Financial Interaction Analysis using Best-Fitted Probability Distribution

Vincent Ang[1]
Specialist Risk Unit, Bank Negara Malaysia (Central Bank of Malaysia)
Email: vincent@bnm.gov.my

Abstract
This paper explores the statistical method to best-fit probability distributions onto data of variables and subsequently performs analysis using the derived distributions. These distributions would represent the characteristics of the variable data. By employing Monte Carlo simulation on the derived distributions to generate values and imputing them into a model or formula that defines the interaction between the variables, we obtain the outcome of their interactions. An example of performing such an analysis onto the US treasury, and subsequently a hypothetical portfolio is included.

Key words: Probability distribution, goodness-of-fit, distribution fitting

1. Introduction

Probability distribution has been used in the financial industry as a way to model and quantify the outcomes of financial interactions – from the start of market activity such as option pricing, to the end of day calculation of a portfolio's Value-at-Risk (VaR). By assuming the distribution of a data, analysts can utilize the characteristics of the distribution to make predictions on outcomes. A commonly used probability distribution is the Normal distribution for its simplicity of having two easily identifiable parameters: mean and variance, and the widespread notion that most populations are distributed normally when sampled at large numbers. However, the Normal distribution assumes some criteria, such as symmetry and excess kurtosis of 0, which might not be true for all dataset. The past crisis has shown us that the Normal distribution is questionable when it comes to the performance of the financial industry. Occurrences of monumental losses causing bankruptcies and bail-outs of some of the largest financial institutions in the world has highlighted that better-fitting distributions ought to be employed in financial modelling to not underestimate any potential risk involved. It is worth highlighting that the aim of the paper is not to discourage the use of Normal distribution or to show the suitability of a distribution onto any data, but instead to encourage the practice to determine the best-fitting distribution and resist assuming data normality.

This paper explores methods to determine the best-fitting probability distribution for a given set of data, in order to use the derived distribution for further analysis. In doing so, the author hopes to also give a practical guide in using probability distribution in day-to-day analysis of financial data. Raw dataset are first matched to probability distributions of interests that reflect a good fit. Subsequently, the optimal parameters for the distribution in context are selected. Goodness of fit tests would then be conducted to determine the

[1] The views expressed in this paper are those of the author and do not necessarily represent the views or policy stance of Bank Negara Malaysia. Author of this paper would like to thank Charles Sandanasamy, Naimi Shuib, Calvin Choo, Cheryl Tan and Sim Jun Yi for their inputs and considerations in the process of writing this paper.

distribution that best fit the data. The distribution can then be used to forecast outcomes of financial interactions based on its characteristics. This paper takes sole interest in continuous random variables (as opposed to discrete random variables), as they enable forecasts at any point within the range supported by the distribution.

The introduction of the *Fundamental review of trading book* by Bank of International Settlement has also made it worthwhile to be proficient in determining the most appropriate loss distribution in embarking on the Expected Shortfall (ES) risk measurement method. As ES is suggested to better capture the tail risk, the shape of the distribution has be first be appropriately specified. Usual convention of taking the tail of a normal distribution[2] could be understating risk. ES would not provide much benefit if the users simply assume a normal distribution as the result would be a mere multiplication of the VaR of the same cut-off by a constant factor. However, after having fitted the distribution that would best match the performance of the underlying data, ES could have a more meaningful result.

2. Method and Materials

The best-fitted distribution for a dataset enables the user to describe the behaviour of the underlying financial data. In fitting the data with a distribution, usually more than one distribution would be of interest in the matching process. In order to identify the distributions that would be fitted onto the data, raw data is grouped into appropriate bins (frequency width) to obtain the density of the empirical data at different points. A histogram can be drawn to depict the shape of the empirical data's distribution for visual representation of the density. There are three general categories of probability distributions that can be employed to fit the distributions: Parametric, Non-parametric, and Semi-parametric. This paper would focus on the parametric and non-parametric distributions only. After determining the probability density function to be used to fit the data, goodness of fit (GoF) tests would be performed to quantitatively select the best fitting distribution.

Parametric Distributions

Parametric distributions (Pd) have pre-determined shapes and behaviours[3]. The user would have to identify Pd with similar shapes to that of the underlying dataset's density. This practice requires good judgement, understanding of the underlying data behaviour, and knowledge of the Pd's description. Non-trivial Pd would require that their parameters[4] be estimated to best fit the identified Pd to the underlying dataset. Methods to estimate parameters that have received widespread reception include Maximum likelihood estimation (MLE), and L-moment.

[2] This is a common practice in measuring Value-at-Risk, the measurement tool that ES is expected to replace.
[3] Behavior meaning the probability of the distribution at any particular point that is unique to itself, like heavy tails or skewness.
[4] Parameters here refer to the constants that would define the shape of the parametric distributions. E.g. mean and variance are two parameters of the normal distribution, while alpha and beta are the two parameters of the Beta distribution.

