χ 2 tests in the Riemannian manifold of negative binomial distributions: a geometrical approach. Application to ecological data

Claude Manté

χ^2 tests in the Riemannian manifold of negative binomial distributions: a geometrical approach. Application to ecological data

Claude Manté

Aix-Marseille Université, Université du Sud Toulon-Var, CNRS/INSU, IRD, MIO,UM 110, Campus de Luminy, Case 901, F13288 Marseille Cedex 09, France

email: claude.mante@gmail.com, claude.mante@mio.osupytheas.fr

Abstract

The statistical analysis of counts of living organisms brings information about the collective behavior of species (schooling, habitat preference, etc), possibly depending on their socio-biological characteristics (aggregation, growth rate, reproductive power, survival rate, *etc*). The negative binomial (NB) distribution is widely used to model such data, but the parametric approach suffers from an important nuisance: the visual distance between parameters cannot be considered as a relevant distance between distributions, because these parameters are not commensurable in general (different ecological meaning, different ranges, ...). On the contrary, considering the Riemannian manifold $NB(D_{\mathcal{R}})$ of NB distributions equipped with the Rao metrics $D_{\mathcal{R}}$, one can compute intrinsic distances between species which can be considered as absolute. Suppose now $D_{\mathcal{R}}(A, B)$ is "small"; does this means that species A and B have similar characteristics?

We first tackle this point by focusing on geometrical aspects of the χ^2 goodness-of-fit test for distributions in $NB(D_{\mathcal{R}})$, and question its feasibility

in connection with the position of the distributions. We afterward focus on a kin problem: performing two-sample χ^2 tests and building confidence regions in the same geometrical setting. Our purpose is illustrated by processing field experiment data studied by Bliss and Fisher in the fifties.
Keywords: Chi square test, Fisher-Rao distance, Riemannian manifolds, Multidimensional scaling

---

Notations

Let's introduce first some notation. Consider a Riemannian manifold $\mathfrak{M}$, and a parametric curve $\alpha : [a, b] \to \mathfrak{M}$; its first derivative will be denoted $\dot{\alpha}$. We will also consider for any $\theta \in \mathfrak{M}$ the local norm $\|V\|_g(\theta)$ associated with the metrics $\mathfrak{g}$ on the tangent space $T_\theta\mathfrak{M}$:

$$\forall\, V \in T_\theta\mathfrak{M},\ \|V\|_{\mathfrak{g}}(\theta) := \sqrt{V^t.\mathfrak{g}(\theta).V}. \tag{1}$$

A parametric probability distribution $\mathfrak{L}^i$ will be identified with its coordinates with respect to some chosen parametrization; for instance, we will write $\mathfrak{L}^i \equiv (\phi^i, \mu^i)$ for some negative binomial distribution. Finally, logical propositions will be combined by using the classical connectors $\vee$ (or) and $\wedge$ (and).

1. Introduction

The statistical analysis of counts of living organisms brings information about the collective behavior of species (schooling, habitat preference, *etc*), possibly associated with their socio-biological characteristics (aggregation, growth rate, reproductive power, survival rate, *etc*). The negative binomial (NB) distribution is widely used to model such data Bliss and Fisher (1953);

O'Neill and Faddy (2003); Vaudor et al. (2011); Manté et al. (2016). It is especially relevant for ecologists, because

1. it arises as a Gamma-Poisson mixture, whose parameters depend on the more or less aggregative behavior of the species, and on the efficiency Fisher et al. (1943); Anscombe (1950); Rao (1971) of the trap used to catch it
2. it arises as the limit distribution of the Kendall (1948) birth-and-death model; in this setting, its parameters depend on the demography of the species (reproductive power, mortality, immigration rate).

But while the parametric approach is quite sound from the ecological point of view (see Manté et al. (2016) and the references therein), it suffers from an important nuisance: the visual distance between the parameters of several distributions is misleading, because on the one hand it depends on the chosen parametrization and, on the other hand, because these parameters are not commensurable in general (different ecological meaning, different ranges, ...). It is possible to remedy this drawback by using the Rao's distance.

In a seminal paper, Rao (2015) noticed that, equipped with the Fisher information metrics denoted $\mathfrak{g}(\bullet)$, a family of probabilities depending on p parameters can be considered as a p-dimensional Riemannian manifold. The associated Riemannian (Rao's) distance between the distributions of parameters θ^1 and θ^2 is given by the integral

$$D_{\mathcal{R}}\left(\theta^1,\theta^2\right) := \int_0^1 \sqrt{\dot{\gamma}^t(t).\mathfrak{g}(\gamma(t)).\dot{\gamma}(t)}dt \tag{2}$$

where γ is a **segment** (minimal length geodesic curve) connecting $\theta^1 = \gamma(0)$ to $\theta^2 = \gamma(1)$ and $\dot{\gamma}(t) := \frac{d\gamma}{dt}(t)$. As any Riemannian distance, $D_{\mathcal{R}}$

is intrinsic (*i.e.* does not depend on the parametrization used). Naturally, Rao (2015, 1989) proposed to use (2) as a distance between populations or for Goodness-Of-Fit (GOF) testing, followed by a number of authors Carter et al. (2009); Galanis et al. (2012); Dodson (2012); Cubedo et al. (2013); Ilea et al. (2015); Manté and Kidé (2016); Kass (1989); Menendez et al. (1995); Cubedo and Oller (2002). Generally, the Rao's distance between members of a family of distributions must be obtained by numerically solving a second-order nonlinear differential equation (the Euler-Lagrange equation) .

2. Few elements of Riemannian geometry

Consider the set $\mathcal{X}(\mathfrak{M})$ of vector fields on $\mathfrak{M}$.

Definition 1. Berger (2003); Amari et al. (2007) A linear connection (or covariant derivative) $\boldsymbol{D}$ on $\mathfrak{M}$ is a bilinear map

$$\begin{cases} \boldsymbol{D} : \mathcal{X}(\mathfrak{M}) \times \mathcal{X}(\mathfrak{M}) \to \mathcal{X}(\mathfrak{M}) \\ (X, Y) \mapsto \boldsymbol{D}_X Y \end{cases}$$

which is linear in X and a derivation on Y.

More concretely Amari et al. (2007), a connection on $\mathfrak{M}$ determines, for neighboring points p and p', a correspondence (linear one-to-one mapping) $\Xi_{p,p'}$ between the tangent spaces $T_p(\mathfrak{M})$ and $T_{p'}(\mathfrak{M})$, possessing the properties above. According to the fundamental theorem of Riemannian geometry (Berger, 2003), there is a unique symmetric connection ∇ compatible with a given metrics $\mathfrak{g}$ (the Levi-Civita or Riemann connection), giving in our case the Rao's distance. It is noteworthy that, while we will work in $(\mathfrak{M}, \boldsymbol{\nabla}, \mathfrak{g})$, other statistically sound (but not Riemannian) connections can be fruitfully considered (see Amari et al. (2007)).

Definition 2. Berger (2003); Gray (1998) Let $\gamma : I \to \mathfrak{M}$ be a curve traced on $\mathfrak{M}$; γ is a geodesic with respect to ∇ if its acceleration $\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t)$ is null $\forall t \in I$.

2.1. The Riemannian measure on some statistical manifold

The Riemannian measure on $(\mathfrak{M}, \boldsymbol{\nabla}, \mathfrak{g})$ associated with $\mathfrak{g}$ is $dV_{\mathfrak{M}}(\theta^0) := \sqrt{\det(\mathfrak{g}_{ij}(\theta^0))}\, d\theta_1 \cdots d\theta_p$; it has been considered from different viewpoints in the literature. Kass (1989) argued that $dV_{\mathfrak{M}}$ can be considered as the **uniform measure** on $\mathfrak{M}$, since "uniform distributions on surfaces are special cases of measures that apply equal mass to sets of equal volume, that is, of measures that are determined by Riemannian metrics". Alternatively, according to Cubedo et al. (2013), $\mathfrak{g}(\theta^0)$ "is a local measure of the maximum mean value change of standardized random variables, in a neighborhood of $\mathfrak{L}^{\theta^0}$". Consequently, $dV_{\mathfrak{M}}(\theta^0)$ measures the **generalized variance** Anderson (1984) that is encountered in some neighborhood $\mathfrak{U}(\theta^0)$ of $\mathfrak{L}^{\theta^0}$. If $dV_{\mathfrak{M}}(\theta^0)$ is large, we should expect an important heterogeneity in $\mathfrak{U}(\theta^0)$ or, in other words, a high sensibility to small variations of the parameters. Computing the Rao's distance between randomly drawn distributions in $\mathfrak{U}(\theta^0)$, we will obtain a larger proportion of big distances than when $dV_{\mathfrak{M}}(\theta^0)$ is small, as Lebanon (2006) pointed.

