1}{n}\boldsymbol{I}^{-1}(\theta)\right)$, where $\boldsymbol{I}(\theta) = \sum_{j=1}^{m} P_j(\theta)\left(\frac{d \ln P_j(\theta)}{d\theta}\right)^2$. Therefore, $Var(\widehat{\theta}_{MLE}) = \frac{\boldsymbol{I}^{-1}(\theta)}{n}$. Now, by definition of asymptotic relative efficiency:

$$\begin{aligned}
ARE(\widehat{\theta}_{MTuM}, \widehat{\theta}_{MLE}) &= \frac{Var(\widehat{\theta}_{MLE})}{Var(\widehat{\theta}_{MTuM})} \\
&= \frac{\boldsymbol{I}^{-1}(\theta)}{(g'_\theta(\widehat{\mu}))^2 \Sigma_\mu}
\end{aligned}$$

$$= \frac{\boldsymbol{I}^{-1}(\theta)}{(g'_\theta(\widehat{\mu}))^2 \boldsymbol{D}_\mu \boldsymbol{\Sigma} \boldsymbol{D}'_\mu}.$$

Note that after finding the derivative, $\boldsymbol{I}(\theta)$ can be expressed as

$$\boldsymbol{I}(\theta) = \sum_{j=1}^{m} P_j(\theta) \left(\frac{c_{j-1} e^{-\frac{c_{j-1}}{\theta}} - c_j e^{-\frac{c_j}{\theta}}}{\theta^2 (e^{-\frac{c_{j-1}}{\theta}} - e^{-\frac{c_j}{\theta}})} \right)^2$$

$$= \sum_{j=1}^{m} \left(\frac{c_{j-1} e^{-\frac{c_{j-1}}{\theta}} - c_j e^{-\frac{c_j}{\theta}}}{\theta^2} \right)^2 \frac{1}{P_j(\theta)}.$$

Chapter 7

Conclusions and Future Outlook

7.1 Concluding Remarks

In this dissertation, we have re-engineered the well-known class of MTM estimators and made them applicable to claim severity models that are fitted to truncated, censored, and insurance payment data. We have first reviewed the most common types of data transformations that appear in insurance contract specifications (due to the loss control strategies used to construct the contracts). In particular, assuming that all observed data satisfy the *i.i.d.* assumption, we have studied: the complete data scenario; left- and right-truncated data; left- and right-censored data; left-truncated, right-censored, and linearly-transformed data (also known as *payment-per-payment* variable); and interval-censored and linearly-transformed data (also known as *payment-per-loss* variable). Taking into account these data transformations, we have specified the corresponding probability distribution functions, including cumulative distribution functions, probability density and/or mass functions, and quantile functions. These probability models have then been used to specify the relevant log-likelihood functions, define sample and population trimmed moments, and describe the procedure for finding MTM estimators.

Further, asymptotic normality theorems have been established for MLE and MTM estimators under all data scenarios and probability models. Consistency of the estimators followed directly from those theorems. Robustness and computational aspects of these estimators have also been discussed. Moreover, several analytic examples based on the exponential and normal distributions have been fully worked out. Furthermore, these estimators have been implemented for the single-parameter Pareto and lognormal models that were fitted to Norwegian fire claims data for the year 1983. Then the effects of model fitting on insurance contract pricing have been investigated. In addition, in Chapter 6 we have explored a number of methodological extensions of the newly designed MTM estimators for complete, grouped and exponentially distributed random variables. Specifically, we have constructed truncated, censored, and insurance payment-type estimators and proved a series of theoretical results about estimators' existence and asymptotic normality. Our analysis has established new connections between data truncation, trimming, and censoring, which paves the way for more effective modeling of non-linearly transformed loss data.

Finally, the results of this dissertation motivate open problems and generate several ideas for further research. First, it is of interest to investigate the finite-sample, not only asymptotic, performance of the newly proposed estimators. Second, the results of Chapter 6 are limited to complete, grouped, and exponentially distributed data, but they could be extended to more general situations and models. Third, additional analysis involving risk measures could be undertaken as well. Finally, all the data scenarios considered in this dissertation are based on the *i.i.d.* assumption, which is a reasonable assumption but not the only one to consider. Some of these problems are discussed in more detail in Section 7.2.

7.2 Future Outlook

7.2.1 Simulation Studies

Asymptotically all the estimators developed in Chapters 4 and 6 are consistent and unbiased, but they might be biased when applied to finite samples. To assess the finite-sample performance of those estimators, Monte Carlo simulations is a standard tool.

By conducting simulation studies, we will aim to determine what sample size is needed to assure that the asymptotic properties of MLE and MTM estimators become valid, under all data scenarios and for selected probability distributions. In particular, we will first study estimators' bias and relative efficiency with respect to the asymptotic performance of MLE for exponential (Pareto I) and normal (log-normal) models. The main reason why MLE should be used as a benchmark is its optimal asymptotic performance in terms of variability (of course, with the usual caveat of "under certain regularity conditions").

Further, predictive modeling is an emerging set of techniques applied in actuarial practice. Scenario simulations are often employed in this area. For example, stress-testing capital allocations via risk measures such as value-at-risk (VaR; which is a quantile of the underlying loss distribution) and tail-VaR (which is a conditional tail expectation) is achieved using simulations. Thus, it is of interest to see how the redesigned MTM estimators and their variants will perform in this context.

Finally, investigation of the estimators' performance for grouped data is also of interest, as more data being released to the end user in grouped format to protect the privacy of customers and clients. Simulation of this type of data might be an especially challenging task because the empirical quantiles are not well-defined in the last (infinite) interval (see Section 6.2.4).

7.2.2 Non *i.i.d.* Data

As mentioned in Section 2.1, the *i.i.d.* assumption is reasonable, and all the estimators developed in this dissertation are based on that assumption, but there exist practical situations when it may be violated. This occurs when data sets contain some explanatory information about the underlying risk variable. For example, homeowners insurance claims database would keep track of not only loss amounts, but various property related characteristics such as location, age and construction type, replacement cost, distance to a body of water, etc. This additional information, if properly taken into account, can improve the accuracy of estimates and predictions. However, it violates the identical distribution assumption in the *i.i.d.* statement and makes losses *heterogenous*. For such data, regression type models need to be employed.

Further, losses may not be independent (e.g., if a group of insured properties were located in close proximity of each other, or if one policyholder insured multiple properties). To address this issue, actuaries use copulas which proved to be a sufficiently flexible tool for handling dependent data (see Frees and Valdez, 1998, for a thorough review of copulas).

Without a doubt, statistical models such as regression, copulas or, more generally, Generalized Linear Models offer sufficient flexibility to handle non *i.i.d.* data. However, if loss data were affected by coverage modifications, these models would require significant revisions. Moreover, the rich structure gained by introducing more parameters, makes such models vulnerable to model misspecification. This serves as additional motivation for development of robust model-fitting procedures. These topics will be pursued in the future.

Bibliography

[1] Brazauskas, V. (2009). Robust and efficient fitting of loss models: diagnostic tools and insights. *North American Actuarial Journal*, **13**(3), 356–369.

[2] Brazauskas, V., Jones, B.L., and Zitikis, R. (2007). Robustification and performance evaluation of empirical risk measures and other vector-valued estimators. *METRON–International Journal of Statistics*, **65**(2), 175–199.

[3] Brazauskas, V., Jones, B.L., and Zitikis, R. (2009). Robust fitting of claim severity distributions and the method of trimmed moments. *Journal of Statistical Planning and Inference*, **139**(6), 2028–2043.

[4] Brazauskas, V. and Kleefeld, A. (2009). Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. *Insurance: Mathematics & Economics*, **45**(3), 424–435.

[5] Brazauskas, V. and Kleefeld, A. (2011). Folded and log-folded-t distributions as models for insurance loss data. *Scandinavian Actuarial Journal*, **2011**(1), 59–74.

[6] Brazauskas, V. and Serfling, R. (2000). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. *North American Actuarial Journal*, **4**(4), 12–27.

[7] Brazauskas, V. and Serfling, R. (2003). Favorable estimators for fitting Pareto models: a study using goodness-of-fit measures with actual data. *ASTIN Bulletin*, **33**(2), 365–381.

[8] Casella, G. and Berger, R.L. (2002). *Statistical Inference*. Second edition. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

[9] Chernoff, H., Gastwirth, J.L., and Johns, Jr., M.V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Annals of Mathematical Statistics*, **38**(1), 52–72.

[10] Clark, D.R. (2013). A note on the upper-truncated Pareto distribution. *Casualty Actuarial Journal E-Forum*.

[11] Cohen, Jr., A.C. (1950). Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. *Annals of Mathematical Statistics*, **21**(4), 557–569.

[12] Ergashev, B., Pavlikov, K., Uryasev, S., and Sekeris, E. (2016). Estimation of truncated data samples in operational risk modeling. *The Journal of Risk and Insurance*, **83**(3), 613–640.

[13] Frees, E. (2017). Insurance portfolio risk retention. *North American Actuarial Journal*, **21**(4), 526–551.

[14] Frees, E.W. and Valdez, E.A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, **2**(1), 1–25.

[15] Hampel, F.R. (1968). *Contributions to the Theory of Robust Estimation.* Thesis (Ph.D.)–University of California, Berkeley.

