Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v12n3p691

An application of transformed distribution: length of stay in hospitals

By Harini et al.

Published: 20 November 2019

Electronic Journal of Applied Statistical Analysis
Vol. 09, Issue 01, November 2019, 691-704
DOI: 10.1285/i20705948v12n3p691

An application of transformed distribution: length of stay in hospitals

Harini S.[a], Subbiah M.[b], Srinivasan M.R.[c], and Gallo M. *[d]

[a,c]*Department of Statistics, University of Madras, Chennai-600004, Tamilnadu, India*
[b]*Department of Mathematics, Presidency College, Chennai-600004, Tamilnadu, India*
[d]*Department of Human and Social Sciences, University of Naples - L'Orientale, P.zza S.Giovanni 30, Napoli-80134, Italy*

Published: 20 November 2019

Length of stay in hospitals are mostly characterized as asymmetric, right skewed and leptokurtic in nature. Earlier studies have considered parametric distributions like gamma, Pareto, lognormal for studying length of stay of patients in hospitals. However, in this study we have proposed transformed distributions to be the best choice for characterizing the length of stay. For this study, we have considered paediatric asthma dataset and identified that transformed Weibull-Pareto as the best fit. For a comparative purpose we have also provided the results of gamma, lognormal, and Pareto distributions. Maximum likelihood approach is considered to estimate the unknown parameters of the Transformed distribution followed by goodness of fit tests to examine the suitability of the fitted distributions. The results provide a direction for modelling the length of stay in hospitals due to different medical problems which require hospitalization.

keywords: Tranformed Ditribution, Weibull-Pareto, Length of Stay, Heavy-Tailed, Light-Tailed

1 Introduction

Length of Stay (LOS) in hospital is important for healthcare management due to ever increasing population. The stay of patients in hospital are not normally distributed as the stay varies from individual to individual. The LOS are found to be asymmetric

*Corresponding author: mgallo@unior.it.

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

which is confirmed empirically and stay of patients in hospitals are right skewed in nature (Marazzi et al., 1998, Faddy and McClean, 1999, Papi et al., 2016). The information regarding the LOS takes a high priority by practitioners and healthcare management for strategic planning. The goal of every hospital management is to meet the demand of patients by providing an effective treatment. Hence, every hospital management is keen on studying the duration of stay of patients in hospital.

For modelling the LOS, there is a need to fit an appropriate distribution, which might help and assist in capacity planning. Faddy and McClean (1999) and Gül and Güneri (2015) have stressed the importance of identifying the best fitted distribution for LOS since they are the backbone for modelling stay in hospital. The consequence of errors which incur while an inappropriate distribution is fitted is discussed by Qualls et al. (2010). Therefore, fitting an appropriate distribution and identifying the best fitted for LOS is the primary objective of the paper.

Parametric distributions such as gamma, log normal, logistic, Pareto are widely considered for studying skewed data sets in the field of finance, hydrology, and insurance (Cooke et al., 2014, Gomes and Guillou, 2015, Thomas et al., 2016). However, distribution like gamma can be considered as heavy or light tailed depending upon the shape parameter (Asmussen and Lehtomaa, 2017). The Mean Excess (ME) plot are widely used for distinguishing the tails which might be helpful in understanding the tail behaviour for right skewed datasets (Coles and Powell, 1996, Drees et al., 2003,Cooke et al., 2014) . Therefore, since LOS is right skewed in nature, we have considered ME plot to model the LOS datasets to understand the nature of the tail.

McClean and Millard (1993) have considered log normal and exponential distribution to study the length of stay. Marazzi et al. (1998) and Faddy and McClean (1999) have considered the lognormal, gamma distributions for analysing the LOS datasets. Gardiner et al. (2014) considered lognormal, gamma, Weibull, and Pareto for studying the stay of patients. Singh and Ladusingh (2010) have considered the finite mixture modelling for the inpatient stay in hospital as the dataset was multimodal and convolutive mixture distribution is considered by Ickowicz and Sparks (2017) to study the short and long stay of patients.

Lin et al. (2013) and Sawilowsky (2016) discussed that large samples tend to reject the null hypothesis and the usual univariate distributions might not perform well. Recently, Harini et al. (2018) suggested that Transformed Gamma-Pareto distribution to be the best fit for diabetes LOS dataset. Therefore, we have considered transformed distribution to study paediatric medical speciality. We also highlighted that these distributions are superior for studying skewed LOS datasets when compared to the usual univariate distributions. In this study, transformed distribution such as Beta-Cauchy (Alshawarbeh et al., 2013), Gamma-Pareto (Alzaatreh et al., 2012), Weibull-Pareto (Alzaatreh et al., 2013, Al-Omari et al., 2016) and Gamma-Exponential-Cauchy (Alzaatreh et al., 2016)) are considered for studying paediatric LOS dataset. For the purpose of comparison and to show that transformed distribution is superior, the results of gamma, Pareto and lognormal distributions are also provided.

