IMS Lecture Notes–Monograph Series

Statistics on Placenta Shapes

Abhishek Bhattacharya

University of Arizona

Abstract: This report presents certain recent methodologies and some new analysis tools for measuring placenta shapes and the statistical analysis of the shape distribution given a sample of placenta coordinates. The placenta shapes have been used to predict features about the new born, for example its birth weight.

Contents

1. Introduction

The statistical analysis of placenta shapes based on a random sample of placentas is important in many areas. By noting key features of the shape, like the position of the **Umbilical Cord Insertion point** (CdInS) relative to the outer and inner perimeters, the thickness of the inter perimeter region, shape of the perimeter boundaries (convexity, roundness etc) and others, one may be able to predict features of the new born, like its gender, birth weight, presence of some disease/abnormality etc.

In this paper, I apply some general principles of nonparametric analysis on Riemannian manifolds to the placenta shape space and do inference based on a random sample of shapes. Some of these principles and techniques have been developed by me with my advisor, Professor Rabi Bhattacharya, which can be studied in detail in [1] and [2]. Also some of the analysis tools are new and have been developed in this report.

In Section 2, I briefly describe some nonparametric statistical analysis tools on general manifolds. I mention the concept of **Fréchet mean** of probability distributions on manifolds, and the sample Fréchet mean obtained from a random sample.

Department of Mathematics, University of Arizona, Tucson, Arizona 85721
Keywords and phrases: Placenta shape, Extrinsic and Intrinsic means, Nonparametric analysis

Depending on the chosen distance metric on the manifold, the Fréchet mean can be either **extrinsic** or **intrinsic** mean. I describe two sample tests carried out to compare the means of two different populations on the manifold, and hence distinguish between them. Also discussed, is a way of estimating the density of a distribution on the manifold.

To analyze placenta shapes, we need a way to measure the shape. For that, I consider Kendall's planer shape space Σ_2^k which is the space of configurations of k points on the plane (not all identical), identified modulo size and Euclidean motions of translation and rotation. The details of statistical analysis on this manifold, Σ_2^k have been described in Section 3.

So to get a placenta shape, I pick a few coordinates or landmarks on the perimeters along with the CdInS and then consider the shape of the picked configuration to be a point on the planer shape space. Then the analysis tools developed in Sections 2 and 3 is used to analyze placenta shapes. These landmarks are picked consistently for different placentas, the details of which is described in Section 4.

In Section 5, I compute the extrinsic and intrinsic mean shapes for the placenta sample using tools mentioned in Section 3. Then I project the sample onto the tangent space of the shape space at the intrinsic mean and hence get shape coordinates for each sample point. By different figures (Figure 6-13), I illustrate how perturbing the mean along different directions on the tangent space changes its shape.

Finally, in Section 6, I study the relation between placenta shape and Foetal Placental Ratio (FPR). I use a regression model to explain how FPR depends on placenta shape. I use the Kernal density estimate developed in Section 2.4, to estimate the distribution of FPR given the placenta shape alone.

Readers without much statistical or geometrical background may skip Sections 2-3 in their first reading and start with Section 4. Later on, whenever they come across new terms or concepts, they may look into Sections 2-3 for details.

2. Statistics on General Manifolds

Let $(\mathbf{M}, \mathbf{g})$ be a d dimensional connected complete Riemannian manifold, g being the Riemannian metric tensor on M. Let ρ be a distance metrizing the topology of M.

2.1. Fréchet Mean of Q

Definition 2.1. For a given probability measure Q on (the Borel sigmafield of) M, we define the **Fréchet function** of Q as

$$F(p) = \int_M \rho^2(p, x) Q(dx), \ p \in M. \tag{2.1}$$

Suppose $F(p) < \infty$ for some $p \in M$. Then the set of all p for which $F(p)$ is the minimum value of F on M is called the **Fréchet mean set** of Q, denoted by C_Q. If this set is a singleton, say $\{\mu_F\}$, then μ_F is called the **Fréchet mean** of Q.

If $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) with common distribution Q, and $Q_n \doteq \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j}$ is the corresponding empirical distribution, then the Fréchet mean set of Q_n is called the **sample Fréchet mean set**, denoted by C_{Q_n}. If this set is a singleton, say $\{\mu_{F_n}\}$, then μ_{F_n} is called the **sample Fréchet mean**.

It has been proved in Theorem 2.3, Bhattacharya and Patrangenaru [3], that if the Fréchet mean of Q, μ_F exists, then any measurable selection from the sample Fréchet mean set, C_{Q_n} is a strongly consistent estimator of μ_F. Proposition 2.1 computes the asymptotic distribution of μ_{F_n} under appropriate assumptions. For a proof, see Theorem 2.1, Bhattacharya and Patrangenaru [4].

Proposition 2.1. *Suppose the following assumptions hold:*
(i) *Q has support in a single coordinate patch, (U, ϕ), $\phi : U \longrightarrow \mathbb{R}^d$ smooth. Let $Y_j = \phi(X_j)$, $j = 1, \ldots, n$.*
(ii) *Fréchet Mean μ_F of Q is unique.*
(iii) *$\forall x$, $y \mapsto h(x, y) = \rho^2(\phi^{-1}x, \phi^{-1}y)$ is twice continuously differentiable in a neighborhood of $\phi(\mu_F) = \mu$.*
(iv) $\mathrm{E}(\mathrm{D}_r h(Y_1, \mu))^2 < \infty$ *$\forall r$.*
(v) $\mathrm{E}(\sup_{|u-v|\leq\epsilon} |\mathrm{D}_s\mathrm{D}_r h(Y_1, v) - \mathrm{D}_s\mathrm{D}_r h(Y_1, u)|) \to 0$ *as $\epsilon \to 0$ $\forall$ r, s.*
(vi) $\Lambda = ((\ \mathrm{E}\,(\mathrm{D}_s\mathrm{D}_r h(Y_1, \mu))\ ))$ *is nonsingular.*
Let μ_{F_n} be a measurable selection from C_{Q_n}, and write $\mu_n = \phi(\mu_{F_n})$. Then under the assumptions (i)-(vi)*,*

$$\sqrt{n}(\mu_n - \mu) \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1}\Sigma\Lambda^{-1}). \tag{2.2}$$

We can use Proposition 2.1 to test whether the population, Q has a specific mean, μ_{0F} given a iid sample $X_1, X_2, \ldots, X_n$ from Q. So we want to test $H_0 : \mu_F = \mu_{0F}$, against $H_1 : \mu_F \neq \mu_{0F}$ where μ_F is the true Fréchet mean of Q. The test statistic used is

$$T_n = n(\mu_n - \mu)^{'}(\hat{\Lambda}^{-1}\hat{\Sigma}\hat{\Lambda}^{-1})^{-1}(\mu_n - \mu) \tag{2.3}$$

where $\mu = \phi(\mu_{0F})$, $\mu_n = \phi(\mu_{F_n})$, μ_{F_n} being a measurable selection from C_{Q_n} and $(\hat{\Sigma}, \hat{\Lambda})$ are consistent sample estimates of (Σ, Λ). We reject H_0 at asymptotic level $1 - \alpha$ if $T_n > \mathcal{X}^2_d(1 - \alpha)$. Here $\mathcal{X}^2_d(1 - \alpha)$ is the upper $(1 - \alpha)^{\text{th}}$ quantile of the chi-squared distribution with d degrees of freedom.

