PARAMETRIC REGRESSION SCHEME FOR DISTRIBUTIONS: ANALYSIS OF DTI FIBER TRACT DIFFUSION CHANGES IN EARLY BRAIN DEVELOPMENT

Anuja Sharma[1] *P. Thomas Fletcher*[1] *John H. Gilmore*[2] *Maria L. Escolar*[3]
Aditya Gupta[2,3] *Martin Styner*[2] *Guido Gerig*[1]

[1] School of Computing, SCI Institute, University of Utah, Salt Lake City, UT, USA
[2] Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA
[3] Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA, USA

ABSTRACT

Temporal modeling frameworks often operate on scalar variables by summarizing data at initial stages as statistical summaries of the underlying distributions. For instance, DTI analysis often employs summary statistics, like mean, for regions of interest and properties along fiber tracts for population studies and hypothesis testing. This reduction via discarding of variability information may introduce significant errors which propagate through the procedures. We propose a novel framework which uses distribution-valued variables to retain and utilize the local variability information. Classic linear regression is adapted to employ these variables for model estimation. The increased stability and reliability of our proposed method when compared with regression using single-valued statistical summaries, is demonstrated in a validation experiment with synthetic data. Our driving application is the modeling of age-related changes along DTI white matter tracts. Results are shown for the spatiotemporal population trajectory of genu tract estimated from 45 healthy infants and compared with a Krabbe's patient.

Index Terms— linear regression, distribution-valued data, spatiotemporal growth trajectory, DTI, early neurodevelopment.

1. INTRODUCTION

Understanding normal development in healthy individuals, along with the natural population variability, is of critical clinical importance. It allows delineation of normative growth trends and a timely identification of individuals deviating from them. These growth curves may quantify evolutions of simple single-valued measures (e.g weight along age) or complex observations (e.g. distribution of the observed blood pressure range for each day in a month). Consequently, a significant portion of medical imaging research is focused on structural and morphological changes in the human brain along time [1, 2, 3, 4]. On the same grounds, our driving application is the characterization of DTI-derived diffusivity changes along white matter tracts in infant neurodevelopment. Tract based changes provide an insight into brain's maturation and are likely to correlate with characteristic cognitive functions. This makes them better suited for assessment related to specific cognitive abnormalities [3]. Previous work on DTI analysis includes parametric and non-parametric methods to model temporal changes in properties derived from spatial neighborhoods in 4D images [2, 3, 4]. These neighborhoods usually correspond to specific anatomical regions-of-interest (ROI), white matter tracts or tissue classes. However, the uncertainty associated with the measurement of the property is often disregarded at early stages by discarding the accompanying variability information observed in these neighborhoods. The analysis then proceeds with scalar statistical summaries representing the central tendencies of the observed distributions of these properties. Therefore, the associated variability information is neither utilized nor recovered for further analysis.

Working with summary statistics is attractive as it simplifies statistical analysis via data reduction. However, it should not introduce large errors at the initial stages of the framework. The choice of statistical summary can be reliable only if the underlying noise model is either known or can be safely assumed, and is homogeneously applicable to the complete data. This is often not the case, specially in images representing complex physiology, image acquisition methods prone to low SNR and low sample size situations. Fig.1 depicts how the choice of mean (red) versus median (green) to summarize and interpolate between two observed distributions of a random variable X can give significantly different results (considering that a linear fit between two points should theoretically have a single solution). Therefore, the reliability of the fit largely depends on the choice of the statistical measure used to summarize the spread of values. Also, the variability information is lost at all the intermediate interpolated points. These are the key issues we address in this paper.

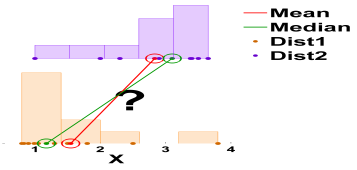

Fig. 1. Two synthetic distributions of a random variable X (dist1:orange, dist2:purple, solid dots:observed values) with the corresponding frequency histograms. Linear interpolation via classic summary statistics (mean(red) vs. median(green)) gives very different results. Data variability (depicted via circles) is discarded and hence lost at interpolated points along the fitted line.

