# Kernel Density Estimation for Claim Size Distributions Using Shifted Power Transformation

### K. M. Sakthivel<sup>1</sup>, C. S. Rajitha<sup>2</sup>

Department of Statistics, Bharathiar University, Coimbatore 641046, Tamilnadu, India

**Abstract:** This paper presents density estimation of univariate claim severity distributions using kernel density estimation. We applied transformations to data prior to implement kernel density estimation so as to ensure the data is symmetry for the purpose of applying Gaussian methods. The paper presents non-parametric method of obtaining density for univariate claim severity distributions with goodness fits analysis for Danish insurance data on fire loss claims.

Keywords: Kernel density; bandwidth; Power transformation; Loss modeling; Cross validation

#### 1. Introduction

Kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. Further it is a fundamental data smoothing techniques to obtain the distributional form using finite sample observations. In non-life insurance, the cost of individual claims follows right skewed distributions due to lots of small claims while only a few large claims. When the data under study is skewed, it is advisable not to use the kernel density directly to get the approximating probability density function for claim costs. Hence, the transformed kernel density estimator is most appropriate for asymmetric claim data.

Bivariate data have total claims will be in two different mutually exclusive claim cost depends on policy and its coverage. Bolance[3] have studied estimation of correlations between claim cost types under bivariate skew normal and bivariate normal distributions. Liu[9] fitted the bivariate lognormal and log-skew-normal distributions for positive and right-skewed claim severity data using bivariate kernel density estimation.

Denault[6], Dhaene[7], Panjer[11] and Wang[18] studied the necessity of density estimation in insurance for the purpose of pricing and capital allocation. Klugman[8] have provided excellent substance on the estimation of univariate and bivariate claims distribution models in insurance. McNeil[10] provided possible risk measures for loss distributions. In this paper, we studied probability density estimation of positive claim costs of 2167 fire losses for the period 1980 to 1990 from a Danish Insurance company Copenhagen Reinsurance.

In section 2, we present the kernel density estimator and the transformed kernel density estimator. In section 3, we present measure of goodness of fit for density estimation. Then, in section 4, we perform the transformed kernel density estimation for the above mentioned data. Lastly, we present the results and conclusions.

#### 2. Kernel Density Estimation

#### 2.1 Classical kernel density estimation

Let  $x_1, x_2, ..., x_n$  be *n* independent and identically distributed random variables from the density estimator is given by

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right).$$
(1)

where k is the kernel and h is the bandwidth (Wand[16]). The bandwidth is h used to smoothing estimated density curve. In this paper, we use the Gaussian kernel which is symmetric with mean and it is given below:

$$K(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$$

The practical implementation of the kernel density estimator requires the selection of the bandwidth h. The unique and best method to choose bandwidth parameters is still an ongoing research. In this paper, we used the rule of thumb bandwidth proposed by Silverman[14] and Unbiased and biased cross-validation methods proposed by Sheather and Jones[13] is also used for selection bandwidth h but the data under considered is asymmetric therefore obtain the optimal value of bandwidth h is not possible under asymmetric setup. Hence we consider the transformation of claim severity data for symmetric purpose.

#### **2.2 Data Transformation**

The claims severity data is mostly follows heavy tail distributions. Many authors have contributed for the heavy tail distributions for the application kernel estimation [see Bolance[1]-[5] and Wand[17]) Further they have proposed number of transformed kernel estimation(TKE) methods. In this paper, we used the shifted power transformation family proposed in Wand[16]

$$T_{\lambda}(x) = \begin{cases} (x + \lambda_1)^{\lambda_2} sign(\lambda_2) \\ ln(x + \lambda_1) \end{cases}$$
(2)

Where  $\lambda = (\lambda_1, \lambda_2), \quad \lambda_1 \ge -\min(x_i, i = 1, ..., n)$  and  $\lambda_2 \le 1$  for right skewed data.

Therefore, the new transformed data have been generated from a symmetric random variable. Hence the application of the classical kernel estimation method to the transformed data will produce the reliable result on density estimation.