2.2. Tangent plane approximation: the χ^2 GOF test

Definition 3. Berger (2003) Let $\mathfrak{M}$ be a Riemann manifold and $x \in \mathfrak{M}$. The **exponential map** of $\mathfrak{M}$ at x is $\exp_x : W_x \to \mathfrak{M}$, defined on some neighborhood W_x of the origin of $T_x\mathfrak{M}$ by:

$$\exp_x(V) := \alpha_{\mathcal{B}(V)}(\|V\|) \tag{3}$$

where $\mathcal{B}(V)$ is the projection of V onto the unit ball and $\alpha_{\mathcal{B}(V)}$ is the unique unit-speed geodesic in $\mathfrak{M}$ such that $\alpha_{\mathcal{B}(V)}(0) = x$ and $\dot{\alpha}_{\mathcal{B}(V)}(0) = \mathcal{B}(V)$.

$D_{\mathcal{R}}$ has been constructively defined by Formula 2; the following proposition shows incidentally that $D_{\mathcal{R}}(\theta^1, \theta^2) = 0 \Rightarrow \theta^1 = \theta^2$ Berger (2003).

Proposition 4. *For ϱ small enough, the exponential map at θ^0 is a local diffeomorphism such that $\exp_{\theta^0}(\mathcal{B}(0, \varrho)) = B_{\mathcal{R}}(\theta^0, \varrho)$, where $\mathcal{B}(0, \varrho)$ (resp. $B_{\mathcal{R}}(\theta^0, \varrho)$) is a ball of radius ϱ of $T_{\theta^0}\mathfrak{M}$ (resp. $\mathfrak{M}$).*

The maximum of these values (named **injectivity radius** of $\mathfrak{M}$ at θ^0) is denoted $\bar{\varrho}(\theta^0)$. We can always write, if θ is close to θ^0:

$$D_{\mathcal{R}}\left(\mathfrak{L}^{\theta^0}, \mathfrak{L}^{\theta}\right) \equiv D_{\mathcal{R}}\left(\theta^0, \theta\right) \approx \sqrt{{}^t(\theta^0 - \theta) . \mathfrak{g}(\theta^0) . (\theta^0 - \theta)} \equiv \left\|\theta^0 - \theta\right\|_{\mathfrak{g}}\left(\theta^0\right) \tag{4}$$

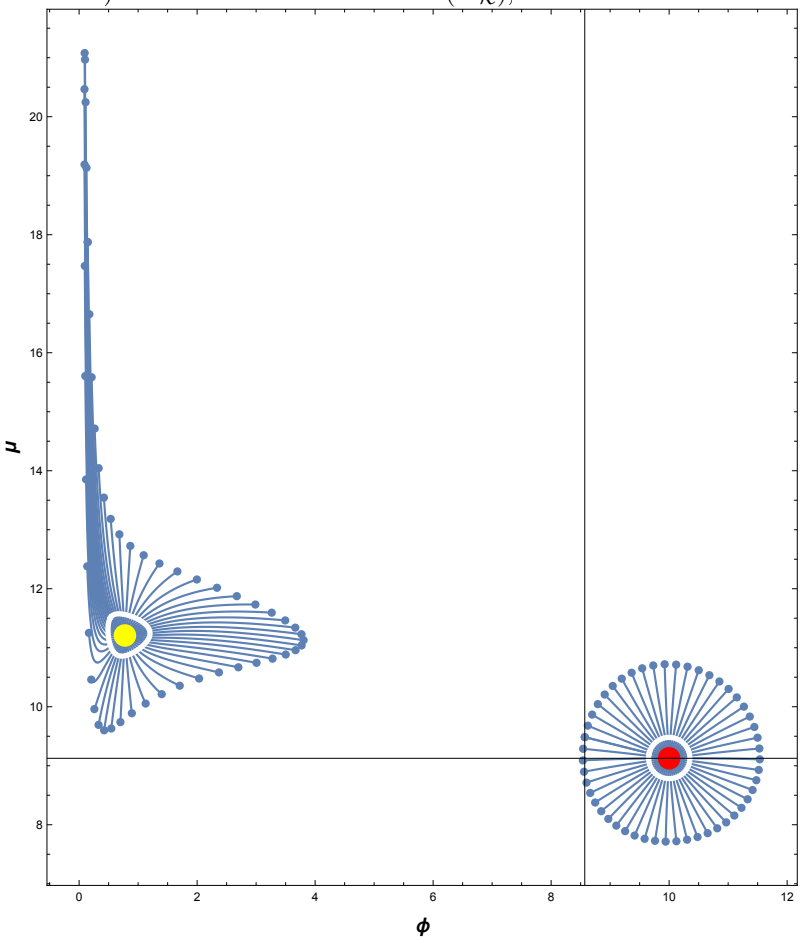
but this result can be refined, thanks to the following proposition.

Corollary 5. *If ϱ is small enough (for instance: $\varrho \leq \bar{\varrho}(\theta^0)$), the metric sphere $S_{\mathcal{R}}(\theta^0, \varrho) := \{\theta \in \mathfrak{M} : D_{\mathcal{R}}(\theta^0, \theta) = \varrho\}$ can be **isometrically identified** with the centered ellipse $\mathcal{E}_{\theta^0}(\varrho)$ of radius $\varrho := \|\theta^0 - \theta\|_{\mathfrak{g}}(\theta^0)$ drawn on $T_{\theta^0}\mathfrak{M}$.*

As an illustration, we displayed on Figure 1 the exponential spray $\exp_\theta(\mathcal{E}_\theta(\varrho)) = S_{\mathcal{R}}(\theta, \varrho)$ for two NB distributions (in the Chua and Ong (2013) parametrization) : $\theta^1 = \{0.7767, 11.2078\}$ (left point, a rather aggregative distribution) and $\theta^2 = \{10., 9.12624\}$ (right point, a bell-shaped distribution). We can see on this figure that for $\varrho = 0.3$ and $\varrho = 1.5$, the geodesics emanating from θ^2 draw circles, while the situation is very different for θ^1 : for $\varrho = 1.5$, the ball is dramatically anisotropic, very far from an ellipse, while for $\varrho = 0.3$ it is nearly spherical.

Consider now the application $\mathcal{T} : \Theta \to \mathfrak{M}$ associating to θ the probability $\mathfrak{L}^\theta$. Suppose first that the Fréchet- Darmois- Cramer- Rao assumptions Rao (2015, 1966) are fulfilled by the family $\left\{\mathfrak{L}^\theta : \theta \in \Theta\right\}$ and, second, that $\widehat{\theta}_N$ is an unbiased first-order efficient estimator of θ, based on some N-sample

Figure 1: Sampled metric balls (50 geodesics) of radius 1.5 (curves) and 0.35 (white contours) for two distributions of $NB(D_{\mathcal{R}})$; see comments in the text.



obeying $\mathfrak{L}^{\theta^0}$. Then, $\mathfrak{g}(\theta^0) \approx V\left(\widehat{\theta}_N\right)^{-1}$ (inverse of the variance matrix). If in addition $\varrho \leq \bar{\varrho}(\theta^0)$, because of Corollary 5, the image $\mathfrak{T}(\mathcal{E}_{\theta^0}(\varrho))$ can be identified with the metric sphere of radius ϱ centered on $\mathfrak{L}^{\theta^0}$. But $\bar{\varrho}(\theta^0)$ is unknown in general; so, what if $\left\|\theta^0 - \widehat{\theta}_N\right\|_{\mathfrak{g}}(\theta^0)$ is too large ($\gg \bar{\varrho}(\theta^0)$)? This is an important issue, in connection with the χ^2 GOF test (see for instance Rao (2015); Menendez et al. (1995); Cubedo and Oller (2002); Ilea et al. (2015)) associated with Formula 4: if a N-sample of counts stems from $\mathfrak{L}^{\theta^0}$, $N\, D^2_{\mathcal{R}}\left(\theta^0, \widehat{\theta}_N\right)$ should asymptotically obey $\chi^2_{(p)}$.

