Bayesian Inference on Multivariate Medians and Quantiles

Indrabati Bhattacharya, *University of Rochester Medical Center*

Subhashis Ghosal, *North Carolina State University*

As mentioned in the main paper, we start with some background and definitions on stochastic processes and empirical processes theory, and then move on to the proofs of Lemmas 2, 3 and 4. Next, we provide some details and results of the real data application, and finally we give the R codes used for our method.

S1 Some concepts on Stochastic Processes and Empirical Processes

For the rest of this sectiom, $L_r(Q)$ denotes the norm $\|f\|_{Q,r} = \left(\int |f|^r dQ\right)^{1/r}$.

Definition 1 (Separability of a Stochastic Process)**.** A stochastic process $\{X(t),\ t \in T\}$, where (T, ρ) is a separable metric space, is separable if there exists a countable subset $S \in T$ and a null set N such that for each $\omega \notin N$ and $t \in T$, there exists a sequence $\{s_m\} \in S$ with $\rho(s_m, t) \to 0$, and

$$|X(s_m, \omega) - X(t, \omega)| \to 0.$$

Definition 2 (Covering Numbers and Uniform Entropy)**.** The covering number $N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the minimal number of balls $\{g : \|g - f\| < \epsilon\}$ of radius ϵ needed to cover $\mathcal{F}$.

A class of functions $\mathcal{F}$ with the envelope function F is said to satisfy the uniform entropy condition if

$$\int_0^\infty \sup_Q \sqrt{\log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))}\, \mathrm{d}\epsilon < \infty, \tag{S1.1}$$

where the supremum has been taken over all finite discrete probability measures on with $\|F\|_{Q,2}^2 = \int F^2 \mathrm{d}Q > 0$.

Definition 3 (Bracketing Numbers)**.** For two functions l and u, the bracket $[l, u]$ is defined to be the set of all functions f with $l \le f \le u$. An ϵ-bracket in $L_r(P)$ is a bracket $[l, u]$ with $\|u - l\|_r \le \epsilon$.

The bracketing number $N_{[\,]}(\epsilon, \mathcal{F}, L_r(P))$ is the minimum number of ϵ-brackets needed to cover $\mathcal{F}$.

Definition 4 (VC Class of Sets)**.** Let $\mathcal{C}$ be a collection of subsets of a set $\mathfrak{X}$. We say that an arbitrary subset $S = \{x_1, x_2, \ldots, x_n\}$ of $\mathfrak{X}$ is shattered by $\mathcal{C}$ if for every subset $S' \subseteq S$, there exists $C \in \mathcal{C}$ such that $S' = S \cap C$.

The VC-index of the class $\mathcal{C}$ is the smallest n for which no set of size n is shattered by $\mathcal{C}$ i.e.

$$V(\mathcal{C}) = \inf\{n : \max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) < 2^n\}, \tag{S1.2}$$

where $\Delta_n(\mathcal{C}, x_1, \ldots, x_n) = \#\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\}$. A collection of measurable sets is called a VC class of sets if its VC-index is finite.

Definition 5 (Glivenko-Cantelli Class)**.** A function class $\mathcal{F}$ for which

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf| \to 0,$$

is called a P-Glivenko-Cantelli class, where the convergence can be in probability or almost surely.

Definition 6 (Donsker Class)**.** For a function class $\mathcal{F}$ and a probability measure P, suppose that

$$\sup_{f \in \mathcal{F}} |f(x) - Pf| < \infty. \tag{S1.3}$$

Let $\mathcal{L}^\infty(T)$ be the set of all functions $f : T \mapsto R$ such that

$$\sup_{t \in T} |f(t)| < \infty.$$

By viewing the empirical process $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ as a map into $\mathcal{L}^\infty(\mathcal{F})$, if

$$\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P) \rightsquigarrow \mathbb{G}, \qquad \text{in } \mathcal{L}^\infty(\mathcal{F}), \tag{S1.4}$$

for a tight Borel measurable element $\mathbb{G}$ in $\mathcal{L}^\infty(\mathcal{F})$, then $\mathcal{F}$ is called a P-Donsker class.