i) Maximum Likelihood Estimation

MLE is a process that attempts to compute the best parameters for any Pd that is hypothesized to fit the given finite number *n*, of empirical data (data points).The idea behind this is that for each different data point, there is a unique probability for that data point to exist/have occurred. Multiplying all *n* probability densities will give a joint probability, which is the likelihood probability, of all *n* data points to coexist/occur together (i.e. exactly the case when the analyst is given a set of empirical data). By selecting different parameters, the data points would result in different probabilities for each point. Hence, a different likelihood probability would be obtained for every different parameter chosen. MLE is then a process to determine the parameters that would result in the highest value of the likelihood probability i.e. highest likelihood for all the given data points to coexist. For instance, let α be a parameter of a Pd; then the likelihood probability is given by

$$\begin{aligned}(\text{Likelihood probability function} \mid \alpha) &= f(x_1|\alpha).f(x_2|\alpha)\dots f(x_n|\alpha) \\ &= \prod_{i=1}^{n} f(x_i|\alpha).\end{aligned}$$

such that $f(x)$ is the probability density function of the parametric distribution, and x_i (for all i from 1 to n) are the data points. The parameters that would maximize the likelihood probability would be the maximum-likelihood estimators (MLE). Since the logarithm[5] of the likelihood probability would be maximized when the likelihood probability is maximized, the derivation of MLE is usually simplified by differentiating the logarithm of the likelihood probability to solve for the MLE. By differentiating

$$\ln \prod_{i=1}^{n} f(x_i|\alpha) = \sum_{i=1}^{n} \ln f(x_i|\alpha)$$

and equating the result to 0, the user can solve for the α that maximizes the likelihood probability. With more parameters, the logarithmic equation would have to be differentiated with respect to all parameters separately, providing each parameter with its own MLE.

ii) L-moments

In order to describe L-moments, it is useful to first explain moments and its use in estimating statistical parameters. Moments are the expected value of the powers of the probability distribution of interest. They are defined as the integration of the probability density function, $f(x)$ with the random variable, x of a certain power:

$$\begin{aligned}\text{first moment} &= \int_{-\infty}^{\infty} x f(x) dx \\ \text{second moment} &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &\dots \\ \text{nth moment} &= \int_{-\infty}^{\infty} x^n f(x) dx\end{aligned}$$

For a Pd that has j number of parameters, the user would need to determine moments up to the power of j to have j equations so that each of the parameters can be solved linearly. From the data points given, the user can derive values needed to be equated to the respective moments to solve for the parameters.

[5] Logarithm function is monotonously increasing and hence the maximum point of any function would also result in the maximum of the logarithm of the function.

L-moments are then linear combinations of probability weighted moments (PWM), such that the PWM are defined as [Greenwood et al. (1979)]

$$\beta_r = E\{X[F(x)]^r\} = \int_{-\infty}^{\infty} xF(x)^r f(x)dx$$

where β_r is the r-th order PWM, $f(x)$ is the probability density function and $F_x(x)$ is the cumulative distribution function of the distribution of interest. The unbiased estimator of β_r is derived as [Landwehr et al. (1979), Hosking and Wallis (1997)]

$$\hat{\beta}_r = \sum_{i=1}^{n} \left[\frac{(i-1)(i-2)(i-3)\ldots(i-r)}{n(n-1)(n-2)(n-3)\ldots(n-r)} \right] x_i$$

with the first four b_i given as:

$$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}\text{, i. e. the mean of the data points}$$

$$\hat{\beta}_1 = \sum_{i=1}^{n-1} \left[\frac{n-i}{n(n-1)} \right] x_i$$

$$\hat{\beta}_2 = \sum_{i=1}^{n-2} \left[\frac{(n-i)(n-i-1)}{n(n-1)(n-2)} \right] x_i$$

$$\hat{\beta}_3 = \sum_{i=1}^{n-3} \left[\frac{(n-i)(n-i-1)(n-i-2)}{n(n-1)(n-2)(n-3)} \right] x_i$$

such that x_i are the n number of ordered data points $x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$. The first four L-moments are then defined as [Hosking (1990)]

$$\lambda_1 = \beta_0$$
$$\lambda_2 = 2\beta_1 - \beta_0$$
$$\lambda_3 = 6\beta_2 - 6\beta_1 - \beta_0$$
$$\lambda_4 = 20\beta_3 - 30\beta_2 - 12\beta_1 - \beta_0$$

By substituting the $\hat{\beta}_i$ into the β_i of the λ equations and solving the system of equations of these L-moments, the user can then obtain the parameters that best fit the Pd of interest.

There exist extensive lists of Pd that can be chosen to fit an empirical dataset. The appendix enumerates the list of Pd, their derived MLE, and the shapes of the Pd that are used in the example that follows.

Non-parametric Distributions

Non-parametric distributions (Npd) do not have a predefined shape or behaviour, and its resultant probability distribution is highly dependent upon the data points provided to be fitted. A notable method of Npd is the Kernel Density Estimation (KDE). KDE [Rosenblatt (1956)] assigns every data point a kernel, $K(x)$, with that kernel itself being a density function with properties of symmetrical at 0, $K(x) \geq 0$ and $\int_{-\infty}^{\infty} K(x)dx = 1$. The individual kernels are then summed up to form a Kernel density estimator that would map closely to the shape of a histogram of the dataset. The KDE function is defined by

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

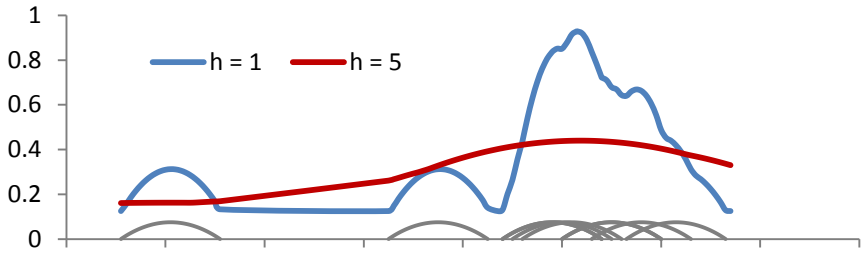


Figure 1.Example of Kernel Density Estimation

such that $h > 0$ is a smoothing parameter called the bandwidth. The KDE can trace the dataset more granularly by having smaller h, but at the trade-off of obtaining a less smoothed curve. In the world of finance, to observe prudency, it is advisable to structure the bandwidth such that the derived Npd has only one global maximum. This would result in $\forall n > 0$, if $f_h(x_t) =$ global maximum , then $f_h(x_t) \geq f_h(x_{t+n})$. This prudency measure makes sure that the further end (i.e. the tail) of the KDE does not have a higher probability than any point before that as it would be counter-intuitive. A list of commonly used kernels is stated in appendix.