3. The special case of $NB(D_{\mathcal{R}})$

There is a large number of parametrizations for the NB distribution, and the most classical one is probably

$$P(X = j; (\phi, p)) = \left(\begin{array}{c} \phi + j - 1 \\ \phi - 1 \end{array} \right) p^j (1-p)^\phi \; j \geq 0 \tag{5}$$

with $(\phi, p) \in \mathbb{R}^+ \times]0, 1[$. Fisher used the following one Fisher et al. (1943); Bliss and Fisher (1953):

$$P(X = j) = \frac{\mathfrak{P}^{\mathfrak{n}}}{(1+\mathfrak{P})^{j+K}} \left(\begin{array}{c} \phi + j - 1 \\ \phi - 1 \end{array} \right) \tag{6}$$

with $(\phi, \mathfrak{P}) \in \mathbb{R}^+ \times \mathbb{R}^+$. In this formulation, specially relevant in Ecology, the so-called "index parameter" ϕ (denoted K by Fisher and many other authors) is an intrinsic parameter of the species, while $\mathfrak{P}$ "depends on the efficiency of the trap" for the species (Rao, 1971). This fact was mentioned by Fisher et al. (1943) and Anscombe (1950), who underscored that the "efficiency of the trap" must include time of exposure.

Nevertheless, because of its orthogonality, we chose instead the parametrization used by Chua and Ong (2013):

$$P\left(X=j;(\phi,\mu)\right)=\left(\begin{array}{c}\phi+j-1\\ j\end{array}\right)\left(\frac{\mu}{\mu+\phi}\right)^{j}\left(1-\frac{\mu}{\mu+\phi}\right)^{\phi},\ j\geq 0 \quad (7)$$

$(\phi,\mu)\in\mathbb{R}^{+}\times\mathbb{R}^{+}$; here, μ denotes the mean of the distribution.

3.1. χ^2 GOF testing in $NB(D_{\mathcal{R}})$

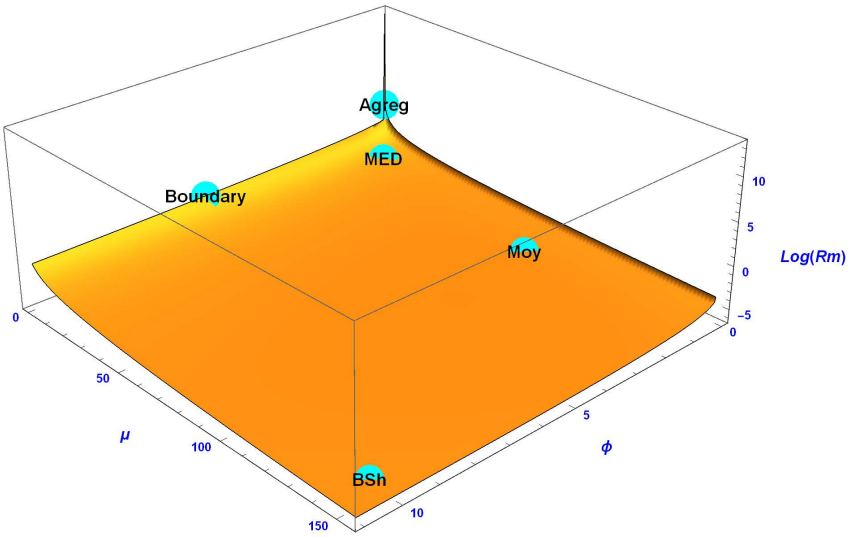
We will now investigate the feasibility of the GOF tests: $\mathfrak{L}^{\hat{\theta}}\stackrel{?}{=}\mathfrak{L}^{\theta^0}$ from a geometrical point of view. Consider as in Section 2.2 some $\theta^0=(\phi^0,\mu^0)\in\Theta:=\mathbb{R}^{+}\times\mathbb{R}^{+}$ and, for $\alpha<1$, the ellipse $\mathcal{E}_{\theta^0}(\alpha,N):=\mathcal{E}_{\theta^0}\left(\chi^2_{(2)}(\alpha)/N\right)$, where N is the sample size and $\chi^2_{(2)}(\alpha)$ denotes the quantile of order α of $\chi^2_{(2)}$ (see Figures 4 and 8). Is its image $\mathcal{T}\left(\mathcal{E}_{\theta^0}(\alpha,N)\right)$ a metric sphere for usual values of α (0.1, 0.5, 0.95,...) and any $\mathfrak{L}^{\theta^0}$? If the answer is negative for some $\alpha\in]0,1[$, the critical locus of the GOF test (the ellipse $\mathcal{E}_{\theta^0}(\alpha,N)$ whose interior is the acceptance region) will be anisotropic, *i.e.* there will be at least a pair of probabilities $\left(\mathfrak{L}^{\underline{\theta}},\mathfrak{L}^{\bar{\theta}}\right)$ with identical probability of rejection $1-\alpha$ such that $D_{\mathcal{R}}\left(\mathfrak{L}^{\theta^0},\mathfrak{L}^{\underline{\theta}}\right)<D_{\mathcal{R}}\left(\mathfrak{L}^{\theta^0},\mathfrak{L}^{\bar{\theta}}\right)$. In such cases the GOF test of level $1-\alpha$ is invalid, and one should choose a greater level (Proposition 4 guaranties the existence of some maximal valid α, depending on θ^0).

3.1.1. The test distributions used

We put in limelight five test distributions:

- "BSh" (for "bell-shaped"), which is $NB(10,144.3)$
- "Moy" $=NB(1.193,87.268)$ is the mean of a bivariate distribution fitting the parameters of non-aggregative species found in some habitat Manté et al. (2016)

Figure 2: The natural logarithm of the Riemannian measure of $NB(D_{\mathcal{R}})$ displayed in the parametrization (7) on a neighborhood of the origin.



- "MED"=$NB\left(0.7767, 11.2078\right)$ is the spatial median Serfling (2004) of the same sample of parameters (the left point on Figure 1)

- "Agreg"=$NB\left(0.01, 0.1443\right)$ corresponds to a theoretical aggregative species (first borderline case: $\phi \approx 0^{+}$)

- "Boundary"=$NB\left(6, 0.05\right)$, designed for investigating the case $\mu \approx 0^{+}$ (second borderline case).

These distributions are depicted from two different manners: as members of $NB(D_{\mathcal{R}})$ on Figure 2, and as probability densities on Figure 3.

Remark 1. The first four distributions were exemplified in (Manté et al., 2016, Appendix 2), but the parametrization used was (6). One can guess half from Figure 2 that the value at the origin of the Riemannian measure, $Rm\left(0, 0\right)$ cannot be consistently defined, as it is proven in Appendix 5.1. Note in addition that both the borderline cases ($\phi \approx 0^{+}$ and $\mu \approx 0^{+}$) correspond to situations where huge samples would be necessary to observe organisms (very small mean count or very large number of zeroes).

3.1.2. Regularly sampled confidence ellipses

Following Critchley and Marriott (2016), we fixed as a theoretical sample size $N = 100$.

For each one of the test distributions above, we determined a set $\mathcal{SE}_{\theta^0}\left(\alpha, N; K\right)$ of $K = 20$ regularly-sampled points upon $\mathcal{E}_{\theta^0}\left(\alpha, N\right) \subset \Theta$, for $\alpha \in \{0.1,\ 0.99\}$. By "regularly-sampled", we mean that the arc length between two neighbor points should be approximately $Length\left(\mathcal{E}_{\theta^0}\left(\alpha, N\right)\right)/K$, where $Length\left(\mathcal{E}_{\theta^0}\left(\alpha, 100\right)\right)$ is approximated by the Gauss-Kummer formula. Regularly-sampled ellipses associated with BSh and $\alpha \in \{0.1,\ 0.99\}$ are represented on Figure 4.

Figure 3: Five typical distributions (see also Figure 2). The index Δ measures the gap between each distribution and the closest Poisson one (see Manté (tted)), which is superposed to the NB distribution when $\Delta < 0.05$; vertical bars are associated with NB probabilities. In the case of the unique Poisson-like distribution Manté (tted),"Boundary", an orange curve is associated with the Poisson distribution.

