Journal of Statistical Planning and Inference 141 (2011) 345–358

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Penalized likelihood estimators for truncated data

Eric W. Cope *

IBM Research – Zurich, Säumerstrasse 4, CH-8803 Rüschlikon, Switzerland

ARTICLE INFO

Article history:
Received 11 December 2009
Received in revised form
9 June 2010
Accepted 9 June 2010
Available online 16 June 2010

Keywords:
Maximum likelihood estimation
Linear penalty function
M-estimator
Location–scale family
Operational risk

ABSTRACT

We investigate the performance of linearly penalized likelihood estimators for estimating distributional parameters in the presence of data truncation. Truncation distorts the likelihood surface to create instabilities and high variance in the estimation of these parameters, and the penalty terms help in many cases to decrease estimation error and increase robustness. Approximate methods are provided for choosing *a priori* good penalty estimators, which are shown to perform well in a series of simulation experiments. The robustness of the methods is explored heuristically using both simulated and real data drawn from an operational risk context.

1. Introduction

Maximum likelihood estimation is known to be asymptotically efficient in many statistical applications, and yet for small samples it can often produce estimators of high variability. This is true in particular when distributional parameters are estimated on the basis of truncated data, in which case the contours of the likelihood function around the maximum may be very oblong in shape, resulting in estimators that may range over a wide set of possible values. Truncated data arise in several applications, such as in financial risk analysis, where a loss distribution must be estimated from data that are subject to a lower reporting threshold. For example, a prominent source of operational loss data for the banking industry, the Operational Riskdata eXchange (ORX) consortium, compiles a database of losses exceeding a fixed threshold level (€20,000) from a large number of international banks. Based on such data, as well as other information sources, certain banks are required under the Basel II Accord (Basel Committee for Banking Supervision, 2004) to measure potential losses at the 99.9th percentile of the annual total loss distribution. In order to meet this regulatory standard, banks must estimate parametric loss severity distributions from a set of left-truncated consortium data. This has been the motivating application for the present work.

This paper addresses some instabilities that are present in the estimation of distributional parameters in the presence of data truncation and model error, and proposes a set of penalty functions for improving estimation accuracy and robustness. These penalty functions are in the spirit of shrinkage estimators such as ridge regression and the lasso (Hastie et al., 2009), which can greatly reduce estimator variance at the price of a small increase in bias. However, rather than always bias the estimators toward zero, the proposed penalty functions are general linear functions of the parameters, that are calibrated according to observed characteristics in the contours of the loss function being minimized. We show how to estimate an appropriate size and direction of the penalty vector, and compare the various estimators in terms of mean

* Tel.: +41 44 724 8423; fax: +41 44 724 8953.
E-mail address: erc@zurich.ibm.com

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.06.010

square error and robustness. Robustness is a key issue in operational risk, as the loss severity distributions are heavy-tailed, and the parameter estimates can be highly sensitive to the most extreme loss values in the data set.

Our use of penalized likelihood estimators also contrasts with other examples in the statistical literature. Penalty functions have also been exploited in a nonparametric maximum likelihood estimation context to impose smoothness conditions on a functional estimate (Eggermont and LaRiccia, 2001), which effectively reduces the variability of the estimate at the price of a small degree of bias. However, this approach is quite distinct from the parametric estimation setting that we address here. In addition, although the estimators are M-estimators of the parameters and achieve robust outcomes, they are distinct from the construction of M-estimators using Huber or other functions that has been extensively applied in the robust estimation literature (Tukey, 1977; Huber, 1981; Hampel et al., 1986). Although the estimator does not improve upon maximum likelihood estimators according to formal criteria of robustness, such as the breakdown point and the gross error sensitivity (Serfling, 2002), we shall provide simulation evidence of increased estimator robustness in the face of varying degrees of data contamination. The use of penalty functions is intended to counteract the particular distortions introduced by data truncation in the shape of the likelihood function, and not as a means to make the estimators resistant to outliers.

