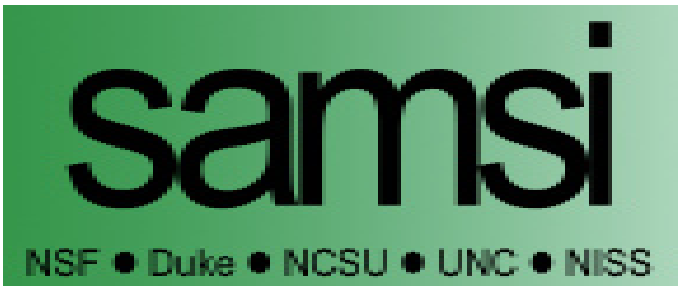


Semiparametric Functional Estimation Using Quantile-Based Prior Elicitation

Elijah Gaioni, Dipak K. Dey
and Mircea Grigoriu

Technical Report #2008-6
June 23, 2008

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

Semiparametric Functional Estimation Using Quantile-Based Prior Elicitation

Elijah Gaioni, Dipak K. Dey and Mircea Grigoriu*

Abstract

In this paper we present an interactive approach to Bayesian functional estimation focusing on applications to extreme value modeling. Since there is often minimal data available pertaining to extreme events, which also tend to be rare events, we present an approach that requires only a very limited amount of quantile information. An adaptive algorithm is used to determine an optimal prior distribution corresponding to the quantile information provided, where the multivariate Bernstein polynomial representation theorem is used in the nonparametric specification of the prior.

Our emphasis is on the interactive nature of the approach between the subject matter expert and the statistician, and the analysis is aided by graphical displays elucidating the relationship between the expert provided information and the resulting functional estimate. By graphically associating the expert-provided information with the estimate, an expert can revisit and revise the information provided in an iterative fashion, thus ensuring that it coincides with the expert's true beliefs.

We detail this estimation procedure for the prior predictive distribution and outline the extensions to posterior predictive distribution and hazard function estimation. A complete example is provided for a previously studied data set consisting of a volatility measure of the NASDAQ 100 over the period 1985-2002.

*Elijah Gaioni is a PhD student (email: elijah.gaioni@uconn.edu) at the University of Connecticut, Dipak K. Dey (email: dipak.dey@uconn.edu.) is Professor and head of the Department of Statistics at the University of Connecticut in Storrs, CT. 06269 - 4120 and Mircea Grigoriu (email: mdg12@cornell.edu.) is Professor of Engineering at Cornell University in Ithaca, NY. 14853. This work was partially completed while the authors were visiting SAMSI, Research Triangle Park, NC-27709.

KEY WORDS: Adaptive algorithm, CDF determination, Expert opinion, Functional estimation, Prior elicitation.

1 INTRODUCTION.

Functional estimation is concerned with the estimation of random curves, e.g., density, hazard and survival curve estimation. The monograph of Ramsay and Silverman (1997) provides a broad range of examples of functional estimation. Consider the random variable X, with associated probability measure P^X . In this paper we are concerned with performing inference under circumstances where a very limited amount of information about P^X is available. Specifically, our methodology will address those circumstances where access to only a small number of quantiles from the distribution of X are provided. We briefly motivate the need for such an estimation procedure with a couple of representative scenarios below.

When studying extreme events, the nature of the event often renders it rare as well as extreme. Such events include flash flooding, disease outbreaks, periods of unusually high volatility in financial markets, etc. Due to the rarity of these events, an insufficient number of observations may be available for purposes of employing traditional inference methods (Behrens et al. 2004). Thus, every additional piece of information that can be incorporated into the analysis is of utmost importance. With respect to these scenarios, such information may be available from an expert in the field of hydrology, epidemiology or finance, respectively. In the absence of such expert-provided information it might prove difficult to perform accurate inference on the quantities of interest.

We demonstrate how one may incorporate this quantile-based expert information into the process of performing functional estimation. Upon obtaining an estimate of the functional of interest, graphical output is utilized to facilitate the expert's understanding of the relationship between the prior information provided and the estimated functional of interest (see Figure 4:(a) in section 4). In this way, the expert and statistician may engage in an interactive exploration, or discovery, of the underlying structure of the phenomenon being modeled.

Arguments supporting our choice of quantile-based elicitation over moment-based elicitation are

provided in Dey and Liu (2007) and O'Hagan (1998). Although we focus on real-valued variables in this presentation, vector-valued variables may also be handled with a similar approach.

This article is organized as follows. In Section 2 we present our methodology. Specifically, in Section 2.1 we present the nature of the information provided by the expert. We then describe the problem formulation in Section 2.2, with a description for the estimation of particular functionals in Section 2.3. Section 3 describes the implementation of our proposed methodology along with a discussion of the key computational issues that arise. We demonstrate how our methodology can be implemented for a data set from the NASDAQ 100 in Section 4. Finally, we provide a discussion in Section 5.

2 METHODOLOGY.

2.1 Sharp versus Vague information.

We consider two modes of classification with respect to the nature of the information provided by the expert. First, the type of information provided may be categorized as being qualitative or quantitative. A further bifurcation of the type of information provided may then be performed on the basis of whether the information is sharp or vague in nature.

In our context, sharp qualitative information consists of specification of the functional form of the distribution up to an unknown parameter vector. For instance, X follows a generalized extreme value distribution (Resnick 2006) with unknown parameter vector (μ, σ, ξ).

Vague qualitative information would be knowledge of features of the distribution that are insufficient for purposes of determining exactly which family of distributions gave rise to the observations. For instance, knowledge of the symmetry or skewness properties of the distribution would represent vague qualitative information.

Sharp quantitative information, on the other hand, consists of the specification of some finite number of quantiles from the marginal distribution: (p_i, q_i), for $i = 1 \ldots m (< \infty)$. There is always a (non-unique) distribution that has these quantiles.

However, this sharp quantitative scenario can be generalized and made more realistic, from an elicitation perspective, by acknowledging that the expert is unlikely to know precisely what the value of any given quantile might be. It is more likely that the expert will be able to provide a set of $K(i)$ possible quantiles, $\{q_{i,1}, \ldots q_{i,K(i)}\}$, for each percentile, p_i, that is provided. For instance, with respect to a Poisson random variable an expert might specify $q_{1,1} = 5$, $q_{1,2} = 6$ or $q_{1,3} = 7$ as possible values for $p_1 = 0.5$, the median. Where more than one percentile is concerned, the number of quantiles provided for each percentile need not be the same, so in general $K(i) \neq K(j)$ for $i \neq j$. Further, the expert may specify the degree of confidence that should be invested in each of these possible quantiles. In this Poisson example each of 5, 6 and 7 might be viewed as equally plausible values for the unknown median, and so one might associate a probability of $1/3^{rd}$ with each of them. This setup, with multiple quantiles provided for each percentile, will be referred to as constituting vague quantitative information.

It should also be noted that in the vague quantitative framework instead of a discrete set of possible quantile values, $\{q_{i,1}, \ldots q_{i,K(i)}\}$, corresponding to a given percentile, p_i, we may take an interval of possible quantile values. The discrete set of probabilities would then be replaced by a continuous distribution over that interval.

A second motivation for the use of vague quantitative information might come from multiple experts. Here, each of the, say K, of them would provide a different set of quantiles. Yet a third way to motivate this framework would be to consider a meta-analysis. The empirical quantiles from different analyses, with associated probabilities based on the sample size of each analysis, would give rise to multiple quantiles as well.

Whatever the motivation, in this paper we proceed in the context of sharp qualitative information in conjunction with vague quantitative information; a circumstance that we believe is encountered relatively frequently. That is, we assume a known functional form with unknown parameter values for the likelihood, and a set of possible quantiles for each of a finite number of percentiles from P^X.

It should be noted that this approach differs from the empirical Bayesian approach (Berger 1983, Robbins 1985) in that we only assume access to a very small set of quantiles from the marginal distribution, as opposed to an entire sample of data points. Of course, were an entire sample available, it could be

incorporated into this mode of analysis by using those sample values to generate empirical estimates of the quantiles.

