International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:14, No:6, 2020

A Methodology for Characterising the Tail Behaviour of a Distribution

Serge Provost, Yishan Zang

***Abstract*—**Following a review of various approaches that are utilized for classifying the tail behavior of a distribution, an easily implementable methodology that relies on an arctangent transformation is presented. The classification criterion is actually based on the difference between two specific quantiles of the transformed distribution. The resulting categories enable one to classify distributional tails as distinctly short, short, nearly medium, medium, extended medium and somewhat long, providing that at least two moments exist. Distributions possessing a single moment are said to be long tailed while those failing to have any finite moments are classified as having an extremely long tail. Several illustrative examples will be presented.

***Keywords*—**Arctangent transformation, change of variables, heavy-tailed distributions, tail classification.

I. Introduction

THIS section features an overview of certain criteria that have been previously introduced in the statistical literature for the purpose of labeling distributional tail behavior. Fairly recently, [1] provided several classification categories for identifying light to heavy-tailed distributions, these being based on moments, hazard rate functions and mean excess loss functions.

Previously, [2] examined the limiting behavior of density quantile functions in terms of a parameter α wherefrom one could identify three types of tail behavior: short, medium and long tails, corresponding to α < 1, α = 1 and α > 1, respectively. In order to refine the tail classification advocated by [2], [3] also relied on a quantity c that depends on the hazard function associated with a given distribution, this enhancement generating five categories of tail behaviour: short $[0 < \alpha < 1]$; medium-short $[\alpha = 1, c = 0]$; medium-medium $[\alpha = 1,\ 0 < c < \infty]$; medium-long $[\alpha = 1,\ c = \infty]$; and long $[\alpha > 1]$. Actually, this criterion has a theoretical connection with the limiting size of extreme spacings. The reader may also refer to [4] whose classification is based on the residual lifetime distributions. The aforementioned classification techniques are reviewed in [5].

As explained in [6], there exists a variety of methodologies for determining whether a distribution has an exponential or a power tail, including QQ-plots, likelihood methods, and plots of the mean residual life functions.

As well, [7] and [8], among others, proposed criteria that take into account the behaviour of a sizeable portion of the distribution at hand. An efficiency-based tail ordering technique was introduced by [9] in the context of location experiments. Another approach consists in relating the limiting distribution of the standardized maximum of a given distribution (when it exists) to that of the Weibull, Gumbel or Fréchet distributions (the three extremal distributions), which leads to categorizing the distribution as having a short, medium or long tail, respectively. It is manifestly desirable to develop procedures that will yield additional categories with a view to identifying more precisely the tail behaviour of a wide array of distributions.

The conceptually simple technique being advocated in this paper results in eight categories. It is applied to numerous theoretical distributions, the resulting tail behaviours being generally found to be consistent with those determined by making use of other criteria. In the case of a sample of observations, one must initially obtain a density estimate to which the proposed approach can then be readily applied. Of course, the larger the sample, the more reliable the results, which is corroborated by a small-scale simulation study presented in Section III. The proposed methodology, which is described in Section II, is applied to an array of widely used distributions in Section IV.

II. A Methodology Based on the Arctangent Transformation

We are proposing to make use of the percentiles of a transformed distribution to define a criterion for characterising the tail behaviour of a given distribution. More specifically, on letting X represent a distribution having finite mean μ, finite variance σ^2 and probability density function (PDF) $f(x)$, the standardized random variable $Y = (X - \mu)/\sigma$ is mapped onto $(-1,\ 1)$ or a subset thereof via the transformation $Z = (2/\pi)$ arctan(Y). The density functions of certain distributions that have been so transformed are plotted in Figs. 1-15. These distributions include the normal, Weibull with parameters 0.5 and 2, extreme value, logistic, exponential, Student t on 3, 5 and 20 degrees of freedom (df), lognormal, Uniform(0,1), Beta(5,2), type-II Beta(5,3) and (50,30), and Gamma(50,1). The dots indicate the 90^{th} and 99.9999^{th} percentiles of Z.