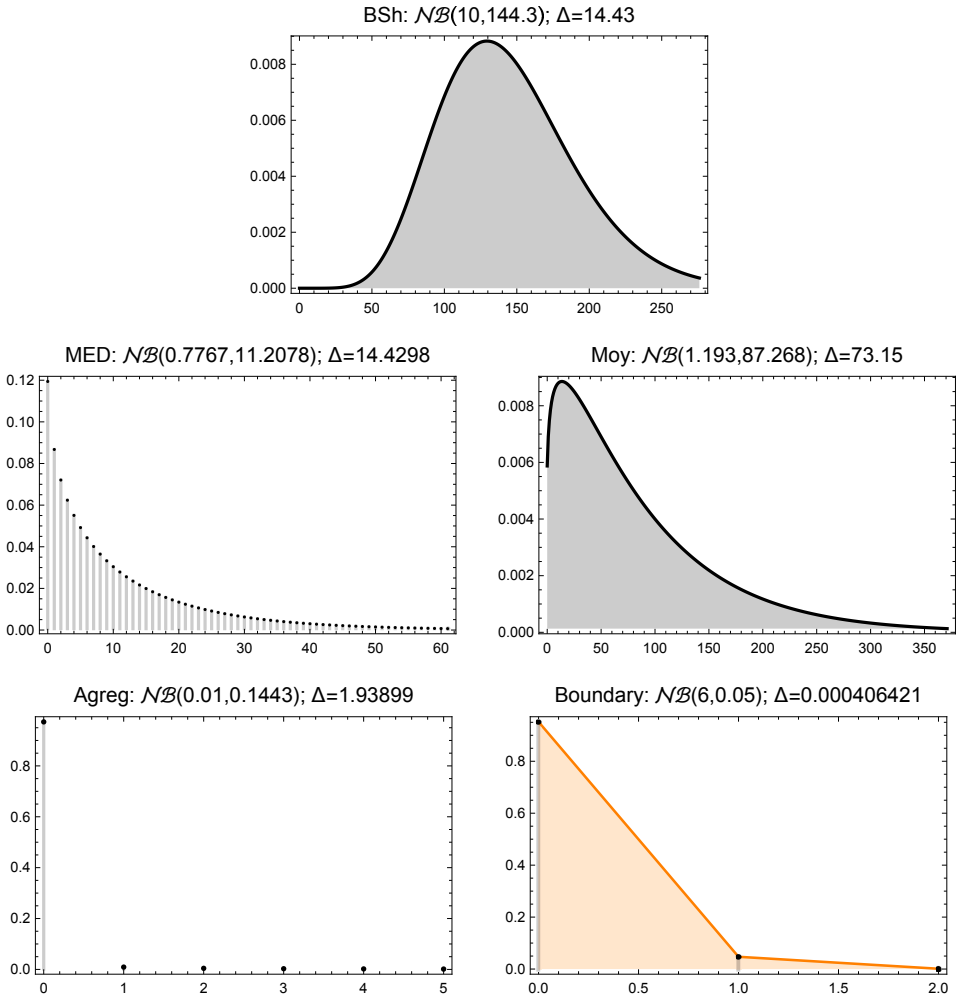


Figure 4: The unit ellipse (gray), $\mathcal{SE}_{\theta^0}(0.99, 100; 20)$ (black) and $\mathcal{SE}_{\theta^0}(0.1, 100; 20)$ in yellow (color figure online), for $\theta^0 = BSh$.

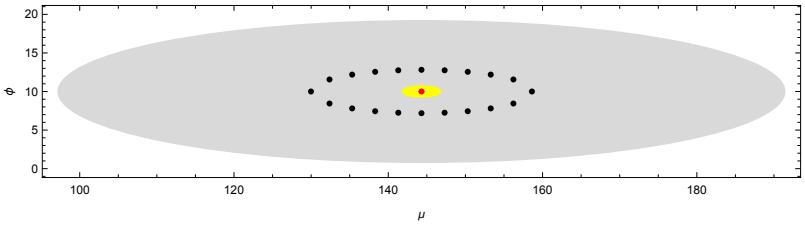
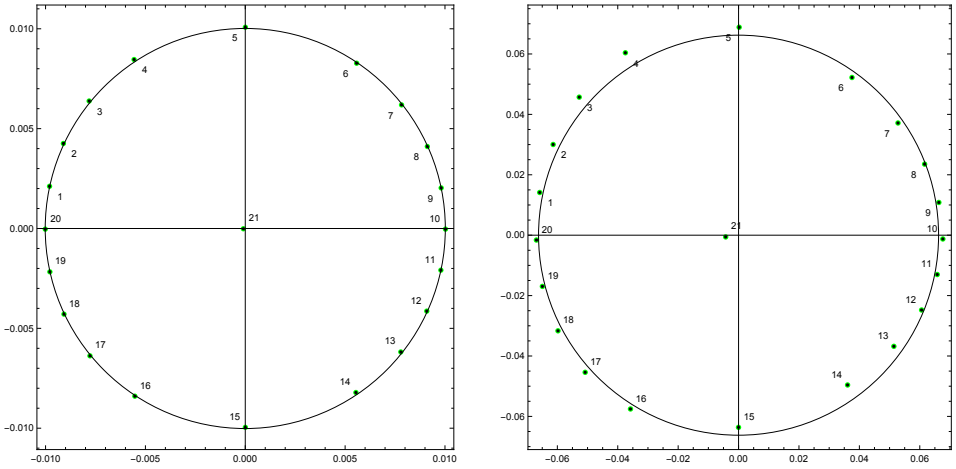


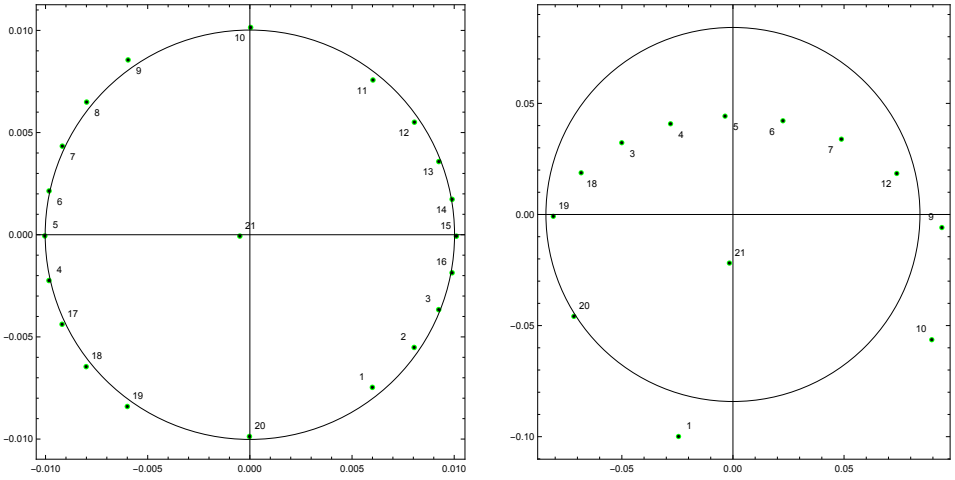
Figure 5: The "BSh" distribution (processed table: Δ^{Exp}). Left (resp. right) panel: representation through MDS of $\mathcal{T}(\mathcal{SE}_{BSh}(0.1, 100; 20))$ (resp. $\mathcal{T}(\mathcal{SE}_{BSh}(0.99, 100; 20))$).



3.1.3. Plotting χ^2 spheres around reference probabilities

For each pair on points $\{\theta^i, \theta^j\} : j \neq i$ of each $\mathcal{SE}_{\theta^0}(\alpha, N; K)$, $D_{\mathcal{R}}(\theta^i, \theta^j)$ was computed according to the methods detailed by Manté (tted); in addition each $D_{\mathcal{R}}(\theta^i, \theta^0)$ was computed too. This gave rise to a 21×21 table $\Delta(\theta^0, \alpha, N; K)$ submitted to MDS. But there is no guarantee that some $\Delta(\theta^0, \alpha, N; K)$ could be **exactly** represented in an Euclidean space by classical MDS (and besides by a sphere) excepted if α is small enough (Corollary 5). It is thus advisable to pre-process each table before performing

Figure 6: The "Boundary" distribution (processed table: Δ^{Exp}). Left (resp. right) panel: representation of $\mathcal{T}\left(\mathcal{SE}_{Boundary}\left(0.1, 100; 20\right)\right)$ (resp. $\mathcal{T}\left(\mathcal{SE}_{Boundary}\left(0.99, 100; 20\right)\right)$).



MDS.

There are several methods for making a distance matrix like Δ Euclidean (*i.e.* find a close distance matrix which can be exactly represented in an Euclidean space) - see for instance Benasseni et al. (2007). The simpler one is the Additive Constant (AC) one Caillez (1983), consisting in adding an optimal positive perturbation c^* to all the extra-diagonal terms of Δ^2. But other pre-processing methods are worth considering Benasseni et al. (2007): one can search for the smallest positive γ_0 such that the power Δ^γ is Euclidean for $\gamma \leq \gamma_0$ Joly and Le Calvé (1986), or the smallest positive γ^* such that $1 - e^{-\gamma \Delta}$ is Euclidean (Exp method). It is noteworthy that both these transformations belong to the class of Schoenberg transformations introduced in Data Analysis by Bavaud (2011). We chose the last one, the Exp method .

Notice now that the representation of $\Delta\left(\theta^0, \alpha, N; K\right)$ should theoretically

consist in a regularly K-sampled circle $SC(\rho, K)$. What is the radius of this circle?

Proposition 6. *The representation through MDS of the sampled circle* $SC(\rho, K)$ *is* $SC\left(\frac{\rho}{\sqrt{K}}, K\right)$.

