Tail Weighted Probability Distribution Parameter Estimation

Malcolm Kemp

Original paper prepared: 12 June 2013
This version: 20 May 2016 (with an updated section 7)

Abstract

In this paper we introduce four ways of estimating probability distribution parameters that target a good fit to a user selected part of the distributional form (e.g. one or both tails). We analyse the characteristics of the resulting parameter estimates for a wide range of commonly used probability distributions. Three of the methodologies reviewed are similar, are motivated by maximum likelihood principles and give the same results as traditional (unweighted) maximum likelihood estimation in the special case where all quantiles within the distributional form are given equal weight. The fourth methodology involves a direct weighted least squares fit to a suitable quantile-quantile plot.

1. Introduction

Traditional parameter estimation techniques generally give equal weight to every observation in the relevant observation set. For example, if observations $X_1, X_2, \ldots, X_n$ come from a normal distribution $N(\mu, \sigma^2)$ then commonly the mean, μ, is estimated as the unweighted average of the observations, i.e. by $\hat{\mu} = \sum_{i=1}^{n} X_i / n$. Two common ways of deriving parameter estimates are least squares and maximum likelihood. In this example, least squares involves selecting $\hat{\mu}$ to minimise $\sum_{i=1}^{n} C_i$ where $C_i = (X_i - \hat{\mu})^2$ and maximum likelihood involves selecting $\hat{\mu}$ to minimise $\sum_{i=1}^{n} C_i$ where $C_i = -\log f(X_i|\mu = \hat{\mu})\ (X_i - \hat{\mu})^2$ and f is the probability density function. The two happen in this example to produce the same estimate, namely the unweighted average as above.

It is relatively straightforward to modify these approaches to give greater weight to different observations as long as the selection of the weights is not dependent on the ranking of the observation in the dataset. For example, we may have a time series of data which we are analysing for risk management purposes and we may believe that more recent data is more relevant to the estimation problem in hand. This can be done by altering the formula to be minimised to be $\sum_{i=1}^{n} w_i C_i$ where the weights are suitably chosen (e.g. here giving greater weight to more recent observations) but the C_i are otherwise unaltered. The w_i might for example be chosen to exhibit an exponential decay backwards in time to give greater weight to more recent data. Formally this can be viewed as a Bayesian style approach where the μ for different observations come from different priors that (here) all have the same underlying mean but have varying underlying standard deviations, so we effectively place more credibility on some observations than on others when estimating the overall value of μ.

However, such an approach does not work if aim is to focus greater weight on a particular part of the distributional form, i.e. on a particular range of quantile values. For example, if we were merely to give a large weight to observations in the bottom quartile of the observation set and a small weight to the remainder of the observation set then the weighted mean as calculated above would be biased downwards relative to the true value characterising the distributional form.

Why might we wish to focus on a particular area of the distributional form? Again risk management provides an obvious example. Typically risk managers are more interested in extreme outcomes (particularly downside extreme outcomes) than they are in more run-of-the-mill outcomes. Capital

requirements for financial institutions may target outcomes with a given low quantile, e.g. the 99.5th percentile 1 year Value-at-Risk (i.e. '1 in 200 year event') targeted by the proposed Solvency II regulatory framework for EU insurers. Whilst practitioners often in practice assume a distributional family and fit to the entire dataset as above, this can often be inappropriate. For example, if we believe that drivers of extreme events are less likely to be present in more 'normal' circumstances then a distributional form estimated primarily from observations outside the relevant tail may be a poor guide to the distributional form within the tail.

Practitioners cognisant of these issues may seek to use more complicated distributional families with additional shape parameters that mainly affect only the shape in the relevant tail. However, there are not many commonly recognised and easy to manipulate distributional families that have such characteristics. So it is more common for such practitioners to use Extreme Value Theory (EVT) techniques. For risk management purposes these generally involve assuming that the tail can be approximated by a generalised Pareto distribution (GPD) and selecting from this distributional family using equally weighted approaches applied to the concatenated dataset formed by discarding all observations outside the relevant tail[1]. However traditional EVT has some weaknesses:

(a) Choice of where the tail starts is potentially arbitrary. In any case an all-or-nothing approach (in which data either is in the tail or it isn't) potentially throws away data relevant to the estimation problem just outside the deemed start of the tail.

