Regression analysis of elliptically symmetric directional data

Zehao Yu, Xianzheng Huang [ID],*

Department of Statistics, University of South Carolina, Columbia, 29201, SC, United States

ARTICLE INFO

ABSTRACT

A comprehensive toolkit is developed for regression analysis of directional data based on a flexible class of angular Gaussian distributions. Informative testing procedures to assess rotational symmetry around the mean direction, and the dependence of model parameters on covariates are proposed. Bootstrap-based algorithms are provided to assess the significance of the proposed test statistics. Moreover, a prediction region that achieves the smallest volume in a class of ellipsoidal prediction regions of the same coverage probability is constructed. The efficacy of these inference procedures is demonstrated in simulation experiments. Finally, this new toolkit is used to analyze directional data originating from a hydrology study and a bioinformatics application.

1. Introduction

Directional data naturally arise in many scientific disciplines, such as flight directions of migrating birds, the directions of wind and waves in the ocean, and geomagnetic field directions. These examples of directional data as the original form of observed data are typically of low dimensions. High dimensional directional data typically result from preprocessing high dimensional features collected in genetic study (Banerjee et al., 2005), computer vision (Ryali et al., 2013), and text analysis (Ennajari et al., 2021), among many other fields of study. In these instances, the raw data vectors in some d-dimensional Euclidean space $\mathbb{R}^d$ are often normalized to lie on a hypersphere $\mathbb{S}^{d-1} = \{\mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}\| = 1\}$, where $\|\mathbf{y}\|$ denotes the Euclidean norm of $\mathbf{y}$.

Regression analysis of directional data is relatively underdeveloped compared to regression analysis of response data in the linear (Euclidean) space. One of the most notable early developments of regression models for directional data is given by Johnson and Wehrly (1978), who formulated parametric models for the joint distribution of a circular response (i.e., $d = 2$) and a linear covariate. Later, Presnell et al. (1998) introduced the spherically projected multivariate linear model based on the projected Gaussian distribution for the circular response with a mean direction depending on covariates linearly. Mimicking the least squares method in regression analysis for a linear response, Lund (1999) proposed a least circular-distance method for regression analysis of a circular response. Scealy and Wood (2019) proposed a transformation of the von Mises-Fisher distribution to study paleomagnetic data, following which they built regression models using the proposed directional distribution. Paine et al. (2018) proposed the elliptically symmetric angular Gaussian distribution (ESAG), focusing on directional data on $\mathbb{S}^2$. In a follow-up work (Paine et al., 2020), the authors formulated regression models based on ESAG of low dimensions.

The formulation of ESAG results from imposing constraints on the mean $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{V}$ of a multivariate Gaussian distribution $\mathcal{N}_d(\boldsymbol{\mu}, \mathbf{V})$ to resolve the identifiability issue. Such an identifiability issue emerges inevitably when normalizing a multivariate Gaussian vector to yield an angular Gaussian random variable, since two Gaussian vectors, $\mathbf{W}$ and $c\mathbf{W}$, are normalized to the same vector when $c > 0$, yet they follow different Gaussian distributions whenever $c \neq 1$. For most angular Gaussian distributions,

---

constraints are imposed on $\mathbf{V}$ that often translate to stringent assumptions on the resultant directional distribution. The constraints on $\boldsymbol{\mu}$ and $\mathbf{V}$ that lead to ESAG give rise to a probability density function (pdf) that does not involve a complicated normalization constant, and the resultant distribution remains flexible in that it is not limited to isotropic distributions. These are virtues of ESAG that make it stand out among many existing named directional distributions. Section 2 provides a brief review of ESAG and its parameterization that facilitates regression analysis of directional data. We then address three inference problems that are theoretically and practically important in the context of directional distributions and regression analysis. More specifically, Section 3 presents a novel diagnostic test for isotropy around the mean. This is equivalent to testing if a distribution supported on a hypersphere is rotationally symmetric around its mean direction. Isotropy is similar to symmetry for a distribution supported on a linear space, and thus testing isotropy is of interest from the perspective of distribution theory like testing the symmetry of a distribution. In Section 4, we propose methods for testing covariate dependence of ESAG model parameters. Section 5 reports simulation experiments for studying the operating characteristics of the proposed testing procedures. Section 6 provides prediction regions of the directional response. We apply these new inference procedures to two real-life applications in Section 7. Section 8 recapitulates the contributions of our study and points out some limitations of the proposed regression framework that motivate follow-up research.

2. The ESAG regression model and likelihood-based inference

2.1. The model and data

A d-dimensional random variable $\mathbf{Y}$ supported on $\mathbb{S}^{d-1}$ follows an angular Gaussian distribution, $\mathrm{AG}(\boldsymbol{\mu}, \mathbf{V})$, if $\mathbf{Y} = \mathbf{W}/\|\mathbf{W}\|$ with $\mathbf{W} \sim \mathcal{N}_d(\boldsymbol{\mu}, \mathbf{V})$. To guarantee the identifiability of the distribution $\mathrm{AG}(\boldsymbol{\mu}, \mathbf{V})$, constraints on $(\boldsymbol{\mu}, \mathbf{V})$ are needed to avoid overparameterization. For example, Presnell et al. (1998) assumed $\mathbf{V} = \mathbf{I}_d$ that leads to an isotropic directional distribution, where $\mathbf{I}_d$ is the d-dimensional identity matrix. Less stringent assumptions are also considered, for example, in Wang and Gelfand (2013) where a sub-block of $\mathbf{V}$ is assumed known. We adopt the ESAG distribution (Paine et al., 2018) resulting from imposing the following constraints that we refer to as ESAG constraints henceforth, $\mathbf{V}\boldsymbol{\mu} = \boldsymbol{\mu}$ and $\det(\mathbf{V}) = 1$, where $\det(\mathbf{V})$ dentoes the determinant of $\mathbf{V}$. These constraints leave more room for flexible modeling of $\mathbf{Y}$ than most previously considered constraints at the price of creating a more complex constrained parameter space. The pdf of the ESAG distribution is

$$f(\mathbf{Y}|\boldsymbol{\mu}, \mathbf{V}) = \frac{(2\pi)^{-(d-1)/2}}{(\mathbf{Y}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{Y})^{d/2}} \exp\left[\frac{1}{2}\left\{\frac{(\mathbf{Y}^{\mathsf{T}}\boldsymbol{\mu})^2}{\mathbf{Y}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{Y}} - \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{\mu}\right\}\right] M_{d-1}\left\{\frac{\mathbf{Y}^{\mathsf{T}}\boldsymbol{\mu}}{\sqrt{\mathbf{Y}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{Y}}}\right\}, \tag{1}$$

where $M_{d-1}(t) = (2\pi)^{-1/2} \int_0^\infty x^{d-1} \exp\{-(x-t)^2/2\} dx$.

We recently reparameterized ESAG by introducing constraint-free parameters $\boldsymbol{\gamma} \in \mathbb{R}^{(d-2)(d+1)/2}$ so that $\mathbf{V}$ that satisfies ESAG constraints can be determined by $(\boldsymbol{\mu}, \boldsymbol{\gamma})$ via an eigendecomposition (Yu and Huang, 2024). This new parameterization is suitable for directional data on $\mathbb{S}^{d-1}$ for an arbitrary $d \geq 3$, the range of dimensions we focus on in this study. Appendix A the Supplementary Material details this parameterization when $d = 4$. Henceforth, we use $\mathbf{Y} \sim \mathrm{ESAG}(\boldsymbol{\mu}, \boldsymbol{\gamma})$ to refer to $\mathbf{Y} \sim \mathrm{AG}(\boldsymbol{\mu}, \mathbf{V})$ with ESAG constraints imposed on $(\boldsymbol{\mu}, \mathbf{V})$.

The benefits of modeling ESAG via constraint-free parameters are at least twofold. First, maximum likelihood estimation of model parameters becomes more straightforward than directly estimating $(\boldsymbol{\mu}, \mathbf{V})$ subject to the nonlinear ESAG constraints, such as the constraint of $\det(\mathbf{V}) = 1$. Second, a covariate-dependent ESAG can be easily formulated without introducing link functions to relate covariates to constrained model parameters as done in earlier regression models for directional responses (Lund, 1999; Scealy and Welsh, 2011, 2017). In this study, we consider an ESAG regression model specified by $\mathbf{Y}|\mathbf{X} \sim \mathrm{ESAG}(\boldsymbol{\mu} = \boldsymbol{\alpha}_0 + \mathbf{A}_1\mathbf{X}, \boldsymbol{\gamma} = \boldsymbol{\beta}_0 + \mathbf{B}_1\mathbf{X})$, where $\mathbf{X} = (X_1, ..., X_q)^{\mathsf{T}}$ is the q-dimensional covariate vector, $\boldsymbol{\alpha}_0$ is the intercept for modeling $\boldsymbol{\mu}$, $\mathbf{A}_1 = [\boldsymbol{\alpha}_1 \mid ... \mid \boldsymbol{\alpha}_q]$ is the $d \times q$ matrix of regression coefficients representing covariates effects on $\boldsymbol{\mu}$, $\boldsymbol{\beta}_0$ is the intercept parameter in $\boldsymbol{\gamma}$, and $\mathbf{B}_1 = [\boldsymbol{\beta}_1 \mid ... \mid \boldsymbol{\beta}_q]$ is the $(d-2)(d+1)/2 \times q$ matrix of covariates effects on $\boldsymbol{\gamma}$, in which $\boldsymbol{\alpha}_k \in \mathbb{R}^d$ and $\boldsymbol{\beta}_k \in \mathbb{R}^{(d-2)(d+1)/2}$, for $k = 0, 1, ..., q$. This regression model generalizes the one in Paine et al. (2020, Section 3.2) that focuses on the case with $d = 3$.

Suppose the observed data include directional responses $\{\mathbf{Y}_1, ..., \mathbf{Y}_n\}$ from n independent experimental units along with their covariates data $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$. Similar to the treatment on covariates data in Scealy and Wood (2019), we standardize covariates data via $(X_{i,k} - X_{(1),k})/(X_{(n),k} - X_{(1),k}) + 1$, for $i = 1, ..., n$, where $X_{(1),k}$ and $X_{(n),k}$ are the minimum and maximum order statistics corresponding to covariate X_k, for $k = 1, ..., q$. The resultant standardized covariates data are more comparable in scale with the response of a unit Euclidean norm, which helps to stabilize the numerical implementation of maximum likelihood estimation without distorting the underlying association between the response and covariates. With a slight abuse of notation, we use $\{\mathbf{X}_i\}_{i=1}^n$ to refer to the standardized covariates data. These standardized covariates data fall in the q-dimensional unit cube $[1, 2]^q$. If a new experimental unit possesses a covariate value falling outside of the original data range, then this subject's standardized covariates data fall outside of the unit cube, which can be problematic when predicting the subject's outcome. To alleviate this concern in prediction, one may consider alternative standardization, such as the more traditional approach of centering and scaling the covariates data to achieve zero-mean and unit-variance. On the other hand, when predicting a directional response based on linear covariates, extrapolation can be even more unreliable than when the response is also linear. Hence, extrapolation using our regression models is especially discouraged, without which the current data standardization poses no complication in prediction.