In case of fitting a distribution, estimation of the parameters play a vital role and most importantly, it is challenging when the data is asymmetric in nature. Gardiner et al.

(2014) considered Bayesian methods for modelling the stay of patients and maximization goodness of fit estimation is considered for fitting generalized Pareto distribution by Luceño (2006). However, the most widely used method is Maximum Likelihood Estimation (MLE) which has been quite often considered for fitting the usual univariate distributions (Faddy et al., 2009, Verburg et al., 2014, Ickowicz and Sparks, 2017). Therefore, for this study, we consider MLE method of estimation for fitting the transformed distributions.

Further, this paper discusses the computational aspect of using the best numerical optimization to estimate the unknown parameters of transformed distributions by comparing the most commonly used methods like Nelder-Mead (NM), Broyden-Fletcher-Goldfarb and Shanno (BFGS) and Limited memory BFGS (L-BFGS) method.We have considered three different goodness of fit measures such as Kolmogorov-Smirnov (KS), Anderson-Darling (AD) and Cramer-von-Mises (CVM) test. The goodness of fit tests are compared to find which suits better and Akaike Information Criteria (AIC) is considered to identify the best distribution for studying paediatric LOS datasets.

The paper is assigned as follows, Section 2 deals with the details of datasets and Section 3 discusses the tail properties. Section 4 details about transformed distributions, MLE method of estimation and numerical optimization procedures. The analysis and results are discussed in Section 5 and Section 6 concludes with the general recommendations.

2 Length of Stay Dataset

In this study dataset represents patients admitted in to paediatric department comprising of asthma complications ($n = 132$) whose stay ranges from 1 to 9 days ((Houchens and Schoeps, 1998)). The spread and frequency of stay for the dataset is presented in Figure 1.

It can be observed from Figure 1 that LOS dataset is asymmetric and right skewed in nature. This is also confirmed from Table 1 since the Moors excess kurtosis exceeds zero and Galton skewness is non-zero. Galton skewness and Moors kurtosis are considered because they are not influenced by the extreme observations.

3 Tail Properties

The knowledge about the tail characteristics are obtained from the ME plot which helps in identifying whether the dataset is heavy-tailed or not (Embrechts and Schmidli, 1994; Coles and Powell, 1996). The mean excess is the tool which can be used to determine the exceedance above a particular level which is commonly referred as a threshold. The sample ME function can be defined as

$$e_n(u) = \frac{\sum_{i=1}^{n}(X - \mu)}{\sum_{i=1}^{n} \mathrm{I}_{((X_i > u))}}$$

where $I = 1$ if $X_i > \mu$ and 0 otherwise. When the plot show an upward trend then, they are heavy-tailed and, it is a light-tailed when they show a downward trend. The

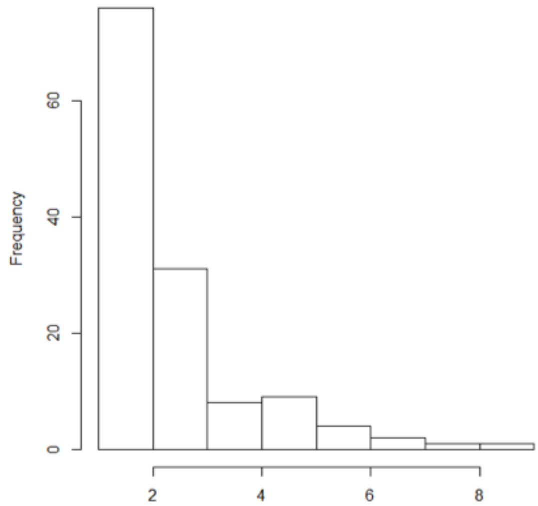

Figure 1: Histogram of Length of Stay in Hospital is asymmetric and skewed.

distribution F is called as heavy-tailed if

$$\int_{-\infty}^{\infty} e^{\lambda x} F(dx) < \infty \quad \text{for all} \quad \lambda > 0$$

if F fails to be heavy-tailed, then they are considered to be light-tailed. Clearly, for any F on $R^+ = [0,)$, all moments are finite.