(b) Whilst often the focus is on one of the tails of the distribution this is not always the case. It would be desirable to identify methodologies that allowed focus to be applied to *any* arbitrary part of the distributional form rather than to just one tail for a univariate distribution. For multivariate distributions we might want to be able to focus on some parts of the joint edge region more than others.

(c) Using EVT as described above for risk management purposes involves a prior view that the tail of the distribution should be well modelled by a suitable member of the GPD family. Whilst theoretical arguments underpinning EVT indicate that commonly used distributional forms nearly always tend to such a distribution in their tail it is relatively easily to construct distributions which do not exhibit this property. EVT assuming a GPD form is, in essence, estimating the tail distributional form by applying a particular form of interpolation (if the quantile of interest is within the range of the observation set) or extrapolation (if it is outside the range of the observation set). There is no inherently good reason why, when applied to real-life data sets coming from an unknown distributional form, GPD based interpolation or extrapolation approaches will be more (or less) accurate than other any plausible approach to interpolation or extrapolation. The choice inevitably involves some application of judgement, i.e. some application of prior views.

The aim of this paper is to introduce alternative more generalised parameter estimation approaches that as far as possible circumvent these issues. Whilst they share some similarities with EVT and may often in practice be applied in similar circumstances they in principle do not need to focus on extreme values and referring to them merely as variants of EVT may therefore be inappropriate. However, for convenience we will refer to the parameter estimates we propose as 'tail weighted'

[1] There are two main EVT results namely 'block maxima' results (which involve the generalised extreme value distribution) and 'peaks over thresholds' results (which involve the generalised Pareto distribution). The latter are generally considered to be more applicable to risk management activities. Selection of the relevant member of the GPD family may be undertaken using so-called Hill estimators or other maximum likelihood style techniques.

parameter estimates since the most obvious area of application is where greater weight is given to the tail of a distribution.

2. Reformulating maximum likelihood to depend on ranked data

As an estimation approach, maximum likelihood has many attractive features. In principle it almost always provides unique parameter estimates[2] that have desirable asymptotic properties[3]. It is a generic approach that can be applied to virtually any distributional family. In theory it can handle multi-parameter families (including cases where some or all of the parameters are not allowed to vary) as easily as it can handle single-parameter ones. For most commonly specified multi-parameter probability distributions it seems able to be implemented relatively robustly using practical computational techniques[4]. It is therefore natural to explore whether such an approach, with suitable modifications, can be applied to the task of creating tail weighted parameter estimates.

Suppose we have a set of observations $X_1, X_2, \dots, X_n$ or when ordered $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Suppose that these are independent draws from a (continuous) probability distribution with probability density function $f(x|\theta)$ and cumulative distribution function $F(x|\theta)$ where each of these is parameterised by a parameter θ (which in general may be a vector, to cater for multi-parameter families). Suppose we have an ordered set of values $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$.

Suppose we define the event Y_r as involving $X_{(1)} \in \left(x_{(2)}, x_{(2)} + dx\right)$, $X_{(2)} \in \left(x_{(2)}, x_{(2)} + dx\right)$, …, $X_{(m)} \in \left(x_{(r)}, x_{(r)} + dx\right)$. For Y_1 to occur, i.e. for $X_{(1)}$ to lie between $x_{(1)}$ and $x_{(1)} + dx$ we need to have 0 observations in the range $\left(-\infty, x_{(1)}\right)$, 1 of the n observations in the range$\left(x_{(1)}, x_{(1)} + dx\right)$ and the remaining $n-1$ observations in the range $\left(x_{(1)} + dx, +\infty\right)$. Therefore the probability of this occurring (for small enough dx) is $P(Y_r|\theta) \equiv P\left(X_{(1)} \in \left(x_{(1)}, x_{(1)} + dx\right)|\theta\right) = p_1 dx$ say, where:

$$p_1 = nf\left(x_{(1)}|\theta\right)\left(1 - F\left(x_{(1)}|\theta\right)\right)^{n-1}$$

More generally the probability of Y_r is $p_r dx^r$ say, where:

$$p_r = \binom{n}{r} f\left(x_{(1)}|\theta\right) \dots f\left(x_{(r)}|\theta\right)\left(1 - F\left(x_{(r)}|\theta\right)\right)^{n-r}$$

where $\binom{n}{i} = \frac{n!}{r!(n-r)!}$ is the binomial coefficient.