[16] Hampel, F.R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**, 383–393.

[17] Hayya, J., Armstrong, D., and Gressis, N. (1975). A note on the ratio of two normally distributed variables. *Management Science*, **21**(11), 1338–1341.

[18] Hongqi, X. and Lixin, S. (2002). Asymptotic properties of MLE for Weibull distribution with grouped data. *Journal of Systems Science and Complexity*, **15**(2), 176–186.

[19] Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**(1), 73–101.

[20] Huber, P.J. and Ronchetti, E.M. (2009). *Robust Statistics.* Second edition. John Wiley & Sons, Inc., Hoboken, NJ.

[21] Johnson, N.L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions. Vol. 1.* Second edition. John Wiley & Sons, Inc., NY.

[22] Kleefeld, A. and Brazauskas, V. (2012). A statistical application of the quantile mechanics approach: MTM estimators for the parameters of t and gamma distributions. *European Journal of Applied Mathematics*, **23**(5), 593–610.

[23] Klugman, S.A., Panjer, H.H., and Willmot, G.E. (2012). *Loss Models: From Data to Decisions.* Fourth edition. John Wiley & Sons, Inc., Hoboken, NJ.

[24] Klugman, S.A. and Parsa, A.R. (1993). Minimum distance estimation of loss distributions. *Proceeding of the Casualty Actuarial Society*, **LXXX**, 250–270.

[25] Lee, G.Y. (2017). General insurance deductible ratemaking. *North American Actuarial Journal*, **21**(4), 620–638.

[26] Maronna, R.A., Martin, R.D., and Yohai, V.J. (2006). *Robust Statistics: Theory and Methods.* John Wiley & Sons, Ltd., Chichester.

[27] Pielke, Jr., R.A. and Landsea, C.W. (1998). Normalized hurricane damages in the United States: 1925-1995. *Weather and Forecasting*, **13**, 621–631.

[28] Reynkens, T., Verbelen, R., Beirlant, J., and Antonio, K. (2017). Modelling censored losses using splicing: A global fit strategy with mixed Erlang and extreme value distributions. *Insurance: Mathematics and Economics*, **77**, 65 – 77.

[29] Serfling, R. (2002). Efficient and robust fitting of lognormal distributions. *North American Actuarial Journal*, **6**(4), 95–109.

[30] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, Inc., NY.

[31] Shah, S.M. and Jaiswal, M.C. (1966). Estimation of parameters of doubly truncated normal distribution from first four sample moments. *Annals of the Institute of Statistical Mathematics*, **18**, 107–111.

[32] Tukey, J.W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics*, pages 448–485. Stanford University Press, Stanford, CA.

[33] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

[34] Verbelen, R., Gong, L., Antonio, K., Badescu, A., and Lin, S. (2015). Fitting mixtures of Erlangs to censored and truncated data using the EM algorithm. *ASTIN Bulletin*, **45**(3), 729758.

[35] Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer-Verlag, NY.

[36] Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer, NY.

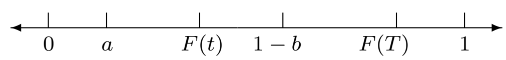
[37] Zhao, Q., Brazauskas, V., and Ghorai, J. (2018a). Robust and efficient fitting of severity models and the method of Winsorized moments. *ASTIN Bulletin*, **48**(1), 275–309.

[38] Zhao, Q., Brazauskas, V., and Ghorai, J. (2018b). Small-sample performance of the MTM and MWM estimators for the parameters of log-location-scale families. *Journal of Statistical Computation and Simulation*, **88**(4), 808–824.

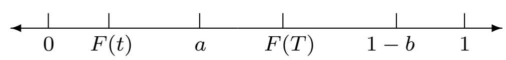
Appendix A

Proofs and Additional Details

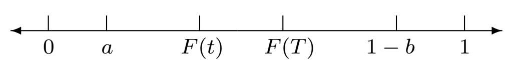
Additional Cases for Note 3.4:

Case 2: $0 \le a \le F(t) \le 1-b < F(T) \le 1$; 

$$\int_a^{1-b} h_j(F_{**}^{-1}(u))\,du \;=\; h_j(t)(F(t|\boldsymbol{\theta}) - a) + \int_{F(t|\boldsymbol{\theta})}^{1-b} h_j(F^{-1}(u|\boldsymbol{\theta}))\,du.$$

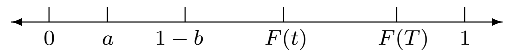
Case 3: $0 \le F(t) \le a < F(T) \le 1-b \le 1$; 

$$\int_a^{1-b} h_j(F_{**}^{-1}(u))\,du \;=\; \int_a^{F(T|\boldsymbol{\theta})} h_j(F^{-1}(u|\boldsymbol{\theta}))\,du + h_j(T)(1-b-F(T|\boldsymbol{\theta})).$$

Case 4: $0 \le a \le F(t) < F(T) \le 1-b \le 1$; 

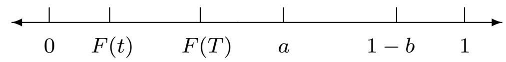
$$\begin{aligned}
\int_a^{1-b} h_j(F_{**}^{-1}(u))\,du &= \int_a^{F(t|\boldsymbol{\theta})} h_j(F_{**}^{-1}(u))\,du + \int_{F(t|\boldsymbol{\theta})}^{F(T|\boldsymbol{\theta})} h_j(F_{**}^{-1}(u))\,du \\
&\quad + \int_{F(T|\boldsymbol{\theta})}^{1-b} h_j(F_{**}^{-1}(u))\,du \\
&= \int_a^{F(t|\boldsymbol{\theta})} h_j(t)\,du + \int_{F(t|\boldsymbol{\theta})}^{F(T|\boldsymbol{\theta})} h_j(F^{-1}(u|\boldsymbol{\theta}))\,du \\
&\quad + \int_{F(T|\boldsymbol{\theta})}^{1-b} h_j(T)\,du \\
&= h_j(t)(F(t|\boldsymbol{\theta}) - a) + \int_{F(t|\boldsymbol{\theta})}^{F(T|\boldsymbol{\theta})} h_j(F^{-1}(u|\boldsymbol{\theta}))\,du \\
&\quad + h_j(T)(1-b-F(T|\boldsymbol{\theta})).
\end{aligned}$$

Case 5: $0 \le a < 1-b \le F(t) < F(T) \le 1$;



$$\int_a^{1-b} h_j(F_{**}^{-1}(u))\,du \;=\; h_j(t)(1-a-b).$$

Case 6: $0 \le F(t) < F(T) \le a < 1-b \le 1$;

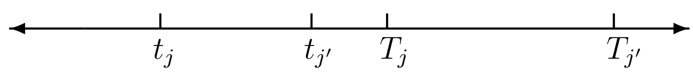


$$\int_a^{1-b} h_j(F_{**}^{-1}(u))\,du \;=\; h_j(T)(1-a-b).$$

It is important to note that the MTM approach does not work in both Cases 5 and 6 as those integral are constants rather than functions of parameters to be estimated.

All four possible scenarios for Section 6.1.2

Scenario 1: $t_j \le t_{j'} < T_j \le T_{j'}$



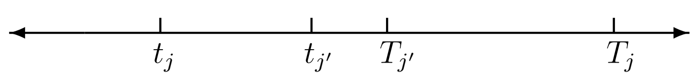
In this case,

$$Y_{jj'} = h_{jj'}(X)\mathbb{1}\{t_{jj'} < X \le T_{jj'}\} = h_{jj'}(X)\mathbb{1}\{t_{j'} < X \le T_j\},$$

$$W_{jj'} = Z_j\mathbb{1}\{t_{j'} < X \le T_j\},$$

$$W_{j'j} = Z_{j'}\mathbb{1}\{t_{j'} < X \le T_j\}.$$

Scenario 2: $t_j \le t_{j'} < T_{j'} \le T_j$



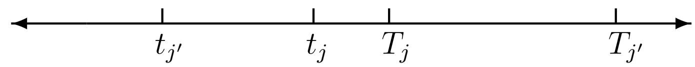
In this case,

$$Y_{jj'} = h_{jj'}(X)\mathbb{1}\{t_{jj'} < X \le T_{jj'}\} = h_{jj'}(X)\mathbb{1}\{t_{j'} < X \le T_{j'}\},$$

$$W_{jj'} = Z_j\mathbb{1}\{t_{j'} < X \le T_{j'}\},$$

$$W_{j'j} = Z_{j'}\mathbb{1}\{t_{j'} < X \le T_{j'}\}.$$

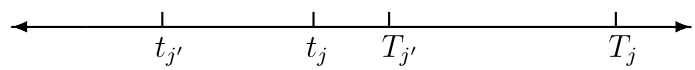
Scenario 3: $t_{j'} \leq t_j < T_j \leq T_{j'}$



In this case,

$$\begin{aligned}
Y_{jj'} &= h_{jj'}(X)\mathbb{1}\{t_{jj'} < X \leq T_{jj'}\} = h_{jj'}(X)\mathbb{1}\{t_j < X \leq T_j\}, \\
W_{jj'} &= Z_j\mathbb{1}\{t_j < X \leq T_j\}, \\
W_{j'j} &= Z_{j'}\mathbb{1}\{t_j < X \leq T_j\}.
\end{aligned}$$