2.2 Problem formulation.

The sharp qualitative information presents us with a cdf, $F(x|\boldsymbol{\theta})$, and the vague quantitative information yields a set of percentiles with possible quantiles,$(p_i, q_{i,1}), \ldots (p_i, q_{i,K(i)})$, $i = 1 \ldots m$. Further, with each possible quantile, $q_{i,k}$, we have an associated probability, $\zeta_{i,k}$, which corresponds to the degree of confidence with which that expert-provided quantile is believed to be the true quantile.

Let $\pi(\boldsymbol{\theta})$ be a prior density consistent with the expert's belief. Under sharp quantitative information we would have

$$p_i = \int F(q_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{2.1}$$

Under vague quantitative information equation (2.1) can be generalized to

$$\sum_k p_i\zeta_{ik} = p_i = \sum_k \zeta_{ik} \int F(q_{ik}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{2.2}$$

by weighted averaging over the quantiles.

Our objectives are to derive an approximation, $\tilde{\pi}(\boldsymbol{\theta})$, to $\pi(\boldsymbol{\theta})$, and then use this approximation to construct the expert prior predictive distribution

$$F_{prp}(x) = \int F(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{2.3}$$

To achieve these objectives, we consider the representation

$$\pi(\boldsymbol{\theta}) = \sum_{r=1}^{\infty} \beta_r \pi_r(\boldsymbol{\theta}), \qquad \sum_{r=1}^{\infty} \beta_r = 1, \quad \beta_r > 0, \ r = 1, \ldots, R, \tag{2.4}$$

where each $\pi_r(\boldsymbol{\theta})$ is a Dirichlet distribution with parameter vector $\boldsymbol{\alpha}_r$, and where the form of the representation is motivated by the multivariate Bernstein polynomial representation theorem (Lorentz 1986). The particular choices of the π_r used in the representation will be discussed in the **IMPLEMENTATION** section. This is simplified to

$$\tilde{\pi}(\boldsymbol{\theta}) = \sum_{r=1}^{R} \beta_r \pi_r(\boldsymbol{\theta}), \qquad \sum_{r=1}^{R} \beta_r = 1, \quad \beta_r > 0, \ r = 1, \ldots, R, \tag{2.5}$$

for calculational purposes. A potentially nonunique set of optimal values of $(\beta_1, \ldots, \beta_R)$ can be obtained by solving the constrained optimization

$$\min_{\beta_1,\ldots,\beta_R} \left(\sum_{i=1}^{m} \varepsilon_i^2 \right), \qquad \sum_{r=1}^{R} \beta_r = 1, \quad \beta_r > 0, \ r = 1, \ldots, R, \tag{2.6}$$

where

$$\varepsilon_i = p_i - \sum_{r=1}^{R} \beta_r \sum_{k=1}^{K(i)} \zeta_{ik} \int F(q_{ik}|\boldsymbol{\theta}) \pi_r(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{2.7}$$

The nonuniqueness of the family of solutions of (2.6), (2.7) is addressed in Section 3.5. Any solution of (2.6), (2.7) will be denoted $\tilde{\pi}^{opt}$.

We note that the combining of expert opinion, as with the weighting in equation (2.2), may may be performed in a variety of ways and has been discussed at length. We follow the Opinion Pool line of Stone (1961).

The above procedure is now applied to estimation of the posterior predictive distribution and the prior and posterior forms of the hazard function in Section 2.3 before turning to the methodological and computational issues arising in the implementation of this approach in Section 3.

2.3 Posterior predictive distribution and hazard function estimation.

We first consider the posterior predictive distribution, F_{pop}, for new observations $\mathbf{z}$ given data (or elicited quantiles) $\mathbf{x}$, defined as

$$F_{pop}(\mathbf{z}|\mathbf{x}) = \int F(\mathbf{z}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta},$$

where

$$\pi(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta}}.$$

We then form the estimate

$$\hat{F}_{pop}(\mathbf{z}|\mathbf{x}) = \int F(\mathbf{z}|\boldsymbol{\theta}) \tilde{\pi}^{opt}(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta},$$

where

$$\tilde{\pi}^{opt}(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\theta})\tilde{\pi}^{opt}(\boldsymbol{\theta})}{\int f(\mathbf{x}|\boldsymbol{\theta})\tilde{\pi}^{opt}(\boldsymbol{\theta}) d\boldsymbol{\theta}}.$$

Considering $\mathbf{x}$ to be the set of expert provided quantiles, say $\mathbf{x}$=$\mathbf{q}$, we have a method of performing predictive inference based on very limited quantile information.

We now consider estimation of the hazard function, h, defined as

$$h(x) = \frac{f(x)}{1 - F(x)},$$

where we require F to be a differentiable CDF, and where f is the associated PDF.

We first note that there are two clear approaches to estimation of the hazard function that follow in a straightforward way from the formulation given in Section 2.2. First, if available, quantiles may be elicited directly on the hazard function. In this case one need only substitute h for F in equation (2.3), and proceed along the same lines as before.

Alternatively, if quantiles from the hazard function are either unavailable or unreliable, then given the same quantile information about the CDF as assumed in Section 2.2, we may form an estimate of h_{prp}, representing the prior predictive form of the hazard function, on the basis of the estimate for F_{prp}. Proceeding from equation (2.3) we define

$$\hat{h}_{prp}(x) = \frac{\hat{f}_{prp}(x)}{1 - \hat{F}_{prp}(x)},$$

where $\hat{F}_{prp}(x)$ is obtained by substituting a solution $\tilde{\pi}^{opt}$ into equation (2.3) and where

$$\hat{f}_{prp}(x) = \frac{d}{dx}\hat{F}_{prp}(x) = \int f(x|\boldsymbol{\theta})\tilde{\pi}^{opt}(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

A posterior predictive estimate of the hazard function may also be obtained as

$$\hat{h}_{pop}(\mathbf{z}|\mathbf{x}) = \frac{\hat{f}_{pop}(\mathbf{z}|\mathbf{x})}{1 - \hat{F}_{pop}(\mathbf{z}|\mathbf{x})},$$

where $\hat{F}_{pop}(\mathbf{z}|\mathbf{x})$ is given in equation (2.3) and

$$\hat{f}_{pop}(\mathbf{z}|\mathbf{x}) = \frac{d}{dx}\hat{F}_{pop}(\mathbf{z}|\mathbf{x}) = \int f(\mathbf{z}|\boldsymbol{\theta})\tilde{\pi}^{opt}(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}.$$

3 IMPLEMENTATION.

3.1 Outline.

In this section we provide the step-by-step procedure by which we form the functional estimate F_{prp}, and discuss in detail those elements warranting an explanation. As noted in Section 2.2, the estimate of F_{prp}

is obtained by determining an optimal prior, $\tilde{\pi}^{opt}(\boldsymbol{\theta})$, corresponding to the expert-provided information. It is thus necessary to first specify a region, say Θ, within the parameter space that forms the support for $\tilde{\pi}^{opt}(\boldsymbol{\theta})$. This is covered in Section 3.2.

As seen in the expression for $\tilde{\pi}(\boldsymbol{\theta})$ given in (2.5), it is necessary to choose which π_r will comprise the terms in a particular mixture. Since we do not know in advance what the appropriate number of terms in representation (2.5) should be, we iteratively solve for $\tilde{\pi}(\boldsymbol{\theta})$ using an increasing number of terms in successive iterations. Let the number of terms in representation (2.5) for the t^{th} iteration be denoted R_t, where $R_{t-1} < R_t$, and with corresponding prior solution $\tilde{\pi}_t(\boldsymbol{\theta})$ at that iteration. We require the discrepancy between solutions in successive iterations, $\tilde{\pi}_{t-1}(\boldsymbol{\theta})$ and $\tilde{\pi}_t(\boldsymbol{\theta})$, to be small before we stop seeking a solution to $\tilde{\pi}^{opt}(\boldsymbol{\theta})$ with more terms in the mixture (2.5). The Kullback-Leibler directed divergence criterion is used for this purpose:

$$KL(\tilde{\pi}_{t-1}, \tilde{\pi}_t) = \int_0^\infty \tilde{\pi}_{t-1}(\boldsymbol{\theta}) \ln \frac{\tilde{\pi}_{t-1}(\boldsymbol{\theta})}{\tilde{\pi}_t(\boldsymbol{\theta})} d\boldsymbol{\theta} < \varepsilon_{KL}.$$

To facilitate a faster rate of convergence and to ease the computational requirements, we introduce an adaptive method of choosing new π_r at each iteration. This approach is described in Section 3.3. The optimization in equations (2.6) and (2.7) entails performing the numerical computations described and discussed in Section 3.4.