---

[2] A commonly used alternative approach is method of moments. If there are m parameters to estimate then the method of moments approach involves choosing suitable parameter values that exactly fit the first m moments of a distribution, i.e. $E(X^r)$ for $r = 1, \dots, m$. However, some families do not have sufficient numbers of finite moments to allow method of moments to be used. Even where the relevant moments do exist the relevant m equations may not be uniquely solvable.

[3] As the sample size increases to infinity maximum likelihood estimation exhibits consistency (i.e. sequences of maximum likelihood estimators converge in probability to the value being estimated), asymptotic normality (i.e. they tend towards a normal distribution which has the mean equal to the value being estimated and covariance matrix equal to the inverse of the Fisher information ratio) and efficiency (i.e. no asymptotically other unbiased estimator has lower asymptotic mean squared error).

[4] Maximum likelihood involves identifying the global maximum of a function. For functions of more than one variable (as would be the case if we are estimating more than one parameter for a given distributional family) this can be inherently challenging computationally even if the function is continuous. Such a function can have many local maxima which the algorithm might select instead of the global maximum. However, it seems that in practice for most commonly specified multi-parameter probability distribution families the likelihood function is well enough behaved in this respect to allow standard maximisation algorithms to reach the global maximum given sensible selection of initial starting values, if the dataset is of a reasonable size.

The (unweighted) likelihood of the overall observation set is $L \equiv p_n/dx^n$. Traditional maximum likelihood estimation involves identifying the value of θ that maximises L or equivalently $\log L$ as L is positive and $\log(.)$ is a monotonic function.

The Y_r are nested, so if Y_r occurs then Y_{r-1} must also have occurred (although not necessarily vice versa). The probability of Y_r conditional on Y_{r-1} (for parameter choice θ) is thus :

$$P(Y_r|Y_{r-1},\theta) = \frac{P(Y_r, Y_{r-1}|\theta)}{P(Y_{r-1}|\theta)} = \frac{p_r dx^r}{p_{r-1}dx^{r-1}} = \frac{p_r}{p_{r-1}}dx$$

$$\Longrightarrow P(Y_r|Y_{r-1},\theta) = \frac{\binom{n}{r} f\left(x_{(r)}|\theta\right)\left(1-F\left(x_{(r)}|\theta\right)\right)^{n-r}}{\binom{n}{r-1}\left(1-F\left(x_{(r-1)}|\theta\right)\right)^{n-(r-1)}}dx$$

Suppose we define the event Y_0 to involve any outcome, so $P(Y_0|\theta) = 1$, and for this purpose view $x_{(0)}$ as being equal to $-\infty$. We may then express the likelihood of the overall dataset as follows:

$$L = \frac{P(Y_1|Y_0,\theta)P(Y_2|Y_1,\theta)\ldots P(Y_n|Y_{n-1},\theta)}{dx^n}$$

$$\Longrightarrow \log L = \sum_{i=1}^{n} r_{(i)} = \sum_{i=1}^{n} k_{(i)} + \sum_{i=1}^{n} C^*_{(i)}$$

where:

$$r_{(i)} = \log\left(\frac{\left(1-F\left(x_{(i)}|\theta\right)\right)^{n-i}}{\left(1-F\left(x_{(i-1)}|\theta\right)\right)^{n-(i-1)}}\frac{(n-i+1)}{i}f\left(x_{(i)}|\theta\right)\right)$$

$$C^*_{(i)} = \log\left(\frac{\left(1-F\left(x_{(i)}|\theta\right)\right)^{n-i}}{\left(1-F\left(x_{(i-1)}|\theta\right)\right)^{n-(i-1)}}\frac{(n-i+1)}{ie^{k_{(n)}}}f\left(x_{(i)}|\theta\right)\right)$$

for arbitrary fixed $k_{(i)}$.

As $\sum_{i=1}^{n} k_{(i)}$ is fixed, maximum likelihood estimation may be expressed as involving minimising $\sum_{i=1}^{n} C^*_{(i)}$. This formulation involves elements that depend on the ordered data. This contrasts with the more usual formulation of maximum likelihood which involves minimising $\sum_{i=1}^{n} C_i$ where $C_i = -\log f(X_i|\mu = \hat{\mu})$ and depends merely on the unordered data.

