POLITECNICO DI TORINO
Repository ISTITUZIONALE

Bayesian size-and-shape regression modelling

Original
Bayesian size-and-shape regression modelling / Di Noia, A., Mastrantonio, G., Jona Lasinio, G.. - In: STATISTICS & PROBABILITY LETTERS. - ISSN 0167-7152. - 204:(2024), pp. 1-4. [10.1016/j.spl.2023.109928]

Availability:
This version is available at: 11583/2981541 since: 2023-09-02T17:53:13Z

Publisher:
Elsevier

Published
DOI:10.1016/j.spl.2023.109928

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Elsevier preprint/submitted version

Preprint (submitted version) of an article published in STATISTICS & PROBABILITY LETTERS © 2024, http://doi.org/10.1016/j.spl.2023.109928

(Article begins on next page)

13 June 2026

Bayesian Size-and-Shape regression modelling

Antonio Di Noia[a,b], Gianluca Mastrantonio[c], Giovanna Jona Lasinio[d]

[a]*Seminar for Statistics, Department of Mathematics, ETH Zurich, Rämistrasse 101, Zurich, 8092, Switzerland*
[b]*Faculty of Economics, Università della Svizzera italiana, Via Giuseppe Buffi 13, Lugano, 6900, Switzerland*
[c]*Department of Mathematical Sciences, Polytechnic of Turin, Corso Duca degli Abruzzi 24, Turin, 10129, Italy*
[d]*Department of Statistical Sciences, Sapienza University, P.le Aldo Moro 5, Rome, 00185, Italy*

Abstract

Building on Dryden et al. (2021), this note presents the Bayesian estimation of a regression model for size-and-shape response variables with Gaussian landmarks. Our proposal fits into the framework of Bayesian latent variable models and, potentially, allows for a highly flexible modelling framework.

Keywords: Bayesian models, Size-and-Shape, Latent variables, Statistics on manifolds

1. Introduction

Shape data naturally arise in various research fields, whenever our attention is on the shape of objects. For example, in biology there is interest in modelling the shape features of organisms and understanding their relationship with environmental conditions. Even magnetic resonance data can be interpreted as containing size-and-shape information (Mardia et al., 2013). Image analysis, computer vision, and bioinformatics are other fields where shape data analysis finds natural applications (Anderson, 1997, Green and Mardia, 2006). In genetics, electrophoretic gel images contain shape information, while chemistry employs three-dimensional coordinates to assess the geometric structure of molecules (Horgan et al., 1992; Czogiel et al., 2011).

Recent advances in this area focus on developing methodologies to capture and explain changes in the shape of objects. For example, Kenobi et al. (2010) proposed smoothing methods, while, on the other hand, regression methodologies for shape and size-and-shape data have been explored in a limited number of papers. For instance, Gutiérrez et al. (2019) propose a Bayesian approach to shape data, starting with a Gaussian distribution on the configuration space and then accounting for location, rotation, and scale effects using a projected normal distribution. Dryden et al. (2021) tackle size-and-shape data, developing their proposal within the likelihood-based framework. An earlier paper (Dryden et al., 2019) proposes the Bayesian estimation of a regression model

where an additional parameter for the rotation matrix is introduced. However, this approach limitation is that it depends on defining a prior distribution on the rotation matrix, making the inference reliant on the rotation information of the original data. On the contrary, for size-and-shape modelling, the inference should be rotation-independent. Additionally, landmarks are assumed independent.

This paper, along with Dryden et al. (2021), treats the rotation as a latent variable, eliminating the need to specify a prior distribution. Additionally, the model considers dependent landmarks. As a result, this study presents an alternative Bayesian regression methodology tailored for size-and-shape response data. The proposed methodology ensures estimation stability through a simple and efficient MCMC (Markov Chain Monte Carlo) algorithm. The model has been implemented in the Julia (Bezanson et al., 2017) package *BayesSizeAndShape* (Mastrantonio and Jona Lasinio, 2023).