Scenario 4: $t_{j'} \leq t_j < T_{j'} \leq T_j$



In this case,

$$\begin{aligned}
Y_{jj'} &= h_{jj'}(X)\mathbb{1}\{t_{jj'} < X \leq T_{jj'}\} = h_{jj'}(X)\mathbb{1}\{t_j < X \leq T_{j'}\}, \\
W_{jj'} &= Z_j\mathbb{1}\{t_j < X \leq T_{j'}\}, \\
W_{j'j} &= Z_{j'}\mathbb{1}\{t_j < X \leq T_{j'}\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mu_{Y_{jj'}} &= \mathbb{E}[Y_{jj'}] & \mu_{W_{jj'}} &= \mathbb{E}[W_{jj'}] \\
&= \int_{F(t_{jj'}|\boldsymbol{\theta})}^{F(T_{jj'}|\boldsymbol{\theta})} h_{jj'}\left(F^{-1}(u|\boldsymbol{\theta})\right) du, & &= \int_{F(t_{jj'}|\boldsymbol{\theta})}^{F(T_{jj'}|\boldsymbol{\theta})} h_j\left(F^{-1}(u|\boldsymbol{\theta})\right) du.
\end{aligned}$$

Proof of Theorem 6.2:

Clearly, $g_{\boldsymbol{V}}(\boldsymbol{\mu_V}) = \left(\frac{\mu_{Y_1}}{p_1}, \ldots, \frac{\mu_{Y_k}}{p_k}\right) =: (\mu_1, \ldots, \mu_k) =: \boldsymbol{\mu}$. From Lemma 6.2, it follows that

$$\boldsymbol{D_V} := \left[\left.\frac{\partial g_j}{\partial x_{j'}}\right|_{\boldsymbol{x}=\boldsymbol{\mu_V}}\right]_{k\times 2k} = \left[d_{\boldsymbol{V},jj'}\right]_{k\times 2k},$$

where

$$d_{\boldsymbol{V},jj'} := \begin{cases} \dfrac{1}{p_{j'}}, & \text{if } 1 \le j = j' \le k; \\ -\dfrac{\mu_{Y_j}}{p_j^2}, & \text{if } j' - j = k; \\ 0, & \text{otherwise.} \end{cases}$$

Now, with an application of the delta method (see Theorem 1.1) corresponding with the function $g_{\boldsymbol{V}}$ above, (see Serfling, 1980, §3.3 Theorem A), we have

$$(\widehat{\mu}_1, \ldots, \widehat{\mu}_k) \sim \mathcal{AN}\left(g_{\boldsymbol{V}}(\boldsymbol{\mu}_{\boldsymbol{V}}) = \boldsymbol{\mu}, \frac{1}{n}\boldsymbol{D}_{\boldsymbol{V}}\boldsymbol{\Sigma}_{\boldsymbol{V}}\boldsymbol{D}_{\boldsymbol{V}}'\right). \qquad \square$$

Proof of Theorem 6.4:

The r.v. Y can be expressed in the form of

$$Y = X \wedge T - T\mathbb{1}\{T < X < \infty\} - X \wedge t + t\mathbb{1}\{t < X < \infty\}.$$

Define, $I_{a,b} := \mathbb{1}\{a < X < b\}$. Therefore,

$$\begin{aligned} \mu_Y &= \mathbb{E}[Y] \\ &= \mathbb{E}[X \wedge T] - \mathbb{E}[TI_{T,\infty}] - \mathbb{E}[X \wedge t] + \mathbb{E}[tI_{t,\infty}] \\ &= \theta(1 - e^{-\frac{T}{\theta}}) - Te^{-\frac{T}{\theta}} - \theta(1 - e^{-\frac{t}{\theta}}) + te^{-\frac{t}{\theta}} \\ &= \theta\left(e^{-\frac{t}{\theta}} - e^{-\frac{T}{\theta}}\right) + te^{-\frac{t}{\theta}} - Te^{-\frac{T}{\theta}}. \end{aligned}$$

Since

$$Y = X \wedge T - T\mathbb{1}\{T < X < \infty\} - X \wedge t + t\mathbb{1}\{t < X < \infty\},$$

then

$$\begin{aligned} Y^2 &= (X \wedge T)^2 + (X \wedge t)^2 + T^2 I_{T,\infty} + t^2 I_{t,\infty} - 2T(X \wedge T)I_{T,\infty} - 2(X \wedge T)(X \wedge t) \\ &\quad + 2t(X \wedge T)I_{t,\infty} + 2T(X \wedge T)I_{T,\infty}(X \wedge t) - 2tTI_{t,\infty}I_{T,\infty} - 2t(X \wedge t)I_{t,\infty} \\ &= (X \wedge T)^2 - (X \wedge t)^2 - 2t\left[X \wedge T - X \wedge t\right] + T^2 I_{T,\infty} + t^2 I_{t,\infty} - 2T^2 I_{T,\infty} \end{aligned}$$

$$
\begin{aligned}
& + 2t\left(XI_{t,T} + TI_{T,\infty}\right) + 2tTI_{T,\infty} - 2tTI_{T,\infty} - 2t^2 I_{t,\infty} \\
= \;& (X \wedge T)^2 - (X \wedge t)^2 - 2t\left[X \wedge T - X \wedge t\right] - T^2 I_{T,\infty} - t^2 I_{t,\infty} \\
& + 2t\left(XI_{t,T} + TI_{T,\infty}\right).
\end{aligned}
$$

Since $X \sim Exp(\theta)$ then $\mathbb{E}[Y^2]$ is computed as below:

$$
\begin{aligned}
\mu_{Y^2} &= \mathbb{E}[Y^2] \\
&= \mathbb{E}[(X \wedge T)^2] - \mathbb{E}[(X \wedge t)^2] - 2t\left[\mathbb{E}[X \wedge T] - \mathbb{E}[X \wedge t]\right] - T^2 \mathbb{E}[I_{T,\infty}] \\
&\quad - t^2 \mathbb{E}[I_{t,\infty}] + 2t\left(\mathbb{E}[XI_{t,T}] + T\mathbb{E}[I_{T,\infty}]\right) \\
&= 2\theta^2 \Gamma\left(3; \frac{T}{\theta}\right) + T^2 e^{-\frac{T}{\theta}} - 2\theta^2 \Gamma\left(3; \frac{t}{\theta}\right) - t^2 e^{-\frac{t}{\theta}} \\
&\quad - 2t\left[\theta\left(1 - e^{-\frac{T}{\theta}}\right) - \theta\left(1 - e^{-\frac{t}{\theta}}\right)\right] - T^2 e^{-\frac{T}{\theta}} - t^2 e^{-\frac{t}{\theta}} \\
&\quad + 2t\left[\theta e^{-\frac{t}{\theta}} + te^{-\frac{t}{\theta}} - \theta e^{-\frac{T}{\theta}} - Te^{-\frac{T}{\theta}} + Te^{-\frac{R}{\theta}}\right] \\
&= 2\theta^2\left(\Gamma\left(3; \frac{T}{\theta}\right) - \Gamma\left(3; \frac{t}{\theta}\right)\right).
\end{aligned}
$$

Therefore,

$$
\sigma_Y^2 = \mu_{Y^2} - \mu_Y^2 = 2\theta^2\left(\Gamma\left(3; \frac{T}{\theta}\right) - \Gamma\left(3; \frac{t}{\theta}\right)\right) - \mu_Y^2. \qquad \square
$$

Proof of Theorem 6.6:

Note that the parameter vector is given by $\boldsymbol{\theta} = (\alpha, x_0)$ with x_0 known in advance. The population version of $\widehat{\mu}$ is given by

$$
\begin{aligned}
\mu_* &= \mathbb{E}[h(X) | t < X \leq T] \\
&= \frac{\mathbb{E}[h(X)\mathbb{1}\{t < X \leq T\}]}{F(T|\boldsymbol{\theta}) - F(t|\boldsymbol{\theta})} \\
&= \frac{\int_t^T h(x) f(x|\boldsymbol{\theta})\, dx}{F(T|\boldsymbol{\theta}) - F(t|\boldsymbol{\theta})} \\
&= \frac{\int_{F(t|\boldsymbol{\theta})}^{F(T|\boldsymbol{\theta})} h(F^{-1}(u|\boldsymbol{\theta}))\, du}{F(T|\boldsymbol{\theta}) - F(t|\boldsymbol{\theta})}
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{\int_{F(t|\boldsymbol{\theta})}^{F(T|\boldsymbol{\theta})} \log(1-u)\,du}{\alpha(F(T|\boldsymbol{\theta}) - F(t|\boldsymbol{\theta}))} \\
&= -\frac{1}{\alpha\left(\left(\frac{x_0}{t}\right)^\alpha - \left(\frac{x_0}{T}\right)^\alpha\right)} \\
&\quad \times \left[F(t|\boldsymbol{\theta}) - F(T|\boldsymbol{\theta}) + \alpha(1-F(t|\boldsymbol{\theta}))\log\left(\frac{x_0}{t}\right) - \alpha(1-F(T|\boldsymbol{\theta}))\log\left(\frac{x_0}{T}\right)\right] \\
&= -\frac{1}{\alpha\left(\left(\frac{x_0}{t}\right)^\alpha - \left(\frac{x_0}{T}\right)^\alpha\right)}\left[\left(\frac{x_0}{T}\right)^\alpha - \left(\frac{x_0}{t}\right)^\alpha + \alpha\left(\frac{x_0}{t}\right)^\alpha \log\left(\frac{x_0}{t}\right) - \alpha\left(\frac{x_0}{T}\right)^\alpha \log\left(\frac{x_0}{T}\right)\right] \\
&= \frac{x_0^\alpha (tT)^\alpha}{\alpha x_0^\alpha (tT)^\alpha (T^\alpha - t^\alpha)}\left[T^\alpha\left(1 - \alpha\log\left(\frac{x_0}{t}\right)\right) - t^\alpha\left(1 - \alpha\log\left(\frac{x_0}{T}\right)\right)\right] \\
&= \frac{1}{\alpha(T^\alpha - t^\alpha)}\left[T^\alpha\left(1 - \alpha\log\left(\frac{x_0}{t}\right)\right) - t^\alpha\left(1 - \alpha\log\left(\frac{x_0}{T}\right)\right)\right] \\
&= \frac{A_{tT}}{\alpha(T^\alpha - t^\alpha)}.
\end{aligned}
$$