Let $q(\alpha)$ represent the $(100 \times \alpha)^{\text{th}}$ quantile of the distribution of Z. We propose to employ the difference between the 90^{th} and 99.9999^{th} percentiles of Z as a criterion for classifying the right-tail behavior of X. We denote this *tail index* by

$$p = q(0.999999) - q(0.90)\,. \quad (1)$$

Serge Provost is Professor of Statistics at the University of Western Ontario, Canada (e-mail: provost@stats.uwo.ca).
Yishan Zang is a Ph.D. student in the Department of Statistical and Actuarial Sciences at the University of Western Ontario, London, Ontario, Canada, N6A5B7 (e-mail: yzang8@uwo.ca).

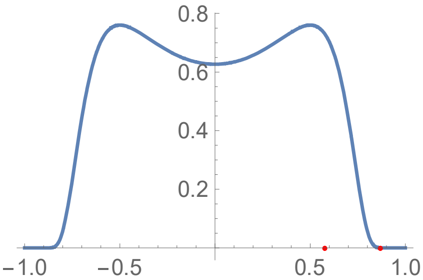

Fig. 1 PDF of *Z* for the normal distribution

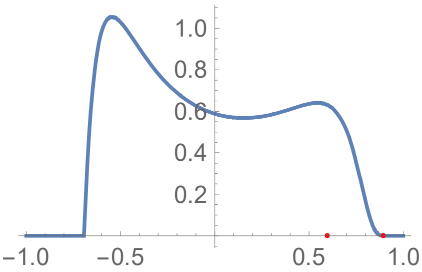

Fig. 2 PDF of *Z* for the Weibull distribution (k = 2)