S2 Proofs

Proof of Lemma 2. Recall that, $\mathcal{F}_R = \mathcal{F}_{1R}\mathcal{F}_{2R}$, where

$$\mathcal{F}_{1R} = \left\{ \frac{|x_j - \theta_j|^{p-1}}{\|x-\theta\|_p^{p-1}} : j = 1, 2, \ldots, k, \|\theta - \theta_0\|_2 \leq R \right\}, \tag{S2.5}$$

$$\mathcal{F}_{2R} = \left\{ \text{sign}(\theta_j - x_j) : j = 1, 2, \ldots, k, \|\theta - \theta_0\|_2 \leq R \right\}. \tag{S2.6}$$

Every $f \in \mathcal{F}_{1R}$ is continuous at each x, hence $\mathcal{F}_{1R}$ has a countable subset $\mathcal{G}$ such that for every $f \in \mathcal{F}_{1R}$ there exists a sequence $g_m \in \mathcal{G}$ such that $g_m(x) \to f(x)$ for every x. Then by Example 2.3.4 of van der Vaart and Wellner (1996), $\mathcal{F}_{1R}$ is P-measurable for every P. Since every $f \in \mathcal{F}_{2R}$ is left-continuous at each x, same conclusion holds for $\mathcal{F}_{2R}$ as well.

A class of functions $\mathcal{F}$ is called a VC-major class of functions if the sets $\{x : f(x) > t\}$ with f ranging over $\mathcal{F}$ and t over $\mathbb{R}$ form a VC-class of sets. By Corollary 2.6.12 of van der Vaart and Wellner (1996), if $\mathcal{F}_{1R}$ is a bounded VC-major class of functions, then it satisfies the uniform entropy condition. It is easy to see that $\mathcal{F}_{1R}$ is bounded. We now show that $\mathcal{F}_{1R}$ is a VC-major class of sets, that is, the sets $\{x : f(x) > t\}$ with f varying over $\mathcal{F}_{1R}$ and t over $\mathbb{R}$ form a VC class of sets. Define the collection of sets $\mathcal{S} = \{S_{\theta,t} : \|\theta - \theta_0\|_2 \leq R,\ t \in \mathbb{R}\}$, where $S_{\theta,t}$ is defined as

$$S_{\theta,t} = \left\{ x : \frac{|x_j - \theta_j|^{p-1}}{\|x-\theta\|_p^{p-1}} > t,\ j = 1, 2, \ldots, k \right\}. \tag{S2.7}$$

We need to show $\mathcal{S}$ is a VC-class of sets. Note that $S_{\theta,t} = \cap_{j=1}^{k} S_{\theta,t}^{j}$, where

$S^j_{\theta,t}$ is defined as

$$S^j_{\theta,t} = \Big\{x : \frac{|x_j - \theta_j|^{p-1}}{\|x-\theta\|_p^{p-1}} > t\Big\}. \tag{S2.8}$$

In view of Lemma 2.6.17 of van der Vaart and Wellner (1996), it is enough to show

$$\mathcal{S}^j = \{S^j_{\theta,t} : \|\theta - \theta_0\|_2 \le R,\ t \in \mathbb{R}\} \tag{S2.9}$$

is a VC-class for every j; because if every $\mathcal{S}^j$ is a VC-class of sets, $\mathcal{S} = \sqcap_{j=1}^k \mathcal{S}^j = \{\cap_{j=1}^k S^j : S^j \in \mathcal{S}^j\}$ is also a VC-class of sets. Hence we only show that $\mathcal{S}^1 = \{S^1_{\theta,t} : \|\theta-\theta_0\|_2 \le R,\ t \in \mathbb{R}\}$ is a VC-class of sets. We can write $S^1_{\theta,t}$ as

$$S^1_{\theta,t} = \Big\{x : |x_1 - \theta_1|^p > \frac{t^{p/(p-1)}}{1 - t^{p/(p-1)}} \sum_{j=2}^k |x_j - \theta_j|^p)\Big\}.$$

Define $R^1_{\theta,c} = \Big\{x : |x_1 - \theta_1|^p > c\sum_{j=2}^k |x_j - \theta_j|^p\Big\}$ and $\mathcal{R}^1 = \{R^1_{\theta,c} : \theta \in \mathbb{R}^k, c \in \mathbb{R}\}$. It is enough to show that $\mathcal{R}^1$ is a VC-class, since $\mathcal{R}^1$ contains $\mathcal{S}^1$. For $i,\ j = 2, \dots, k,\ i \ne j$, we define

$$\begin{aligned} A_{0;\theta,c} =& \{x : (x_1 - \theta_1)^p > c\sum_{j=2}^k (x_j - \theta_j)^p\} \bigcap_{j-1}^k \{x :\ x_j - \theta_j \ge 0\}, \\ A'_{0;\theta,c} =& \{x : (\theta_1 - x_1)^p > c\sum_{j=2}^k (x_j - \theta_j)^p,\ \theta_1 - x_1 \le 0\} \bigcap_{j=2}^k \{x : x_j - \theta_j \ge 0\}, \end{aligned}$$