Then composition scheme of T-X family is as follows $F_X(x)$: $\mathbb{R} \to [0, 1]$, $W : [0, 1] \to \mathbb{R}$, $F_T(t)$: $\mathbb{R} \to [0, 1]$, Now, H is the transformed variable $F_H(x) = F_T \circ W$; $\Re \to [0, 1]$ where ?$\circ$? denotes composition of two functions. The range of the function W is the domain of $F_T(t)$ which is actually the support of random variable T and PDF of transformed variable is obtained by integrating $F_H(h)$. Similarly, T-Y-X family of random variable can be achieved using composition of the CDF of T and X with a quantile of new random variable Y. $F_X(x) : \mathbb{R} \to [0, 1], F_Y(y) : \mathbb{R} \to [0, 1], Q_Y : [0, 1] \to \mathbb{R}, F_T(t) : \mathbb{R} \to [0, 1]$, $F_H(x) = F_T \circ QY \circ F_X; \mathbb{R}x \to [0, 1]$ The role of W is significant in these transformation and its shape influences the choice of T from a family of distributions and vice-versa in some cases. Few choices of W and its relation with T are discussed in Table 2.

The role of W is suitably modified as quantile function of Y to construct $T - Y - X$ family of distribution (Table 2 of Alzaatreh et al., 2016). In this paper, we have considered three members of $T - X$ family (Beta-Cauchy, Gamma-Pareto, and Weibull-Pareto) and one member of $T - Y - X$ family (Gamma-Exponential-Cauchy) to fit the LOS datasets. The details of the transformed distributions are discussed in Table 2 and additional details pertaining to these four distributions are provided in Appendix.

Parameters of the BC, GP, WP, and GEC are estimated using MLE procedure in R. The numerical optimization techniques are adopted for estimating the parameters of transformed distributions which are detailed before proceeding with the estimation

Table 1: Statistical characterization of paediatric Length of Stay in Hospitals

Measures	**Values**
Mean	2.63
Median	2.00
Maximum	9.00
Skewness	1.47
Excess Kurtosis	5.47
Galton Skewness	1.00
Moors Kurtosis	2.63

S.No	**Random Variable**	**Support**	**PDF**	**CDF**
1	X	$\mathcal{A}_X$	$f_X(x)$	$F_X(x)$
2	Y	$\mathcal{A}_Y$	$f_Y(y)$	$F_Y(y)$
3	T	$\mathcal{A}_T$	$f_T(t)$	$F_T(t)$
4	H	$\mathcal{A}_H$	$f_H(h)$	$F_H(h)$

procedures of transformed distribution.

3.1 Numerical Optimization Methods

To estimate the parameters, we consider the numerical optimization technique. Since the likelihood is not of closed form, hence to solve them we consider the optimization techniques to estimate the parameters. As there are plethora of optimization techniques, we consider the most commonly used Nelder-Mead, BFGS and L-BFGS methods (Nocedal, 2004; Moré and Wild, 2009; Nash, 2014).
Nelder proposed the method which is gradient free and it is also considered as a derivative free approach. This method is considered to be efficient in finding local minima in the case of high dimensional, convex and non-linearly constrained problems (Nocedal, 2004). For estimating the parameters, MLE method uses the direct search method and further, they are helpful in maximizing the likelihood of transformed distributions. For every iteration, they get terminated with a new function value satisfying the descent condition to the previous simplex. However, it is to be noted that they have slower convergence property hence the computational time might be higher (Nash, 2014).
Quasi Newton method is helpful since they do not require the computations of Hessian matrix and they attain rate of convergence rapidly when compared to NM method (Nocedal, 2004). We have considered BFGS and L-BFGS methods in this study which are

Table 2: Relation between different forms of W and random variable T for transformed distribution

S.No	**Forms of W**	**Range of W = Domain of $F_T(t)$**
1	$F_X(x)$	$[0,1]$
2	$-log[1-F_X(x)]$	$[0,\infty)$
3	$F_X(x)/(1-F_X(x))$	$[0,\infty)$
4	$log[-log(1-F_X(x))]$	$(-\infty,\infty)$
5	$log[F_X(x)/(1-F_X(x))]$	$(-\infty,\infty)$
6	$-1/log[1-F_X(x)]$	$(-\infty,\infty)$

Table 3: Transformed distribution - using random variables X, T, Y with their support and W is the differentiable and monotonically non-decreasing function.