Also we can perform a nonparametric test to test if two distributions Q_1 and Q_2 have the same Fréchet mean μ_F. That can help us distinguish between the two populations. Let $\mu = \phi(\mu_F)$. Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be iid observations from Q_1 and Q_2 respectively. Let Q_n and Q_m be the empirical distributions and μ_{n1} and μ_{m2} be the corresponding sample mean coordinates. We want to test $H_0 : \mu_{1F} = \mu_{2F} = \mu_F$ say, against $H_1 : \mu_{1F} \neq \mu_{2F}$ where μ_{1F} and μ_{2F} are the true intrinsic means of Q_1 and Q_2 respectively. Then the test statistic used is

$$T_{nm} = (n + m)(\mu_{n1} - \mu_{m2})'\hat{\Sigma}^{-1}(\mu_{n1} - \mu_{m2}), \tag{2.4}$$

$$\hat{\Sigma} = (m + n)\left(\tfrac{1}{n}(\hat{\Lambda}_1^{-1}\hat{\Sigma}_1\hat{\Lambda}_1^{-1} + \tfrac{1}{m}\hat{\Lambda}_2^{-1}\hat{\Sigma}_2\hat{\Lambda}_2^{'}\right), \tag{2.5}$$

(Λ_1, Σ_1) and (Λ_2, Σ_2) being the parameters in the asymptotic distribution of $\sqrt{n}(\mu_{n1} - \mu)$ and $\sqrt{m}(\mu_{m2} - \mu)$ respectively as defined in Proposition 2.1. $(\hat{\Lambda}_1, \hat{\Sigma}_1)$ and $(\hat{\Lambda}_2, \hat{\Sigma}_2)$

are consistent sample estimates. In case $n, m \to \infty$ such that $\frac{n}{m+n} \to \theta$, $0 < \theta < 1$; then under the hypothesis of Proposition 2.1, assuming H_0 to be true,

$$\sqrt{n+m}(\mu_{n1} - \mu_{m2}) \xrightarrow{\mathcal{L}} N_d(0, \frac{1}{\theta}\Lambda_1^{-1}\Sigma_1\Lambda_1^{-1} + \frac{1}{1-\theta}\Lambda_2^{-1}\Sigma_2\Lambda_2^{-1}). \tag{2.6}$$

So $T_{nm} \xrightarrow{\mathcal{L}} \mathcal{X}^2_d$. We reject H_0 at asymptotic level $1-\alpha$ if $T_{nm} > \mathcal{X}^2_d(1-\alpha)$.

2.2. Intrinsic Mean

The natural choice for the distance on M is $\rho = d_g$, the geodesic distance under g. Then the Fréchet mean (set) of a distribution, Q is called its **intrinsic mean** (set). If $X_1, X_2, \ldots, X_n$ are iid observations from Q, then the sample Fréchet mean (set) is called the **sample intrinsic mean** (set). The existence of the intrinsic mean and the asymptotic distribution of the sample intrinsic mean in Proposition 2.1 depend on the curvatures of M. To indicate the role curvature plays in this endeavor, let $r_* = min\{inj(M), \frac{\pi}{\sqrt{\overline{C}}}\}$, where $\overline{C}$ is an upper bound of sectional curvatures of M if this upper bound is positive, and $\overline{C} = 0$ otherwise. Also, $inj(M) \equiv \inf\{d_g(p, C(p)) : p \in M\}$ is the **injectivity radius** of M, where $C(p)$ is the **cut locus** of p, i.e., the set of points of the form $\gamma(t_0)$, where γ is a geodesic and t_0 is the supremum of all $t > 0$ such that the geodesic from p to $\gamma(t)$ is distance minimizing. From a result due to Kendall [6], if $supp(Q) \subseteq B(p, \frac{r_*}{2})$, then Q has a unique intrinsic mean μ_I in $B(p, \frac{r_*}{2})$. If in addition, $supp(Q) \subseteq B(\mu_I, \frac{r_*}{2})$, the assumptions in Proposition 2.1 hold with $\phi = \exp_{\mu_I}^{-1} : B(\mu_I, \frac{r_*}{2}) \longrightarrow T_{\mu_I}M(\approx \mathbb{R}^d)$, and hence the sample intrinsic mean has asymptotic Normal distribution. For a proof and expression for the asymptotic dispersion in terms of curvature, see Theorem 2.2, Bhattacharya and Bhattacharya [1].

2.3. Extrinsic Mean

Another notion of mean, which is much easier to compute and exists under much broader conditions is called the **extrinsic mean** on manifolds. To get that we embed M isometrically into some higher dimensional euclidean space via some map, $\Phi : M \to \mathbb{R}^k$. We choose the distance on M as: $\rho(x, y) = \|\Phi(x) - \Phi(y)\|$, where $\|.\|$ denotes Euclidean norm ($\|u\|^2 = \sum_{i=1}^k {u_i}^2$, $u = (u_1, u_2, .., u_k)'$). Let Q be a probability measure on M with finite Fréchet function. The Fréchet mean (set) of Q with respect to the above distance, is called the **extrinsic mean**(set) of Q. If X_j $(j = 1, \ldots, n)$ are iid observations from Q, then the sample Fréchet mean(set) is called the **extrinsic sample mean**(set).

Let $\tilde{M} = \Phi(M) \subset \mathbb{R}^k$. For every $u \in \mathbb{R}^k$ there exists a compact set of points in $\tilde{M}$ whose distance from u is the smallest among all points in $\tilde{M}$. We denote this set by

$$Pu \equiv P_{\tilde{M}}u = \{x \in \tilde{M} : \|x - u\| \leq \|y - u\| \ \forall y \in \tilde{M}\}.$$

If this set is a singleton, u is said to be a **nonfocal point** of $\mathbb{R}^k$ (with respect to $\tilde{M}$); otherwise it is said to be a **focal point** of $\mathbb{R}^k$. Let $\tilde{Q} = Q \circ \Phi^{-1}$ be the image of Q on $\mathbb{R}^k$. If $\tilde{\mu} = \int_{R^k} u\tilde{Q}(du)$ is the mean of $\tilde{Q}$, then the extrinsic mean set of Q is given by $\phi^{-1}(P\tilde{\mu})$. Hence Q has an extrinsic mean iff $\tilde{\mu}$ is a nonfocal point of $\mathbb{R}^k$. For a proof, see Proposition 3.1 in Bhattacharya and Patrangenaru [3].

2.4. Nonparametric Density Estimation

To discriminate between two different distributions on the manifold, usually the mean comparision test devised in Section 2.1, is sufficient. If not, more elaborate procedures such as nonparametric density estimation becomes necessary. This can be thought of as a generalization of the common Kernal density estimation in Euclidean spaces. For that we assume that the distribution Q has a density $f(x)$ with respect to the volume measure V on M. Then $f(x)$ can be approximated as

$$f(x) = \lim_{\sigma \to 0} \frac{\int_M f(y) e^{-\frac{1}{2}\frac{d_g^2(y,x)}{\sigma^2}} V(dy)}{\int_M e^{-\frac{1}{2}\frac{d_g^2(y,x)}{\sigma^2}} V(dy)} \tag{2.7}$$

If M is symmetric, i.e. given any $p, p' \in M$, there exists an isometry, say $\psi : M \to M$ which maps p to p', then the denominator of (2.7) is independent of $x \in M$. Then

$$f_\sigma(x) = \frac{\int_M f(y) e^{-\frac{1}{2}\frac{d_g^2(y,x)}{\sigma^2}} V(dy)}{\int_M e^{-\frac{1}{2}\frac{d_g^2(y,x)}{\sigma^2}} V(dy)} \tag{2.8}$$

represents a probability density for all $\sigma > 0$. Given an iid sample, $X_1, \ldots, X_n$ from Q, we estimate $f_\sigma(x)$ by

$$\hat{f}_\sigma(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{e^{-\frac{1}{2}\frac{d_g^2(X_j,x)}{\sigma^2}}}{\int_M e^{-\frac{1}{2}\frac{d_g^2(y,x)}{\sigma^2}} V(dy)} \tag{2.9}$$

$\hat{f}_\sigma(x)$ is a consistent estimator of $f_\sigma(x)$ and for suitably choosen σ, is close to $f(x)$.

