The Rao’s distance between negative binomial distributions for Exploratory Analyses and Goodness-of-Fit Testing
Claude Manté

To cite this version:
Claude Manté. The Rao’s distance between negative binomial distributions for Exploratory Analyses and Goodness-of-Fit Testing. 61st World Statistics Congress - ISI2017, Jul 2017, Marrakech, Morocco. hal-01632444

HAL Id: hal-01632444
https://hal.archives-ouvertes.fr/hal-01632444
Submitted on 10 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The Rao’s distance between negative binomial distributions for Exploratory Analyses and Goodness-of-Fit Testing

Manté Claude
Aix-Marseille Université, Université du Sud Toulon-Var, CNRS/INSU, IRD, MIO, UM 110, Marseille, France 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 biological characteristics (growth rate, reproductive power, survival rate, etc). The negative binomial distribution (NB) is widely used to model such data but the parametric approach is ill-suited from an exploratory point of view. Indeed, the “visual” distance between parameters is not relevant, because it depends on the chosen parametrization! On the contrary, considering the Riemannian manifold $NB(D_R)$ of negative binomial distributions equipped with the Fisher-Rao metrics, it is possible to compute intrinsic distances between species.

In this work, we focus on geometrical aspects of the $\chi^2$ goodness-of-fit (GOF) test for distributions in $NB(D_R)$, in connection with the position of the reference distribution. We show that this position is critical for performances of this test, as Critchley & Marriott (2016) noticed in a different setting.

Keywords: Information geometry, Riemannian manifold, $\chi^2$ test, overdispersed data.

1. Introduction

In a seminal paper, Rao (1945) noticed that, equipped with the Fisher information metrics denoted $g(\bullet)$, 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 $\theta^1$ and $\theta^2$ is given by:

$$D_R(\theta^1, \theta^2) := \int_0^1 \sqrt{\dot{\gamma}(t)} g(\gamma(t)) \dot{\gamma}(t) dt$$

(1)

where $\gamma$ is a segment (minimal length curve) connecting $\theta^1 = \gamma(0)$ to $\theta^2 = \gamma(1)$. As any Riemannian distance, $D_R$ is intrinsic (i.e. does not depend on the parametrization used). Following this pioneer work, a number of authors used the Rao’s distance to deal with various statistical topics: exploratory methods such as data visualization, clustering and classification, or hypothesis testing problems (Menendez et al., 1995; Cubedo & Oller, 2002).

We will focus on the latest topic (GOF tests), in the setting of the Riemannian manifold $NB(D_R)$ of negative binomial distributions equipped with this distance. A probability distribution $\mathcal{L}$ will be identified with its coordinates with respect to some chosen parametrization; for instance, we will write $\mathcal{L} \equiv (\phi^i, \mu^i)$.

2. Essential elements of Riemannian geometry

Consider a Riemannian manifold $\mathcal{M}$, and a parametric curve $\alpha : [a, b] \to \mathcal{M}$. Its first derivative with respect to “time” will be denoted $\dot{\alpha}$. We will also consider for any $\theta \in \mathcal{M}$ the local norm $\|V\|_g(\theta)$ associated with the metrics $g$ on the tangent space $T_\theta \mathcal{M}$:

$$\forall V \in T_\theta \mathcal{M}, \|V\|_g(\theta) := \sqrt{g(V, V)}.$$ 

(2)

Definition 1. (Berger, 2003) Let $\gamma : [0, 1] \to \mathcal{M}$ be a curve traced on $\mathcal{M}$, and $D$ be a connection on $\mathcal{M}$. $\gamma$ is a geodesic with respect to $D$ if its acceleration $D_{\gamma(t)} \dot{\gamma}(t)$ is null $\forall t \in [0, 1]$. In other words, a geodesic has constant speed in the local norm (2):

$$\|\dot{\gamma}\|_g := \|\dot{\gamma}(\bullet)\|_g(\gamma(\bullet)) = \sqrt{\dot{\gamma}(\bullet), g(\gamma(\bullet), \dot{\gamma}(\bullet)).}$$
Proposition 1. (Berger, 2003) For is really a metric, at least locally. Remember that $D^\chi_3$. Tangent plane approximation and where $\Theta$.