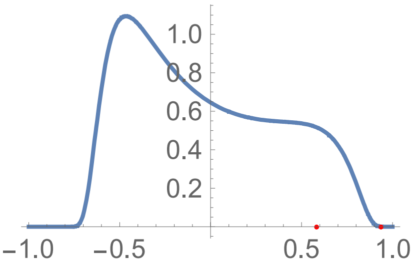

Fig. 3 PDF of *Z* for the extreme value distribution

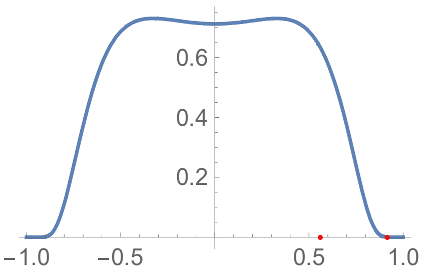

Fig. 4 PDF of *Z* for the logistic distribution

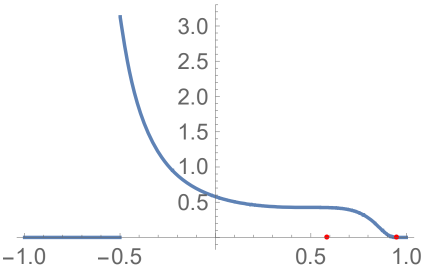

Fig. 5 PDF of *Z* for the exponential distribution

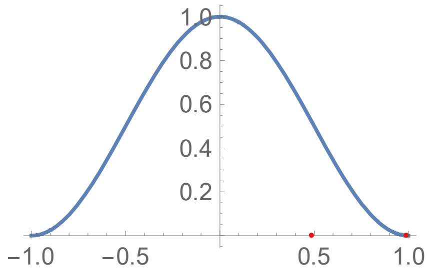

Fig. 6 PDF of *Z* for the *t* distribution on 3 df

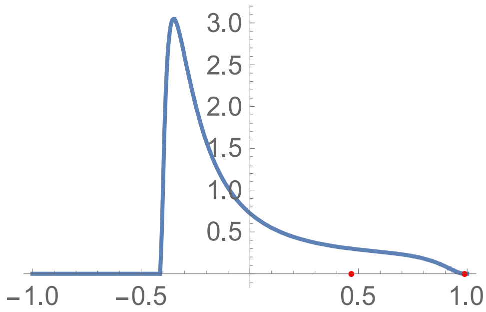

Fig. 7 PDF of *Z* for the lognormal distribution

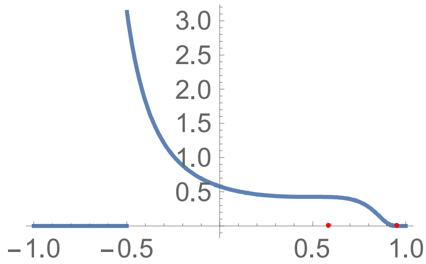

Fig. 8 PDF of *Z* for the Weibull distribution (k = 0.5)

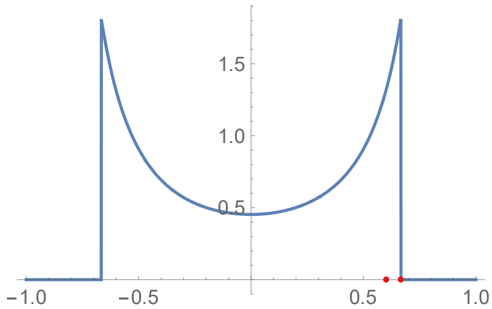

Fig. 9 PDF of *Z* for the Uniform(0,1) distribution

We define distributions whose mean is finite but whose variance is infinite as having a long tail and distributions whose mean is undefined as having an extremely long tail.

Generally, the fewer the number of finite moments a distribution possesses, the heavier its tail is. The proposed tail behavior categories and their associated tail index ranges are:

- Distinctly Short: $p < 0.1$
- Short: $0.1 \leq p < 0.2$
- Nearly Medium: $0.2 \leq p < 0.3$

- Medium: $0.3 \leq p < 0.4$
- Extended Medium: $0.4 \leq p < 0.5$
- Relatively Long: $p \geq 0.5$
- Long: Indefinite second moment
- Extremely Long: Indefinite first moment.