2.2. Maximum likelihood estimation

To parameterize $\mathbf{V}$ in AG($\boldsymbol{\mu}$, $\mathbf{V}$) that satisfies ESAG constraints, we introduced longitude and latitude angle parameters to specify eigenvectors of $\mathbf{V}$ after $\boldsymbol{\mu}$ is specified (as explained in detail in Appendix A in the Supplementary Material). We showed that $\boldsymbol{\gamma}$ or a certain subvector of it being zero amounts to some latitude angles falling on the boundary of 0 or π and some other latitude and longitude angles being non-identifiable (see Appendix B in Yu and Huang, 2024, for details). This suggests violations of regularity conditions in the context of drawing likelihood-based inference for model parameters even though the parameter space of ESAG($\boldsymbol{\mu}$, $\boldsymbol{\gamma}$) is the entire real space $\mathbb{R}^{(d-1)(d+2)/2}$. The irregularity carries over to the ESAG regression model. As a result, maximum likelihood estimators (MLE) of some regression coefficients may converge in distribution to Gaussian at a slower rate than $\sqrt{n}$, or may not be asymptotically Gaussian, depending on where the true model parameters fall in the parameter space. Regardless, numerical implementation of maximum likelihood estimation is straightforward under the current parameterization of ESAG, as demonstrated in our earlier work (and thus omitted here), and a simple resample-based bootstrap procedure can be used to quantify the uncertainty of the MLEs.

When it comes to hypothesis testing, the conventional likelihood ratio test is inadequate when regularity conditions are not satisfied because the asymptotic null distribution of a likelihood ratio (LR) statistic is no longer a χ^2 (Chernoff, 1954). Most existing strategies for addressing this complication with LR aim at estimating the exact distribution of LR or its limiting distribution under the null using some simulation-based methods, such as the method proposed by Drton (2009) and the approach developed in Mitchell et al. (2019). Instead of using LR, we propose different test statistics that exploit unique properties of the ESAG distribution. These are elaborated in the next two sections, one focusing on testing for isotropy, and the other considering tests for covariate dependence of $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$.

3. Hypothesis testing for isotropy

If $\mathbf{Y}$ follows an isotropic distribution around its mean, then $\mathbf{RY}$ and $\mathbf{Y}$ are identically distributed for any given $d \times d$ rotation matrix $\mathbf{R}$ such that $\mathbf{R}\boldsymbol{\mu} = \boldsymbol{\mu}$. By the parameterization of $\mathbf{V}$ via $\boldsymbol{\gamma}$ after $\boldsymbol{\mu}$ is specified, ESAG($\boldsymbol{\mu}$, $\boldsymbol{\gamma}$) is isotropic when $\boldsymbol{\gamma} = \mathbf{0}$, which gives $\mathbf{V} = \mathbf{I}_d$. Hence, testing isotropy is relevant to inferring correlations of the components in $\mathbf{W}$, i.e., the pre-normalization version of $\mathbf{Y}$, and also relates to model selection between the more parsimonious isotropic ESAG and a generic ESAG distribution. Outside of the ESAG family, isotropy can be manifested in different forms other than having $\mathbf{V} = \mathbf{I}_d$. In what follows, we propose a strategy for testing the null hypothesis $H_0^{(\mathbf{V})} : \mathbf{Y} \sim \text{ESAG}(\boldsymbol{\mu}, \boldsymbol{\gamma} = \mathbf{0})$, where potential dependence of $\boldsymbol{\mu}$ on covariates $\mathbf{X}$ is suppressed for notational simplicity. The proposed strategy is motivated by properties of the MLE for the concentration parameter in the presence of model misspecification.

3.1. Concentration estimation

For the distribution ESAG($\boldsymbol{\mu}$, $\boldsymbol{\gamma}$), $\|\boldsymbol{\mu}\|$ quantifies the overall concentration of the distribution, with $\mathbf{V}$ (or $\boldsymbol{\gamma}$) controlling the variation in different subspaces on the unit sphere. A data cloud generated from an isotropic ESAG supported on $\mathbb{S}^{d-1}$ visually tends to be ball-shaped, whereas a data cloud from an anisotropic ESAG distribution takes the shape of an elliptical disc. Intuitively, when fitting an isotropic ESAG model to data from an anisotropic ESAG, one essentially tries to find a ball that can best resemble (in some sense) the elliptical disc. To accomplish this, the radius of the ball tends to be some weighted average of the axes of the elliptical disc, leading to a lower concentration of the fitted isotropic ESAG compared to the concentration of the true anisotropic distribution. In the context of model comparison, two ESAG distributions, ESAG($\boldsymbol{\mu}_1$, $\boldsymbol{\gamma}_1 = \mathbf{0}$) and ESAG($\boldsymbol{\mu}_2$, $\boldsymbol{\gamma}_2 \neq \mathbf{0}$), are more alike when $\|\boldsymbol{\mu}_1\| < \|\boldsymbol{\mu}_2\|$ than when $\|\boldsymbol{\mu}_1\| \geq \|\boldsymbol{\mu}_2\|$. We demonstrate this phenomenon next by exploiting the properties of MLEs in the presence of model misspecification.

Let P denote a generic ESAG distribution with pdf $P(\mathbf{Y}; \boldsymbol{\mu}_a, \boldsymbol{\gamma}_a)$, which specifies the true data-generating mechanism with true model parameters $\boldsymbol{\mu}_a$ and $\boldsymbol{\gamma}_a$. Let Q denote an isotropic ESAG distribution with pdf $Q(\mathbf{Y}; \boldsymbol{\mu})$. Using the density function in (1), we have $P(\mathbf{Y}; \boldsymbol{\mu}_a, \boldsymbol{\gamma}_a) = f(\mathbf{Y}|\boldsymbol{\mu}_a, \mathbf{V}(\boldsymbol{\gamma}_a))$ and $Q(\mathbf{Y}; \boldsymbol{\mu}) = f(\mathbf{Y}|\boldsymbol{\mu}, \mathbf{I}_d)$, where we use $\mathbf{V}(\boldsymbol{\gamma}_a)$ to highlight the dependence of $\mathbf{V}$ on $\boldsymbol{\gamma}_a$ after $\boldsymbol{\mu}_a$ is specified in the true pdf. The Kullback–Leibler divergence of Q from P is defined as $D_{\text{KL}}(P\|Q; \boldsymbol{\mu}) = E_P[\log\{P(\mathbf{Y}; \boldsymbol{\mu}_a, \boldsymbol{\gamma}_a)/Q(\mathbf{Y}; \boldsymbol{\mu})\}]$, where the subscript "$P$" signifies that the expectation is with respect to the distribution P. Under regularity conditions (White, 1982), if one fits the model Q to data from P, then the MLE for $\boldsymbol{\mu}$ converges in probability to $\boldsymbol{\mu}_0 = \text{argmin}_{\boldsymbol{\mu}} D_{\text{KL}}(P\|Q; \boldsymbol{\mu}) = \text{argmax}_{\boldsymbol{\mu}} E_P\{\log Q(\mathbf{Y}; \boldsymbol{\mu})\}$. We show next that $\|\boldsymbol{\mu}_0\| \leq \|\boldsymbol{\mu}_a\|$, or, equivalently, in the presence of model misspecification (i.e., $P \neq Q$), $E_P\{\log Q(\mathbf{Y}; \boldsymbol{\mu})\}$ is maximized when the ratio of concentrations (RoC) $\|\boldsymbol{\mu}_a\|/\|\boldsymbol{\mu}_0\|$ exceeds 1. To single out the concentration, we view $\boldsymbol{\mu}_a = c_a \mathbf{R}_a \boldsymbol{\mu}^*$ and $\boldsymbol{\mu}_0 = c_0 \mathbf{R}_0 \boldsymbol{\mu}^*$ for some rotation matrices, $\mathbf{R}_a$ and $\mathbf{R}_0$, and some positive constants, c_a and c_0, where $\boldsymbol{\mu}^*$ is a unit vector. In other words, $\boldsymbol{\mu}_a$ and $\boldsymbol{\mu}_0$ may differ in concentration, quantified by c_a and c_0 respectively, or differ in orientation when $\mathbf{R}_a \neq \mathbf{R}_0$. Using this factorization of the mean direction parameter, we have $\|\boldsymbol{\mu}_a\|/\|\boldsymbol{\mu}_0\| = c_a/c_0$ since $\|\mathbf{R}_a\boldsymbol{\mu}^*\|/\|\mathbf{R}_0\boldsymbol{\mu}^*\| = 1$. Now we re-express the density $P(\cdot; \boldsymbol{\mu}_a, \boldsymbol{\gamma}_a)$ as $P(\cdot; c_a, \mathbf{R}_a, \boldsymbol{\gamma}_a)$, and similarly write the density $Q(\cdot; \boldsymbol{\mu})$ as $Q(\cdot; c, \mathbf{R})$, where the dependence of both distributions on $\boldsymbol{\mu}^*$ is suppressed because they depend on the same $\boldsymbol{\mu}^*$. Viewing P as the reference distribution, we let $\boldsymbol{\mu}^* = \boldsymbol{\mu}_a/\|\boldsymbol{\mu}_a\|$, $\mathbf{R}_a = \mathbf{I}_d$, and $c_a = \|\boldsymbol{\mu}_a\|$. Fitting Q to data from P now amounts to, in limit as $n \to \infty$, maximizing $E_P\{\log Q(\mathbf{Y}; c, \mathbf{R})\}$ with respect to $(c, \mathbf{R})$, which cannot be done analytically but can be simulated using large samples.

To simulate this maximization problem, we generate a random sample of size $n = 10^4$ from $P(\cdot; c_a, \mathbf{R}_a, \boldsymbol{\gamma}_a)$ with $c_a = \sqrt{63}$ resulting from setting $\boldsymbol{\mu}_a = (2, -5, 3, 5)^{\text{T}}$, $\mathbf{R}_a = \mathbf{I}_4$, and $\boldsymbol{\gamma}_a$ taking one of the following three values, $\boldsymbol{\gamma}^{(1)} = \mathbf{0}$, $\boldsymbol{\gamma}^{(2)} = (1.5, 0, 1.5, 0, 0)^{\text{T}}$, and $\boldsymbol{\gamma}^{(3)} = 2\boldsymbol{\gamma}^{(2)}$, with the first value creating a scenario where $P = Q$, and the latter two creating increasing degree of anisotropy in P. We then use the log-likelihood function $\ell(c, \mathbf{R}) = n^{-1}\sum_{i=1}^{n} \log Q(\mathbf{Y}_i; c, \mathbf{R})$ as an empirical version of $E_P\{\log Q(\mathbf{Y}; c, \mathbf{R})\}$ to demonstrate
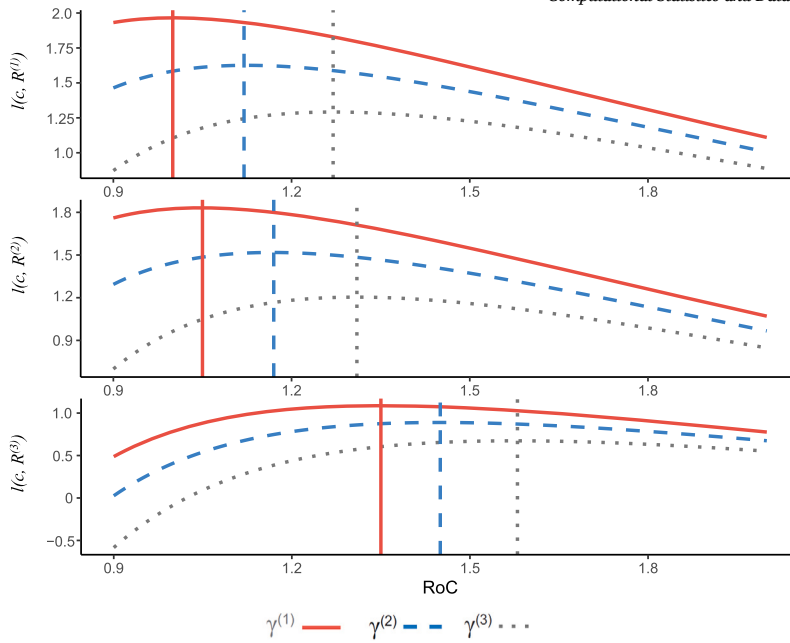
Fig. 1. The empirical version of $E_P\{\log Q(\mathbf{Y};c,\mathbf{R})\}$, $\ell(c,\mathbf{R})$, based on a random sample of size $n = 10^4$ from an isotropic ESAG (solid lines), anisotropic ESAG with $\gamma = \gamma^{(2)} \neq \mathbf{0}$ (dashed lines), and $\gamma = \gamma^{(3)} = 2\gamma^{(2)}$ (dotted lines), versus RoC when $\mathbf{R}$ is set at $\mathbf{R}^{(1)} = \mathbf{I}_d$ (top panel), $\mathbf{R}^{(2)} \neq \mathbf{I}_d$ (middle panel), and $\mathbf{R}^{(3)}$ that deviates from $\mathbf{I}_d$ even more (bottom panel). Vertical lines mark the value of RoC where the corresponding function $\ell(c,\mathbf{R})$ is maximized.