Now, to establish the proof of the statement, it is enough to prove that the function g_{tT} is strictly decreasing with respect to α. For that,

$$
\begin{aligned}
g'_{tT}(\alpha) &= \frac{dg_{tT}(\alpha)}{d\alpha} \\
&= \frac{(Tt)^\alpha \alpha^2 \left(\log\left(\frac{T}{t}\right)\right)^2 - (T^\alpha - t^\alpha)^2}{\alpha^2 (T^\alpha - t^\alpha)^2}.
\end{aligned}
$$

Now, in order to show that $g'_{tT}(\alpha) < 0$, it is enough to establish $(Tt)^\alpha \alpha^2 \left(\log\left(\frac{T}{t}\right)\right)^2 - (T^\alpha - t^\alpha)^2 < 0$ which is equivalent to establish that $(Tt)^{\frac{\alpha}{2}} \alpha \log\left(\frac{T}{t}\right) < T^\alpha - t^\alpha$. Now,

$$
\begin{aligned}
& (Tt)^{\frac{\alpha}{2}} \alpha \log\left(\frac{T}{t}\right) < T^\alpha - t^\alpha \\
\iff\quad & \alpha \log\left(\frac{T}{t}\right) < \left(\frac{T}{t}\right)^{\frac{\alpha}{2}} - \left(\frac{T}{t}\right)^{-\frac{\alpha}{2}} \\
\iff\quad & \alpha \log\left(\frac{T}{t}\right) < 2\sinh\left(\frac{\alpha}{2}\log\left(\frac{T}{t}\right)\right) \\
\iff\quad & \frac{\alpha}{2} \log\left(\frac{T}{t}\right) < \sinh\left(\frac{\alpha}{2}\log\left(\frac{T}{t}\right)\right).
\end{aligned}
$$

But we know that $x < \sinh x$ for all $x > 0$, therefore, $g'_{tT}(\alpha) < 0$ for all $\alpha > 0$ which implies that g_{tT} is strictly decreasing. Also, note that

$$\lim_{\alpha \to 0+} g_{tT}(\alpha) = \frac{(\log(T))^2 - (\log(t))^2 - 2\log(T)\log\left(\frac{x_0}{t}\right) + 2\log(t)\log\left(\frac{x_0}{T}\right)}{2\log\left(\frac{T}{t}\right)},$$

$$\lim_{\alpha \to \infty} g_{tT}(\alpha) = -\log\left(\frac{x_0}{t}\right). \qquad \square$$

Appendix B

MATLAB Code for Exponential ARE Computation

```
function [ARE, Ft, FT] = Exp_MTuM_MLE_ARE(t, T, theta)
% This program evaluates the asymptotic relative efficiency of MTuM for
% Exponential complete data w.r.t. MLE.
%
% INPUT:
%   t:      Vector of left truncation points
%   T:      Vector of right truncation points
%   theta:  Parameter of Exp(theta)
%
% OUTPUT:
%   ARE: A (length(t)+2) x (length(T)+2) matrix as below
%        Read carefully!
%   Ft:  Truncated proportion on the left tail.
%   FT:  Truncated proportion on the right tail.
%   Ft:  Truncated proportion on the left tail.
%   FT:  Truncated proportion on the right tail.
%
% External Programs Called: Non!

AREMatrix = zeros(length(t),length(T));
for i = 1:length(t)
    horARE = nan(1,length(T));
    for j = 1:length(T)
        p = FExp(T(j), theta) - FExp(t(i), theta);
        et = exp(-t(i)/theta);
        eT = exp(-T(j)/theta);
        MeanY = theta*(et-eT) + t(i)*et - T(j)*eT;
        EY2 = 2*(theta^2)*(gammainc((T(j)/theta),3) - gammainc((t(i)/theta),3));
        % Please note that the parametrization of incomplete gamma function is just
        % opposite in MATLAB and KPW.
        popMu = MeanY/p;
        VarY = EY2 - MeanY^2;

        % The following is the derivative of g_{\theta}
        A2 = t(i)*exp(-t(i)/theta)*(theta+t(i)) - T(j)*exp(-T(j)/theta)*(theta+T(j)) ...
            + (theta^2)*p - popMu*(t(i)*exp(-t(i)/theta) - T(j)*exp(-T(j)/theta));
        gThetaDer = (theta^2)*p/A2;

        format short
        AA =  (gThetaDer^2)*(VarY*p - (1 - p)*MeanY^2);
        ARE1 = round(1000*((theta^2)*(p^3))/AA)/1000;

        if t(i) < T(j)
            horARE(j) = ARE1;
```

```
        end
    end
    AREMatrix(i,:) = horARE;
end

Ft = round(100*(FExp(t, theta)))/100;
FT = round(100*(1-FExp(T, theta)))/100;
TT = [[nan nan Inf T(2:length(T))]; [nan nan FT]];
tt = [t' Ft'];

ARE = [TT;[tt AREMatrix]];

end

%%% ———————— Auxiliary Functions ———————— %%%

function distF = FExp(x, theta)
  distF = 1-exp(-x/theta);
end
```

```
function [ARE, Ft, FT] = Exp_MTuMWin_MLE_ARE(t, T, theta)
% This program evaluates the asymptotic relative efficiency of MTuM through
% Winsorized approach for Exponential complete data w.r.t. MLE.
%
% INPUT:
%   t:       Vector of left truncation points
%   T:       Vector of right truncation points
%   theta:   Parameter of Exp(theta)
%
% OUTPUT:
%   ARE: A (length(t)+2) x (length(T)+2) matrix as below
%          Read carefully!
%   Ft:  Truncated proportion on the left tail.
%   FT:  Truncated proportion on the right tail.
%
% External Programs Called: Non!

AREMatrix = zeros(length(t),length(T));
for i = 1:length(t)
    horARE = nan(1,length(T));
    for j = 1:length(T)
        et = exp(-t(i)/theta);
        eT = exp(-T(j)/theta);
        MeanY = theta*(et-eT) + t(i)*et - T(j)*eT;
        EY2 = 2*theta^2*(gammainc((T(j)/theta),3) - gammainc((t(i)/theta),3));
        % Please note that the parametrization of incomplete gamma function is just
        % opposite in MATLAB and KPW.

        popMu = t(i)*FExp(t(i),theta) + MeanY + T(j)*(1-FExp(T(j),theta));
        popMu2 = t(i)^2*FExp(t(i),theta) + EY2 + T(j)^2*(1-FExp(T(j),theta));
        popVar = popMu2 - popMu^2;
        % The following is the derivative of g_{\theta}
        A1 = exp(-t(i)/theta)*(theta+t(i)) - exp(-T(j)/theta)*(theta+T(j));
        gThetaDer = theta/A1;

        format short
        AA =  (gThetaDer^2)*(popVar);
        ARE1 = round(1000*((theta^2)/(AA)))/1000;

        if t(i) < T(j)
            horARE(j) = ARE1;
        end
    end
    AREMatrix(i,:) = horARE;
end

Ft = round(100*(FExp(t, theta)))/100;
FT = round(100*(1-FExp(T, theta)))/100;
```

```
TT = [[nan nan Inf T(2:length(T))]; [nan nan FT]];
tt = [t' Ft'];

ARE = [TT;[tt AREMatrix]];

end

%%% -------------- Auxiliary Functions -------------- %%%

function distF = FExp(x, theta)
  distF = 1-(exp(-x/theta));
end
```

```
function [ARE, Ft, FT] = Exp_MTuMLTRC_MLE_ARE(t, T, theta)
% This program evaluates the asymptotic relative efficiency of left
% truncated and right censored Exponential data w.r.t. MLE.
%
% INPUT:
%   t:      Vector of left truncation points
%   T:      Vector of right truncation points
%   theta:  Parameter of Exp(theta)
%
% OUTPUT:
%   ARE: A (length(t)+2) x (length(T)+2) matrix as below
%        Read carefully!
%   Ft:  Truncated proportion on the left tail.
%   FT:  Truncated proportion on the right tail.
%
% External Programs Called: Non!

AREMatrix = zeros(length(t),length(T));
for i = 1:length(t)
    p = 1-FExp(t(i),theta);
    horARE = nan(1,length(T));
    for j = 1:length(T)
        et = exp(-t(i)/theta);
        eT = exp(-T(j)/theta);
        EY = theta*(et-eT) + t(i)*et - T(j)*eT;
        EY2 = 2*theta^2*(gammainc((T(j)/theta),3) - gammainc((t(i)/theta),3));
        % Please note that the parametrization of incomplete gamma function is just
        % opposite in MATLAB and KPW.

        EW = EY + T(j)*(1-FExp(T(j),theta));
        EW2 = EY2 + T(j)^2*(1-FExp(T(j),theta));
        VarW = EW2 - EW^2;
        popMu = EW/p;

        % s12 = EY+T(j)*exp(-T(j)/theta)-EW*exp(-t(i)/theta);
        popVar = (VarW/p^2) - (EW^2*(1-p)/p^3);
        % The following is the derivative of g_{\theta}
        A2 = theta*((t(i)*exp(-t(i)/theta)) - (T(j)*exp(-T(j)/theta))) + ...
            theta^2*(exp(-t(i)/theta)-exp(-T(j)/theta)) + t(i)^2*p;
        gThetaDer = (theta^2*p)/(A2-t(i)*popMu*p);

        format short
        AA =  (gThetaDer^2)*(popVar);
        ARE1 = round(1000*((theta^2)/(AA)))/1000;

        if t(i) < T(j)
            horARE(j) = ARE1;
        end
    end
    AREMatrix(i,:) = horARE;
end

Ft = round(100*(FExp(t, theta)))/100;
FT = round(100*(1-FExp(T, theta)))/100;
TT = [[nan nan Inf T(2:length(T))]; [nan nan FT]];
```

```
tt = [t' Ft'];

ARE = [TT;[tt AREMatrix]];

end

%%% ------------ Auxiliary Functions ------------ %%%

function distF = FExp(x, theta)
  distF = 1-(exp(-x/theta));
end
```