3. The Planer Shape space

Consider a set of k points on the plane, not all points being the same. We will assume $k > 2$ and refer to such a set as a **k-ad** or a set of k **landmarks**. For convenience we denote a k-ad by k complex numbers ($z_j = x_j + iy_j, 1 \le j \le k$), i.e., we will represent a k-ad by a set of k points on the complex plane. By the **shape** of a k-ad $\mathbf{z} = (z_1, z_2, \ldots, z_k)$, we mean the equivalence class, or orbit of $\mathbf{z}$ under the euclidean motions of translation, rotation and scaling. To remove translation, one may substract $\langle \mathbf{z} \rangle \equiv (\langle z \rangle, \langle z \rangle, \ldots, \langle z \rangle)$ ($\langle z \rangle = \frac{1}{k}\sum_{j=1}^k z_j$) from $\mathbf{z}$ to get $\mathbf{z} - \langle \mathbf{z} \rangle$. Rotation of the k-ad by an angle θ and scaling by a factor $r > 0$ are achieved by multiplying $\mathbf{z} - \langle \mathbf{z} \rangle$ by the complex number $\lambda = re^{i\theta}$. Hence one may represent the shape of the k-ad as a complex line in $\mathbb{C}^k$, passing through $\mathbf{z} - \langle \mathbf{z} \rangle$, namely, $\{\lambda(\mathbf{z} - \langle \mathbf{z} \rangle) \colon \lambda \in \mathbb{C} \setminus \{0\}\}$. Thus the space of k-ads is the set of all complex lines on the (complex $(k-1)$-dimensional) hyperplane, $H^{k-1} = \{w \in C^k \setminus \{0\} \colon \sum_1^k w_j = 0\}$. Therefore the shape space Σ_2^k of planer k-ads has the structure of the **complex projective space** $\mathbb{C}P^{k-2}$: the space of all complex lines through the origin in $\mathbb{C}^{k-1}$. As in the case of $\mathbb{C}P^{k-2}$, it is convenient to represent the element of Σ_2^k corresponding to a k-ad $\mathbf{z}$ by the curve

$$\pi(z) = [z] = \{e^{i\theta} \frac{(z - \langle \mathbf{z} \rangle)}{\|z - \langle \mathbf{z} \rangle\|} \colon 0 \le \theta < 2\pi\} \tag{3.1}$$

on the unit sphere in H^{k-1}. So if we denote by u the quantity $\frac{(\mathbf{z}-\langle \mathbf{z}\rangle)}{\|\mathbf{z}-\langle \mathbf{z}\rangle\|}$; called the **preshape** of the shape of $\mathbf{z}$, then u lies on what is called the **preshape sphere** and denoted by S_2^k. S_2^k is isometric to the unit sphere of (real) dimension $2k-3$, S^{2k-3}. Then the map $\pi : S_2^k \to \Sigma_2^k$, is a Riemannian submersion. This makes Σ_2^k a complete Reimannian manifold of dimension $2k-4$. Σ_2^k has all sectional curvatures bounded between 1 and 4 and its injectivity radius is $\frac{\pi}{2}$. From a result due to Kendall [6], if Q is a probability distribution on Σ_2^k with support in a geodesic ball of radius $\frac{\pi}{4}$, then it has an intrinsic mean, μ_I, in its support. Also from a result due to Bhattacharya and Bhattacharya [1], the sample mean from an iid sample has asymptoticaly Normal distribution, if $\text{supp}(Q) \subset B(\mu_I, R)$, where R is the unique solution of $tan(x) = 2x$, $x \in (0, \frac{\pi}{2})$. For expressions for the asymptotic distribution parameters, see Theorem 3.1, Bhattacharya and Bhattacharya [1].

Another representation of Σ_2^k is via the **Veronese-Whitney embedding** Φ into the space $S(k,\mathbb{C})$ of all $k \times k$ complex Hermitian matrices. $S(k,\mathbb{C})$ is viewed as a (real) vector space of dimension k^2. The embedding Φ is given by

$$\Phi \colon \Sigma_2^k \to S(k,\mathbb{C}), \tag{3.2}$$

$$\Phi([z]) = uu^* \ (u = \frac{(\mathbf{z}-\langle \mathbf{z}\rangle)}{\|\mathbf{z}-\langle \mathbf{z}\rangle\|}) \tag{3.3}$$

$$= ((u_i \bar{u}_j))_{1 \le i,j \le k}. \tag{3.4}$$

Define the **extrinsic distance** ρ on Σ_2^k by that induced from this embedding, namely,

$$\rho^2([z],[w]) = \|uu^* - vv^*\|^2 \ , u = \frac{z-\langle \mathbf{z}\rangle}{\|z-\langle \mathbf{z}\rangle\|} \ , v = \frac{w-\langle \mathbf{w}\rangle}{\|w-\langle \mathbf{w}\rangle\|} \tag{3.5}$$

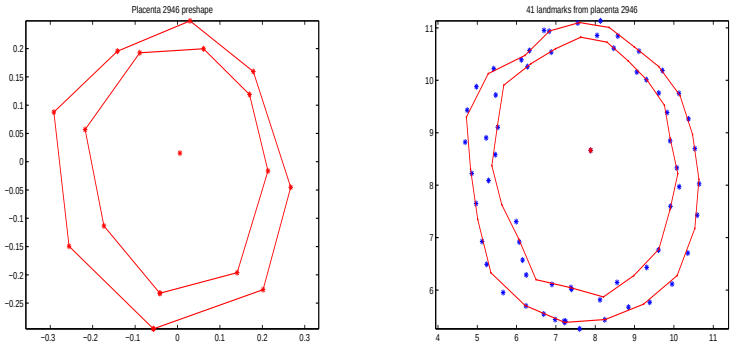
where for arbitrary $k \times k$ complex matrices A, B

$$\|A-B\|^2 = \sum_{j,j'} |a_{jj'} - b_{jj'}\|^2 = \text{Trace}(A-B)(A-B)^* \tag{3.6}$$

is just the squared euclidean distance between A and B regarded as elements of $\mathbb{C}^{k^2}$ (or, $\mathbb{R}^{2k^2}$). Let Q be a probability measure on Σ_2^k, and let $\tilde{\mu}$ denote the mean vector of $\tilde{Q} \doteq Q \circ \Phi^{-1}$, regarded as a probability measure on $\mathbb{C}^{k^2}$ (or, $\mathbb{R}^{2k^2}$). Then it can be shown that the extrinsic mean set of Q is the orbit under rotation of the space of unit eigenvectors for the largest eigen value of $\tilde{\mu}$. It follows that the extrinsic mean μ_E, say, of Q is unique if and only if the eigenspace for the largest eigenvalue of $\tilde{\mu}$ is (complex) one dimensional, and then $\mu_E = [\mu]$, $\mu(\neq 0) \in$ the eigenspace of the largest eigenvalue of $\tilde{\mu}$. Also it can be shown that in that case, any measurable selection from the sample extrinsic mean set, is a strongly consistent estimator of μ_E, and has asymptotic Normal distribution with mean μ_E. For expressions for the asymptotic dispersion, we refer to Section 3.3, Bhattacharya and Bhattacharya [2].