that $c_a/c^* > 1$ when $\gamma_a \neq \mathbf{0}$, where $c^* = \operatorname{argmax}_{c>0} \ell(c,\mathbf{R}^*)$ for some arbitrary rotation matrix $\mathbf{R}^*$. For concreteness, we consider three values for $\mathbf{R}^*$ given by $\mathbf{R}^{(1)} = \mathbf{R}_a$, $\mathbf{R}^{(2)}$ resulting from replacing the upper 2×2 block of $\mathbf{I}_4$ with the two-dimensional rotation matrix $R^*(0.1)$, and $\mathbf{R}^{(3)}$ similarly defined using $R^*(0.3)$ to produce a rotation matrix that deviates from $\mathbf{R}_a$ further than $\mathbf{R}^{(2)}$ does. A two-dimensional rotation matrix $R^*(\theta)$ has diagonal entries equal to $\cos(\theta)$, and $[1,2]$ and $[2,1]$ entries equal to $-\sin(\theta)$ and $\sin(\theta)$, respectively.

The top panel of Fig. 1 depicts $\ell(c,\mathbf{R}^{(1)})$ as a function of RoC $= c_a/c$ when the data-generating mechanism P has γ_a set at $\gamma^{(1)} = \mathbf{0}$ (isotropy), $\gamma^{(2)} \neq \mathbf{0}$ (mild anisotropy), and $\gamma^{(3)} = 2\gamma^{(2)}$ (severe anisotropy), respectively. With $\mathbf{R}^{(1)} = \mathbf{R}_a$, the mean directions of P and Q have the same orientation. When P is isotropic, $E_P\{\log Q(\mathbf{Y};c,\mathbf{R}^{(1)})\}$ is expected to be maximized at $c^* = c_a$, resulting in $D_{\mathrm{KL}}(P\|Q;\mu_0) = 0$. This is indeed (empirically) justified by the curve of $\ell(c;\mathbf{R}^{(1)})$ that reaches its peak at around RoC $= c_a/c^* = 1$. Once P exhibits anisotropy by having γ_a deviating from $\mathbf{0}$, one witnesses a drop in the likelihood $\ell(c,\mathbf{R}^{(1)})$, which is maximized at some RoC that exceeds 1, indicating that $c^* < c_a$. The inflation in RoC, i.e., the attenuation in c^*, becomes more substantial as γ_a deviates from $\mathbf{0}$ further. This implies that misspecifying γ in the ESAG distribution by assuming isotropy can be manifested in a larger-than-1 RoC. The bottom two panels in Fig. 1 show $\ell(c,\mathbf{R}^{(2)})$ and $\ell(c,\mathbf{R}^{(3)})$ versus RoC, where all the previously observed phenomena for $\ell(c,\mathbf{R}^{(1)})$ remain except for that, even with γ_a set at $\mathbf{0}$, $\ell(c,\mathbf{R}^*)$ is also maximized when RoC is larger than 1.

Comparing the three panels in Fig. 1 reveals a clear trend of RoC increasing as model misspecification becomes more severe by having γ_a further away from zero or having the orientation of μ_a mismatch more with the orientation of μ_0. The latter observation suggests that RoC can be used to test assumptions regarding μ as well, which is a point we come back to in a later section on testing assumptions on the mean direction parameter.

3.2. A test for isotropy

Inspired by the above findings regarding concentration estimation, we propose the statistic for testing isotropy defined by

$$\mathrm{RoC} = \frac{1}{n}\sum_{i=1}^{n}\frac{\|\hat{\mu}_{ai}\|}{\|\hat{\mu}_{0i}\|}, \tag{2}$$

where $\hat{\mu}_{0i}$ is the restricted MLE of the mean direction of $\mathbf{Y}_i$ given $\mathbf{X}_i$ under $H_0^{(\mathbf{V})}$ that assumes $\gamma = \mathbf{0}$ and covariate-dependent μ, and $\hat{\mu}_{ai}$ is the unrestricted MLE under the alternative hypothesis that allows anisotropy. If the true data-generating mechanism is consistent with $H_0^{(\mathbf{V})}$, then RoC is expected to be close to 1; otherwise, RoC tends to be larger than 1.

Algorithm 1 below gives the parametric bootstrap procedure to estimate the p-value associated with RoC to assess its statistical significance. The goal of the bootstrap procedure is to estimate the null distribution of RoC by simulating realizations of RoC under the null. To this end, we repeatedly compute RoC based on data generated from an isotropic ESAG distribution $\hat{Q}(\cdot;\hat{\mu}_{0i})$ for the i-th experimental unit, for $i = 1, \ldots, n$. This distribution is an estimate of Q that is closest to the unknown true model P for each experimental unit. An estimated p-value can then be obtained by comparing the RoC computed based on the raw data with the

simulated RoC's. As seen here and in the testing procedures proposed later for other purposes, the distribution of a test statistic under any hypothesized ESAG model can be easily approximated via parametric bootstrap because it is straightforward to simulate data from any ESAG distribution, which is yet another virtue of the ESAG distribution family and the constraint-free parameterization.

Algorithm 1 Hypothesis testing for isotropy based on RoC defined in (2).

1: **procedure** COMPUTE ROC BASED ON THE OBSERVED DATA
2:     Input data $\{(\mathbf{Y}_i, \mathbf{X}_i)\}_{i=1}^n$, find the restricted MLE $\hat{\boldsymbol{\mu}}_{0i}$ under $H_0^{(\mathbf{V})}$ and the unrestricted MLE $(\hat{\boldsymbol{\mu}}_{ai}, \hat{\boldsymbol{\gamma}}_{ai})$, for $i = 1, \ldots, n$.
3:     Compute the test statistic $\text{RoC} = (1/n) \sum_{i=1}^n \|\hat{\boldsymbol{\mu}}_{ai}\| / \|\hat{\boldsymbol{\mu}}_{0i}\|$.
4: **end procedure**
5: **procedure** BOOTSTRAP PROCEDURE TO ESTIMATE THE NULL DISTRIBUTION OF ROC
6:     Set B = number of bootstraps
7:     Initiate $s = 0$
8:     **for** b in $1, \ldots, B$ **do**
9:         Generate the b-th bootstrap sample $\{\mathbf{Y}_i^{(b)}\}_{i=1}^n$, where $\mathbf{Y}_i^{(b)}|\mathbf{X}_i \sim \text{ESAG}(\hat{\boldsymbol{\mu}}_{0i}, \boldsymbol{\gamma}_i = \mathbf{0})$, for $i = 1, \ldots, n$.
10:         Repeat Steps 2–3 using data $\{(\mathbf{Y}_i^{(b)}, \mathbf{X}_i)\}_{i=1}^n$. Denote the resultant value of RoC as $\text{RoC}^{(b)}$.
11:         **if** $\text{RoC}^{(b)} > \text{RoC}$ **then** $s = s + 1$
12:     **end for**
13:     Output s/B as an estimated p-value associated with RoC from Step 3.
14: **end procedure**

---

4. Tests for covariates effects

4.1. Testing covariates dependence of $\boldsymbol{\mu}$

For a directional response, a practically interesting question is whether or not its mean direction depends on covariates. For concreteness, let us consider testing the null $H_0^{(\boldsymbol{\mu})}$: $\mathbf{Y}|\mathbf{X} \sim \text{ESAG}(\boldsymbol{\mu} = \boldsymbol{\alpha}_0, \boldsymbol{\gamma} = \boldsymbol{\beta}_0 + \mathbf{B}_1\mathbf{X})$ versus the alternative H_1 : $\mathbf{Y}|\mathbf{X} \sim \text{ESAG}(\boldsymbol{\mu} = \boldsymbol{\alpha}_0 + \mathbf{A}_1\mathbf{X}, \boldsymbol{\gamma} = \boldsymbol{\beta}_0 + \mathbf{B}_1\mathbf{X})$. If the alternative is true with $\mathbf{A}_1 \neq \mathbf{0}$, the fitted $\boldsymbol{\mu}$ under the null is expected to differ from the fitted value that allows covariates dependence of $\boldsymbol{\mu}$. The difference can lie in their directions, or their norms, i.e., the concentrations of the two fitted distributions. This motivates the following test statistic that captures both sources of discrepancies,

$$D = \frac{1}{n} \sum_{i=1}^n \left(2 - \frac{\hat{\boldsymbol{\mu}}_{0i}^\top \hat{\boldsymbol{\mu}}_{ai}}{\|\hat{\boldsymbol{\mu}}_{0i}\| \|\hat{\boldsymbol{\mu}}_{ai}\|} \right) \frac{\|\hat{\boldsymbol{\mu}}_{ai}\|}{\|\hat{\boldsymbol{\mu}}_{0i}\|}, \tag{3}$$

where, for the i-th data point, $\hat{\boldsymbol{\mu}}_{0i}$ is the restricted MLE obtained under the null that assumes covariate-independent $\boldsymbol{\mu}$, and $\hat{\boldsymbol{\mu}}_{ai}$ is the unrestricted MLE obtained under the alternative. In (3), $\hat{\boldsymbol{\mu}}_{0i}^\top \hat{\boldsymbol{\mu}}_{ai}/(\|\hat{\boldsymbol{\mu}}_{0i}\| \|\hat{\boldsymbol{\mu}}_{ai}\|)$ is known as the cosine similarity between two vectors, $\hat{\boldsymbol{\mu}}_{0i}$ and $\hat{\boldsymbol{\mu}}_{ai}$, which is equal to 1 if they have the same direction, and is equal to -1 if the directions are opposite. Hence the first factor in the summand in (3) quantifies the dissimilarity in direction between $\hat{\boldsymbol{\mu}}_{0i}$ and $\hat{\boldsymbol{\mu}}_{ai}$. The second factor of the summand in (3) contrasts the concentrations of the two estimates for $\boldsymbol{\mu}$ as in RoC. By construction, under the null $H_0^{(\boldsymbol{\mu})}$, D is expected to be close to 1; and a realization of D larger than 1 can imply the observed data coming from a model that violates the null.

Algorithm 2 provides detailed steps for implementing the test based on the newly proposed test statistic, where we again use a parametric bootstrap procedure to estimate the p-value associated with D.

Algorithm 2 Hypothesis testing regarding $\boldsymbol{\mu}$ based on D defined in (3).

