Computing the Rao distance for Gamma distributions

F. Reverter*, J.M. Oller

Department of Statistics, University of Barcelona, Diagonal, 645, 08028 Barcelona, Spain

Abstract

This paper deals with the calculation of the Rao distance between Gamma distributions, which is the Riemannian distance induced by the Fisher information matrix on the Gamma statistical model. In this case no closed form expression of the Rao distance is available and a numerical approach is thus necessary to compute the distance. A computer program based on a simple shooting algorithm is presented and discussed.
© 2003 Elsevier B.V. All rights reserved.

1. Introduction

Metrics and distances between probability distributions play a fundamental role in problems of statistical inference and data analysis to study affinities within a given set of populations. See for instance the works of Matusita [13], Bhattacharyya [2], Mahalanobis [12] and Cuadras [9] among others.

We first introduce some notation. Let χ a sample space, a a σ-algebra of subsets of χ and μ a positive measure on the measurable space (χ, a). A parametric statistical model is defined as the triple $\{(\chi, a, \mu); \Theta; f\}$ where: (χ, a, μ) is a measure space, Θ, called the parameter space, is a manifold and f is a measurable map $f : \chi \times \Theta \to \mathbb{R}$ such that $f \geqslant 0$ and $P_\theta(\mathrm{d}x) = f(x; \theta)\mu(\mathrm{d}x)$ is a probability measure on (χ, a), $\forall \theta \in \Theta$. In general Θ can be any manifold although in many applications Θ will be an n-dimensional C^∞ real manifold, Hausdorff and connected. Moreover, for many applications Θ is an open set of $\mathbb{R}^n$ and in this case it is customary to use the same symbol (θ) to denote points and coordinates.

---

* Corresponding author.
E-mail address: reverter@porthos.bio.ub.es (F. Reverter).

Additionally, if we assume that f satisfies certain regularity conditions, see for instance [3], the elements

$$g_{ij}(\theta) = \int_\chi \frac{\partial \log f(x;\theta)}{\partial \theta^i} \frac{\partial \log f(x;\theta)}{\partial \theta^j} f(x;\theta)\mu(\mathrm{d}x), \quad i,j = 1,\ldots,n,$$

of the Fisher information matrix are all finite and the information matrix is strictly positive definite. Moreover,

$$\mathrm{d}s^2 = \sum_{i,j=1}^n g_{ij}(\theta)\,\mathrm{d}\theta^i\,\mathrm{d}\theta^j \tag{1.1}$$

is a Riemannian metric on Θ, known as the information metric of the model, see [17]. Given two probability measures P_{θ_1} and P_{θ_2} which belong to the statistical model, the Rao distance between P_{θ_1} and P_{θ_2} is defined as the Riemannian distance between $\theta_1, \theta_2 \in \Theta$. For a more detailed account, see [3,6,7,16]. The Rao distance has been computed for several statistical models, see Atkinson and Mitchell [1], Mitchell and Krzanowski [14], Burbea and Oller [4] among others. There are some statistical models, such as Gamma, Beta and multivariate Normal, for which a closed form of the Rao distance is not available, in which cases a numerical approach may be appropriate. Then we may compute the distance by solving numerically a system of nonlinear equations obtained through the corresponding geodesic equations. Here we attempt to calculate the Rao distance between Gamma distributions.

2. Description of the method

Let M be an n-dimensional Riemannian manifold, M_p be the tangent space of M at $p \in M$, ∇ be the Levi–Civita connection on M and $\omega : I \subset \mathbb{R} \to M \in C^l$, $l \geqslant 2$, be a curve on M. To write geodesic equations corresponding to ∇ in coordinates, we let $z : U \to \mathbb{R}^n$ be the chart, set $\omega^j = z^j \circ \omega$, $j = 1,\ldots,n$, and let Γ^l_{jk} denote the Christoffel symbols. Then, ω is a geodesic if and only if is solution of

$$\ddot{\omega}^l + \sum_{j,k=1}^n \Gamma^l_{jk}(\omega)\,\dot{\omega}^j\dot{\omega}^k = 0, \quad l = 1,\ldots,n. \tag{2.1}$$

In order to compute the Rao distance between p and q, $d(p,q)$, we must solve system (2.1) with the boundary conditions

$$\omega(0) = p \in M, \qquad \omega(1) = q \in M, \tag{2.2}$$

because the length of the resulting geodesic $\omega(t)$ is the Riemannian distance, see for instance [20]. In general, geodesics or even minimal geodesics are not unique but if M is simply connected and complete and all sectional curvatures are nonpositive then any two points on M are joined by a unique minimal geodesic. Using the exponential map, see for instance [8], and considering the map $F : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$F(x) = z \circ \exp_p \circ \tilde{z}_p^{-1}(x) - z(q), \tag{2.3}$$

where $\tilde{z}_p : M_p \rightarrow \mathbb{R}^n$ denotes the chart on M_p induced by z, we can express the two-point boundary value problems (2.1) and (2.2) through the system of equations