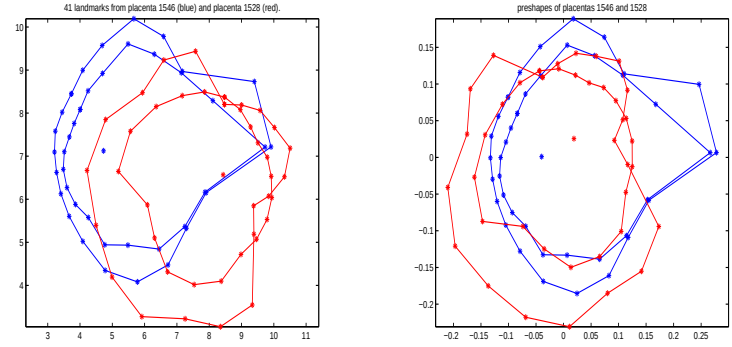
4. Measuring Placenta Shapes

To get a k-ad out of a placenta image, I pick a few landmarks on the outer perimeter, an equal number along the inner perimeter and the **Umbilical Cord Insertion point** (CdIns) for each image. The landmarks are picked at equiangular basis, i.e.



(a) (b)

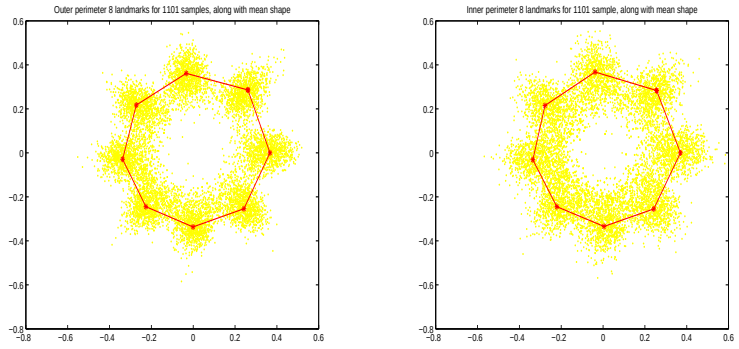
FIG 1. *(a) Preshape of 17 landmarks on Placenta 2946, (b) All landmarks (blue) along with the selected 41 landmarks (red) on Placenta 2946.*



(a) (b)

FIG 2. *(a) 41 landmarks on placentas 1546 (blue) and 1528 (red). (b) Their preshapes.*

the line joining each landmark along the outer and inner perimeters to the CdInS makes an angle from $0, 2\pi/k, \ldots, 2\pi(k-1)/k$ to the line joining the CdInS and the **point of rupture** (SOR). Here k is the number of landmarks chosen on the outer and inner perimeters respectively. If there is no point on a perimeter for some particular angle, I get that point by linear interpolation. That gives a configuration of $2k+1$ points for each placenta image. I try $k = 8$ and $k = 20$. When $k = 8$, the shape of the 17 landmark configuration lies in the planer shape space, Σ_2^{17} which has real dimension 30; and when $k = 20$, the shape of the 41 landmark configuration lies in Σ_2^{41} which has real dimension 78. Note that since I pick landmarks on equiangular basis, the shape of the 17-ads lie on a 15 dimensional space, while the 41-ad shapes lie on 39 dimensional space. However for computational simplicity, I work on the shape spaces of dimensions 30 and 78 respectively. Figure 1a shows the **preshape** of placenta 2946 using 17 landmarks. Figure 1b shows the whole placenta along with the selected 41 landmarks. Figure 2a shows 41 landmarks on placentas 1546 and 1528. Figure 2b shows their preshapes. Placenta 1528 has been rotated to bring it closest to the placenta 1546 preshape.



(a) (b)

FIG 3. *(a): 8 landmark outer perimeter mean shape along with sample outer perimeters. (b): 8 landmark inner perimeter mean shape along with sample inner perimeters.*

The results of shape analysis for the two k's are consistent as we shall see in the subsequent sections.

5. Placenta Mean Shapes

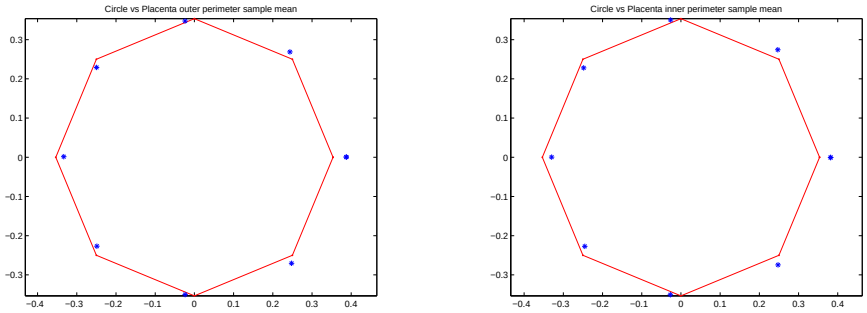
Given a sample of 1101 placenta configurations, I compute the **extrinsic sample mean** shapes as described in Section 3. Figure 3a,b show the preshapes of the extrinsic sample means of 8 inner and outer landmarks respectively along with the corresponding sample landmarks. The sample preshapes have been rotated and scaled so as to minimize their euclidean distances from the mean preshape. The figures suggest that both the outer and inner mean shapes are close to being circular, i.e. the 8-ad population mean shapes should be regular octagons. To test that I perform one sample tests as described in Section 2. The value of the test statistics, T_1 and T_2 for outer and inner octagonal mean tests and the p-values for the tests are:

$$T_1 = 228.33, \text{ P-value} = P(\mathcal{X}^2_{12} > T_1) < 10^{-16}$$
$$T_2 = 156.97, \text{ P-value} = P(\mathcal{X}^2_{12} > T_2) < 10^{-16}$$

The very small p-values force me to accept the alternative hypothesis that the sample shapes come from a population whose mean shape is different from a regular octagon. Figure 4 shows the plot of the 2 means along with octagons.

Next I compute the **sample intrinsic mean** shapes, using the 17 and 41 landmarks and compare them with the extrinsic mean shapes. Figure 5 shows the preshapes of the extrinsic and intrinsic sample means for 41 landmarks. The geodesic distance between the two means is 0.0019. Hence they are almost indistinguishable in the figure. Thus we will get very close results whether we use extrinsic or intrinsic distances in our analysis.

Having got the intrinsic mean shape, I project the data onto the **tangent space** of the planer shape spaces, Σ^k_2, $k = 17, 41$ at the respective intrinsic means using the **inverse exponential map**. That gives $2k-4$ dimensional coordinates for each



(a) (b)

FIG 4. *(a): 8 landmark outer perimeter mean shape along with a regular octagon. (b): 8 landmark inner perimeter mean shape along with a regular octagon. Red represents octagon edges, blue are the mean shape landmarks*

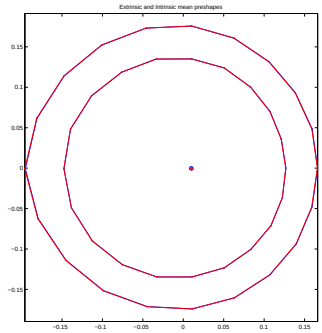


FIG 5. *Blue is the preshape of the extrinsic mean using 41 landmarks, red is the intrinsic mean preshape.*

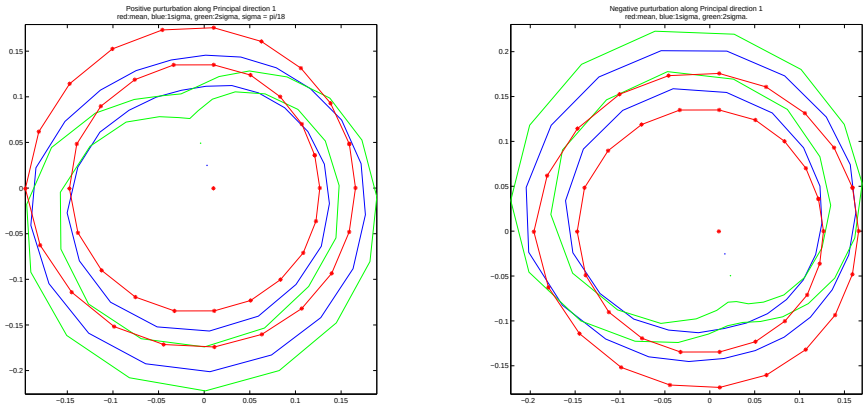
placenta shape, known as **normal coordinates**. Each placenta shape is a single point in the tangent space and therefore the 1101 sample placentas form a cloud of points in this high dimensional space. (As noted before, there is a 2-fold redundancy, so that the cloud of points actually occupy 15 and 39 dimensional spaces, depending on the value of k).