For the classical kernel density estimator for transformed data is

$$\hat{f}_{Y}(y) = \frac{1}{n} \sum_{i=1}^{n} K_{b}(y - y_{i})$$
(3)

Where K is the kernel function and b is the bandwidth. The original density estimator can be obtained by using the transformed density estimator:

$$\hat{f}_{X}(x,\lambda) = T_{\lambda}'(x)\hat{f}_{Y}(y) = \frac{T_{\lambda}'(x)}{n}\sum_{i=1}^{n}K_{b}\left\{T_{\lambda}(x) - T_{\lambda}(x_{i})\right\}$$
(4)

Here the transformation parameters,  $\lambda = (\lambda_1, \lambda_2)$  and the bandwidth, *b* is necessary for the implementation of transformation approach.

## **2.3** Criteria for selection of the bandwidth and the transformation parameters

The selection of values for transformation parameters,  $\lambda = (\lambda_1, \lambda_2)$  is subject to zero skewness for the transformed data. Further it also minimizing the mean integrated square error (MISE) of estimator of transformed density

$$MISE_{Y}(\hat{f}_{Y}) = E_{Y} \left[ \int_{-\infty}^{\infty} (\hat{f}_{Y}(y) - f_{Y}(y)^{2}) dy \right]$$
(5)

The transformation parameters that minimize asymptotically  $MISE_Y$  will also minimize  $MISE_X$  of  $\hat{f}_X(x,\lambda)$  in (4). The selection of estimator for the bandwidth *b* can be obtained by using Silverman[14].

#### 3. Measuring the Goodness of Fit

Let  $\hat{f}_{X}(x)$  be an estimate of the density and let  $x_{1}, x_{2}, ..., x_{n}$  be *n* i.i.d random variable. Then the log-likelihood function is

$$\ln \hat{L}(\hat{f}_{X}(.);x_{1},x_{2}...,x_{n}) = \sum_{i=1}^{n} \ln \hat{f}_{X}(x_{i})$$

If the transformation method were used to give estimated density  $\hat{f}_X(x,\hat{\lambda};\hat{b})$  then the estimated log-likelihood function is

$$\ln \hat{L}(\hat{f}_{X}(.);T_{\hat{\lambda}}(.);x_{1},x_{2}...,x_{n}) = \sum_{i=1}^{n} \ln \hat{f}_{X}(x_{i},\hat{\lambda};\hat{b})$$

The integrated square error (ISE) is used to measure for evaluating the quality of kernel density estimators

$$ISE_{X}(\hat{f}_{X}) = \int_{-\infty}^{\infty} (\hat{f}_{X}(x) - f_{X}(x))^{2} dx$$

where  $f_X(x)$  be a kernel estimation of  $f_X(x)$ 

It is proved that minimizing  $ISE_x$  is equivalent to minimizing the cross-validation function:

$$CV_{X} = \int_{-\infty}^{\infty} [\hat{f}_{X}(x)]^{2} dx - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{i}(x_{i})$$
(6)

We can generalize the definition of log-likelihood given previously by providing a statistic that gives more weight to the right tail of the distribution. A weighted log-likelihood can be estimated if weights  $w_i$ , i = 1,...,n are included preceeding each summation term as:

$$\ln_{w} \hat{L}(\hat{f}_{X}(.); x_{1}, x_{2}..., x_{n}) = \sum_{i=1}^{n} w_{i} \ln \hat{f}_{X}(x_{i})$$

If  $w_i, i = 1, ..., n$ , then we would have the usual loglikelihood expression. We can also use some distance from zero as a weight, so that observations that are located close to zero have much less importance than those located in the right tail.