$$F(x) = 0. \tag{2.4}$$

This is a nonlinear system of equations that has to be solved iteratively. To compute F we have a set of standard methods such as the Taylor series, Runge–Kutta, multistep methods and many others. The problem of finding a good starting approximation may be the most difficult step for a general system, and it will be discussed later, specifically for the Gamma statistical model. The basic approach for solving (2.4) is the Newton method or one of its variants. We have to calculate the $n \times n$ elements of the Jacobian matrix, $DF(x)$, which can be computed by solving the equations of variation corresponding to (2.1). Equations of variation may be obtained by considering the closed relation between the derivative of the exponential map and the Jacobi fields. Observe that from (2.3) the Jacobian matrix $DF(x)$ coincides with the matrix of the derivative map $(\exp_p)_*$ at ξ, $(\exp_p)_{*|\xi}$, expressed in local coordinates. On the other hand, if Y denotes a Jacobi field along the geodesic $\omega(t) = \exp_p(t\xi)$, with $p \in M$, $\xi \in M_p$, determined by the initial conditions $Y(0) = 0$ and $(\nabla/dt)Y(0) = \eta$, then

$$(\exp_p)_{*|t\xi}(t\eta) = Y(t), \tag{2.5}$$

see for instance [8]. Therefore, integrating Jacobi's equation

$$\frac{\nabla^2}{dt^2} Y + R(\dot{\omega}, Y)\dot{\omega} = 0, \tag{2.6}$$

we can compute the $(\exp_p)_{*|\xi}$. Eq. (2.5) holds for any η in M_p, in particular, for the coordinate basis vectors of M_p, $(\partial/\partial z^1)_p, \ldots, (\partial/\partial z^n)_p$. Then the vectors $Y_k(1)$, $k = 1, \ldots, n$, which result by integrating Jacobi's equation with initial conditions $Y_k(0) = 0$ and $(\nabla/dt)Y_k(0) = (\partial/\partial z^k)_p$ along $\omega(t) = \exp_p(t\xi)$ give us the kth column of the matrix associated with $(\exp_p)_{*|\xi}$. By straightforward calculus we can write Jacobi's equation (2.6) in coordinates. In matrix notation,

$$\ddot{Y}_k(t) = A Y_k(t) + B \dot{Y}_k(t), \tag{2.7}$$

where Y_k and $\dot{Y}_k$ are $n \times 1$ vectors with the components and the derivatives with respect to t of the field, respectively, A and B are $n \times n$ matrices not depending on the Y_k and $\dot{Y}_k$, of the form

$$A = \begin{pmatrix} -\frac{\partial \Gamma^1_{ij}}{\partial \omega^1} \dot{\omega}^i \dot{\omega}^j & \cdots & -\frac{\partial \Gamma^1_{ij}}{\partial \omega^n} \dot{\omega}^i \dot{\omega}^j \\ \vdots & & \vdots \\ -\frac{\partial \Gamma^n_{ij}}{\partial \omega^1} \dot{\omega}^i \dot{\omega}^j & \cdots & -\frac{\partial \Gamma^n_{ij}}{\partial \omega^n} \dot{\omega}^i \dot{\omega}^j, \end{pmatrix} \qquad B = \begin{pmatrix} -2\Gamma^1_{i1} \dot{\omega}^i & \cdots & -2\Gamma^1_{in} \dot{\omega}^i \\ \vdots & & \vdots \\ -2\Gamma^n_{i1} \dot{\omega}^i & \cdots & -2\Gamma^n_{in} \dot{\omega}^i \end{pmatrix},$$

where we have employed the summation of repeated index convention and we omit the dependence of t for shortness. Then the kth column of $DF(x)$ results by integrating the system (2.7) with the initial conditions

$$Y_k(0) = (0, \ldots, 0, \ldots, 0)', \qquad \dot{Y}_k(0) = (0, \ldots, 1, \ldots, 0)',$$

where 1 is placed at the kth row, along the geodesic $\omega(t)$.

Remark 2.1. From (2.5) the local coordinates of $(\nabla/\mathrm{d}t)Y(0) \in M_p$ are $-\mathrm{D}F(x)^{-1}F(x)$, the correction of the Newton method, when the Jacobi field $Y(t)$ along the geodesic $\omega(t)$ is determined by the boundary conditions

$$Y(0) = 0, \qquad Y(1) = -F(x),$$

given in local coordinates.

3. Applications to the Gamma statistical model

In this section we describe the more interesting features of the program developed to compute Rao's distance between Gamma distributions based on a simple shooting method algorithm. We pay special attention to two stages: the choice of a good starting approximation and the numerical implementation of the iterative procedure by Newton's method. Another important stage of the algorithm is the resolution of the subsequent initial value problem by numerical integration of the geodesic equations. In our case, since we have obtained the second, third and fourth derivatives of the geodesics we have used a modification of the Taylor series method, see [19], to integrate the geodesics with a $O(h^8)$ method. We first summarize the fundamental quantities of the Gamma statistical model.