2. A Bayesian model for size-and-shape

Let $\tilde{\mathbf{X}}_i \in \mathbb{R}^{(k+1)\times p}$, with $i = 1, \ldots, n$ and $k \geq p$, be a collection of random configuration matrices. In statistical shape analysis, the configuration matrix contains the Euclidean coordinates of $k+1$ landmarks ("a landmark is a point of correspondence on each object that matches between and within populations", Dryden and Mardia (2016)) in a p-dimensional space, where p is typically 2 or 3. To perform size-and-shape inference, it is essential to eliminate information about location and orientation from every configuration matrix $\tilde{\mathbf{X}}_i$. To remove location information, we use the Helmert submatrix $\mathbf{H}$ (see Dryden and Mardia, 2016), where $\mathbf{H}$ has dimension $k \times (k+1)$, and, setting $d_j = 1/\sqrt{j(j+1)}$, its j-th row has the first j elements equal to $-d_j$, element $j+1$ is equal to jd_j, and the remaining $k-j$ elements are equal to zero. By pre-multiplying the i-th configuration matrix with $\mathbf{H}$, we obtain the Helmertized configuration $\tilde{\mathbf{X}}_i^H := \mathbf{H}\tilde{\mathbf{X}}_i$. Following the approach in Dryden et al. (2021), we can decompose each $\tilde{\mathbf{X}}_i^H$ using the singular value decomposition, which results in $\tilde{\mathbf{X}}_i^H = \mathbf{U}_i\boldsymbol{\Delta}_i\tilde{\mathbf{R}}_i^\top$, where $\tilde{\mathbf{R}}_i \in \mathrm{SO}(p)$, and $\mathrm{SO}(p)$ represents the $p \times p$-dimensional Special Orthogonal group. The matrix $\tilde{\mathbf{R}}_i$ contains all the information about the i-th object's orientation. On the other hand, $\mathbf{Y}_i := \mathbf{U}_i\boldsymbol{\Delta}_i$ represents the size-and-shape version of the original configuration $\tilde{\mathbf{X}}_i$, which is the actual focus of the inference. By assuming $\tilde{\mathbf{R}}_i \in \mathrm{SO}(p)$, we preserve reflection information that would otherwise be lost if $\tilde{\mathbf{R}}_i$ belongs to the space of the $p \times p$-dimensional Orthogonal group, denoted as $\mathrm{O}(p)$.

2.1. The model

In our scenario, each $\mathbf{Y}_i$ is coupled with a vector of d covariates, $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{id})^\top$, and we are interested in understanding the relationship between $\mathbf{z}_i$ and $\mathbf{Y}_i$. However, the size-and-shape space, where $\mathbf{Y}_i$ takes values, is a non-Euclidean manifold with a complex geometric structure, and directly specifying a probability distribution for $\mathbf{Y}_i$ is not a simple task. We adopt the

idea proposed in Dryden et al. (2021) to address these challenges, extending it to a Bayesian setting. The approach involves introducing a latent variable $\mathbf{R}_i \in \mathrm{SO}(p)$ and defining a regressive-type relationship between $\mathbf{X}_i := \mathbf{Y}_i\mathbf{R}_i^\top$ and $\mathbf{z}_i$ as follows:

$$\mathrm{vec}(\mathbf{X}_i) \sim \mathcal{N}_{kp}\Big(\mathrm{vec}\Big(\sum_{h=1}^{d} z_{ih}\mathbf{B}_h\Big), \mathbf{I}_p \otimes \mathbf{\Sigma}\Big),\ i = 1, \ldots, n, \tag{1}$$