1: **procedure** COMPUTE D BASED ON THE OBSERVED DATA
2:     Input data $\{(\mathbf{Y}_i, \mathbf{X}_i)\}_{i=1}^n$, find the restricted MLEs $(\hat{\boldsymbol{\mu}}_{0i}, \hat{\boldsymbol{\gamma}}_{0i})$, and the unrestricted MLEs $(\hat{\boldsymbol{\mu}}_{1i}, \hat{\boldsymbol{\gamma}}_{1i})$, for $i = 1, \ldots, n$.
3:     Compute $D = (1/n) \sum_{i=1}^n [\{2 - (\hat{\boldsymbol{\mu}}_{0i}^\top \hat{\boldsymbol{\mu}}_{ia})/(\|\hat{\boldsymbol{\mu}}_{0i}\| \|\hat{\boldsymbol{\mu}}_{ai}\|)\}(\|\hat{\boldsymbol{\mu}}_{ai}\| / \|\hat{\boldsymbol{\mu}}_{0i}\|)]$.
4: **end procedure**
5: **procedure** BOOTSTRAP PROCEDURE TO ESTIMATE THE NULL DISTRIBUTION OF D
6:     Set B = number of bootstraps
7:     Initiate $s = 0$
8:     **for** b in $1, \ldots, B$ **do**
9:         Generate the b-th bootstrap sample $\{\mathbf{Y}_i^{(b)}\}_{i=1}^n$, where $\mathbf{Y}_i^{(b)}|\mathbf{X}_i \sim \text{ESAG}(\hat{\boldsymbol{\mu}}_{0i}, \hat{\boldsymbol{\gamma}}_{0i})$, for $i = 1, \ldots, n$.
10:         Repeat Steps 2–3 using data $\{(\mathbf{Y}_i^{(b)}, \mathbf{X}_i)\}_{i=1}^n$. Denote the resultant test statistic as $D^{(b)}$.
11:         **if** $D^{(b)} > D$ **then** $s = s + 1$
12:     **end for**
13:     Output s/B as an estimated p-value associated with D from Step 3.
14: **end procedure**

As indicated in Section 3.1, RoC can be used to test hypotheses about $\boldsymbol{\mu}$, such as testing covariate dependence of it by adopting Algorithm 1 with the restricted MLEs for $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ obtained under the current null $H_0^{(\boldsymbol{\mu})}$. Moreover, because D incorporates direction comparison between two fitted values of $\boldsymbol{\mu}$ besides concentration comparison that RoC focuses on, one can combine the two test statistics to gain more insight into the underlying data-generating mechanism. If D is significantly larger than RoC when testing covariate dependence of $\boldsymbol{\mu}$, one may interpret it as data evidence for the direction of $\boldsymbol{\mu}$ depending on some covariate. Having D close

to RoC can imply that the direction of $\boldsymbol{\mu}$ may not depend on covariates, although its norm may depend on covariates. This exemplifies the versatility and additional insight our proposed test statistics can offer compared to LR.

4.2. Testing covariates dependence of $\boldsymbol{\gamma}$ and beyond

Unique to our parameterization of $\text{ESAG}(\boldsymbol{\mu}, \boldsymbol{\gamma})$, parameters in $\boldsymbol{\gamma}$ control variation of the distribution in different subspaces on the hypersphere besides (an)isotropy. It is thus of interest to test if such distributional features depend on covariates. For instance, one may consider testing the null $H_0^{(\boldsymbol{\gamma})} : \mathbf{Y}|\mathbf{X} \sim \text{ESAG}(\boldsymbol{\mu} = \boldsymbol{\alpha}_0 + \mathbf{A}_1\mathbf{X}, \boldsymbol{\gamma} = \boldsymbol{\beta}_0)$ versus the alternative $H_1 : \mathbf{Y}|\mathbf{X} \sim \text{ESAG}(\boldsymbol{\mu} = \boldsymbol{\alpha}_0 + \mathbf{A}_1\mathbf{X}, \boldsymbol{\gamma} = \boldsymbol{\beta}_0 + \mathbf{B}_1\mathbf{X})$. Because $\boldsymbol{\gamma}$ as a whole relates to (an)isotropy of the distribution, RoC that is initially proposed for testing isotropy has its natural appeal for testing hypotheses about $\boldsymbol{\gamma}$. When a violation of $H_0^{(\boldsymbol{\gamma})}$ adversely affects inferences for $\boldsymbol{\mu}$, the test statistic D designed for testing assumptions on $\boldsymbol{\mu}$ also has the potential to detect covariates dependence of $\boldsymbol{\gamma}$. With the restricted MLEs $\hat{\boldsymbol{\mu}}_{0i}$ and $\hat{\boldsymbol{\gamma}}_{0i}$ now reflecting $H_0^{(\boldsymbol{\gamma})}$ used in Algorithm 1 or Algorithm 2, one can carry out the test based on RoC or D for testing $H_0^{(\boldsymbol{\gamma})}$ versus H_1.

Fixing H_1 at the above saturated ESAG model, to test other null hypotheses, say, $\boldsymbol{\alpha}_k = \mathbf{0}$ for a given $k \in \{1, \ldots, q\}$ (with all other covariates in the null model), RoC and D can be used with the restricted MLEs in Algorithms 1 and 2 revised accordingly when computing $\hat{\boldsymbol{\mu}}_{0i}$ to reflect the specific null hypothesis under consideration. Similarly, if one considers a different alternative ESAG model that allows $\boldsymbol{\mu}$ or $\boldsymbol{\gamma}$ to depend on covariates nonlinearly, the tests based on RoC and D remain applicable with $\hat{\boldsymbol{\mu}}_{ai}$ obtained under such alternative model. We thus argue that the proposed testing strategies are more versatile than the strategies of testing covariate dependence of $\boldsymbol{\mu}$ or $\boldsymbol{\gamma}$ based on the magnitude of regression-coefficient matrices $\mathbf{A}_1$ or $\mathbf{B}_1$. Even if one adopts an angular Gaussian distribution that is not ESAG, as long as the mean vector $\boldsymbol{\mu}$ has the same interpretations as that in $\text{ESAG}(\boldsymbol{\mu}, \boldsymbol{\gamma})$, RoC and D remain meaningful statistics for testing assumptions on $\boldsymbol{\mu}$ or other model assumptions that inferences for $\boldsymbol{\mu}$ are sensitive to. One simply needs to revise the bootstrap procedures to adapt to the assumed angular Gaussian distribution.

Lastly, RoC and D depend on both the restricted and unrestricted MLEs of model parameters, which in turn add to the computational burden in Algorithms 1 and 2 where these MLEs are obtained based on each bootstrap sample. We thus propose yet another testing strategy that only requires computing the restricted MLEs that is based on a second moment estimation, with the test statistic given by

$$M = \left\| \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbf{Y}_i^2 - \widehat{\mathbf{E}_0\left(\mathbf{Y}_i^2\right)} \right\} \right\|, \tag{4}$$

where $\mathbf{Y}_i^2$ is the element-wise quantity square of $\mathbf{Y}_i$, and $\widehat{\mathbf{E}_0\left(\mathbf{Y}_i^2\right)}$ is an empirical mean of $\mathbf{Y}^2$ given $\mathbf{X} = \mathbf{X}_i$ computed using a random sample simulated from an estimated null model. Unlike RoC and D, the construction of M is not motivated by (and thus does not target at testing) a particular aspect of the model specification; instead, M can serve as an overall goodness-of-fit test statistic. In fact, $\mathbf{Y}^2$ can be viewed as a compositional vector in a simplex, of which (non-negative) entries sum to one, and hence M can be interpreted as a prediction error of the compositions under a null model. This interpretation also reveals that lots of information of $\mathbf{Y}$ regarding orientations is lost in M (by taking the element-wise quantity square), and thus M is insensitive to model misspecification that impacts inferences on such information. Regardless, in the absence of model misspecification, M is expected to be close to zero, and a larger M serves as data evidence of a worse fit of a null model for the observed data. As an example, Algorithm 3 below gives the algorithm for using M to test the null model that assumes an isotropic ESAG, with an estimated p-value obtained via parametric bootstrap as an output.

Algorithm 3 Hypothesis testing for isotropy based on M defined in (4).

1: **procedure** COMPUTE M BASED ON THE OBSERVED DATA
2:     Input data $\{(\mathbf{Y}_i, \mathbf{X}_i)\}_{i=1}^n$, find the restricted MLE $\hat{\boldsymbol{\mu}}_{0i}$, for $i = 1, \ldots, n$.
3:     For $i = 1, \ldots, n$, generate $\{\tilde{\mathbf{Y}}_{i,m}\}_{m=1}^{10^4}$ from $\text{ESAG}(\hat{\boldsymbol{\mu}}_{0i}, \boldsymbol{\gamma}_i = \mathbf{0})$, compute $\widehat{\mathbf{E}_0(\mathbf{Y}_i^2)} = 10^{-4} \sum_{m=1}^{10^4} \tilde{\mathbf{Y}}_{i,m}^2$.
4:     Compute $M = \|(1/n) \sum_{i=1}^n \{\mathbf{Y}_i^2 - \widehat{\mathbf{E}_0(\mathbf{Y}_i^2)}\}\|$.
5: **end procedure**
6: **procedure** BOOTSTRAP PROCEDURE TO ESTIMATE THE NULL DISTRIBUTION OF M
7:     Set B = number of bootstraps
8:     Initiate $s = 0$
9:     **for** b in $1, \ldots, B$ **do**
10:         Generate the b-th bootstrap sample $\{\mathbf{Y}_i^{(b)}\}_{i=1}^n$, where $\mathbf{Y}_i^{(b)}|\mathbf{X}_i \sim \text{ESAG}(\boldsymbol{\mu}_i = \hat{\boldsymbol{\mu}}_{0i}, \boldsymbol{\gamma}_i = \mathbf{0})$, for $i = 1, \ldots, n$.
11:         Repeat Steps 2–4 using data $\{(\mathbf{Y}_i^{(b)}, \mathbf{X}_i)\}_{i=1}^n$. Denote the resultant test statistic as $M^{(b)}$.
12:         **if** $M^{(b)} > M$ **then** $s = s + 1$
13:     **end for**
14:     Output s/B as an estimated p-value associated with M from Step 4.
15: **end procedure**

Table 1

Data-generating mechanisms (DGM) designed for testing the power of tests for each considered null hypothesis regarding ESAG(μ, γ), along with values of model parameters in these DGMs.