3.1. Gamma statistical model

Let $\{(\chi, a, \mu); \Theta; f\}$ be the Gamma statistical model where; $\chi \equiv \mathbb{R}^+$, a is a σ-algebra on the Borel sets of χ and μ is the Lebesgue measure restricted to $\mathbb{R}^+$, $\Theta \equiv \mathbb{R}^+ \times \mathbb{R}^+$ and $f: \chi \times \Theta \to \mathbb{R}$ parametrized by

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x}, \tag{3.1}$$

where (α, β) are the coordinates given by the chart corresponding to identity. If we consider a set of new coordinates on Θ given by the transformation $\alpha = \alpha$, $\eta = \log(\alpha/\beta)$ then the information metric is given by

$$\mathrm{d}s^2 = \frac{\phi(\alpha)}{\alpha} (\mathrm{d}\alpha)^2 + \alpha(\mathrm{d}\eta)^2 \tag{3.2}$$

with $\phi(\alpha) = \alpha \Psi'(\alpha) - 1$ where $\Psi'(\alpha) = (\log \Gamma(\alpha))''$ denotes the usual trigamma function. The Gauss curvature is given by

$$\kappa(\alpha) = -\frac{1}{4} \left(\frac{1}{\phi(\alpha)} \right)', \tag{3.3}$$

see for instance [11]. As pointed out in [5], $\kappa(\alpha)$ is a strictly negative function for $\alpha \in (0, \infty)$. Then any two points on the manifold Θ can be joined by only one geodesic and the length of the resulting geodesic gives the Riemannian distance between these points. The geodesic equations, see for instance [11,21], are of the form

$$\ddot{\alpha} + \frac{1}{2} \left(\frac{\phi'(\alpha)}{\phi(\alpha)} - \frac{1}{\alpha} \right) \dot{\alpha}^2 - \frac{1}{2} \frac{\alpha}{\phi(\alpha)} \dot{\eta}^2 = 0, \qquad \ddot{\eta} + \frac{1}{\alpha} \dot{\alpha}\dot{\eta} = 0, \tag{3.4}$$

with the property $(\phi(\alpha)/\alpha)\dot{\alpha}^2 + \alpha\dot{\eta}^2 = \rho^2$, with $\rho = 1$ when the geodesic is parametrized by arc length. Our problem is to solve system (3.4) given the boundary conditions

$$(\alpha(0), \eta(0)) = (\alpha_1, \eta_1), \qquad (\alpha(1), \eta(1)) = (\alpha_2, \eta_2).$$

Integrating the geodesic equations once gives

$$\dot{\alpha} = \pm\sqrt{\frac{\rho^2\alpha - A^2}{\phi(\alpha)}}, \qquad \dot{\eta} = A/\alpha, \tag{3.5}$$

where A is a constant of integration. Further analytical integration of equations seems difficult which suggests that we should proceed numerically.

3.2. Starting approximation

The information metric ds^2, given by (3.2), is bounded by Poincaré metrics, see [5] for a detailed account. If we consider new coordinates on Θ given by the bijection $(\alpha, \eta) \mapsto (\xi, \eta)$ where $\xi = \sqrt{b}/\sqrt{\alpha}$ and $\xi > 0$, with $b = 2$ or 4, then ds^2 is bounded by

$$\frac{b}{\xi^2}((d\xi)^2 + (d\eta)^2), \tag{3.6}$$

which are Poincaré metrics on Θ, with lower bound for $b = 2$ and upper bound for $b = 4$. To obtain a good starting approximation, we shall compute the exact solution for the present problem using a Poincaré type metric (3.6) instead of (3.2). The geodesics corresponding to the Poncairé metric are, see [4],

$$\xi(t) = C^{-1}\operatorname{sech}\left(\frac{\rho t}{\sqrt{b}} + \varepsilon\right), \tag{3.7}$$

$$\eta(t) = \sqrt{b}C^{-2}B\tanh\left(\frac{\rho t}{\sqrt{b}} + \varepsilon\right) + D, \tag{3.8}$$

and

$$\xi(t) = \exp\left(\frac{\rho t}{\sqrt{b}} + \varepsilon\right),$$

$$\eta(t) = D,$$

when $\eta_1 = \eta_2$, where ε, B and D are real constants of integration, and $C = \sqrt{b}|B|$, ρ denotes the Poincaré distance between the boundary points (ξ_1, η_1) and (ξ_2, η_2), that is, $\rho = \sqrt{b}\log((1 + \Delta)/(1 - \Delta))$, with

$$\Delta = \sqrt{\frac{(\eta_2 - \eta_1)^2 + b((1/\sqrt{\alpha_2}) - (1/\sqrt{\alpha_1}))^2}{(\eta_2 - \eta_1)^2 + b((1/\sqrt{\alpha_2}) + (1/\sqrt{\alpha_1}))^2}}.$$

Observe that as a consequence of (3.6) the above Poincaré distances are upper and lower bounds for the Rao distance.

Then, by differentiating (3.7) and (3.8), and evaluating at $t = 0$, we obtain the starting tangent vector of the geodesic between points with $\eta_1 \neq \eta_2$,

$$\dot{\xi}(0) = -\frac{C^{-1}}{\sqrt{b}} \rho \sinh(\varepsilon) \cosh^{-2}(\varepsilon),$$

$$\dot{\eta}(0) = C^{-2} B \rho \cosh^{-2}(\varepsilon).$$

Expressing the vector in the boundary points and taking into account the contravariant law to describe the change of tangent vectors we obtain

$$\dot{\alpha}(0) = -\frac{\rho}{\sqrt{b}} \frac{1 - \Delta^2}{\Delta} \alpha_1 \sqrt{\alpha_2} \left(-\frac{1}{\sqrt{\alpha_1}} + \frac{1}{\sqrt{\alpha_2}} \frac{1 + \Delta^2}{1 - \Delta^2} \right), \tag{3.9}$$

$$\dot{\eta}(0) = \frac{\rho}{\sqrt{b}} \frac{1 - \Delta^2}{2\Delta} (\eta_2 - \eta_1) \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}. \tag{3.10}$$

The vector with components (3.9) and (3.10) is the exact solution in the Poincaré metric and supplies a good starting approximation for the problem formulated in the information metric, as can be seen in Section 4 where several examples are presented.