Appendix C

MATLAB Code for Real Data Illustrations

```
function [Estimators, KS_GOF, AD_GOF, PPrem, LNPrem] = ...
    RealDataIllustrationsB(YY,ETH,LNTH,a,b,dataDed,polDed,pLim,c,SigLevel,dmu,
        n_pVal)
% Typical Run:
%   a = [10^-7 10^-7 10^-7 10^-7 .05 0.10 0.25];
%   b = [10^-7 0.05 0.10 0.25 .05 0.10 0.25];
%   [Estimators, KS_GOF, AD_GOF, PPrem, LNPrem] =
%   RealDataIllustrationsB(83,1,0,a,b,500,1500,14000,0.8,0.05,1000,10^3)
tic
h = waitbar(0,'Please wait...');
%==========================================================================%
% INPUT:
%   YY:          Year.
%   ETH:         Threshold for Pareto I distributions.
%   LNTH:        Threshold for Lognormal distributions.
%   a:           Left side trimming proportion vector.
%   b:           Right side trimming proportion vector.
%               NOTE: a and b should be of same length.
%   dataDed:     Observed left truncation, deductible.
%   polDed:      Researcher produced deductible. Note that this is not
%                coming from the data.
%   pLim:        Researcher produced policy limit. Note that this is not
%                coming from the data.
%               NOTE: All dataDed, polDed, and pLim should be in thousand not in
%                million.
%   c:           Co-insurance factor.
%   SigLevel:    Significance level for confidence interval construction.
%   dmu:         Data measure unit. For example, for NFC it is 1000.
%   n_pVal:      Total iteration for p-value calculation.
%
% OUTPUT:
%   Estimators:      A leangth(a)-by-3 matrix. See the matrix for details.
%   KS_GOF:          KolmogorovSmirnov goodness-of-fit statistics.
%   AD_GOF:          Anderson-Darling goodness-of-fit statistics.
%   PPrem:           Pareto I based and empirical premiums with CIs.
%   LNPrem:          Lognormal based and empirical premiums with CIs.
%
% External Function Called: Non! See auxiliary functions!
%==========================================================================%

x = importdata('nfc.txt'); % This is the entire data set.
xYY = x(x(:,2) == YY); % This can extract required data for a particular year.
lenD = sum(xYY == dataDed);
if lenD > 0
    for dataDegroup = 1:lenD
```

```matlab
        xYY(dataDegroup) = dataDed+(dataDed+0.5-dataDed)*dataDegroup/(lenD+1);
    end
end

n = length(xYY);

disp(sprintf('############################################'));
disp(sprintf('Year = 19%d: %d data (%dth) are degrouped\n',YY,lenD,dataDed));
disp(sprintf('Sample Size: n = %d \n',length(xYY)));
disp(sprintf('############################################'));

I0 = zeros(length(a),1); I1 = zeros(length(a),1); J = zeros(length(a),1);
for i = 1:length(a)
    I0(i) = b(i).*(1-log(b(i)))+(1-a(i)).*(log(1-a(i))-1);
    I1(i) = (a(i)+b(i)-1) + log((1-a(i))./b(i));
    J(i) = (1-b(i)-a(i)).*(a(i)+log(1-a(i))) - I0(i) +(-b(i)).*I1(i);
end

aData = sort(xYY.*dmu);
% This is the sorted and scalled (by dmu) data!
uData = Inf;
expData = log(aData./ETH); % Log Transformation of the data.
dExp = log((dataDed.*dmu)./ETH); % Deductible in terms of Exponential.
uExp = log((uData.*dmu)./ETH); % Policy limit in terms of Exponential.

figure(1);
histogram(xYY./1000,'FaceColor','b');
xlabel('Losses (million)','Interpreter','Latex','FontSize', 11);
ylabel('Frequency', 'Interpreter','Latex','FontSize', 11);

figure(2)
histogram(log(xYY.*1000),'FaceColor','b');
xlabel('Log(Actual Losses)','Interpreter','Latex','FontSize', 11);
ylabel('Frequency', 'Interpreter','Latex','FontSize', 11);

%==========================================================================%
%                          Model Fitting                                   %
%==========================================================================%
% Pareto I MLE fit!
thetaHatMLEExp = mean(expData)-dExp;
alphaHatMLE = 1./thetaHatMLEExp;

% Lognormal MLE Fit
nData = log(aData-LNTH); % Normal data.
dNormal = log(dataDed.*dmu-LNTH);
uNormal = log(uData.*dmu-LNTH);
mu1MLEHatLN = mean(nData);
mu2MLEHatLN = mean((nData).^2);
Z = @(tz)(2.*normpdf(tz)./(erfc(tz./sqrt(2)))); % This is Z1 from Cohen (1950)!
opt = optimoptions(@fsolve,'Display','off');
A_EPUS = ((mu2MLEHatLN-mu1MLEHatLN.^2)./((mu1MLEHatLN-dNormal).^2));
G = @(t)(erfc(t./sqrt(2)).*power(2*normpdf(t) - t.*erfc(t./sqrt(2)), -1)...
    .*(erfc(t./sqrt(2)).*power(2*normpdf(t) - t.*erfc(t./sqrt(2)), -1) - t));
MLEobjFun = @(t)(G(t)-1-A_EPUS);
tz0 = (dNormal-mu1MLEHatLN)./(sqrt(mu2MLEHatLN-mu1MLEHatLN.^2));
tzHat = fsolve(MLEobjFun, tz0, opt);
sigmaMLEHatLN = (mu1MLEHatLN-dNormal)./(feval(Z,tzHat)-tzHat);
thetaMLEHatLN = dNormal - sigmaMLEHatLN.*tzHat;
MLESol = [thetaMLEHatLN sigmaMLEHatLN];
an = floor(a.*n);
bn = floor(b.*n);

% MTM Fitting.
% First, Exponential Fit!
muHatExp = zeros(length(a),1); % Sample trimmed mean vector.
ExpThetaHat = zeros(length(a),1); % Exponential trimmed estimated vector.
for i = 1:length(a)
    muHatExp(i) = (1./(n-an(i)-bn(i))).*sum(expData((an(i)+1):(n-bn(i))));
```

```
    ExpThetaHat(i) = -((muHatExp(i)-dExp).*(1-a(i)-b(i)))./I0(i);
end
% Second, Normal Fit!
thSgHat = zeros(length(a),2); % Lognormal trimmed estimated vector.
LNMoments = zeros(length(a),2);
for i = 1:length(a)
    mu1HatLN = (1./(n-an(i)-bn(i))).*sum(nData((an(i)+1):(n-bn(i))));
    mu2HatLN = (1./(n-an(i)-bn(i))).*sum((nData((an(i)+1):(n-bn(i)))).^2);
    theta0 = mu1HatLN;
    sigma0 = sqrt(mu2HatLN-mu1HatLN.^2);
    x0 = [theta0 sigma0];
    MTMobjFun = @(x)OnlyLeftTruncatedNormalMTM2Eqns(x,dNormal,[mu1HatLN mu2HatLN],a(i),b(i));
    MTMSol = fsolve(MTMobjFun, x0, opt);
    LNMoments(i,:) = [mu1HatLN mu2HatLN];
    thSgHat(i,:) = MTMSol;
end

%===================================================================%
% Calculation of Klmogorov-Smirnov and Anderson-Darling Test Statistics
% and p-values for Exp and Normal Fitting!
%===================================================================%
KS_TS = zeros(length(a),2);
% KS-test statistic matrix.
% First column represents for Exponential model.
% Second column represents for Normal model.
AD_TS = zeros(length(a),2);
% AD-test statistic matrix.
% First column represents for Exponential model.
% Second column represents for Normal model.
for i = 1:length(a)
    FExpMTMFitted = @(x)((exp(-dExp./ExpThetaHat(i)) - exp(-x./ExpThetaHat(i))) ...
        ./exp(-dExp./ExpThetaHat(i)));
    FNormalMTMFitted = @(x)((normcdf((x-thSgHat(i,1))./thSgHat(i,2))-normcdf((dNormal-thSgHat(i,1))./thSgHat(i,2))) ...
        ./(1-normcdf((dNormal-thSgHat(i,1))./thSgHat(i,2))));
    KS_ExpBase = KolmogorovSmirnovTS(FExpMTMFitted,expData);
    KS_NormalBase = KolmogorovSmirnovTS(FNormalMTMFitted,nData);
    AD_ExpBase = AndersonDarlingTS(FExpMTMFitted,expData',dExp,uExp);
    AD_NormalBase = AndersonDarlingTS(FNormalMTMFitted,nData',dNormal,uNormal);
    KS_TS(i,:) = [KS_ExpBase KS_NormalBase];
    AD_TS(i,:) = [AD_ExpBase AD_NormalBase];
end
%===================================================================%
% Calculation of p-values:
ExpKSADpArray = zeros(n_pVal,2,length(a));
NormalKSADArray = zeros(n_pVal,2,length(a));
% This is the array of dimension "n_pVal-by-2-by-length(a)", that is, there
% will be length(a) many matrices where the first column represents the
% KS-test value and the second column represent the AD-test value.
ExpKSADpValMatrix = zeros(length(a),2);
NormalKSADpValMatrix = zeros(length(a),2);
% This will produce the KS and AD final values.
for j = 1:n_pVal
    uRand = rand(n,1);