We propose a framework which retains and utilizes the rich information embedded in the variability to estimate a linear regression trend. Unlike classic regression, all observations of the dependent and independent variables are distribution-valued with their density functions represented empirically via histograms [5]. This avoids early parametric assumptions regarding the underlying noise

Supported by NIH grants NA-MIC U54 EB005149, R01 HD055741, R01 MH070890, Twin R01 MH070890, Conte Center MH064065, BRP R01 NS055754, NINDS R01NS61965, DANA Foundation, NIH P30 HD03110 (UNC IDDRC), USTAR initiative at the University of Utah.

model for the random variables and the need for scalar summaries. Moreover, the method can statistically predict the expected value as well as the variance of the dependent variable. To the best of our knowledge, this is the first parametric tract-based DTI growth model to incorporate the complete spatial information by employing distribution-valued observations. In contrast to our previous nonparametric approach to the same application in [6], the current method allows the growth trajectory to be compactly parametrized by the regression coefficients (estimated as functions of arc length along the tract), making further statistical inference much simpler.

In the context of our driving application, scalar diffusion values like FA (Fractional Anisotropy) along 3D DTI tract geometries are available from healthy subjects scanned in age-groups centered around 1 month, 1 and 2 years. Our aim is to characterize a continuous spatiotemporal population trajectory of diffusivity changes from birth to 2 years. Linear regression is performed with FA distributions as the dependent variable and the age distributions within each age-group as the independent predictor variables. Section 2.1 explains the proposed model. However, prior to applying regression, two application specific sources of variability need to be accounted for. i) Measurement variability stemming from FA distributions along a tract's length in a single DTI scan (Fig.2). ii) The natural variability of subjects within an age-group (Fig.3a, 3b). Section 2.2 describes a kernel regression scheme and barycenter histogram estimation to incorporate both of these. Section 3 includes validation using synthetic data and results of application to pediatric DTI data.

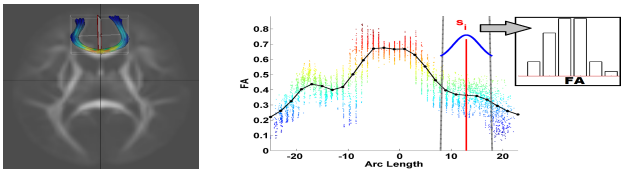

Fig. 2. Left: 3D visualization of the genu white matter tract from a single DTI scan. Right: Diffusion values (FA) assembled along genu tract's length (scatter points colored by FA) with the corresponding cross-sectional average FA curve (black) [6]. Kernel weights are applied to compute a histogram of the spatial FA distribution within a kernel window centered at arc-length location s_i along the tract.

2. METHOD

2.1. Linear regression for distribution-valued data

We begin by representing the probability distribution associated with a random variable Z as normalized histograms. Since histograms summarize data variability empirically, they do not require prior information of the underlying distribution [5, 6]. Formally,

$$h^l_Z(z) = (\pi^l, c^l, r^l), \tag{1}$$

where $l = [1 \cdots L]$ is the histogram bin index, L is the total number of bins and z is the variable which the histogram represents (for e.g. DTI diffusion property like FA). Each bin l is characterized by the bin frequency π^l, location of the bin's center c^l and half the bin width as bin radius r^l. r^l can vary across l allowing unequal bin widths.

Classic linear regression for simple scalar-valued variables X (independent) and Y (dependent) has the stochastic formulation $Y = \beta_0 + X\beta_1 + \epsilon$ where ϵ is the residual error defined as $\hat{\epsilon} = Y - \hat{Y}$ and β_1 (slope), β_0 (intercept) are the two unknown regression coefficients. For N paired observations of (X_n, Y_n), $(n = 0, \ldots N)$, the least squares estimate for the regression coefficients is obtained by minimizing the sum of squared residuals $\Sigma_{n=1}^N \hat{\epsilon}_n^2$ giving,

$$\hat{\beta}_1 = \frac{Cov(X,Y)}{Var(X)} \ , \ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \ . \tag{2}$$

This closed form solution holds under certain assumptions: $E(\epsilon_n) = 0$, homoscedasticity ($Var(\epsilon_n) = \sigma^2$), $Cov(\epsilon_n, \epsilon_m) = 0$ for $n \neq m$ and X measured without error (where $E()$ is the expectation).