Remark 3.1. We see no direct reasons to prefer $b = 2$ instead of $b = 4$ or vice versa. In fact, the better choice may depend on the problem. We combine the solution obtained for both values of b. To combine the starting vectors we average their physical components with respect to each approximate Poincaré metrics, then normalize the resulting vector with respect to the information metric and scale it by the average of the two previously obtained Rao distance bounds.

3.3. Newton iteration

As we have introduced in Section 2 from (2.5) we can compute the correction proposed by the Newton method if we compute $(\nabla / \mathrm{d}t)Y(0)$ by integrating the Jacobi field $Y(t)$ defined by the boundary conditions (in coordinates)

$$Y(0) = 0, \qquad Y(1) = -F(x),$$

along the geodesic corresponding to each iteration. To this end we describe the numerical integration of the Jacobi fields corresponding to the geodesics on the manifold related to the Gamma statistical model. We can decompose any Jacobi field along a geodesic into its normal component and its tangential component: $Y(t) = Y^{\tan}(t) + Y^{\mathrm{nor}}(t)$, both components also being Jacobi fields, see [8]. Moreover for a tangential Jacobi field, we can express its initial values in terms of the known boundary values, that is,

$$Y^{\tan}(0) = 0, \qquad \frac{\nabla}{\mathrm{d}t} Y^{\tan}(0) = \langle Y(1), \gamma'(1) \rangle \gamma'(0).$$

Then the derivative of the Jacobi field at the origin satisfies

$$\frac{\nabla}{\mathrm{d}t} Y(0) = \langle Y(1), \gamma'(1) \rangle \gamma'(0) + \frac{\nabla}{\mathrm{d}t} Y^{\mathrm{nor}}(0).$$

Thus we need only to compute $(\nabla/\mathrm{d}t)Y^{\mathrm{nor}}(0)$ by integrating the Jacobi field equation for the normal component with boundary conditions

$$Y^{\mathrm{nor}}(0) = 0, \qquad Y^{\mathrm{nor}}(1) = -F(x) - Y^{\mathrm{tan}}(1). \tag{3.11}$$

For the Gamma model, from (2.6), the Jacobi system of equations in coordinates is

$$\ddot{u} + \left(\frac{\phi'(\alpha)}{\phi(\alpha)} - \frac{1}{\alpha} \right) \dot{\alpha}\dot{u} - \frac{\alpha}{\phi(\alpha)}\dot{\eta}\dot{v} - \frac{1}{2} \left(\frac{\phi(\alpha) - \alpha\phi'(\alpha)}{\phi(\alpha)^2} \right) \dot{\eta}^2 u$$

$$+ \frac{1}{2} \left(\frac{\phi''(\alpha)\phi(\alpha) - \phi'(\alpha)^2}{\phi(\alpha)^2} + \frac{1}{\alpha^2} \right) \dot{\alpha}^2 u = 0, \tag{3.12}$$

$$\ddot{v} + \frac{\dot{\eta}\dot{u}}{\alpha} + \frac{\dot{\alpha}\dot{v}}{\alpha} - \frac{\dot{\alpha}\dot{\eta}u}{\alpha^2} = 0, \tag{3.13}$$

where we assume that $(u(t), v(t))$ are normal components. We integrate Eqs. (3.12) and (3.13) with conditions (3.11) which we write for shortness as

$$(u(0), v(0)) = (0, 0), \tag{3.14}$$

$$(u(1), v(1)) = (u_1, v_1). \tag{3.15}$$

Taking into account the equation for $\ddot{\eta}$ in (3.4) and substituting it in (3.13) we obtain

$$\alpha\dot{v} + \dot{\eta}u = C, \tag{3.16}$$

where C is a constant of integration to be determined from the boundary conditions. Observe that,

$$\dot{v} = \frac{1}{\alpha}(C - \dot{\eta}u). \tag{3.17}$$

On the other hand, the normal components satisfy

$$\frac{\phi(\alpha)}{\alpha}\dot{\alpha}u + \alpha\dot{\eta}v = 0$$

and using the equation for $\dot{\eta}$ in (3.5) we obtain

$$\phi(\alpha)\dot{\alpha}u + A\alpha v = 0, \tag{3.18}$$

where A is a known constant for given initial values of the geodesic associated to the iteration. Differentiating (3.18), we obtain

$$(\phi'(\alpha)\dot{\alpha}^2 + \phi(\alpha)\ddot{\alpha})u + \phi(\alpha)\dot{\alpha}\dot{u} + A\dot{\alpha}v + A\alpha\dot{v} = 0. \tag{3.19}$$

On the other hand, from the equation for $\ddot{\alpha}$ in (3.4) and taking into account that $\rho^2 = (\phi(\alpha)/\alpha)\dot{\alpha}^2 + \alpha\dot{\eta}^2$, where ρ^2 denotes the square norm of the tangent vector of the geodesic corresponding to the iteration, we obtain

$$\phi(\alpha)\ddot{\alpha} = \frac{1}{2} \left(\alpha\dot{\eta}^2 - \phi'(\alpha)\dot{\alpha}^2 + \frac{\phi(\alpha)\dot{\alpha}^2}{\alpha} \right) = \frac{1}{2}(\rho^2 - \phi'(\alpha)\dot{\alpha}^2). \tag{3.20}$$