$$\begin{aligned}
A_{i;\theta,c} = & \{x : (x_1-\theta_1)^p > c(\theta_i - x_i)^p + c\sum_{\substack{j=2\\ j\neq i}}^{k}(x_j-\theta_j)^p, \\
& x_1-\theta_1\geq 0,\ \theta_i - x_i \geq 0,\ x_j-\theta_j\geq 0,\ j=2,\ldots,k,\ j\neq i\}, \\
A'_{i;\theta,c} = & \{x : (\theta_1-x_1)^p > c(\theta_i - x_i)^p + c\sum_{\substack{j=2\\ j\neq i}}^{k}(x_j-\theta_j)^p, \\
& \theta_1-x_1\geq 0,\ \theta_i - x_i \geq 0,\ x_j-\theta_j\geq 0,\ j=2,\ldots,k,\ j\neq i\}, \\
A_{ij;\theta,c} = & \{x : (x_1-\theta_1)^p > c(\theta_i - x_i)^p + c(\theta_j-x_j)^p + c\sum_{\substack{l=2\\ l\neq i,j}}^{k}(x_l-\theta_l)^p, \\
& x_1-\theta_1\geq 0,\ \theta_i - x_i \geq 0,\ \theta_j-x_j\geq 0,\ x_l-\theta_l\geq 0,\ l=2,\ldots,k, \\
& l\neq i,j\}, \\
A'_{ij;\theta,c} = & \{x : (\theta_1-x_1)^p > c(\theta_i - x_i)^p + c(\theta_j-x_j)^p + c\sum_{\substack{l=2\\ l\neq i,j}}^{k}(x_l-\theta_l)^p, \\
& \theta_1-x_1\geq 0,\ \theta_i - x_i \geq 0,\ \theta_j-x_j\geq 0,\ x_l-\theta_l\geq 0,\ l=2,\ldots,k, \\
& l\neq i,j\}.
\end{aligned}$$

Continuing this pattern, finally

$$\begin{aligned}
A_{n-1;\theta,c} = & \{x : (x_1-\theta_1)^p > c\sum_{j=2}^{k}(\theta_j-x_j)^p,\ x_1-\theta_1\geq 0,\ \theta_j-x_j\geq 0, \\
& j=2,\ldots,k\},
\end{aligned}$$

$$A'_{n-1;\theta,c} =\{x : (\theta_1 - x_1)^p > c\sum_{j=2}^{k}(\theta_j - x_j)^p,\ \theta_1 - x_1 \geq 0,\ \theta_j - x_j \geq 0,$$
$$j = 2, \ldots, k\}.$$

Using the preceding notations, $R^1_{\theta,c}$ can be written as

$$R^1_{\theta,c} =A_{0;\theta,c} \cup A'_{0;\theta,c} \cup \{\bigcup_{l=1}^{n-2} B_{l;\theta,c}\} \cup \{\bigcup_{l=1}^{n-2} B'_{l;\theta,c}\} \cup A_{n-1;\theta,c} \cup A'_{n-1;\theta,c}, \tag{S2.10}$$

where

$$B_{1;\theta,c} = \bigcup_{i=2}^{k} A_{i;\theta,c};\ \ B'_{1;\theta,c} = \bigcup_{i=2}^{k} A'_{i;\theta,c};\ \ B_{2;\theta,c} = \bigcup_{\substack{i=2\\i<j}}^{k}\bigcup_{j=2}^{k} A_{ij;\theta,c},$$

and so on. Since all the sets on the right hand side of (S2.10) are in the same form, if $\mathcal{C} = \{A_{0;\theta,c} : \theta \in \mathbb{R}^k, c \in \mathbb{R}\}$ forms a VC-class of sets, then $\mathcal{R}^1$ also forms a VC-class of sets. This follows from Lemma 2.6.17 of van der Vaart and Wellner (1996), which says that if $\mathcal{F}$ and $\mathcal{G}$ are VC-classes, then $\mathcal{F} \sqcup \mathcal{G} = \{F \cup G : F \in \mathcal{F},\ G \in \mathcal{G}\}$ also forms a VC-class. We can write $A_{0;\theta,c}$ as

$$A_{0;\theta,c} = \{x : (x_1 - \theta_1)^p > c\sum_{j=2}^{k}(x_j - \theta_j)^p\} \bigcap_{j=1}^{k}\{x : x_j - \theta_j \geq 0\}. \tag{S2.11}$$