S.No	**Dist**	**X**	**T**	**Y**	**W**	$\mathcal{A}_X$	$\mathcal{A}_Y$	$\mathcal{A}_T$
1	BC	Cauchy	Beta	NA	$F_X(x)$	$(-\infty,\infty)$	NA	$(0,1)$
2	GP	Pareto	Gamma	NA	$-log[1-F_X(x)]$	$(0,\infty)$	NA	$[0,\infty)$
3	WP	Pareto	Weibull	NA	$-log[1-F_X(x)]$	$(0,\infty)$	NA	$[0,\infty)$
4	GEC	Cauchy	Gamma	Exponential	$Q_Y(y)$	$(0,\infty)$	$[0,\infty)$	$[0,\infty)$

NA: Not Applicable.

widely used Quasi-Newton methods. The gradient of the objective function is required which needs to be provided at every iteration. In this approach, the second order derivative are not required hence they are said to be efficient. In this approach, the upper and lower limits for the parameters for maximizing the likelihood needs to be provided and, it should be noted that this approach converges rapidly and are robust in nature when compared to NM.

L-BFGS method is helpful in computing hessian matrices as they just save few length of vectors. In this approach, similar approach of BFGS is carried out, but the details on curvature are only considered from the recent iteration. The approach of L-BFGS is identical to that of BFGS except the way of handling the Hessian approximation. Computationally it is also indicated that L-BFGS method is robust, rapid than other conjugate gradient methods. Among BFGS and L-BFGS, it has been highlighted that L-BFGS is more efficient since it is limited memory Quasi Newton methods. It is helpful when there is a need to solve large problems as it is difficult to compute the Hessian matrices.

In this study, we have compared NM, BFGS and L-BFGS optimization methods for estimating the parameter. The number of iterations carried out by each method to estimate the parameters are detailed in analysis and result section. These methods are helpful in the estimating the parameters of transformed distribution which are discussed in the next section.

3.2 Parameter Estimation of Transformed Distributions

To estimate the parameters of BC, the Cauchy distribution (θ and λ) parameters are estimated from the data. Then, these estimates are considered as initial values to estimate the BC parameters. In addition, suitable initial values are considered for the parameters (α, β) of the Beta distribution for each of the numerical optimization methods. Finally, distribution of BC parameters are estimated starting from these initial values. Smith (1985) method is considered to estimate the parameters of Pareto distribution for both GP(α, c, θ) and WP(c,β, θ). Subsequently, they are helpful in estimating the parameters which are discussed below

Gamma - Pareto

1. Estimate shape parameter θ of the Pareto distribution which is the sample minimum $x_{(1)}$
2. Apply transformation $Z_i = log(x_i/x_{(1)}$ for all $x_i \neq x_{(1)}$
3. Initial value for the parameters α and c of the Gamma distribution are computed as $\alpha_0 = \overline{z}^2/s_z^2$ and $c_0 = s_z^2/\overline{z}$, where $\overline{z}$ and s_z^2 are the sample mean and variance.

Weibull - Pareto

1. Estimate shape parameter θ of the Pareto distribution which is the sample minimum $x_{(1)}$
2. Apply transformation for $Z_i = \theta e^{x_i}$ for all $x_i \neq x_{(1)}$
3. Initial value for the parameters c and β are computed as $c_0 = \frac{\Pi}{\sqrt{6} s_Z^2}$ and $\beta_0 = exp(-\overline{z} - \frac{\gamma}{c_0})$, where $\overline{z}$ and s_Z^2 are the sample mean and variance.

Then, parameters of the GP and WP are estimated using these initial values. In a similar way, an iterative scheme is adopted in estimating the parameters of the GEC (α, β, θ) which involves the following steps;

1 Initial value of θ is considered as θ_0.

2 Apply the transformation to the data as $\mathrm{U}_i = -log[0.5 - \pi^1 arctan(x_i/\theta_0)]$ for $i = 1...n$, where n is the sample size.

3 Initial values for the parameters α and β of the Gamma distribution are computed as $\alpha_0 = \overline{\mathrm{U}}^2/s_{\mathrm{U}}^2$ and $\beta_0 = s_{\mathrm{U}}^2/\overline{\mathrm{U}}$, where $\overline{\mathrm{U}}$ and s_{U}^2 are sample mean and variance.

4 Then, the parameters of the GEC (α, β and θ) are estimated starting from these initial values. The natural sequence after estimating parameters is to check the adequate fit, hence most widely used goodness of fit such as Kolmogorov-Smirnov (KS), Anderson-Darling (AD) and Cramer-Von-Mises (CVM) test are used.