Combining (3.20), (3.17) and (3.18) into (3.19) we have

$$\phi(\alpha)\dot{\alpha}\dot{u} + \left(\frac{\rho^2}{2} + \frac{\phi'(\alpha)}{2}\dot{\alpha}^2 - \frac{\phi(\alpha)}{\alpha}\dot{\alpha}^2 - A\dot{\eta} \right) u + AC = 0. \tag{3.21}$$

By considering again (3.5) we could write $\rho^2 = (\phi(\alpha)/\alpha)\dot{\alpha}^2 + A\dot{\eta}$, then (3.21) can be written as

$$\phi(\alpha)\dot{\alpha}\dot{u} + \left\{ \frac{\phi'(\alpha)}{2}\dot{\alpha}^2 - \frac{\rho^2}{2} \right\} u + AC = 0, \tag{3.22}$$

and finally,

$$\dot{u} - \frac{\ddot{\alpha}}{\dot{\alpha}}u + \frac{AC}{\phi(\alpha)\dot{\alpha}} = 0, \tag{3.23}$$

we comment later the case $\dot{\alpha} = 0$. Next, to compute $\dot{u}(0)$ and $\dot{v}(0)$ we consider three cases depending on the initial values $\dot{\alpha}(0)$ and $\dot{\eta}(0)$ of the geodesic corresponding to the iteration.

Case 1: $\dot{\eta}(0) = 0$.

Then $A = 0$ and from (3.18) results

$$\phi(\alpha)\dot{\alpha}u = 0. \tag{3.24}$$

The above equation must hold along the geodesic and since in this case $\dot{\alpha}(t) \neq 0$ for all t values, then $u(t) = 0$, $\dot{u}(t) = 0$ and then $\dot{u}(0) = 0$. On the other hand, since $\dot{\alpha}(t) \neq 0$ for all t, from (3.17) and (3.18) we obtain

$$\dot{v} - \frac{A\dot{\eta}}{\phi(\alpha)\dot{\alpha}}v - \frac{C}{\alpha} = 0. \tag{3.25}$$

Therefore, in this case,

$$\dot{v}(t) = \frac{C}{\alpha(t)}. \tag{3.26}$$

Taking into account that there is a linear relation between the values $\dot{v}(0)$ and $v(1)$, then we can write

$$v(1) = b\dot{v}(0), \tag{3.27}$$

where $b \in \mathbb{R}$ is an unknown constant to be determined from boundary conditions. Notice that, if we fix the value $\dot{v}(0) = 1$ or equivalently $C = \alpha(0)$ and we integrate numerically Eq. (3.26) along the geodesic, using a method with global error of order $O(h^p)$, where h denotes the steplength for the geodesic integration, then at $t = 1$ we will obtain

$$v(1) = \tilde{v} + O(h^p). \tag{3.28}$$

Therefore, from (3.27) and (3.28)

$$b = \tilde{v} + O(h^p).$$

Consequently, for the normal Jacobi field with conditions (3.14) and (3.15), from (3.27), we have

$$v_1 = \tilde{v}\dot{v}(0) + O(h^p).$$

and thus

$$\dot{v}(0) = \frac{v_1}{\tilde{v}} + O(h^p).$$

Case 2: $\dot{\alpha}(0) = 0$.

Observe that, from (3.21), at $t=0$, then $AC=0$. Thus $C=0$. Taking into account that the component α of the geodesic is a convex function on t as can be deduced from the geodesic equations, then $\dot{\alpha}(t) > 0$ for $t \neq 0$, and, therefore, from (3.23) we have

$$\dot{\alpha}\dot{u} - \ddot{\alpha}u = 0. \tag{3.29}$$

Hence, $u/\dot{\alpha} = B$, where B is a constant that should be determined from boundary conditions. Then

$$u(t) = B\dot{\alpha}(t), \tag{3.30}$$

$$\dot{u}(t) = B\ddot{\alpha}(t). \tag{3.31}$$

We can determine B by using the conditions at $t = 1$, from (3.30)

$$B = \frac{u_1}{\dot{\alpha}(1)}.$$

and from (3.31)

$$\dot{u}(0) = \frac{u_1}{\dot{\alpha}(1)}\,\ddot{\alpha}(0).$$

On the other hand, if we consider (3.16) at $t = 0$, since $C = 0$ results $\alpha(0)\dot{v}(0) = 0$. Consequently $\dot{v}(0) = 0$.

  Case 3: $\dot{\alpha}(0) \neq 0$, $\dot{\eta}(0) \neq 0$.