with $\mathbf{X}_i \perp \mathbf{X}_{i'}$ if $i \neq i'$. In the above expression, $\mathrm{vec}(\cdot)$ denotes the vectorization of a matrix, $\mathbf{\Sigma}$ is a non-singular $k \times k$ covariance matrix and $\mathbf{B}_h$, $h = 1, \ldots, d$, is a $k \times p$ matrix of regressive coefficients. Remark that $\mathbf{R}_i$ is latent, and it must not be confused with $\tilde{\mathbf{R}}_i$, which is the rotation matrix of the original data. The latent variable approach avoids dependence on the rotation and leads to a proper inference depending only on $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. To describe our Bayesian model, we must introduce the prior distributions for the model's unknown parameters. To simplify notation and the derivation of the full conditionals (i.e., the distribution of one parameter given all the others and the data) we use $\mathbf{X}_{i,l}$ and $\mathbf{B}_{h,l}$ to indicate the $l-$th column of $\mathbf{X}_i$ and $\mathbf{B}_h$, respectively, we assume $\boldsymbol{\beta}_l := (\mathbf{B}_{1,l}^\top, \mathbf{B}_{2,l}^\top, \ldots, \mathbf{B}_{d,l}^\top)^\top$, $\boldsymbol{\beta} := (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_p^\top)^\top$, and we introduce the design matrix $\mathbf{Z}_i := \mathbf{I}_k \otimes \mathbf{z_i}^\top$. Since the l-th column of $\mathrm{vec}(\sum_{h=1}^{d} z_{ih}\mathbf{B}_h)$ is equal to $\mathbf{Z}_i\boldsymbol{\beta}_l$, we define the Bayesian model as

$$\begin{aligned} \mathbf{X}_{i,l}|\boldsymbol{\beta}, \mathbf{\Sigma} &\sim \mathcal{N}_k(\mathbf{Z}_i\boldsymbol{\beta}_l, \mathbf{\Sigma}), \quad i = 1, \ldots, n, \quad l = 1, \ldots, p, \\ \boldsymbol{\beta}_l &\sim N_{kd}(\mathbf{M}_l, \mathbf{V}_l), \quad \mathbf{\Sigma} \sim IW(\nu, \mathbf{\Psi}), \end{aligned} \tag{2}$$

with $\mathbf{X}_{i,l}|\boldsymbol{\beta}, \mathbf{\Sigma} \perp \mathbf{X}_{i',l'}|\boldsymbol{\beta}, \mathbf{\Sigma}$ if $i \neq i'$ or $l \neq l'$ and IW denotes the Inverse Wishart distribution. It should be noted that, once we condition on $(\boldsymbol{\beta}, \mathbf{\Sigma})$, models (1) and (2) induce the same distribution on the configuration space.
Remark that an identification problem arises from the model specification, which has not been highlighted in Dryden et al. (2021). To make it evident, we show that the sets of parameters $\{\mathbf{B}_1, \ldots, \mathbf{B}_d, \mathbf{\Sigma}\}$ and $\{\mathbf{B}_1\mathbf{\Lambda}, \ldots, \mathbf{B}_d\mathbf{\Lambda}, \mathbf{\Sigma}\}$, where $\mathbf{\Lambda} \in \mathrm{SO}(p)$ is a rotation matrix, induce the same probability density function over $(\mathbf{U}_1, \Delta_1, \ldots, \mathbf{U}_n, \Delta_n)$. Let $\boldsymbol{\mu}_i := \sum_{j=1}^{d} z_{ij}\mathbf{B}_j$ and f denotes a probability density function, we need to prove that

$$\frac{f(\mathbf{U}_i, \Delta_i; \boldsymbol{\mu}_i, \mathbf{\Sigma})}{f(\mathbf{U}_i, \Delta_i; \boldsymbol{\mu}_i\mathbf{\Lambda}, \mathbf{\Sigma})} = \exp\Big(-\frac{\mathrm{tr}(\boldsymbol{\mu}_i^\top\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i) - \mathrm{tr}(\mathbf{\Lambda}^\top\boldsymbol{\mu}_i^\top\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i\mathbf{\Lambda})}{2}\Big) = 1, \tag{3}$$