But geodesics can be as well be computed by solving (3) under the alternative constraints
\[
\{ \gamma (0) = \theta^1, \gamma (1) = \theta^2 \}.
\]

But geodesics can be as well be computed by solving (3) under the alternative constraints
\[
\{ \gamma (0) = \theta^1, \gamma (0) = V \in \mathbb{R}^2 \}\]

where $V$ is the initial velocity of the geodesic; this solution is associated with the exponential map at $\theta^1$.

3. Tangent plane approximation and $\chi^2$ GOF tests
Remember that $D_R$ has been constructively defined by formula (1); the following proposition shows that it is really a metric, at least locally.

Proposition 1. (Berger, 2003) For $\varrho$ small enough, the exponential map at $\theta^0$ is a local diffeomorphism, such that $\exp_{\theta^0} (B (0, \varrho)) = B_R (\theta^0, \varrho)$ (metric balls of radius $\varrho$ of $T_{\theta^0} \mathcal{M}$ and $\mathcal{M}$, respectively).

The maximal admissible value of $\varrho$ (named injectivity radius of $\mathcal{M}$ at $\theta^0$) is denoted $\varrho_0 (\theta^0)$.

Considering the first-order approximation of $D_R$ we can write, if $\theta$ is close enough to $\theta^0$:
\[
D_R (\varrho^0, \varrho) = D_R (\theta^0, \theta) \approx \sqrt{\gamma (\theta^0 - \theta)} = \| \varrho - \theta \|_g (\theta^0).
\]

Thus, if $D_R (\varrho^0, \varrho) = \varrho \leq \varrho_0 (\theta^0)$ is small enough, the metric sphere $S_R (\theta^0, \varrho) := \{ \theta \in \mathcal{M} : D_R (\theta^0, \theta) = \varrho \}$ can be isometrically identified with the centered ellipsoid $E_{\varrho^0} (\varrho)$ drawn on $T_{\theta^0} \mathcal{M}$.

This is illustrated on Figure 1, where we displayed $\exp_{\theta} (E_{\varrho} (\varrho)) = S_R (\theta, \varrho)$ for NB distributions of parameters $\theta^1 = \{ 0.7767, 11.2078 \}$ (yellow point, a rather aggregated distribution) and $\theta^2 = \{ 10., 9.12624 \}$ (red point, a bell-shaped distribution). We can see on Figure 1 that for $\varrho = 0.3$ and $\varrho = 1.5$, the geodesics emanating from $\theta^2$ draw circles, while the situation is very different for $\theta^1$: for $\varrho = 1.5$, the ball is dramatically anisotropic, very far from an ellipse while for $\varrho = 0.3$, the tangent approximation (6) seems reasonable.

Consider now the application $\mathfrak{F} : \Theta \rightarrow \mathcal{M}$ associating to $\theta$ the probability $\varrho^0$. Suppose that the classical Fréchet- Darmois- Cramer- Rao (FDCR) assumptions (Rao, 1945, 1965, chapter 5) are fulfilled by the familyinion 1, where we displa yed $\exp_{\theta} (E_{\varrho} (\varrho)) = S_R (\theta, \varrho)$ for NB distributions of parameters $\theta^1 = \{ 0.7767, 11.2078 \}$ (yellow point, a rather aggregated distribution) and $\theta^2 = \{ 10., 9.12624 \}$ (red point, a bell-shaped distribution). We can see on Figure 1 that for $\varrho = 0.3$ and $\varrho = 1.5$, the geodesics emanating from $\theta^2$ draw circles, while the situation is very different for $\theta^1$: for $\varrho = 1.5$, the ball is dramatically anisotropic, very far from an ellipse while for $\varrho = 0.3$, the tangent approximation (6) seems reasonable.