The solution is then reviewed by the subject matter expert for plausibility. Any discrepancy between the solution obtained and the anticipated solution may be addressed in two straightforward ways. Either the expert may revise the vague quantiles in light of their impact on the solution, or the statistician may refocus on a different region of the parameter space believed to hold solutions whose behavior coincides more nearly with the expert's beliefs. Here we implicitly assume that the expert possesses more knowledge about P^X than can captured by the quantiles alone. We summarize the procedure involved in solving a particular problem in the following steps:

Step1: Expert Information

- *Specify sharp qualitative information* $F(x|\boldsymbol{\theta})$
- *Specify vague quantitative information* $\{p_i, q_{ik}\}$ *for* $i = 1 \ldots m$ *and* $k = 1 \ldots K(i)$

Step2: Adaptive Algorithm Input

- *Choose a region, Θ, for the support of the prior, $\tilde{\pi}(\boldsymbol{\theta})$*
- *Choose π_r for $r = 1 \ldots R_1$ in the representation of $\tilde{\pi}_1(\boldsymbol{\theta})$ in the initial iteration*
- *Choose the convergence criterion, ε_{KL}, for successive prior solutions, $\tilde{\pi}_t(\boldsymbol{\theta})$*

Step3: Examine Output

- *Run the adaptive algorithm until convergence of successive solutions is achieved*
- *Investigate the relationship between the expert-provided information in Step 1 and both $\tilde{\pi}^{opt}(\boldsymbol{\theta})$, and F_{prp}. If the expert deems the solution implausible, explore solutions in other regions in the parameter space or revisit the information in Step 1 to accommodate the expert's belief.*

The implementation of these steps is illustrated in an application to financial data in Section 3. The following subsections now elaborate upon some of the steps in the outline provided above.

3.2 Parameter region determination.

Since we require $\tilde{\pi}^{opt}(\boldsymbol{\theta})$ to determine F_{prp}, we first must settle on a region of support for the prior. The values of the expert-provided quantiles used in conjunction with expert knowledge of the process should suggest which region to focus on in the parameter space. For example, in the GEV case, where $F(x|\mu, \sigma, \xi) = \exp\{-[1 + \xi(x - \mu)/(\sigma)]^{-1/\xi}\}$, if one is dealing with a heavy upper-tailed, nonnegative quantity, we know to focus only on a region where the support, $\mu - \sigma/\xi$, for such a variable is > 0. Non-negativity and heavy upper tails indicate $\xi > 0$, corresponding to the special case of the GEV distribution known as the Frchet distribution. Furthermore, if certain physical limitations prohibit excessively large values, then σ and ξ may be restricted to reasonable intervals that do not lead to a disproportionate amount of large values and unacceptably heavy tail behavior (Resnick 2006).

However, it is not necessary to know in advance precisely which parameter region is best suited to the process. The nature of our approach allows one to go back and revise the parameter region if the estimate obtained using the current region is unsatisfactory. So as a first step one may take a conservatively large parameter region with the knowledge that it may be refined later.

Based on such considerations assume one settles on an initial finite, rectangular region, Θ, consisting of plausible values for the parameters. In the general case with p parameters, this region will consist of the cartesian product of intervals

$$\Theta = (\theta_1^L, \theta_1^U) \times \ldots \times (\theta_p^L, \theta_p^U).$$

All points in Θ must then be standardized suitably to become arguments for the Dirichlet distributions in the nonparametric representation given in equation (2.5). Corresponding to points $(\theta_1, \ldots, \theta_p) \in \Theta$, the arguments to the Dirichlet distributions consist of the standardized values θ_i', where

$$\theta_i' = \frac{\theta_i - \theta_i^L}{p(\theta_i^U - \theta_i^L)}, \qquad i = 1 \ldots p.$$

In this way all points in Θ (as well as some points outside Θ) are standardized to belong to the standard $p-$dimensional simplex, $\{(\theta_1, \ldots, \theta_{p+1}) : \sum_{i=1}^{p+1} \theta_i = 1, \;\; 0 \leq \theta_i \;\; \forall i\}$. This particular choice of standardization is motivated by the fact that it maximizes the volume of the region in equation (3.2) as a proportion of the total volume of the region in Θ that lies in the support simplex for $\tilde{\pi}^{opt}(\theta)$, as discussed in the appendix. Thus, the standardization we use in the example given in section 4 for the GEV distribution is

$$\mu' = \frac{\mu - \mu^L}{3(\mu^U - \mu^L)}, \;\; \sigma' = \frac{\sigma - \sigma^L}{3(\sigma^U - \sigma^L)}, \;\; \xi' = \frac{\xi - \xi^L}{3(\xi^U - \xi^L)}.$$

3.3 Adaptive basis element generation.

Having chosen the support, Θ, we now specify an initial set of Dirichlet distributions, $S_1 = \{\pi_r, \; r = 1 \ldots R_1\}$, used to estimate $\tilde{\pi}_1(\boldsymbol{\theta})$ in the 1^{st} iteration. Consider a Dirichlet distribution π for $(\theta_1, \ldots, \theta_p)$, with standard parametrization

$$f(\boldsymbol{\theta}) = \frac{\Gamma(\sum_{i=1}^{p+1} \alpha_i)}{\prod_{i=1}^{p+1} \Gamma(\alpha_i)} \prod_{i=1}^{p+1} \theta_i^{\alpha_i - 1},$$

where $\theta_{p+1} = 1 - \sum_{i=1}^{p} \theta_i$. Instead of directly specifying the parameter vector $\alpha = (\alpha_1, \ldots, \alpha_{p+1})$, we first specify the means, α_i/α_0, where $\alpha_0 = \sum_{i=1}^{p+1} \alpha_i$, of each argument θ_i, providing a measure of the center of the distribution. Then we choose a constant, $c(1)$, which multiplies the parameter vector: $c(1)\alpha$, controlling the dispersion of the distributions in the 1^{st} iteration. In this way we may first spread out the $\pi_r, \;\; r = 1 \ldots R_1$ according to their means, and then ensure adequate coverage by choosing a multiplier that yields a sufficiently large amount of dispersion that no region in Θ is neglected.

The choices for new π_r at subsequent iterations plays the crucial role in limiting the computational burden of the approach. As shown in Section 3.4, the number of terms, say R_T, in equation (2.5) at the final iteration plays a pivotal role in this regard. The following adaptive approach to allocating new π_r at each iteration keeps R_T small.

After the solution, $\tilde{\pi}_t$, at the t^{th} iteration has been obtained, we evaluate the density over a fine lattice of points in Θ. The goal is to determine those subregions within Θ where the prior density appears to be highest, and to concentrate on those regions in the following iterations. Therefore, at the subsequent iteration we add new π_r, say $S_{t+1}^{new} = \{\pi_r,\ r = R_t + 1, \ldots, R_{t+1}\}$, to the previous set of π_r, $S_t = \{\pi_r,\ r = 1, \ldots, R_t\}$. These distributions are centered (via their means) at the lattice points corresponding to regions of high prior density at the current iteration. Utilizing information in $\tilde{\pi}_t$ when determining S_{t+1}^{new} allows us to avoid using a naive approach where one might evenly spread out a considerably larger number of new elements over the entire region at each iteration- an approach that grows increasingly inefficient with increasing parameter space dimension.

Here we note that since we haven't yet established convergence of the prior sequence $(\tilde{\pi}_t)_t$, we don't want to invest full confidence in $\tilde{\pi}_t$ and reallocate all of the π_r at the $t + 1^{st}$ iteration based solely on the regions in Θ where $\tilde{\pi}_t$ has high density. Thus, we retain coverage over all regions of Θ that looked promising at prior iterations, as captured by the distributions in S_t, and add new elements, S_{t+1}^{new}, that focus on those subregions that look promising at the current iteration. This leads to a nested sequence $S_t \subset S_{t+1} = S_t \cup S_{t+1}^{new}$, which also proves to be computationally advantageous, as discussed in Section 3.4.

