Some illustrations of information geometry in biology and physics

Dodson, CTJ

2011

MIMS EPrint: **2011.68**

Manchester Institute for Mathematical Sciences

School of Mathematics

The University of Manchester

Some illustrations of information geometry
in biology and physics[*]

C.T.J. Dodson

School of Mathematics, University of Manchester UK

Abstract

Many real processes have stochastic features which seem to be representable in some intuitive sense as 'close to Poisson', 'nearly random', 'nearly uniform' or with binary variables 'nearly independent'. Each of those particular reference states, defined by an equation, is unstable in the formal sense, but it is passed through or hovered about by the observed process. Information geometry gives precise meaning for nearness and neighbourhood in a state space of processes, naturally quantifying proximity of a process to a particular state via an information theoretic metric structure on smoothly parametrized families of probability density functions. We illustrate some aspects of the methodology through case studies: inhomogeneous statistical evolutionary rate processes for epidemics, amino acid spacings along protein chains, constrained disordering of crystals, distinguishing nearby signal distributions and testing pseudorandom number generators.

Keywords: Families of probability densities, gamma distributions, information geometry, amino acids, constrained disordering, evolution, epidemic, inhomogeneous rate process, entropy, analytic and numerical computation.

1   Introduction

A question: *"We already use statistical modeling, why should we bother with information geometry?"* Information geometry is concerned with the natural geometrization of smoothly parametrized families of discrete probability or continuous probability density functions; the naturality stems from the fact that the metric structure arises from the covariance matrix of gradients of probability. This metric yields a smooth Riemannian structure on the space of parameters, so adding the geometric concepts of curvature and arc length to the analytic tools for studying trajectories through probability distributions as statistical models evolve with time or during changes of system conditions. The development of the subject over the past 65 years has been substantially due to the work of C.R. Rao and S-I. Amari and coworkers; see for example [Rao, 1945], [Amari, 1963], [Amari, 1968], [Amari, 1985], [Amari et al, 1987], [Amari & Nagaoka, 2000] and references therein. Information geometry and its applications remain vigorous research areas, as witness for example the series of international conferences of the same name [IGA Conference, 2010]. In phenomenological modeling applications, information geometric methods complement the standard statistical tools with techniques of representation similar to those used in physical field theories where the analysis of curved geometrical spaces have contributed to the understanding of phenomena and development of predictive models.

In many statistical models of practical importance there is a small range of probability density functions that has very wide application as a result of general theorems, and the spaces of these families have just a small number of dimensions. For example, the families of Gaussian and gamma distributions and their bivariate versions are widely applied and moreover their information geometry is easily tractable, [Arwini & Dodson, 2008]. In particular, the family of gamma distributions is ubiquitous in modeling natural processes that involve scatter of a positive random variable around a target state, such as for inhomogeneous populations or features of elements in a collection. The reason for this ubiquity is that a defining characteristic of the gamma distribution is for the sample standard deviation to be proportional to the sample mean. In practice, that property is commonly found to varying degrees of approximation; the case when the standard deviation *equals* the mean corresponds to the exponential distribution associated with a Poisson process, which is the fundamental reference process for statistical models. Sums of independent gamma random variables (hence also sums of independent exponential random variables) follow a gamma distribution and products of gamma random variables have distributions closely approximated by gamma distributions. We shall provide below more details about the properties of the gamma family and its associated families which include the uniform distribution, approximations to truncated Gaussians and a wide range of others.

Our case studies will show that gamma distributions model well the spacings between successive occurrences of each of the 20 different amino acids which with differing abundances lie along a protein chain [Cai et al, 2002]. Figure 6 illustrates the information distance in the space of gamma distributions for amino acid spacings along protein chains, measured from the exponential (Poisson) case $\kappa = 1$; intuitively we might expect that they would be scattered around the reference exponential case. In fact they all lie on the clustered side of the distribution, all have more variance than that expected by chance—the exponential case.

In typical real situations it is of interest to depict the changing state of a statistical model through the trajectory its representative distribution follows in the appropriate family of distributions, under the influence of external influences or some internal evolutionary imperative. An example is shown in Figure 1 for the integral curves starting from different initial gamma distributions of the entropy (ie the 'mean log probability density') gradient in the space of gamma distributions; there the unconstrained disordering means that the asymptote coincides with the exponential distribution, when the standard deviation equals the mean and we have maximal entropy (and hence maximal disorder). Information geometry provides the correct measures of 'information distance' along or between such trajectories, and along any other arbitrary curves, and it defines parallelism and perpendicularity as well as minimal distance curves (geodesics). Sometimes the degeneration of order is constrained by conditions in the model and then the process does not tend to the maximal entropy but only to a lower level. An example of a class of stochastic phenomena that involves the degeneration of a structure from more orderly to less orderly, through an external application of disruptive statistical influences, is the heating of a crystalline structure. Here there is what might be called a constrained degeneration into disorder, where the constraints are the structural rules that define the material. We shall discuss in the sequel the results from [Lucarini, 2009], which simulates the progression of the disordering degeneration of 3D BCC crystal structure via surface area distributions of Voronoi cells.

Some applications require bivariate statistical models in which correlation is a parameter. Examples are when these variables are lengths of adjacent polygon edges in stochastic fibre networks, and the void and capillary size in stochastic porous media; [Arwini & Dodson, 2008] provides details for such cases. Here, by extending an existing product epidemic model [Britton & Lindenstrand, 2009] we shall illustrate a statistically similar situation using a bivariate gamma representation of the random variables for periods of latency and infectiousness; see [Andersson & Britton, 2000] for more details on the modeling of epidemics. Epidemics in inhomogeneous populations have disease transmission susceptibilities and infectiousness that vary within the population, so we have an inhomogeneous stochastic rate process and this admits interesting representation through information geometric methods. In the sequel, Figures 4 and 5, show how we can depict the parameters in the joint distribution of periods of latency and infectiousness as surfaces of distance, measured from the two reference cases for the evolution of the epidemic, namely starting from Poisson processes for each variable. On such surfaces could be represented data on the progress of epidemics under

different intervention schemes, or simulations of such scenarios.

Another situation where we encounter an inhomogeneous stochastic rate process is that of an evolutionary development of a population through the stochastic de-selection of individuals with unfavourable features. Here we are interested in how the distribution of features develops and what its effect is on the properties of the population. Particular solutions for the cases of initial densities that were Poisson, gamma or uniform were given by [Karev, 2003]. Elsewhere [Dodson, 2010b] reported a study of evolution from initial log-gamma densities that determine a neighbourhood of the uniform distribution [Arwini & Dodson, 2008], so recovering the solution for the uniform distribution as a special case. For a recent account of numerical methods in modeling evolution see [Roff, 2010].

Pseudorandom number generators are common tools in software engineering and we can use information geometry to compare them. Cryptological attacks on encryption/decryption devices may be defended against by obscuring algorithms that overlay randomizing procedures; then there is a need to compare nearby signal distributions and again the information metric can help. Tests for randomness of such sequences have been studied extensively and the NIST Suite of tests [Rushkin, Soto et al, 2001] for cryptological purposes is widely employed. Information theoretic methods also are used, for example see [Grzegorzewski & Wieczorkowski, 1999] also [Ryabko & Monarev, 2005] and references therein for recent work. Here we add to the latter by outlining how pseudorandom sequences may be tested quickly and easily using information geometry by computing distances in the gamma manifold to compare maximum likelihood parameters for separation statistics of sequence elements.

Finally, we utilize the log-gamma family and compare nearby truncated Gaussian-like distributions of arbitrarily small variance and also perturbations of the uniform distribution, as illustrated in Figure 8. Such comparisons are used in a variety of signal analytic applications and in particular we have applied them to studies of cryptological security of smartcards [Dodson & Thompson, 2000], providing simple approximate information distance formulae.

1.1   Software

In the case studies reported in this chapter we have used the computer algebra package *Mathematica* [Wolfram, 1996] for the necessary analytic differential geometry, to perform computer simulations and to represent graphically the typically important process features for a variety of phenomena. The analytic mathematical and numerical computational methodology used here is supported by open source code in the form of *Mathematica* interactive notebooks [Dodson, 2010c] which can be downloaded from the author's webpage http://www.maths.manchester.ac.uk/kd/mmaprogs/InfoGeomMMANotebooks/
Such notebooks allow interface and exchange of data with many other programs including C, C^{++} and MatLab as well as export in a variety of graphical, animated and tabular formats. The book [Rose & Smith, 2002] provides a wide range of mathematical statistics coding in *Mathematica*. [Gray, 1998] provides *Mathematica* code for the geometry of curves and surfaces.

1.2   Summary of properties of gamma distributions

We begin with some notes on properties of the family of gamma distributions, which has certain distinguishing characteristics that give it a central role in a wide range of applications because it generalizes the duality between the discrete Poisson event distribution and its complementary continuous exponential event spacing distribution.
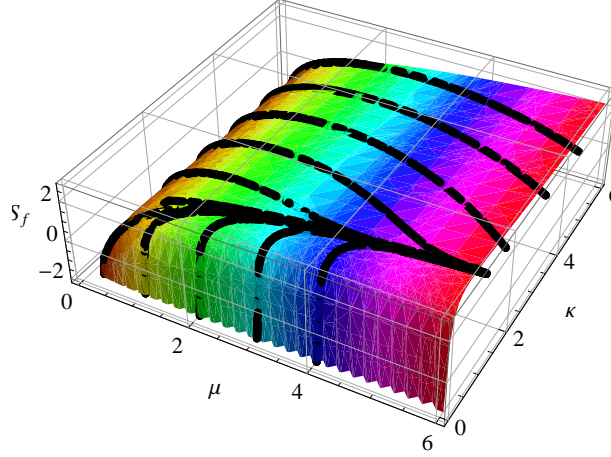
Figure 1: *Entropy function S_f from (2) for the gamma family with entropy gradient flow and integral curves as a surface. On the surface, the dashed asymptote at $\kappa = 1$ is the exponential case of maximum disorder.*

The family of gamma distributions with random variable x in event space $\Omega = \mathbb{R}^+$ has a collection of probability density functions given by

$$\left\{ f(x; \mu, \kappa) = \left(\frac{\kappa}{\mu} \right)^\kappa \frac{x^{\kappa-1}}{\Gamma(\kappa)} e^{-x\kappa/\mu} \mid \mu, \kappa \in \mathbb{R}^+ \right\} \equiv \mathbb{R}^+ \times \mathbb{R}^+. \tag{1}$$

These probability density functions depend in a smoothly differentiable way on parameters $\mu, \kappa \in \mathbb{R}^+$, with mean $E[x] = \mu$ and variance $E[x^2] - E[x]^2 = \sigma^2 = \frac{\mu^2}{\kappa}$. The Shannon entropy S_f of (1) and some gradient flow curves, satisfying $\dot{c} = \nabla S_f$, are shown in Figure 1 from the definition

$$\begin{aligned}
S_f &= -\int_0^\infty f \log f \; dx : \mathbb{R}^{2+} \to \mathbb{R} \\
&: (\mu, \kappa) \mapsto \kappa - \log\left(\frac{\kappa}{\mu} \right) + \log(\Gamma(\kappa)) - (\kappa - 1)\frac{d \log \Gamma(\kappa)}{d\kappa}.
\end{aligned} \tag{2}$$