Consider now the application $\mathfrak{F} : \Theta \rightarrow \mathcal{M}$ associating to $\theta$ the probability $\varrho^0$. Suppose that the classical Fréchet- Darmois- Cramer- Rao (FDCR) assumptions (Rao, 1945, 1965, chapter 5) are fulfilled by the family
Figure 1: Sampled metric balls (50 geodesics) of radius 1.5 (blue curves) and 0.35 (white contours) for two distributions from \(NB(D_R)\); see comments in the text.

\[ \{ \Theta_0 : \theta \in \Theta \} \text{ and that } \hat{\theta} \text{ is an unbiased first-order efficient estimator of } \theta. \text{ Then, } g(\theta^0) \approx V(\hat{\theta})^{-1} \text{ and, if } \varrho \leq \bar{\varrho}(\theta^0), \text{ because of the Mahalanobis-like relationship (6), } \mathcal{P}(\mathcal{E}_{\theta^0}(\varrho)) \text{ can be confused with the metric sphere of radius } \varrho \text{ centered on } \mathcal{L}_{\theta^0}. \text{ But } \bar{\varrho}(\theta^0) \text{ is unknown in general; so, what if } \|\theta^0 - \hat{\theta}\|_g(\theta^0) \text{ is too large} (\gg \bar{\varrho}(\theta^0))? \text{ This is an important issue because of the } \chi^2 \text{ GOF test (see for instance } \text{ Rao, 1945; Menendez et al., 1995; Cubedo & Oller, 2002) associated with formula (6): if } \hat{\theta}_N \text{ is the estimation obtained from some } N \text{-sample of } \mathcal{L}_{\theta^0}, \text{ } N D_R^2(\theta^0, \hat{\theta}_N) \text{ should obey } \chi^2(p), \text{ asymptotically.}

\textbf{4. The geometry of } NB(D_R) \\
\text{There is a number of parametrizations for the negative binomial distribution; because of its orthogonality, we chose the one used by Chu & Ong (2013):}

\begin{equation}
P(X = j; (\phi, \mu)) = \binom{\phi + j - 1}{j} \left( \frac{\mu}{\mu + \phi} \right)^j \left( \frac{\mu}{\mu + \phi} \right)^\phi, j \geq 0
\end{equation}

\((\phi, \mu) \in \mathbb{R}^+ \times \mathbb{R}^+; \text{ here, } \mu \text{ is the mean of the distribution. } D_R(\mathcal{L}_1, \mathcal{L}_2) \text{ cannot be obtained in a closed-form but must be computed by finding the numerical solution of the Euler-Lagrange equation (3), completed in the parametrization (7) by the boundary conditions}

\begin{equation}
\{ \gamma(0) = (\phi^1, \mu^1), \gamma(1) = (\phi^2, \mu^2) \}
\end{equation}

\text{Geodesics can be as well be computed by solving (3) under the alternative constraints}

\begin{equation}
\{ \gamma(0) = (\phi^1, \mu^1), \gamma(0) = V \in \mathbb{R}^2 \}
\end{equation}

\text{where } V \text{ is the initial velocity of the geodesic; this solution is associated with the exponential map at } (\phi^1, \mu^1).
neighbourhood should be tested. Distributions above, we determined a set $\mathcal{E}_{\theta^0} (\alpha, N)$ of 20 regularly-sampled points upon the ellipsoid $\mathcal{E}_{\theta^0} (\alpha, N) \subset \Theta$, for $\alpha \in \{0.1, 0.99\}$. By "regularly-sampled", we mean that the arc length between two neighbor points should be $\text{Length} (\mathcal{E}_{\theta^0} (\alpha, N)) /20$. Afterwards, for each pair on points $\{\theta^1, \theta^2\}$ of each $\mathcal{S} \mathcal{E}_{\theta^0} (\alpha, N)$, $D_R (\theta^1, \theta^2)$ was computed by solving (3) under (8); in addition each $D_R (\theta^1, \theta^0)$ was computed too. This gave rise to a $21 \times 21$ table $\Delta (\theta^0, \alpha, N)$ submitted to MDS.