Our investigation focuses on distributions that are location–scale families, either in the raw scale or the logarithmic scale of the data such as the Lognormal and Loglogistic distributions, and that are subject to data truncation at various percentile points of the distribution. Let $\boldsymbol{\theta}=(\mu,\sigma)$ be a vector of location and scale parameters, and let $\boldsymbol{x}=x_1,\ldots,x_n$ denote a set of data drawn independently from a distribution $F(\cdot;\boldsymbol{\theta})$ that is either left- or right-truncated at a level t, such that the sampling distribution is $\tilde{F}(\cdot;\boldsymbol{\theta})$, i.e.,

$$\tilde{F}(x;\boldsymbol{\theta})=\begin{cases}\dfrac{F(x;\boldsymbol{\theta})-F(t;\boldsymbol{\theta})}{1-F(t;\boldsymbol{\theta})},\ x\geq t & \text{for left}-\text{truncated data,}\\ \dfrac{F(x;\boldsymbol{\theta})}{F(t;\boldsymbol{\theta})},\ x\leq t & \text{for right}-\text{truncated data.}\end{cases} \tag{1}$$

The following sections describe the maximum likelihood methods for fitting the parameters $\boldsymbol{\theta}$, and introduce penalty functions that are appropriate for each of these fitting procedures. A series of simulation experiments and a real data example taken from an operational risk consortium demonstrate the effectiveness of the penalty methods in reducing error and increasing robustness.

2. Distributional fitting and penalty functions

In maximum likelihood estimation (MLE) with truncated data, the objective function is

$$L_n(\boldsymbol{\theta};\boldsymbol{x})=-\frac{1}{n}\sum_{i=1}^{n}[\log(f(x_i;\boldsymbol{\theta}))-\log(\overline{F}(t;\boldsymbol{\theta}))]\equiv-\frac{1}{n}\sum_{i=1}^{n}l(\boldsymbol{\theta};x_i), \tag{2}$$

where $f(x_i;\boldsymbol{\theta})$ is the density function of the sample value at $\boldsymbol{\theta}$, and $\overline{F}(t;\boldsymbol{\theta})=1-F(t;\boldsymbol{\theta})$ for left-truncated data, and $\overline{F}(t;\boldsymbol{\theta})=F(t;\boldsymbol{\theta})$. Note that the objective function is expressed so that it should be minimized in order to determine the estimate $\hat{\boldsymbol{\theta}}$. Maximum likelihood estimators are widely used due to their well-known asymptotic efficiency properties; however, they can suffer from high sensitivity to sample outliers and a lack of robustness to model misspecification (Hampel et al., 1986; Serfling, 2002).

We can see how the estimates are distributed from Fig. 1, which displays the maximum likelihood estimates for 1000 random samples of size 1000 from a Lognormal distribution with parameters $\mu=10$ and $\sigma=2$. The variance of these estimates is such that large estimates of the scale parameter σ are quite possible. In risk applications, this inflation can have drastic consequences, as regulatory capital estimates are highly sensitive to the extreme quantiles of the loss severity distribution, and especially to the value of the scale parameter. For example, a bank that experiences 10 losses per year that are distributed as Lognormal(10, 2) would estimate capital at the 99.9th percentile of the distribution of total annual losses as \$39 million if they were to actually estimate the parameters to be (10, 2). However, if they were instead to estimate the parameters to be (9.43, 2.24), the predicted capital would be \$52 million. This represents a shift from the true value within the typical range of estimation error, based on a sample size of 1000 from the truncated distribution.

One motivation for introducing a penalized likelihood function is to reduce the variance and estimation error associated with these fitting routines as well as to reduce the chance of grossly overestimating the scale parameters. Although in risk applications, methods based on extreme value theory (EVT) are very popular, in particular the peaks-over-threshold (POT) method of fitting generalized Pareto distributions to the data exceeding high quantile level of the distribution (Embrechts et al., 1997), studies have shown that for many distributions, such as the Lognormal and g-and-h distribution (Hoaglin, 1985; Dutta and Perry, 2006), straightforward application of EVT methods can lead to highly inaccurate quantile estimators (Degen et al., 2007). In such cases, maximum likelihood fitting of distributional parameters remains a viable means of estimating quantiles. Of course, the utility of efficient parameter estimation is not confined to the estimation of any one functional of the distribution such as quantiles; rather, quantile estimation is used merely as an example of such a function that is relevant to certain estimation contexts.