I perform a Principal Component Analysis (PCA) on the cloud of points. When I use $k = 17$ landmarks, the first two principal components explain about 64% of variation in shape and the first 11 components explain more than 90% of variation in shape. Table 1 shows the percent variation and the cumulative percent variation explained by the first 12 principal components(PC) when $k = 41$. These results suggest that placenta shapes lie on a much smaller dimensional submanifold of the shape space, which means that the landmarks are highly correlated.

What does the distribution of points along the principal directions tell us about placenta shapes? How are placenta shapes affected by movement along principal directions? Figures 6-13 illustrate the change in the 41-ad intrinsic mean shape

TABLE 1
Percent variation(V) and cumulative percent variation(CV) explained by first few PCs

PC	1	2	3	4	5	6	7	8	9	10	11	12
V	37.8	32.9	7.4	6.7	2.3	2.0	1.7	1.1	0.9	0.9	0.8	0.5
CV	37.8	70.7	78.1	84.8	87.1	89.0	90.7	91.8	92.8	93.6	94.4	95.0



(a) (b)

FIG 6. *(a) Perturbation along Principal direction 1 by 1σ, 2σ from the mean shape. (b) Perturbation by -1σ, -2σ. Red curve is the intrinsic mean of 41 landmarks, blue curves are perturbation by $\mp 1\sigma$, green are $\mp 2\sigma$. $\sigma = \frac{\pi}{18}$*

caused by perturbation along the first 8 principal directions. The perturbations are measured for times $\mp\sigma$ along each principal direction, where σ is the standard deviation for that principal component.

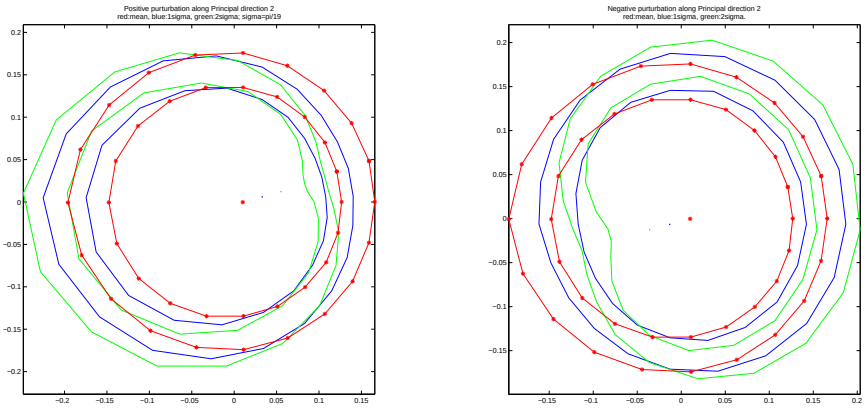
It is interesting to note that the Principal directions 1 and 2 seem to be reflection of each other. Perturbation of the mean along both these directions causes it to lose its convexity and the CdIns point moves more towards the inner perimeter edge.

Perturbation along Principal directions 3 and 4 makes the perimeters concave, however the position of CdIns does not change considerably. This is also true of the directions, 5-8.

Positive perturbation along direction 5 causes the 2 perimeters to get closer at certain points, and there is a cross over of the two perimeters when there is excessive perturbation. This is not true of negative perturbation.

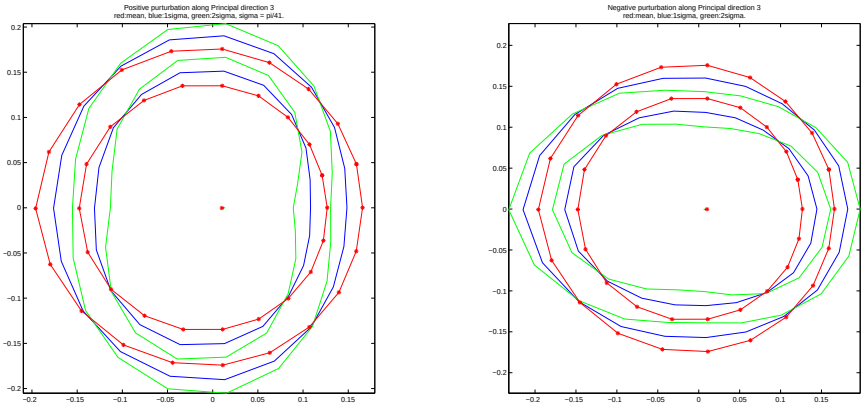
6. Relation between placenta shape and FPR

The objective of this paper is to study placenta shapes and use them to predict key features of the new born baby, for example, its birth weight, sex, presence of some disease etc. In this section, we study the relation between placenta shape and **Foetal Placental Ratio** (FPR). I have 41-ad shapes and FPR values for 1084 placentas. Figure 14 shows the histogram of the distribution of geodesic distances of sample shapes from the intrinsic mean. The mean distance is 0.2540, and the standard deviation is 0.1201. Figure 15 is the scatter plot of the geodesic distances against the corresponding FPR values. The plot suggests correlation between FPR



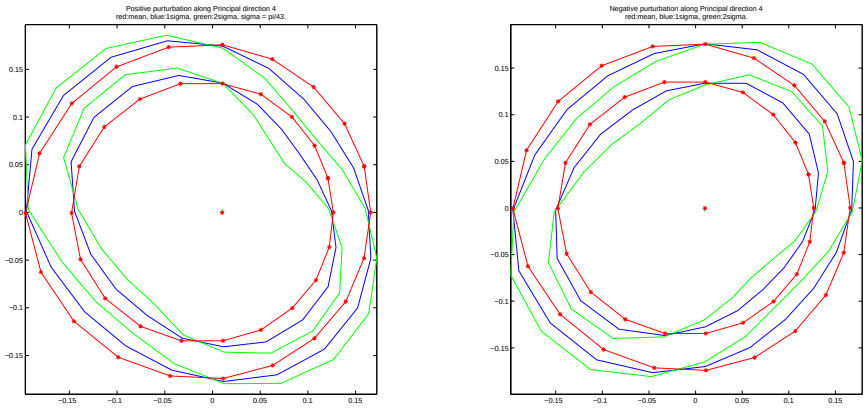
(a) (b)

FIG 7. *Perturbation along Principal direction 2 by* $\mp\sigma$, $\mp 2\sigma$ *from the mean shape.* $\sigma = \frac{\pi}{19}$