Null hypothesis	ESAG data-generating mechanism
$H_0^{(\mathbf{V})}: \; \boldsymbol{\mu} = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 X, \; \boldsymbol{\gamma} = \mathbf{0}$	$\mathrm{DGM}_1^{(\mathbf{V})}: \; \boldsymbol{\mu} = \boldsymbol{\alpha}_0^* + \boldsymbol{\alpha}_1^* X, \; \boldsymbol{\gamma} = \boldsymbol{\beta}_{0r}^*$
	$\mathrm{DGM}_2^{(\mathbf{V})}: \; \boldsymbol{\mu} = \boldsymbol{\alpha}_0^* + \boldsymbol{\alpha}_1^* X, \; \boldsymbol{\gamma} = \boldsymbol{\beta}_{0r}^* + \boldsymbol{\beta}_{1r}^* X$
$H_0^{(\boldsymbol{\mu})}: \; \boldsymbol{\mu} = \boldsymbol{\alpha}_0, \; \boldsymbol{\gamma} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 X$	$\mathrm{DGM}_1^{(\boldsymbol{\mu})}: \; \boldsymbol{\mu} = \boldsymbol{\alpha}_0^* + \boldsymbol{\alpha}_{1r}^* X, \; \boldsymbol{\gamma} = \mathbf{0}$
	$\mathrm{DGM}_2^{(\boldsymbol{\mu})}: \; \boldsymbol{\mu} = \boldsymbol{\alpha}_0^* + \boldsymbol{\alpha}_{1r}^* X, \; \boldsymbol{\gamma} = \boldsymbol{\beta}_0^*$
	$\mathrm{DGM}_3^{(\boldsymbol{\mu})}: \; \boldsymbol{\mu} = \boldsymbol{\alpha}_0^* + \boldsymbol{\alpha}_{1r}^* X, \; \boldsymbol{\gamma} = \boldsymbol{\beta}_0^* + \boldsymbol{\beta}_1^* X$
$H_0^{(\boldsymbol{\gamma})}: \; \boldsymbol{\mu} = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 X, \; \boldsymbol{\gamma} = \boldsymbol{\beta}_0$	$\mathrm{DGM}_1^{(\boldsymbol{\gamma})}: \; \boldsymbol{\mu} = \boldsymbol{\alpha}_0^* + \boldsymbol{\alpha}_1^* X, \; \boldsymbol{\gamma} = \boldsymbol{\beta}_{1r}^* X$
	$\mathrm{DGM}_2^{(\boldsymbol{\gamma})}: \; \boldsymbol{\mu} = \boldsymbol{\alpha}_0^* + \boldsymbol{\alpha}_1^* X, \; \boldsymbol{\gamma} = \boldsymbol{\beta}_0^* + \boldsymbol{\beta}_{1r}^* X$
Values of model parameters	$\boldsymbol{\alpha}_0^* = (2, -5, 3, 5)^\top,$
	$\boldsymbol{\alpha}_1^* = (2, 1, 2, 1)^\top, \; \boldsymbol{\alpha}_{1r}^* = \frac{r}{2}\mathbf{1}_4,$
	$\boldsymbol{\beta}_0^* = (3, 5, -3, -4, 2)^\top, \; \boldsymbol{\beta}_{0r}^* = \frac{r}{\sqrt{5}}\mathbf{1}_5$
	$\boldsymbol{\beta}_1^* = (4, 2, 5, -2, 3)^\top, \; \boldsymbol{\beta}_{1r}^* = \frac{r}{\sqrt{5}}\mathbf{1}_5$

5. Simulation study

5.1. Design of simulation experiments

We are now in the position to study empirically operating characteristics of the proposed testing procedures for testing $H_0^{(\mathbf{V})}$, $H_0^{(\boldsymbol{\mu})}$, and $H_0^{(\boldsymbol{\gamma})}$ versus the alternative $H_1 : \; \mathbf{Y}|\mathbf{X} \sim \mathrm{ESAG}(\boldsymbol{\mu} = \boldsymbol{\alpha}_0 + \mathbf{A}_1 \mathbf{X}, \; \boldsymbol{\gamma} = \boldsymbol{\beta}_0 + \mathbf{B}_1 \mathbf{X})$. To this end, we design several data-generating mechanisms (DGM) for each null hypothesis. A random sample of size $n \in \{100, 200, 400\}$ is generated according to each DGM, based on which the proposed test statistics and their estimated p-values are computed following Algorithms 1–3 with $B = 300$. As a benchmark testing procedure to compare with ours, we also test each null using LR, with the corresponding p-value estimated via parametric bootstrap as opposed to assuming a χ^2 null distribution for LR as in Paine et al. (2020). This experiment is repeated 200 times at each simulation setting specified by the null hypothesis, DGM, and the level of n. Common in all settings, we consider a scalar covariate, with n realizations $\{X_i'\}_{i=1}^n$ generated from $\mathcal{N}(0, 1)$, followed by standardization via $X_i = (X_i' - X_{(1)}')/(X_{(n)}' - X_{(1)}') + 1$, for $i = 1, \ldots, n$. Given the covariate data $\{X_i\}_{i=1}^n$, response data $\{\mathbf{Y}_i\}_{i=1}^n$ are generated according to each DGM designed for inspecting the power of a test (see Table 1) or for checking the size of a test (see Table B.1 in Appendix B).

For each considered null hypothesis, we include one or multiple DGM's consistent with the null (see Table B.1 in the Supplementary Material). This allows for inspecting the size of a test. For each considered null, as one can see in Table 1, we also design several DGMs with increasing model complexity compared to the null. The values of some regression coefficients depend on a quantity r that we vary in the simulation to control the severity of model misspecification under a null, with a larger r leading to a more pronounced deviation of the DGM from a null. This allows for monitoring the power of a test as the true model deviates from the null model further.

The metric we record in the simulation study is the relative frequency of a considered test rejecting the current null across 200 Monte Carlo replicates at a pre-specified significance level. In what follows, we present these rejection rates associated with different tests for testing each of the three null hypotheses tabulated in Table 1.

5.2. Simulation results

The rejection rates of RoC, D or M, and LR versus the nominal significance level based on data generated from an ESAG regression model consistent with a null hypothesis are provided in Appendix B of the Supplementary Material. Empirically, the sizes of the proposed tests are mostly well controlled provided that the sample size is moderate or large (e.g., $n > 100$). A slight inflation in the test size often occurs when $n = 100$ or when some parameters (e.g., $\boldsymbol{\gamma}$) take true values (as zeros) that lead to irregular MLEs when fitting the full model. Such inflation is more evident for LR, especially when testing covariate dependence of model parameters. For example, when testing $H_0^{(\boldsymbol{\mu})}$ using a random sample of size 100 from an isotropic ESAG distribution with covariate-free $\boldsymbol{\mu}$, the rejection rates are 0.085, 0.080, 0.005, and 0.115 for the tests based on RoC, D, M, and LR, respectively, in contrast to the nominal level of 0.05. This can be where the size of LR fails to approach the nominal level asymptotically even when its p-value is estimated by the conventional parametric bootstrap, which is a phenomenon described in Drton and Williams (2011). One shall thus interpret the empirical power of LR with caution. For this reason, we omit to report the empirical power of LR for testing covariate dependence.

Table 2 presents the empirical power of various tests for testing each of the three null hypotheses at a significance level of 0.05 based on data from different true ESAG models specified in Table 1. When using RoC, M, and LR to test isotropy, the three tests are comparable in their power to detect anisotropy, with the power increasing steadily as n grows bigger or as the true value of $\boldsymbol{\gamma}$ deviates from zero further (by having a larger r). Having a covariate-dependent $\boldsymbol{\gamma}$ in the true regression model also enhances the power of these tests.

Table 2

Rejection rates of tests for $H_0^{(V)}$, $H_0^{(\mu)}$, and $H_0^{(\gamma)}$ at nominal level 0.05, with the highest rejection rate at each level of r when $n = 100$ highlighted in bold.

$\{n\}$	100	200	400	100	200	400	100	200	400
Testing $H_0^{(V)}$: $\mu = \alpha_0 + \alpha_1 X, \gamma = 0$									
	RoC			M			LR		
r	DGM$_1^{(V)}$: $\mu = \alpha_0^* + \alpha_1^* X, \gamma = \beta_{0r}^*$								
0.1	0.065	0.075	0.110	**0.075**	0.095	0.105	0.055	0.070	0.100
0.2	0.085	0.145	0.245	0.070	0.145	0.280	**0.095**	0.145	0.305
0.4	0.160	0.440	0.870	**0.215**	0.360	0.745	0.175	0.470	0.900
r	DGM$_2^{(V)}$: $\mu = \alpha_0^* + \alpha_1^* X, \gamma = \beta_{0r}^* + \beta_{1r}^* X$								
0.1	0.090	0.210	0.420	**0.110**	0.190	0.390	0.080	0.215	0.470
0.2	0.325	0.675	0.990	0.290	0.525	0.890	**0.345**	0.710	1.000
0.4	0.895	1.000	1.000	0.740	0.980	1.000	**0.910**	1.000	1.000
Testing $H_0^{(\mu)}$: $\mu = \alpha_0, \gamma = \beta_0 + \beta_1 X$									
	RoC			D			M		
r	DGM$_1^{(\mu)}$: $\mu = \alpha_0^* + \alpha_{1r}^* X, \gamma = 0$								
0.5	**0.085**	0.075	0.065	**0.085**	0.055	0.070	0.005	0.005	0.025
1	0.160	0.200	0.265	**0.165**	0.195	0.240	0.005	0.005	0.025
2	0.485	0.660	0.800	**0.490**	0.635	0.815	0.030	0.115	0.160
r	DGM$_2^{(\mu)}$: $\mu = \alpha_0^* + \alpha_{1r}^* X, \gamma = \beta_0^*$								
0.5	**0.245**	0.440	0.760	0.235	0.350	0.635	0.070	0.075	0.105
1	**0.675**	0.935	0.995	0.655	0.930	0.995	0.125	0.205	0.160
2	**0.980**	1.000	1.000	**0.980**	0.995	1.000	0.375	0.470	0.525
r	DGM$_3^{(\mu)}$: $\mu = \alpha_0^* + \alpha_{1r}^* X, \gamma = \beta_0^* + \beta_1^* X$								
0.5	**0.460**	0.850	0.975	0.435	0.745	0.935	0.080	0.140	0.145
1	**0.980**	1.000	1.000	0.975	1.000	1.000	0.230	0.215	0.240
2	**1.000**	1.000	1.000	**1.000**	1.000	1.000	0.655	0.775	0.880
Testing $H_0^{(\gamma)}$: $\mu = \alpha_0 + \alpha_1 X, \gamma = \beta_0$									
	RoC			D			M		
r	DGM$_1^{(\gamma)}$: $\mu = \alpha_0^* + \alpha_1^* X, \gamma = \beta_{1r}^* X$								
0.5	**0.065**	0.120	0.090	**0.065**	0.125	0.090	0.040	0.045	0.035
1	**0.105**	0.135	0.130	**0.105**	0.135	0.125	0.020	0.030	0.070
2	**0.135**	0.145	0.170	**0.135**	0.155	0.160	0.050	0.070	0.090
r	DGM$_2^{(\gamma)}$: $\mu = \alpha_0^* + \alpha_1^* X, \gamma = \beta_0^* + \beta_{1r}^* X$								
0.5	**0.070**	0.045	0.055	0.065	0.050	0.050	0.025	0.070	0.075
1	0.070	0.105	0.095	**0.075**	0.105	0.090	0.055	0.060	0.055
2	**0.155**	0.165	0.340	0.150	0.165	0.325	0.025	0.065	0.045

According to Table 2, the tests based on RoC and D enjoy higher power to detect covariate dependence of μ when the true model also has a covariate-dependent γ (as in DGM$_3^{(\mu)}$) than when it has an intercept-only model for γ (as in DGM$_2^{(\mu)}$). Noting that obtaining the unrestricted MLE for γ using data from DGM$_2^{(\mu)}$ creates an irregular maximum likelihood estimation, but the same estimation using data from DGM$_3^{(\mu)}$ is a regular case, we believe that having irregular MLEs for model parameters can compromise the power of RoC and D. When testing $H_0^{(\gamma)}$, the power of the proposed tests does not increase as quickly as when testing $H_0^{(\mu)}$ when n increases or when the covariate dependence becomes stronger. We conjecture that, once we allow μ to depend on covariates, inferences for the concentration are less sensitive to the assumption of covariate-independent for γ, and thus RoC and D may lack high power to detect the dependence of γ on covariates unless when the dependence is very strong.

The moment-based test using M is much less powerful than the tests based on RoC and D for testing covariate dependence of model parameters. By solely focusing on the fit for the mean of $\mathbf{Y}^2$, the power M to detect model misspecification heavily hinges on the impact of the misspecification on second-moment estimation. The observed phenomenon suggests some level of robustness of the second-moment estimation to covariate dependence of ESAG model parameters. In additional simulation study not reported here where we generate covariate data from different distributions, we observe that likelihood-based estimation of $E(\mathbf{Y}^2)$ is more sensitive to violation of $H_0^{(\mu)}$ or $H_0^{(\gamma)}$ when the covariate distribution is skewed, and, consequently, M becomes more powerful in detecting covariate dependence of μ or γ.