    for i = 1:length(a)
        expSimData = dExp + ExpVaR(uRand, ExpThetaHat(i));
        % Left truncated Exp quantile.
        expSimSData = sort(expSimData);
        muHatpVal = (1./(n-an(i)-bn(i))).*sum(expSimSData((an(i)+1):(n-bn(i))));
        thetaHatpVal = -((muHatpVal-dExp).*(1-a(i)-b(i)))./I0(i);
        FExpMTMFittedpVal = @(s)((exp(-dExp./thetaHatpVal) - exp(-s./thetaHatpVal))
            ...
        ./exp(-dExp./thetaHatpVal));

        nSimData = thSgHat(i,1)+thSgHat(i,2).*norminv(uRand+(1-uRand).*normcdf((dNormal-thSgHat(i,1))./thSgHat(i,2)));
        nSimSData = sort(nSimData);
```

```matlab
        mu1HatLNpVal = (1./(n-an(i)-bn(i))).*sum(nSimSData((an(i)+1):(n-bn(i))));
        mu2HatLNpVal = (1./(n-an(i)-bn(i))).*sum((nSimSData((an(i)+1):(n-bn(i)))).^2);
        theta0pVal = mu1HatLNpVal;
        sigma0pVal = sqrt(mu2HatLNpVal-mu1HatLNpVal.^2);
        x0pVal = [theta0pVal sigma0pVal];
        MTMobjFunpVal = @(x)OnlyLeftTruncatedNormalMTM2Eqns(x,dNormal,[mu1HatLNpVal mu2HatLNpVal],a(i),b(i));
        MTMSolpVal = fsolve(MTMobjFunpVal, x0pVal, opt);
        thSgHatpVal = MTMSolpVal;
        FNormalMTMFittedpVal = @(x)((normcdf((x-thSgHatpVal(1))./thSgHatpVal(2))-normcdf((dNormal-thSgHatpVal(1))./thSgHatpVal(2))) ...
        ./(1-normcdf((dNormal-thSgHatpVal(1))./thSgHatpVal(2))));

        ExpKSADpArray(j,1,i) = KolmogorovSmirnovTS(FExpMTMFittedpVal,expSimSData);
        ExpKSADpArray(j,2,i) = AndersonDarlingTS(FExpMTMFittedpVal,expSimSData',dExp,uExp);

        NormalKSADArray(j,1,i) = KolmogorovSmirnovTS(FNormalMTMFittedpVal,nSimSData);
        NormalKSADArray(j,2,i) = AndersonDarlingTS(FNormalMTMFittedpVal,nSimSData',dNormal,uNormal);
    end
    waitbar(j/n_pVal);
end
for i = 1:length(a)
    KSEntryExp = (sum(ExpKSADpArray(:,1,i) > KS_TS(i,1)))./n_pVal;
    ADEntryExp = (sum(ExpKSADpArray(:,2,i) > AD_TS(i,1)))./n_pVal;

    KSEntryN = (sum(NormalKSADArray(:,1,i) > KS_TS(i,2)))./n_pVal;
    ADEntryN = (sum(NormalKSADArray(:,2,i) > AD_TS(i,2)))./n_pVal;

    ExpKSADpValMatrix(i,:) = [KSEntryExp ADEntryExp];
    NormalKSADpValMatrix(i,:) = [KSEntryN ADEntryN];
end

%=====================================================================%
%                 Premium Calculation!                                %
%=====================================================================%
% Emperical premium.
% Left truncated data with introduced deductible.
xYYT = xYY(xYY > polDed);
n1 = length(xYYT);
% Left truncated and right censored data.
xYYTC = zeros(n1,1);
for i = 1:length(xYYT)
    if (xYYT(i) <= pLim)
        xYYTC(i) = xYYT(i);
    else
        xYYTC(i) = pLim;
    end
end
dataTC = xYYTC.*dmu;
dedPrem = polDed*dmu; % Policy deductible level.
polLimPrem = pLim.*dmu;
EmTCPrem = mean(c.*(dataTC-dedPrem));
EmVar = (mean((dataTC-dedPrem).^2)-((mean(dataTC-dedPrem)).^2))./n1;
EmPremLow = EmTCPrem+norminv(SigLevel./2).*sqrt(EmVar);
EmPremUp = EmTCPrem+norminv(1-(SigLevel./2)).*sqrt(EmVar);
EmPremVec = [EmPremLow EmTCPrem EmPremUp];

alphaHat = 1./ExpThetaHat;
PPVec = zeros(length(a),3);
PLNVec = zeros(length(a),3);
for i = 1:length(a)
    % Probability of beging bigger than deductible.
    pPareto = (ETH./dedPrem).^alphaHat(i);
    pLN = 0.5.*erfc((log(dedPrem-LNTH)-thSgHat(i,1))./(thSgHat(i,2).*sqrt(2)));
```

```

    [ParetoXminU, LNXminU] = ParetoLogNormalXminX(ETH,alphaHat(i),thSgHat(i,1),thSgHat(i,2),polLimPrem,2);
    [ParetoXminD, LNXminD] = ParetoLogNormalXminX(ETH,alphaHat(i),thSgHat(i,1),thSgHat(i,2),dedPrem,2);

    deltaDer = ParetoPDDer(alphaHat(i),dedPrem,polLimPrem,c);
    DVar = (1./n).*(J(i)./(I0(i).^2)).*(alphaHat(i).^2).*(deltaDer.^2);
    PremiumParetoEst = c.*(ParetoXminU(1)-ParetoXminD(1))./pPareto;
    PremuimParetoCILow = PremiumParetoEst+norminv(SigLevel./2).*sqrt(DVar);
    PremuimParetoCIUp = PremiumParetoEst+norminv(1-(SigLevel./2)).*sqrt(DVar);

    [~, ~, ~, LNPremVarDelta] = ...
        MTMSigmaDForPPData(thSgHat(i,1),thSgHat(i,2),a(i),b(i),c,log(dedPrem),log(polLimPrem),LNMoments(i,1),LNMoments(i,2));
    LNPremVar = LNPremVarDelta./n;
    PremiumLNEst = c.*(LNXminU(1)-LNXminD(1))./pLN;
    PremuimLNCILow = PremiumLNEst+norminv(SigLevel./2).*sqrt(LNPremVar);
    PremuimLNCIUp = PremiumLNEst+norminv(1-(SigLevel./2)).*sqrt(LNPremVar);

    PPVec(i,:) = [PremuimParetoCILow PremiumParetoEst PremuimParetoCIUp];
    PLNVec(i,:) = [PremuimLNCILow PremiumLNEst PremuimLNCIUp];
end

%==========================================================================%
%                      Preparation for Plots!                              %
%==========================================================================%

ffs = 10; % Universal Figure Font Size!
for w = 1:n
    sql(w) = (w-0.5)./n; % Standard quantile level - (i-.5)/n.
    eQ(w) = dExp + ExpVaR(sql(w),1); % Standard truncated Exp quantile.
    nQ(w) = norminv(sql(w)+(1-sql(w)).*normcdf(dNormal)); % Standard truncated normal quantile.
end
fittedDataExp = zeros(n,length(a));
% This "n-by-length(a)" matrix stores the fitted data along columns.
fittedDataN = zeros(n,length(a));
% This "n-by-length(a)" matrix stores the fitted data along columns.
for i = 1:length(a)
    fittedDataExp(:,i) = dExp + ExpVaR(sql,ExpThetaHat(i));
    fittedDataN(:,i) = thSgHat(i,1) + thSgHat(i,2).*norminv(sql+(1-sql).*normcdf((dNormal-thSgHat(i,1))./thSgHat(i,2)));
end

%==========================================================================%
%                      Quantile-quantile plots!                            %
%==========================================================================%
xla = 13; xrb = 19;
xVal1 = (xla):0.001:(xrb-1.0);
xVal2 = (xla):0.001:(xrb);
xValL = length(xVal2);
for i = 1:length(a)
    figure(0.*length(a)+i+2)
    subplot(1,2,1); axis equal;
    plot(fittedDataExp(:,i), sort(expData),'*b'); hold on;
    line(xVal2,ones(xValL,1).*log(dedPrem./ETH),'LineStyle','--','Color','green','LineWidth',1.0); hold on;
    line(xVal2,ones(xValL,1).*log(polLimPrem./ETH),'LineStyle','--','Color','green','LineWidth',1.0); hold on;
    line(xVal1,xVal1,'Color','red','LineWidth',1.5); hold off;
    xlim([xla xrb]); ylim([xla xrb]);
    text(17.4,log(dedPrem./ETH)+0.20,'$\log(d^{*}/x_{0})$','Interpreter','Latex','FontSize',ffs);
    text(17.5,log(polLimPrem./ETH)+0.20,'$\log(u/x_{0})$','Interpreter','Latex','FontSize',ffs);
    xlabel(['MIM' num2str(i)' ' Fitted Exponential Quantiles'], 'Interpreter','Latex','FontSize',ffs);
```

```
    ylabel('$\log\left(\frac{data}{x_{0}}\right)$', 'Interpreter','Latex','FontSize',ffs);
    subplot(1,2,2);
    plot(fittedDataN(:,i), sort(nData),'*b'); hold on;
    line(xVal2,ones(xValL,1).*log(dedPrem),'LineStyle','--','Color','green','LineWidth',1.0); hold on;
    line(xVal2,ones(xValL,1).*log(polLimPrem),'LineStyle','--','Color','green','LineWidth',1.0); hold on;
    line(xVal1,xVal1,'Color','red','LineWidth',1.5); hold off;
    xlim([xla xrb]); ylim([xla xrb]);
    text(fittedDataN(1,i)-0.0,log(dedPrem)+0.20,'$\log(d^{*})$','Interpreter','Latex','FontSize',ffs);
    text(fittedDataN(1,i)-0.0,log(polLimPrem)+0.20,'$\log(u)$','Interpreter','Latex','FontSize',ffs);
    xlabel(['MTM' num2str(i)' ' Fitted Normal Quantiles'], 'Interpreter','Latex','FontSize',ffs);
    ylabel('$\log\left(data\right)$', 'Interpreter','Latex','FontSize',ffs);
end
format long
Estimators = [alphaHatMLE MLESol; 1./ExpThetaHat thSgHat];
KS_GOF = round(1000.*[KS_TS ExpKSADpValMatrix])./1000;
AD_GOF = round(1000.*[AD_TS NormalKSADpValMatrix])./1000;