Since p is a positive integer greater than 1, we can write

$$\{x : (x_1 - \theta_1)^p > c \sum_{j=2}^{k} (x_j - \theta_j)^p\}$$
$$= \{x : \textstyle\sum_{r=0}^{p} \binom{p}{r} x_1^{p-r} (-1)^r \theta_1^r > c \sum_{j=2}^{k} \sum_{r=0}^{p} \binom{p}{r} x_j^{p-r} (-1)^r \theta_j^r\}$$
$$= \{x : x_1^p - p x_1^{p-1}\theta_1 + \cdots + (-1)^p \theta_1^p - c \textstyle\sum_{j=2}^{k} (x_j^p - p x_j^{p-1} \theta_j$$
$$+ \cdots + (-1)^p \theta_j^p) > 0\}.$$

Consider the map $x \mapsto \phi(x)$, where

$$\phi(x) = \big\{x_1^p, x_1^{p-1}, x_1^{p-2}, \ldots, x_1, \sum_{j=2}^{k} x_j^p, \sum_{j=2}^{k} x_j^{p-1}, \ldots, \sum_{j=2}^{k} x_j, 1\big\}. \tag{S2.12}$$

Note that the class of functions $\{g_a(x) = a^T \phi(x) : a \in \mathbb{R}^{2p+1}\}$ is a finite dimensional vector space. The collection of sets

$$\{x : (x_1 - \theta_1)^p > c \sum_{j=2}^{k} (x_j - \theta_j)^p,\ \theta \in \mathbb{R}^k,\ c \in \mathbb{R}\}$$

is the same as $\mathcal{C}_1 = \{x : g_a(x) > 0,\ a \in \mathbb{R}^{2p+1}\}$ and $\mathcal{C}_1$ is a VC-class of sets by Lemma 2.6.15 of van der Vaart and Wellner (1996). Each of the classes $\{x : x_j - \theta_j \geq 0\}$ for $j = 1, 2, \ldots, k$, is a sub-collection of VC classes of sets $\mathcal{C}_2 = \{x : a^T x + b \geq 0,\ a \in \mathbb{R}^k,\ b \in \mathbb{R}\}$. Hence by Lemma 2.6.15 of van der Vaart and Wellner (1996), $\mathcal{C}$ forms a VC-class of sets.

Thus we proved that $\mathcal{F}_{1R}$ is a bounded VC major class of functions. Hence $\mathcal{F}_{1R}$ satisfies the uniform entropy condition.

For $\mathcal{F}_{2R}$, we see that the class of functions $\{x \mapsto \theta_j - x_j : \|\theta - \theta_0\|_2 \leq R\}$ belongs to a finite dimensional vector space and hence is a VC class. From the stability properties of VC classes (Example 3.3.9, van der Vaart and Wellner (1996)), the class of functions $\{x \mapsto \theta_j - x_j : \|\theta - \theta_0\|_2 \le R\}$ is also VC. Hence from Lemma 2.6.15 of van der Vaart and Wellner (1996), $\mathcal{F}_{2R}$ is a bounded VC major class of functions and satisfies the uniform entropy condition. Therefore $\mathcal{F}_R = \mathcal{F}_{1R}\mathcal{F}_{2R}$ is P_0-Donsker. □

Proof of Lemma 3. Define $M(P_0, \theta) = P_0 m_\theta = P_0(\|X - \theta\|_p - \|X\|_p)$. We show that for $0 < \epsilon < 1/4$, there exists $K > 0$ such that $\|\theta\|_p \ge 3K$ implies $M(P_0, \theta) > 0$. If $\|X\|_p \le K$ and $\|\theta\|_p \ge 3K$, then

$$\|X - \theta\|_p \ge \|\theta\|_p - \|X\|_p \ge \frac{2\|\theta\|_p}{3} + K - \|X\|_p \ge \frac{2\|\theta\|_p}{3},$$

Hence as $\|X\|_p \le K \le \|\theta\|_p/3$,

$$\|X - \theta\|_p - \|X\|_p \ge \frac{2\|\theta\|_p}{3} - \frac{\|\theta\|_p}{3} = \frac{\|\theta\|_p}{3}.$$

Now since always $\big|\|X - \theta\|_p - \|X\|_p\big| \le \|\theta\|_p$, we can write

$$\begin{aligned} M(P_0, \theta) &= \int_{\|X\|_p \le K} (\|X - \theta\|_p - \|X\|_p) \mathrm{d}P_0 + \int_{\|X\|_p > K} (\|X - \theta\|_p - \|X\|_p) \mathrm{d}P_0 \\ &\ge \|\theta\|_p \big(\frac{1}{3} P_0(\|X\|_p \le K) - P_0(\|X\|_p > K)\big) \\ &= \|\theta\|_p \big(\frac{1}{3} - \frac{4}{3} P_0(\|X\|_p > K)\big) \\ &\ge \|\theta\|_p \big(\frac{1}{3} - \frac{4}{3}\epsilon\big) > 0. \end{aligned}$$