At fixed κ, the entropy increases like $\log \mu$; at fixed mean μ, the maximum entropy is given by $\kappa = 1$, which determines the exponential distribution case of maximal disorder or chaos. Integral curves are shown through the space of parameters for gamma distributions to illustrate trajectories of unconstrained disordering processes that converge to the line $\kappa = 1$. Constrained disordering, such as for heated crystals, converges not to $\kappa = 1$ but to a higher value, as we shall see below. Given a set of identically distributed, independent data values $X_1, X_2, \ldots, X_n$, the 'maximum likelihood' or 'maximum entropy' parameter values $\hat{\mu}, \hat{\kappa}$ for fitting the gamma distribution (1) are computed in terms of the mean and mean logarithm of the X_i by maximizing the likelihood function

$$Lik_f(\mu, \kappa) = \prod_{i=1}^n f(X_i; \mu, \kappa).$$

By taking the logarithm and setting the gradient to zero we obtain

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \tag{3}$$

$$\begin{aligned}
\log \hat{\kappa} - \frac{\Gamma'(\hat{\kappa})}{\Gamma(\hat{\kappa})} &= \log \bar{X} - \frac{1}{n} \sum_{i=1}^n \log X_i \\
&= \log \bar{X} - \overline{\log X}.
\end{aligned} \tag{4}$$

The smooth family of log-gamma distributions has probability density functions of form

$$h(a, \nu, \tau) = \frac{a^{\nu-1} \nu^{\tau} \left| \log \left(\frac{1}{a} \right) \right|^{\tau-1}}{\Gamma(\tau)} \tag{5}$$

for random variable $a \in (0, 1]$ and parameters $\nu, \tau > 0$, see Figure 8. The mean and variance are given by

$$E(a) = \left(\frac{\nu}{1 + \nu} \right)^{\tau} \tag{6}$$

$$\sigma^2(a) = \left(\frac{\nu}{\nu + 2} \right)^{\tau} - \left(\frac{\nu}{1 + \nu} \right)^{2\tau}. \tag{7}$$

In this family the locus of those with central mean $E(a) = \frac{1}{2}$ satisfies

$$\nu(2^{\frac{1}{\tau}} - 1) = 1. \tag{8}$$

The uniform density is the special case with $\tau = \nu = 1$. The log-gamma density (5) actually arises from the gamma density

$$f(x, \nu, \tau) = \frac{x^{\tau-1} \nu^{\tau}}{\Gamma(\tau)} e^{-x\nu}. \tag{9}$$

via the change of variable $x = -\log a$.

The gamma distribution can be characterized by an important uniqueness property which we have in the form of the following theorem—the proof of which could be omitted in a first reading but the method of proof is important:

Theorem 1 *For independent positive random variables with a common probability density function f, having independence of the sample mean and the sample coefficient of variation is equivalent to f being the gamma distribution.*

This characterization is one of the main reasons for the large number of applications of gamma distributions, because many near-random natural processes have observed standard deviation approximately proportional to the mean, as we illustrate in [Arwini & Dodson, 2008]. A proof of the Theorem 1 was given in [Hwang & Hu, 1999] but in fact the result seems to have been known much earlier and here we give a proof partly based on the article [Laha, 1954], of interest for the methodology using Laplace Transforms, which have a close relation to moment generating functions in statistics [Feller, 1971].

Proof of Theorem 1 Let f be a probability density function for a positive random variable x, so f is positive and of unit measure:

$$\int_0^{\infty} f(x) \, dx = 1.$$

The expectation operator E gives the mean m and variance $\sigma^2 < \infty$ by

$$E[x] = \int_0^{\infty} x \, f(x) \, dx = m \quad \text{and} \quad E[(x - m)^2] = \int_0^{\infty} x^2 \, f(x) \, dx - m^2 = \sigma^2.$$

The Laplace Transform of f is therefore the expectation of e^{-tx} which we write as

$$\phi_x(t) = E[e^{-tx}] = \int_0^{\infty} e^{-tx} f(x) \, dx. \tag{10}$$

Its nth derivative is

$$\phi_x^{(n)}(t) = (-1)^n \int_0^\infty x^n \, e^{-tx} \, f(x) \, dx$$

and so in particular

$$\phi_x'(0) = -m, \ \phi_x''(0) = \sigma^2 + m^2.$$

If y is another positive random variable, with probability density function g and y is independent of x then from (10) the probability density function of $z = x + y$ has Laplace Transform

$$\phi_z(t) = E[e^{-tz}] = E[e^{-tx}e^{-ty}] = E[e^{-tx}]E[e^{-ty}] = \phi_x(t)\phi_y(t)$$

It follows that if $x_1, x_2, \ldots, x_n$ are independent random variables all with the same probability density function f, then the pdf of the sum $w = \sum_{j=1}^n x_j$ has Laplace Transform $\phi_w(t) = (\phi_x(t))^n$.

Now let $x_1, x_2, \ldots, x_n$ be independent positive random variables all with the same probability density function f. If the sum $\sum_{j=1}^n x_j$ is independent of the ratio

$$(\sum_{j=1}^n x_j^2)/(\sum_{j=1}^n x_j)^2,$$

then we have to show that f is a gamma probability density function.

By the hypothesis of independence,

$$E\left[\frac{\sum_{j=1}^n x_j^2}{\left(\sum_{j=1}^n x_j\right)^2} (\sum_{j=1}^n x_j)^2 \, e^{t\sum_{j=1}^n x_j} \right] = E\left[\frac{\sum_{j=1}^n x_j^2}{(\sum_{j=1}^n x_j)^2} \right] E\left[(\sum_{j=1}^n x_j)^2 \, e^{t\sum_{j=1}^n x_j} \right].$$

Using the above result and dividing through by ϕ^n,

$$n\phi_x''/\phi_x = K\left(n\phi_x''/\phi_x + n(n-1)\left(\phi_x'/\phi_x \right)^2 \right)$$

Now put $\psi(t) = \log \phi_x(t)$ and simplify to give

$$A\psi'' + B(\psi')^2 = 0, \quad \text{with } \psi(0) = 0, \psi'(0) = m, \psi''(0) = \sigma^2 \quad \text{so } \frac{\psi''(t)}{\psi'(t)^2} = -\frac{\sigma^2}{m^2}. \tag{11}$$

Put $\xi(t) = \psi'(t)$ then rewrite (11) as

$$-\frac{\psi''(t)}{\psi'(t)^2} = \frac{d}{dt} 1/\xi(t) = \frac{\sigma^2}{m^2} \quad \text{hence } \xi(t) = \frac{m}{1 + (\sigma^2/m)t}.$$

Integrating $\xi(t) = \psi'(t)$ gives

$$\psi(t) = \log \phi_x(t) = (m^2/\sigma^2) \log[1 + (\sigma^2/m)t] \quad \text{finally } \phi_x(t) = \left(1 + (\sigma^2/m)t \right)^{m^2/\sigma^2}. \tag{12}$$

It can be checked that this ϕ_x is the Laplace Transform of the gamma probability density function

$$f(x) = \left(\frac{m}{\sigma^2} \right)^{(m/\sigma)^2} \frac{e^{-mx/\sigma^2}}{\Gamma(m^2/\sigma^2)} x^{(m/\sigma)^2 - 1}$$

which completes the proof of Theorem 1.

2 Background information geometry for stochastic processes

There are many books on differential geometry, we mention [Dodson, & Poston, 1991] and [Gray, 1998] which introduce the ideas of Riemannian geometry and the synthesis of smooth n-manifolds by smoothly piecing together subsets of $\mathbb{R}^n$, and include many examples. Let $\Theta \subseteq \mathbb{R}^n$ be the parameter space of a smooth family of probability density functions defined on some fixed event space Ω (typically $\mathbb{R}$, $\mathbb{R}^+$ or products thereof)

$$\{p_\theta | \theta \in \Theta\} \quad \text{with} \quad \int_\Omega p_\theta = 1 \quad \text{for all } \theta \in \Theta. \tag{13}$$

The covariance matrix $[g_{ij}]$ is the expectation of the matrix of derivatives of the log-likelihood function $l = \log p_\theta$, with respect to parameters (θ^i), it is positive definite and hence defines a Riemannian information metric on the smooth n-manifold of probability density functions with coordinates these parameters (θ^i). The components of the Riemannian metric are, subject to certain regularity conditions, cf. [Arwini & Dodson, 2008],

$$[g_{ij}] = \left[E\left(\frac{\partial l}{\partial \theta^i} \frac{\partial l}{\partial \theta^j} \right) \right] = -\left[E\left(\frac{\partial^2 l}{\partial \theta^i \partial \theta^j} \right) \right], \quad \text{giving arc length function } ds^2 = \sum_{i,j} g_{ij}\, d\theta^i\, d\theta^j. \tag{14}$$

We say that the family (13) is an exponential family if the p_θ admit expression in terms of functions $\{C, F_1, ..., F_n\}$ on Ω and a function φ on Θ as:

$$p_\theta(x) = e^{\{C(x) + \sum_i \theta_i\, F_i(x) - \varphi(\theta)\}}, \tag{15}$$

then we say that these (θ_i) are its natural parameters, and φ is the potential function. By normalization we obtain:

$$\int p(x; \theta)\, dx = 1, \quad \text{so } \varphi(\theta) = \log \int e^{\{C(x) + \sum_i \theta_i\, F_i(x)\}}\, dx. \tag{16}$$

In this case the n-manifold of probability density functions can be represented by a natural affine immersion (cf. [Dodson & Matsuzoe, 2003], [Arwini & Dodson, 2008]) in $\mathbb{R}^{n+1}$ via

$$(\theta) \in \mathbb{R}^n \mapsto ((\theta), \varphi(\theta)) \in \mathbb{R}^{n+1}. \tag{17}$$

For an exponential family (15) we have a simpler method to compute the information metric (14) from the log-likelihood function $l(\theta, x) = \log p_\theta(x)$ as follows.

$$\partial_i l(\theta, x) = F_i(x) - \partial_i \varphi(\theta) \tag{18}$$

and

$$\partial_i \partial_j l(\theta, x) = -\partial_i \partial_j \varphi(\theta). \tag{19}$$

Then the information metric g on the n-dimensional space of parameters $\Theta \subset \mathbb{R}^n$, equivalently on the set $\{p_\theta | \theta \in \Theta \subset \mathbb{R}^n\}$, has components:

$$[g_{ij}](\theta) = -\int_\Omega [\partial_i \partial_j l(\theta, x)]\, p_\theta(x)\, dx = \partial_i \partial_j \varphi(\theta) = [\varphi_{ij}](\theta). \tag{20}$$