Goodness of Fit Tests

After obtaining a satisfactory probability density function with a shape that resembles the histogram drawn (by simple observation), the user has to measure quantitatively how good the derived distribution matches the given dataset. The following section enumerates a list of GoF tests that would be major decision making tools in choosing the best-fitting distribution.

i) *Pearson's Chi-Squared test*

The Pearson's Chi-Squared test measures the GoF between the derived distributions against the data points. It does so by comparing the observed frequencies of occurrence (frequency of data points in their respective histogram bins) to the expected frequency using the derived distribution. This test produces a p-value statistics that can be used to compare against the Chi-Squared distribution to test the null hypothesis, h_o: difference in frequencies is equal to zero. The value of the test statistics is defined by [Pearson (1900)]:

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

such that O_i is the observed frequency of the dataset, and E_i is the expected frequency according the to the derived distribution. The test statistics would have *n-p* degree of freedom, with *p* being number of distribution's parameters plus 1. The advantage of this test is that it enables the user to test the GoF to obtain a certain degree of confidence level using the Chi-Square by itself (without the need to compare against another distribution). A low p-value would mean that the observed frequency deviated from the expected frequency with statistical significance. Functionally, a low p-value indicates that the distribution is not suitable for the dataset, and 0.05 (i.e. 5%) is usually a benchmark of confidence.

ii) *Akaike Information criterion*

The Akaike Information criterion (AIC) provides a relative measure of GoF among different distributions of interest. This means that using AIC on one single distribution provides no information about the GoF. AIC measures the fitness by emphasizing on GoF while deterring

overfitting. The distribution that produces the smallest AIC value is the preferred choice. The AIC value is given by [Akaike (1974)]

$$AIC = 2k - 2\ln L$$

such that k is the number of parameters of the distribution and L is the likelihood function mentioned above for the derived distribution. To compare distributions that has small number of data, the corrected AIC (AICc) is used [Hurvich and Tsai (1989)]

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

The AIC and AICc however does not measure an absolute GoF and hence if none of the chosen distributions is not a good fit, this test does not show that.

iii) *Cramer-von Mises test*

The Cramer-von Mises test measures the distance between the cumulative density function of the derived distribution against the dataset's cumulative histogram. It is defined by [Cramer (1928)]

$$w^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x)$$

with $F_n(x)$ being the cumulative frequency of the data points and $F(x)$ being the expected cumulative probability from the derived distribution. It tests for the null hypothesis that the dataset comes from the derived distribution. The test statistics for this test is

$$T = nw^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[\frac{2i-1}{2n} - F(x_i)\right]^2$$

such that n is the number of data points and x_i is the data points in increasing order. If the value of the test statistics is larger than the tabulated value, then null hypothesis is rejected suggesting that the derived distribution is not a good fit [Anderson (1962)].

iv) *Anderson-Darling test*

Taking a step further on the Cramer-von Mises test by adding a weighing factor of $\psi\big(F(x)\big) = [F_n(x)(1-F(x))]^{-1}$, Anderson-Darling test takes the form of [Anderson-Darling (1954)]

$$W_n^2 = n\int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi\big(F(x)\big)\, dF(x)$$

$$= n\int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{[F_n(x)(1-F(x))]}\, dF(x)$$

The test statistics then becomes

$$W_n^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\log F(x_i) + \log(1 - F(x_{n-i+1}))]$$

such that n is the number of data points, x_i is the data points in increasing order and $F(x_i)$ is the expected cumulative frequency. The Anderson-Darling's weighing factor naturally places more weight on the tail end of the data and hence would serve better purpose in trying to

determine the GoF of heavy-tailed distributions. Just like the Cramer-von Mises test, if the test statistics is greater than the tabulated value, the null hypothesis is rejected and the derived distribution is not a good fit. This test however, needs a different test statistics table for each distribution of interest requiring more resources than other GoF tests. The author did not include this test in the example, but is highlighting that this test is a beneficial one in particular to test the GoF of heavy tailed events, which has been discussed as a crucial characteristic in financial analysis.

v) *Mean Square Error (MSE) and Mean Absolute Deviation (MAD)*

Both the MSE and MAD measure the difference in density between the derived distribution and the empirical data. MSE squares the error terms, also in a way weighing the individual errors by the size of the errors themselves, while MAD sums the absolute value of the differences. They are given by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}(x) - f(x))^2 \qquad MAD = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{f}(x) - f(x)\right|$$

such that $\hat{f}(x)$ is the derived density and $f(x)$ is the empirical density. Naturally, MSE would put more weight on the errors that are bigger, resulting in bigger values of MSE for empirical data with outliers. The distributions with the smaller value of MSE and MAD are preferred.