Proof. see Appendix 5.2 □

As a consequence, the representation of each $\Delta(\theta^0, \alpha, N; K)$ had to be compared to the circle of radius $\sqrt{\chi^2_{(2)}(\alpha)/N K}$. We gathered in Tables 1 and 2 several quality criteria for representation of confidence ellipses through MDS. In the cases investigated here, we found that the solution associated with the *Exp* method, Δ^{Exp} was by far the best one. In the last column of each table, we added a visual appreciation about the sphericity of the representation of $\mathcal{T}(\mathcal{SE}_{\theta^0}(\alpha, N; K))$ in the first principal plane of MDS: **G**ood, **D**eformed or **B**ad; the reader can examine such sampled metric spheres on Figures 5 and 6, or in Manté (2017). On these plots, points sampled on each ellipse were labeled $\{1, \cdots, 20\}$, while the 21^{st} point corresponded to the reference distribution $\mathfrak{L}^{\theta^0}$, which should theoretically occupy the center of the circle. Clearly, for $\alpha = 0.1$ (see Table 1), the representation was always rather good (even in the case of Agreg; see also Figure 3 of Manté (2017)) and the distortions associated with Δ^{Exp} were much smaller than those associated with the traditional Additive Constant method, Δ^{AC} .

With a much lower level ($\alpha = 0.99$), there were smaller discrepancies in distortion between Δ^{Exp} and Δ^{AC} (see Table 2). The spheres were more or less deformed (see the right panel of Figure 5 or Figure 2 of Manté (2017)) , or squarely destroyed in the case of borderline distributions like "Agreg" or "Boundary" (see the right panel of Figure 6, or Figure 3 of Manté (2017)).

Table 1: Quality criteria of the representation for $N = 100$ and $\alpha = 0.1$; MedDist() denotes the median distortion.

Distribution	$c^*/\overline{\Delta^2}$ (%)	$MedDist(\Delta^{AC})$	$MedDist(\Delta^{Exp})$	Spherical?
BSh	0.00615466	0.0000254661	8.18752*10^-6	G
Moy	0.00697764	0.0000249921	7.72381*10^-6	G
MED	0.00303372	0.0000106903	7.65055*10^-6	G
Agreg	0.315966	0.00133039	8.72335*10^-6	D
Boundary	0.0855695	0.000387506	8.64581*10^-6	G

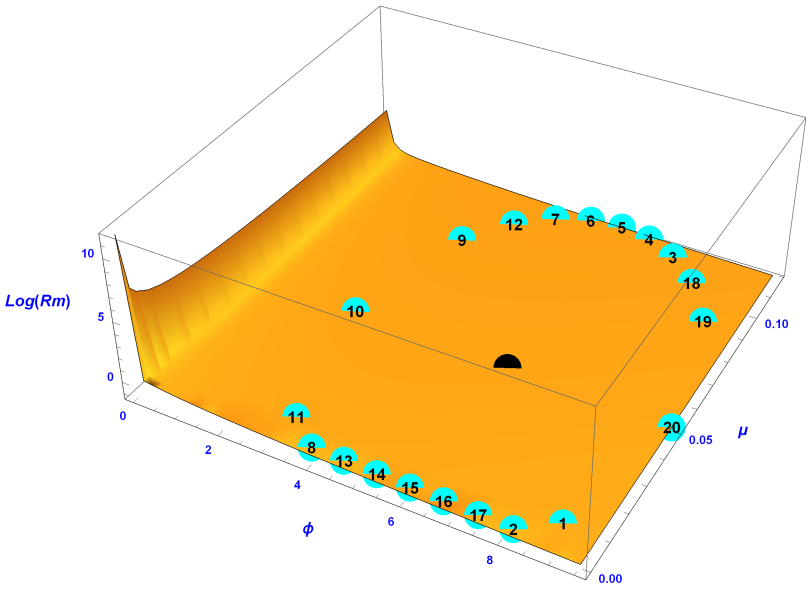
Table 2: Quality criteria of the representation for $N = 100$ and $\alpha = 0.99$; MedDist() denotes the median distortion.

Distribution	$c^*/\overline{\Delta^2}$ (%)	$MedDist(\Delta^{AC})$	$MedDist(\Delta^{Exp})$	Spherical?
BSh	0.246803	0.000989106	0.000360119	D
Moy	0.371856	0.00160617	0.00198782	D
MED	0.168661	0.000723894	0.000788534	D
Agreg	0.0719816	0.000582904	0.000280619	B
Boundary	2.6131	0.0152534	0.00688024	B

Remark 2. Comparing Tables 1 and 2, one can see that both median distortions and $c^*/\overline{\Delta^2}$ were generally smaller for $\alpha = 0.1$, which is quite natural: the more α is small, the more the geometry is Euclidean in the neighborhood of the reference probability.

Let's now examine further the "Boundary" case. On Figure 7, we plotted the Riemannian measure $Rm(\theta)$ for values sampled on the closed ellipse $\mathcal{E}_{Boundary}(0.99, 100)$, cut by the axis $\mu = 0$. The points far from the frontier are $\{3, \cdots, 7, 12, 18, \cdots, 20\}$, while points close to the $\phi = 0$ borderline are $\{9, 10, 11\}$ and points on the $\mu = 0$ borderline (or very close to it) are $\{1, 2, 8, 11, 13, \cdots, 17\}$. Most points close to the $\mu = 0$ borderline, $\{2, 8, 11, 13, \cdots, 17\}$ are missing on the right panel of Figure 6: this is due to time limits encountered in the computations, causing the incompleteness of the table $\Delta(\theta^0, \alpha, N; K)$. On the contrary, on this figure, most of the points

Figure 7: Plot of the Riemannian measure inside $\mathcal{SE}_{Boundary}(0.99, 100)$ (truncated ellipse); the reference distribution $\theta^{Boundary}$ is represented by the black point.



far from boundaries are inside the circle, and only 19 and 20 are on the circle. In addition, even the reference distribution (point 21) is ill-placed. Indeed, as Critchley and Marriott (2016); Marriott et al. (2016) noticed, one should always expect such problems near the frontier $\left(\Theta - \mathring{\Theta}\right)$, because usual statistical methods "break down" in such regions. Incidentally, a similar phenomenon happens in large neighborhoods of the "MED" distribution (see Figure 1).

3.2. The χ^2 two-sample test: processing some data of Bliss and Fisher (1953)

Remember now that the Kullback-Leibler divergence between two close distributions verifies Kass (1989): $K(\mathcal{D}_i, \mathcal{D}_j) \approx \frac{1}{2} D^2_{\mathcal{R}}(\mathcal{D}_i, \mathcal{D}_j)$; as a consequence $\frac{N}{2} D^2_{\mathcal{R}}(\mathcal{D}_i, \mathcal{D}_j)$ should asymptotically obey $\chi^2_{(2)}$ when $\mathcal{D}_i \approx \mathcal{D}_j$ Ilea et al. (2015). The application of this result to the two-sample test: $\mathfrak{L}^{\hat{\theta}_i} \stackrel{?}{=} \mathfrak{L}^{\hat{\theta}_j}$

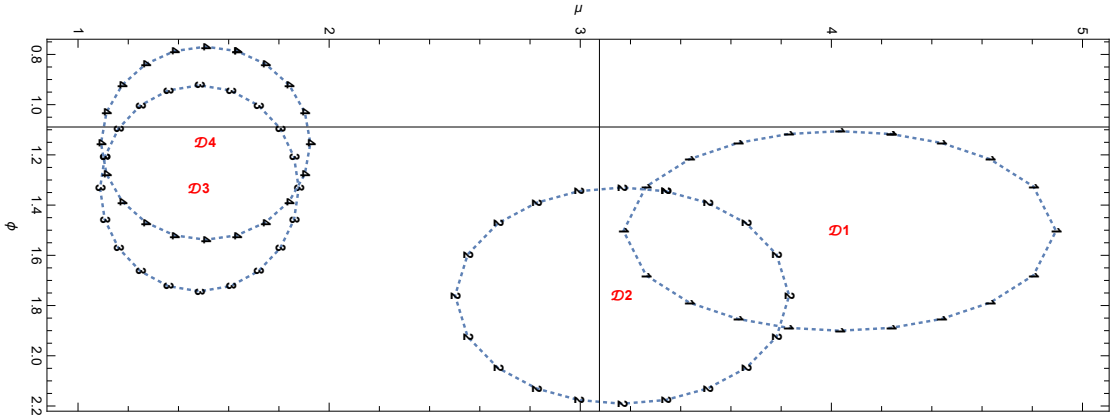
is straightforward; we will now exemplify this fact on data from the literature.