Table 3
Rejection rates of tests for isotropy using RoC, the scatter test, and the hybrid test at nominal level 0.05, with the highest rejection rate at each setting highlighted in bold.

n	r	RoC	Scatter Test	Hybrid Test
200	0.2	**0.253**	0.170	0.133
	0.4	**0.763**	0.523	0.450
400	0.2	**0.473**	0.313	0.263
	0.4	**0.983**	0.850	0.787

5.3. A comparative study on tests for isotropy

A referee brought to our attention tests for rotational symmetry about a location of a directional distribution proposed in García-Portugués et al. (2020), where the study is not limited to a particular distribution family. The location is $\boldsymbol{\mu}$ in our context, and rotational symmetry about $\boldsymbol{\mu}$ amounts to isotropy considered in our study. Their proposed tests are based on a special form of $\boldsymbol{\mu}$-dependent "projection" of $\mathbf{Y}$ that falls in $\mathbb{R}^{d-1}$ and, after normalization to possess a unit norm, is uniformly distributed on $\mathbb{S}^{d-2}$ if $\mathbf{Y}$ is rotationally symmetric about $\boldsymbol{\mu}$. This normalized projection is known as the multivariate sign, denoted by $\mathbf{u}_{\boldsymbol{\mu}}(\mathbf{Y})$. Test statistics based on $\mathbf{u}_{\boldsymbol{\mu}}(\mathbf{Y})$ are constructed to assess the discrepancies between the first two moments of $\mathbf{u}_{\boldsymbol{\mu}}(\mathbf{Y})$ suggested by the data and the first two moments of a uniform distribution on $\mathbb{S}^{d-2}$. To compare their moment-based strategies (developed in the non-regression setting) with our RoC-based isotropy test, we use the R package, rotasym (García-Portugués et al., 2023), to implement the test called the scatter test, which is based on the second-moment discrepancy, and the hybrid test, which accounts for discrepancies in both the first and second moments.

In the comparative study, we generate random samples of size $n \in \{200, 400\}$ from ESAG$(\boldsymbol{\mu}, \boldsymbol{\gamma})$, where $\boldsymbol{\mu} = (2, -5, 3, 5)^{\top}$ and $\boldsymbol{\gamma} = (r/\sqrt{5})\mathbf{1}_5$, for $r \in \{0.2, 0.4\}$, with a higher value of r leading to a greater degree of anisotropy. Table 3 provides the rejection rates of the three considered tests across 300 Monte Carlo replicates in each simulation setting. By exploiting the concentration estimation with and without the assumption of isotropy under the ESAG family, the proposed RoC test achieves higher power than the two competing methods that do not assume a particular distribution family as the alternative or full model. This may be partly thanks to the true data-generating mechanism being ESAG in this experiment. The competing methods have the advantage of having the test statistics follow some χ^2 distribution asymptotically, and thus the corresponding p-values can be easily obtained without a bootstrap procedure as for RoC. Systematically theoretical and empirical comparisons between the RoC test and these tests when the true data-generating mechanism deviates from ESAG are interesting directions to pursue in follow-up research.

6. Prediction regions

Following the estimation of all model parameters in an ESAG regression model, one can predict the outcome of the directional response $\mathbf{Y}$. If all model parameters are known, similar to the prediction region for a multivariate Gaussian distribution (Chew, 1966), a sensible $100(1 - a)\%$ prediction region that reflects the elliptical symmetry of ESAG$(\boldsymbol{\mu}, \boldsymbol{\gamma})$ is an ellipsoidal ball given by

$$\text{PR}_a = \left\{ \mathbf{y} \in \mathbb{S}^{d-1} : (\mathbf{y} - \boldsymbol{\mu}/\|\boldsymbol{\mu}\|)^{\top} \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}/\|\boldsymbol{\mu}\|) \leq q_{\alpha} \right\}, \tag{5}$$

where q_a is chosen such that $P(\mathbf{Y} \in \text{PR}_a) = 1 - a$. We show in Appendix C of the Supplementary Material that PR_a defined in (5) has the smallest volume in a class of ellipsoidal prediction regions centering around $\boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ with the nominal coverage probability of $1 - a$.

When the model parameters are unknown, we evaluate $\boldsymbol{\mu}$ and $\mathbf{V}$ at their MLEs, $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{V}}$, in (5), and estimate q_a by $\hat{q}_a$ that is obtained using bootstrap samples from the estimated ESAG distribution. This leads to a $100(1 - a)\%$ prediction region defined as

$$\widehat{\text{PR}}_a = \left\{ \mathbf{y} \in \mathbb{S}^{d-1} : (\mathbf{y} - \hat{\boldsymbol{\mu}}/\|\hat{\boldsymbol{\mu}}\|)^{\top} \hat{\mathbf{V}}^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}/\|\hat{\boldsymbol{\mu}}\|) \leq \hat{q}_a \right\}. \tag{6}$$

Algorithm 4 below provides the detailed computational path leading to $\widehat{\text{PR}}_a$ when $\mathbf{X} = \mathbf{x}_0$. Appendix D of the Supplementary Material presents a simulation study where we follow Algorithm 4 to compute prediction regions of different nominal coverage probabilities based on samples of size $n \in \{200, 400, 800\}$. The simulation results suggest that the empirical coverage probabilities of the resultant prediction regions match closely with the nominal levels.

7. Real-life data applications

We now put into action the regression analysis toolkit on data examples from two real-life applications.

7.1. Hydrochemical data

We analyzed in a recent work (Yu and Huang, 2024) the relative abundance of two major ions, K^+ and Na^+, and two minor ions, Ca^{2+} and Mg^{2+}, in water samples collected from two sets of locations between the summer of 1997 and the spring of 1999:

Algorithm 4 Compute the prediction region in (6).

1: **procedure** PARAMETRIC BOOTSTRAP ACCOUNTING FOR VARIATION OF ESAG
2:     Given the observed data $\{(\mathbf{Y}_i, \mathbf{X}_i)\}_{i=1}^n$, compute the MLEs for regression coefficients, $\hat{\boldsymbol{\alpha}}_0$, $\hat{\mathbf{A}}_1$, $\hat{\boldsymbol{\beta}}_0$, and $\hat{\mathbf{B}}_1$, assuming an ESAG model for $\mathbf{Y}_i$ conditioning on $\mathbf{X}_i$.
3:     Compute $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\alpha}}_0 + \hat{\mathbf{A}}_1 \mathbf{x}_0$ and $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\beta}}_0 + \hat{\mathbf{B}}_1 \mathbf{x}_0$, obtain the corresponding $\hat{\mathbf{V}}$.
4:     Set m = the number of bootstrap samples. Generate a random sample, $\{\mathbf{Y}'_j\}_{j=1}^m$, from ESAG($\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\gamma}}$).
5:     Compute $q_j = (\mathbf{Y}'_j - \hat{\boldsymbol{\mu}}/\|\hat{\boldsymbol{\mu}}\|)^{\mathsf{T}} \hat{\mathbf{V}}^{-1} (\mathbf{Y}'_j - \hat{\boldsymbol{\mu}}/\|\hat{\boldsymbol{\mu}}\|)$, for $j = 1, ..., m$.
6: **end procedure**
7: **procedure** NONPARAMETRIC BOOTSTRAP ACCOUNTING FOR VARIATION OF MLES
8:     Set B = the number of bootstrap samples.
9:     **for** b in $1, ..., B$ **do**
10:         Generate the b-th bootstrap sample $\{\mathbf{Y}_i^{(b)}, \mathbf{X}_i^{(b)}\}_{i=1}^n$ via sampling with replacement from the raw data.
11:         Repeat Steps 2–5 using data $\{(\mathbf{Y}_i^{(b)}, \mathbf{X}_i^{(b)})\}_{i=1}^n$. Denote the bootstrap version of q_j as $q_j^{(b)}$.
12:     **end for**
13:     Viewing $\{q_j, q_j^{(1)}, ..., q_j^{(B)}\}_{j=1}^m$ as a sample of size $m \times (B + 1)$, find the $(1 - a)$-quantile of this sample. Denote this sample quantile as $\hat{q}_a$.
14:     Output a $100(1 - a)\%$ prediction region when $\mathbf{X} = \mathbf{x}_0$ given by $\{\mathbf{y} \in \mathbb{S}^{d-1} : (\mathbf{y} - \hat{\boldsymbol{\mu}}/\|\hat{\boldsymbol{\mu}}\|)^{\mathsf{T}} \hat{\mathbf{V}}^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}/\|\hat{\boldsymbol{\mu}}\|) \leq \hat{q}_a\}$.
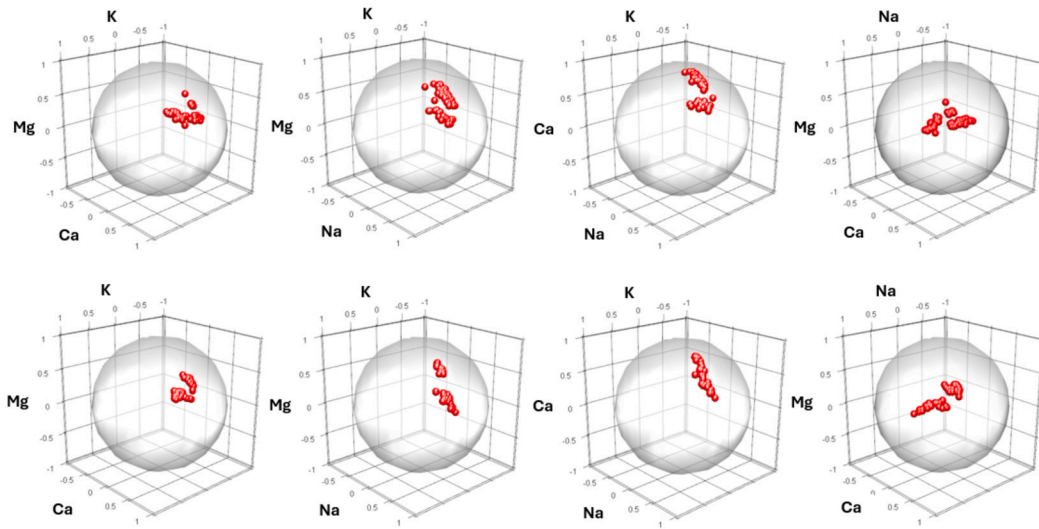15: **end procedure**

Fig. 2. Directional data on $\mathbb{S}^2$ corresponding to each triplet of components from tributaries of Anoia (upper panels) and those from the lower Llobregat course (lower panels).