Application on Financial Interaction Analysis

Having determined the best fitting distribution for the empirical data, the analysis of financial interaction that the author intends to introduce here is the application of Monte Carlo simulation onto the selected distribution to cross relates different distributions under different random scenarios. By generating random numbers bounded by (0,1),we use the inverse of the derived probability distribution's density function to generate the x-values. From the x-values, we can analyse the data according to the objective of the analysis.

For the best fitting exercise, the author has used 1000 positive daily yield hike i.e. price drop of the 1Y US treasury yield. The daily yield hikes are tabulated into bins of 50 basis points (bps) from 0 to 3500bps. These data are then fitted with different probability distribution of interests: Rayleigh, Pareto, Folded Normal, Weibull, Exponential and Gamma, all using MLE for parameter estimation. An additional Npd using the KDE is also fitted onto the data to obtain a resultant curve. These fitted distributions are then tested for GoF using the Pearson Chi-Square, Akaike Information Criteria, Cramer-von-Mises, and Mean Square Error & Mean Absolute Deviation. The best fit is then used in an example of estimating the ES of a portfolio.

In this example, the author is interested in calculating the ES of a hypothetical portfolio consisting of two assets (1Y US treasuries and 10Y US AAA corporate bonds[6]) with equal weights. The Gamma distribution is used to generate the x-values (i.e. yield hikes) of US treasuries while the Log-normal distribution is used to generate the x-values of US AAA Corporate bonds, by computing the inverses of both the distributions using randomly

[6] In this case the US AAA Corporate bond is only assumed to have a best-fitted distribution of log-normal without the process of best-fitting as done for the US treasuries.

generated numbers within (0,1) as inputs. By simulating 1000 occurrences of yield hike using each of the derived distributions for the two assets, the portfolio's overall expected loss can be calculated by averaging the tail end after a 95% cut-off. Repeating this exercise of calculating ES for 1000 times gave a good estimate of expected shortfall of this portfolio.

3. Results and Discussion

The result from the GoF test shows different best fit depending on the test used. However, in our particular cases, Gamma distribution stands out as the best fitted line for the US 1Y treasuries. Figure 4 in the appendix shows the fitted line of the distributions onto the histogram of the empirical data.

It is observed that the KDE would give the closest estimates of the empirical data. This however does not give much meaning as the KDE is built upon each individual points given and hence would naturally be a closer match than any of the Pd. Among the Pd, Gamma is lowest in AIC and has relatively low CvM and MSE & MAD.

	Rayleigh	Pareto	Folded Normal	Weibull	Exponential	Gamma	Kernel
Chi-Square	NA	646*	NA	7891	3330	3637	33*
AIC	5559	4257	4470	3955	12747	3949^	3901^
CvM	168.2	16.8	NA	64.2	1.2	1.3	0.7*
MSE & MAD	118.80 0.94	24.57 0.53	11.74 0.35	1.84^ 0.16	2.71 0.18	2.77 0.19	1.08^ 0.13

*Does not reject the distribution at 99% confidence interval[7]

^ Top 2 relatively most suitable distributions

Table 1. Goodness of Fit tests

Having determined the best fitted probability with their respective optimal parameters, the distributions produced expected shortfalls via repeated iterations to provide a result in Table 2 below. The portfolio ES obtained is then a reliable estimate of a potential yield hike that can be experienced by the portfolio.

Asset Type	US Treasuries		AAA Corporate Bonds	
Derived distribution	Gamma Distribution		Log Normal	
Optimized parameters	k	1.0106	mean	1.8053
	θ	2.6111	standard deviation	2.0277
Expected shortfall (yield hike %)	12.3321%		35.2846%	
Portfolio ES (yield hike %)	23.8083%			

Table 2. Results of ES from best fitted distribution

[7] The densities of the derived distribution and the empirical data are not statistically significant different at the 1% level.

Concluding Remarks

Despite the fact that statistical tools are primarily developed for application on physical science, the financial sector would stand to benefit from more robust probability modelling and application. Using distribution fitting and subsequently Monte Carlo simulation in determining the possible outcomes of financial interactions (not limited to just ES like in this paper) is a useful method to the provide user with forecasts of the data based on its empirical behaviour. This example is in no way to show that the US treasury or AAA Corporate yield hike takes any particular shape. Instead, it serves as an example of the process flow for the user to obtain a suitable probability distribution. Such analysis can be extended further to other datasets to build its distribution.

On a separate note, the Anderson-Darling GoF method is not applied in the example as this test requires extensive resources in building a test statistics table for each individual distribution. The author wishes to note that this area of research would be beneficial for further undertaking as the AD test puts emphasis on the tail end. Recent discussions on the financial industry appearing to have fat-tail distributions, especially given large losses happening more frequently than expected, is a supporting factor for possible future research in fatter tail distributions and methods to test the goodness of fit for such dataset.

References

Myung, I.J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology* **47**, 90-100.

Hosking, J.R.M. (1990). L-moments: Analysis and Estimation of Distribution using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society, Series B* **52**, 105-124.

Hosking, J.R.M. (1992). Moments or L moments? An example comparing two measures of distributional shape. *The American Statistician* **46** (3), 186-189.

Hosking, J.R.M., and Wallis J.R. (1997). Regional Frequency Analysis – An Approach Based on L-Moments. *Cambridge University Press.*

Greenwood, J.A., et al. (1979). Probability-weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research* **15** (4), 1049-1054.

Landwehr, J.M., Matalas, N.C., and Wallis, J.R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources* Research **15** (5), 1055-1064.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics* **27** (3), 832-837

Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. *Monographs on Statistics and Applied Probability*.