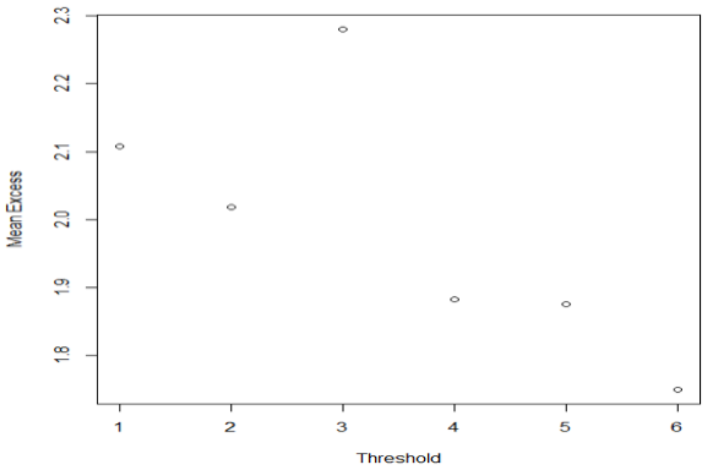

Figure 2: Mean excess plot for distinguishing heavy and light tail for Length of Stay datasets.

3.3 Goodness of Fit Test

The next natural process after estimating the parameter is to assess the goodness of fit and test the hypothesis whether the dataset follows specified distribution. The KS and AD test are based on the empirical cumulative distribution function and test statistic for KS test is $D_n = \sup_x |F_n(x) - F_X(x)|$, D_n is usually referred as KS distance. Similarly, the AD distance is referred as A^2 which is defined as $A^2 = -n - S$ and $S = \sum_{i=1}^{n} \frac{2i-1}{n} \left[ln(F_X(x_i)) + ln(1 - F_X(x_{n+1-i}))\right]$. CVM test is based on minimum distance method and defined as $W^2 = \int_{-\infty}^{\infty} |F_n(x) - F_X(x)|\, dF(x)$. The hypothesis of the distributional form has been rejected at given level of significance if D_n, A^2 and W^2 are greater than critical value obtained from the table. Generally, for most of the studies level of significance are chosen to be 5% to evaluate the hypothesis. If the p-value is greater than the level of significance then the distribution fits the data, else they do not fit (Murdoch et al., 2008).

4 Analysis and Results

As discussed in the methodology section, ME plot is discussed in Figure 2 which shows that the LOS datasets are light-tailed as they have a downward trend. Hence, the tail property of LOS dataset is identified to be light-tailed in this study. We further proceed fitting transformed distributions to both the datasets and the results are discussed in Table 4.

It can be observed from Table 4 that transformed WP distribution is identified as best fit. In general, Table 4 highlights that transformed distributions are the best fit for studying LOS datasets with the least AIC values when compared to the usual univariate

Table 4: Parameter estimates, Number of Iterations, Goodness of Fit tests and AIC for the Paediatric Length of Stay Dataset.

Measures	Gamma	Lognormal	Pareto	GP	BC	WP	GEC
	$\alpha = 3.342$	$\mu = 0.809$	$\alpha = 0.871$	$\alpha = 7.879$	$\lambda = 0.580$	$\beta = 0.848$	$\alpha = 38.829$
Parameter	(0.393)	(0.049)	(0.064)	(1.079)	(1.044)	(0.029)	(6.389)
Estimates	$\beta = 1.272$	$\sigma^2 = 0.559$	$\beta = 1.000$	$c = 0.133$	$\theta = 0.240$	$c = 2.819$	$\beta = 0.091$
(S.E)	(0.161)	(0.034)		(0.018)	(0.038)	(0.127)	(0.015)
				$\theta = 1.00$	$\alpha = 51.6822$	$k = 1.00$	$\theta = 0.218$
					$\beta = 4.4792$		
					(0.1523)		
Number of Iterations							
NM	NA	NA	NA	59	252	76	164
BFGS	NA	NA	NA	39	142	39	85
L-BFGS	NA	NA	NA **22**	**123**	**26**	**42**	
Log Likelihood	-221.075	-217.423	-297.50	-146.566	-218.358	-153.807	-161.851
KS	0.190	0.193	0.348	0.576	0.521	0.385	0.954
(p-value)	(0.001)	(0.001)	(< 0.001)	(0.047)	(0.084)	(0.141)	(0.052)
AD	4.649	4.882	0.807	4.437	5.462	4.481	1.647
(p-value)	(0.004)	(0.003)	(< 0.001)	(0.056)	(0.053)	(0.061)	(0.072)
CVM	3.799	4.882	6.218	2.378	3.367	1.767	2.646
(p-value)	(0.007)	(0.006)	(< 0.001)	(0.061)	(0.057)	(0.081)	(0.060)
AIC	446.151	438.846	617.23	**297.132**	**314.776**	**156.415**	**340.801**