67 samples from tributaries of Anoia and 43 samples from tributaries of the lower Llobregat course in Spain (Otero et al., 2005). The complete data are available in the R package, compositions (Van den Boogaart and Tolosana-Delgado, 2008). The relative abundance of (K^+, Na^+, Ca^{2+}, Mg^{2+}) is an example of compositional data in a 3-dimensional simplex, $\mathbb{C}^3 = \{\mathbf{y}^* \in \mathbb{R}^4 : \mathbf{1}_4^{\mathsf{T}} \mathbf{y}^* = 1 \text{ and } \mathbf{e}_j^{\mathsf{T}} \mathbf{y}^* \geq 0, \text{ for } j = 1, ..., 4\}$, where $\mathbf{1}_4$ is the 4×1 vector of ones, and $\mathbf{e}_j$ is the unit vector with the j-th entry being 1. We transformed the compositional data by taking the square-root of $\mathbf{y}^* \in \mathbb{C}^3$ element-wise to directional data in $\mathbb{S}^3$. Previous analyses of the directional data from each set of locations suggested an adequate fit of an intercept-only ESAG model, but a poor fit for the combined data of size $n = 110$ from two sets of locations. Fig. 2 presents the directional data associated with triplets of components from each set of locations on the three-dimensional sphere $\mathbb{S}^2$. Even though the directional data to be analyzed are on $\mathbb{S}^3$, a spherical space hard to visualize, the partial data plotted on $\mathbb{S}^2$ in Fig. 2 seem to suggest that data from tributaries of Anoia are more variable than those from tributaries of the lower Llobregat course. The lack of circular shape of (many of) the data clouds may also suggest anisotropy of the underlying distribution.

These earlier findings and data visualization motivate a location-dependent ESAG model for all data from these locations, where we incorporate a covariate X indicating locations, with $X = 0$ corresponding to tributaries of Anoia (At), and $X = 1$ representing tributaries of the lower Llobregat course (LLt). Fitting the regression model, $Y_i | X_i \sim \text{ESAG}(\boldsymbol{\mu}_i = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 X_i, \boldsymbol{\gamma}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 X_i)$, to the data, we arrive at the following estimates for the ESAG model parameters,

$$\hat{\boldsymbol{\mu}}_i = \begin{bmatrix} 1.99 \\ 5.74 \\ 7.95 \\ 4.59 \end{bmatrix} + \begin{bmatrix} 1.28 \\ 2.83 \\ 1.06 \\ 1.20 \end{bmatrix} X_i, \qquad \hat{\boldsymbol{\gamma}}_i = \begin{bmatrix} -0.67 \\ 0.15 \\ -0.82 \\ 6.12 \\ 0.64 \end{bmatrix} + \begin{bmatrix} 2.43 \\ -0.22 \\ 10.17 \\ -20.19 \\ 0.47 \end{bmatrix} X_i.$$

Hence, for the directional response associated with tributaries of Anoia, the mean direction is estimated to be $\hat{\boldsymbol{\mu}}_{\text{At}} = (1.99, 5.74, 7.95, 4.59)^{\top}$, and, for the directional response coming from tributaries of the lower Llobregat course, the estimated mean direction is $\hat{\boldsymbol{\mu}}_{\text{LLt}} = (3.27, 8.57, 9.01, 5.79)^{\top}$. These estimates lead to the estimated concentration at each set of locations that provide data evidence suggesting that the latter set of locations exhibits a higher concentration than the former. These are coherent with results in our previous analysis when we analyzed data from one set of locations at a time, and also in line with the visual impression from Fig. 2. Estimates for $\boldsymbol{\gamma}_i$ when $X_i = 0$ and 1 are also aligned with our earlier analyses (and are omitted here), based on which estimates for $\mathbf{V}$ corresponding to two sets of locations, $\hat{\mathbf{V}}_{\text{At}}$ and $\hat{\mathbf{V}}_{\text{LLt}}$, can be obtained.

For model diagnosis, we carry out tests for isotropy and covariate dependence of $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ based on the three proposed test statistics. All tests suggest statistically significant evidence of location-dependent model parameters in the ESAG distribution that is anisotropic for the (transformed) compositions of (K^+, Na^+, Ca^{2+}, Mg^{2+}). All estimated p-values are less than 10^{-3} except for that associated with M when testing covariate dependence of $\boldsymbol{\gamma}$, which returns an estimated p-value less than 0.01 (although larger than 10^{-3}). This is consistent with findings in existing literature reporting that the hydrochemical profile of Anoia and that of the lower Llobergat course are substantially different because the two sets of tributaries pass through zones that are differently populated with vastly different distributions of agricultural and industrial areas (González et al., 2012). Zooming in on the tests for covariate dependence of $\boldsymbol{\mu}$, we have $D = 1.062$ that is somewhat higher than $\text{RoC} = 1.059$. This can be data evidence indicating that not only the norm of the mean direction depends on X, that is, the concentration varies across locations, but also the orientation of the mean direction differs between locations. When testing covariate dependence of $\boldsymbol{\gamma}$, i.e., testing $H_0^{(\gamma)}$ versus the full model, the two statistics are nearly equal (at around 1.090). This suggests that, once we acknowledge a location-dependent $\boldsymbol{\mu}$ in the null model, allowing $\boldsymbol{\gamma}$ to depend on X in the alternative model mostly helps to distinguish the variability of data across different locations but it may not contribute to capturing the discrepancy in the orientation of $\boldsymbol{\mu}$ in different locations.

Lastly, applying Algorithm 4 with $x_0 = 0$ and 1, we obtain the prediction regions for the two sets of locations given by

$$\widehat{\text{PR}}_a^{(\text{At})} = \{\mathbf{y} \in \mathbb{S}^3 \,:\, (\mathbf{y} - \hat{\boldsymbol{\mu}}_{\text{At}}/\|\hat{\boldsymbol{\mu}}_{\text{At}}\|)^{\top}\hat{\mathbf{V}}_{\text{At}}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}_{\text{At}}/\|\hat{\boldsymbol{\mu}}_{\text{At}}\|) \leq \hat{q}_a^{(\text{At})}\},$$
$$\widehat{\text{PR}}_a^{(\text{LLt})} = \{\mathbf{y} \in \mathbb{S}^3 \,:\, (\mathbf{y} - \hat{\boldsymbol{\mu}}_{\text{LLt}}/\|\hat{\boldsymbol{\mu}}_{\text{LLt}}\|)^{\top}\hat{\mathbf{V}}_{\text{LLt}}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}_{\text{LLt}}/\|\hat{\boldsymbol{\mu}}_{\text{LLt}}\|) \leq \hat{q}_a^{(\text{LLt})}\},$$

with $\hat{q}_a^{(\text{At})} = 0.029, 0.036, 0.050$ and $\hat{q}_a^{(\text{LLt})} = 0.018, 0.023, 0.031$ for the nominal level of 70%, 80%, and 90%, respectively. At each considered nominal level, having $\hat{q}_a^{(\text{LLt})} < \hat{q}_a^{(\text{At})}$ is in line with the finding that the distribution of directional data from the lower Llobregat course exhibits a higher concentration (i.e., lower variability) than that for Anoia.

7.2. Microbiome data

We now turn to a dataset regarding the gut microbiota of elderly adults. Besides gut microbiome compositions of 160 elderly adults, also recorded in this data include the residence types, age, body mass index (BMI), diet, and gender. A similar dataset has been analyzed by Claesson et al. (2012), where the authors carried out a principal component analysis to study correlations of the relative abundance of various microorganisms in the gut. Shen et al. (2022) used the Gaussian chain graph model for the data to infer the effects of one's diet and residence type on gut microbiome composition. For illustration purposes, we study the potential association between two covariates, one's age and BMI, and the directional response on $\mathbb{S}^3$ defined as the square root of the relative abundance of four genera of bacteria found in the gut: Blautia, Caloramator, Clostridium, and Faecalibacterium.

We first fit the directional response data to the ESAG regression model, for $i = 1, \dots, 160$,

$$\mathbf{Y}_i | (\text{Age}_i, \text{BMI}_i) \sim \text{ESAG}(\boldsymbol{\mu}_i = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 \text{Age}_i + \boldsymbol{\alpha}_2 \text{BMI}_i, \, \boldsymbol{\gamma}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \text{Age}_i + \boldsymbol{\beta}_2 \text{BMI}_i), \tag{7}$$

where $\mathbf{Y}_i = (Y_{i,1}, Y_{i,2}, Y_{i,3}, Y_{i,4})^{\top}$, with $Y_{i,j}$ equal to the squared root of the relative abundance of Blautia, Caloramator, Clostridium, and Faecalibacterium, for $j = 1, 2, 3, 4$, respectively, for subject i, $\text{Age}_i = (\text{subject } i\text{'s age} - \text{the youngest subject's age})/(\text{age range}) + 1$, and BMI_i is similarly computed by standardizing the BMI data. Maximum likelihood estimation yields

$$\hat{\boldsymbol{\mu}}_i = \begin{bmatrix} 1.76 \\ 0.62 \\ 5.27 \\ 3.23 \end{bmatrix} + \begin{bmatrix} -1.46 \\ 0.59 \\ -2.70 \\ -2.93 \end{bmatrix} \text{Age}_i + \begin{bmatrix} 1.11 \\ -0.63 \\ 1.63 \\ 2.39 \end{bmatrix} \text{BMI}_i,$$

$$\hat{\boldsymbol{\gamma}}_i = \begin{bmatrix} -6.39 \\ 0.31 \\ 3.12 \\ 0.63 \\ 0.69 \end{bmatrix} + \begin{bmatrix} 24.08 \\ 2.54 \\ -4.29 \\ -3.31 \\ 0.61 \end{bmatrix} \text{Age}_i + \begin{bmatrix} -21.15 \\ -2.35 \\ 3.07 \\ 3.20 \\ -0.79 \end{bmatrix} \text{BMI}_i. \tag{8}$$

We then carry out the residual-based goodness-of-fit test proposed in an earlier work (Yu and Huang, 2024). We showed there that, if $\mathbf{Y} \sim \text{ESAG}(\boldsymbol{\mu}, \boldsymbol{\gamma})$, then $T = (\|\boldsymbol{\mu}\|^2 + \sum_{j=1}^d \lambda_j)\mathbf{r}\mathbf{V}^{-1}\mathbf{r}$ follows χ_{d-1}^2 approximately, where $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{d-1}$ and $\lambda_d = 1$ are the eigenvalues of $\mathbf{V}$, and $\mathbf{r} = (\mathbf{I}_d - \hat{\mathbf{Y}}\hat{\mathbf{Y}}^{\top})\mathbf{Y}$ is the directional residual (Jupp, 1988) associated with the prediction $\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}}/\|\hat{\boldsymbol{\mu}}\|$. Fig. 3 shows the residual-based quantities T evaluated at the MLEs of unknown parameters, $\{\hat{T}_i\}_{i=1}^{160}$, where $\hat{T}_i = (\|\hat{\boldsymbol{\mu}}_i\|^2 + \sum_{j=1}^3 \hat{\lambda}_{i,j} + 1)\mathbf{r}_i\hat{\mathbf{V}}_i^{-1}\mathbf{r}_i$. In particular, the empirical distribution of T depicted by the histogram of $\{\hat{T}_i\}_{i=1}^{160}$ appears to resemble χ_3^2, even though the scatter
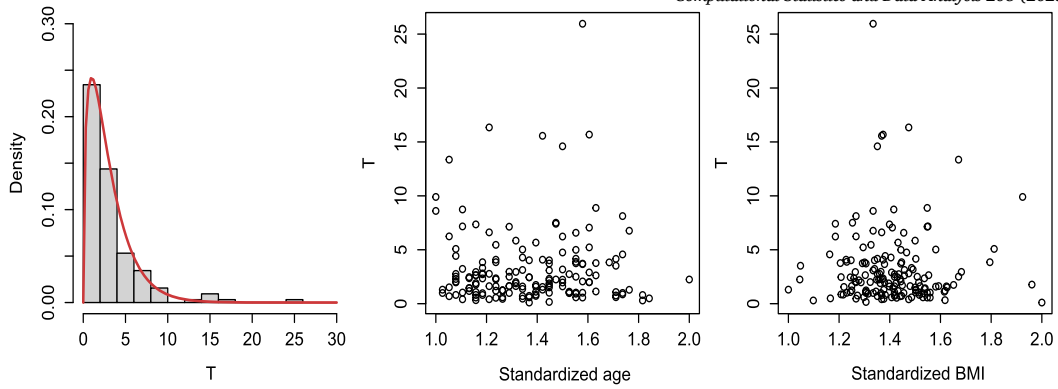
Fig. 3. The histogram of $\{\hat{T}_i\}_{i=1}^{160}$ with the density of χ_3^2 superimposed (in the left panel), the scatter plot of $\{(\text{Age}_i, \hat{T}_i)\}_{i=1}^{160}$ (in the middle panel), and the scatter plot of $\{(\text{BMI}_i, \hat{T}_i)\}_{i=1}^{160}$ (in the right panel) based on microbiome data modeled by (7).
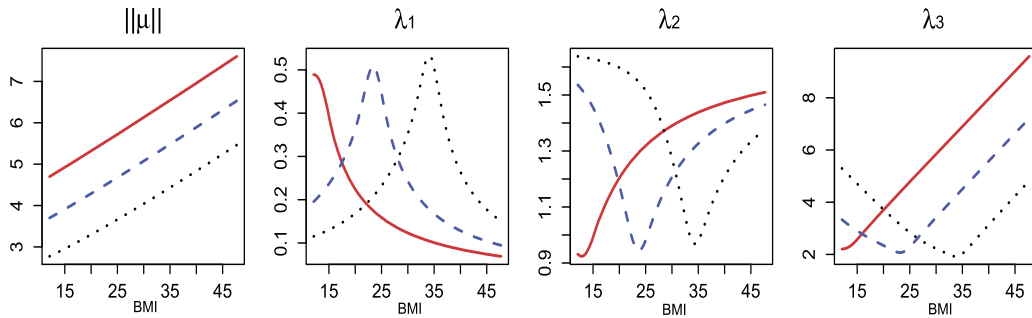