Epanechnikov, V.A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability and its Application* **14**, 153-158.

Mohammad, M., Osati, K., Sadeghi, S.A., Karimi, B., and Mobaraki, J. (2010). Determining Suitable Probability Distribution Models for Annual Precipitation Data. *Journal of Sustainable Development* **3** (1), 159-168.

Deka, S., Borah, M., and Kakaty, S.C. (2009). Distribution of Annual Maximum Rainfall Series of North-East India. *European Water (E.W.) Publications* **27/28** (1), 3-14.

Embrechts, P., Klüppelberg, C., Mikosch, T. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag Berlin Heidelberg 1997 Print, 152-167,283-370.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5* **50** (302), 157-175.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19** (6), 716-723.

Hurvich, C.M., and Tsai, C.L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297-307.

Anderson, T.W., and Darling, D.A. (1954). A Test of Goodness of Fit. *American Statistical Association Journal*, 765-769.

Anderson ,T.W. (1962). On the Distribution of the Two-Sample Cramer-von Mises Criterion. *The Annals of Mathematical Statistics, 1148-1159.*

Csorgo, S., and Faraway, J.J. (1994). The Exact and Asymptotic Distributions of Cramer-von Mises Statistics. *Journal of Royal Statistical Society* **58** (1), 221-234

Siddiqui, M.M. (1964). Statistical inference for Rayleigh distributions. *The Journal of Research of the National Bureau of Standands* **68D** (9), 1007

Newman, M.E.J. (2004). Power laws, Pareto distribution and Zipf's law. *Contemporary Physics* **46**, 323.

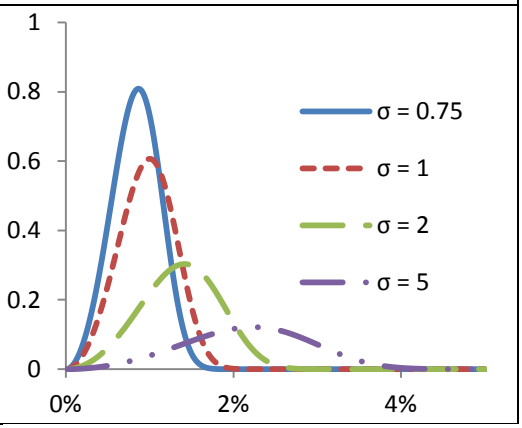
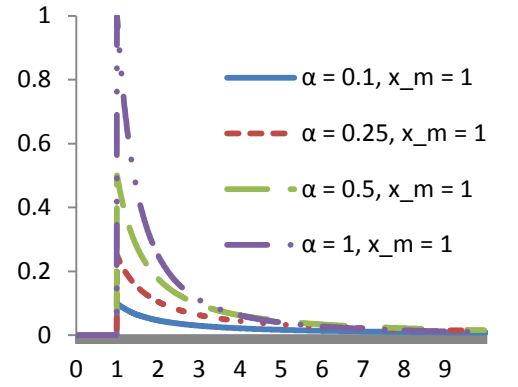
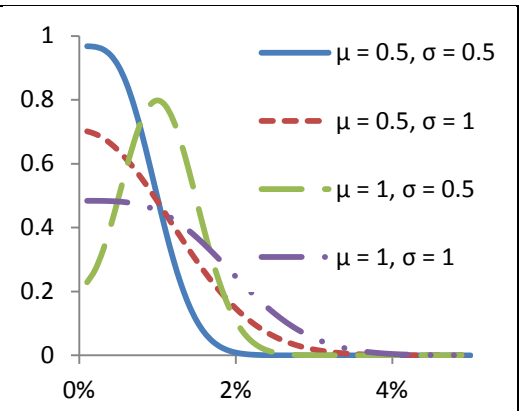
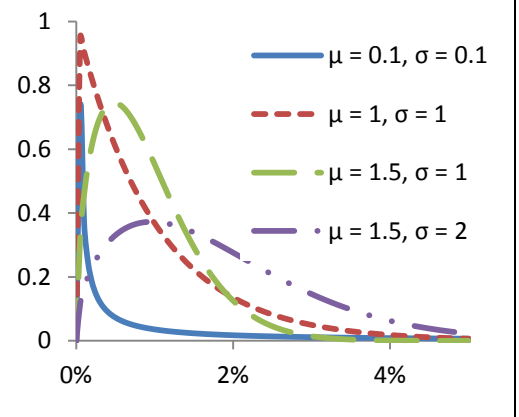
Johnson, N.L. (1962). The folded normal distribution: accuracy of the estimation by maximum likelihood. *Technometrics* **4** (2), 249.

Lei, Y. (2008). Evaluation of three methods for estimating the Weibull distribution parameters of Chinese Pine (*Pinus tabulaeformis*). *Journal of Forest Science* **54** (12), 566-571.