We assume that $0 < \epsilon < 1/4$ and $K > 0$ have been chosen so that P_0 satisfies $P_0(\|X\|_p \le K) > 1 - \epsilon$. Hence $\|\theta(P_0)\|_p \le 3K$. Also, since $\mathbb{P}_n \rightsquigarrow P_0$, for some $0 < \epsilon < 1/4$ and $K > 0$, $\mathbb{P}_n$ satisfies $\mathbb{P}_n(\|X\|_p \leq K) > 1-\epsilon$ with high probability. Hence $\|\hat{\theta}_n\|_p \leq 3K$ with high probability. Similarly, we know that $\mathbb{B}_n - \mathbb{P}_n \overset{P_0^\infty \times \mathbb{B}_n}{\to} 0$ in the weak topology. Hence with high joint probability, $\mathbb{B}_n$ satisfies $\mathbb{B}_n(\|X\|_p \leq K) > 1-\epsilon$, leading to $\|\theta(\mathbb{B}_n)\|_p \leq 3K$ with high joint probability.

□

Proof of Lemma 4. We need to show that $\mathcal{S}^1 = \{S^1_{\theta,t} : \|\theta - \theta_0\|_2 \leq R,\ t \in \mathbb{R}\}$ is a VC class for any fixed $p > 1$, where $S^1_{\theta,t}$ is as defined in (S2.8). For $k = 2$, $S^1_{\theta,t}$ can be written as

$$S^1_{\theta,t} = \left\{x : |x_1 - \theta_1| > \left(\frac{t^{p/(p-1)}}{1 - t^{p/(p-1)}}\right)^{1/p} |x_2 - \theta_2|\right\}. \tag{S2.13}$$

Define $R^1_{\theta,c} = \{x : |x_1 - \theta_1| > c|x_2 - \theta_2|\}$ and $\mathcal{R}^1 = \{R^1_{\theta,c} : \theta \in \mathbb{R}^2, c \in \mathbb{R}\}$.

It is enough to show that $\mathcal{R}^1$ is a VC class, since $\mathcal{R}^1$ contains $\mathcal{S}^1$. We can write $R^1_{\theta,c}$ as

$$\begin{aligned} R^1_{\theta,c} =&\{x : (x_1 - \theta_1) > c(x_2 - \theta_2),\ x_1 \geq \theta_1,\ x_2 \geq \theta_2\} \\ &\bigcup \{x : (\theta_1 - x_1) > c(x_2 - \theta_2),\ \theta_1 \geq x_1,\ x_2 \geq \theta_2\} \\ &\bigcup \{x : (x_1 - \theta_1) > c(\theta_2 - x_2),\ x_1 \geq \theta_1,\ \theta_2 \geq x_2\} \\ &\bigcup \{x : (\theta_1 - x_1) > c(\theta_2 - x_2),\ \theta_1 \geq x_1,\ \theta_2 \geq x_2\}. \end{aligned}$$

Define $C_{\theta,c} = \{x : (x_1 - \theta_1) > c(x_2 - \theta_2),\ x_1 \geq \theta_1,\ x_2 \geq \theta_2\}$. By the same argument used in the previous proof, it is enough to show that $\mathcal{C} = \{C_{\theta,c} :$

$\theta \in \mathbb{R}^2, c \in \mathbb{R}\}$ forms a VC class of sets. This follows since $C_{\theta,c}$ can be written as

$$C_{\theta,c} = \{x : x_1 - \theta_1 > c(x_2 - \theta_2)\} \bigcup \{x : x_1 \geq \theta_1\} \bigcup \{x : x_2 \geq \theta_2\}.$$

Each of the sets in the right hand side of the above expression is a sub-collection of $\mathcal{C}_2 = \{x : a^T x + b \geq 0 : a \in \mathbb{R}^2, b \in \mathbb{R}\}$ which is a VC class by Lemma 2.6.15 of van der Vaart and Wellner (1996). Hence by Lemma 2.6.17 of van der Vaart and Wellner (1996), $\mathcal{C}$ is also a VC-class. □

S3 Analysis of Fisher's iris data

Fisher's iris data consists of three plant species, namely, Setosa, Virginica and Versicolor and four features, namely, sepal length, sepal width, petal length and petal width measured for each sample. The object of interest is the 4-dimensional spatial median of the above mentioned features. As mentioned in the main paper, we have considered a non-parametric Bayesian framework with a DP(α) prior with $\alpha = 2 \times \mathrm{N}_4(0_4, 10I_4)$. Then we compute the 95% credible ellipsoid of the 4-dimensional multivariate ℓ_1-median with $p = 2$ and report its four principal axes in Table 1. Also, for the purpose of illustration, in Figures 1, 2 and 3, we plot 6 pairs of features for each species and the credible ellipsoids for the corresponding two dimensional

Table 1: Principal axes of 95% credible ellipsoid of spatial median

1st axis	2nd axis	3rd axis	4th axis
0.0580	–0.3129	–0.6747	–0.6629
–0.1461	0.2193	–0.6143	–0.7437
–0.2965	0.8626	0.4089	–0.0252
0.9420	0.3252	–0.0081	–0.0824

ℓ_1-medians with $p = 2$.