for all $i = 1, \ldots, n$, where the joint density of $(\mathbf{U}_i, \Delta_i)$ is derived by Dryden et al. (2021) in Theorem 1. From the properties of the trace operator, and since $\mathbf{\Lambda}^\top = \mathbf{\Lambda}^{-1}$, we have that $\mathrm{tr}(\mathbf{\Lambda}^\top\boldsymbol{\mu}_i^\top\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i\mathbf{\Lambda}) = \mathrm{tr}(\boldsymbol{\mu}_i^\top\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i)$, which shows that (3) holds true for any $i = 1, \ldots, n$. For this reason, an identification constraint is needed to prevent any arbitrary rotation of $\boldsymbol{\mu}_i$. This can be achieved by assuming that for one of the $\mathbf{B}_h$, e.g., the first, we have that

$$[\mathbf{B}_1]_{wl} = 0, \ \ l > w, \quad [\mathbf{B}_1]_{ll} \geq 0, \ \ l = 1, \ldots, p-1. \tag{4}$$

These are a version of the transformations proposed by Dryden et al. (2021) to identify and isolate the size-and-shape information of the mean configuration. To impose these constraints effectively, two different approaches can be used. The first approach involves modifying the prior distributions, altering the parameter space to adhere to the identification constraints. The second approach relies on the MCMC algorithm to explore the posterior freely, and after obtaining each posterior sample $\mathbf{B}_h^b$, where b indicates the b-th sample, a remapping step is performed. This remapping transforms each $\mathbf{B}_h^b$ to an identified version $\tilde{\mathbf{B}}_h^b$ using the map $\mathbf{B}_h^b \mapsto \tilde{\mathbf{B}}_h^b := \mathbf{B}_h^b \boldsymbol{\Lambda}^b$. Here, $\boldsymbol{\Lambda}^b = g(\mathbf{B}_h^b) \in \mathrm{SO}(p)$ represents an appropriate rotation matrix defined by a function $g : \mathbb{R}^{k \times p} \to \mathrm{SO}(p)$ such that $\tilde{\mathbf{B}}_h^b$ satisfies the desired identification constraints (4). We employ the second approach with a Gram-Schmidt construction to define $\boldsymbol{\Lambda}^b$. This method allows for a more straightforward MCMC algorithm, while respecting the constraints.

2.2. The Markov chain Monte Carlo algorithm

To implement the MCMC algorithm, we need to derive the full conditional distributions of $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_n$. Owing to the model specification given in (2), we can easily see that the full conditional of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ (indicated, respectively, as $\boldsymbol{\beta}_l| \ldots$ and $\boldsymbol{\Sigma}| \ldots$) are the same that we would obtain in the case of a standard Bayesian regression, i.e.,

$$\boldsymbol{\beta}_l| \cdots \sim N_{kd}(\mathbf{M}_l^*, \mathbf{V}_l^*), \qquad \boldsymbol{\Sigma}| \cdots \sim IW(\nu^*, \boldsymbol{\Psi}^*),$$