To extend this regression model to work with distribution-valued observations, basic descriptive statistics used in eq.(2) are adapted appropriately [5]. Variables X and Y themselves are random variables now with each of their realizations being a probability distribution, instead of a single scalar value. Hence, the respective marginal distributions of each observed joint distribution can be represented as a paired histogram observation $(h^l_{X_n}(x), h^l_{Y_n}(y))$. The variability information of X_n and Y_n is now embedded in their histogram representations. Therefore, while keeping the assumptions and solution framework of classic regression the same, the definitions of mean, variance and covariance needed in eq.(2) can be adapted to deal with histogram observations [5]. Assuming each histogram bin has a uniform distribution, bin's interval boundaries and frequency are used to incorporate the variability within each distribution-valued observation as well as variability across observations. Using eq.(1), let the bin's interval boundaries be $a^l = c^l - r^l$, $b^l = c^l + r^l$. Then,

$$\bar{X} = \frac{1}{2N} \sum_{n=1}^{N} \sum_{l=1}^{L} \pi^l_{X_n} (a^l_{X_n} + b^l_{X_n}).$$

$$Var(X) = [\ \frac{1}{3N} \sum_{n=1}^{N} [\ \sum_{l=1}^{L} ((b^l_{X_n})^2 + (b^l_{X_n} a^l_{X_n}) + (a^l_{X_n})^2) \pi^l_{X_n}]\ ] \ - \ [\bar{X}]^2.$$

$$Cov(X,Y) = \frac{1}{3N} \sum_{n=1}^{N} \sum_{l_X=1}^{L_X} \sum_{l_Y=1}^{L_Y} \pi^{l_X}_{X_n} \pi^{l_Y}_{Y_n} G_{X_n} G_{Y_n} \sqrt{O_{X_n} O_{Y_n}},$$

$$O_{X_n} = (a^{l_X}_{X_n} - \bar{X})^2 + (a^{l_X}_{X_n} - \bar{X})(b^{l_X}_{X_n} - \bar{X}) + (b^{l_X}_{X_n} - \bar{X})^2,$$

$$G_{X_n} = -1 \text{ if} \bar{X}_n \leq \bar{X} \ , \ 1 \text{ if } \bar{X}_n > \bar{X}.$$

Similar equations can be derived for Y. (For $a^l = b^l$ and $\pi^l = 1$, they reduce to classic regression definitions). Now, by plugging these into eq.(2), estimates of the scalar regression coefficients are obtained. We can now statistically predict Y, given a new observation X_m. If X_m is a scalar value, $\hat{Y}_m$ is also scalar. If X_m is a distribution, then the complete distribution of $\hat{Y}_m$ can either be estimated via Monte Carlo simulations or the first two statistical moments of $\hat{Y}_m$ can be estimated by applying the rules of random variable algebra, $E(Y_m) = \beta_0 + E(X_m)\beta_1$, $Var(Y_m) = \beta_1^2 Var(X_m)$.

2.2. Setting up 4D DTI data for linear regression

Our paper is driven by a major limitation of the current ROI and fiber-tract based analyses which reduce local regional and tract properties to mean values used for group statistics, thus discarding data variability which is important for statistical testing and inference. Moreover, taking the mean assumes normal distribution and unimodality, which is not a proper model for FA and tract locations showing mixtures of fiber bundles. We now formulate 4D DTI tract data in the context of a distribution-valued regression problem.

