Tail Fitting of a Normal Distribution

One approach for fitting the tail of a distribution is to select an appropriate distributional family and then to select the parameters characterising the distribution in a manner that provides the best fit of the observed (cumulative) distribution function and/or quantile-quantile plot in the relevant tail.

Suppose that the observations are x_i for $i = 1, \ldots, n$. When ordered these are say $x_{(i)}$. Weights given to each observation in the curve fitting process are $w_{(i)}$. Typically we might expect the $w_{(i)}$ to be non-zero (and then typically constant) only for i sufficiently small, or for i sufficiently large, although this is not strictly necessary.

A common way of carrying out curve fitting is least squares, so a natural way of implementing this approach to fit a (univariate) *Normal* distribution to the data might be:

Any Normal distribution is characterised by a mean, μ, and standard deviation, σ. We might therefore derive, $y_{(i)}$, the expected value for the observation $x_{(i)}$, using the following formula:

$$y_{(i)} = \mu + \sigma q_{(i)} \text{ where } q_{(i)} = N^{-1}(i - 1/2)$$

[Note, the expected value of *j*'th quantile of a Normal distribution is not precisely q_j as defined above because the pdf is not flat, see e.g. Expected Worst Loss Analysis]

We would then identify estimates of the mean, $\hat{\mu}$, and standard deviation, $\hat{\sigma}$, that together minimise the following least squares computation:

$$Y = \sum_{i=1}^{n} w_{(i)} \left(y_{(i)} - x_{(i)}\right)^2 = \sum_{i=1}^{n} w_{(i)} \left(\mu + \sigma q_{(i)} - x_{(i)}\right)^2$$

This is minimised when $\frac{\partial Y}{\partial \mu} = 0$ and $\frac{\partial Y}{\partial \sigma} = 0$, i.e. for the values of $\hat{\mu}$ and $\hat{\sigma}$ where:

$$\sum_{i=1}^{n} w_{(i)} \left(\hat{\mu} + \hat{\sigma} q_{(i)} - x_{(i)}\right) = 0 \text{ and } \sum_{i=1}^{n} w_{(i)} q_{(i)} \left(\hat{\mu} + \hat{\sigma} q_{(i)} - x_{(i)}\right) = 0$$

If $W = \sum w_{(i)}$, $W_q = \sum w_{(i)} q_{(i)}$, $W_{qq} = \sum w_{(i)} q_{(i)} q_{(i)}$, $W_x = \sum w_{(i)} x_{(i)}$ and $W_{qx} = \sum w_{(i)} q_{(i)} x_{(i)}$ then these equations simplify to:

$$W\hat{\mu} + W_q \hat{\sigma} = W_x \text{ and } W_q \hat{\mu} + W_{qq} \hat{\sigma} = W_{qx}$$

$$\therefore \hat{\mu}_{TF} = \frac{W_{qq} W_x - W_q W_{qx}}{W W_{qq} - W_q^2} \text{ and } \hat{\sigma}_{TF} = \frac{-W_q W_x + W W_{qx}}{W W_{qq} - W_q^2}$$

Whilst this type of approach is primarily designed to be used merely in the tail of the distribution (i.e. with $w_{(i)}$ non-zero, perhaps constant, only for i suitably small or, for the other tail, only for i suitably close to n), we can also consider what answer this approach would give if it were applied to the *entire* distributional form, e.g. using $w_{(i)} = 1$ for all $i = 1, \ldots, n$. As the $q_{(i)}$ are symmetric around 0.5, we have $W_q = 0$ so $\hat{\mu} = W_x / W$, i.e. $\hat{\mu}$ is then the usual maximum likelihood estimator $\bar{x}$. By, say, carrying out a simulation exercise we can also confirm that $\hat{\sigma}$ is also typically close to the relevant maximum likelihood estimator if n is not very small.