3. Introducing ranking dependent weights

As noted previously, to introduce weights into maximum likelihood in the case where weights are *not* ranking dependent involves minimising $\sum_{i=1}^{n} w_i C_i$ for suitably chosen w_i. One way of introducing weights into maximum likelihood in cases where the weights *are* ranking dependent is instead to minimise $\sum_{i=1}^{n} w^*_{(i)} C^*_{(i)}$ for suitably chosen $w^*_{(i)}$. For example we might give a weight of 1 to any observation in the lower quartile and a weight of zero to any observation in the upper three quartiles if we wanted to identify a fit that targeted behaviour in the lower quartile and was indifferent to fit elsewhere. We might call this approach *tail weighted maximum likelihood estimation* (TWMLE).

As characterised above, the $C^*_{(i)}$'s are not independent of choice of $k_{(i)}$. However, choice of $k_{(i)}$ does not actually affect the resulting TWMLE estimator since:

$$\sum_{i=1}^{n} w_i^* C^*_{(i)} = \sum_{i=1}^{n} w^*_{(i)} \log\left(\frac{\left(1 - F\left(x_{(i)}|\theta\right)\right)^{n-i}}{\left(1 - F\left(x_{(i-1)}|\theta\right)\right)^{n-(i-1)}} \frac{(n-i+1)}{i} f\left(x_{(r)}|\theta\right) \right) - \sum_{i=1}^{n} w^*_{(i)} k_{(i)}$$

and $\sum_{i=1}^{n} w^*_{(i)} k_{(i)}$ is fixed so $\sum_{i=1}^{n} w_i^* C^*_{(i)}$ will be maximised for the same θ whatever the selection of $k_{(i)}$.

As we will generally only be interested in the end answer, we will assume in this paper that $k_{(i)} = 0\ \forall i$ unless otherwise stated. However, if needed, we might choose values for $k_{(i)}$ that have particular presentational attractions. For example, we might choose them so that $C^*_{(i)} - \log f\left(x_{(i)}|\theta\right)$ is approximately constant for different i if the $F\left(x_{(i)}|\theta\right)$ are approximately equally spaced (which on average should be approximately the case). In situations where all the $w^*_{(i)}$ are nearly equal then such a presentation would have the advantage of more obviously appearing to give nearly the same prominence to each $x_{(i)}$.

In some circumstances we may have a weighting element that is ranking dependent *as well as* some component that that is independent of rank. For example, we might want to focus on distributional fit in the lower tail (e.g. lower quartile) because we think that it is more relevant for risk management purposes *and* we might also view more recent data as more relevant to the estimation task at hand. This can be incorporated by minimising $\sum_{i=1}^{n} w_i w^*_{(i)} C^*_{(i)}$ with the w_i suitably chosen based on (here) time observation occurred.

4. Up, down and averaged tail weighted maximum likelihood estimators

There is, however, a nicety that complicates the above picture. With unranked maximum likelihood, each C_i depends only on the probability density of a single observation, i.e. x_i and does not depend on how the data is ordered. However the $C^*_{(i)}$ introduced above depend not just on $x_{(i)}$ but also on the next lower observation in the ordering (via the term in $\left(1 - F\left(x_{(r-1)}|\theta\right)\right)^{n-(r-1)}$). Thus in the case where the w_i^* vary we will get a different answer depending on which way we order the observations or equivalently whether we use the above prescription to fit $-x_i$ (with the shape of the distribution inverted) or x_i (with the shape of the distribution not inverted).

This direction dependency seems theoretically undesirable and we can circumvent it as follows:

(a) We introduce one TWMLE approach which is the same as above, i.e. an 'up direction' tail TWMLE estimator that has $C^*_{(i)} = r^{up}_{(i)}$, where:

$$r^{up}_{(i)} = \log\left(\frac{\left(1 - F\left(x_{(i)}|\theta\right)\right)^{n-i}}{\left(1 - F\left(x_{(i-1)}|\theta\right)\right)^{n-(i-1)}} \frac{(n-i+1)}{i} f\left(x_{(i)}|\theta\right) \right)$$