Nonparametric kernel regression along DTI tracts: To obtain histogram descriptions of diffusion-value distributions, we use a kernel based weighting function within a moving kernel window along a

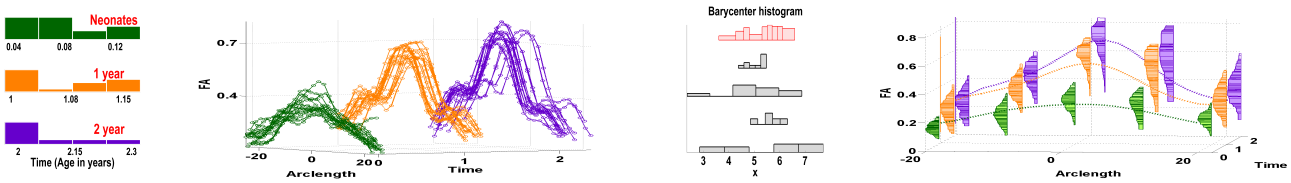

Fig. 3. Left to right: a) Distribution of ages of 15 healthy subjects within each age-group:neonate(green), 1 year(orange), 2 year(purple). b) Average FA curves from subjects within each age group with two inherent sources of variability highlighted. First, circles on each curve depict the local spatial FA distribution. Second, within each age-group, FA curves vary amongst subjects due to natural population variability. c) A barycenter histogram (topmost-red) calculated as an 'average' of the bottom four synthetic histograms. d) At each arc length location along the tract, FA histograms of all subjects within an age-group are averaged to create a barycenter histogram (only a few locations shown).

tract's length (Fig.2). Within a given kernel window, the unweighted bin frequencies π_Z^l of the diffusion histogram are weighted by a normalized Gaussian kernel with standard deviation σ, $K_\sigma(s, s_i) \propto exp[\frac{-(s-s_i)^2}{2\sigma^2}]$. The weights account for the inherent functional correlation of diffusion along tracts related to the fiber bundle geometry. These diffusion histograms are estimated at each tract location s_i where s is a continuous spatial variable and i indexes discrete arc-length parameterized locations along the tract's total length [6].

Barycenter histogram estimation: To account for the natural population variability within an age group, we create an 'average' of diffusion histograms from all subjects within that age-group, for each tract location s_i. (DTI scans from all subjects and age-groups are already co-registered and spatially normalized prior to this step [6]). For this, we use the concept of a barycenter histogram which minimizes the Mallow's distance metric between itself and other histograms [6] and reflects translation, changing width and shape of the individual histograms (Fig.3c). Mallow's distance is the L2 norm of the difference between u-quantiles of two distributions. This minimization problem can be conveniently expressed in terms of the bin centers and radii of the participating histograms. For more details, refer to [6]. Kernel regression followed by barycenter histogram estimation creates barycenter FA histograms along the tract's length for each age-group in the population (Fig.3d). These can now be regressed on the corresponding age-distributions (Fig.3a).

3. EXPERIMENTS

3.1. Validation using synthetic data

A linear regression experiment is conducted using two synthetic samples of bivariate distributions of random variables X and Y (Fig.4). To simulate the scenario of an unknown underlying joint distribution, we instead generate two observed marginal distributions for each: ($X_1 \sim \mathcal{N}(1, 0.3)$, $X_2 \sim \mathcal{N}(2, 0.3)$, $Y_1 \sim \mathcal{N}(3, 0.7)$, $Y_2 \sim \mathcal{N}(5, 0.7)$). Each marginal distribution has ten observations creating 20 pairs representing the two bivariate distribution samples ((x_{1i}, y_{1i}):orange, (x_{2i}, y_{2i}):purple, $i = 1, \ldots, 10$). A linear regression fit is estimated between the two distribution-valued samples using four methods. a) All Observations (Black): Uses all 20 scalar observation pairs (x_{ji}, y_{ji}), j=[1,2] to fit a line without applying any statistical summarization at early stages. In the absence of known underlying noise models, this scenario is as close to 'true' as may be possible. b) Distribution-valued (Blue): Uses the proposed framework for regression on histograms of marginal distributions as summary statistics. c) Mean (Red): Uses means of the two distribution-valued samples as a scalar summary. d) Median (Green): Uses medians as a summary. RMSE is calculated with respect to all observations (x_{ji}, y_{ji}) to allow comparison between different methods. Fig.4 shows that due to high amount of noise and small sample size per joint distribution, mean and median fits don't agree with each other and have higher RMSE values. The estimated correlation coefficient is a perfect 1, because regression is now exposed to two scalar points only owing to an early discarding of variability information. Our proposed method's performance closely matches the all-observations' result. They both incorporate the variability and accordingly lower the correlation value.