Additionally, we choose an increasing sequence of multipliers, $\{c(t)\}_{1 \leq t \leq T}$, resulting in a decreasing dispersion of the π_r in Θ. This is observed by noting that if the mean of a Dirichlet distribution has been specified as α_i/α_0 for $i = 1 \ldots p$, the application of multiplier $c(t)$ to parameter vector $\alpha = (\alpha_1, \ldots, \alpha_{p+1})$ leaves the mean unchanged but reduces the variance to

$$Var\{\theta_i\} = \frac{\alpha_i(\alpha_0 - \alpha_i)}{(\alpha_0)^2(c(t)\alpha_0 + 1)},$$

where the only difference is the presence of $c(t)$ in the denominator. The introduction of π_r with smaller dispersion into the mixture distribution in equation (2.5) allows us to focus more precisely on small

regions in Θ than if only highly dispersed π_r were incorporated into the mixture.

3.4 Computational considerations.

In order to perform the optimization in equations (2.6) and (2.7) it is necessary to calculate

$$a_{irk} = \zeta_{ik} \int F(q_{ik}|\boldsymbol{\theta})\pi_r(\boldsymbol{\theta})d\boldsymbol{\theta}, \qquad i = 1 \ldots m, \quad r = 1 \ldots R_t, \quad k = 1 \ldots K(i),$$

from the RHS of equation (2.7). This requires an integration over the $p-$dimensional region Θ. The benefit of choosing the nested sequence, $(S_t)_{1 \le t \le T}$, of π_r in the manner specified in Section 3.3 is that we needn't recalculate all of the a_{irk} at each iteration. Instead, we only have to calculate those a_{irk} with indices r in $R_{t-1} + 1, \ldots, R_t$, corresponding to those π_r in S_t^{new}, added at the most recent iteration.

Assuming convergence of the $\tilde{\pi}_t$ occurs at the $t = T^{th}$ iteration, we must perform $R_T \sum_{i=1}^m K(i)$ numerical integrations in total. If instead we had used a progressive algorithm that chose new π_r at each iteration, and assuming that n new π_r are added at each iteration, a total of $[TR_1 + (T-1)Tn/2] \sum_{i=1}^m K(i)$ numerical integrations would have been required. The savings resulting from the nested choice of $(S_t)_{1 \le t \le T}$ is therefore $(T-1)[R_1 + (T-2)n/2] \sum_{i=1}^m K(i)$, clarifying the crucial role of the adaptive algorithm in limiting the computer time necessary to solve for $\tilde{\pi}^{opt}(\boldsymbol{\theta})$. We also note that the computation of any particular a_{irk} is independent of the other a_{irk}, so they may also be computed in parallel.

3.5 Nonuniqueness of $\tilde{\pi}^{opt}$.

While any solution to equations (2.6), (2.7) must be checked by the expert for reasonableness, it may be argued that a certain subclass of the entire family of solutions is preferable. Below we propose one criterion for reducing the family of solutions to a preferable subfamily.

The goal is to minimize

$$f(\boldsymbol{\beta}) = (\mathbf{p} - \mathbf{A}\boldsymbol{\beta})'(\mathbf{p} - \mathbf{A}\boldsymbol{\beta}),$$

subject to

$$\sum_{r=1}^{R} \beta_r = 1 \quad and \quad \beta_r \ge 0, r = 1 \ldots R,$$

where $\mathbf{p} = (p_1, \ldots, p_m)'$ contains the percentiles and $\mathbf{A}_{ir} = \mathbf{a}_{ir}$ contains the numerically calculated integrals:

$$a_{ir} = \int \sum_k \zeta_{ik} F(q_{ik}|\boldsymbol{\theta}) \pi_r(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Let $\boldsymbol{\beta}^* = (\beta_1, \ldots, \beta_{R-1})'$. Taking $\beta_R = 1 - \sum_{r=1}^{R-1} \beta_r = 1 - \mathbf{1}'\boldsymbol{\beta}^*$, write $\mathbf{A} = [\mathbf{A}^*, \mathbf{a}]$ where $\mathbf{a}$ is the R^{th} column of $\mathbf{A}$. We then wish to minimize

$$f^*(\boldsymbol{\beta}^*) = \left(\mathbf{p} - [\mathbf{A}^*, \mathbf{a}] \begin{pmatrix} \boldsymbol{\beta}^* \\ 1 - \mathbf{1}'\boldsymbol{\beta}^* \end{pmatrix} \right)' \left(\mathbf{p} - [\mathbf{A}^*, \mathbf{a}] \begin{pmatrix} \boldsymbol{\beta}^* \\ 1 - \mathbf{1}'\boldsymbol{\beta}^* \end{pmatrix} \right), \tag{3.8}$$

subject to

$$\beta_r^* \geq 0, \ r = 1 \ldots R - 1. \tag{3.9}$$

This yields the unrestricted family of solutions

$$\Im = \{\boldsymbol{\beta} : \boldsymbol{\beta} = ((\mathbf{A}^* - \mathbf{a}\mathbf{1}')'(\mathbf{A}^* - \mathbf{a}\mathbf{1}'))^-(\mathbf{A}^* - \mathbf{a}\mathbf{1})'(\mathbf{p} - \mathbf{a})\}.$$

Denote those members of $\Im$ for which at least one $\beta_r < 0$ by $\Im^-$. The restricted family of solutions satisfying equations (3.8) and (3.9) is then $\{\boldsymbol{\beta} : \boldsymbol{\beta} \in \Im \backslash \Im^-\}$.

To facilitate the following calculations, we use the representation

$$\boldsymbol{\theta} = \Theta \mathbf{I}$$

where

$$\Theta = [\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^R]; \quad \boldsymbol{\theta}^r = (\theta_1^r, \ldots, \theta_p^r)' \sim Dirichlet(\boldsymbol{\alpha}_r), \ \ for r = 1 \ldots R,$$

and $\mathbf{I}$ has the discrete distribution

$$\mathbf{I} = \mathbf{e}_r \ \ \textit{with probability} \ \ \beta_r, \ \ r = 1 \ldots R,$$

for elementary unit vectors $\mathbf{e}_r$. So $(\boldsymbol{\theta}|\mathbf{I} = \mathbf{e}_r) = \boldsymbol{\theta}^r \sim Dirichlet(\boldsymbol{\alpha}_r)$, and unconditionally $P(\boldsymbol{\theta} \leq \mathbf{t}) = \sum_{r=1}^R \pi_r^{CDF}(\boldsymbol{\theta})\beta_r$, where π_r^{CDF} denotes the CDF of the r^{th} Dirichlet distribution in the mixture.

We then reduce the family of solutions in (3.8) and (3.9) to only those which minimize the trace of the resultant covariance matrix of $\tilde{\pi}^{opt}$:

$$\hat{\boldsymbol{\beta}} = argin f_{\boldsymbol{\beta}: \boldsymbol{\beta} \in \Im \backslash \Im^-} tr(\Sigma_{\boldsymbol{\theta}}(\boldsymbol{\beta})),$$

where $\Sigma_{\boldsymbol{\theta}}(\boldsymbol{\beta})$ is the covariance matrix of the parameter vector $\boldsymbol{\theta}$ viewed as a function of $\boldsymbol{\beta}$. Since $\Sigma_{\boldsymbol{\theta}}(\boldsymbol{\beta}) = E\{Cov(\Theta\mathbf{I}|\mathbf{I})\} = \sum_{r=1}^{R} \beta_r \Sigma_{\boldsymbol{\theta}^r}$ where $\{\Sigma_{\boldsymbol{\theta}^r}\}_{ii} = \alpha_i^r(\alpha_0^r - \alpha_i^r)/(\alpha_0^r)^2(\alpha_0^r + 1)$, we have

$$tr(\Sigma_{\boldsymbol{\theta}}(\boldsymbol{\beta})) = tr(\sum_{r=1}^{R} \beta_r \Sigma_{\boldsymbol{\theta}^r}) = \sum_{r=1}^{R} \beta_r tr(\Sigma_{\boldsymbol{\theta}^r}).$$

Denoting $\mathbf{v} = (tr(\Sigma_{\boldsymbol{\theta}^1}), \ldots, tr(\Sigma_{\boldsymbol{\theta}^R}))$ the solutions are given by

$$\hat{\boldsymbol{\beta}} = argin f_{\boldsymbol{\beta} \in \Im \setminus \Im^-} \boldsymbol{\beta}' \mathbf{v}$$

Alternatively, this can be expressed in terms of the generalized inverse G^- as

$$\hat{\boldsymbol{\beta}} = argin f_{\boldsymbol{\beta} \geq 0} (\mathbf{v}'\boldsymbol{\beta}) \;\; s.t. \;\; \boldsymbol{\beta} = G^-(\mathbf{A}^* - \mathbf{a1})'(\mathbf{p} - \mathbf{a}) \;\; for \; some \; G^-,$$

where $G^- = ((\mathbf{A}^* - \mathbf{a1}')'(\mathbf{A}^* - \mathbf{a1}'))^-$.