Minka, T. (2002). Estimating a Gamma distribution. *Microsoft Research*. http://research.microsoft.com/en-us/um/people/minka/papers/minka-gamma.pdf

Data source: US Government Treasury website and Bloomberg

Appendix

Distribution Name	Probability Density Function	Shape
Supported on partial real line, [0,∞]		
Rayleigh distribution	$f(x;\sigma) = \frac{x}{\sigma^2}\exp(\frac{-x^2}{2\sigma^2})$ σ = scale parameter MLE (Siddiqui 1964) $\hat{\sigma} = \sqrt{\frac{1}{2N}\sum_{i-1}^{N} x_i^2}.\frac{4^N N!\,(N-1)!\sqrt{N}}{(2N)!\sqrt{\pi}}$	
Pareto distribution	$f(x;x_m,\alpha) = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}$ x_m = scale parameter α = shape parameter MLE (Newman 2004) $\widehat{x_m} = \text{Min}\{x_i\}$ $\hat{\alpha} = \frac{N}{\sum_{i=1}^{N}(\ln x_i - \ln \widehat{x_m})}$	
Folded normal distribution	$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\left[\exp\left(-\frac{(-x-\mu)^2}{2\sigma^2}\right) + \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right]$ μ = location parameter σ = scale parameter MLE μ = mean/mod σ = standard deviation	
Weibull distribution	$f(x;k,\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}.\exp(-(\frac{x}{\lambda})^k)$ λ = scale paramter k = shape parameter MLE (Lei, 2008) $\hat{\lambda} = \left[\frac{1}{N}\sum_{i=1}^{N}(x_i^k - x_N^k)\right]^{-k}$ $\widehat{k^{-1}} = \frac{\sum_{i=1}^{N}(x_i^k \ln x_i^k - x_i^k \ln x_N)}{\sum_{i=1}^{N}(x_i^k - x_N^k)} - \frac{1}{N}\sum_{i=1}^{N}\ln x_i$ (k has to be solved using iterative procedure, as there is no closed form solution)	

Appendix

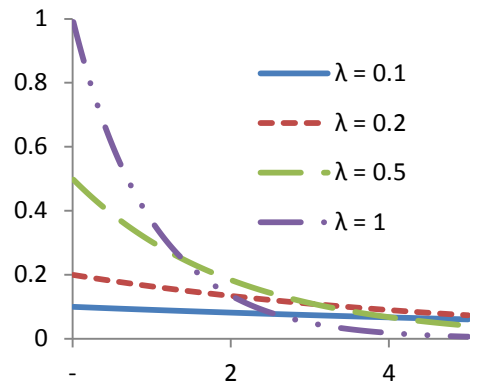
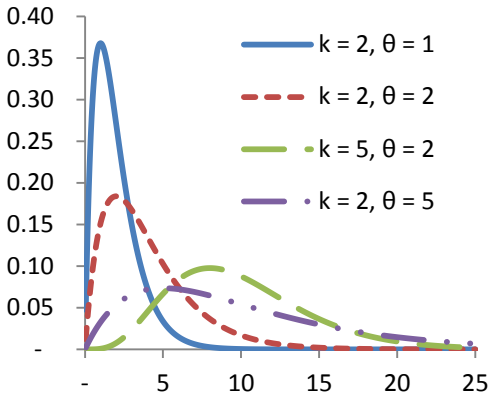
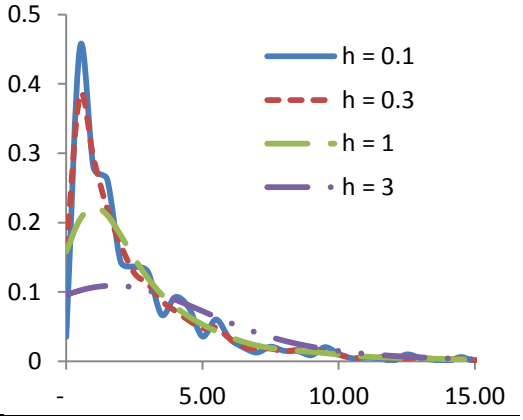
Distribution Name	Probability Density Function	Shape
Supported on partial real line, [0,∞]		
Exponential distribution	$f(x;\lambda) = \lambda\exp(-\lambda x)$ λ = rate parameter MLE: $\hat{\lambda} = \frac{1}{\bar{x}}$	
Gamma distribution	$f(x;k,\theta) = \frac{x^{k-1}\exp(-\frac{x}{\theta})}{\theta^k\Gamma(k)}$ k = scale parameter > 0 θ = shape parameter > 0 MLE (Minka 2002) $\hat{\theta} = \frac{1}{kN}\sum_{i=1}^{N} x_i$ $\hat{k} = \frac{3 - s + \sqrt{(s-3)^2 + 24s}}{12s}$ such that $s = \ln\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) - \frac{1}{N}\sum_{i=1}^{N}\ln(x_i)$	
Kernel Density Estimator	$f_h(x) = \frac{1}{hN}\sum_{i=1}^{N} K(\frac{x - x_i}{h})$ h = bandwidth	