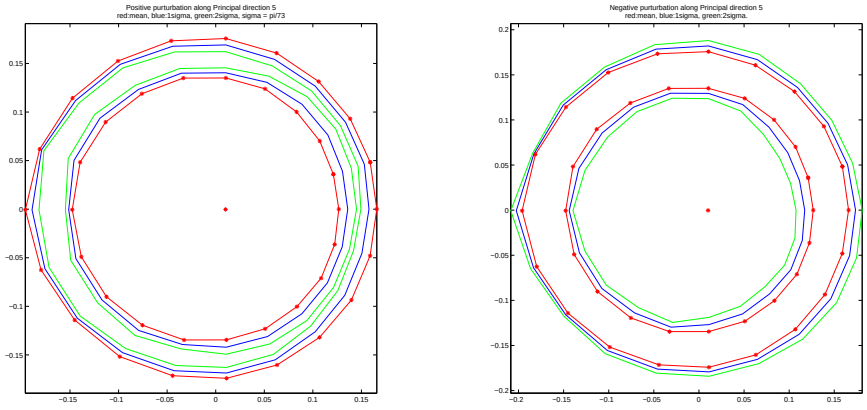
(a) (b)

FIG 8. *Perturbation along Principal direction 3 by* $\mp\sigma$, $\mp 2\sigma$ *from the mean shape.* $\sigma = \frac{\pi}{41}$



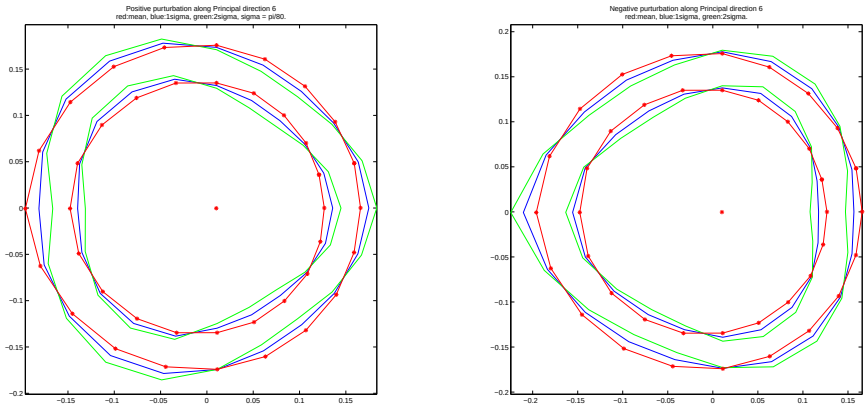
(a) (b)

FIG 9. *Perturbation along Principal direction 4 by* $\mp\sigma$, $\mp 2\sigma$ *from the mean shape.* $\sigma = \frac{\pi}{43}$



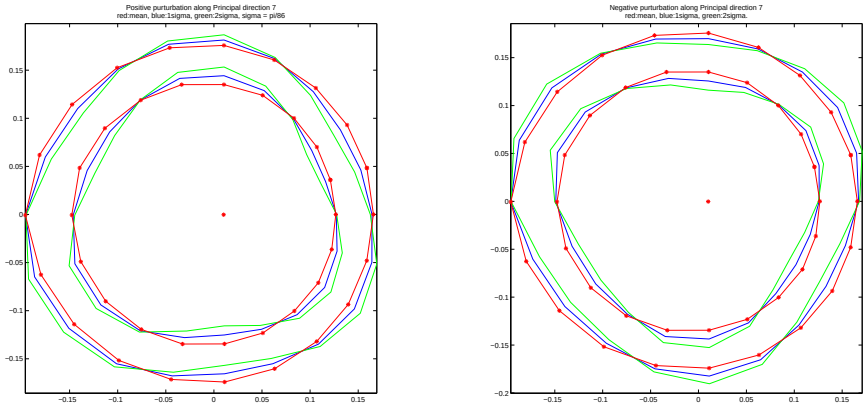
(a) (b)

FIG 10. *Perturbation along Principal direction 5 by* $\mp\sigma$, $\mp 2\sigma$ *from the mean shape.* $\sigma = \frac{\pi}{73}$



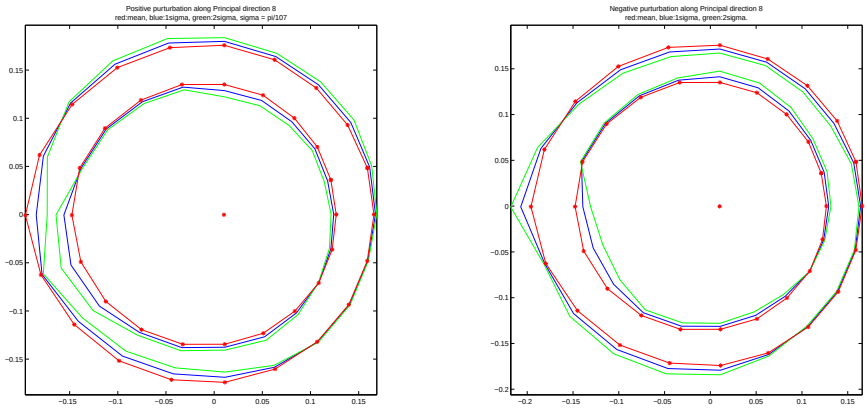
(a) (b)

FIG 11. *Perturbation along Principal direction 6 by* $\mp\sigma$, $\mp 2\sigma$ *from the mean shape.* $\sigma = \frac{\pi}{80}$



(a) (b)

FIG 12. *Perturbation along Principal direction 7 by* $\mp\sigma$, $\mp 2\sigma$ *from the mean shape.* $\sigma = \frac{\pi}{86}$



(a) (b)

FIG 13. *Perturbation along Principal direction 8 by* $\mp\sigma$, $\mp 2\sigma$ *from the mean shape.* $\sigma = \frac{\pi}{107}$

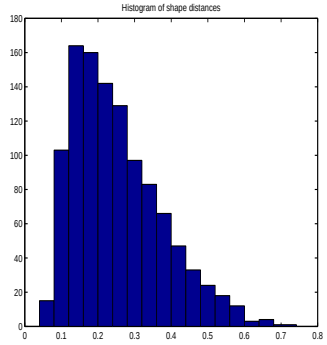


FIG 14. *Histogram of shape distance from mean shape*

and placenta shape.

To see how the two are correlated, firstly I regress FPR, say y, on the first few principal components of placenta shape, say $\mathbf{x} = (x_1, x_2, \ldots, x_t)$. I try a quadratic model as follows:

$$y = a_0 + \sum_{j=1}^{t} a_j x_j + \sum_{j=1}^{t} b_j x_j^2 + \sum\sum_{1 \le i < j \le t} c_{ij} x_i x_j + \epsilon \tag{6.1}$$

For the model (6.1), I estimate the coefficients as $\{\hat{a}_j\}_{j=0}^t$, $\{\hat{b}_j\}_{j=1}^t$ and $\{\hat{c}_{ij}\}_{1 \le i < j \le t}$ using the Least Squares method, and obtain 95% confidence intervals for the coefficients assuming a Normal distribution. I use the confidence intervals to test which coefficients are non zero at level 5%. I also compute the proportion of variation in y explained by the model, R^2 and test whether the model has any non zero coefficient other than a_0, i.e. if there is any interaction between y and $\mathbf{x}$. Table 2 shows the results of my analysis. Column 1 shows the shape components used in