Bliss and Fisher (1953) fitted by NB probabilities four distributions of insect pest (the corn borer *Pyrausta nubilalis*, for the connoisseurs) obtained from field experiment. In this experiment, 15 plots were determined and 8 hills of corn were randomly selected into each plot and differently treated. Afterward, the number of hills with $k = 0, \cdots, 5, \cdots$ borers was determined, for each treatment - it is noteworthy that "Treatment 1" was the **untreated** case (control) and that the number of counts was $N = 15 * 8 = 120$ for each treatment. As Bliss and Fisher (1953) noticed, the composite distribution from the 15 plots could likely be NB (as Gamma-Poisson mixtures), because the level of infestation varied significantly from plot to plot. In their paper, they showed that the NB model fits very well these four distributions (high P Values of the χ^2 test), unlike the Neyman type A "contagious" distribution.

We denoted $\mathcal{D}_i$ the parameters (or the probability itself) of the NB distribution associated with "Treatment i". Using the above χ^2 test in $NB(D_{\mathcal{R}})$, we found that $\mathcal{D}_1$ and $\mathcal{D}_2$ are significantly different from $\mathcal{D}_3$ and $\mathcal{D}_4$ (P Values $< 10^{-6}$), while $\mathcal{D}_2$ is slightly different from $\mathcal{D}_1$ (P Values ≈ 0.08) and $\mathcal{D}_3$ and $\mathcal{D}_4$ are quite similar (P Values ≈ 0.77). We can thus conclude (with a slight delay) from these tests that treatments 3 and 4 were equally and strongly efficient, while treatment 2 had only a slight efficiency (remember $\mathcal{D}_1$ was a control). Clearly, treatments 3 and 4 diminished the mean number of borers, while treatment 2 did not (but the distribution of borers was probably altered by this treatment).

In addition, we plotted on Figure 8 the corresponding regularly sampled confidence ellipses of level 0.95 centered on the four distributions. We rep-

Figure 8: The four confidence ellipses $\mathcal{SE}_{\mathcal{D}_i}(0.95, 120; 20)$; $\mathcal{D}_i$ corresponds to the parameters of the i^{th} data set (treatment).



resented on Figure 9 the first principal plane (more than 98% of the total variance) issued from MDS of the image of these data transformed under $\mathcal{T}$. Notice that the clouds centered on $\mathcal{D}_2$ and $\mathcal{D}_1$ are slightly deformed. With a much smaller confidence level (0.1), we got the smaller spheres displayed on Figure 10; the percentage of variance was higher (about 99.93) and the clouds are perfectly spherical (illustration of Corollary 5).

Remark 3. Consider any pair of distributions, $(\mathcal{D}_i, \mathcal{D}_j)$, and the straight line $\lambda_{i,j}(t) := t\,\mathcal{D}_i + (1-t)\,\mathcal{D}_j$ linking them. We have necessarily that

$$D_{\mathcal{R}}(\mathcal{D}_i, \mathcal{D}_j) \leq D_{\Lambda}(\mathcal{D}_i, \mathcal{D}_j) := \int_0^1 \sqrt{\dot{\lambda}_{i,j}^{\ t}(t) \,.\mathfrak{g}\left(\lambda_{i,j}(t)\right).\dot{\lambda}_{i,j}(t)}dt.$$

It is noteworthy that we had for any pair of distributions: $D_{\Lambda}(\mathcal{D}_i, \mathcal{D}_j) \leq 1.004\, D_{\mathcal{R}}(\mathcal{D}_i, \mathcal{D}_j)$. Thus, the curvature is very close to zero in this region! That is likely why we obtained excellent Euclidean approximations on Figures 9 and 10.

4. Conclusion

Notice first that while this work focuses on the Riemannian manifold $NB(D_{\mathcal{R}})$ of negative binomial distributions, it could be extended to other

Figure 9: Metric spheres and images of the ellipses transformed under $\mathcal{T}$ (chosen table: Δ^{Exp}).

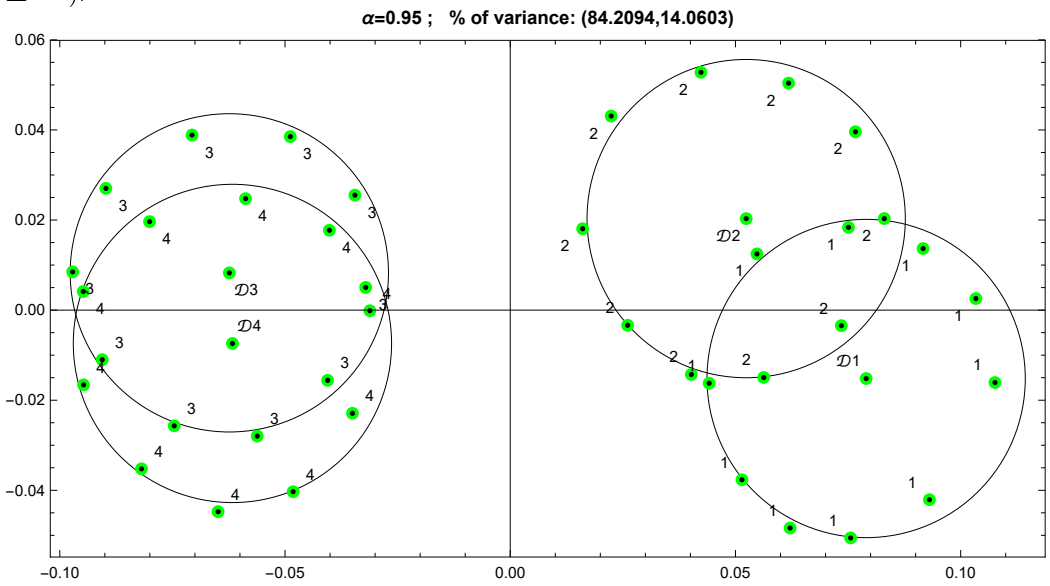
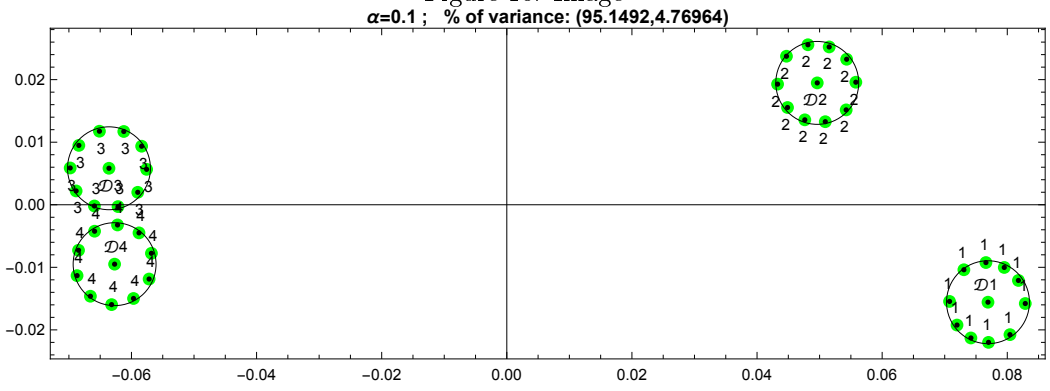


Figure 10: Image



families of probability distributions.

In the setting of $NB(D_{\mathcal{R}})$, with regard to the χ^2 GOF test, we have shown that:

- in the parameters space Θ, the critical locus of level $1-\alpha$ associated with the GOF test $\mathfrak{L}^{\hat{\theta}} \stackrel{?}{=} \mathfrak{L}^{\theta^0}$ is an ellipse $\mathcal{E}_{\theta^0}(\alpha)$ which depends (center, eccentricity) on θ^0
- the image of $\mathcal{E}_{\theta^0}(\alpha)$ under $\mathcal{T} : \Theta \to NB(D_{\mathcal{R}})$ associating to θ the probability $\mathfrak{L}^{\theta}$ is (theoretically) a metric sphere
- $\mathcal{T}(\mathcal{E}_{\theta^0}(\alpha))$ is (approximately) spherical only when $\mathcal{E}_{\theta^0}(\alpha)$ is far enough from the frontier of Θ - in other words, if $\mathcal{E}_{\theta^0}(\alpha) \cap \left(\Theta - \mathring{\Theta}\right) = \emptyset$.

When θ^0 is too close from a frontier of Θ, the GOF test $\mathfrak{L}^{\hat{\theta}} \stackrel{?}{=} \mathfrak{L}^{\theta^0}$ is not feasible for large values of α. Notice (see Figures 2 and 7) that in such cases, the values of $dV_{NB(D_{\mathcal{R}})}$ in some neighborhood $\mathfrak{U}(\theta^0)$ of θ^0 is very large (high local heterogeneity/generalized variance).