with $\nu^* = \nu + np$, $\boldsymbol{\Psi}^* = \boldsymbol{\Psi} + \sum_{i=1}^n \sum_{l=1}^p (\mathbf{X}_{i,l} - \mathbf{Z}_i \boldsymbol{\beta}_l)(\mathbf{X}_{i,l} - \mathbf{Z}_i \boldsymbol{\beta}_l)^\top$, $\mathbf{M}_l^* = \mathbf{V}_l^* (\sum_{i=1}^n \mathbf{Z}_i^\top \boldsymbol{\Sigma}^{-1} \mathbf{X}_{i,l} + \mathbf{V}_l^{-1} \mathbf{M}_l)$, and $\mathbf{V}_l^* = (\sum_{i=1}^n \mathbf{Z}_i^\top \boldsymbol{\Sigma}^{-1} \mathbf{Z}_i + \mathbf{V}_l^{-1})^{-1}$. To derive the full conditional distribution of $\mathbf{R}_i$, we can refer to the computation presented in Dryden et al. (2021). According to Theorem 1, the distribution of $\mathbf{R}_i$ has density proportional to $\exp(\mathrm{tr}(\mathbf{R}_i \mathbf{A}_i^\top))$ where $\mathbf{A}_i = \boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}^{-1} \mathbf{Y}_i$, i.e. a Matrix Fisher distribution with parameter $\mathbf{A}_i$ (Mardia and Jupp, 2000). Sampling directly from a Matrix Fisher distribution is typically not straightforward, and ad hoc techniques are required; some proposals can be found in Kent et al. (2013) and Hoff (2009). Here, we handle the case $p = 2$ by expressing the rotation matrix as a function of the rotation angle $\theta_i \in [0, 2\pi)$. It can be shown that the distribution of θ_i is von-Mises, since its density is proportional to $\exp(\kappa_i \cos(\theta_i - \eta_i))$, and its parameters (η_i, κ_i) can be derived by setting the equation $\mathrm{tr}(\mathbf{R}_i \mathbf{A}_i^\top) = \kappa_i \cos(\theta_i - \eta_i)$. When $p = 3$, we express $\mathbf{R}_i$ as function of the Euler angles $\theta_{i,1} \in [0, 2\pi)$, $\theta_{i,2} \in [0, \pi)$, $\theta_{i,3} \in [0, 2\pi)$, by representing $\mathbf{R}_i$ as the product of elementary rotations. Then, samples from the Matrix Fisher are obtained using the approach proposed by Green and Mardia (2006).

3. Simulation experiments

The simulation experiment has been designed to assess the model's capacity to recover the underlying parameters used to generate the data. It covers a wide range of settings, providing a comprehensive evaluation of the model's performance under diverse conditions. We generate the data from equation

(2), assuming that $\boldsymbol{\Sigma}$ is equal to $\lambda\boldsymbol{\Sigma}^*$ and λ is a parameter we change in the simulations, with $d = 3$. Variable $z_{i,1} = 1, \forall\, i = 1, \ldots, n$, defines the intercepts, a continuous covariate $z_{i,2}$ is simulated from a normal distribution with mean 10 and standard deviation 1, and a categorical variable $z_{i,3}$ with two levels is included in a corner-point representation. Each $z_{i,3}$ has an equal probability of assuming one of the two levels. Parameter $\boldsymbol{\Sigma}$ is simulated from an $IW(k+2, 5\mathbf{I}_k)$, and the regressive coefficients are generated from a $N(5, 1)$. To comply with the identifiability constraint in (4), coefficients are transformed to their identifiable version by multiplication of each $\mathbf{B}_h$ with an appropriate rotation matrix $\boldsymbol{\Lambda}$. We generate 100 datasets for each of 16 different settings, considering all possible combinations of $k \in \{10, 20\}$, $n \in \{20, 100\}$, $\lambda \in \{1, 10\}$, and $p \in \{2, 3\}$. The total number of regressors to be estimated, taking into consideration the identifiability constraints, are $kdp-1$ if $p = 2$ and $kdp-3$ if $p = 3$, which ranges from a minimum of 59 to a maximum of 177, while the number of elements in $\boldsymbol{\Sigma}$ is $(k^2 + k)/2$, which is equal to 55 if $k = 10$ and 210 if $k = 20$. We assume $\mathbf{M}_l = \mathbf{0}_{kd}$, $\mathbf{V}_l = 10^4\mathbf{I}_{kd}$, $\nu = k + 2$, and $\boldsymbol{\Psi} = \mathbf{I}_k$. We run the algorithm for 90000 iterations, discarding the first 30000 iterations as burn-in, and keep every 30th iteration, retaining 2000 samples for inferential purposes. The model is implemented in Julia 1.8.2.