The optimization can then be performed over all $(R-1) \times (R-1)$ matrices $\mathbf{Z}$ given a particular solution, say G_0, using a categorization of all generalized inverses (Harville, 1997)

$$\hat{\boldsymbol{\beta}} = argin f_{\{\mathbf{Z}: \boldsymbol{\beta} \geq 0\}} (\mathbf{v}'[G_0 + \mathbf{Z} - \mathbf{G_0 A Z A G_0}]((\mathbf{A}^* - \mathbf{a1}')')(\mathbf{p} - \mathbf{a})).$$

We now illustrate this methodology with an example.

4 APPLICATION ON NASDAQ 100 DATA.

The example we use to illustrate our approach has been investigated in Behren's et. al. (2004). In total there are 4,394 observations over the period 1985 to 2002 consisting of financial data with response variable X recording the daily percentage deviation of the Nasdaq 100 index from the previous day's price

$$X_t = 100|P_t/P_{t-1} - 1|.$$

Our goal is to investigate the upper tail of this measure of volatility, which comprises a small yet important portion of the observations. We proceed according to the steps outlined in Section 3.1.

Step1: Expert Information

- *Specify sharp qualitative information* $F(x|\boldsymbol{\theta})$

Table 1: Vague quantitative information

p_i	q_i	$\rightarrow$	q_{i1}	q_{i2}	ζ_{i1}	ζ_{i2}
0.90	2.11	$\rightarrow$	2.02	2.20	0.5	0.5
0.95	2.92	$\rightarrow$	2.77	3.07	0.5	0.5
0.99	5.35	$\rightarrow$	5.05	5.65	0.5	0.5
0.999	9.00					

Since we are investigating tail behavior, we use the heavy-tailed GEV distribution in an attempt to find a good empirical fit to the larger of the observations.

• *Specify vague quantitative information* $\{p_i, q_{ik}\}$ *for* $i = 1 \dots m$ *and* $k = 1 \dots K(i)$

In lieu of actual expert opinion, we use quantiles provided in the Behren's (2004) paper as the starting point for our analysis. The data we use for our modeling is found in the first two columns in Table 1, providing a limited amount of information about the upper tail of the process.

The actual data consists of only one quantile per percentile, which constitutes sharp quantitative information. Consequently, in order to illustrate the vague quantitative approach we use those values to derive vague quantitative information by splitting each of the first three quantiles into two values, each centered at the actual quantile. For instance, $q_1 = 2.11 \rightarrow \{2.02, 2.20\} = \{q_{1,1}, q_{1,2}\}$, etc. Each of those derived quantiles was then assigned probability $\zeta_{ik} = 1/2$ $k = 1, 2$.

We only use the first three pairs of percentile-quantile values as input, and reserve the 99.9^{th} quantile, 9.00, to assess the accuracy of our approach by comparing it with the estimate $F_{pred}(9.00)$.

Step2: Adaptive Algorithm Input

• *Choose a region,* Θ*, for the support of the prior,* $\pi(\boldsymbol{\theta})$

We propose the following region:

$$\Theta = (\mu^L, \mu^U) \times (\sigma^L, \sigma^U) \times (\xi^L, \xi^U) = (0, 0.25) \times (0, 1) \times (0, 0.5). \tag{4.10}$$

The following two reasons lead us to believe this is a conservatively large region. First, X, lies on the positive real line, so we needn't concern ourselves with covering negative values. Second, we observe that for a GEV distribution with parameters $(\mu, \sigma, \xi) = (0.25, 1, 0.5)$

$$F^{-1}(0.9|\mu = 0.25, \sigma = 1, \xi = 0.5) = 4.4$$

$$F^{-1}(0.95|\mu = 0.25, \sigma = 1, \xi = 0.5) = 7.1$$

$$F^{-1}(0.99|\mu = 0.25, \sigma = 1, \xi = 0.5) = 18.2,$$

all of which comfortably exceed the corresponding vague quantiles given in Table 1. Other regions in (μ, σ, ξ) space tend either to have upper tails that are too heavy, or to place too much mass on negative values to be of use in modeling the phenomenon generating this volatility measure.

- *Choose π_r for $i = 1 \ldots R_0$ in the representation of $\pi(\boldsymbol{\theta})$ in the initial iteration*

We space nine π_r symmetrically about the center of the region Θ. Since we standardize according to equation (3.3), the standardized values of points in Θ lie in the rectangular region $(0, 1/3) \times (0, 1/3) \times (0, 1/3)$, as seen in Figure 1:(a). We refine this region after observing the results in Step 3.

- *Choose the convergence criterion for successive prior solutions, $\tilde{\pi}_t(\boldsymbol{\theta})$*

When the KL-divergence between successive solutions, $\tilde{\pi}_{t-1}(\boldsymbol{\theta})$ and $\tilde{\pi}_t(\boldsymbol{\theta})$, is less than 0.1, we deem convergence to have occurred. The last solution, $\tilde{\pi}_t^{opt}(\boldsymbol{\theta})$, is then taken as the optimal prior, $\tilde{\pi}^{opt}(\boldsymbol{\theta})$.

Step3: Examine Output

- *Run the adaptive algorithm until convergence of successive solutions is achieved*

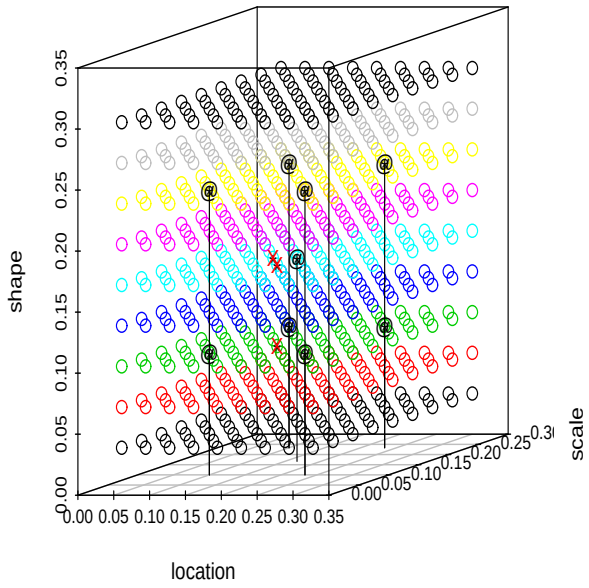
The locations of the means of the π_r used at each iteration are shown in Figures 1:(a) through 1:(f). The different colored planes of lattice points are used to assist in the location of points in the 3-dimensional figure. The centers of the new π_r allocated at each iteration appear to be located predominantly near the face of the cube where μ takes its largest values, so we expect to find $\tilde{\pi}_t^{opt}(\boldsymbol{\theta})$ concentrated in that region of the parameter space.

The KL-divergence between priors in successive iterations is shown in Figure 2. Since $KL(\tilde{\pi}_5(\boldsymbol{\theta}), \tilde{\pi}_6(\boldsymbol{\theta})) < 0.1$, we take $\tilde{\pi}^{opt}(\boldsymbol{\theta}) = \tilde{\pi}_6(\boldsymbol{\theta})$.