  For the component $v(t)$, from (3.19), at $t = 0$ we have

$$\dot{v}(0) = -\frac{\phi(\alpha(0))\dot{\alpha}(0)}{A\alpha(0)}\,\dot{u}(0). \tag{3.32}$$

Observe that $\dot{v}(0)$ depends linearly on $\dot{u}(0)$. Moreover, there is a linear relationship between $\dot{u}(0)$ and $u(t)$ at each t. At $t = 1$ we could write

$$u(1) = a\dot{u}(0), \tag{3.33}$$

where a is a real unknown constant that is determined by the boundary conditions. With a similar reasoning as *Case* 1, if we fix $\dot{u}(0) = 1$ and then integrate numerically the differential equation of $\dot{u}$ in (3.23) we will obtain at $t = 1$ the value $\tilde{u}$ such that

$$u(1) = \tilde{u} + O(h^p), \tag{3.34}$$

where $O(h^p)$ denotes the order of the final error. Then, from (3.33) and (3.34) we obtain $a = \tilde{u} + O(h^p)$. Then for the Jacobi field with boundary conditions (3.14) and (3.15) we have $u_1 = \tilde{u}\dot{u}(0) + O(h^p)$ and in consequence

$$\dot{u}(0) = \frac{u_1}{\tilde{u}} + O(h^p). \tag{3.35}$$

With (3.35) and (3.32) we obtain the desired quantities.

Remark 3.2. We have integrated Eq. (3.23) by using a modified Taylor of order $O(h^5)$, see [18,19]. We have observed numerically that the behaviour of the component $u(t)$ is stable in spite of $\dot{\alpha}(t)$ being so small. A singularity will appear only at $t = t_m$ with $\dot{\alpha}(t_m) = 0$ but we could avoid this eventuality. When the program detects $\dot{\alpha}(t) \simeq 0$, we approximate $u(t_m + h)$ by a Taylor expansion at

t_m, using the coefficients at $t_m - h$, instead of using expression (3.23). We should store, temporarily, at each step the approximate values of all derivatives.

Remark 3.3. From numerical calculations we have observed that the Gaussian curvature κ is bounded by $-\frac{1}{2} < \kappa < -\frac{1}{4}$. Then from the geometric Rauch theorem, see [8], we have

$$\frac{\rho}{\sqrt{2}\sinh(\rho/\sqrt{2})}\,\|Y\|(1) \leqslant \|(\nabla/\mathrm{d}t)Y\|(0) \leqslant \frac{\rho}{2\sinh(\rho/2)}\,\|Y\|(1).$$

Observe that the above bounds allow to stabilize the iterative procedure because if eventually the norm of the computed $(\nabla/\mathrm{d}t)Y(0)$ exceeds the bounds we can scale conveniently the vector.

4. Behaviour of the program

We have written a computer program, using Microsoft Developer Studio 97 for Visual Fortran [10], based on the ideas introduced above. In order to illustrate the behaviour of the program we have considered two features: the range of applicability over the parameter space and time spent in computing any distance. With this aim we describe two kinds of examples.

4.1. Example 1

For the values of the parameters α and β corresponding to the vertex of the six rectangles defined by $[10^{-k}, 10^k] \times [10^{-k}, 10^k]$, $k = 1, \ldots, 6$, Tables 1–6 shows for every rectangle the performance of the program in solving the six boundary problems determined by the two diagonals and the four sides of each rectangle. Tables summarize in columns: N, the number of iterations needed to solve the problem, ρ, the Rao's distance between the boundary points (α_1, β_1) and (α_2, β_2), error, the length of the final deviation vector measured in the metric of (α_2, β_2), time, the time in seconds to solve the problem, v, speed as the ratio between distance ρ and time. Observe that:

1. Every problem has been solved by the program. The range of the parameter space where the program converges to the solution is clearly wider than the range of standard programs like the D02HAF routine of NAG library, see [15]. This indicates the goodness of the initial approximation and the global accuracy of the iterative scheme.

Table 1
Rectangle: $[10^{-1}, 10^1] \times [10^{-1}, 10^1]$

(α_1, β_1)	(α_2, β_2)	N	ρ	Error	Time (s)	v
$(10^{-1}, 10^{-1})$	$(10^1, 10^1)$	5	3.76091	$0.591\ 10^{-7}$	$0.606\ 10^{-2}$	620.5
$(10^{-1}, 10^1)$	$(10^1, 10^{-1})$	5	6.08427	$0.141\ 10^{-7}$	$0.105\ 10^{-1}$	579.0
$(10^{-1}, 10^{-1})$	$(10^{-1}, 10^1)$	4	1.42335	$0.315\ 10^{-6}$	$0.211\ 10^{-2}$	672.9
$(10^{-1}, 10^1)$	$(10^1, 10^1)$	4	4.67064	$0.555\ 10^{-6}$	$0.655\ 10^{-2}$	712.9
$(10^{-1}, 10^{-1})$	$(10^1, 10^{-1})$	4	4.67064	$0.555\ 10^{-6}$	$0.654\ 10^{-2}$	714.0
$(10^1, 10^{-1})$	$(10^1, 10^1)$	4	6.90379	$0.948\ 10^{-6}$	$0.968\ 10^{-2}$	712.8