distributions. Pareto distribution did not fit the dataset; similarly, all the three usual univariate distributions failed to fit the dataset in all the three goodness of fit tests KS, AD and CVM.
The L-BFGS method of optimization is found to be efficient and superior when compared to NM and BFGS methods in estimating the parameters of transformed distributions. The time taken by NM method for convergence is twice the iterations of L-BFGS and BFGS method. It needs to be noted L-BFGS methods converges rapidly for both the datasets which can be observed from Table 4. Indeed, BFGS also converge quickly with lesser iterations when compared to NM method.
Among the three measures of goodness of fit, CVM test provides much scope for handling asymmetric datasets as the acceptance level of significance is higher when compared with KS and AD tests. For both the transformed and usual univariate distributions CVM test is superior. Hence, CVM might be an ideal choice for fitting any light-tailed datasets as they provide consistent results when compared to the other goodness of fit tests considered in this study.

5 Conclusion

The study of length of stay in hospital is very much important for the healthcare management. The hospital administrators are highly interested in studying the LOS due to the increasing number of patients. They are also very much interested in understanding the distribution of LOS since it is useful for modelling or prediction of length of stay in hospitals. The main aim of this study is to understand the underlying statistical distribution for length of stay. Further, this study proposes transformed distribution to

be a better choice when compared to the usual univariate distributions.
In this study, transformed distributions are compared with the usual univariate distributions and the result shows that transformed distribution to be more ideal choice when compared to the latter. Earlier study identified Gamma-Pareto as the best for diabetes LOS dataset. However, in this study, it has been identified that Weibull-Pareto as the best fit for paediatric asthma patients. Due to advent of computers, computational aspect also plays a vital role. Hence, for estimating the parameter, numerical optimization has been considered, and this study identified L-BFGS method to be superior as it converges rapidly when compared to NM and BFGS. This will be very much helpful when dealing with large datasets since the time and cost involved might be minimal due to rapid convergence.
In conclusion, it may be noted from the analysis from this study that transformed distribution is an ideal choice for fitting the length of stay in hospitals. However, the choice of transformed distribution might vary based on the speciality or different stay of patients in hospitals. Further studies can be attempted for different LOS datasets of varied specialities using other transformed distributions available in the literature.

5.1 Acknowledgements

Funding for this project was provided by 2017 funds of the University of Naples - L'Orientale (I).

References

Al-Omari, A. I., Al-Nasser, A. D., and Gogah, F. S. (2016). Double acceptance sampling plan for time truncated life tests based on transmuted new weibull-pareto distribution. *Electronic Journal of Applied Statistical Analysis*, 9(3):520–529.

Alshawarbeh, E., Famoye, F., and Lee, C. (2013). Beta-cauchy distribution: some properties and applications. *Journal of Statistical Theory and Applications*, 12(4):378–391.

Alzaatreh, A., Famoye, F., and Lee, C. (2012). Gamma-pareto distribution and its applications. *Journal of Modern Applied Statistical Methods*, 11(1):7.

Alzaatreh, A., Famoye, F., and Lee, C. (2013). Weibull-pareto distribution and its applications. *Communications in Statistics-Theory and Methods*, 42(9):1673–1691.

Alzaatreh, A., Lee, C., Famoye, F., and Ghosh, I. (2016). The generalized cauchy family of distributions with applications. *Journal of Statistical Distributions and Applications*, 3(1):12.

Asmussen, S. and Lehtomaa, J. (2017). Distinguishing log-concavity from heavy tails. *Risks*, 5(1):10.

Coles, S. G. and Powell, E. A. (1996). Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review/Revue Internationale de Statistique*, pages 119–136.

Cooke, R. M., Nieboer, D., and Misiewicz, J. (2014). *Fat-tailed distributions: Data, diagnostics and dependence*, volume 1. John Wiley & Sons.

Drees, H. et al. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli*, 9(4):617–657.

Embrechts, P. and Schmidli, H. (1994). Modelling of extremal events in insurance and finance. *Zeitschrift für Operations Research*, 39(1):1–34.