Figure 1: 95% Credible ellipsoids of two-dimensional spatial medians for the species Setosa

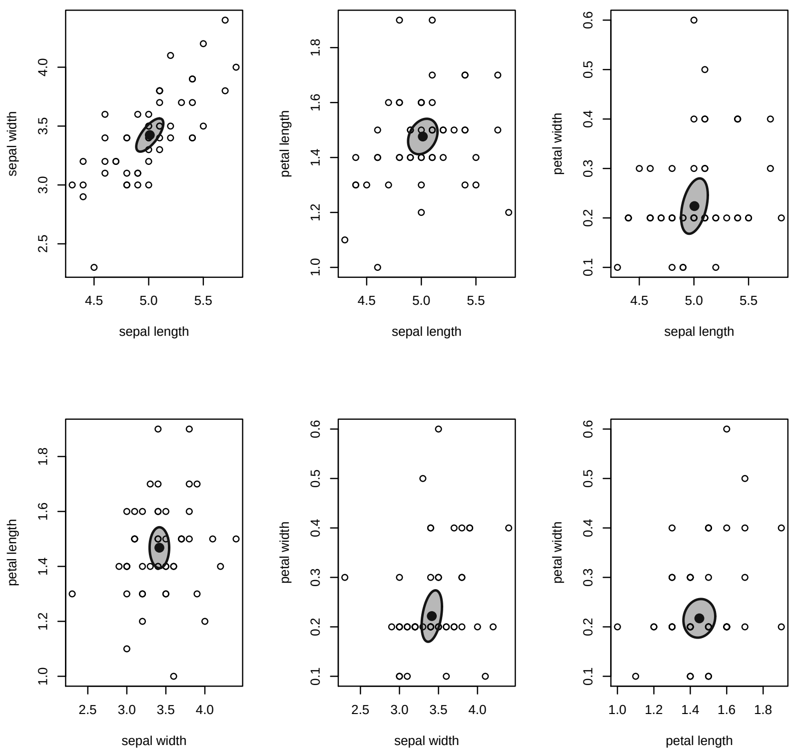


Figure 2: 95% Credible ellipsoids of two-dimensional spatial medians for the species Virginica

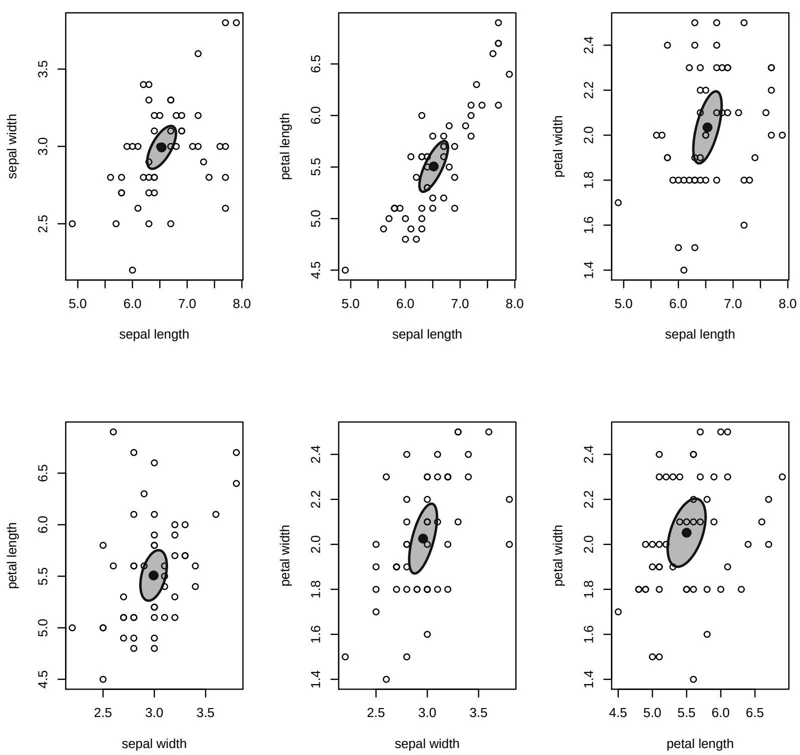
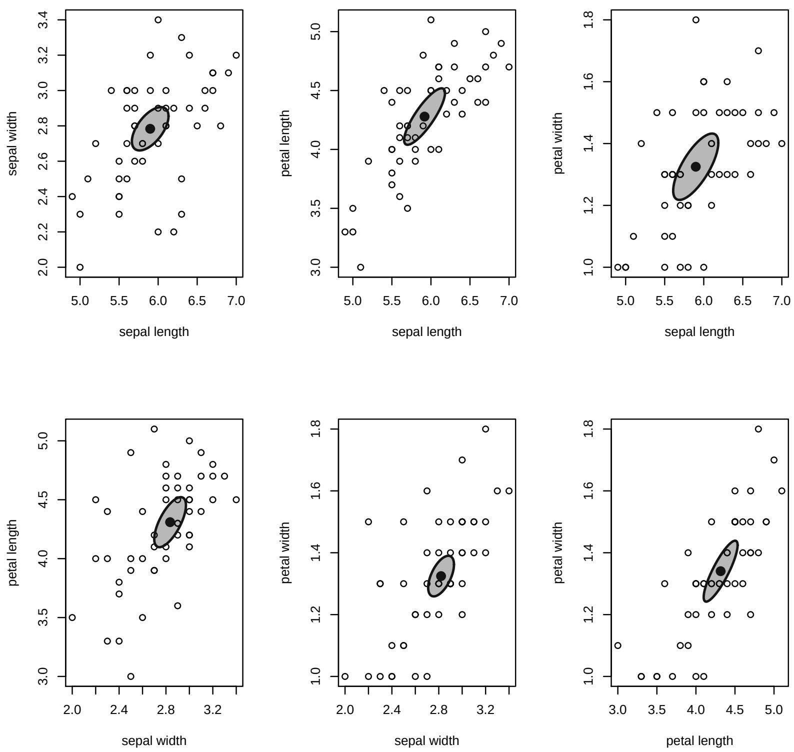