Surprisingly, (14) is tractable for many important smoothly parametrized distributions, in particular for Gaussian, exponential and gamma, as well as their bivariate versions. For example, the space $\mathcal{G}$ of gamma probability density functions (1) is a smooth 2-manifold (here a curved surface) with Riemannian metric

$$[g_{ij}](\mu, \kappa) = \begin{bmatrix} \frac{\kappa}{\mu^2} & 0 \\ 0 & \frac{d^2}{d\kappa^2}\log(\Gamma) - \frac{1}{\kappa} \end{bmatrix} \quad \text{and arc length function } ds^2 = \frac{\kappa}{\mu^2}\, d\mu^2 + \left(\frac{d^2}{d\kappa^2}\log(\Gamma) - \frac{1}{\kappa} \right) d\kappa^2. \tag{21}$$
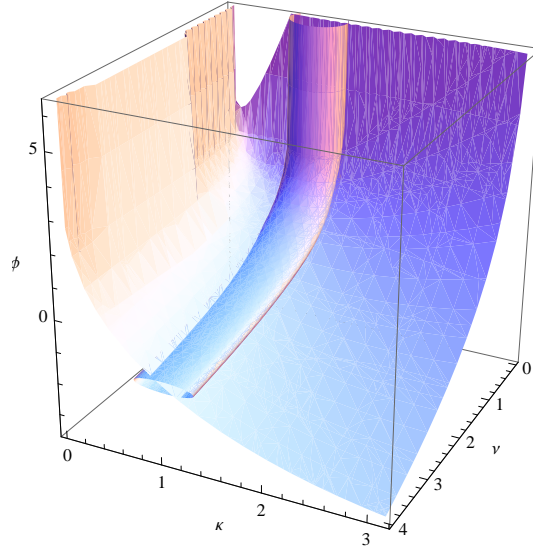
Figure 2: *Tubular neighbourhood of the curve $\kappa = 1$ of all exponential random processes, in the surface representing the gamma 2-manifold in $\mathbb{R}^3$ as an affine immersion $\phi(\nu, \kappa) = \log \Gamma(\kappa) - \kappa \log \nu$ in natural coordinates $\nu = \kappa/\mu, \kappa$.*

A curve in the space $\mathcal{G}$ of gamma distributions

$$c : [0, 1] \to \mathcal{G} : t \mapsto \{c_1(t), c_2(t)\} \tag{22}$$

has tangent vector $\{\dot{c}_1(t), \dot{c}_2(t)\}$ where the dot signifies differentiation by t and its information length up to $t = T$ is

$$L(T) = \int_0^T \sqrt{\sum_{i,j} \dot{c}_i(t) \dot{c}_j(t) \, g_{ij}(c_1(t), c_2(t))} \, dt \tag{23}$$

It is easy to show that $(\nu = \kappa/\mu, \kappa)$ are natural parameters for the gamma distribution and its potential function is

$$\varphi(\nu, \kappa) = \log \Gamma(\kappa) - \kappa \, \log \nu. \tag{24}$$

In these coordinates the components of the metric are given by

$$[g_{ij}] (\nu, \kappa) = \begin{bmatrix} \frac{\kappa}{\nu^2} & -\frac{1}{\nu} \\ -\frac{1}{\nu} & \frac{d^2}{d\kappa^2} \log(\Gamma) \end{bmatrix} . \tag{25}$$

Figure 2 shows the corresponding affine immersion in $\mathbb{R}^3$ of the 2-manifold $\mathcal{G}$ of gamma distributions, together with a tubular neighbourhood of the curve $\kappa = 1$ which contains all exponential distributions. This is significant because exponential distributions are in one-to-one correspondence with Poisson processes— Poisson processes on a line have exponentially distributed intervals between events and exhibit maximal disorder.

By inspection of (14), the same information geometry arises from the logarithmic version in each case. Also, among the gamma distributions is included the exponential distribution as a special case so among the log-gamma distributions we have also the uniform distribution, as well as approximations to truncated Gaussians of arbitrarily small variance, among other asymmetric unimodal distributions.

The log-gamma family yields a smooth Riemannian 2-manifold with coordinates $(\nu, \tau) \in \mathbb{R}^+ \times \mathbb{R}^+$ and metric tensor given by

$$[g_{ij}](\nu, \tau) = \begin{bmatrix} \frac{\tau}{\nu^2} & -\frac{1}{\nu} \\ -\frac{1}{\nu} & \frac{d^2}{d\tau^2} \log(\Gamma) \end{bmatrix}, \quad \text{with } ds^2 = \frac{\tau}{\nu^2} d\nu^2 + \frac{d^2}{d\tau^2} \log[\Gamma] d\tau^2 - \frac{2}{\nu} d\nu d\tau. \tag{26}$$

In fact, this 2-manifold of log-gamma densities is an isometric diffeomorph (smooth 1 to 1 correspondence) of the 2-manifold of gamma densities (9), via natural coordinates (ν, τ) (cf. [Amari & Nagaoka, 2000], [Arwini & Dodson, 2008]) for which $\tau = 1$ corresponds to the subfamily of exponential densities with mean $\frac{1}{\nu}$. So the information geometry of gamma and log-gamma families coincide and the correspondence maps the exponential distribution with unit mean to the uniform distribution which has $(\nu = 1, \tau = 1)$.

2.1 Characterization of important neighbourhoods of process states

Some other useful general results that arise in this context and have wide application are given by the following theorems from [Arwini & Dodson, 2008]

Theorem 2 *Every neighbourhood of a Poisson process contains a neighbourhood of processes subordinate to gamma probability density functions.*
A Poisson process defines a unique exponential distribution, the exponential distributions are special cases of gamma distributions and the information geometry of the gamma family determines a metric structure for neighbourhoods. Figure 2 shows a tubular neighbourhood in $\mathbb{R}^3$ of the curve of exponential distributions in an immersion of the 2-dimensional space of gamma distributions.

Theorem 3 *Every neighbourhood of a uniform process contains a neighbourhood of processes subordinate to log-gamma probability density functions.*
The log-gamma distributions contain the uniform distribution as a special case; logarithmic transformation of random variables leaves unaltered the information geometry so this follows from Theorem 2. Thus a spherical neighbourhood in $\mathbb{R}^3$ centred on the the uniform distribution in an immersion of the space of log-gamma distributions provides a neighbourhood of uniformity.

Theorem 4 *Every neighbourhood of an independent pair of identical Poisson processes contains a neighbourhood of bivariate processes subordinate to Freund bivariate exponential probability density functions.*
The 4-dimensional space of Freund bivariate mixtures of exponential distributions [Freund, 1961] contains subfamilies that include correlated identical exponential distributions, so the Freund information geometry provides a metric structure for neighbourhoods of independent exponential distributions and consequently for their corresponding complementary Poisson distributions.

Theorem 5 *The 5-dimensional space of bivariate Gaussians admits a 2-dimensional subspace through which can be provided a neighbourhood of independence for bivariate Gaussian processes.*
The information geometry of the bivariate Gaussian family determines a 2-dimensional subfamily that contains the space of independent Gaussians, so determining a neighbourhood of independence.

Theorem 6 *Via the Central Limit Theorem, by continuity, the tubular neighbourhoods of the curve of zero covariance for bivariate Gaussian processes will contain all limiting bivariate processes sufficiently close to the independence case for all processes with marginals that converge in probability density function to Gaussians.*

These results are rather general and their qualitative features make them widely applicable. We shall illustrate some aspects in this chapter through several case studies in the following sections for processes that

have stochastic features which seem to be representable in some intuitive sense as 'close to Poisson', 'nearly random', 'nearly uniform' or with binary variables 'nearly independent'. Those particular reference states, being defined by an equation, are unstable in the mathematical statistical sense, but are passed through or hovered about by many important observed processes and hence we need to consider neighbourhoods. The fact that we have the topological property of geometrical 'nearness' means that qualitative results remain true even if the actual models (choice of distribution family) is only an approximation. We look in some detail at inhomogeneous statistical rate processes for epidemics and evolution, and outline applications to amino acid spacings, constrained disordering, distinguishing nearby signal distributions and testing pseudorandom number generators. This should not be viewed as a substitution for other methods of study but rather as an additional methodology that may be useful in representation or evaluation of phenomena.

3   Case Study 1: Modeling a stochastic epidemic rate process

Epidemiology is a big subject with a long history and a large literature; standard texts on modeling include [Bailey, (1975)], [Andersson & May, 1991], [Diekmann & Heesterbeek, 2000], note also the new volume on numerical methods [Andersson & Britton, 2000].

The spreading of an infectious disease involves a (large) population in which an initially small number of individuals are infected. Each infected individual is for a period of latency L not yet infectious but at the end of the latent period the individual becomes infectious for a period I; models provide distributions for the random variables L and I. The rest of the population is susceptible to infection from infectious individuals; this susceptibility is often taken to be a constant but later we shall consider a population with an evolving inhomogeneous distribution of susceptibilities to infection. An infectious individual has random infectious contacts at a rate λ and contact with a susceptible individual results in infection and then the latent period of that individual commences. The so-called basic reproduction number is $R_0 = \lambda \mu_I$, the product of the rate of infectious contacts and the mean period of infectiousness μ_I.

[Lloyd, 2001] discussed the sensitivity of dynamical properties of an epidemic model to the choices of formulation and made use of a gamma distribution for the period of infectiousness and allowed for optional seasonality. [Cauchemez & Ferguson, 2008] introduced a new approach to the analysis of epidemic time series data to take account of partial observation of latency and the temporal aggregation of observed data. They showed that homogeneous standard models can miss key features of epidemics in large populations. Also, [Nishiura et al, 2009] devised an estimate of reproduction number in terms of coarsely reported epidemic data, showing that an ideal reporting interval is the mean generation time rather than a fixed chronological interval. See also recent work by [He et al, 2009] for related results on partially observed data and by [Miller et al, 2009] on general distributions of generating intervals.

[Chowell et al, 2009] have edited a new collection of articles on mathematical and statistical approaches to epidemic modelling and Chapter 2 there, by G. Chowell and F. Bauer, gives a detailed study of the basic reproduction rate in a variety of epidemic models. [Wallinga & Lipsitch, 2007] addressed the sensitivity of the reproduction number to the shape of the distribution of generation intervals and obtained upper bounds even in the situation of no information on shape.

Recently [Britton & Lindenstrand, 2009] described a model where the period of latency L and the period of infectiousness I have independent gamma distributions. They found that variability in these random variables had opposite effects on the epidemic growth rate. That rate increased with greater variability in L but decreased with greater variability in I. Here we extend their result by using the McKay bivariate gamma distribution [McKay, 1934] for the joint distribution of L and I, recovering the above effects of variability but allowing in case it may be of relevance the possibility of correlation. One might imagine that in the case of a

disease in which the physical changes during latency lead to longer future infectiousness if the period of their development is longer, then the random variables L and I may have a positive correlation. We use methods of stochastic rate processes to obtain explicit solutions for the growth of the epidemic and the evolution of the inhomogeneity and information entropy. This admits a closed analytic solution to the evolution of the distribution of the number of uninfected individuals as the epidemic proceeds, and a concomitant expression for the decay of entropy. The family of McKay bivariate gamma distributions has a tractable information geometry which provides a framework in which the evolution of distributions can be studied as the outbreak grows, with a natural distance structure for quantitative tracking of progress.