Faddy, M., Graves, N., and Pettitt, A. (2009). Modeling length of stay in hospital and other right skewed data: comparison of phase-type, gamma and log-normal distributions. *Value in Health*, 12(2):309–314.

Faddy, M. and McClean, S. (1999). Analysing data on lengths of stay of hospital patients using phase-type distributions. *Applied Stochastic Models in Business and Industry*, 15(4):311–317.

Gardiner, J. C., Luo, Z., Tang, X., and Ramamoorthi, R. (2014). Fitting heavy-tailed distributions to health care data by parametric and bayesian methods. *Journal of Statistical Theory and Practice*, 8(4):619–652.

Gomes, M. I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: a review. *International Statistical Review*, 83(2):263–292.

Gül, M. and Güneri, A. F. (2015). Forecasting patient length of stay in an emergency department by artificial neural networks. *Journal of Aeronautics and Space Technologies*, 8(2):43–48.

Harini, S., Subbiah, M., and Srinivasan, M. (2018). Fitting length of stay in hospitals using transformed distributions. *Comm Stat Case Stud Data Anal Appl*, 4(1):1–8.

Houchens, R. L. and Schoeps, N. (1998). Comparison of hospital length of stay between two insurers for patients with pediatric asthma. *Statistical Case Studies: A Collaboration Between Academe and Industry*, pages 45–64.

Ickowicz, A. and Sparks, R. (2017). Modelling hospital length of stay using convolutive mixtures distributions. *Statistics in medicine*, 36(1):122–135.

Lin, M., Lucas Jr, H. C., and Shmueli, G. (2013). Research commentary—too big to fail: large samples and the p-value problem. *Information Systems Research*, 24(4):906–917.

Luceño, A. (2006). Fitting the generalized pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics & Data Analysis*, 51(2):904–917.

Marazzi, A., Paccaud, F., Ruffieux, C., and Beguin, C. (1998). Fitting the distributions of length of stay by parametric models. *Medical care*, 36(6):915–927.

McClean, S. and Millard, P. (1993). Patterns of length of stay after admission in geriatric medicine: an event history approach. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 42(3):263–274.

Moré, J. J. and Wild, S. M. (2009). Benchmarking derivative-free optimization algorithms. *SIAM Journal on Optimization*, 20(1):172–191.

Murdoch, D. J., Tsai, Y.-L., and Adcock, J. (2008). P-values are random variables. *The American Statistician*, 62(3):242–245.

Nash, J. C. (2014). *Nonlinear parameter optimization using R tools*. John Wiley & Sons.

Nocedal, J. (2004). Wright. sj, 1999. numerical optimization.

Papi, M., Pontecorvi, L., and Setola, R. (2016). A new model for the length of stay of hospital patients. *Health care management science*, 19(1):58–65.

Qualls, M., Pallin, D. J., and Schuur, J. D. (2010). Parametric versus nonparametric statistical tests: the length of stay example. *Academic Emergency Medicine*, 17(10):1113–1121.

Sawilowsky, S. (2016). Rao-lovric and the triwizard point null hypothesis tournament. *Journal of Modern Applied Statistical Methods*, 15(2):4.

Singh, C. H. and Ladusingh, L. (2010). Inpatient length of stay: a finite mixture modeling analysis. *The European Journal of Health Economics*, 11(2):119–126.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.

Thomas, M., Lemaitre, M., Wilson, M. L., Viboud, C., Yordanov, Y., Wackernagel, H., and Carrat, F. (2016). Applications of extreme value theory in public health. *PloS one*, 11(7):e0159312.

Verburg, I. W., de Keizer, N. F., de Jonge, E., and Peek, N. (2014). Comparison of regression methods for modeling intensive care length of stay. *PloS one*, 9(10):e109684.

6 Appendix A

6.1 Beta-Cauchy Distribution

Here, $X \sim Cauchy(\theta, \lambda)$ and $T \sim Beta(\alpha, \beta)$ with PDF $f_X(x)$ and $f_T(t)$ respectively. By assuming $W = F_X(x)$, CDF of X, we get CDF of H using the transformation $F_T \circ W$. That is

$$\frac{1}{B(\alpha,\beta)} \int_0^{F_X(x)} t^{\alpha-1}(1-t)^{\beta-1}dt; \quad 0 < \alpha,\ \beta < \infty$$

where $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ and hence PDF of transformed random variable H is