Table 2
Rectangle: $[10^{-2}, 10^2] \times [10^{-2}, 10^2]$

(α_1, β_1)	(α_2, β_2)	N	ρ	Error	Time (s)	v
$(10^{-2}, 10^{-2})$	$(10^2, 10^2)$	4	7.65804	$0.334\ 10^{-6}$	$0.954\ 10^{-2}$	802.9
$(10^{-2}, 10^2)$	$(10^2, 10^{-2})$	5	8.89581	$0.286\ 10^{-7}$	$0.146\ 10^{-1}$	609.5
$(10^{-2}, 10^{-2})$	$(10^{-2}, 10^2)$	5	0.91301	$0.222\ 10^{-6}$	$0.218\ 10^{-2}$	418.8
$(10^{-2}, 10^2)$	$(10^2, 10^2)$	5	8.04868	$0.660\ 10^{-7}$	$0.133\ 10^{-1}$	605.2
$(10^{-2}, 10^{-2})$	$(10^2, 10^{-2})$	5	8.04868	$0.660\ 10^{-7}$	$0.132\ 10^{-1}$	609.8
$(10^2, 10^{-2})$	$(10^2, 10^2)$	5	12.5209	$0.162\ 10^{-6}$	$0.201\ 10^{-1}$	621.9

Table 3
Rectangle: $[10^{-3}, 10^3] \times [10^{-3}, 10^3]$

(α_1, β_1)	(α_2, β_2)	N	ρ	Error	Time (s)	v
$(10^{-3}, 10^{-3})$	$(10^3, 10^3)$	4	11.5798	$0.315\ 10^{-6}$	$0.141\ 10^{-1}$	822.7
$(10^{-3}, 10^3)$	$(10^3, 10^{-3})$	5	11.9328	$0.623\ 10^{-7}$	$0.195\ 10^{-1}$	610.2
$(10^{-3}, 10^{-3})$	$(10^{-3}, 10^3)$	3	0.43602	$0.410\ 10^{-7}$	$0.757\ 10^{-3}$	575.9
$(10^{-3}, 10^3)$	$(10^3, 10^3)$	5	11.6756	$0.569\ 10^{-7}$	$0.187\ 10^{-1}$	625.6
$(10^{-3}, 10^{-3})$	$(10^3, 10^{-3})$	5	11.6756	$0.569\ 10^{-7}$	$0.188\ 10^{-1}$	622.3
$(10^3, 10^{-3})$	$(10^3, 10^3)$	6	17.2655	$0.709\ 10^{-7}$	$0.323\ 10^{-1}$	535.6

Table 4
Rectangle: $[10^{-4}, 10^4] \times [10^{-4}, 10^4]$

(α_1, β_1)	(α_2, β_2)	N	ρ	Error	Time (s)	v
$(10^{-4}, 10^{-4})$	$(10^4, 10^4)$	4	15.5092	$0.476\ 10^{-6}$	$0.193\ 10^{-1}$	803.1
$(10^{-4}, 10^4)$	$(10^4, 10^{-4})$	5	15.5754	$0.462\ 10^{-7}$	$0.254\ 10^{-1}$	614.2
$(10^{-4}, 10^{-4})$	$(10^{-4}, 10^4)$	3	0.18414	$0.357\ 10^{-10}$	$0.757\ 10^{-3}$	243.1
$(10^{-4}, 10^4)$	$(10^4, 10^4)$	5	15.5259	$0.170\ 10^{-7}$	$0.261\ 10^{-1}$	593.8
$(10^{-4}, 10^{-4})$	$(10^4, 10^{-4})$	5	15.5259	$0.169\ 10^{-7}$	$0.253\ 10^{-1}$	612.6
$(10^4, 10^{-4})$	$(10^4, 10^4)$	10	21.6093	$0.484\ 10^{-6}$	$0.652\ 10^{-1}$	331.3

2. The program took little time to solve all the cases even when the points were far apart. This is an advantage in situations, like simulation procedures, when a large number of distances must be computed.

4.2. Example 2

We performed an additional test. By means of a random number generator we have obtained 10 000 pairs of points independently and uniformly distributed on the rectangle $[10^{-6}, 10^6] \times [10^{-6}, 10^6]$. The program converges to the solution for every pair of points. In Table 7 we show the summary

Table 5
Rectangle: $[10^{-5}, 10^5] \times [10^{-5}, 10^5]$

(α_1, β_1)	(α_2, β_2)	N	ρ	Error	Time (s)	v
$(10^{-5}, 10^{-5})$	$(10^5, 10^5)$	4	19.4378	$0.720\ 10^{-6}$	$0.247\ 10^{-1}$	785.4
$(10^{-5}, 10^5)$	$(10^5, 10^{-5})$	5	19.4478	$0.141\ 10^{-7}$	$0.323\ 10^{-1}$	600.6
$(10^{-5}, 10^{-5})$	$(10^{-5}, 10^5)$	2	0.07281	$0.379\ 10^{-7}$	$0.543\ 10^{-3}$	134.1
$(10^{-5}, 10^5)$	$(10^5, 10^5)$	5	19.4399	$0.721\ 10^{-8}$	$0.319\ 10^{-1}$	608.1
$(10^{-5}, 10^{-5})$	$(10^5, 10^{-5})$	5	19.4399	$0.721\ 10^{-8}$	$0.321\ 10^{-1}$	604.3
$(10^5, 10^{-5})$	$(10^5, 10^5)$	7	25.7231	$0.225\ 10^{-6}$	$0.537\ 10^{-1}$	478.5