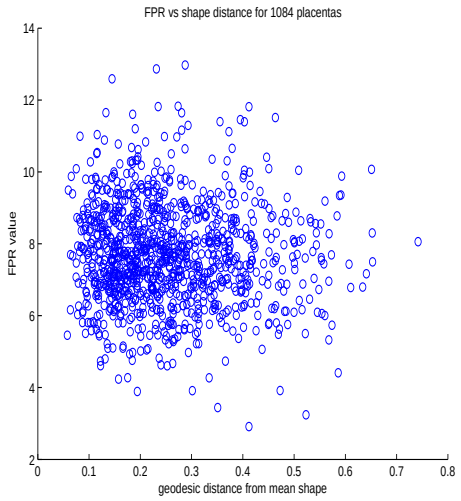


FIG 15. *FPR against placenta shape distance from mean shape*

TABLE 2
*Regression of FPR(y) on shape(***x***)*

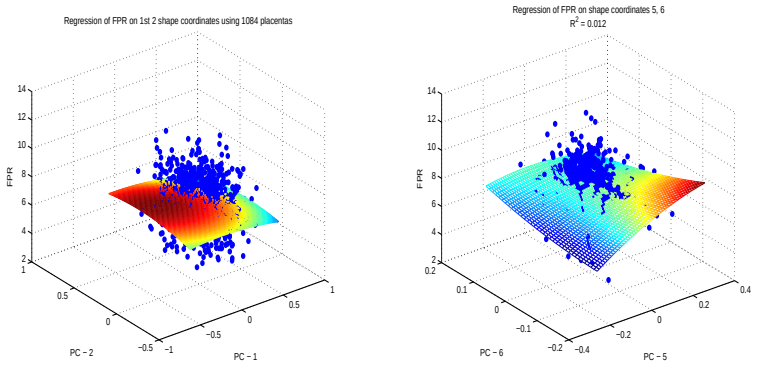
x	**Significant Coefficients**	R^2	**P-value**
x_1	$\hat{a}_0 = 7.7$, $\hat{a}_1 = -0.64$	0.0069	0.02
(x_1, x_2)	$\hat{a}_0 = 7.75$, $\hat{a}_1 = -0.63$	0.0089	0.0859
(x_1, x_2, x_3, x_4)	$\hat{a}_0 = 7.7$	0.0172	0.1768
$(x_1, x_2, x_3, x_4,)$ $x_5, x_6)$	$\hat{a}_0 = 7.7$, $\hat{a}_5 = 2.6$, $\hat{a}_6 = -2.5$	0.0338	0.1
$(x_1, x_2, x_3, x_4,$ $x_5, x_6, x_7, x_8)$	$\hat{a}_0 = 7.7$, $\hat{a}_5 = 3.24$, $\hat{a}_6 = -3.35$, $\hat{c}_{28} = 25.7$, $\hat{c}_{38} = 43.7$, $\hat{c}_{47} = -25.9$, $\hat{c}_{48} = 33.5$, $\hat{c}_{68} = -65.1$	0.0668	0.004
(x_5, x_6)	$\hat{a}_0 = 7.7$, $\hat{a}_5 = 2.5$	0.0119	0.02

the model explaining FPR as a function of shape, column 2 lists the estimates of the coefficients in that model that are found to be non zero (at level 5%), column 3 is R^2 and column 4 is the p-value for the F-test carried out to test for interaction between y and $\mathbf{x}$. If that p-value is less than 5%, I accept the hypothesis that the model has some non zero coefficient other than a_0 and hence is a good model.

The table suggests that I should use the model

$$y = a_0 + \sum_{j=1}^{8} a_j x_j + \sum_{j=1}^{8} b_j x_j^2 + \sum_{1 \le i < j \le 8} \sum c_{ij} x_i x_j + \epsilon \tag{6.2}$$

This model explains about 6.7% of variation in FPR and the p-value is 0.004 which is fairly small. It is interesting to note that the only non zero linear coefficients are that of principal components 5 and 6 and there are no non zero quadratic terms. This suggests that FPR depends on shape linearly through components 5 and 6. Of course there are non zero interaction terms like the coefficients of $x_2x_8, x_3x_8, x_4x_7, x_4x_8$ and x_6x_8. Figure 16a shows FPR as a function of principal components 1 and 2. This model explains 0.89% of FPR variation. Figure 16b shows FPR as a function of principal components 5 and 6. This model explains 1.19% of



(a) (b)

FIG 16. *(a) Scatter plot of FPR against x_1, x_2 along with best quadratic model. (b) Scatter plot of FPR against x_5, x_6 along with best quadratic model.*

FPR variation.

Next I use the nonparametric density estimation technique on manifolds, described in Section 2.4, to estimate the posterior distribution of FPR (y) given the placenta shape. To do that, I divide the FPR values into 5 ordered classes, say, $(-\infty, a_1]$, $(a_1, a_2]$, $(a_2, a_3]$, $(a_3, a_4]$ and (a_4, ∞) and then estimate the probability that considering the placenta shape alone, the placenta will fall into a particular FPR class.

To get the partition points dividing the FPR classes, $a_1, \ldots, a_p$, $p = 4$, I maximize the weighted sum of squared distances between the means of the $p+1$ groups, or equivalently, minimize the weighted sum of within group variations. The weights are proportional to the probability of the groups. Mathematically, I choose $\mathbf{a} = (a_1, \ldots, a_p)$, $p = 4$ so as to maximize

$$\phi(\mathbf{a}) = \sum_{i=1}^{p+1} P(a_{i-1}, a_i] \left(\mathrm{E}(Y | a_{i-1} < Y \leq a_i) - \mathrm{E}(Y)\right)^2 \tag{6.3}$$

with respect to $\mathbf{a}$. In (6.3), P denotes the FPR probability distribution and Y has distribution P. There $a_0 = -\infty$ and $a_{p+1} = \infty$. Given an iid sample $Y_1, \ldots, Y_n$ with common distribution P, I get sample estimate for $\mathbf{a}$, $\hat{\mathbf{a}} = (\hat{a}_1, \ldots, \hat{a}_p)$ by replacing P in (6.3) by the sample empirical distribution, P_n. For this specific sample,

$$\hat{a}_1 = 5.985,\ \hat{a}_2 = 7.245, \hat{a}_3 = 8.354,\ \hat{a}_4 = 9.71.$$

Figure 17 shows the histogram of the FPR values along with the 5 classes. Red lines denote class boundary. Note that there are other ways of classifying FPR values, for example that can be done based on some biological consideration, but for this case, I use purely statistical reasoning.

If the prior probabilities of the $p+1$ classes are $\pi = (\pi_1, \ldots, \pi_{p+1})$ ($\pi_j = P(a_{j-1}, a_j]$, $j = 1, \ldots, p+1$), then their posterior probabilities given a shape $\mathbf{x}$ are $\varpi(\mathbf{x}) = (\varpi_1(\mathbf{x}), \ldots, \varpi_{p+1}(\mathbf{x}))$ where

$$\varpi_j(\mathbf{x}) = \frac{f(\mathbf{x} | Y \in (a_{j-1}, a_j])\pi_j}{\sum_{j=1}^{p+1} f(\mathbf{x} | Y \in (a_{j-1}, a_j])\pi_j},\ j = 1, \ldots, p+1. \tag{6.4}$$