Bolance[3] have tried two different expressions for weights. The first one gives more weight to those observations that are distant from zero. Since our data are always positive. The second form is weighted with a squared distance. The form of the weights is

$$w_{i}^{(1)} = \frac{nx_{i}}{\sum_{i=1}^{n} x_{i}}$$
$$w_{i}^{(2)} = \frac{nx_{i}^{2}}{\sum_{i=1}^{n} x_{i}^{2}}$$

When a transformation is used, the corresponding expression is:

$$\ln_{w} \hat{L}(\hat{f}_{X}(.);T_{\hat{\lambda}}(.);x_{1},x_{2}...,x_{n}) = \sum_{i=1}^{n} w_{i} \ln \hat{f}_{X}(x_{i},\hat{\lambda};\hat{b})$$

Similarly, we can approximate a weighted  $ISE_X$  ( $WISE_X$ ), weighting by x or by  $x^2$ .

$$WISE^{1}x(\hat{f}_{X}) = \int_{-\infty}^{\infty} (\hat{f}_{X}(x) - f_{X}(x))^{2} x \, dx$$
$$WISE^{2}x(\hat{f}_{X}) = \int_{-\infty}^{\infty} (\hat{f}_{X}(x) - f_{X}(x))^{2} x^{2} \, dx$$

that can be approximated with:

$$WCV_{1} = \int_{-\infty}^{\infty} [\hat{f}_{X}(x)]^{2} x dx - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{i}(x_{i}) x_{i}$$
(7)

$$WCV_{2} = \int_{-\infty}^{\infty} [\hat{f}_{X}(x)]^{2} x^{2} dx - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{i}(x_{i}) x_{i}^{2}$$
(8)

Here we used to different methods were used for obtain the parameters. Method 1 reduces mean sequare error and Method 2 deals with zero skewness (i.e.,) symmetric of data.

#### 4. Data Analysis

In this paper, we used the Copenhagen Reinsurance fire losses data expressed in millions of Danish Krone of 2167 observations for the period of 1980 to 1990. It is also

#### Volume 5 Issue 4, April 2016

<u>www.ijsr.net</u>

available in the R software in the name of Danish dataset in QRM package.

 Table 1: Decriptive Statistics of the Claim data set (in millions of Danish Krone)

	Mean	SD	Skewness	Kurtosis	Min	Max
Claim Size	3.385	8.508	18.737	482.198	1	263.250

#### 4.1 Fitting of Kernel densities

In Figure 1, we show histogram and fit of the classical kernel density estimates of claim data. In Figure 2, we show histogram and classical kernel fit for log-transformed claim data



Figure 1: Histograms and classical kernel fit



Figure 2: Histograms and classical kernel fit to logtransformed data



$\lambda = (\lambda_1, \lambda_2)$					
	Method 1	Method2			
Fire losses Claims	(-0.7761,-0.3905)	(-0.7621,-0.35)			

In Figure 3, we show the kernel density estimates for the transformed variable using method 2



Figure 3: Histogram and classical kernel fit to optimally transformed data



Table 3: Log-likelihood and weighted log-likelihoods

	$\ln \hat{L}$	$\ln \hat{L}$	$\ln \hat{L}$
		(weight 1)	(weight 2)
Classical kernel	-4759.31	-8811.86	-17717.35
TKE log transformed	-8838.62	-14306.07	-26754.85
TKE Method1	-3364.00	-8946.35	-21708.02
TKE Method2	-3364.00	-8946.35	-21708.02

 Table 4: Cross-validation

 CV
 WCV1

	CV	WCV1	WCV2
Classical kernel	-0.1790	-1.1355	-0.6002
TKE log transformed	-0.1078	-0.2316	-0.1473
TKE Method1	-0.2564	-0.0735	-0.1602
TKE Method2	-0.2560	-0.0735	-0.1521

#### 4.2 Results of Goodness of fit

In table 4, the results of CV, WCV1 and WCV2 to compare the fit of classical kernel and TKE. Note that the values CV, WCV1 and WCV2 can be negative. The minimum values of CV, WCV1 and WCV2 will indicate the better fit. Hence to Danish fire losses claim data, TKE Methods with shifted power transformation fits better compare to the method of classical kernel density estimation. Finally, for claims severity data with right skewed, TKE with transformation especially with shifted power transformation give better fit compare to classical kernel density estimation.