3.1   Inhomogeneous Malthusian epidemic models

In their discussion of epidemic modelling, [Britton & Lindenstrand, 2009] highlighted aspects when stochastic features are more important than deterministic ones. In particular, they described the importance of admitting random variables to represent the period of latency L and the period of infectiousness I, Their standard susceptible-exposed-infectious-removed (SEIR) epidemic model was elaborated using independent gamma distributions for L and I with means μ_L, μ_I and standard deviations σ_L, σ_I. The basic (mean) reproduction number is given by

$$R_0 = \lambda \mu_I \tag{27}$$

where λ is the rate of infectious contacts and μ_I is the mean length of infectious period. An epidemic becomes a major outbreak if $R_0 > 1$ and then the number infected increases exponentially,

$$n_I(t) \sim e^{rt} \tag{28}$$

where the Malthusian parameter r satisfies the equation

$$E\left(e^{-rt}\lambda \; Prob\{L < t < L + I\}\right) = 1. \tag{29}$$

Their independent bivariate model expresses $\frac{r}{R_0}$ in terms of the parameters of the two gamma distributions. They used means, μ_L, μ_I and coefficients of variation $\tau_L = \frac{\sigma_L}{\mu_L}, \tau_I = \frac{\sigma_I}{\mu_I}$ to deduce

$$r = \frac{R_0}{\mu_I} \left(1 + r\tau_L^2\mu_L\right)^{-1/\tau_L^2} \left(1 - \left(1 + r\tau_I^2\mu_I\right)^{-1/\tau_I^2}\right). \tag{30}$$

Then [Britton & Lindenstrand, 2009] found from numerical analysis of (30) that, at fixed R_0, the growth rate r is monotonically decreasing with μ_L, μ_I and τ_I, but it is increasing with τ_L. So increased variability in latency period increases the epidemic growth rate whereas increased variability in infectious period decreases the epidemic growth rate.

3.2   Bivariate gamma distribution of periods of latency and infectiousness

The model described here adds to the work of [Britton & Lindenstrand, 2009] in which they used *independent* univariate gamma distributions for the periods of latency and infectiousness in an epidemic model that they illustrated with data from the SARS outbreak [WHO, 2003]. They used numerical methods to obtain approximate solutions. Our contribution is to use a bivariate gamma distribution which allows positive correlation between the random variables representing the periods of latency and infectiousness. That could represent a situation where physical changes during the latency period lead to longer future infectiousness if the period of their development is longer. We obtain a closed analytic solution and show that the same qualitative features persist in the presence of such correlation. This makes available the analytic information geometry of the space of probability densities, allowing comparison of possible trajectories for the epidemic

against, for example, exponential distributions for periods of infectiousness or of latency, corresponding to underlying Poisson processes.

Somewhat surprisingly, it is rather difficult to devise bivariate versions of Poisson, exponential distributions or more generally gamma distributions that have reasonably simple form [Arwini, 2004], [Arwini & Dodson, 2008] and indeed only Freund bivariate exponential [Freund, 1961] and McKay [McKay, 1934] bivariate gamma distributions seem to have tractable information geometry. The family of McKay bivariate gamma density functions is defined on $0 < x < y < \infty$ with parameters $\alpha_1, \sigma_{12}, \alpha_2 > 0$ and probability density functions,

$$f(x, y; \alpha_1, \sigma_{12}, \alpha_2) = \frac{\left(\frac{\alpha_1}{\sigma_{12}}\right)^{\frac{(\alpha_1 + \alpha_2)}{2}} x^{\alpha_1 - 1} (y - x)^{\alpha_2 - 1} e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}} y}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \ . \tag{31}$$

Here σ_{12}, which must be positive, is the covariance of x and y and $f(x, y)$ is the probability density for the two random variables x and $y = x + z$ where x and z both have gamma density functions.

We obtain the means, standard deviations and coefficients of variation by direct integration:

$$\text{Means}: \qquad \mu_x = \sqrt{\alpha_1 \sigma_{12}}, \quad \mu_y = \frac{(\alpha_1 + \alpha_2)\sqrt{\sigma_{12}}}{\sqrt{\alpha_1}}, \quad \mu_z = \frac{\alpha_2 \sqrt{\sigma_{12}}}{\sqrt{\alpha_1}} \tag{32}$$

$$\text{SDs}: \qquad \sigma_x = \sqrt{\sigma_{12}}, \quad \sigma_y = \sqrt{\frac{\sigma_{12}(\alpha_1 + \alpha_2)}{\alpha_1}}, \quad \sigma_z = \sqrt{\frac{\alpha_2 \sigma_{12}}{\alpha_1}} \tag{33}$$

$$\text{CVs}: \qquad \tau_x = \frac{1}{\sqrt{\alpha_1}}, \quad \tau_y = \frac{1}{\sqrt{\alpha_1 + \alpha_2}}, \quad \tau_z = \frac{1}{\sqrt{\alpha_2}} \tag{34}$$

The correlation coefficient, and marginal probability density functions of x and y are given by

$$\rho \quad = \quad \sqrt{\frac{\alpha_1}{\alpha_1 + \alpha_2}} > 0 \tag{35}$$

$$f_1(x) \quad = \quad \frac{\left(\frac{\alpha_1}{\sigma_{12}}\right)^{\frac{\alpha_1}{2}} x^{\alpha_1 - 1} e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}} x}}{\Gamma(\alpha_1)}, \quad x > 0 \tag{36}$$

$$f_2(y) \quad = \quad \frac{\left(\frac{\alpha_1}{\sigma_{12}}\right)^{\frac{(\alpha_1 + \alpha_2)}{2}} y^{(\alpha_1 + \alpha_2) - 1} e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}} y}}{\Gamma(\alpha_1 + \alpha_2)}, \quad y > 0 \tag{37}$$

The marginal probability density functions of latency period x and infectiousness period y are gamma with shape parameters α_1 and $\alpha_1 + \alpha_2$, respectively. It is not possible to choose parameters such that both marginal functions are exponential, so the two random variables cannot both arise from Poisson processes in this model.

3.3 Stochastic rate processes

For a detailed monograph on stochastic epidemic models see [Andersson & Britton, 2000]. We consider here a class of simple stochastic rate processes where a population N, of uninfected individuals, is classified by a smooth family of time-dependent probability density functions $\{P_t, t \geq 0\}$ with random variable $a > 0$, having at time t mean $E_t(a)$ and variance $\sigma^2(t)$. This situation was formulated by [Karev, 2003], [Karev, 2010b] in the following way. Let $l_t(a)$ represent the frequency at the a-cohort, then we have

$$N(t) \quad = \quad \int_0^\infty l_t(a)\, da \quad \text{and} \ P_t(a) = \frac{l_t(a)}{N(t)} \tag{38}$$

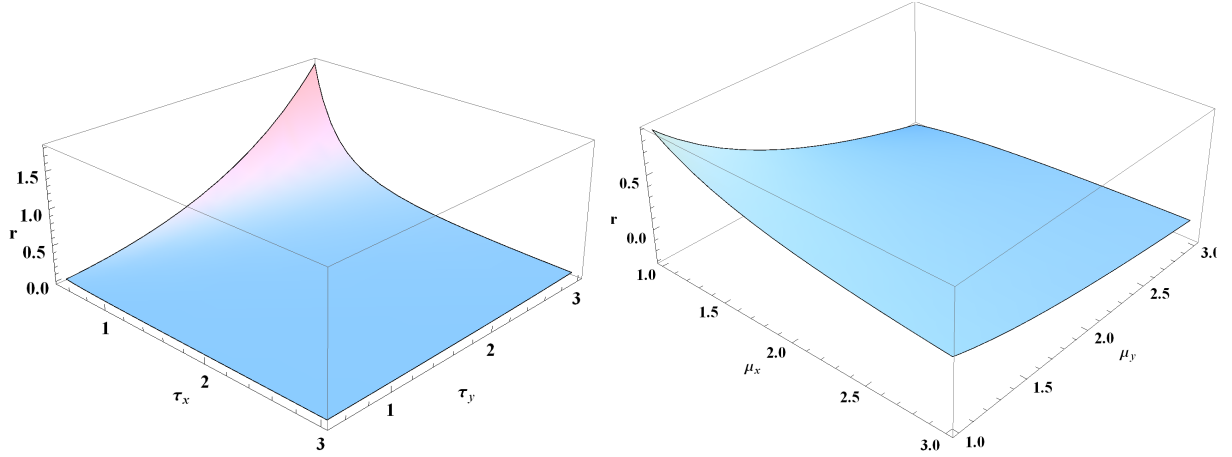$$\frac{dl_t(a)}{dt} \quad = \quad -a l_t(a) \quad \text{so} \ l_t(a) = l_0(a) e^{-at} \tag{39}$$

Figure 3: *Left: Plot of the Malthusian parameter r against the coefficients of variation of periods of latency τ_x and infectiousness τ_y with means $\mu_x = 3$, $\mu_y = 2$ and $R_0 = 2.2$ from [Britton & Lindenstrand, 2009] for the SARS data [WHO, 2003]. So the exponential growth rate of infection decreases with variability in latency period (τ_x) but increases with variability in infectiousness period (τ_y). Right: Plot of the Malthusian parameter r against the means of periods of latency μ_x and infectiousness μ_y with coefficients of variation $\tau_x = \tau_y = 4/7$ and $R_0 = 2.2$ from [Britton & Lindenstrand, 2009] for the SARS data [WHO, 2003]. So the exponential growth rate of infection decreases with mean latency period (μ_x) and with mean infectiousness period (μ_y).*

General solutions for these equations were given in [Karev, 2003], from which we obtain

$$N(t) \quad = \quad N(0)L_0(t) \quad \text{where } L_0(t) = \int_0^\infty P_0(a)e^{-at}\, da \tag{40}$$

$$\frac{dN}{dt} \quad = \quad -E_t(a)\, N \quad \text{where } E_t(a) = \int_0^\infty a\, P_t(a)\, da = -\frac{d\log L_0}{dt} \tag{41}$$

$$\frac{dE_t(a)}{dt} \quad = \quad -\sigma_t^2(a) = (E_t(a))^2 - E_t(a^2) \tag{42}$$

$$P_t(a) \quad = \quad e^{-at}\frac{P_0(a)}{L_0(t)} \quad \text{and } l_t(a) = e^{-at}L_0(t) \tag{43}$$

$$\frac{dP_t(a)}{dt} \quad = \quad P_t(a)(E_t(a) - a). \tag{44}$$

Here $L_0(t)$ is the Laplace transform of the initial probability density function $P_0(a)$ and so conversely $P_0(a)$ is the inverse Laplace transform of the population (monotonic) decay solution $\frac{N(t)}{N(0)}$. See [Feller, 1971] for more discussion of the existence and uniqueness properties of the correspondence between probability densities and their Laplace transforms. In this section we shall use $N(t)$ to represent the decreasing population of *uninfected* individuals as an epidemic grows. In our context of an epidemic model we might view the random variable a as a feature representing susceptibility to infection in the population; in general this distribution will evolve during the epidemic. The model can be reformulated for a vector $(N^i(t))$ representing a composite population with a vector of distributions (P_t^i) and a matrix of variables $[a_{ij}]$.