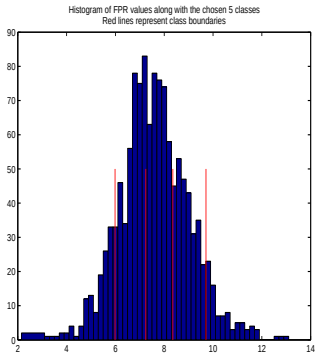


FIG 17. *Histogram of FPR values classified into 5 classes*

Here $f(\mathbf{x}|Y \in (a_{j-1}, a_j])$ represents the conditional shape density for the class $Y^{-1}(a_{j-1}, a_j]$. We estimate that by the Kernal density estimate described in Section 2.4,

$$\hat{f}_{\sigma j}(\mathbf{x}) = \frac{1}{n_j} \sum_{j:Y_j \in (\hat{a}_{j-1}, \hat{a}_j]} \frac{e^{-\frac{1}{2}\frac{d_g^2(X_j, \mathbf{x})}{\sigma^2}}}{\int_{\Sigma_2^{41}} e^{-\frac{1}{2}\frac{d_g^2(y, \mathbf{x})}{\sigma^2}} V(dy)}, \ j = 1, \ldots, p+1 \tag{6.5}$$

for appropriately chosen σ. Here $X_1, \ldots, X_n$ is the iid shape sample and n_j is the the number of shapes in the class $C_j = \{X_i : Y_i \in (\hat{a}_{j-1}, \hat{a}_j]\}$. Then we estimate the posterior probability, $\varpi_j(\mathbf{x})$ by

$$\hat{\varpi}_j(\mathbf{x}) = \frac{\hat{f}_{\sigma j}(\mathbf{x})\hat{\pi}_j}{\sum_{j=1}^{p+1} \hat{f}_{\sigma j}(\mathbf{x})\hat{\pi}_j}, \ j = 1, \ldots, p+1. \tag{6.6}$$

Here $\hat{\pi}_j$ is the proportion of Y_j's in $(\hat{a}_{j-1}, \hat{a}_j]$. For our sample, they are

$$\hat{\pi}_1 = 0.11, \ \hat{\pi}_2 = 0.30, \hat{\pi}_3 = 0.31, \ \hat{\pi}_4 = 0.21, \ \hat{\pi}_5 = 0.07.$$

Table 3 shows the posterior probabilities $\hat{\varpi}_j$, $j = 1, 2, \ldots, 5$ for a few shapes when I take $\sigma = 0.07$. The first 10 placentas in the table are the ones with shapes closest to the (intrinsic) mean shape. The next 10 are the ones with shape distance in the middle, and the last 10 have shapes furthest from the mean. Note how the FPR distribution changes for the 3 shape groups. The first group placenta shapes seem to have the most homogeneous conditional FPR distribution, while for the last 10, the distribution seems to be the least homogeneous. In the table, Placenta 1806 with shape distance 0.64 belongs to class 2 , i.e. the one with FPR in (6.0, 7.2], with probability 1. That seems to be consistent with Figure 15 because placenta 1806 has an FPR value of 7.17. However placenta 3244 with shape distance 0.65 belongs to class 1, i.e. has FPR less than 6 with probability 1 which does not seem to be consistent with Figure 15, because this placenta has a FPR of 10.07 (class 5). This discrepency may be justified, if we note that our probability estimates depend on the entire shape and not just on the shape distance from the mean.

TABLE 3
Posterior probabilities ($\hat{\varpi}$) for a few placentas

Placenta Id	**Geodesic Distance from mean shape**	$\hat{\varpi}_1$	$\hat{\varpi}_2$	$\hat{\varpi}_3$	$\hat{\varpi}_4$	$\hat{\varpi}_5$
2946	0.0577	0.1015	0.2887	0.3239	0.2216	0.0642
2163	0.0596	0.0960	0.2983	0.3257	0.2166	0.0634
2919	0.0636	0.1061	0.2878	0.3202	0.2214	0.0645
2639	0.0637	0.0966	0.2974	0.3265	0.2139	0.0657
2830	0.0653	0.0869	0.3062	0.3125	0.2319	0.0625
3363	0.0678	0.0891	0.3020	0.3281	0.2189	0.0619
2561	0.0682	0.1059	0.2973	0.3259	0.2131	0.0578
3062	0.0746	0.0973	0.2989	0.3140	0.2116	0.0782
2645	0.0750	0.0966	0.3107	0.3203	0.2140	0.0584
2957	0.0750	0.0954	0.3036	0.3185	0.2088	0.0737
3303	0.2300	0.1469	0.2413	0.3114	0.1747	0.1257
2630	0.2306	0.1006	0.2788	0.3117	0.2288	0.0801
2620	0.2308	0.1268	0.4238	0.2330	0.1767	0.0397
2007	0.2309	0.0655	0.3128	0.2580	0.2772	0.0865
3396	0.2314	0.1476	0.2720	0.2931	0.1640	0.1233
2938	0.2318	0.4820	0.1625	0.1839	0.1035	0.0681
1730	0.2321	0.1166	0.3423	0.2932	0.1660	0.0818
2788	0.2321	0.0714	0.4759	0.1959	0.2054	0.0515
2648	0.2325	0.0302	0.1373	0.1312	0.6827	0.0186
2126	0.2326	0.1119	0.3083	0.2540	0.2491	0.0765
2732	0.5915	0.0001	0.9423	0.0462	0.0011	0.0103
1976	0.5929	0.0000	0.9993	0.0000	0.0006	0.0000
1762	0.6073	0.0000	1.0000	0.0000	0.0000	0.0000
1832	0.6105	0.0000	0.0155	0.0053	0.0019	0.9773
2776	0.6338	0.0148	0.9807	0.0004	0.0002	0.0040
1806	0.6410	0.0000	1.0000	0.0000	0.0000	0.0000
3244	0.6510	1.0000	0.0000	0.0000	0.0000	0.0000
2107	0.6525	0.0000	0.0014	0.0187	0.9798	0.0001
3061	0.6531	0.0000	0.0006	0.0035	0.0115	0.9844
1921	0.7419	0.0000	0.0000	0.0000	0.9999	0.0000

Further Work

A lot more needs to be done in this area. One may carry out two sample tests described in Section 2 to discriminate between placenta shapes of the two sex. Also we may carry out the regression of FPR on shape, described in Section 6 separately for the two sex and may get very different results.

Another important analysis will be to use placenta size and shape information to predict new born features. A measure of placenta size can be the placenta weight. Using size and shape information together, we may get much stronger models explaining FPR.

Also one may consider the flat norm signature of the annulus like inter perimeter region and use that to describe placenta shape. For the concept of flat norm and the details on how to get it, I refer to Morgan and Vixie [7].

To get more information on shape, one may consider the shape of 3-D configurations from the whole placentas. For recent results on statistical analysis of 3-D shapes, I refer to Dryden et al.[5].

Finally, to measure placenta shape more accurately, one may also include the position of the blood vessels, nerve endings etc in the shape. These features can tell us a lot about the new born.

Acknowledgements

I am very thankful to my mentors, especially Professors Simon P. Morgan, Robert S. Sarracino and Kevin R. Vixie for their kind and helpful suggestions in preparing this report. I am also thankful to the Los Alamos National Labs for their support during the 2007 DDMA summer school which made this project possible.

References

[1] BHATTACHARYA, A. AND BHATTACHARYA, R. (2007). Statistics on Riemannian Manifolds: Asymptotic Distribution and Curvature. *To appear.*

[2] BHATTACHARYA, A. AND BHATTACHARYA, R. (2007). Nonparametric Statistics on Manifolds with Applications to Shape Spaces. *IMS Lecture Notes.* To appear.

[3] BHATTACHARYA, R. N. AND PATRANGENARU, V. (2003). Large sample methods of intrinsic and extrinsic sample means on manifolds-I. *Ann. Statist.* **31** 1-29.

[4] BHATTACHARYA, R. AND PATRANGENARU, V. (2005). Large sample methods of intrinsic and extrinsic sample means on manifolds-II. *Ann. Statist.* **33** 1225-1259.

[5] DRYDEN, I. L.; LE, H. AND WOOD, A. (2007). The MDS model for shape. *To appear.*

[6] KENDALL, W. S. (1990). Probability, convexity, and harmonic maps with small image-I. Uniqueness and the fine existence. Proc. *London Math. Soc.* **61** 371-406.

[7] MORGAN, S. P. AND VIXIE K. R. (2007). L^1TV computes flat norm for boundaries.