However, since a single random experiment is inconclusive, we repeat it 10,000 times (Fig.5). The linear fits with means and medians have a much larger variability with the fit swinging wildly with some extremely off results. In contrast, despite summarizing data as marginal histograms, our method performs comparable to 'All-Observations', exhibiting robustness to noise and low sample size. The spread of regression coefficients (β_0, β_1) is also considerably smaller indicating smaller standard errors (even when extreme outliers with mean/median methods have been omitted from the plot).

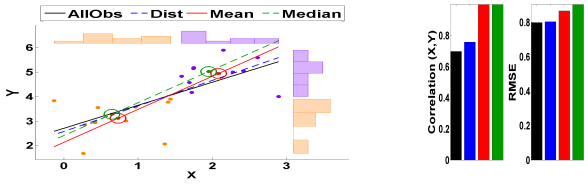

Fig. 4. Linear regression between two synthetically generated samples (orange, purple) of a bivariate distribution of random variables X and Y. Line color:: a)Black- Uses all observations (solid dots) for a linear fit. b)Blue- Proposed distribution-valued regression using marginal distributions (shown as histograms). c)Red- Using means as scalar summary statistics and d)Green- Using medians. Corresponding estimates of Correlation(X,Y) and RMSE also shown.

3.2. Application to pediatric DTI data

We apply our method to pediatric DTI data (FA values for the genu tract (Fig.2) from 45 healthy infants registered to a common atlas space. Subjects' ages at the time of scan fall into age-distributions grouped as neonates, 1 year and 2 years (15 subjects per age-group). These age distributions are the independent variables used for regression (Fig.3a). The dependent variables are the corresponding barycenter FA histograms estimated along the length of the genu tract (Fig.3d, Section 2.2). Regression is performed at each tract location, with the three pairs of age-histograms and location-specific FA barycenter histograms (with one pair per age-group). This provides a growth trajectory for FA continuously in space and time,

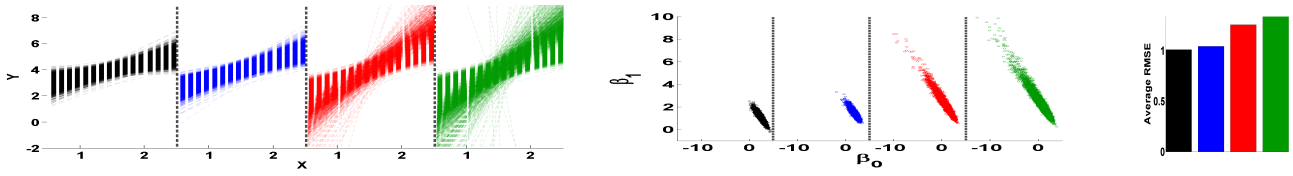

Fig. 5. 10,000 repetitions of synthetic regression experiment (Fig.4). Regression results: All Observations (black), Distribution-valued (blue), mean (red), median (green). Left to right: a) Linear fits obtained from 10,000 repetitions. b) Estimated regression coefficients β_1:slope, β_0:intercept (extreme outlier coefficients from mean/median regression omitted in the plot). c) Average of 10,000 RMSE estimates.

parametrized by regression coefficients estimated at each tract location. Note that most of the barycenters (Fig.3d) do not follow a Gaussian trend, thereby highlighting the need for our framework to avoid early, uninformed and possibly inaccurate assumptions.

Fig.6 shows the estimated normative growth surface (mean FA). Standard deviation bounds (also estimated for each point on the surface) are also shown for the three age-groups where DTI scans were originally available (neonate, 1 and 2 years). Overall, FA increases over time as a result of the expected brain maturation in early neurodevelopment which agrees with clinically observed patterns [2, 6]. The surface also exhibits localized temporal trends along the length of the genu tract. To further exhibit the clinical relevance of this framework, we compare the normative trend with a Krabbe's patient (with serial DTI scans at 14 days, 6 months and 1 year mapped into the common atlas space). Krabbe's disease is degenerative in nature affecting the myelin of the nervous system and is often fatal without early diagnosis and intervention in infants. Comparison at 14 days and 1 year is straightforward as control data is already available at these timepoints (Fig.3). For comparison at 6 months, we generate a gaussian age distribution for 6 months $\pm$ 2 weeks and estimate the corresponding FA mean and variance. Fig.6 shows the comparison of normative estimates with the mean FA curves of the Krabbe's subject at matched timepoints. The FA values from the latter are close to the expected temporal trend till 6 months but then show significant reduction towards 1 year. This scenario highlights the clinical need to enable comparison of individual patients with population models at timepoints where the control data is not available. In such cases, our method allows straightforward statistical predictions of FA mean and variance continuously along space and time.