Figure 3: 95% Credible ellipsoids for the species Versicolor



S4 R Codes

```
# Constructing a 95% credible ellipsoid for a 2-dimensional
# non-affine equivariant spatial median

set.seed(120)

if (!require(MASS)) install.packages('MASS')
if (!require(ICSNP)) install.packages('ICSNP')
if (!require(lcmix)) install.packages('lcmix')

library(MASS)
library(ICSNP)
library(lcmix)


# Constructing the posterior credible ellipsoid for the spatial median

cred_med <- function(x,c,p,n1,alpha){
c <- 2
n <- nrow(x)
k <- ncol(x)
prb <- c/(c+n)
nB <- 0.2*n1  # Burn-in
nRem <- n1-nB # Remaining samples
index<- sample(c(1,2),1,prob=c(prb,(1-prb)))
```

```
if(index==2)
{
  # Bayesian bootstrap weight generation
  weights <- matrix( rexp(n * n1, 1) , ncol = n, byrow = TRUE)
  weights <- weights / rowSums(weights)
  bb <- matrix(NA,nrow=n1,ncol=k)
  for(i in 1:n1){
    myfun <- function(para){
      a <- 0
      for (j in 1:k)
      {
      a <- a+(abs(para[j]-t(x)[j,]))^p
      }
      sum(a^(1/p)*weights[i,])
    }
    fit <- optim(rep(0.5,k),myfun)
    bb[i,] <- fit$par
  }
}
if (index==1){
  b <- rbeta(n, 1, c)
  stick_break <- numeric(n)
  stick_break[1] <- b[1]
  stick_break[2:n] <- sapply(2:n, function(i) b[i] * prod(1 - b[1:(i-1)]))
  y <- mvrnorm(n,Sigma=10*diag(k),mu=rep(0,k))
```

```
  theta <- sample(1:n, prob = stick_break, replace = TRUE, size=n1)
  bb <- y[theta,]
}
  samples <- matrix(NA,nrow=nRem,ncol=k)
  samples[1:nRem,] <- bb[(nB+1):n1,]
center <- colMeans(samples) #Center of the ellipsoid
S <- cov(samples)  # Scale matrix of the ellipsoid
val <- c()
for(i in 1:nRem){
  val[i] <- (samples[i,]-center)%*%solve(S)%*%
    matrix(samples[i,]-center)
}
rad <- quantile(val,alpha) # 'Radius' of the ellisoid
return(list(c=center,S=S,rad=rad))
}


# Data generation from a mixture distribution
ind <- sample(x=c(1,2),size=1,prob=c(.9,.1))
if(ind==1){
  x <- mvrnorm(n,mu=c(1,1),Sigma=diag(2))
} else {
  x <- rmvgamma(n,shape=1,rate=1,diag(2))
}


sim <- cred_med(x,2,2,5000,.95)
```

```
print(sim$c)
print(sim$S)
print(sim$rad)

# For iris data
data(iris)
myIris <- cred_med(iris[,1:4],2,2,5000,.95)

# Vectors representing principal axes of the credible ellipsoid
print(eigen(solve(myIris$S))$vectors)

# Constructing a 95% credible ellipsoid for a 2-dimensional
# affine equivariant spatial median

set.seed(120)

if (!require(MASS)) install.packages('MASS')
if (!require(ICSNP)) install.packages('ICSNP')
if (!require(lcmix)) install.packages('lcmix')

library(MASS)
library(ICSNP)
library(lcmix)

# Transforming the data
```

```
