

Modeling and Fitting Quantile Distributions and Regressions

Warren Gichrist

Auburn March 2006

Exercises

The first two exercises have been taken from Gilchrist (2000).

1. Page 152 question 4

Investigate the construction and quantile properties of the following distributions:

(a) $S(p) = (2p - 1)/[p(1 - p)] + 2 \ln[p/(1 - p)]$.

(b) $S(p) = p + (\beta/2k)(2p - 1)^k$.

(c) $S(p) = -p \ln(1 - p)$.

(d) $S(p) = p^2(3 - 2p)$.

(e) $S(p) = -\ln(1-p^\alpha)$, considering $\alpha = 0.5$ and $\alpha = 2$.

2. Page 153 question 8

A measure of the severity of an illness has a distribution $Q(p) = \lambda + \eta S(p)$ over a population of those diagnosed as having the illness. A new treatment is to be tried. What modified models might be appropriate to the treated population if:

(a) All respond in the same way by a constant reduction in severity?

(b) The treatment has an effect only on the most severely ill half of the population?

(c) The treatment works for all but in general has an effect which increases with the severity of the illness?

3. Some data used by Hald (1952) gives the stopping distance, y , of vehicles as a function of their speed, x . The Excel file 'Exercise3' gives the data in columns B and C respectively.

The model recommended by Hald is $Y_i = \beta x_i + \gamma x_i^2 + \eta x_i e_i$, with e_i being i.i.d. $N(0, 1)$. Note: as the model has a scale factor containing x the residuals $(y - y^*)$ are scaled by x before calculations.

Examine the formulae that lead to the calculation of the deterministic residual sum of absolutes and the residual sum of squares. Compare these to the formulae used to calculate their distributional equivalents. Comment on what the results obtained for the normal model (Sheet1) and the skew logistic model (Sheet2).

4. The data in the Excel file 'Exercise4' gives data from Karian and Dudewicz (2000).

a. The sheets 'KidMoM' and 'KidDLA' look at the Horse Kidney Data from page 96 of the text.

In KidMoM, method of moments estimates of the parameters for the four parameter lambda distribution (presented in the text) are used to undertake calculations for the sum of absolute deviations (SAD) and to produce a fit-observation plot and a plot of the distributional residuals.

In KidDLA, 'solver' is used to find values of the four parameters to 'minimize' the SAD.

i. Examine the formulae leading to the calculation of the SAD in KidDLA, make a note of the 'optimal' value and compare it with that in KidMoM.

ii. Use solver to minimize the SAD using the starting values: $\lambda_1 = 58$ and $\lambda_2 = \lambda_3 = \lambda_4 = 0.01$. Comment on the results obtained.

- b. The sheets 'ScanMoM' and 'ScanDLA' look at the MRI Scan Data from page 98 of the text.
The two sheets give the corresponding calculations to those in 'KidMoM' and 'KidDLA' for example 4a above.
- i. Examine the formulae leading to the calculation of the SAD in ScanDLA, make a note of the 'optimal' value and compare it with that in ScanMoM.
 - ii. Use solver to minimize the SAD using the starting values: $\lambda_1 = 107$ and $\lambda_2 = \lambda_3 = \lambda_4 = 0.01$. Comment on the results obtained.
5. The data in the Excel file 'Exercise5' gives data from various sources as follows:
Sheets 1-3 give an example from Hald (1952) showing his results (1), minimisation by Distributional least squares (2) and a fit using the logistic (3).
Sheet 4 gives an example from Guttman et al (1971).
Sheets 5-7 give another Hald example, where on sheet 7 the use of a skew logistic gives significant improvement over the normal and logistic.
Sheet 8 uses data from Hand (1993) (set 5) to explore a number of graphics.
Sheets 9-12 use data set 84 from this book and look at various graphics and a new distribution (the symmetric Weibull (sheet 12)) to fit. A u-shaped symmetric lambda gives a significantly better fit than the others tried (sheet 11).

References

- Gilchrist, W. G. (2000) *Statistical Modelling with Quantile Functions*. Boca Raton. FL. Chapman and Hall / CRC Press.
- Guttman, I., Wilks, S. and Stuart Hunter, J. (1971) *Introductory Engineering Statistics*. John Wiley & Sons Inc
- Hald, A. (1952) *Statistical Theory with Engineering Applications*. Wiley, New York.
- Hand, D. J. (1993) *Handbook of Small Data Sets*. Chapman & Hall/CRC.
- Karian, Z. A. and Dudewicz, E. J. (2000). *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*, CRC Press, Boca Raton, FL.