PPrem = ([EmPremVec;PPVec])./dmu;
LNPrem = ([EmPremVec;PLNVec])./dmu;

delete(h);
toc;
end

%=====================================================================%
%                        Auxiliary Functions!                         %
%=====================================================================%

function ExQ = ExpVaR(u, theta)
    ExQ = -theta.*log(1-u);
end

function f = OnlyLeftTruncatedNormalMTM2Eqns(x,d,muHat,a,b)
c1Fun = @(x,y,u)(norminv(u+(1-u).*normcdf((d-x)./y)));
c2Fun = @(x,y,u)((norminv(u+(1-u).*normcdf((d-x)./y))).^2);
c1 = @(x,y)((1./(1-a-b)).*integral(@(u)c1Fun(x,y,u),a,1-b));
c2 = @(x,y)((1./(1-a-b)).*integral(@(u)c2Fun(x,y,u),a,1-b));

f1 = x(1)-muHat(1)+c1(x(1),x(2)).*x(2);
f2 = (x(2).^2).*(c2(x(1),x(2))-((c1(x(1),x(2))).^2))-muHat(2)+((muHat(1)).^2);
f = [f1;f2];
end

function [ParetoXminx, LNXminx] = ParetoLogNormalXminX(theta,alpha,mu,sigma,x,k)
%   This program computes the expectation of X minimum x for single parameter
%   Pareto with parameters alpha and theta, and for lognormal with parameters
%   mu and sigma. You may see the formulas from KPW.
    ParetoIExpVec = zeros(k,1);
    LNExpVec = zeros(k,1);
    for i = 1:k
        ParetoIExpVec(i) = ((alpha.*(theta.^i))./(alpha-i)) - ...
            ((i.*(theta.^alpha))./((alpha - i).*(x.^(alpha-i))));
        LNExpVec(i) = exp(i.*mu+((i.*sigma).^2)./2).*0.5.*erfc(-(log(x)-mu-i.*(sigma.^2))./(sigma.*sqrt(2))) ...
            + (x.^i).*(0.5.*erfc((log(x)-mu)./(sigma.*sqrt(2))));
    end
    ParetoXminx = ParetoIExpVec;
    LNXminx = LNExpVec;
end

function KSTestSTAT = KolmogorovSmirnovTS(ModelDis,x)
%=====================================================================%
% This program computes the Kolmogorov-Smirnov Test Statistics for
```

```
% individual data.
%
% INPUT:
%   ModelDis:       Fited distribution function.
%   x:              Sample left truncated and right censored data.
%
% OUTPUT:
%   KSTestSTAT:     KS test statistics.
%==========================================================================%
n = length(x);
FStarVec = feval(ModelDis, sort(x));
KS_Matrix = zeros(n,6);
for i = 1:n
    KS_Matrix(i,1) = FStarVec(i);
    KS_Matrix(i,2) = (i-1)./n;
    KS_Matrix(i,3) = i./n;
    KS_Matrix(i,4) = abs(KS_Matrix(i,1) - KS_Matrix(i,2));
    KS_Matrix(i,5) = abs(KS_Matrix(i,1) - KS_Matrix(i,3));
    KS_Matrix(i,6) = max(KS_Matrix(i,4),KS_Matrix(i,5));
end
KSTestSTAT = max(KS_Matrix(:,6));
end

function ADTestSTAT  = AndersonDarlingTS(ModelDis,x,t,T)
%==========================================================================%
% This program computes the Anderson-Darling Test Statistics for
% individual data which is either both sides truncated, OR, left truncated
% and right censored, OR, single left rtuncated.
%
% INPUT:
%   ModelDis:       Fited distribution function.
%   x:              Sample data, should be in a row vector form.
%   t:              Left truncation or deductible.
%   T:              Right truncation and/or censored point, policy limit.
%
% OUTPUT:
%   ADTestSTAT:     AD test statistics.
%==========================================================================%

n = length(x);
xx = [t x T];
xxx = unique(xx);
k = length(xxx);
FnX = @(w)((sum(x <= w))./n);
% 'FnX' is Empirical Distribution function of x.
ZeroToK = zeros(k-1,1);
OneToK = zeros(k-2,1);
for j = 1:(k-1)
    if (ModelDis(xxx(j+1)) ~= 1)
        ZeroToK(j) = ((1-FnX(xxx(j))).^2).*(log(1-ModelDis(xxx(j))) - log(1-
            ModelDis(xxx(j+1))));
    else
        ZeroToK(j) = ((1-FnX(xxx(j))).^2).*log(1-ModelDis(xxx(j)));
    end
end
for j = 1:(k-2)
    OneToK(j) = ((FnX(xxx(j+1))).^2).*log(ModelDis(xxx(j+2))./ModelDis(xxx(j+1)));
end
ADTestSTAT = -n.*ModelDis(T) + n.*sum(ZeroToK) + n.*sum(OneToK);
end

function ParetoPremDelta = ParetoPDDer(alpha,d,u,c)
ParetoPremDelta = -(c.*d)./((alpha-1).^2) ...
    +((c.*d)./(alpha-1)).*((d./u).^(alpha-1)).*((1./(alpha-1))-log(d./u));
end

function [SigmaMatrix, DMatrix, Prod, PremVar] = MTMSigmaDForPPData(theta,sigma,a,b
    ,c,d,u,mu1,mu2)
```

```
dz = (d-theta)./sigma;
% Calculation of the trimmed moments variance-covariance matrix.
h1 = @(x,y)((min(x,y)-x.*y)./(normpdf(norminv(y+(1-y).*normcdf(dz))).*normpdf(norminv(x+(1-x).*normcdf(dz)))));
h2 = @(x,y)(((min(x,y)-x.*y).*norminv(x+(1-x).*normcdf(dz))) ...
    ./(normpdf(norminv(y+(1-y).*normcdf(dz))).*normpdf(norminv(x+(1-x).*normcdf(dz)))));
h3 = @(x,y)(((min(x,y)-x.*y).*norminv(x+(1-x).*normcdf(dz)).*norminv(y+(1-y).*normcdf(dz))) ...
    ./(normpdf(norminv(y+(1-y).*normcdf(dz))).*normpdf(norminv(x+(1-x).*normcdf(dz)))));

sigma112 = ((sigma.^2).*((1-normcdf(dz)).^2)./((1-a-b).^2)).*integral2(h1,a,1-b,a,1-b);
sigma122 = 2.*theta.*sigma112 + (2.*(sigma.^3).*((1-normcdf(dz)).^2)./((1-a-b).^2)).*integral2(h2,a,1-b,a,1-b);
sigma222 = 4.*(theta.^2).*sigma112 + ...
    (8.*theta.*(sigma.^3).*((1-normcdf(dz)).^2)./((1-a-b).^2)).*integral2(h2,a,1-b,a,1-b) + ...
    (4.*(sigma.^4).*((1-normcdf(dz)).^2)./((1-a-b).^2)).*integral2(h3,a,1-b,a,1-b);

% Calculation of ck's
fc1 = @(x)(norminv(x+(1-x).*normcdf(dz)));
fc2 = @(x)((norminv(x+(1-x).*normcdf(dz))).^2);

c1 = (1./(1-a-b)).*integral(fc1,a,1-b);
c2 = (1./(1-a-b)).*integral(fc2,a,1-b);

% Calculation of derivatives of ck's
fdc1 = @(x)((1-x)./normpdf(norminv(x+(1-x).*normcdf(dz))));
fdc2 = @(x)(((1-x).*norminv(x+(1-x).*normcdf(dz)))./normpdf(norminv(x+(1-x).*normcdf(dz))));

dc1t = -((normpdf(dz))./(sigma.*(1-a-b))).*integral(fdc1,a,1-b); % Derivative of c1 w.r.t. theta.
dc1s = -(((d-theta).*normpdf(dz))./((sigma.^2).*(1-a-b))).*integral(fdc1,a,1-b); % Derivative of c1 w.r.t. sigma.
dc2t = -((2.*normpdf(dz))./(sigma.*(1-a-b))).*integral(fdc2,a,1-b); % Derivative of c2 w.r.t. theta.
dc2s = -((2.*(d-theta).*normpdf(dz))./((sigma.^2).*(1-a-b))).*integral(fdc2,a,1-b); % Derivative of c2 w.r.t. sigma.

% Calculation of f11, f12, f21, and f22.
f11 = 1+sigma.*dc1t;
f12 = c1+sigma.*dc1s;
f21 = dc2t-2.*c1.*dc1t;
f22 = dc2s-2.*c1.*dc1s;

% Calculation of the gradient matrix, D.
K = 0.5.*(sqrt((c2-c1.^2)./(mu2-mu1.^2)));
d21 = -(K.*(2.*f11.*mu1.*(c2-c1.^2)+f21.*(mu2-mu1.^2))) ...
    ./(f11.*((c2-c1.^2).^2)+K.*(mu2-mu1.^2).*(f11.*f22-f12.*f21));
d22 = (K.*f11.*(c2-c1.^2)) ...
    ./(f11.*((c2-c1.^2).^2)+K.*(mu2-mu1.^2).*(f11.*f22-f12.*f21));
d11 = (1-f12.*d21)./f11;
d12 = -(f12.*d22)./f11;

PremDeltaMatrix = LNPremGradientVec(theta,sigma,d,u,c);

SigmaMatrix = [sigma112 sigma122; sigma122 sigma222];
DMatrix = [d11 d12; d21 d22];
Prod = (DMatrix*SigmaMatrix*DMatrix');
PremVar = (PremDeltaMatrix*Prod*PremDeltaMatrix');
end

function PremiumDeltaMat = LNPremGradientVec(theta,sigma,d,u,c)
Tz = (u-theta)./sigma;
tz = (d-theta)./sigma;
```

```
K0t = 1-normcdf(tz);
K0T = 1-normcdf(Tz);
H1 = normcdf(Tz-sigma)-normcdf(tz-sigma);

dH1t = -(1./sigma).*normpdf(Tz-sigma)+(1./sigma).*normpdf(tz-sigma);
dK0Tt = (1./sigma).*normpdf(Tz);
dK0tt = (1./sigma).*normpdf(tz);

dgt = c.*(((exp(theta+0.5.*sigma.^2).*H1+exp(theta+0.5.*sigma.^2).*dH1t+u.*dK0Tt-d.*dK0tt).*K0t ...
    -(exp(theta+0.5.*sigma.^2).*H1+u.*K0T-d.*K0t).*dK0tt)./(K0t.^2));

dH1s = -normpdf(Tz-sigma).*((Tz./sigma)+1)+normpdf(tz-sigma).*((tz./sigma)+1);
dK0Ts = (Tz./sigma).*normpdf(Tz);
dK0ts = (tz./sigma).*normpdf(tz);

dgs = c.*(((sigma.*exp(theta+0.5.*sigma.^2).*H1+exp(theta+0.5.*sigma.^2).*dH1s+u.*dK0Ts-d.*dK0ts).*K0t ...
    -((exp(theta+0.5.*sigma.^2).*H1+u.*K0T-d.*K0t).*dK0ts))./(K0t.^2));

PremiumDeltaMat = [dgt dgs];
end
```