For each parameter, we evaluate the percentage of times that the 95% credible interval (CI) contains the value used to simulate the data, and the total length of the CI. Summary tables are shown in the online supplementary material. The proportion of correctly estimated parameters (parameters inside the associated CI) within the simulated datasets ranges from 0.917 to 0.974 under all settings. The CI lengths increase as k, p, and (especially) λ increase. On the other hand, interval lengths decrease as n increases, as expected.

4. Further developments

In this paper we presented the size-and-shape regression model within a Bayesian latent variable framework, addressing the identifiability issues. Our work opens new possibilities for modelling size-and-shape data, particularly in formalizing complex dependence structures which might be handled within the Bayesian framework. Future research will explore dependence among landmarks using lattice-type modelling approaches. We also aim to account for temporal dependence to model the shape evolution over time (Fontanella et al., 2018). Recognizing the limitations of linearity, we plan to explore more flexible functional relationships between covariates and mean configuration.

Acknowledgement

The work of the second author was partially carried out within the FAIR - Future Artificial Intelligence Research and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 –

D.D. 1555 11/10/2022, PE00000013). This manuscript reflects only the author's view and opinion, neither the European Union nor the European Commission can be considered responsible for them.

References

Anderson, C.R., 1997. Object recognition using statistical shape analysis. PhD thesis, University of Leeds.

Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B., 2017. Julia: A fresh approach to numerical computing. SIAM Review 59, 65–98.

Czogiel, I., Dryden, I.L., Brignell, C.J., 2011. Bayesian matching of unlabeled marked point sets using random fields, with an application to molecular alignment. Annals of Applied Statistics 5, 2603–2629.

Dryden, I.L., Kim, K.R., Le, H., 2019. Bayesian linear size-and-shape regression with applications to face data. Sankhya A 81, 83–103.

Dryden, I.L., Kume, A., Paine, P.J., Wood, A.T.A., 2021. Regression modeling for size-and-shape data based on a gaussian model for landmarks. Journal of the American Statistical Association 116, 1011–1022.

Dryden, I.L., Mardia, K., 2016. Statistical Shape Analysis: With Applications in R. Wiley Series in Probability and Statistics. 2nd ed., John Wiley & Sons, Chichester.

Fontanella, L., Ippoliti, L., Kume, A., 2018. The Offset Normal Shape Distribution for Dynamic Shape Analysis. Journal of Computational and Graphical Statistics .

Green, P.J., Mardia, K.V., 2006. Bayesian alignment using hierarchical models, with applications in protein bioinformatics. Biometrika 93, 235–254.

Gutiérrez, L., Gutiérrez-Peña, E., Mena, R.H., 2019. A Bayesian Approach to Statistical Shape Analysis via the Projected Normal Distribution. Bayesian Analysis 14, 427 – 447.

Hoff, P.D., 2009. Simulation of the matrix bingham–von mises–fisher distribution, with applications to multivariate and relational data. Journal of Computational and Graphical Statistics 18, 438–456.

Horgan, G.W., Creasey, A., Fenton, B., 1992. Superimposing two-dimensional gels to study genetic variation in malaria parasites. Electrophoresis 13, 871–875.

Kenobi, K., Dryden, I.L., Le, H., 2010. Shape Curves and Geodesic Modelling. Biometrika 97(3), 567–584.

Kent, J.T., Ganeiber, A.M., Mardia, K.V., 2013. A new method to simulate the Bingham and related distributions in directional data analysis with applications. arXiv preprint arXiv:1310.8110v1 .

Mardia, K., Jupp, P., 2000. Directional Statistics. 1st ed., John Wiley & Sons, New York.

Mardia, K.V., Bookstein, F.L., Kent, J.T., 2013. Alcohol, babies and the death penalty: Saving lives by analysing the shape of the brain. Significance 10(3), 12–16.

Mastrantonio, G., Jona Lasinio, G., 2023. BayesSizeAndShape : a Julia package for Bayesian estimation of size and shape data regression models, in: Proceedings of the GRASPA 2023 Conference.