$$f_H(x) = F'_H(x) = \frac{1}{B(\alpha,\beta)} F_X(x)^{\alpha-1}(1-F_X(x))^{\beta-1} f_X(x)$$

$$F_H(x) = \frac{\lambda}{\pi B(\alpha,\beta)} \left(\frac{1}{2} + \frac{1}{\pi} arctan\left(\frac{x-\theta}{\lambda}\right)\right)^{\alpha-1} \left(\frac{1}{2} - \frac{1}{\pi} arctan\left(\frac{x-\theta}{\lambda}\right)\right)^{\beta-1} \frac{1}{\lambda^2 + (x-\theta)^2}$$

$$-\infty < x < \infty;\ 0 < \alpha,\ \beta,\ \lambda < \infty;\ \lambda < \infty;\ -\infty < \theta < \infty.$$

6.2 Gamma-Pareto Distribution

Here, $X \sim Pareto$ Type 1 (θ, k) and $T \sim Gamma(\alpha, \beta)$ with PDF $f_X(x)$ and $f_T(t)$ respectively. By assuming $W = -log[1 - F_X(x)]$, CDF of X, we get CDF of H using the transformation $F_T \circ W$. That is

$$F_H(x) = \frac{1}{\beta^\alpha \Gamma\alpha} \int_0^{-log[1-F_X(x))]} t^{\alpha-1} e^{-\frac{t}{\beta}} dt; \quad -\infty < t < \infty; \quad 0 < \alpha,\ \beta,\ \lambda < \infty$$

Where where $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ and hence PDF of transformed random variable H is

$$f_H(x) = F'_H(x) = \frac{1}{\Gamma\alpha, \beta^\alpha} \left(-log(1 - F_X(x))\right)^{\alpha-1} \left(1 - F_X(x)\right)^{\frac{1}{\beta}-1} f_X(x)$$

$$f_H(x) = \frac{1}{x\Gamma(\alpha)c^\alpha} \left(\frac{\theta}{x}\right)^{\frac{1}{c}} \left[log(\frac{x}{\theta})\right]^{\alpha-1};$$

where $c = \frac{\beta}{k}$; $\alpha,\ c,\ \theta > 0$; $0 < x < \infty$.

6.3 Weibull-Pareto

Here, $X \sim Pareto$ Type 1 (θ, k) and $T \sim Weibull(\alpha, \beta)$ with PDF $f_X(x)$ and $f_T(t)$ respectively. By assuming $W = -log[1 - F_X(x)]$, CDF of X, we get CDF of H using the transformation $F_T \circ W$. That is

$$F_H(x) = \int_0^{-log[1-F_X(x)]} \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^\alpha} dt; \quad -\infty < t < \infty; \quad 0 < \alpha,\ \beta,\ \lambda < \infty$$

$$f_H(x) = F'_H(x) = \frac{\alpha}{\beta}\frac{f_X(x)}{1-F_X(x)}\left[\frac{-log(1-F_X(x))}{\beta}\right]^{\alpha-1} exp\left[-(\frac{-log(1-F_X(x))}{\beta})^{\alpha}\right]$$
$$f_H(x) = \frac{\tau\alpha}{x}\left[\tau log(\frac{x}{k})\right]^{\alpha-1} exp\left[-\left(\tau log\left(\frac{x}{k}\right)\right)^{\alpha}\right]; \quad x > k \text{ and } \alpha,\ \tau,\ k > 0.$$

6.4 Gamma -Exponential - Cauchy Distribution

Let $X \sim Cauchy(\theta, \lambda)$, $T \sim \Gamma(\alpha, \beta)$ a random variable Y is defined using a quantile function Q_Y which follows exponential distribution. By assuming $Q_Y = -log[1 - F_X(x)]$,we get CDF of H using the transformation $F_T \circ Q_Y \circ F_X$. That is

$$F_H(x) = \int_0^{Q_Y(y)} f(t)dt = \int_0^{-log[1-F_X(x)]} f(t)dt$$
$$f_H(x) = \frac{\gamma\left[\alpha - \beta^{-1} log\left(0.5 - \pi^{-1} arctan(x/\theta)\right)\right]}{\Gamma\alpha}; \quad -\infty < x < \infty;\ 0 < \theta,\ \alpha,\ \beta < \infty$$

and hence PDF of transformed random variable H is

$$f_H(x) = \frac{f_X(x)}{1-F_X(x)} f_T(-log(1 - F_x(x))); \quad -\infty < x < \infty;\ 0 < \theta,\ \alpha,\ \beta < \infty.$$