(b) We introduce a second, 'down direction' TWMLE approach which uses the same underlying approach but with the order of observations inverted, so if we adopt the convention that $x_{(n+1)} = +\infty$ this has $C^*_{(i)} = r^{down}_{(i)}$, where:

$$r^{down}_{(i)} = \log\left(\frac{F(x_{(i)}|\theta)^{i-1}}{F(x_{(i+1)}|\theta)^{i}}\frac{i}{(n-i+1)}f(x_{(i)}|\theta)\right)$$

(c) We introduce a third 'average' TWMLE approach which involves the average of 'up' and 'down' approaches and is thus independent of order direction, involving $C^*_{(i)} = r^{average}_{(i)}$, where:

$$r^{average}_{(i)} = \frac{1}{2}\left(r^{up}_{(i)} + r^{down}_{(i)}\right)$$

5. Generalising to multivariate distributions

Standard maximum likelihood can be applied to multivariate as well as univariate distributions. In principle we can also apply tail weighted maximum likelihood to multivariate distributions. However, when doing so we need to define an ordering. One way of doing this is to target a particular tail direction although we could also define an ordering which referred to distance from the middle of the distribution.

6. Least squares approaches

An alternative approach is motivated by the fitted cubic quantile approach described in e.g. Section 9.5.5 of Kemp (2009) or Section 2.4.5 of Kemp (2011). A fitted polynomial estimation approach involves the following:

(a) We assume that the plot of the observed versus expected quantile is a specific polynomial (e.g. a cubic), where 'expected' is e.g. the quantile value assuming that the distribution were normal with mean and standard deviation equal to the sample mean and sample standard deviation

(b) We then choose the polynomial coefficients to minimise the (weighted) least squared divergence between the observed quantiles and the expected quantiles, i.e. we choose $a_0, a_1, \ldots, a_m$ to minimise $\sum_{i=1}^{n} w^{**}_{(i)} C^{**}_{(i)}$ where typically:

$$C^{**}_{(i)} = \left(X_{(i)} - Z_{(i)}\right)^2$$

where:

$$Z_{(i)} = a_0 + a_1 y_{(i)} + \cdots + a_m y^m_{(i)}$$

$$y_{(i)} = m + sN^{-1}\left(\frac{i-1/2}{n}\right)$$

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - m)^2}$$

and $N^{-1}(x)$ is the inverse Normal cumulative distribution function.

(c) As in most cases most observations clump towards the middle of the distribution and we would typically be interested mainly in tail behaviour, we might choose weights that scaled in line with distance between consecutive expected quantile values and/or only included some parts of the distributional form, e.g. merely the lower quartile of observations. Using weights that scale in line with distance between consecutive expected quantile values means that visually the fit (for a quantile-quantile plot) will appear reasonably good across the entire distributional form, if the *x*-axis of our chart relates to these quantile values. Superposition of a weighting schema that focuses only on specific parts of the distributional form (e.g. between specific quantile ranges) allows us to target the best (visual) fit merely for that part of the distributional form in such charts.

An advantage of a fitted polynomial (to the Normal quantile function) approach is that it is relatively simple and easy to implement given merely access to a function returning the inverse normal cdf, such as the NORMSINV() function in Mcrosoft Excel. However it has some disadvantages including:

(i) Not all distributions have finite first or second moments, so the 'expected' normal distribution against which the actual observations might best be compared is not always well-posed.

(ii) Not all cubics (or other polynomials) that the fitting might generate correspond to actual cumulative distribution functions

(iii) The methodology lacks clear asymptotic properties that add lustre to maximum likelihood.

(iv) The choice of weighting schema is relatively arbitrary and is difficult to interpret in a formal statistical sense. This is especially true for the weight to ascribe to the very last data point, as the distance between it and the next (notional) consecutive point further out into the tail is ill-posed. But even for the remaining data points it is not obvious what meaning in a statistical sense to ascribe to any particular weighting schema, because all that the approach is targeting is a good visual fit to the data (plotted in an appropriate way) in a some chosen part of the distributional form. At the edges of the distribution the least squares prescription being adopted may also be unsound as it may be assuming the wrong 'expected' value for the relevant quantile. For example, the expected value of the $1/n$'th quantile of a large number of samples each of n observations may not be necessarily be approximately $F^{-1}\left(\frac{1}{2n}\right)$ (indeed this expected value may not exist for some distributions).