trans_matrix <- function(x){
  n <- nrow(x)
  k <- ncol(x)
  xT <- t(x)
  xbar <- colMeans(x)
  D1 <- x-xbar
  sigmahat <- t(D1)%*%D1/n   #Estimate of sigma
  sigmainv <- solve(sigmahat)
  A <- matrix(unlist(combn(1:n,(k+1),simplify=F)),ncol=(k+1),byrow=T)
  ratio <- c()
  for(i in 1:nrow(A)){
    xmat <- cbind(xT[,A[i,2]]-xT[,A[i,1]],xT[,A[i,3]]-xT[,A[i,1]])
    xmat1 <- t(xmat)%*%sigmainv%*%xmat
    ratio[i] <- sum(diag(xmat1))/det(xmat1)
  }
  ind <- which.min(ratio)
  xalpha <- cbind(xT[,A[ind,2]]-xT[,A[ind,1]],xT[,A[ind,3]]-xT[,A[ind,1]])
  newx <- x[-A[ind,],]
  z <- t(solve(xalpha)%*%t(newx))
  return(list(xalpha=xalpha,newx=newx))
}


cred_medAE <- function(newx,xalpha,c,p,n1,alpha){
  z <- t(solve(xalpha)%*%t(newx))
  c <- 2
```

```
n <- nrow(z)
k <- ncol(z)
prb <- c/(c+n)
nB <- 0.2*n1  # Burn-in
nRem <- n1-nB # Remaining samples
index<- sample(c(1,2),1,prob=c(prb,(1-prb)))
if(index==2)
{
  weights <- matrix( rexp(n * n1, 1) , ncol = n, byrow = TRUE)
  weights <- weights / rowSums(weights)
  bb <- matrix(NA,nrow=n1,ncol=k)
  for(i in 1:n1){
    # This function has to be modified if dimension is more than 2
    myfun <- function(para){
      a <- 0
      for (j in 1:k)
      {
        a <- a+(abs(para[j]-t(z)[j,]))^p
      }
      sum(a^(1/2)*weights[i,])
    }
    fit <- optim(c(.5,.5),myfun)
    bb[i,1] <- fit$par[1]
    bb[i,2] <- fit$par[2]
  }
```

```
  }
  if (index==1){
    b <- rbeta(n, 1, c)
    stick_break <- numeric(n)
    stick_break[1] <- b[1]
    stick_break[2:n] <- sapply(2:n, function(i) b[i] * prod(1 - b[1:(i-1)]))
    y <- mvrnorm(n,Sigma=10*diag(k),mu=rep(0,k))
    theta <- sample(1:n, prob = stick_break, replace = TRUE,size=n1)
    bb <- y[theta,]
  }
  samples <- matrix(NA,nrow=nRem,ncol=2)
  samples[1:nRem,] <- xalpha %*% t(bb[(nB+1):n1,])
  center <- colMeans(samples) #Center of the ellipsoid
  S <- cov(samples)  # Scale matrix of the ellipsoid
  val <- c()
  for(i in 1:nRem){
    val[i] <- (samples[i,]-center)%*%solve(S)%*%
      matrix(samples[i,]-center)
  }
  rad <- quantile(val,alpha) # 'Radius' of the ellisoid
  return(list(c=center,S=S,rad=rad))
}

n <- 100
ind <- sample(x=c(1,2),size=1,prob=c(.9,.1))
```

```
if(ind==1){
  x <- mvrnorm(n,mu=c(1,1),Sigma=diag(2))
} else {
  x <- rmvgamma(n,shape=1,rate=1,diag(2))
}

xalpha <- trans_matrix(x)$xalpha
newx <- trans_matrix(x)$newx
sim <- cred_medAE(newx,xalpha,2,2,5000,.95)
print(sim$c)
print(sim$S)
print(sim$rad)
```

Bibliography

Van Der Vaart, A. W. & Wellner, J. A. (1996), Weak convergence and empirical processes, Springer.