Table 3. List of probability distributions used

Appendix

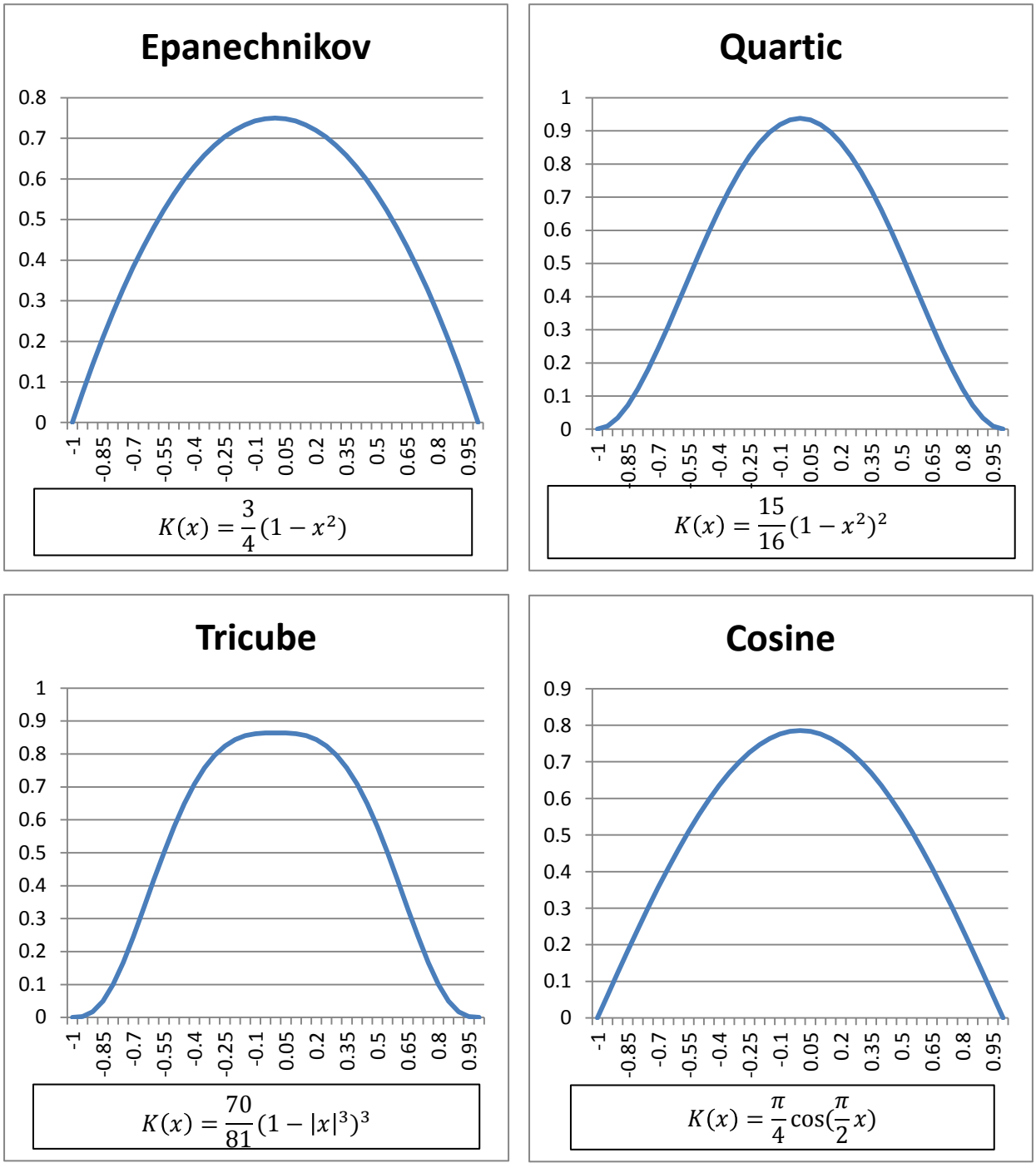
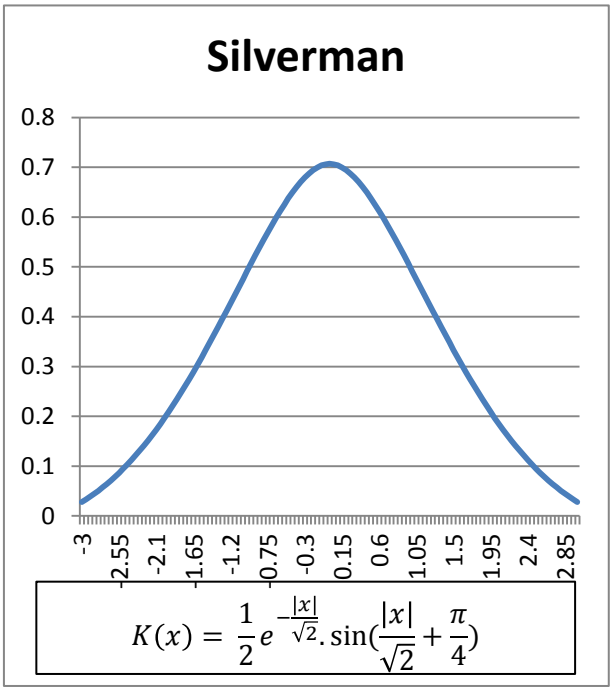
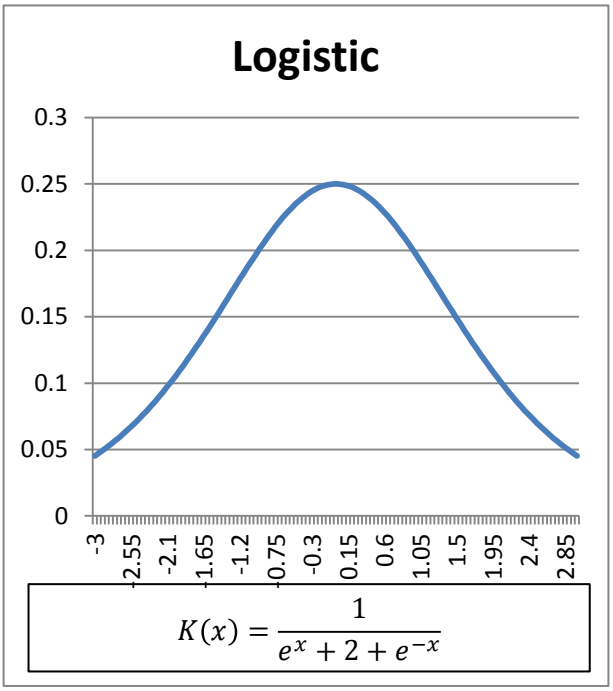