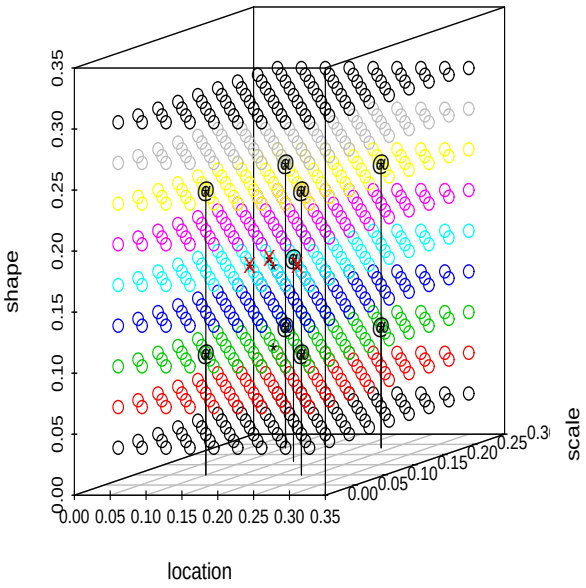
Table 2: Mixture components π_r along with weights β_i from equation 2.5, which contribute to the optimal prior $\tilde{\pi}^{opt}$. Each π_r is identified by its parameter vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{p+1})$.

r	$\boldsymbol{\alpha}_r$	β_r	r	$\boldsymbol{\alpha}_r$	β_r
1	(5,5,5,15)	.02	12	(14,14,8,24)	.12
2	(3,3,3,21)	<.01	13	(21,21,12,36)	.04
3	(3,7,3,17)	<.01	14	(18,21,12,39)	.05
4	(3,3,7,17)	.01	15	(27,21,9,33)	.01
5	(3,7,7,13)	.02	16	(36,28,12,44)	.07
6	(7,3,3,17)	<.01	17	(32,28,12,48)	<.01
7	(7,7,3,13)	.07	18	(36,32,16,36)	<.01
8	(7,3,7,13)	<.01	19	(40,35,15,60)	.06
9	(7,7,7,9)	<.01	20	(45,35,15,55)	.12
10	(14,14,4,28)	.31	21	(25,35,15,75)	<.01
11	(12,14,6,28)	.10			

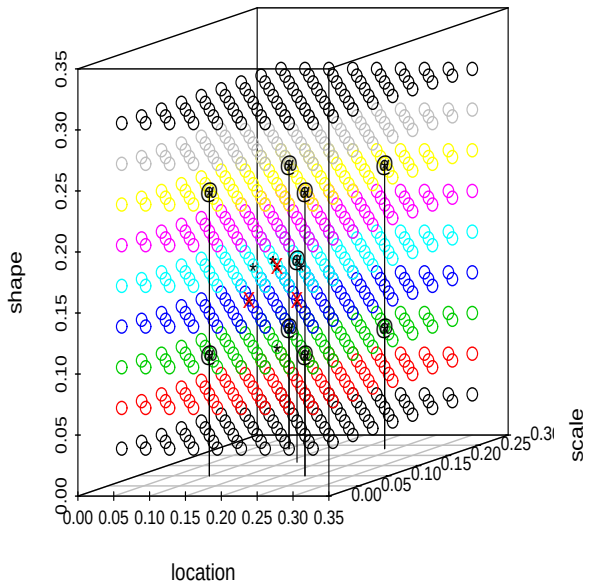
(a)


shape
location
scale
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35

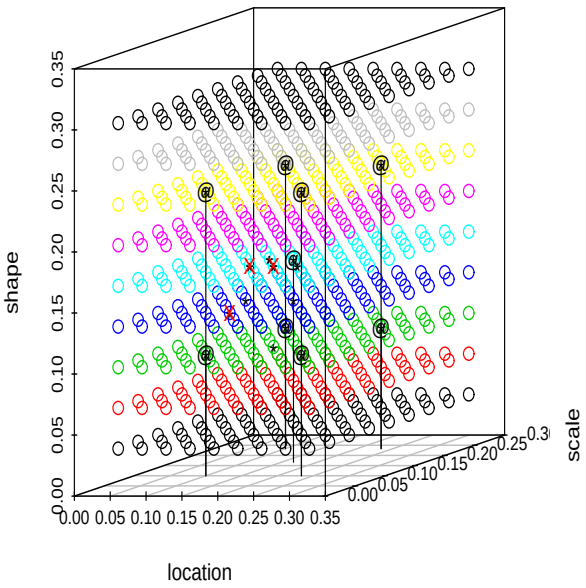
(b)


shape
location
scale
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35

(c)


shape
location
scale
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35

(d)


shape
location
scale
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35

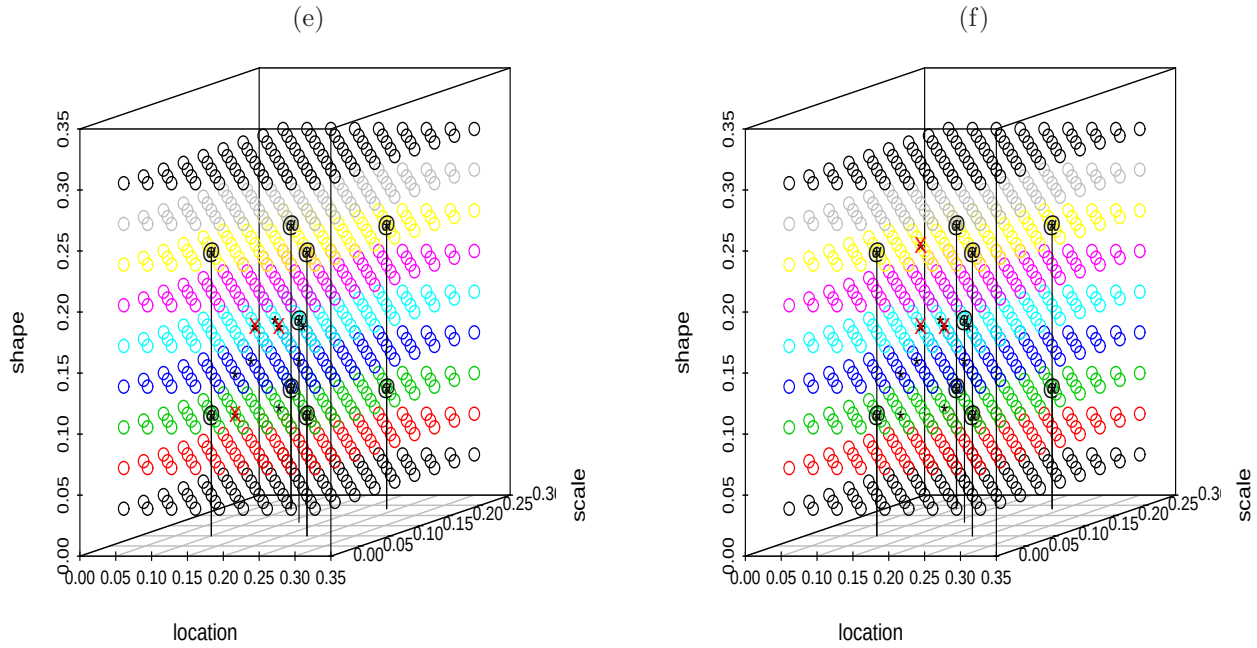


Figure 1: Locations of the means of π_r in successive iterations- (a):iteration 1...(f):iteration 6. The means of π_r used in the 1^{st} iteration are marked "@", means of new π_r to be included at the subsequent iteration are marked "**X**", and means of from previous iterations are marked "*".

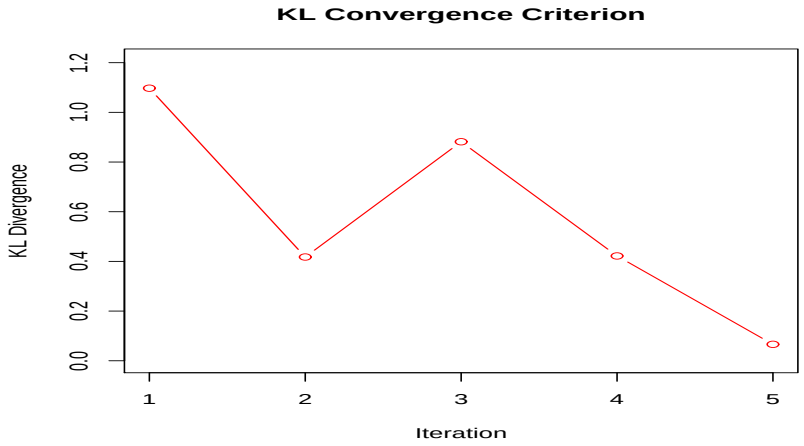


Figure 2: KL-divergence between the $\tilde{\pi}_t$ in consecutive iterations of the adaptive algorithm.