In [Dodson, 2010b] a similar situation was used in an evolutionary selection process, where the random variable a represents unfitness and the population $N(t)$ declines with time as $P_t(a)$, the distribution of unfitness, evolves towards greater mean fitness. There solutions were found for several initial distributions from the log-gamma family, so investigating a neighbourhood of the uniform initial case.

It is easy to deduce the rate process for entropy from Karev's model. The Shannon entropy at time t is

$$S_t = -E_t\left(\log P_t(a)\right) = -E_t\left(\log \frac{P_0(a)e^{-at}}{L_0(t)}\right) \tag{45}$$

which reduces to

$$S_t = S_0 + \log L_0(t) + E_t(a)\, t. \tag{46}$$

By using $\dot{E}_t(a) = -\sigma^2(t)$, the decay rate is then

$$\frac{dS_t}{dt} = -t\; \sigma^2(t). \tag{47}$$

This result shows how the variance controls the entropy change during quite general inhomogeneous population processes. In fact equation (47) and further related results were given also in subsequent papers [Karev, 2010a], [Karev, 2010b]. We note that the reverse process of population growth may have applications in constrained disordering type situations [Dodson, 2010a].

Initial growth rate
We follow the method of [Britton & Lindenstrand, 2009] in their §3.2, to compute the initial exponential growth rate of the epidemic from equation (29) which we write for bivariate x, y in the form

$$\int_0^\infty \int_y^x e^{-r(y-x)}\, \lambda\, f(x,y)\; dy\, dx = 1 \tag{48}$$

Here, from [Britton & Lindenstrand, 2009], $R_0 = \lambda \mu_x$ for the average number of infections per infective, so λ is the contact rate; this gives the Malthusian parameter analytically in explicit form as

$$r = \frac{1}{\mu_x \tau_x^2}\left(\left(\frac{R_0}{\mu_y}\right)^{\tau_y^2} - 1 \right). \tag{49}$$

Thus, r is monotonically decreasing with μ_x, μ_y and τ_x but increasing with τ_y; typical values from the SARS epidemic [WHO, 2003] as used by [Britton & Lindenstrand, 2009] are $\mu_x = 3$ and $r = 0.053$. So the bivariate gamma model reveals that the result of [Britton & Lindenstrand, 2009] for the dependence of growth rate on variability in the periods of latency and infectiousness in the independent case persists also in the presence of correlation between these two random variables. Such a correlation may be relevant in particular applications, when physical changes evolve during the latent period and influence the length of the subsequent infectiousness period.

We can estimate also the evolution of an inhomogeneous distribution of susceptibility a, as the population $N(t)$ of uninfected individuals declines with time t. For example, the case when the initial distribution $P_0(a)$ is a gamma distribution with parameters s, k was solved in [Karev, 2003] giving the result

$$P_t(a) = \frac{P_0(a)}{L_0(t)} e^{-at} = \frac{(s+t)^k\, a^{k-1}}{\Gamma(k)}\, e^{-a(s+t)}, \quad \text{for time } t \geq 0. \tag{50}$$

Then the time dependences of mean, standard deviation and coefficient of variation are given by

$$\mu_a(t) = \frac{k}{s+t} \tag{51}$$

$$\sigma_a(t) = \frac{\sqrt{k}}{s+t} \tag{52}$$

$$\tau_a(t) = \frac{1}{\sqrt{k}}. \tag{53}$$

From (47), we can see that the *rate* of entropy decrease is greater for more variability in susceptibility.

3.4 Information geometry of the space of McKay bivariate gamma distributions

Information geometry of the smooth family M of McKay bivariate gamma probability density functions, which is of exponential type, has been studied in detail in [Arwini & Dodson, 2008] Chapter 4. This
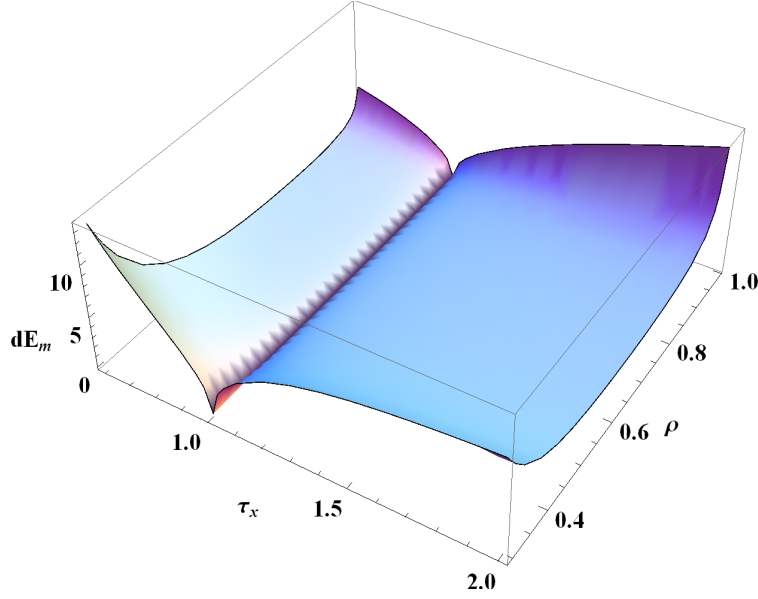
Figure 4: *Approximate information distances $dE_M = \sqrt{E_M}$ (equation (54)) in the McKay manifold, measured from distributions T_0 with exponential marginal distribution for x so $\alpha_1 = 1$ and $\tau_x = 1$. So the surface represents distances from an exponential process for periods of latency.*

provides a Riemannian metric on M, yielding a curved 3-manifold so the affine immersion is a 3-dimensional object in $\mathbb{R}^4$, which we can only represent in $\mathbb{R}^3$ through its 2-dimensional submanifolds. Here we illustrate how the geometry may nevertheless be used to provide a natural distance structure on the space of the McKay distributions used in our epidemic model.

First we measure distances from distributions with exponential marginal distributions—those for which $\alpha_1 = 1$, $\tau_x = 1$ when the latency periods are controlled by a Poisson event process. The derivation of a distance from distribution T_0 is given in [Arwini & Dodson, 2008], and yields in terms of τ_x and ρ

$$
\begin{aligned}
E_M(\tau_x, \rho)_{|[T_0:\alpha_1=1]} &= \frac{\left(\rho^2+1\right)^2}{16\rho^6}\left|\frac{1}{\tau_x^2}-1\right| \\
&+ \frac{1}{4}\left|\left(1-\frac{1}{\tau_x^2}\right)\left(1-\frac{1}{\rho^2}\right)+3\log\left(\tau_x^2\right)\right| \\
&+ \left|\psi\left(\frac{1}{\tau_x^2}\frac{1}{\rho^2}-1\right)-\psi\left(\frac{1}{\rho^2}-1\right)\right| \\
&+ \left|\psi\left(\frac{1}{\tau_x^2}\right)+\gamma\right|
\end{aligned}
\tag{54}
$$

where $\psi(u) = \frac{d\log\Gamma(u)}{du}$ is the digamma function and γ is the Euler gamma constant—with numerical value about 0.577. Figure 4 shows a plot of $dE_M = \sqrt{E_M(\tau_x, \rho)}$ from equation (54). This is an approximation to the Riemannian distance but it represents the main features of the information distance of arbitrary latency period distributions T_1 from the curve of distributions T_0 with $\alpha_1 = 1$, $\tau_x = 1$.

Repeating the above procedure for the case when T_0 has $(\alpha_1 + \alpha_2) = 1$, which corresponds to an exponential
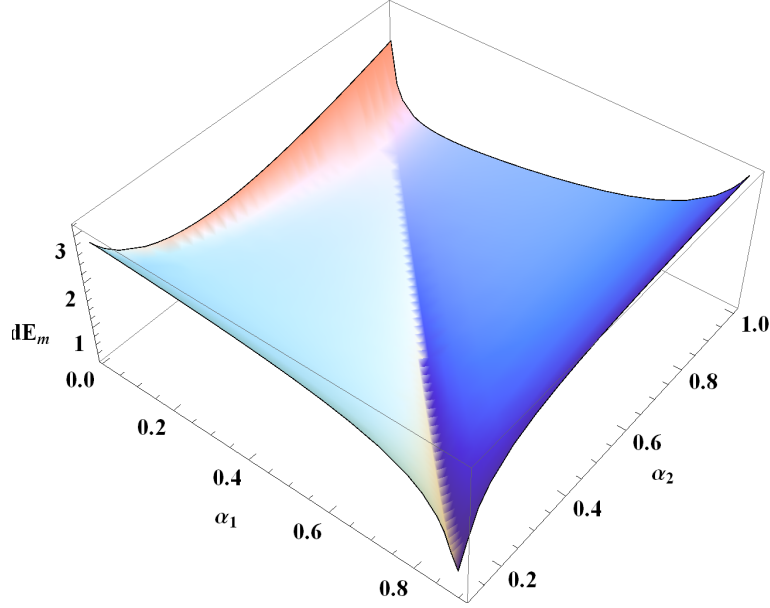
Figure 5: *Approximate information distances $dE_M = \sqrt{E_M}$ (equation (55)) in the McKay manifold, measured from distributions T_0 with exponential marginal distribution for y, so $\alpha_1 + \alpha_2 = 1$ and $\tau_y = 1$. So the surface represents distances from a Poisson process for infectivity.*

infectiousness period distribution (and a Poisson process of infections) we obtain

$$
\begin{aligned}
E_M(\alpha_1, \alpha_2)_{|[T_0:\alpha_1+\alpha_2=1]} &= |\psi(\alpha_2) - \psi(1-\alpha_1)| \\
&+ \frac{1}{4}\left|\frac{(2\alpha_1 + \alpha_2)^2}{4\alpha_1} - \frac{1}{2}(\alpha_1 + 1)\right|.
\end{aligned}
\tag{55}
$$

This is plotted in Figure 5. The two graphics, Figures 4 and 5, show how we can depict the parameters in the joint distribution of periods of latency and infectiousness as surfaces of distance, measured from the two reference cases for the evolution of the epidemic starting from Poisson processes, respectively. On such surfaces could be represented data on the progress of epidemics under different intervention schemes, or simulations of such scenarios.