The two main computational weaknesses of this approach, i.e. (i) and (ii), can both be circumvented, if inverse functions for target distributional families are readily available, by setting $Z_{(i)}$ to the corresponding inverse function for the selected distributional family, i.e. by setting $Z_{(i)} = F^{-1}\left(\frac{i-1/2}{n}|\theta\right)$ and then selecting θ to minimise $\sum_{i=1}^{n} w_{(i)}^{**} C_{(i)}^{**}$. We might call an approach that includes such a refinement a *tail weighted least squares estimation* (TWLS) approach.

However, this refinement does not circumvent the remaining weaknesses including the difficulty of interpreting the end answer. For many commonly used distributional families it is also more time consuming to evaluate inverse functions than to evaluate pdfs and cdf so the resulting computations may also be more computer resource intensive than for TWMLE.

7. Fitting to selected quantile points

A specialised use of the tail weighted least squares approach arises if the input involves (a typically small number of) VaR/quantile values at specified quantile points and our aim is to fit the distributional form so that it reproduces as closely as possible these quantile values at these quantile points. This might arise if we are seeking to obtain the distributional form via 'expert judgement' and the experts providing the judgemental inputs are expected to be best able to express judgements using VaR values, e.g. that 10% of outcomes will be greater than x_1, 50% greater than x_2, 90% greater than x_3, say. If we have at least as many such judgement expressions as we have free parameters in the selection of the distributional form then it is relatively straightforward to modify the TWLS approach to identify the distributional family that best fits these judgemental inputs. The only specific change needed is to amend the formula for $Z_{(i)}$ to be as follows:

$$Z_{(i)} = F^{-1}\big(q_{(i)}|\theta\big)$$

where the $q_{(i)}$ are the quantile points to which the $X_{(i)}$ are assumed to relate. The approach described in 6 is then a special case of the above with $q_{(i)} = (i - 1/2)/n$.

However, we have noted above that the TWLS approach has weaknesses relative to the TWMLE approaches described earlier. It would therefore be attractive to re-express the TWMLE approaches so that they too were capable of being applied to arbitrary quantile points. This can be done as follows, if the quantile points for which inputs are available are $q_{(i)}$ $(0 < q_{(1)} < \cdots < q_{(n^*)} < 1)$:

$$r_{(i)}^{up} = n\Big(\big(1 - \bar{q}_{(i+1/2)}\big)\log\Big(1 - F\big(x_{(i)}|\theta\big)\Big) - \big(1 - \bar{q}_{(i-1/2)}\big)\log\Big(1 - F\big(x_{(i-1)}|\theta\big)\Big)\Big) + \big(\log\big(1 - \bar{q}_{(i-1/2)}\big) - \log\big(\bar{q}_{(i+1/2)}\big)\big) + \log f\big(x_{(i)}|\theta\big)$$

$$r_{(i)}^{down} = n\Big(\big(1 - \bar{q}_{(i-1/2)}\big)\log\Big(F\big(x_{(i)}|\theta\big)\Big) - \big(1 - \bar{q}_{(i+1/2)}\big)\log\Big(F\big(x_{(i+1)}|\theta\big)\Big)\Big) + \big(\log\big(\bar{q}_{(i+1/2)}\big) - \log\big(1 - \bar{q}_{(i-1/2)}\big)\big) + \log f\big(x_{(i)}|\theta\big)$$

$$r_{(i)}^{average} = n\Big(\big(1 - \bar{q}_{(i+1/2)}\big)\log\Big(1 - F\big(x_{(i)}|\theta\big)\Big) - \big(1 - \bar{q}_{(i-1/2)}\big)\log\Big(1 - F\big(x_{(i-1)}|\theta\big)\Big)\Big) + n\Big(\big(1 - \bar{q}_{(i-1/2)}\big)\log\Big(F\big(x_{(i)}|\theta\big)\Big) - \big(1 - \bar{q}_{(i+1/2)}\big)\log\Big(F\big(x_{(i+1)}|\theta\big)\Big)\Big) + \log f\big(x_{(i)}|\theta\big)$$

where $\bar{q}_{(i+1/2)} = \frac{1}{2}\big(q_{(i)} + q_{(i+1)}\big)$ for $\frac{1}{2} \le i + \frac{1}{2} \le n^* - \frac{1}{2}$ and we adopt the convention that $\bar{q}_{(-1/2)} = 0$ and $\bar{q}_{(n^*+1/2)} = 1$, $F\big(x_{(0)}|\theta\big) = 0$ and $F\big(x_{(n+1)}|\theta\big) = 1$

These formulae then exactly reproduce those in section 4(c) in the special case where $q_{(i)} = (i - 1/2)/n$.