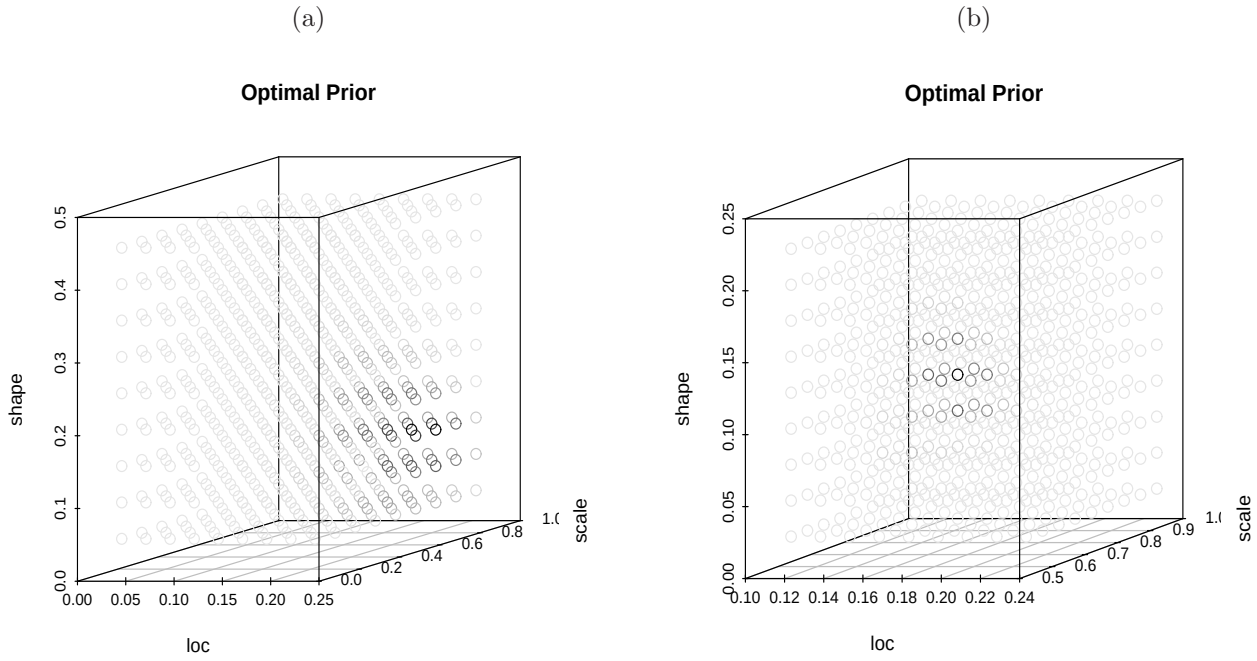


Figure 3: Representation of the optimal prior $\tilde{\pi}^{opt}(\mu, \sigma, \xi)$ using the definition of Θ given (a): in equation 4.10, and (b): in equation 4.11. Darker points indicate regions of higher density.

• *Investigate the relationship between the expert-provided information in Step 1 and both* $\tilde{\pi}^{opt}(\boldsymbol{\theta})$*, and* F_{prp}*. If the expert deems the solution implausible, explore solutions in other regions in the parameter space or revisit the information in Step 1 to accommodate the expert's belief.*

The pdf of the optimal prior, $\tilde{\pi}^{opt}(\boldsymbol{\theta}) = \tilde{\pi}_6(\boldsymbol{\theta})$, is shown in Figure 3:(a). This distribution is composed of the mixture of Dirichlet distributions given in Table 2. The component distributions, π_1 through $\pi_2 1$, are shown in the order in which they were introduced: π_1 to π_9 in the 1^{st} iteration, and 3 more at each subsequent iteration. As indicated in equation (3.4), the dispersion of the component distributions added at later iterations is diminished.

The density is highest in the region near $(\mu, \sigma, \xi) = (0.2, 0.6, 0.15)$, and it drops off relatively quickly, with very little mass in those subregions within Θ where μ and σ are near 0, and at the upper end of the range of ξ. Therefore, we narrow our focus and redefine Θ as

$$\Theta = (\mu^L, \mu^U) \times (\sigma^L, \sigma^U) \times (\xi^L, \xi^U) = (0.1, 0.25) \times (0.5, 1) \times (0, 0.25), \tag{4.11}$$

and reapply the same methodology. Upon rerunning the algorithm with the same convergence measure

(a) (b)

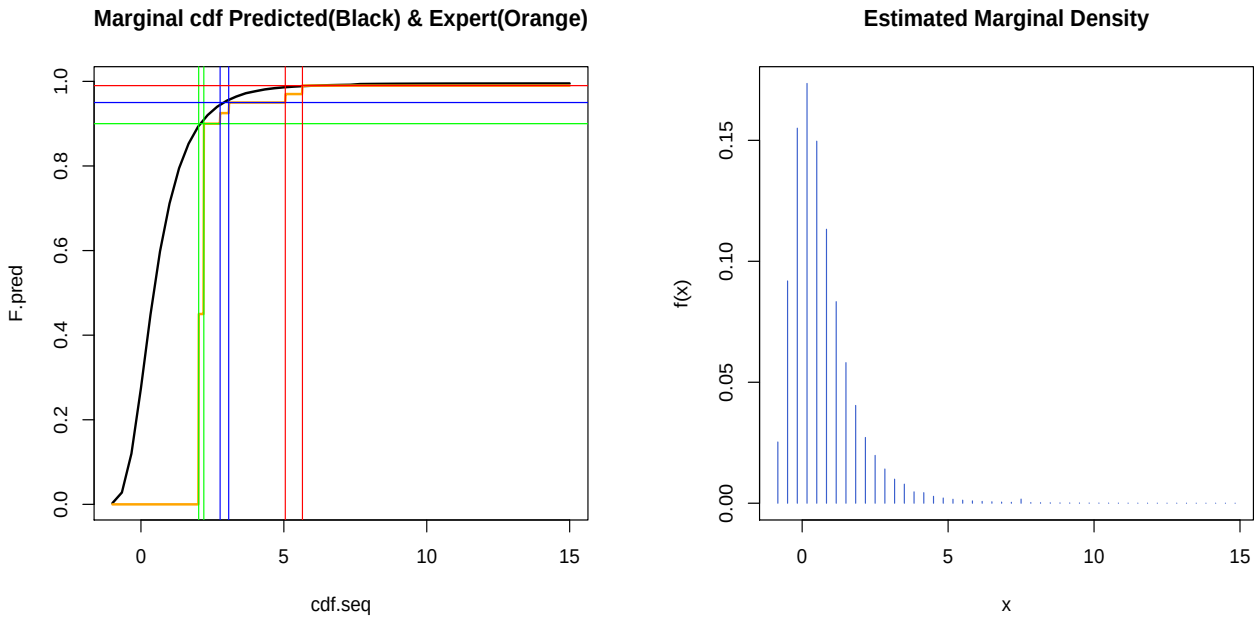


Figure 4: (a): The cdf indicates the expert provided percentiles (horizontal red, green and blue lines), the associated quantiles (vertical red, green and blue lines), the stepwise cdf corresponding to the expert provided vague quantitative information (orange), and the cdf resulting from equation 2.3 with $\tilde{\pi}^{opt}(\boldsymbol{\theta})$ substituted for $\pi(\boldsymbol{\theta})$. (b): The pdf associated with F_{pred}.

and the same relative initial distribution of π_r with respect to the new rectangular bounds on Θ, we obtain $\tilde{\pi}^{opt}(\boldsymbol{\theta})$ as shown in Figure 3:(b). The density now appears to be comfortably situated near the center of the parameter region of interest, suggesting that the region has now been specified adequately.

