Semi-parametric fitting of cost distributions with tails based on extreme value theory

Caterina Conigliani and Andrea Tancredi
University of Roma3 and University of Roma "La Sapienza"
Italy

1 Introduction

In cost evaluations conceived to have an impact on medical policy, the main interest is the total healthcare cost, so that it is inference on population mean costs that is informative. However, cost data obtained for individual patients in health economic studies typically exhibit highly skew and heavy tailed distributions, and many problems arise with the various approaches currently available for analysing such data.

In fact, as discussed in O'Hagan and Stevens (2002a and 2002b), nonparametric methods, such as those based on the asymptotic normality of the sample mean or nonparametric bootstrapping, may be inefficient and their justification breaks down in small samples. On the other hand, parametric modelling may lead to more efficient inference, but is dependent on the population distribution matching the model adequately. The main difficulty in this sense is that the high skewness and kurtosis usually found in cost data imply that the population mean can be very sensitive to the tail of the distribution beyond the range of the data; one consequence of this is that parametric models that fit the data equally well can produce very different answers. Another problem related to the parametric modelling of cost data concerns possible transformations of the data; in fact, as discussed in Thompson and Barber (2000) and Briggs and Gray (1998), mean values and confidence limits may be difficult to interpret on the transformed scales, and back-transformation onto the original scale is not always straighforward.

Here we consider a Bayesian approach, and for the simple problem of estimating the population mean cost from data on a single sample of patients, we model the bulk of the data and the tails separately. More specifically, we consider a distribution composed of a piecewise constant density up to an unknown endpoint, and a generalised Pareto distribution (GPD) for the remaining tail. Note this model has been used to model extreme and non extreme environmental data by Tancredi *et al.* (2002).

The first component of the model, the step function, is very flexible, in the sense that it has the appealing property of catching all the relevant features of the data; if for instance the data exhibit multimodality, the corresponding model will be multimodal. However, the step function will hardly give any weight to values beyond the range of the data; for this reason, we introduce a different model for the right tail of the distribution, the GPD, that is often used in extreme value theory to model tail data (see, for instance Coles, 2001a). Recall that extreme value theory is mainly concerned with quantifying the stochastic behaviour of a process at unusually large (or small) levels, and provides a class of models to enable estimation of the probability of events that are more extreme than any that have already been observed.

2 The model

Let $h\,(x\,|\cdot)$ be a piecewise constant density on $(0,\alpha)$ with unknown number of steps s at positions $a_1 = 0 < a_2 < ... < a_s < a_{s+1} = \alpha$ taking value ω_i on the subinterval $[a_i, a_{i+1})$; if $a^{(s)} = (a_2, ..., a_s)$ denotes the vector of unknown step positions and $\omega^{(s)} = (\omega_1, ..., \omega_s)$ denotes the vector of unknown heights, we can write:

$$h\left(x\left|s,\omega^{(s)},a^{(s)},\alpha\right.\right)=\sum_{i=1}^{s}\omega_i I_{[a_i,a_{i+1})}(x) \tag{1}$$

with the constraint

$$\sum_{i=1}^{s}\omega_i\left(a_{i+1}-a_i\right)=1.$$

The step function (1) has been analysed in a Bayesian context by Robert (1998) and Robert and Casella (1999), and can be seen as a mixture of s uniform distributions $U_{[a_i,a_{i+1})}$:

$$h\left(x\left|s,p^{(s)},a^{(s)},\alpha\right.\right)=\sum_{i=1}^{s}p_i U_{[a_i,a_{i+1})}$$

where $p^{(s)}=(p_1,...,p_s)$, $p_i=\omega_i\left(a_{i+1}-a_i\right)$, and $\sum_{i=1}^{s}p_i=1$.

Moreover, let $g\left(x\left|\alpha,\sigma,\xi\right.\right)$ be the generalized Pareto density with threshold α, scale parameter σ and shape parameter ξ, all unknown:

$$g\left(x\left|\alpha,\sigma,\xi\right.\right)=\frac{1}{\sigma}\left[1+\frac{\xi\left(x-\alpha\right)}{\sigma}\right]^{-\frac{1}{\xi}-1}, \tag{2}$$

defined on

$$\left\{x:x>\alpha \text{ and } 1+\frac{\xi\left(x-\alpha\right)}{\sigma}>0\right\}.$$

Then we model cost data $x_1,...,x_n$ as *i.i.d.* observations from a distribution with density

$$f\left(x_i\left|s,p^{(s)},a^{(s)},\alpha,\sigma,\xi,\omega\right.\right)=\begin{cases}(1-\omega)\,h\left(x_i\left|s,p^{(s)},a^{(s)},\alpha\right.\right) & 0<x<\alpha\\ \omega\,g\left(x_i\left|\alpha,\sigma,\xi\right.\right) & \alpha\le x<\infty\end{cases} \tag{3}$$

where ω is the probability that an observation is greater than α.