Geodesic curves in Riemannian manifolds give minimal arc length and examples are given in [Dodson, 2010a] for manifolds of Weibull, gamma and McKay bivariate gamma distributions, together with gradient flow curves for entropy. More details of the information geometry of uniform, exponential, gamma, Gaussian, and bivariate versions with applications are provided in [Arwini & Dodson, 2008].

[Britton & Lindenstrand, 2009] highlighted aspects when stochastic features are important and used independent gamma random variables to represent inhomogeneity of latency and infectiousness periods. In this papersection we have a bivariate inhomogeneous epidemic process, modeled by correlated gamma distributions and we can use similar methods to depict and quantify departures from exponential periods of latency and infectiousness. This shows that the result of [Britton & Lindenstrand, 2009] for the dependence of growth rate on variability in the periods of latency and infectiousness in the independent case persists also in the presence of correlation between the two random variables. Moreover, the information theoretic distance from the two reference scenarios of exponential distributions of periods latency and infectiousness, Figures 4 and 5, provide natural quantitative representations for comparing different parametric data.

[Britton & Lindenstrand, 2009] used independent gamma distributions for periods of latency and infectiousness, from which the reproduction rate can be estimated, with applications for example to the SARS out-
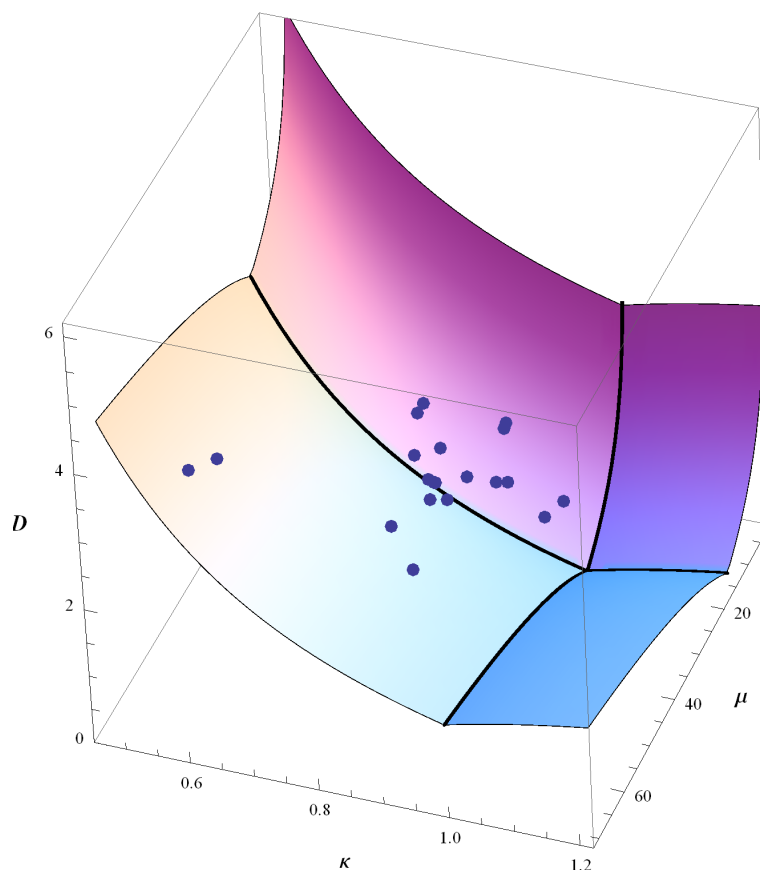
Figure 6: *Distances D in the space of gamma models, using a geodesic mesh [Arwini & Dodson, 2008]. The surface height represents upper bounds on distance D from the grand mean point $(\mu, \kappa) = (18, 1)$, the Poisson random case with mean $\mu = 18$. Depicted also are the 20 data points for the amino acid sequences from a large database of 6294 proteins with sequence lengths up to 4092. All amino acids show clustering to differing degrees by lying to the left of the Poisson random curve $\kappa = 1$, some very substantially so.*

break [WHO, 2003]. Here we have used a bivariate gamma distribution which allows a corresponding reproduction rate to be computed. Also, we considered the case when the susceptibility to infection is not uniform and illustrated with the case when it begins as a gamma distribution then evolves as the epidemic proceeds. Other models could be used for the initial distribution of susceptibilities, including asymmetric distributions. A wide range of such other cases using log-gamma distributions was considered in [Dodson, 2010b], for a similar rate process applied to an evolutionary model when the random variable a represents unfitness (like susceptibility to infection) in a population.

4 Case Study 2: Amino acids self-cluster along protein chains

[Cai et al, 2002] analysed for each of the 20 amino acids X the statistics of spacings between consecutive occurrences of X within the *Saccharomyces cerevisiae* genome, from protein chains with differing sequence lengths n. The process as modelled is like having beads of colours labelled $i = 1, 2, \ldots, 20$, each colour i having a relative abundance p_i. We string the beads onto a long line of total length n beads, then count for each colour i the spacing lengths between successive occurrences in the sequence of beads, yielding 20 empirical distributions which of course depend on the rules used to make the bead choices along the line.

Amino Acid i	Occurrences	Abundance p_i	Mean separation μ_i	Variance σ_i^2	κ_i
A Alanine	163376	0.055	17	374	0.81
C Cysteine	38955	0.013	55	5103	0.59
D Aspartate	172519	0.058	16	346	0.77
E Glutamate	192841	0.065	15	292	0.73
F Phenylalanine	133737	0.045	21	554	0.78
G Glycine	147416	0.049	19	487	0.74
H Histidine	64993	0.022	39	1948	0.79
I Isoleucine	195690	0.066	15	222	0.95
K Lysine	217315	0.073	14	240	0.74
L Leucine	284652	0.095	10	122	0.85
M Methionine	62144	0.021	46	2461	0.85
N Asparagine	182314	0.061	16	277	0.87
P Proline	130844	0.044	21	587	0.77
Q Glutamine	116976	0.039	24	691	0.81
R Arginine	132789	0.045	21	565	0.78
S Serine	269987	0.091	11	136	0.85
T Threonine	176558	0.059	16	307	0.85
V Valine	166092	0.056	17	315	0.93
W Tryptophan	31058	0.010	62	6594	0.58
Y Tyrosine	100748	0.034	27	897	0.79

Table 1: *Experimental data from [Cai et al, 2002] for amino acid occurrences in sequences from 6294 protein chains of the Saccharomyces cerevisiae genome of length up to $n = 4092$. This gives relative abundance, mean spacing, variance of spacing, and maximum likelihood gamma parameter κ_i for each amino acid $i = 1, 2, \ldots, 20$. The grand mean relative abundance was $\overline{p_i} \approx 0.05$ and the grand mean interval separation was $\overline{\mu_i} \approx 18$. Surprising is that the range $0.59 \leq \kappa_i \leq 0.95$ lies below unity, which is expected by chance.*

Analytically, we derived the solution if the beads were chosen completely at random after mixing in a large bag; this gave us a reference Poisson case and the effects of the number n of amino acids in the sequence and the relative abundance of each type [Arwini & Dodson, 2008]. In these reference cases the standard deviation was approximately equal to the mean spacing, increasing with sequence length and decreasing with abundance. So again we have a family of distributions contained approximately in a small neighbourhood of the exponential distributions in the space of gamma distributions. Intuitively, in real proteins with 20 types of amino acids distributed at different abundances along a chain, it might be expected that some would be more clustered ($\kappa < 1$) and others would be more evenly spread ($\kappa > 1$) compared to a Poisson process—which latter case would have exponential spacings and $\kappa = 1$. From a large database with mean spacing $\overline{\mu_i} \approx 18$, the observed spacing distributions of the amino acids, though discrete of course, were all well approximated by gamma distributions (1), cf. [Cai et al, 2002] for details of goodness of fit. We might have expected that the 20 amino acids would have spacing distributions scattered more or less isotropically around the Poisson case, like that for a pseudorandom number generator in Figure 7. Table 1 summarizes some 3 million experimentally observed occurrences of the 20 different amino acids within the *Saccharomyces cerevisiae* genome from the analysis of 6294 protein chains with sequence lengths up to $n = 4092$. Listed also for each amino acid are the relative abundances p_i and mean separation μ_i; the grand mean relative abundance was $p \approx 0.05$ and the grand mean interval separation was $\bar{\mu}_i \approx 18$. The gamma parameter range $0.59 \leq \kappa_i \leq 0.95$, revealed that each amino acid was more clustered and so had higher variance of spacings than expected by chance ($\kappa = 1$).

Specifically, we found that maximum-likelihood estimates of parametric statistics show that all 20 amino acids tend to cluster ($\kappa_i < 1$), some substantially, and so had greater variance than would result from a Poisson process. In other words, the frequencies of shorter gap lengths tends to be higher and the variance of the gap lengths is greater than expected by chance. This may be because localizing amino acids with

the same properties may favour secondary structure formation or transmembrane domains [Cai et al, 2002]. Figure 6 represents the information distance in the space of gamma distributions for amino acids along the protein chains, measured from the Poisson case ($\kappa = 1$) at the grand mean value $\mu = 18$, with data points superimposed. This finding revealed and quantified an important qualitative property: a universal self-clustering that is stable over long sequences for all of these amino acids. The biological significance is yet to be elaborated but it is intriguing to consider the phenomenon as providing a long-range rule that acts as a check during DNA synthesis. Clearly, the maximum likelihood gamma distributions fit only stochastic features and in that respect view the data as exhibiting transient behaviour at small gap sizes; other methods are available for interpretation of such deterministic features and we concentrate here on representation of whole sequences as a stochastic process. Our approach contributes to the characterization of whole sequences by extracting and quantifying stable stochastic features. More detail concerning the data and further discussion is provided in [Cai et al, 2002] and [Arwini & Dodson, 2008].

5   Case Study 3: Constrained degeneration of crystalline order

The degeneration of crystal order was simulated in [Lucarini, 2009] through perturbations of the simple 3D cubic crystal lattices simple, body-centred, face-centred (SC, BCC, FCC) by an increasing spatial Gaussian noise applied to vertices. Physically, the perturbing spatial noise intensity corresponds to a lattice temperature in the structural symmetry degeneration. The statistical parameters of the evolving changes were analyzed through those of the (convex) cells in the Voronoi tessellations, which are optimal partitions of the space from the given set of generating vertices of the structure. In all cases the gamma distribution (1) was an excellent model for the observed probability density functions of all metric and topological properties. With the onset of noise, quite quickly the three tessellations became indistinguishable. With intense noise they all converged to the 3D Poisson-Voronoi tessellations, for which exact analytic results are known [Finch, (2003)]. The observations showed the evolution of the mean and standard deviation of these properties as they converge asymptotically towards the Poisson-Voronoi case, illustrating the rapid degeneration of crystallinity from $\kappa \sim \infty$ for crystalline structures to much lower values as the noise increased, meanwhile the mean area of Voronoi cells reduced only by some ten percent.