Table 6
Rectangle: $[10^{-6}, 10^6] \times [10^{-6}, 10^6]$

(α_1, β_1)	(α_2, β_2)	N	ρ	Error	Time (s)	v
$(10^{-6}, 10^{-6})$	$(10^6, 10^6)$	4	23.3667	$0.919\ 10^{-6}$	$0.303\ 10^{-1}$	772.2
$(10^{-6}, 10^6)$	$(10^6, 10^{-6})$	5	23.3682	$0.133\ 10^{-7}$	$0.394\ 10^{-1}$	593.9
$(10^{-6}, 10^{-6})$	$(10^{-6}, 10^6)$	2	0.02763	$0.782\ 10^{-9}$	$0.542\ 10^{-3}$	50.9
$(10^{-6}, 10^6)$	$(10^6, 10^6)$	5	23.3671	$0.588\ 10^{-6}$	$0.393\ 10^{-1}$	595.4
$(10^{-6}, 10^{-6})$	$(10^6, 10^{-6})$	5	23.3671	$0.588\ 10^{-6}$	$0.394\ 10^{-1}$	593.9
$(10^6, 10^{-6})$	$(10^6, 10^6)$	7	29.6856	$0.555\ 10^{-6}$	$0.603\ 10^{-1}$	492.5

Table 7
Summary statistics from a sample of size 10000

	Average	Std. deviation	Minimum	Maximum
Iterations	5.6166	1.0456	2	13
Distance	11.326	5.7267	0.0017	27.766
Time (s)	0.0219	0.0129	0.0005	0.1399
Speed	547.66	118.12	3.1430	1701.3

statistics of the results after simulation. Observe that: for great enough average distance the program needed a small number of iterations on average and also the average time taken to solve the problem is low.

5. Summary

A computer program to evaluate Rao's distance between Gamma populations is implemented. In summary the main features of the program are; (1) using geometric properties of the Gamma manifold a good starting value for the iterative process is supplied, then the program shows a greater range of convergence than other standard routines, (2) the Newton approximation at each iteration is

obtained by solving the variational equations by the integration of normal Jacobi fields determined by the boundary conditions, yielding a fast iterative procedure. A FORTRAN program to compute Rao distance between Gamma distributions is available on request.

Acknowledgements

References

[1] C. Atkinson, A.F.S. Mitchell, Rao's distance measure, Sankhyā A 43 (1981) 345–365.
[2] A. Bhattacharyya, On a measure of divergence between two statistical populations, Bull. Calcutta Math. Soc. 35 (1943) 99–109.
[3] J. Burbea, Informative geometry of probability spaces, Exposition. Math. 4 (1986) 347–378.
[4] J. Burbea, J.M. Oller, The information metric for univariate linear elliptic models, Statist. Decisions 6 (1988) 209–221.
[5] J. Burbea, J.M. Oller, F. Reverter, Some geometrical quantities of the Gamma statistical model, preprint (Universitat de Barcelona, 2000).
[6] J. Burbea, C.R. Rao, Entropy differential metric, distance and divergence measures in probability spaces: a unified approach, J. Multivariate Anal. 12 (1982) 575–596.
[7] J. Burbea, C.R. Rao, Differential metrics in probability spaces, Probab. Math. Statist. 3 (1984) 241–258.
[8] I. Chavel, Riemannian Geometry: A Modern Introduction, Cambridge University Press, Cambridge, 1993.
[9] C.M. Cuadras, Distance analysis in discrimination and classification using both continuous and categorical variables, in: Y. Dodge (Ed.), Recent Developments in Statistical Data Analysis and Inference, Elsevier, Amsterdam, 1989, pp. 459–474.
[10] DIGITAL visual fortran professional, Digital, 1997.
[11] S.L. Lauritzen, The Gamma Manifold, in: S. Amari, O.E. Barndoff-Nieldon, R.E. Kass, S.L. Lauritzen, C.R. Rao (Eds.), Differential Geometry in Statistical Inference, Lectures Notes—Monograph Series, Vol. 10, Institute of Mathematical Statistics, Hayward, California, 1987, pp. 203–205.
[12] P.C. Mahalanobis, On the generalized distance in statistics, Proc. Nat. Inst. Sci. 12 (1936) 49–55.
[13] K. Matusita, Distance and decision rules, Ann. Inst. Statist. Math. 16 (1964) 305–315.
[14] A.F.S. Mitchell, W.J. Krzanowski, The Mahalanobis distance and elliptic distributions, Biometrika 72 (1985) 464–467.
[15] The NAG Fortran Library Manual-Mark 17 (NAG Eds), Oxford University Press, Oxford, 1995.
[16] J.M. Oller, C.M. Cuadras, Rao's distance for negative multinomial distributions, Sankhyā A 47 (1985) 75–83.
[17] C.R. Rao, Information and accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc. 37 (1945) 81–91.
[18] F. Reverter, Métodos computacionales para el cálculo de distancias Riemannianas, Aplicaciones a la geometria informativa correspondiente a la distribución Gamma, Ph.D. Thesis, Universitat de Barcelona, Barcelona, 1999.
[19] F. Reverter, J.M. Oller, A modified Taylor series method for solving initial value problems in ordinary differential equations, Internat. J. Comput. Math. 65 (1997) 231–246.
[20] M. Spivak, A Comprehensive Introduction to Differential Geometry, Publish or Perish, Berkeley, 1980.
[21] D. Wauters, L. Vermeire, Intensive numerical and symbolic computing in parametric test theory, in: W. Hardle, L. Siman (Eds.), Computer Intensive Methods in Statistics, Physica-Verlag, Heidelberg, 1993, pp. 62–72.