With regard to the χ^2 two-sample test: $\mathfrak{L}^{\hat{\theta}_i} \stackrel{?}{=} \mathfrak{L}^{\hat{\theta}_j}$, we have shown on an example from the literature that its implementation is straightforward . Since the NB distributions studied were far from the frontier of the parameters space, we found that the images under $\mathcal{T}$ of confidence ellipses were metric spheres (possibly slightly deformed, depending on the chosen confidence level).

5. Appendices

5.1. Behavior at the origin of the Riemannian measure

It can be shown after some computation that in the case of $NB(D_{\mathcal{R}})$, the Riemannian measure expressed in the parametrization (7) is given by:

$$Rm\left(\phi,\mu\right)=\sqrt{-\frac{\phi(\mu+\phi)\left(\left(\frac{\phi}{\mu+\phi}\right)^{\phi}-1\right)\psi^{(1)}(\phi)+\mu}{\mu(\mu+\phi)^{2}}} \tag{8}$$

where $\psi^{(1)}(\phi)$ denotes the derivative of the Trigamma function Abramowicz and Stegun (2002).

Notice first that ϕ and μ are linked by the relation: $\mu=\mathfrak{P}\,\phi$, where the hidden Fisher's parameter $\mathfrak{P}$ depends on the efficiency of the trap for the species. There are four routes (free of indeterminate) towards $(0,0)$:

1. $\phi=\frac{\mu}{\mathfrak{P}}\searrow 0^{+}$, which implies either $(\mu\searrow 0^{+})\wedge(+\infty>\mathfrak{P}>0)$ (a) or $(\mu=O\left(1\right))\wedge(\mathfrak{P}\nearrow+\infty)$ (b)
2. $\mu=\mathfrak{P}\,\phi\searrow 0^{+}$, which implies either $(\phi\searrow 0^{+})\wedge(\mathfrak{P}=O\left(1\right))$ (a) or $(\mathfrak{P}\searrow 0^{+})\wedge(\phi=O\left(1\right))$ (b).

Since $\phi\searrow 0^{+}\iff\mu\searrow 0^{+}$ when $\mathfrak{P}>0$, the cases 1 (a) and 2 (a) are equivalent.

5.1.1. Case 2 (a)

The first terms of the Taylor series near 0 of $Rm\left(\phi,\mathfrak{P}\,\phi\right)$ with respect to ϕ is:

$$\sqrt{\frac{-\mathfrak{P}+\mathfrak{P}\left(-\log\left(\frac{1}{\mathfrak{P}+1}\right)\right)-\log\left(\frac{1}{\mathfrak{P}+1}\right)}{\mathfrak{P}(\mathfrak{P}+1)^{2}\phi^{2}}}$$

and the function $(-\mathfrak{P}-Log[1/(1+\mathfrak{P})]-\mathfrak{P}Log[1/(1+\mathfrak{P})])/(\mathfrak{P}((1+\mathfrak{P})^2))$ is strictly positive for any $\infty > \mathfrak{P} > 0$; thus we have generally that

$$\lim_{\phi \to 0^+} (Rm(\phi, \mathfrak{P}\,\phi)) = +\infty$$

excepted if $\mathfrak{P} = 0$ or $\mathfrak{P}$is infinite (indeterminate cases).

5.1.2. Case 1 (b)

Suppose now that $\phi \searrow 0^+$ and $\mathfrak{P} \nearrow +\infty$ in such a way that $\mathfrak{P}\,\phi = \mu$ is $O(1)$. Then, $Rm\left(\frac{\mu}{\mathfrak{P}}, \mu\right) = \sqrt{-\frac{\mathfrak{P}^2+\mu(\mathfrak{P}+1)\left(\left(\frac{1}{\mathfrak{P}+1}\right)^{\frac{\mu}{\mathfrak{P}}}-1\right)\psi^{(1)}\left(\frac{\mu}{\mathfrak{P}}\right)}{\mu^2(\mathfrak{P}+1)^2}}$, and $\lim_{\mathfrak{P}\to+\infty} Rm\left(\frac{\mu}{\mathfrak{P}}, \mu\right) = +\infty$.

5.1.3. Case 2 (b): $\mu = \mathfrak{P}\,\phi \searrow 0^+$, $\mathfrak{P} \searrow 0^+$, $0 < \phi = O(1)$

Since $0 < \phi = O(1)$, $\mu = \mathfrak{P}\,\phi \searrow 0^+ \iff \mathfrak{P} \searrow 0^+$ and we can merely consider $\lim_{\mu \to 0^+} Rm(\phi, \mu)$. Inserting the Taylor series with respect to μ

$$\phi(\mu+\phi)\left(\left(\frac{\phi}{\mu+\phi}\right)^{\phi}-1\right) = -\mu\,\phi^2 + \frac{1}{2}(\phi-1)\,\mu^2\phi + O[\mu]^3$$

(valid for $(\phi > 0) \wedge (\mu \approx 0)$) into equation (8), we obtain after simplifications that

$$\lim_{\mu \to 0^+} Rm(\phi, \mu) = \frac{\sqrt{\phi^2\psi^{(1)}(\phi)-1}}{\phi}. \tag{9}$$

Then, because the Trigamma function verifies (Abramowicz and Stegun, 2002, p. 260) $\psi^{(1)}(\phi) = \frac{1}{\phi^2} + \sum_{k\geq 1} \frac{1}{(\phi+k)^2} : \phi \notin \{0, -1, \cdots -n, \cdots\}$, we have that for $\phi \in]0,1[$, $Rm(\phi, 0) = \sqrt{\sum_{k\geq 1} \frac{1}{(\phi+k)^2}}$, and (Abramowicz and Stegun, 2002, p. 807)

$$\lim_{\phi \to 0^+}\left(\lim_{\mu \to 0^+} Rm(\phi, \mu)\right) = \sqrt{\zeta(2)} = \pi/\sqrt{6}.$$

Consequently $Rm(0,0)$ cannot be consistently defined $\square$

Remark 4. Both the conditions $\mu \approx 0^+$ or $\phi \approx 0^+$ correspond to situations where huge samples would be necessary to observe something (very small mean count, or very large number of zeroes. Finally, notice the first limit case, $\mu = 0$, corresponds to the null distribution, while the second one, $\phi \approx 0^+$, is associated with situations where the data should be probably better fitted by the log-series distribution of Fisher et al. (1943).

5.2. Multidimensional Scaling of $SC(\rho, K)$

Consider the squared distance matrix $\mathcal{D}_K$ obtained by traveling along some "regularly-K-sampled" circle $SC(\rho, K)$ of fixed radius ρ. Clearly, $\mathcal{D}_K$ is a circulant matrix, completely determined by its first column $\mathcal{D}_K^{(1)} = \mathcal{C} := (\mathcal{C}_1, \cdots, \mathcal{C}_{K-1}, 0)$, where K is the sample size, $\mathcal{C}_1$ is the square of the chord distance between the (arbitrary) first point and its nearest neighbor, *etc.* Each column of $\mathcal{D}_K$ is obtained by cyclic permutation of $\mathcal{C}$, with the column index as an offset. Let's compute the spectrum of the operator

$$W_K := -\frac{1}{2} A_K . \mathcal{D}_K . A_K$$

where $A_K := I_K - \frac{1}{K} 1_K \otimes 1_K$. Since $\mathcal{D}_K$ and A_K are circulant, W_K has the same property; consequently (Brillinger, 1981, Section 3.7) its eigenvalues only depend on the discrete Fourier transform (we adopted the same convention as Brillinger (1981))

$$\mathcal{F}_d \left(W_K^{(1)} \right) [p] := \sum_{k=1}^{K} W_k^1 \exp\left(-2\, i\, \pi\ (p-1)\ (k-1)\, /K \right);\ 1 \leq p \leq K$$

of the fist column $W_K^{(1)}$ of W_K, and its eigenvectors are columns of the matrix Ψ of general term

$$\psi_{j,k} := \frac{1}{\sqrt{K}} \exp\left(\frac{-2i\pi\, j\, k}{K} \right) :\ 0 \leq j \leq K-1,\ \ 0 \leq k \leq K-1.$$

After some algebra, it can be shown that **in our case**

$W_K^{(1)} = -\frac{1}{2}\left(\mathcal{D}_K^{(1)} - \frac{1}{K}\sum_{k=1}^{K}\mathcal{C}_k\right)$. Thus, the unique difference between $\mathcal{F}_d\left(W_K^{(1)}\right)$ and $\frac{-1}{2}\mathcal{F}_d\left(\mathcal{D}_K^{(1)}\right)$ is that the first one is centered, and the eigenvalues of W_K are given by the non-null coordinates of $\frac{-1}{2}\mathcal{F}_d(\mathcal{C})$. Observe now that since