Of course, the constraint of remaining tessellations, albeit highly disordered ones, precludes convergence down to the maximum entropy limit $\kappa = 1$. In fact the Poisson Voronoi limiting values are $\kappa \approx 16$ for number of vertices and the same for number of edges and $\kappa \approx 22$ for the number of faces; actually these are discrete random variables and the gamma model is not strictly appropriate. However, for the positive real random variables, polyhedron volume in the Poisson Voronoi limit has $\kappa \approx 5.6$ and polygon face area $\kappa \approx 16$. The constrained degeneration process descends much more rapidly from high κ values than is the case for entropy gradient flow curves shifted to the same asymptote.

[Lucarini, 2008] had reported similar findings for the 2D case of perturbations of the three regular tessellations of the plane: square, hexagonal and triangular. There also the gamma distribution gave a good fit for the distributions during the degeneration of the regular tessellations to the 2D Poisson-Voronoi case; the limiting values were $\kappa \approx 16$ for the perimeter of polygons and $\kappa \approx 3.7$ for areas. These lower limiting values reflect the fewer constraints in the 2D tessellations than in the 3D case. For comparison with other natural apparently constrained disordered structures, $\kappa \approx 4.25$ is the gamma fitted value to the spacings of eigenvalues in the spectrum of Gaussian unitary infinite random matrices (GUE), which model quantum chaotic spacings between energy levels of complex nuclei and is a very good fit also to the spacings between the zeros of the Riemann zeta function [Dodson, 2008].
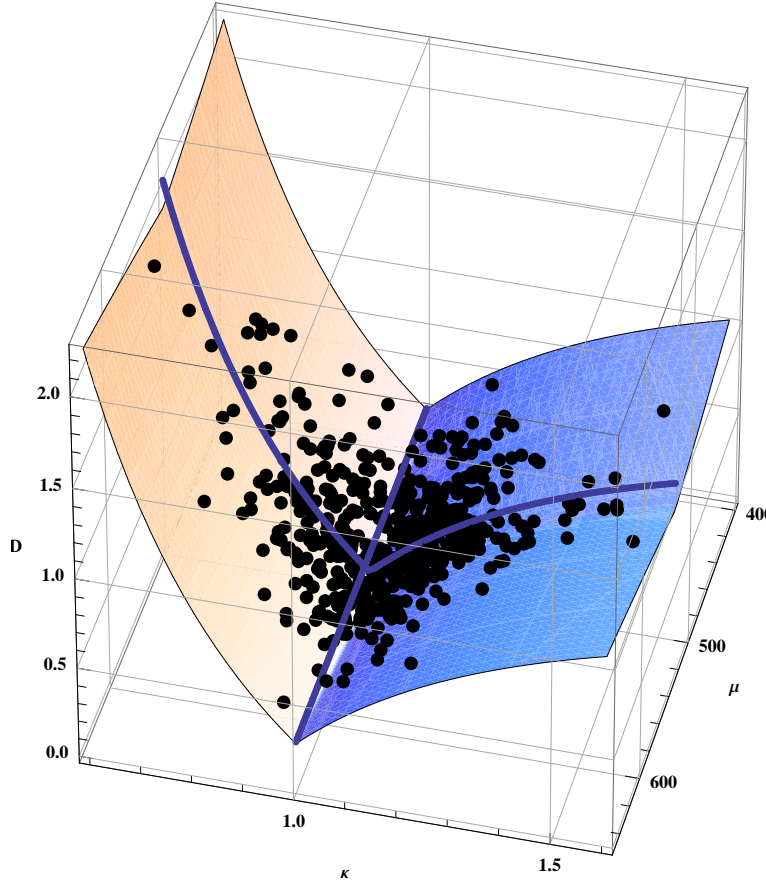
Figure 7: *Distances in the space of gamma models, using a geodesic mesh. The surface height represents upper bounds on distances from $(\mu, \kappa) = (511, 1)$ from equation (56). Also shown are data points from simulations of Poisson random sequences of length 100000 for an element with expected separation $\mu = 511$. In the limit as the sequence length tends to infinity and the element abundance tends to zero we expect the gamma parameter κ to tend to 1.*

6   Case Study 4: Testing pseudorandom number generators

In a variety of contexts in cryptology for encoding, decoding or for obscuring procedures, sequences of pseudorandom numbers are generated. Tests for randomness of such sequences have been studied extensively and the NIST Suite of tests [Rushkin, Soto et al, 2001] for cryptological purposes is widely employed. Information theoretic methods also are used, for example see [Grzegorzewski & Wieczorkowski, 1999] also [Ryabko & Monarev, 2005] and references therein for recent work. Here we can show how pseudorandom sequences may be tested using information geometry by using distances in the gamma manifold to compare maximum likelihood parameters for separation statistics of sequence elements.

Mathematica [Wolfram, 1996] simulations were made of Poisson random sequences with length $n = 100000$ and spacing statistics were computed for an element with abundance probability $p = 0.00195$ in the sequence. In the data from 500 simulations the ranges of maximum likelihood gamma distribution parameters were $419 \leq \mu \leq 643$ and $0.62 \leq \kappa \leq 1.56$. The surface height in Figure 7 represents upper bounds on information geometric distances from $(\mu, \kappa) = (511, 1)$ in the gamma manifold:

$$Distance[(511, 1), (\mu, \kappa)] \leq \left| \frac{d^2 \log \Gamma}{d\kappa^2}(\kappa) - \frac{d^2 \log \Gamma}{d\kappa^2}(1) \right| + \left| \log \frac{511}{\mu} \right|. \tag{56}$$

This employs the geodesic mesh function we described in [Arwini & Dodson, 2008]. Also shown in Figure 7 are data points from the *Mathematica* simulations of Poisson random sequences of length 100000 for an element with expected separation $\mu = 511$. As expected, the simulated data scatters more or less isotropically around the grand mean.

In the limit, as the sequence length tends to infinity and the abundance of the element tends to zero, we expect the gamma parameter τ to tend to 1. However, finite sequences must be used in real applications and then provision of a metric structure allows us, for example, to compare real sequence generating procedures against an ideal Poisson random model.

7 Case Study 5: Comparing nearby signal distributions and drifts from uniformity

The typical application context here is in the protection of public-key encryption methods involving variations of the RSA procedure, which employs modular arithmetic with a very large modulus. It is necessary to compute

$$R \equiv y^e \ (mod \, m) \ \ \text{or} \ \ R \equiv y^d \ (mod \, m) \tag{57}$$

for, respectively, encoding or decoding a message y. The very large modulus m and the encryption key e are made public; the decryption key d is secret. The modulus m is chosen to be the product of two large prime numbers p, q, also secret, then choose d, e such that

$$ed \equiv 1 \ (mod \, (p-1)(q-1)). \tag{58}$$

Encoding and decoding computations both involve repeated numerical exponentiation procedures. Then, some knowledge of the design of an implementation and information on the timing or power consumption during computational stages could yield clues to the decryption key d. [Canvel, 1999], [Canvel & Dodson, 2000] showed how timing analyses of the modular exponentiation algorithm quickly reveal the private key, regardless of its length. An obscuring procedure could mask the timing information but that may not be straightforward for some small memory devices. It is important to be able to assess departures from Poisson randomness of underlying or overlaid procedures that are inherent in devices and here we outline some information geometric methods to add to the standard tests [Rushkin, Soto et al, 2001].

In cryptographic attacks, differential Power Analysis (DPA) methods and Statistical Zero-Knowledge (SZK) proofs depend on discrimination between noisy samples drawn from pairs of closely similar distributions. In many cases the distributions resemble truncated Gaussians; sometimes one distribution is uniform. A log-gamma family of probability density functions provides a 2-dimensional metric space of distributions on $(0, 1]$, ranging from the uniform distribution to symmetric unimodular distributions of arbitrarily small variance. Illustrative calculations are provided here; more discussion is given in [Arwini & Dodson, 2008].

In practical signal comparison situations, we obtain statistical data for an observable that is defined on some finite interval. We shall use as our model the family (5) of log-gamma probability density functions defined for random variable $a \in (0, 1]$. The choice of log-gamma model is due to the fact that it contains a neighbourhood of the uniform distribution, for parameter values $(\nu, \tau) = (1, 1)$ in (5), and for parameter values $\tau >> 1$ it has approximations to Gaussians truncated to domain $(0, 1]$ and with arbitrarily small variance. Figure 8 illustrates symmetric such cases with mean value $E(a) = \frac{1}{2}$. From the information metric (26) on the space of these probability density functions we can obtain information distances between nearby distributions as follows.

Suppose that we record data on amplitude $a \in (0, 1]$ for two cases with parameters (ν_0, τ_0) and $(\nu_0 + \Delta\nu, \tau_0 +$
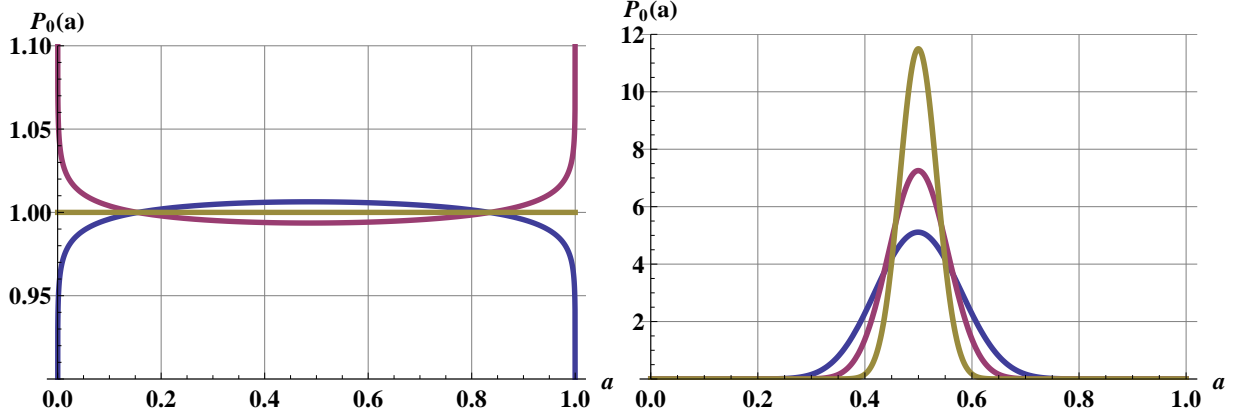
Figure 8: *Examples from the log-gamma family of probability densities with central mean $E(a) = \frac{1}{2}$. Left: near to uniform, $\tau = 0.99, 1, 1.1$. Right: approximations to truncated Gaussians, $\tau = 20, 40, 100$.*

$\Delta\tau$) for small $\Delta\nu, \Delta\tau$. Then the information distance Δs between these distributions is approximated from

$$\Delta s^2 \approx \frac{\tau_0}{\nu_0^2}\,\Delta\nu^2 - \frac{2}{\nu_0}\,\Delta\nu\Delta\tau + \frac{d^2 \log\Gamma}{d\tau^2}(\tau_0)\,\Delta\tau^2. \tag{59}$$

Two particular cases are of interest:

Near to the uniform distribution Here we have $(\nu_0 = 1, \tau_0 = 1)$ and (59) reduces to

$$\Delta s^2 \approx \Delta\nu^2 - 2\Delta\nu\Delta\tau + 1.645\Delta\tau^2. \tag{60}$$

Two nearby unimodular distributions Here we have $\tau_0 >> 1$ and (59) reduces to

$$\Delta s^2 \approx \frac{\tau_0}{\nu_0^2}\,\Delta\nu^2 - \frac{2}{\nu_0}\,\Delta\nu\Delta\tau + \Delta\tau^2. \tag{61}$$

For example,some data on power measurements from a smartcard leaking information during processing of a '0' and a '1', at a specific point in process time, yielded two data sets C, D. These had maximum likelihood parameters $(\nu_C = 2.506, \tau_C = 1.816)$ and $(\nu_D = 4.527, \tau_D = 1.757)$.