The first four distributions were already exemplified in Manté et al. (2016, Appendix 2). Just as Critchley & Marriott (2016), we fixed the sample size to $N = 100$; then, for each one of the test distributions above, we determined a set $\mathcal{S} \mathcal{E}_{\theta^0} (\alpha, N)$ of 20 regularly-sampled points upon the ellipsoid $\mathcal{E}_{\theta^0} (\alpha, N) \subset \Theta$, for $\alpha \in \{0.1, 0.99\}$. By "regularly-sampled", we mean that the arc length between two neighbor points should be $\text{Length} (\mathcal{E}_{\theta^0} (\alpha, N)) /20$. Afterwards, for each pair on points $\{\theta^1, \theta^2\}$ of each $\mathcal{S} \mathcal{E}_{\theta^0} (\alpha, N)$, $D_R (\theta^1, \theta^2)$ was computed by solving (3) under (8); in addition each $D_R (\theta^1, \theta^0)$ was computed too. This gave rise to a $21 \times 21$ table $\Delta (\theta^0, \alpha, N)$ submitted to MDS. The reader can examine on Figures 2&3 several sampled metric "spheres" obtained through MDS. On these plots, the 21st point corresponds to the reference $\mathcal{L}^{\theta^0}$, and should occupy the center of the circle; we superimposed the "true" circle obtained through MDS from the regularly sampled circle, with $N = 100$. For $\alpha = 0.1$, the representation was always rather good (even in the case of Agreg: see Figure 3). With a lower confidence level ($\alpha = 0.99$), the spheres are more or less deformed (see the right panel of Figure 2), or
Figure 3: The case of Agreg. Left (resp. right) panel: representation of $\Psi(S\mathcal{E}_{Agreg}(0.1, 100))$ (resp. $\Psi(S\mathcal{E}_{Agreg}(0.99, 100))$).

squarely destroyed in the case of borderline distributions like “Agreg” or “Boundary” (see the right panel of Figure 3).

Figure 4: Plot of the Riemannian measure inside $\mathcal{SE}_{Agreg}(0.99, 100)$ (truncated ellipsoid); the reference distribution $\theta^{Agreg}$ is represented by the black point.

Let’s examine a little further the case of “Agreg”. On Figure 4 we plotted the Riemannian measure $dV_M(\theta) := \sqrt{\det(g_{ij}(\theta))} \, d\theta_1 \cdots d\theta_p$ associated with the metrics $g$, for values inside $\mathcal{E}_{Agreg}(0.99, 100)$ cut by the axes $\phi = 0$ and $\mu = 0$. The points far from the frontiers are labeled by $\{1, 15, \cdots, 20\}$; we can see on the right panel of Figure 3 that these points are exactly those which appear on an arc of ellipsoid while the other points
collapse, forming two clusters: \{2, 4, 5\}, associated with the \(\mu = 0\) borderline, and \{8, \ldots, 14\}, associated with the \(\phi = 0\) borderline. In fact, as Critchley & Marriott (2016) noticed, problems are met near the frontier \((\Theta - \hat{\Theta})\). Incidentally, a similar phenomenon happened in large neighborhoods of the “MED” distribution (see Figure 1).

6. Conclusions

We have shown that:

- the critical locus of level \(1 - \alpha\) associated with the GOF test \(L_\theta^2 = L_{\theta_0}\) is an ellipsoid \(E_{\theta_0}(\alpha) \subset \Theta\) which depends (centre, eccentricity) on \(\theta_0\)
- the image of \(E_{\theta_0}(\alpha)\) under \(\Psi : \Theta \rightarrow NB(D_R)\) associating to \(\theta\) the probability \(L^\theta\) is theoretically a metric sphere
- \(\Psi(E_{\theta_0}(\alpha))\) can be considered as spherical only when \(E_{\theta_0}(\alpha)\) is “far enough” from the frontier of \(\Theta\) (see Figure 1, or Critchley & Marriott (2016)).

References