#### 5. Conclusions

In this paper, we fitted the univariate distributions to Danish data sets on fire losses claim costs. It is observed that the classical kernel density estimates are not providing better fit to the claim size data which is generally right skewed. But the power transformation of classical kernel density provides the better fit and smoothed version of empirical distribution for weighted likelihoods when more weights are assigned to the right tail of the distribution. Minimization of ISE, WISE and Cross-validation is better criteria for the goodness of the estimated densities. Finally we conclude that the shifted

Volume 5 Issue 4, April 2016 <u>www.ijsr.net</u> Licensed Under Creative Commons Attribution CC BY

In Figure 4, we plot the TKE of pdf of the fire losses claims

power transformation works and suits well to fit claim severity distribution which is usually having heavy tails.

#### References

- Bolance,C.(2010) Optimal Inverse Beta(3,3) Transformation in kernel density estimation, SORT Statistics and Operations Research Transaction,34, 223-238.
- [2] Bolance, C., Guillen, M. and Nielsen, J.P. (2009) Transformation kernel estimation of insurance claim cost distribution, in Corazza, M. and Pizzi, C. (Eds), Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer, Roma, 223-231.
- [3] Bolance, C., Guillen, M., Pelican, E. and Vernic, R. (2008) Skewed bivariate models and nonparametric estimation for the CTE risk measure, Insurance:Mathematics and Economics, 43, 386-393.
- [4] Bolance, C., Guillen, M. and Pitt, D. (2014) Nonparametric models for univariate claim severity distributions-an approach using R, UB Riskcenter Working papers Series, 01,1-34.
- [5] Bolance, C., Guillen, M. and Nielsen, J.P., (2003) Kernel density estimation of actuarial loss functions, Insurance: Mathematics and Economics, 32, 19-36.
- [6] Denault, M. (2001): Coherent allocations of risk capital, Working paper, Ecole des H.E.C., Montreal.
- [7] Dhaene, J., Goovaerts, M.J. and Kaas, R. (2003). Economic capital allocation derived from risk measures, North American Actuarial Journal, 7, 44-59.
- [8] Klugman, S.A., Panjer, H.H, and Willmot, G.E., (2008). Loss models: from data to decisions. 3<sup>rd</sup> Edition. John Wiley \$ Sons, New Jersey.
- [9] Liu, Q., Pitt, D., Zhang, X. and Wu, X. (2011) A Bayesian approach to parameter estimation for kernel density estimation via transformations, Annals of Actuarial Science, 5, 181-193.
- [10] McNeil, A.J., Frey. R. and Embrechts, P. (2005) Quantitative risk management: concepts, techniques and tools. Princeton University Press, Princeton Series in Finance, London.
- [11] Panjer, H.H (2002) Measurement of risk, solvency requirements and allocation of capital within financial conglomerates, 27<sup>th</sup> International Congress of Actuaries, Cancun 2002.
- [12] Reiss, R.D., (1981) Nonparametric estimation of smooth distribution functions, Scandinavian Journal of Statistics 8,116-119.
- [13] Sheather, S.J. and Jones, M.C. (1991) A reliable databased bandwidth selection method for kernel density estimation, Journal of the Royal statistical Society, Serial B, 53,683-690.
- [14] Silverman, B.W. (1986) Density Estimation for Statistics and Data Analysis, Chapman & Hall, London.
- [15] Terrel, G.R. (1990) The maximal smoothing principle in density estimation, Journal of the American Statistical Association, 85, 270-277.
- [16] Wand, M.P. and Jones, M.C.(1995) Kernel Smoothing. Chapman & Hall, London.
- [17] Wand, P., Marron, J.S. and Rupport, D. (1991) Transformations in density estimation, Journal of the American Statistical Association, 86, 343-361.

[18] Wang, S. (2002) A set of new methods and tools for enterprise risk capital management and portfolio optimization, Working paper, SCOR reinsurance company.