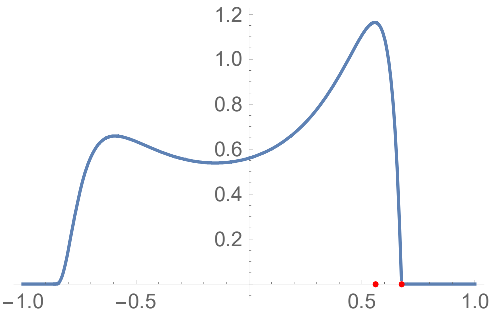

Fig. 10 PDF of Z for the Beta(5,2) distribution

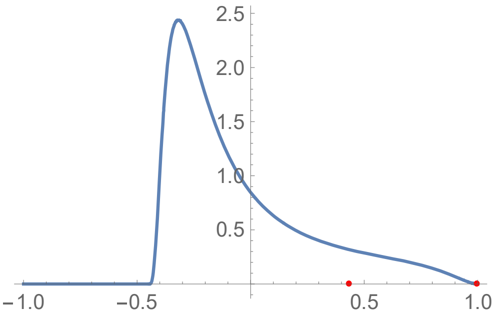

Fig. 11 PDF of Z for the type-II Beta(5,3) distribution

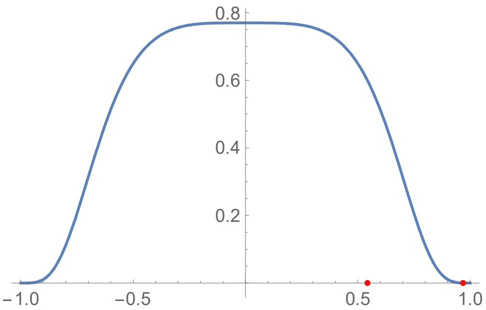

Fig. 12 PDF of Z for the t distribution on 3 df

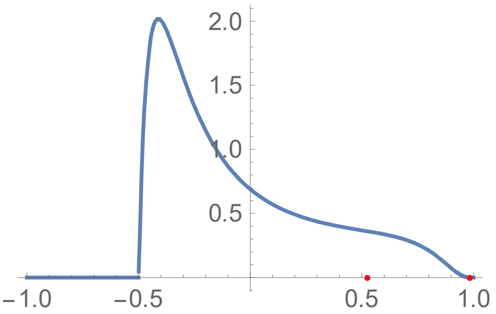

Fig. 13 PDF of Z for the Gamma(50,1) distribution

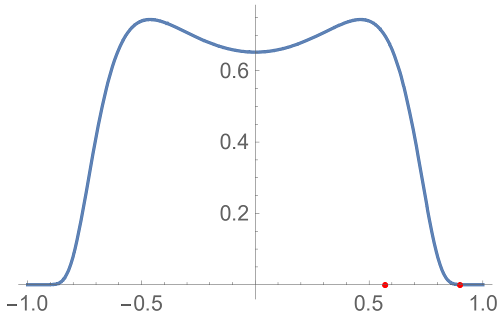

Fig. 14 PDF of Z for the t distribution on 20 df

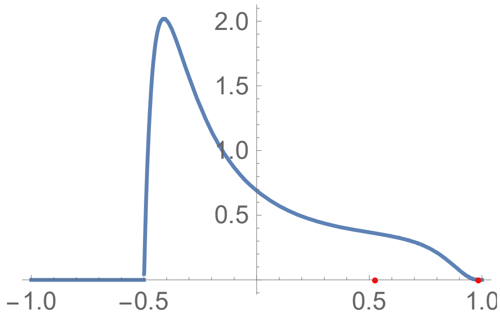

Fig. 15 PDF of Z for the type-II Beta(50,30) distribution

The left-tail behavior of a distribution is similarly characterized by defining the corresponding tail index as

$$p^* = q(0.10) - q(0.000001). \quad (2)$$

The specified ranges for p also apply to p^*.

When this methodology is implemented, a distribution is deemed to be heavy tailed if it belongs to one of the following categories: Extended Medium ($0.4 \leq p < 0.5$), Relatively Long ($p \geq 0.5$), Long or Extremely Long. For instance, the lognormal, Weibull(k) with $k < 1$, Pareto, Student t with fewer than 6 degrees of freedom and Cauchy are such distributions. Heavy-tailed distributions belong to the medium-long and long tail categories in Schuster's classification.

III. A Simulation Study

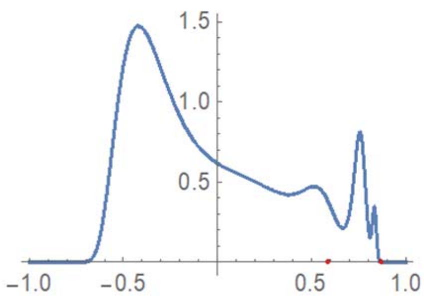

Fig. 16 PDF of Z, 100 Exp(1) points

We generated samples of sizes 100, 1000, 50000 and 1000000 from the standard exponential distribution and obtained the following tail indices from their kde's: 0.281119, 0.309942, 0.342787 and 0.361315. We observe that these values tend to the theoretical *p*-index for the standard exponential distribution, which is 0.3672. The corresponding density functions of *Z* are plotted in Figs. 16-19.

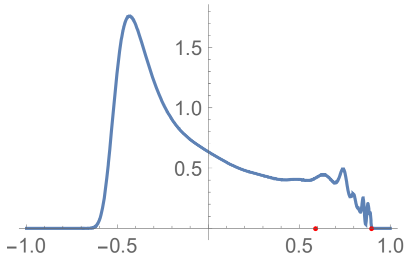

Fig. 17 PDF of *Z*, 1,000 Exp(1) points

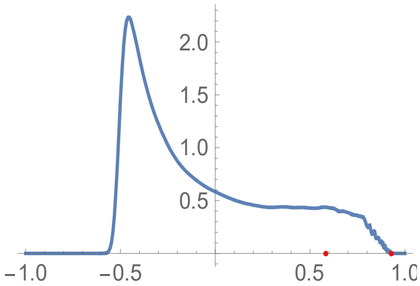

Fig. 18 PDF of *Z*, 50,000 Exp(1) points

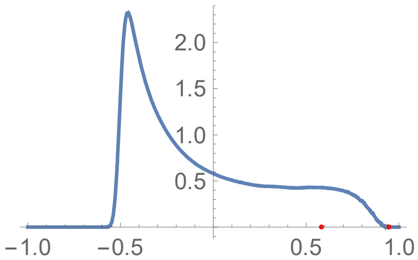

Fig. 19 PDF of *Z*, 1,000,000 Exp(1) points

IV. Application of the Tail Index Criterion

Some illustrative classification results resulting from the application of the proposed tail index criterion are presented in Table I for certain commonly utilized distributions. It should be noted that for the beta, Weibull and lognormal distributions, the shapes of the standardized densities and thus the tail behavior and the associated value of the tail index vary with the parameters.

V. Conclusion

While being readily implementable, the proposed approach to characterising distributional tail behaviour produces easily identifiable categories that, generally, prove consistent with those obtained by making use of other criteria.

TABLE I
The Tail Behavior of Certain Distributions

Distribution	Tail Behavior	Tail Index
Uniform	Distinctly Short	0.0646
Beta(5, 2)	Short	0.1152
Normal	Nearly Medium	0.2898
Rayleigh	Nearly Medium	0.2998
Type-II Beta(50,30)	Medium	0.3444
Extreme value	Medium	0.3549
Logistic	Medium	0.3562
Exponential	Medium	0.3672
Student *t* on 5 df	Extended Medium	0.4244
Type-II Beta(2,5)	Extended Medium	0.4591
Student *t* on 3 df	Relatively Long	0.5071
Lognormal	Relatively Long	0.5201
Type-II Beta(5,3)	Relatively Long	0.5609
Weibull	Relatively Long	0.5800
Student *t* on 2 df	Long	**
Cauchy	Extremely Long	**

** Moment-based.

References

[1] S. A. Klugman, H. H. Panjer, and G. E. Willmot, "Loss Models: from Data to Decisions, 4th Ed." John Wiley & Sons, New York, 2012.
[2] E. Parzen, "Nonparametric statistical data modeling," Journal of the American Statistical Association, vol. 74, pp. 105-121, 1979.
[3] E. F. Schuster, "Classification of probability laws by tail behavior," Journal of the American Statistical Association, vol. 79, pp. 936-939, 1984.
[4] J. Rojo, "On tail categorization of probability laws," Journal of the American Statistical Association, vol. 91, pp. 378-384, 1996.
[5] J. Rojo and R. C. Ott, "Testing for tail behavior using extreme spacings," arXiv:1011.6458, 2010.
[6] C. C. Heyde and S. G. Kou,"On the controversy over tailweight of distributions," Operations Research Letters, vol. 32, pp. 399–408, 2004.
[7] W.-Y. Loh. "Bounds on are's for restricted classes of distributions defined vai tail-orderings," Annals of Statistics, vol.12, pp. 685–701, 1984.
[8] K. Doksum,."Starshaped transformations and the power of rank tests," The Annals of Mathematical Statistics, vol.40, pp.1167–1176, 1969.
[9] E. Lehmann, "Comparing location experiments," The Annals of Statistics, vol. 16, pp.521–533, 1988.