Note that the qualitative behaviour of the GPD depends significantly on the value of the the shape parameter ξ. In particular, $g\left(x\left|\alpha,\sigma,\xi\right.\right)$ is an increasing function of x when $\xi<-1$, is constant when $\xi=-1$, while is a decreasing function of x for $\xi>-1$; moreover, it has an upper bound of $\alpha-\sigma/\xi$ if $\xi<0$, while it has no upper limit if $\xi\ge 0$; for this reason, when modelling cost data such those arising in the evaluation of health care technologies it seems appropriate to impose the constraint $\xi\ge 0$. Furthermore, the GPD has finite expected value

$$E\left(x\left|\alpha,\sigma,\xi\right.\right)=\alpha+\frac{\sigma}{1-\xi}$$

only if ξ is less than 1; as our main interest is the population mean μ, in what follows we will also impose the constraint $\xi<1$, so that

$$\mu=(1-\omega)\sum_{i=1}^{s}p_i\frac{a_{i+1}+a_i}{2}+\omega\left[\alpha+\frac{\sigma}{1-\xi}\right]. \tag{4}$$

Moreover, it is important to notice that the scale parameter of the GPD, σ, depends on the threshold α, and one should take this into account when introducing a prior distribution for the parameters of the GPD. However, as pointed out in Tancredi *et al.* (2002) and in Coles and Tawn (1996), this problem can easily be avoided by reparametrizing model (3) in terms of the parameters of the generalized extreme value distribution (GEV) corresponding to the GPD (Coles, 2001a). In particular we can write σ and ω as:

$$\sigma=\psi+\xi\left(\alpha-m\right)$$

$$\omega = \frac{1}{n}\left[1+\frac{\xi\left(\alpha-m\right)}{\psi}\right]^{-\frac{1}{\xi}}$$

where m and ψ are the location and scale parameters of the GEV, ξ is the shape parameter of the GEV, that coincides with the shape parameter of the GPD, α is the threshold and n is the sample size; this parametrization has the advantage of making the parameters m, ψ, ξ independent of α.

Finally, note that Bayesian inference for model (3) is possible using *Markov chain Monte Carlo* (MCMC) methods (see, for instance, Robert and Casella, 1999). If $\theta = \left(s, p^{(s)}, a^{(s)}, \alpha, m, \psi, \xi\right)$ denotes the vector of unknown parameters of model (3), a complete updating of θ is made through the following moves:

a) updating the weights $p^{(s)}$;
b) updating the step positions $a^{(s)}$;
c) updating the threshold α;
d) updating the GEV parameters m, ψ, ξ;
e) splitting one step in two or combining two steps into one.

In particular, moves a), b), c) and d) do not modify the dimension of θ, and their implementation is quite standard. For example, for the updating of $p^{(s)}$ we use a Gibbs kernel, since it is possible to simulate directly from the posterior distribution of $p^{(s)}$ conditional of all the other parameters. Instead, for moves b), c) and d) we apply the Metropolis-Hastings algorithm; as proposal distributions we consider a uniform distribution between a_{i-1} and a_{i+1} to update each $a_i \in a^{(s)}$, and a normal random walk for α. For the updating of the GEV parameters, Coles (2001b) notes that moving with proposal transition orthogonal to the m, ψ, ξ axes produces slow mixing, and suggests instead moves orthogonal to the ω, σ, ξ axes.

Instead, move e) modifies the parameter s, and consequently the dimension of the vector θ; it follows that it requires the reversible jump methodology (Green, 1995).

References

Briggs, A. and Gray, A. (1998) The distribution of health care costs and their statistical analysis for economic evaluation. *J. Health Serv. Res. Policy*, 3, 233-245.

Coles, S.G. (2001a) *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London.

Coles, S.G. (2001b) The use and misuse of extreme value models in practice. *Technical Report, University of Bristol, UK.*

Coles, S.G. and Tawn, J.A. (1996) A Bayesian analysis of extreme rainfall data. *Appl. Stat.*, 45, 463-478.

Green, P.J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82, 711-732.

O'Hagan, A. and Stevens, J.W. (2002a) Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? *Health Economics*, 12, 33-49.

O'Hagan, A. and Stevens, J.W. (2002b) Bayesian methods for design and analysis of cost-effectiveness trials in the evaluation of health care technologies. *Statistical Methods in Medical Research*, 11, 469-490.

Robert, C.P. (1998) *Discretization and MCMC convergence assessment.* Lecture Notes in Statistics 135, Springer-Verlag, New York.

Robert, C.P. and Casella, G. (1999) *Monte Carlo Statistical Methods* Springer-Verlag, New York.

Tancredi, A. Anderson, C. and O'Hagan, A. (2002) A Bayesian model for threshold uncertainty in extreme value theory, submitted.

Thompson, S.G. and Barber, J.A. (2000) How should cost data in pragmatic randomised trials be analysed? *Br. Med. J.*, 329, 1197-1200.