The prior predictive distribution, F_{prp}, is shown in Figure 4:(a), along with the associated pdf in Figure 4:(b) evaluated over an evenly spaced interval of points. The estimate of the 99.9th percentile resulting from the optimal predictive CDF is $F_{prp}(9.00) = 0.998$, which is in very near agreement with the empirical value of 0.999. We note that F_{prp} does cover negative values. however, this is not to be unexpected since we only used upper-tailed quantiles to inform our analysis. Had we wished to jointly model the extreme and non-extreme components of the process simultaneously we could have adjusted our likelihood accordingly (Tancredi, Anderson, O'Hagan 2006).

5 DISCUSSION

As summarized in Gosling et al. (2007), expert opinion can play a significantly more influential role in statistical analyses pertaining to extreme and rare events than it does in typical situations. In recognition of this fact, we used such expert opinion-based information as the basis for our general functional estimation procedure. However, while our methodology enables the estimation of functionals based on very limited information, caution must be taken with respect to the sensitivity of the functional estimate to each piece of information provided. An approach to mitigating the undue influence of any one particular expert's opinion on the resulting estimate is to consult with multiple experts (Gelfand et al. 1995). In this multiple expert scenario, rather than the democratic Opinion Pool approach advocated by Stone (1961), one may choose to downweight any outlying expert opinions for each quantile, corresponding to a smaller value of $\zeta_{i,k}$.

As an alternative to optimizing equations (2.5), (2.6), where percentiles play the crucial role, one may prefer to incorporate the values of the associated quantiles into the optimization procedure. This change has potentially large consequences in extreme value scenarios where small percentile changes in the tails may lead to significant changes in the quantiles. Insofar as these quantiles correspond to large floods, dangerous disease outbreaks or financially damaging scenarios, it is reasonable that the magnitude

of these events should play a role in the estimating procedure. Such a change is readily incorporated into the framework presented.

Though our methodology was motivated by circumstances typical of extreme value scenarios, and the example given in Section 5 used the GEV distribution, other likelihoods corresponding to mixture distributions, regressions, etc., are accommodated equally well. The limiting factor is the dimension of the parameter space, though the potential for computing the a_{irk} values in parallel, in conjunction with the lessening of the computational burden by choosing nested S_t, helps to address this difficulty.

APPENDIX.

Here we motivate the choice of the standardization

$$\theta_i' = \frac{\theta_i - \theta_i^L}{p(\theta_i^U - \theta_i^L)} \qquad i = 1 \dots p. \tag{5.12}$$

Consider a simplex S with vertices $\{v_0, v_1, \dots, v_p\}$, and define $w_i = v_i - v_0$ for $i = 1 \dots p$. It's known (Gritzmann & Klee 1994) that the p-dimensional volume of S is given by

$$Vol_p(S) = \frac{|det(WW^t)|^{1/2}}{p!} \tag{5.13}$$

where W is the $p \times p$ matrix with i^{th} row equal to w_i for $i = 1 \dots p$.

Recalling equation 3.2, let $v_0 = (\theta_1^L, \dots, \theta_p^L)$ and $v_i = (\theta_1^L, \dots, \theta_{i-1}^L, \theta_i^U, \theta_{i+1}^L, \dots, \theta_p^L)$ for $i = 1 \dots p$, and denote the upper corner of Θ by

$$v^* = (\theta_1^U, \dots, \theta_p^U). \tag{5.14}$$

The goal is to ensure that after standardization all points in Θ, including v^*, are in the support of $\tilde{\pi}(\boldsymbol{\theta})$.

Let $\boldsymbol{S}$ consist of all simplices with vertices of the form $\{\overline{\theta}_0, \overline{\theta}_1, \dots, \overline{\theta}_p\}$ such that

$$\overline{\theta}_0 = v_0 \quad \textit{and} \quad \overline{\theta}_i = v_0 + M_i(v_i - v_0) \quad \textit{for} \quad i = 1 \dots p, \tag{5.15}$$

where conditions on the M_i will be derived shortly.

We can then restate the goal as searching for the simplex $S^* \in \boldsymbol{S}$ such that:

1. $v^* \in S^*$ and

2. $Vol_p(S^*) \leq Vol_p(S) \qquad \forall\ S \in \boldsymbol{S}$ such that $v^* \in S$.

If by 1. $v^* \in S$ then $\exists(t_0, \ldots, t_p)$ such that *(i)*: $\sum_{i=0}^{p} t_i = 1$, and $0 \leq t_i$, where

$$v^* = \sum_{i=1}^{p} t_i \overline{\theta}_i.$$

Then from equations 5.14 and 5.15 we find

$$(\theta_1^U, \ldots, \theta_p^U) = (\theta_1^L + t_1 M_1(\theta_1^U - \theta_1^L), \ldots, \theta_p^L + t_p M_p(\theta_p^U - \theta_p^L)) \tag{5.16}$$

which implies $t_i = 1/M_i$ for $i = 1 \ldots p$. From *(i)* it follows that *(ii)*: $\sum_{i=1}^{p} 1/M_i = 1$ and $M_i \geq 1$.

Defining $w_i = \overline{\theta}_i - \overline{\theta}_0 = M_i(v_i - v_0) = M_i(0, \ldots, 0, \theta_i^U - \theta_i^L, 0, \ldots, 0)$ we use equation 5.13 to calculate

$$Vol_p(S) = \frac{1}{p!} \prod_{i=1}^{p} M_i(\theta_i^U - \theta_i^L). \tag{5.17}$$

Proceeding by the Lagrange multiplier method, with $\mathbf{M} = (M_1, \ldots, M_p)$ we define

$$f(\mathbf{M}) = \prod_{i=1}^{p} M_i(\theta_i^U - \theta_i^L) \quad and \quad g(\mathbf{M}) = \sum_{i=1}^{p} \frac{1}{M_i} = 1.$$

Then

$$\nabla f(\mathbf{M}) = \lambda g(\mathbf{M}) \quad and \quad g(\mathbf{M}) = 1$$

imply

$$f(\mathbf{M}) = \frac{-\lambda}{M_i} \quad and \quad p f(\mathbf{M}) = -\lambda,$$

which yield the solution $M_i = p$ for $i = 1 \ldots p$, which was the claim we wished to establish.

References

[1] Behrens, C. N., Lopes, H. F., and Gamerman, D. (2004), "Bayesian analysis of extreme events with threshold estimation," *Statistical Modelling*, **4**, 227–244.

[2] Berger, J.: *Statistical Decision Theory and Bayesian Analysis*, Springer, New York, 1985.

[3] Dey, D., and Liu, J. (2007), "A quantitative study of quantile based direct prior elicitation from expert opinion," *Bayesian Analysis*, **2**, 137-166.

[4] Gelfand, A. E., Mallick, B. K., and Dey, D. K. (1995), "Modeling Expert Opinion Arising as a Partial Probabilistic Specification," *Journal of the American Statistical Association,*, **90**, 598-604.

[5] Gosling, J. P., Oakley, J. E., and O'Hagan, A. (2007), "Nonparametric elicitation for heavy-tailed prior distributions," *Bayesian Analysis*, **2**, 693-718.

[6] Gritzmann, P., and Klee, V. (1994), "On the complexity of some basic problems in computational convexity. II: Volume and mixed volumes," *Polytopes: abstract, convex and computational*, (Scarborough, Ontario, Canada) (T. Bisztriczky et al., eds.), NATO ASI Series, Ser. C, Math. Phys. Sci., no. 440, NATO Advanced Study Institute, Kluwer Academic Publishers, pp. 373-466.

[7] Lorentz, G. G.: *Bernstein Polynomials*. Chelsea, New York, 1986.

[8] O'Hagan, A. (1998), "Eliciting expert beliefs in substantial practical applications," *The Statistician*, **47**, 21-35.

[9] Ramsay, J. O., and Silverman, B. W. (1997), *Functional Data Analysis*, Berlin: Springer-Verlag.

[10] Resnick, S.: *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer-Verlag, New York, 2006.

[11] Robbins, H. (1983), "Some Thoughts on Empirical Bayes Estimation," *Ann. Statist*, **11**, 713-723.

[12] Stone, M. (1961), "The Opinion Pool,' *The Annals of Mathematical Statistics*, Vol. 32, No.4, pp. 1339-1342.

[13] Tancredi, A., Anderson, C., and O'Hagan, A. (2006), "Accounting for threshold uncertainty in extreme value estimation," *Extremes*, **9**, 87-106.