$\mathcal{C}_m = \left(2\rho\sin\left(\frac{m\pi}{K}\right)\right)^2$, they are indeed given by the discrete Fourier transform of the function $\mathcal{C}(x) := (2\rho\sin(\pi x))^2$, which is strictly equivalent to the Fourier series $\widehat{\mathcal{C}}(m) := \frac{\rho^2}{2\pi}\int_{-\pi}^{\pi}\mathcal{C}(x)\ e^{-i\,m\,x}dx$ of $\mathcal{C}(x)$ because $\widehat{\mathcal{C}}$ is supported by $[-2,2]$ (Shannon-Kotelnikov theorem). Actually, $\widehat{\mathcal{C}}(m)$ is null excepted for three indices: $\widehat{\mathcal{C}}(0) = 2\rho^2$ and $\widehat{\mathcal{C}}(\pm 2) = -\rho^2$ (thus, $\mathcal{C}(x) = 2\rho^2 - \rho^2(\exp(-2\,i\,x) + \exp(2\,i\,x))$). As a consequence $\frac{-1}{2}\mathcal{F}_d(\mathcal{C})$ is null, at the exception of two frequencies whose coefficient are $\gamma := {}^{K\rho^2}/_2$. Now, classical MDS indeed consists Caillez (1983) in computing the spectral decomposition of the operator $W_K.D_P$, where P is the vector of weights attributed to the K "individuals" and D_P is the associated diagonal matrix. Of course, when $P = 1/K$, the eigenvector are the same as those of W_K, while the vector of eigenvalues is $\lambda := \gamma/K$. Finally, the representation $SC(\rho, K)$ is obtained by combining two columns of Ψ (K points on the circle of radius $1/\rho$) with the square root of the double eigenvalue $\lambda = {}^{\rho^2}/_2$. Thus we get a circle of radius $\sqrt{2\,\lambda}\,\frac{1}{\sqrt{K}} = \frac{\rho}{\sqrt{K}}$ □

Remark 5. It is **mandatory** for this proof that the circle has been regularly-sampled; otherwise, $\mathcal{D}_K$ wouldn't be circulant.

References

References

C. I. Bliss, R. A. Fisher, Fitting the Negative Binomial distribution to biological data, Biometrics 9 (1953) 176–200.

M. F. O'Neill, M. J. Faddy, Use of binary and truncated negative binomial modelling in the analysis of recreational catch data, Fisheries Research 60 (2003) 471–477.

L. Vaudor, N. Lamouroux, J.-M. Olivier, Comparing distribution models for small samples of overdispersed counts of freshwater fish, Acta Oecologica-International Journal of Ecology 37 (2011) 170–178.

C. Manté, S. O. Kidé, A.-F. Yao-Lafourcade, B. Merigot, Fitting the truncated negative binomial distribution to count data. A comparison of estimators, with an application to groundfishes from the Mauritanian Exclusive Economic Zone, Environmental and Ecological Statistics 23 (2016) 359–385.

R. A. Fisher, A. Corbet, C. B. Williams, The relation between the number of species and the number of individuals in a random sample of an animal population, J. anim. Ecol 12 (1943) 42–58.

F. J. Anscombe, Sampling theory of the negative binomial and logarithmic series distributions, Biometrika 36 (1950) 358–382.

C. R. Rao, Some comments on the logarithmic series distribution in the analysis of insect trap data., in: W. W. G. Patil, E. Pielou (Ed.), Statistical ecology, Vol. 1, Pensylvania State University Press, University Park, 1971, pp. 131–142.

D. G. Kendall, On some modes of population growth leading to R. A. Fisher's logarithmic series distribution, Biometrika 35 (1948) 6–15.

C. R. Rao, Information and the Accuracy Attainable in the Estimation of Statistical Parameters, Resonance-Journal of Science Education 20 (2015) 78–90.

C. R. Rao, Comment to Kass' paper, Statistical Science 4 (1989) 229–231.

K. M. Carter, R. Raich, W. G. Finn, A. O. Hero, FINE: Fisher Information Nonparametric Embedding, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 2093–U195.

G. Galanis, P. C. Chu, G. Kallos, Y.-H. Kuo, C. T. J. Dodson, Wave height characteristics in the north Atlantic ocean: a new approach based on statistical and geometrical techniques, Stochastic Environmental Research and Risk Assessment 26 (2012) 83–103.

C. T. J. Dodson, Some illustrations of information geometry in biology and physics, in: J. Leng, W. W. Sharrock (Eds.), Handbook of research on computational science and engineering, Engineering Science Reference, 2012, pp. 287–315.

M. Cubedo, A. Minarro, J. M. Oller, A dissimilarity based on relevant population features, Journal of Statistical Planning and Inference 143 (2013) 346–355.

I. Ilea, L. Bombrun, C. Germain, I. Champion, R. Terebes, M. Borda, Statistical Hypothesis Test for Maritime Pine Forest Sar Images Classification Based on the Geodesic Distance, in: 2015 IEEE International Geoscience and Remote Sensing Symposium (igarss), IEEE, New York, 2015, pp. 3215–3218.

C. Manté, S. O. Kidé, Approximating the Rao's distance between negative binomial distributions. Application to counts of marine organisms, in: Proceedings of COMPSTAT 2016, A. Colubi, A. Blanco and C. Gatu., Oviedo (Spain), 2016, pp. 37–47. URL: https://hal.archives-ouvertes.fr/hal-01357264.

R. E. Kass, The geometry of asymptotic inference, Statistical Science 4 (1989) 188–234.

M. Menendez, D. Morales, L. Pardo, M. Salicru, Statistical Tests Based on Geodesic Distances, Applied Mathematics Letters 8 (1995) 65–69.

M. Cubedo, J. M. Oller, Hypothesis testing: a model selection approach, Journal of Statistical Planning and Inference 108 (2002) 3–21.

M. Berger, A Panoramic View of Riemannian Geometry, Springer, Berlin, Heidelberg, 2003.

S.-i. Amari, H. Nagaoka, D. Harada, Methods of information geometry, number 191 in Translations of mathematical monographs, nachdr. ed., American Math. Soc. [u.a.], Providence, RI, 2007.

A. Gray, Modern differential geometry of curves and surfaces with Mathematica, 2nd ed ed., CRC Press, Boca Raton, 1998.

T. W. Anderson, An introduction to multivariate statistical analysis, Wiley series in probability and mathematical statistics, 2nd ed ed., Wiley, New York, 1984.

G. Lebanon, Metric learning for text documents, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (2006) 497–508.

K. C. Chua, S. H. Ong, Test of misspecification with application to negative binomial distribution, Computational Statistics 28 (2013) 993–1009.

C. R. Rao, Linear statistical inference and its applications, Wiley series in probability and mathematical statistics, 2. print ed., Wiley, New York, 1966.

R. Serfling, Nonparametric multivariate descriptive measures based on spatial quantiles, Journal of Statistical Planning and Inference 123 (2004) 259–278.

C. Manté, Computing the Rao's distance between negative binomial distributions. Application to Exploratory Data Analysis, Journal of Multivariate Analysis (submitted). URL: `https://hal.archives-ouvertes.fr/hal-02130199`.

F. Critchley, P. Marriott, Computing with Fisher geodesics and extended exponential families, Statistics and Computing 26 (2016) 325–332.

J. Benasseni, M. B. Dosse, S. Joly, On a General Transformation Making a Dissimilarity Matrix Euclidean, Journal of Classification 24 (2007) 303–304.

F. Caillez, The analytic solution of the Additive Constant problem, Psychometrika 48 (1983) 305–308.

S. Joly, G. Le Calvé, Etude des puissances d'une distance, Statistique et Analyse des Données 11 (1986) 30–50.

F. Bavaud, On the Schoenberg transformations in data analysis: Theory and illustrations, Journal of Classification 28 (2011) 297–314.

C. Manté, The Rao's distance between negative binomial distributions for Exploratory Analyses and Goodness-Of-Fit Testing, in: Proceedings of ISI2017, Marrakech (Marroco), 2017, pp. X–X+5. URL: https://hal.archives-ouvertes.fr/hal-01632444.

P. Marriott, R. Sabolova, G. Van Bever, F. Critchley, The Information Geometry of Sparse Goodness-of-Fit Testing, Entropy 18 (2016) 421.

M. Abramowicz, I. A. Stegun, Handbook of mathematical functions with formulas, graphs and mathematical tables, Knovel, London, 2002.

D. R. Brillinger, Time Series: data analysis and theory, McGraw-Hill, New York, 1981.