8   Conclusions

Information geometry allows us to formulate precise meaning for nearness and neighbourhood in a state space of processes and it quantifies the proximity of a process to a particular state by means of a natural information theoretic metric structure on smoothly parametrized families of probability density functions. Information geometric methods complement standard statistical tools with techniques of representation similar to those used in physical field theories where the analysis of curved geometrical spaces have contributed to the understanding of phenomena and development of predictive models. The role of information theory itself is important, attributing significance to the geometry and giving access to the concepts of entropy and likelihood. In many statistical models of practical importance there is a small range of probability density functions that has very wide application as a result of general theorems, and the spaces of these families

have just a small number of dimensions. Families of Gaussian and gamma distributions and their bivariate versions are widely applied and moreover their information geometry is easily tractable. Gamma distributions are ubiquitous in modeling natural processes that involve scatter of a positive random variable around a target state, such as for inhomogeneous populations or features of elements in a collection. A defining characteristic of the gamma distribution is for the sample standard deviation to be proportional to the sample mean, commonly found to varying degrees of approximation; the case when the standard deviation equals the mean corresponds to the exponential distribution associated with a Poisson process which is the fundamental reference process. We have described some general theorems about information neighbourhoods of randomness, independence, and uniformity for processes and their qualitative features make them widely applicable, as our case studies illustrate.

The case of an inhomogeneous statistical rate processes for epidemics showed how a bivariate process may be involved, to reveal and quantify qualitative features. For example, we demonstrated a stable property that persisted in the presence of correlation between infectiousness and latency periods. The epidemic exponential growth rate increased with variability of infectiousness period but decreased with variability of latency period. The information metric provided a distance structure to quantify proximity to the natural reference states—those for exponential distributions for one of the two variables. A gamma model for amino acid spacings along protein chains revealed the qualitative stable property that all 20 of the amino acids are more clustered together than would be expected at random, over a range of abundances and lengths of protein chains. Simulations of the constrained disordering of crystals in two and three dimensions again revealed the dominance of the gamma family of distributions for features such as polygon areas, polygon perimeters and polyhedral volumes; this means that the progress of degeneration of order may be tracked through the information geometry of the space of gamma distributions. Distinguishing nearby signal distributions is important in a variety of contexts, particularly now in cryptological studies. The family of log-gamma distributions provides a neighbourhood of the uniform distribution as well as good approximations to truncated Gaussians of arbitrarily small variance; such distributions are commonly needed for testing purposes.

Clearly, there is plenty of scope for further applications of information geometry in the modeling of real phenomena, and in particular for the elaboration of inhomogeneous biological rate processes. One example of the latter is in the neural modeling of the process of neuron and synapse evolution during early development. Typically, embryo brains seem to create a large number of neurons very rapidly [Hamburger, 1975] but some die out to reduce the adult number to about 60% of the maximum. In humans, there is a rise in number of synapses that seem randomly created during the first two years; this is followed by a decline to 60% of the maximum by the time of adolescence [Huttenlocher, 1984]. This gives rise to a large neural net model for populations of nodes (neurons) and edges (synapses) declining through a rate process controlled by early usage statistics from neuronal stimulation and activity [Izhikevich, 2003]; such a study is in progress. In another direction, modern small chip-controlled interactive devices, that need encryption to protect data but have limited computational capability, could benefit from 'quick and rough' software applications that can use information geometry to detect patterns in fraudulent attempts at access. Similarly, in the development of devices it could be useful to monitor externally detectable non-random process activity statistics, informing the need for shielding or obscuring procedures [Canvel & Dodson, 2000]. Constrained disordering on a small scale may be viewed as a minor degradation in a structure, a device or a service; this may be as a result of ageing or externally caused damage and may be amenable to automatic monitoring through information geometry associated with the defining structural features and through which the degradation is detectable and measurable. More developments in the theory and applications of information geometry can be gleaned from the papers presented at the latest international conference, [IGA Conference, 2010].

References

[Amari, 1963] Amari, S-I. (1963). *Diakoptics of Information Spaces.* Doctoral Thesis, University of Tokyo.

[Amari, 1968] Amari, S-I. (1968). *Theory of Information Spaces—A Geometrical Foundation of the Analysis of Communication Systems.* Research Association of Applied Geometry Memoirs 4, 171-216.

[Amari, 1985] Amari, S-I. (1985). *Differential Geometrical Methods in Statistics.* Berlin Springer Lecture Notes in Statistics 28, Springer-Verlag.

[Amari et al, 1987] Amari, S-I., Barndorff-Nielsen, O.E., Kass, R.E., Lauritzen, S.L. & Rao, C.R. (1987). *Differential Geometry in Statistical Inference.* Lecture Notes Monograph Series, Institute of Mathematical Statistics, Volume 10, Hayward California, Springer-Verlag.

[Amari & Nagaoka, 2000] Amari, S-I. & Nagaoka, H. (2000). *Methods of Information Geometry.* Oxford, American Mathematical Society, Oxford University Press.

[Andersson & May, 1991] Andersson, H. & May, R.M. (1991). *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, Oxford.

[Andersson & Britton, 2000] Andersson, H. & Britton, T. (2000). *Stochastic epidemic models and their statistical analysis.* Lecture Notes in Statistics Volume 151, Springer-Verlag, New York.

[Arwini, 2004] Arwini, K. (2004). *Differential geometry in neighbourhoods of randomness and independence.* PhD thesis, Department of Mathematics, University of Manchester Institute of Science and Technology.

[Arwini & Dodson, 2008] Arwini, K. & Dodson, C.T.J. (2008). *Information Geometry Near Randomness and Near Independence.* Lecture Notes in Mathematics. New York, Berlin, Springer-Verlag.

[Britton & Lindenstrand, 2009] Britton, T. & Lindenstrand, D. (2009). *Epidemic modelling: aspects where stochasticity matters.* Mathematical Biosciences 222, 2 109-116. Cf. also http://arxiv.org/abs/0812.3505 3 January 2009.

[Baake & Georgii, 2006] Baake E. & Georgii, H.-O. (2006). *Mutation, selection, and ancestry in branching models: a variational approach.* http://arxiv.org/abs/q-bio/0611018 3 December 2006.

[Bailey, (1975)] Bailey, N.T.J. (1975). *The Mathematical Theory of Infectious Diseases and its Applications.* London, Griffin.

[Cai et al, 2002] Cai, Y. Dodson, C.T.J. Wolkenhauer O. & Doig, A.J. (2002). *Gamma Distribution Analysis of Protein Sequences shows that Amino Acids Self Cluster.* Journal Theoretical Biology 218, 4 409-418.

[Canvel, 1999] Canvel, B. (1999). *Timing Tags for Exponentiations for RSA.* MSc Thesis, Department of Mathematics, University of Manchester Institute of Science and Technology, Manchester.

[Canvel & Dodson, 2000] Canvel, . & C.T.J. Dodson, C.T.J. (2000). *Public Key Cryptosystem Timing Analysis.* CRYPTO 2000, Rump Session Santa Barbara, 20-24 August 2000.
http://www.maths.manchester.ac.uk/ kd/PREPRINTS/rsatim.ps 27 August 2000.

[Cauchemez & Ferguson, 2008] Cauchemez, S. & Ferguson, N.M. (2008). *Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London.* Journal Royal Society Interface 5 885-897.

[Chowell et al, 2009] Chowell, G., Hyman, J. M., Bettencourt, L. M. A. & Castillo-Chavez, C. (2009). Editors. *Mathematical and Statistical Estimation Approaches in Epidemiology.* Springer Dordrecht, Heidelberg, London, New York.

[Cziko, 1995] Cziko, G. (1995). *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution.* MIT Press, Cambridge, Massachusetts.

[Diekmann & Heesterbeek, 2000] Diekmann, O. & Heesterbeek, J.A.P. (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation.* John Wiley, Chichester.

[Dodson, 2008] Dodson, C.T.J. (2009). *A note on quantum chaology and gamma manifold approximations to eigenvalue spacings for infinite random matrices.* In C. H. Skiadas and I. Dimotikalis and C. Skiadas (Eds.) *Topics in Chaotic Systems: Selected Papers From Chaos 2008*, World Scientific, Singapore 2009, pp 96-103.

[Dodson, 2009] Dodson, C.T.J. (2009). *Information geometry and entropy in a stochastic epidemic rate process.* http://arxiv.org/abs/0903.2997 17 December 2010.

[Dodson, 2010a] C.T.J. Dodson, (2010). *On the entropy flows to disorder.* In C. H. Skiadas and I. Dimotikalis, (Eds.), *Chaotic Systems: Theory and Applications* World Scientific, Singapore, 2010 pp 75-84.

[Dodson, 2010b] Dodson, C.T.J. (2010). *An inhomogeneous stochastic rate process for evolution from states in an information geometric neighbourhood of uniform fitness.* Invited paper at 3rd Conference on Information Geometry and its Application, Leipzig 2-6 August 2010. Cf also: http://arxiv.org/abs/1001.4177v1 . 18 January 2010.

[Dodson, 2010c] Dodson, C.T.J. (2010). *Mathematica Notebooks.*
http://www.maths.manchester.ac.uk/kd/mmaprogs/InfoGeomMMANotebooks/ . 19 October 2010.

[Dodson & Matsuzoe, 2003] Dodson, C.T.J. & Matsuzoe, H. (2003). *An affine embedding of the gamma manifold.* Applied Sciences (5) 1 1-6.
http://www.maths.manchester.ac.uk/ kd/PREPRINTS/affimm.pdf 3 November 2003.

[Dodson, & Poston, 1991] Dodson, C.T.J. & Poston, T. (1991) *Tensor Geometry.* Second Edition, Springer-Verlag, Berlin, New York, Graduate Texts in Mathematics 120.

[Dodson & Thompson, 2000] Dodson, C.T.J. & Thompson, S.M. (2000). *A metric space of test distributions for DPA and SZK proofs.* Poster Session, Eurocrypt 2000, Bruges, 14-19 May 2000.
http://www.maths.manchester.ac.uk/ kd/PREPRINTS/mstd.pdf 14 May 2000.

[Feller, 1971] