Stable Distributions for Heavy-Tailed Data and Their Application in Asset Health Monitoring

Sundaram S, McDonald K
Optimized Systems and Solutions Limited
174F, Brook House, Milton Park, Oxfordshire OX14 4SE
+44 1235 433 574
srini.sundaram@o-sys.com

Clifton D A
Institute of Biomedical Engineering,
University of Oxford, Oxford, UK, OX3 7DQ

Abstract

Probabilistic approaches used in asset health monitoring applications attempt to capture the properties exhibited by the random variables describing the operational behaviour of the asset. In a standard parametric approach, the underlying data density is modelled using distributions. The model parameters are fitted to the observed data (in some approaches with an assumption that the data follows a Gaussian distribution) using empirical or maximum likelihood methods. The model is then used to make decisions on out-of-sample data for any potentially extreme (and “abnormal”) events. In practical applications, large datasets obtained from assets exhibit heavier tails than Gaussian distributions, and have non-zero skewness. The tail behaviour of such distributions offers valuable information for decision support and risk management tools used by fleet engineers. This paper introduces the application of stable distributions to a dataset of performance parameters acquired from an aerospace gas-turbine engine. Stable distributions are a class of probability distributions that generalise Gaussian distributions, and which can accommodate heavy tails and skewness. The results show that stable distributions accurately describe the tail events of real-life data in an accurate manner and can be used as an alerting tool in fleet management.

1. Introduction

In high integrity systems, decision support tools are widely used to monitor the operational characteristics and state of the health of the asset. These tools can take the form of data-driven models that estimate the structure of asset data in a probabilistic manner. Such approaches aim to identify changes in the structure of the data due to extreme events and raise alerts if the out-of-sample data (or “test” data) deviate from the learnt structure of the training data. These models allow the detection of precursors of asset failure such that unplanned shutdown can be prevented. In this context, Gaussian-based models are widely used in many applications such as engineering[1], finance, and medicine. However, the assumption of Gaussianity is often inappropriate for modelling data distributed with skewness and heavy tails. In such cases, the Gaussian distribution decays too rapidly (according to the negative exponential form of the Gaussian), which is inadequate when the data exhibit slow-decaying heavy tails[2]. Many risk management practices that make the assumption of Gaussianity have failed to handle extreme events
during the 2007-2008 financial crisis, for example. It is not hard to imagine that the 
dynamics behind human decision-making reveals that such processes exhibit similar 
heavy-tailed behaviour. To this end, this paper applies a family of leptokurtic 
distributions, named stable distributions, to asset health monitoring. The approach is 
illustrated using an application to aerospace gas-turbine engine health monitoring.

2. Univariate Stable Distributions and Heavy Tails

2.1 Definition and Parameterisation

If \( X, X_1, X_2, X_3, \ldots, X_n \) are random variables that are independent and identically 
distributed, they are termed stable if the shape of their distribution is retained after 
summation; i.e., for every \( n \),

\[
X_1 + X_2 + X_3 + \ldots + X_n \overset{d}{=} c_n X + d_n
\]

where \( c_n > 0 \) and \( d_n \) are constants and \( \overset{d}{=} \) stands for equality in distribution. The class of 
distributions with the above property can be termed stable distributions\(^{[2]} \). The above 
form as in equation (1) is termed sum stable, because the stability is defined in a 
summation sense. A well known example is the Gaussian family, whereby a mixture of 
Gaussians is itself Gaussian. This can also be extended into multiplication stable, 
min-stable, and max-stable. The latter two leads to extreme value distributions\(^{[10]} \). The stable 
framework can also be extended into geometric equivalents as geometric sum stable, 
multiplication stable, min stable, and max stable\(^{[3]} \). Let \( X_i \) be a random variable at 
period \( t = t_0 + i \) with a distribution function \( F \). There exists a small probability in any 
period that an event can alter the probabilistic structure of the underlying process. Let us 
denote the time at which such event occurs be \( T(p) \). This time \( T(p) \) is assumed to a 
random variable following a geometric distribution \( P\{T(p) = k\} = (1 - p)^{k-1} p \). The 
geometric sum can be defined as the accumulation of all \( X_i \)’s up to the event \( t_0 + T(p) \). 
i.e. \( G(p) = \sum_{i=1}^{T(p)} X_i \). The distribution function \( F \) of the random variable is geometric 
stable if there exists constants \( a = a(p) > 0 \) such that \( aG(p)^d = X_i \). More detailed 
definitions can be found in [3].

The stable distribution is characterised by four parameters:

- **Index of stability** or **characteristic exponent** \( \alpha \not\in (0, 2] \),
- **skewness** parameter, \( \beta \not\in [-1, 1] \),
- **gamma or scale** parameter, \( \gamma > 0 \), and
- **location** parameter, \( \delta \in \mathbb{R} \).
The characteristic exponent \( \alpha \) determines the rate at which the tails decay. When \( \alpha = 2 \), the stable distribution becomes a Gaussian distribution. For \( \alpha < 2 \), the decay characteristics follows a power-law. The \( \delta \) parameter shifts the distribution to the left or right on the \( x \)-axis, while the \( \gamma \) parameter compresses or expands the distribution about \( \delta \) in proportion to \( \gamma \). The skewness parameter \( \beta \) along with the index of stability (\( \alpha \)) determines the shape of the distribution. Often for analysis, the stable random variable \( X \) is used as a transformed variable according to \( (X - \delta)/\gamma \), because the transformation results in a stable distribution due to the property shown in (1).

### 2.3 Parameter Estimation from Data

The methods commonly employed to estimate the values of the four parameters using a training dataset are (i) the method of moments or (ii) maximum likelihood estimation. This paper uses a two-stage method for estimating the four parameters: firstly, the initial values of the parameters are estimated using McCulloch’s quantile-based method[6]. These initial estimates are then refined using the regression-based method described by Koutrovelis[7].

#### 2.3.1 Initial estimate using McCulloch’s method

McCulloch showed that the four parameters can be estimated consistently from the predetermined sample quantiles for \( \alpha \) \([0.6, 2]\) and \( \beta \) \([-1, 1]\). Using the five population quantiles \( x_p \) where \( p = 0.05, 0.25, 0.5, 0.75 \) and 0.95, McCulloch showed that reliable estimates of the values of the four parameters can be found[6]. First, the functions \( v_\alpha \) and \( v_\beta \) were found using the population quantiles as follows:

\[
v_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}} \quad \text{and} \quad v_\beta = \frac{x_{0.95} - x_{0.05} - 2x_{0.5}}{x_{0.95} - x_{0.05}}.
\]

A set of tables containing the values of \( v_\alpha \) as a function of \( \phi_1(\alpha, \beta) \) and \( v_\beta \) as functions of \( \phi_2(\alpha, \beta) \) The functions values \( \phi_1 \) and \( \phi_2 \) were stored as look up tables by MuCulloch[6]. For an estimated \( v_\alpha \), the equivalent \( \alpha \) and \( \beta \) can be obtained from the tables (reversing the relationship). Similarly \( v_\gamma = \frac{x_{0.75} - x_{0.25}}{\gamma} \) was used and its relationship with a function \( \phi_3(\alpha, \beta) \) was used to calculate \( \gamma \). Using \( \gamma \) and functions \( \phi_3(\alpha, \beta) \) and \( v_\delta = \frac{\delta - x_{0.5}}{\gamma} \), \( \delta \) was estimated.
2.3.2. Characteristic functions and distribution functions

In continuous the sense, the distribution function

\[ F = P[X \leq x] = \int_{-\infty}^{x} f(y)dy \]  

(2)
defines the probability that a random variable (r.v.) \( y \) realises a value in the interval \( (-\infty, x] \), where \( f(.) \) denotes the density function of the r.v. The distributions can then be used to make decisions about the asset variable exceeding a certain value with a degree of confidence using probability measures. This means that it is preferable to estimate the distributions in closed form, and which is one of the reasons that makes the conventional use of the Gaussian distribution a popular choice, along with its analytic ease of treatment. Among the stable family of distributions, only three types of distributions have closed form expression densities: the Gaussian (\( \alpha = 2 \)), the Levy (\( \alpha = 0.5 \)), and the Cauchy (\( \alpha = 1 \)). For other values of \( \alpha \), the densities and distribution functions can be estimated using the characteristic function approach. A characteristic function for a random variable \( X = \{x_1, x_2, \ldots, x_n\} \) can be defined as

\[ \phi_n(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(itx_j) \]  

(3)

It is possible to obtain distribution functions from the characteristic function either using the inversion theorem[5] or integral transforms. Stable laws can be defined by their characterisation equation

\[ \phi(t) = \exp \left( i\delta t - \gamma |t|^\alpha [1 + i\beta \text{sgn}(t)\omega(t, \alpha)] \right) \]  

(4)

where \( \omega(t, \alpha) = \tan(\pi \alpha / 2) \) for \( \alpha \neq 1 \) and \( \omega(t, \alpha) = (2/\pi) \log |t| \) for \( \alpha = 1 \).

Koutrovolis showed that if we take \( z = \log(-\log|\phi(t)|^\gamma) = \log(2\gamma^\alpha) + \alpha \log(t) \), \( z \) depends only on \( \alpha \) and \( \gamma \). From the above expression, he showed that it is possible to estimate the parameters using a two step procedure.

- \( \alpha \) and \( \gamma \) are estimated by regressing \( \log(-\log|\phi(t)|^\gamma) \) onto \( w = \log |t| \) in the model \( \log(2\gamma^\alpha) + \alpha w_k + \epsilon_k \). Here \( k = 1, 2, 3, \ldots, K \) denotes appropriate set of points chosen from a lookup table for various sample sizes \( N \) and \( \alpha \). \( \epsilon_k \) denotes the regression error term. The characteristic function \( \phi(t) \) and \( w \) are evaluated at points \( t_k = \pi k / 25 \).

- Similarly \( \beta \) and \( \delta \) can be estimated by regressing \( \arctan(\text{img}(\phi(u))/\text{real}(\phi(u))) \) onto \( u \) and \( \text{sign}(u) |u|^\alpha \) in the model \( \delta u_l - \beta |u|^\alpha \tan(\pi \alpha / 2) \text{sgn}(u_l) |u_l|^\alpha + \epsilon_l \). Here \( l = 1, 2, 3, \ldots, L \) denotes appropriate set of points chosen using a lookup table for various sample sizes \( N \) and \( \alpha \). \( \epsilon_l \) denotes the regression error term. The \( \arctan(.) \) term and regression model are evaluated at points \( u_l = \pi l / 50 \).
2.2 Stable Density and Distributions

Many parameterisations were proposed in the literature to compute stable densities using characteristic functions similar to that shown in equation (4). Once such parameterisation, called Zolotarev’s (M) parameterisation\[^4\], is used for the study described in this paper. According to this method, the characteristic function takes the form

\[
E(e^{itX}) = \begin{cases} 
\exp \left\{ -\left| t \right|^\alpha \left[ 1 + \beta(t)(\tan \frac{\pi \alpha}{2})\left(\left| t \right|^{\alpha - 1} - 1 \right) \right] \right\} & \alpha \neq 1 \\
\exp \left\{ -\left| t \right|^\alpha \left[ 1 + \beta(t)(\frac{2}{\pi})\left(\ln \left| t \right| \right) \right] \right\} & \alpha = 1 
\end{cases} \tag{5}
\]

The above parameterisation is widely used as it ensures the characteristic function is jointly continuous in all four parameters, and that the densities and the distributions derived from it remain continuous. Zolotarev’s integral formulae\[^9\] were used to compute the density \( f(x; \theta) \) and distribution function \( F(x; \theta) \) of a random variable with a characteristic function of the form as given in equation (5), where \( \theta = \{\alpha, \beta, \gamma, \delta\} \). Using the following definitions below,

\[
\zeta = \begin{cases} 
-\beta \tan \frac{\pi \alpha}{2} & \alpha \neq 1 \\
0 & \alpha = 1 
\end{cases}, \quad \theta_0 = \begin{cases} 
\frac{1}{\alpha} \arctan \left( \beta \tan \frac{\pi \alpha}{2} \right) & \alpha \neq 1 \\
\frac{\pi}{2} & \alpha = 1 
\end{cases}, \quad c = \begin{cases} 
\frac{1}{\pi} \left( \frac{\pi}{2} - \theta_0 \right) & \alpha < 1 \\
0 & \alpha = 1 \\
1 & \alpha > 1 
\end{cases}
\]

\[
\phi(\theta) = \begin{cases} 
\cos \left( \alpha \theta_0 \right)^{\frac{1}{\alpha - 1}} \left( \frac{\cos \theta}{\sin \alpha \left( \theta_0 + \theta \right)} \right)^{\frac{\alpha}{\alpha - 1}} \cos \left( \alpha \theta_0 + (\alpha - 1) \theta \right) & \alpha \neq 1 \\
\frac{2}{\pi} \left( \frac{0.5 \pi + \beta \theta}{\cos \theta} \right) \exp \left( \frac{1}{\beta} \left( \frac{\pi}{2} + \beta \theta \right) \tan \theta \right) & \alpha = 1, \beta \neq 0 
\end{cases}
\]

Compared to equation (5), expressing the characteristic function \( \phi(\theta) \) as a function of alternative skewness \( \theta_0 \), makes the calculation direct.

The densities and distributions are given as follows.

For \( \alpha \neq 1 \) and \( x > \zeta \)

\[
f(x; \theta) = \frac{\alpha(x - \zeta)^{\alpha - 1}}{\pi[\alpha - 1]} \int_{-\theta_0}^{\frac{\pi}{2}} \phi(\theta) \exp(-x(x - \zeta)^{\alpha - 1} \phi(\theta)) d\theta \tag{6}
\]

\[
F(x; \theta) = c_1(\alpha, \beta) + \frac{\text{sign}(1 - \alpha)}{\pi} \int_{-\theta_0}^{\frac{\pi}{2}} \exp(-x(x - \zeta)^{\alpha - 1} \phi(\theta)) d\theta \tag{7}
\]
For $\alpha \neq 1$ and $x = \gamma$

$$f(x; \theta) = \frac{\Gamma(1 + \frac{1}{\alpha})(\cos \theta_0)}{\pi(1 + \gamma^2)^{(1/2\alpha)}}$$  \hspace{1cm} (8)

$$F(x; \theta) = \frac{1}{\pi} \left( \frac{\pi}{2} - \theta_0 \right)$$  \hspace{1cm} (9)

Detailed proofs can be found in [8]. The effect of the four parameters of stable densities is shown in figure 1.

![Graphs showing the effect of parameters on stable densities](image)

(a) and (b) show the effect of $\alpha$ and $\gamma$ on the distribution, respectively. (c) and (d) illustrate the impact of $\beta$ and $\delta$.

Figure 1. Effect of parameters on stable distributions

3. Application to Asset Health Monitoring

Industrial assets are often equipped with a number of sensors measuring variables describing the asset behaviour. Traditional conditional monitoring approaches typically use an ad-hoc alerting threshold (derived using domain-based expert knowledge) to monitor the variables for abnormal events. The data-driven approach is a useful
alternative to this, in which we construct a probabilistic model to capture the structure of the data and aim to identify any precursors of failures in the out-of-sample data as being “improbable” events. In the latter, inappropriate assumptions about the distribution of the random variables may result in underestimation of the tail mass, as discussed earlier. This will lead to inaccurate classification of out-of-sample data, and so careful consideration should be given to such assumptions during model development. In this section, stable distributions were applied to data acquired from an aerospace gas-turbine engine, which were observed to be distributed with heavy tails. These parameters are descriptive parameters of an aerospace gas-turbine engine performance. They were also applied as a tool to identify the variables that exhibit tail events (i.e., “abnormalities” with respect to the distribution of the “normal” training data). This will enable the automated testing of large volumes of data to identify heavy tailed data for further analysis. The following section presents some of the results of this application.