Figure 2: Kernels that are supported only on [0,1]

Appendix





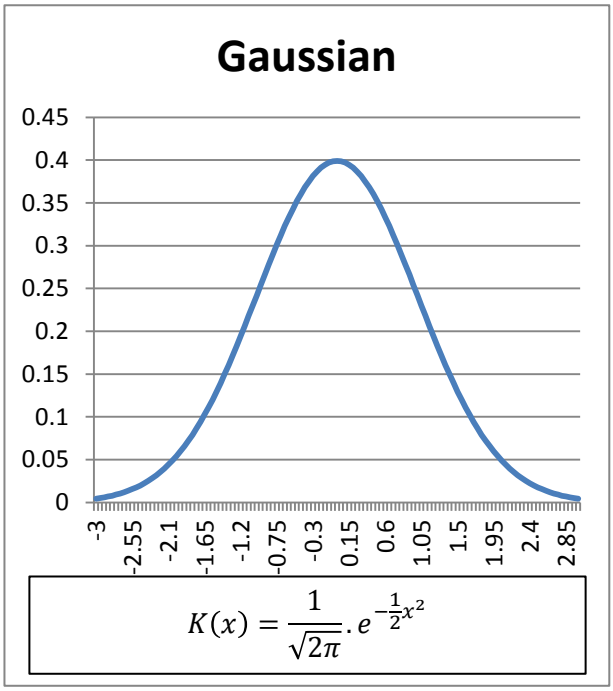


Figure 3: Kernels that are supported on the entire Real line

Appendix

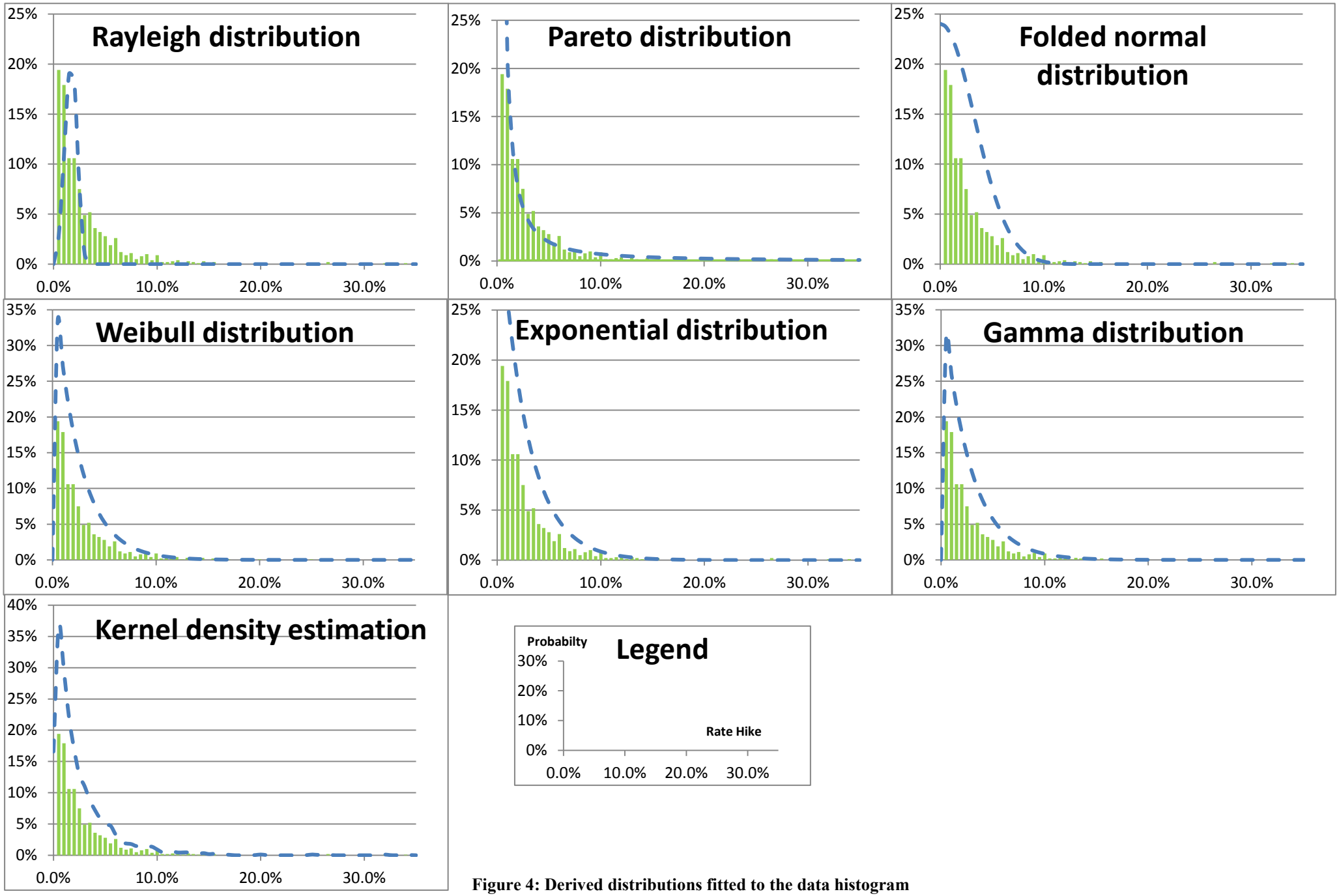


Figure 4: Derived distributions fitted to the data histogram