4. DISCUSSION AND CONCLUSION

The synthetic experiments validate the increased stability of our method when compared with regression based on scalar measures like mean and median. In the case of well-behaved white noise with high SNR, they exhibit similar performances. At the same time, our method increases reliability by avoiding possible errors introduced early in the process due to parametric assumptions and over-simplified summary statistics. Moreover, despite using a compact summary measure (histogram representations of data distributions), the performance of our method closely agrees with the results achieved by using all observations which serve as a benchmark in the absence of ground truth. This is an advantage because even if all observations are available to the framework, it is often necessary to reduce data size when working with complex, high dimensional data (e.g. 4D DTI tract data) in order to make downstream statistical analysis more manageable. For DTI analysis, the method allows comparison at timepoints where no original scans were available by providing a smooth spatiotemporal trajectory. It also inherently accounts for naturally accelerated or delayed growth trends by providing mean and variance estimates in a given age window (e.g. 6 months $\pm$ 2 weeks). To conclude, the proposed method uses distribution-valued measurements in a closed-form, linear regression model leading to improved robustness. Observed distributions are summarized as histograms of marginal distributions instead of classic scalar statistical summaries like mean and median. Future work would involve an evaluation of the method's bias-variance trade-off, improvement in sensitivity to identify statistically significant group differences and application to large clinical studies.

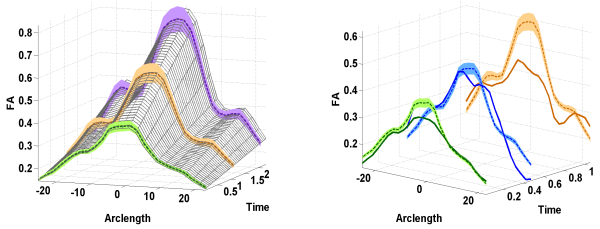

Fig. 6. Left: Continuous spatiotemporal normative FA evolution (genu tract). Estimated mean FA: gray, standard deviation bounds: neonate (green), 1 year (orange), 2 year (purple). Right: Krabbe's subject's average FA curves (solid lines) at 14 days, 6 month (blue), 1 year. Corresponding population mean (dashed line) and standard deviation shown for matched age-groups (blue: 6 month $\pm$ 2 weeks).

5. REFERENCES

[1] M. Kuklisova-Murgasova, P. Aljabar, L. Srinivasan, et al., "A dynamic 4d probabilistic atlas of the developing brain," *NeuroImage*, vol. 54, no. 4, pp. 2750–2763, 2011.

[2] Y. Yuan, J. Gilmore, H. Zhu, et al., "A longitudinal functional analysis framework for analysis of white matter tract statistics," in *IPMI*. Springer, 2013, pp. 220–231.

[3] J. Dubois, L. Hertz-Pannier, D. Le Bihan, et al., "Assessment of the early organization and maturation of infants' cerebral white matter fiber bundles: a feasibility study using quantitative diffusion tensor imaging and tractography," *Neuroimage*, vol. 30, no. 4, pp. 1121–1132, 2006.

[4] S. Smith, M. Jenkinson, P. Matthews, et al., "Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data," *Neuroimage*, vol. 31, no. 4, pp. 1487–1505, 2006.

[5] L. Billard and E. Diday, *Smybolic Data Analysis: Conceptual Statistics and Data Mining*, John Wiley, 2006.

[6] A. Sharma, P.T. Fletcher, G. Gerig, et al., "Spatiotemporal modeling of distribution-valued data applied to dti tract evolution in infant neurodevelopment," in *ISBI*, 2013, pp. 684–687.