4. Results

First, the quantile-quantile plots of a number of engine performance parameters (acquired during a period of approximately one year) against the best-fit Gaussian distribution is shown. If the distributions of the data are actually Gaussian, the plotted data will follow the red-dashed line closely. Figure 2 shows that data exhibit non-Gaussian behaviour in most of the tails of the distributions of each variable. These data were then used to fit stable distributions, and the results are shown in Figure 3 and Table 1.

![Figure 2- Q-Q Plot of the variables exhibiting non Gaussian behaviour](image-url)
Figure 3 – Stable density and distribution estimates for a variable exhibiting normal behaviour (a,c) and an event (b,d)

Table 1. Estimated Stable parameters

<table>
<thead>
<tr>
<th>Variable</th>
<th>Alpha ($\alpha$)</th>
<th>Beta ($\beta$)</th>
<th>Gamma ($\gamma$)</th>
<th>Delta ($\delta$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-1</td>
<td>0.3039</td>
<td>-2.8791</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>0.6630</td>
<td>-4.329</td>
</tr>
<tr>
<td>3</td>
<td>1.7333</td>
<td>-0.6461</td>
<td>0.3235</td>
<td>0.2464</td>
</tr>
<tr>
<td>4</td>
<td>1.8536</td>
<td>-1</td>
<td>0.1466</td>
<td>-1.8655</td>
</tr>
<tr>
<td>5</td>
<td>1.8274</td>
<td>-0.92422</td>
<td>0.2558</td>
<td>-5.0187</td>
</tr>
<tr>
<td>6</td>
<td>1.9618</td>
<td>-1</td>
<td>0.1695</td>
<td>-0.8034</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>-1</td>
<td>1.5353</td>
<td>70.3441</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>-1</td>
<td>4.0646</td>
<td>119.5177</td>
</tr>
</tbody>
</table>

Figure 3 shows that the stable family of distributions is able to fit the heavy tails in the data more closely than the Gaussian distribution. The departure from Gaussianity is also shown in Table 1, where it may be seen that $\alpha < 2$ (recalling that $\alpha = 2$ for Gaussianity). In a second test (the results of which are shown in figure 4), the time-variation of $\alpha$ is shown during periods of “normal” operation and during a period in
which asset failure occurs. The event is clearly identified by rapid changes in the value of $\alpha$ during the period of abnormal asset condition which may be seen between $t=690$ to $t=705$ in figure 4.

![Figure 4 Index of Stability for an Engine with normal behaviour and failure event](image)

5. Conclusions

The application of stable distributions to model the heavy tailed data is presented, in which we illustrated the processes of parameter estimation from training data. The stable framework applied to the dataset from an aerospace gas-turbine engine showed that it was a more appropriate distribution for use in modelling the tails of real-life data than the conventional choice of the Gaussian distribution. Noting that one aim of health monitoring is to detect “abnormal” events, which typically lie in the tails of distributions (by definition, being improbable with respect to the distribution of “normal” data), a close fit to the tails of the distribution is of critical importance, hence our use of the stable distribution.

We note that this method has considered unimodal distributions, and has limited its scope for this introductory paper to illustrating the deviation of one of the four stable parameters. In practice, we would consider the deviation of all four parameters, in the unimodal case, and could consider mixtures of stable distributions such that multimodal data could be modelled (analous to the application of Gaussian mixture models).

Acknowledgement

The authors would like to acknowledge Rolls-Royce Plc and Optimized Systems and Solutions for the support in this study. David Clifton was supported by the Centre of Excellence in Personalised Healthcare funded by the Wellcome Trust and EPSRC under grant number WT 088877/Z/09/Z.
References