However, a subtlety arises that was not specifically covered in the original June 2013 version of this paper. If the $q_{(i)}$ are arbitrary then it ceases to be clear what value to use for n in the above formulae, as it cannot be directly derived from n^*, the number of data points supplied to the algorithm. An approach that approximates to the position applying if there are equally and completely spaced quantiles in situations where only a contiguous subset of them is available involves the following:

$$n_{effective} = \left|\frac{n^* - 1}{q_{(n^*)} - q_{(1)}}\right|$$

$$\bar{q}_{(-1/2)} = max\left(0, \bar{q}_{(1/2)} - \frac{1}{n_{effective}}\right)$$
$$\bar{q}_{(n^*+1/2)} = min\left(1, \bar{q}_{(n^*-1/2)} + \frac{1}{n_{effective}}\right)$$
$$F\left(x_{(0)}|\theta\right) = max\left(0, F\left(x_{(1)}|\theta\right) - \frac{1}{n_{effective}}\right)$$
$$F\left(x_{(n+1)}|\theta\right) = min\left(1, F\left(x_{(n)}|\theta\right) + \frac{1}{n_{effective}}\right)$$

Please note that 'edge' effects mean that this approach will not exactly reproduce the results that would otherwise have been returned had a more complete set of quantiles been passed to the algorithm but with some of these quantiles then given a weight of zero when carrying out the maximisation process.

8. Practical application of these tools

The online toolkit on Nematrian website (www.nematrian.com) includes four web service functions that implement TWLS and TWMLE. Each accepts arbitrary $q_{(i)}$ as inputs (as long as $0 < q_{(1)} < \cdots < q_{(n)} < 1$), i.e. each implements the generalised version of the relevant approach as set out in section 7. Each returns a vector of distributional parameter estimates that identifies the member of the relevant distributional family that optimises the relevant target fit characteristic.

The relevant functions are:

(a) MnProbDistTWLS. Returns relevant TWLS estimates. Inputs are *DistributionName* (more than 30 underlying distributions are currently supported as well as shift and scale adjusted variants of these distributions), *InputValues* (i.e. the $X_{(i)}$ as above), Weights (i.e. the $w^*_{(i)}$ or $w^{**}_{(i)}$ as above), *QuantilePoints* (i.e. the $q_{(i)}$ as above) and *OpeningVariableChoices* (i.e. seed values for the search algorithm, e.g. if we were apply the algorithm to the normal distribution, so *DistributionName* = "normal" then these would be seeds for the mean and standard deviation respectively).

(b) MnProbDistTWLSRestricted. Same as (a) but with an additional input array *AllowParameterToVary* of the same length as *OpeningParameterChoices*. If relevant entry in *AllowParameterToVary* is set to FALSE then parameter optimisation proceeds on the basis that the relevant distributional parameter cannot vary from its value as defined in *OpeningVariableChoices*.

(c) MnProbDistTWMLE. Targets TWMLE. Inputs are the same as for (a) except that there is an additional input *TWMLEDirection* used to define whether the TWMLE refers to "up", "down" or "average" TWMLE as above.

(d) MnProbDistTWMLERestricted. Inputs are as per (c) but again with an additional input array allowing estimation process to apply to some but not all of the distributional parameters.

The main reason for including (b) and (d) is to cater for situations where we wish to constrain one or more of the distributional parameters to a fixed value, e.g. we might want to force the mean to have a specific value.

Corresponding quantiles for the input quantile points can then be derived using the MnProbDistQuantile web service function, which takes as inputs the *DistributionName*, a quantile point, *q*, and a vector of parameter values, *ParamValues*, defining the which member of the distributional family is being used.

References

Kemp, M.H.D. (2009). *Market consistency: Model calibration in imperfect markets*. John Wiley & Sons

Kemp, M.H.D. (2010). *Extreme Events. Robust Portfolio Construction in the Presence of Fat Tails*. John Wiley & Sons