Curriculum Vitae

Full Name: Chudamani Poudyal
Place of Birth: Dhungesanghu, Taplejung, Nepal

Education

2018 **Ph.D., Mathematics,** *University of Wisconsin-Milwaukee.*
Concentrations: Actuarial Science; Statistics.
Dissertation Title: Robust Estimation of Parametric Models for Insurance Loss Data.
Advisor: Professor Vytaras Brazauskas.

2013 **M.S., Mathematics,** *New Mexico State University.*
Passed Ph.D. qualifying exam in Algebra, Real Analysis, and Topology.

2009 **M.A., Mathematics,** *Tribhuvan University*, Nepal.
Thesis Title: Bottleneck Just-in-Time Sequencing for Mixed-Model Production Systems.
Advisor: Professor Tanka Nath Dhamala.

2004 **B.Ed., Mathematics Education,** *Tribhuvan University*, Nepal.

Research Interests

Actuarial Science; Computational Statistics; Financial Mathematics; Quantitative Risk Management; Robust Statistics; Stochastic Optimization Problems.

Actuarial Education

Society of Actuaries, Schaumburg, IL.
Professional Examinations Passed: P; FM; MFE; C.
VEE Credits Earned: Applied Statistics; Corporate Finance; and Economics.

Experience

2013–2018 **Graduate Teaching Assistant,** *University of Wisconsin-Milwaukee.*
2017 **Actuarial Summer Intern,** *CUNA Mutual Group*, Madison, WI.
2011–2013 **Graduate Teaching Assistant,** *New Mexico State University*.
2010–2011 **Lecturer,** *Malpi Institute (affiliated with University of Cambridge, UK)*, Nepal.
2009–2011 **Assistant Lecturer,** *Tribhuvan University*, Nepal.

Publications

Dhamala, T. N., Khadka, S. R., and Poudyal, C. (2011). Optimal bottleneck mixed-model just-in-time production sequences, *Journal of Institute of Science and Technology*, 17, 81–102. Tribhuvan University, Nepal.

Poudyal, C. (2009). A concept on computational complexity theory, *Epsilon-Delta Yearly Mathematical Magazine*, 5, 85–91. Tribhuvan University, Nepal.

Working Paper

Poudyal, C. and Brazauskas, V. (2018). T- and W-estimation for claim severity data and models. *In preparation*.

Professional Society Memberships

2016–present Member, American Statistical Association (ASA).

2013–present Member, American Mathematical Society (AMS).

2012–present Life Member, Association of Nepalese Mathematicians in America (ANMA).

2009–present Life Member, Nepal Mathematical Society (NMS).

Conference Presentation

2017 T-Estimation for Insurance Loss Data, *52nd Actuarial Research Conference*, Atlanta, GA.

Professional Activities

Referee For: Decision (2016); Insurance: Mathematics and Economics (2017); Scandinavian Actuarial Journal (2017).

2016 Attended *Joint Statistical Meetings*, Chicago, IL.

2015 Attended *3rd Annual Midwest Actuarial Student Conference*, University of Michigan, Ann Arbor, MI.

2015 Attended *SOA Actuarial Teaching Conference*, Indianapolis, IN.

2013 Attended *Data Intensive Summer School*, University of Texas at El Paso, El Paso, TX.

2010 Attended *CIMPA-UNESCO-NEPAL Research School on Number Theory in Cryptography and its Applications,* Nepal.

Honors and Awards

2018 **Mark Lawrence Teply Award** - *in recognition of research potential, Department of Mathematical Sciences, University of Wisconsin-Milwaukee.*

2017 **1st Place Prize, Student Presentation Competition,** *52nd Actuarial Research Conference.*

2017 **Research Excellence Award,** *Department of Mathematical Sciences, University of Wisconsin-Milwaukee.*

2017 **Actuarial Research Conference Travel Grant,** *Society of Actuaries, Schaumburg, IL.*

2017 **Graduate Student Travel Award,** *University of Wisconsin-Milwaukee.*

2013–2018 **Chancellor's Graduate Student Award,** *University of Wisconsin-Milwaukee.*

2011–2013 **Erasmus Mundus Europe Asia Scholarship,** *European Commission.*

Scholarship for two years to attend ALGANT Master Program at the University of Milan, Italy.