Fig. 4. Estimates of $\|\boldsymbol{\mu}\|$, λ_1, λ_2, and λ_3 versus BMI when one is 70 (solid lines), 80 (dashed lines), and 90 (dotted lines) years of age.

plots of $\{\hat{T}_i\}_{i=1}^{160}$ versus the covariates values seem to suggest several outliers in the sample. To estimate the null distribution of T without approximating its distribution by χ_{d-1}^2, the authors also developed a bootstrap-based test for assessing the adequacy of an ESAG model. This test applied to the current dataset yields an estimated *p*-value of 0.58, suggesting insufficient evidence for the lack of fit of the current model. In addition, the tests for isotropy based on RoC and M, and the tests for covariates dependence of ESAG model parameters based on RoC and D all produce estimated *p*-values less than 0.01. We thus conclude significant covariates effects on the ESAG model parameters and recommend against opting for a regression model more parsimonious than (7). To further demonstrate the versatility and informativeness of RoC and D, we test the significance of age in modeling $\boldsymbol{\mu}$ given that $\boldsymbol{\gamma}$ depends on both covariates as in (7) and that $\boldsymbol{\mu}$ depends on BMI. These tests yield RoC = 1.068 and $D = 1.070$, with estimated *p*-values less than 0.001. These can serve as data evidence indicating the significance of age in modeling (the BMI-dependent) $\boldsymbol{\mu}$ and that age potentially impacts the direction of $\boldsymbol{\mu}$ (besides its norm) significantly. These findings on the significance of the covariate effect and overall goodness-of-fit remain unchanged when we use the normalized covariate data with mean zero and variance one.

To further elucidate the effects of age and BMI on ESAG model features, we present in Fig. 4 estimates of the concentration and three eigenvalues of $\mathbf{V}$, $(\lambda_1, \lambda_2, \lambda_3)$, versus BMI when one is 70, 80, and 90 years of age. As age increases, we observe in Fig. 4 a decrease in the estimate of $\|\boldsymbol{\mu}\|$, corresponding to an increase in the estimated overall variation of $\mathbf{Y}$. The finding of highly variable directional distribution can imply highly variable in the composition of the gut microbiota among the elderly, which is a finding reported in existing literature but has been mostly stated in comparison with younger (than 65) healthy adults that are found to have a more stable composition of intestinal microorganisms (Claesson et al., 2012). Our results here can be evidence for that, even among the elderly, the trend of higher variability in microbiome composition as one ages persists. Moreover, a higher BMI also leads to a more variable distribution. Examining the estimated eigenvalues of $\mathbf{V}$, one can see two change points in BMI: one at BMI of nearly 25 for an 80-year-old and the other at BMI of around 35 for a 90-year-old. The first change point separates healthy weight (BMI $\in (18.5, 24.9)$) and overweight (BMI $\in (25.0, 29.9)$); the second change point belongs to the obese range (https://www.cdc.gov/healthyweight/assessing). Because $\lambda_1 = \lambda_2 = \lambda_3 = 1 (= \lambda_4)$ implies $\mathbf{V} = \mathbf{I}_4$, the proximity of the three considered eigenvalues to 1 implies (nearly) isotropy of the directional distribution and also relates to weak correlations between the four genera of bacteria. The aforementioned change points are where the estimates for these eigenvalues are closest to 1, and thus the distribution of $\mathbf{Y}$ tends to be more isotropic when $\mathbf{X} = (\text{Age} = 80, \text{BMI} \approx 25)$ and $(\text{Age} = 90, \text{BMI} \approx 35)$. This can also imply a reduction in the correlation between the relative abundance of the four considered genera of bacteria at these change points.

8. Discussion

We develop in this study a complete package of regression analysis for directional response built upon the ESAG distribution family indexed by constraint-free parameters. We consider a full range of statistical inference problems, including parameter estimation, testing hypotheses on model features, and prediction. The uncertainty of parameter estimation can be assessed via bootstrap. Parametric bootstrap is also heavily involved in all proposed inference procedures, which is straightforward to implement owing to the formulation and parametrization of ESAG that allow for easy data generation from an ESAG distribution. Computer programs for implementing all proposed methods are available at https://github.com/Zehaoyu217/ESAG/blob/main/ESAG_Project2.R. We also demonstrate the use of this package for analyzing two datasets from different fields of applications.

The number of parameters in an ESAG regression model can be large in an application since the dimension of the parameter space grows quadratically in the dimension of $\mathbf{Y}$, d, and linearly in the number of covariates, q. For example, in microbiome analysis, d is the dimension of the compositional response, which typically is much larger than four, and one may wish to consider many covariates relating to the host's physiological characteristics. We have started developing penalized likelihood-based methods to deal with high-dimensional directional data. Besides this ongoing follow-up research, another interesting topic is compositional data analysis the two case studies in Section 7 relate to. The idea of relating compositional data on a simplex to directional data on a hypersphere has been explored (Scealy and Welsh, 2011, 2017; Li et al., 2023) but with many open questions yet to be addressed. In this particular context, more components tend to have zero or nearly zero relative abundance as d increases, which is a data pattern ESAG and most existing named directional distributions tend to fit poorly. Interpretations and implications of model parameters of a directional distribution that are practically meaningful for the corresponding compositional data also demand further systematic investigation.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2025.108167.

References

Banerjee, A., Dhillon, I.S., Ghosh, J., Sra, S., Ridgeway, G., 2005. Clustering on the unit hypersphere using von Mises-Fisher distributions. J. Mach. Learn. Res. 6.

Chernoff, H., 1954. On the distribution of the likelihood ratio. Ann. Math. Stat. 25, 573–578.

Chew, V., 1966. Confidence, prediction, and tolerance regions for the multivariate normal distribution. J. Am. Stat. Assoc. 61, 605–617.

Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'connor, E.M., Cusack, S., Harris, H., Coakley, M., Lakshminarayanan, B., O'sullivan, O., et al., 2012. Gut microbiota composition correlates with diet and health in the elderly. Nature 488, 178–184.

Drton, M., 2009. Likelihood ratio tests and singularities. Ann. Stat. 37, 979–1012.

Drton, M., Williams, B., 2011. Quantifying the failure of bootstrap likelihood ratio tests. Biometrika 98, 919–934.

Ennajari, H., Bouguila, N., Bentahar, J., 2021. Combining knowledge graph and word embeddings for spherical topic modeling. IEEE Trans. Neural Netw. Learn. Syst. 34, 3609–3623.

García-Portugués, E., Paindaveine, D., Verdebout, T., 2020. On optimal tests for rotational symmetry against new classes of hyperspherical distributions. J. Am. Stat. Assoc. 115, 1873–1887.

García-Portugués, E., Paindaveine, D., Verdebout, T., 2023. rotasym: Tests for Rotational Symmetry on the Hypersphere. https://CRAN.R-project.org/package=rotasym, r package version 1.1.5.

González, S., López-Roldán, R., Cortina, J.L., 2012. Presence and biological effects of emerging contaminants in Llobregat River basin: a review. Environ. Pollut. 161, 83–92.

Johnson, R.A., Wehrly, T.E., 1978. Some angular-linear distributions and related regression models. J. Am. Stat. Assoc. 73, 602–606.

Jupp, P., 1988. Residuals for directional data. J. Appl. Stat. 15, 137–147.

Li, B., Yoon, C., Ahn, J., 2023. Reproducing kernels and new approaches in compositional data analysis. J. Mach. Learn. Res. 24, 1–34.

Lund, U., 1999. Least circular distance regression for directional data. J. Appl. Stat. 26, 723–733.

Mitchell, J.D., Allman, E.S., Rhodes, J.A., 2019. Hypothesis testing near singularities and boundaries. Electron. J. Stat. 13, 2150.

Otero, N., Tolosana-Delgado, R., Soler, A., Pawlowsky-Glahn, V., Canals, A., 2005. Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. Water Res. 39, 1404–1414.

Paine, P.J., Preston, S.P., Tsagris, M., Wood, A.T., 2018. An elliptically symmetric angular Gaussian distribution. Stat. Comput. 28, 689–697.

Paine, P.J., Preston, S., Tsagris, M., Wood, A.T., 2020. Spherical regression models with general covariates and anisotropic errors. Stat. Comput. 30, 153–165.

Presnell, B., Morrison, S.P., Littell, R.C., 1998. Projected multivariate linear models for directional data. J. Am. Stat. Assoc. 93, 1068–1077.

Ryali, S., Chen, T., Supekar, K., Menon, V., 2013. A parcellation scheme based on von mises-fisher distributions and Markov random fields for segmenting brain regions using resting-state FMRI. NeuroImage 65, 83–96.

Scealy, J., Welsh, A., 2011. Regression for compositional data by using distributions defined on the hypersphere. J. R. Stat. Soc., Ser. B, Stat. Methodol. 73, 351–375.

Scealy, J., Welsh, A., 2017. A directional mixed effects model for compositional expenditure data. J. Am. Stat. Assoc. 112, 24–36.

Scealy, J., Wood, A.T., 2019. Scaled von Mises–Fisher distributions and regression models for paleomagnetic directional data. J. Am. Stat. Assoc.

Shen, Y., Solís-Lemus, C., Deshpande, S.K., 2022. Sparse Gaussian chain graphs with the spike-and-slab LASSO: Algorithms and asymptotics. arXiv preprint. arXiv:2207.07020.

Van den Boogaart, K.G., Tolosana-Delgado, R., 2008. "Compositions": a unified R package to analyze compositional data. Comput. Geosci. 34, 320–338.

Wang, F., Gelfand, A.E., 2013. Directional data analysis under the general projected normal distribution. Stat. Methodol. 10, 113–127.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.

Yu, Z., Huang, X., 2024. A new parameterization for elliptically symmetric angular gaussian distributions of arbitrary dimension. Electron. J. Stat. 18, 301–334.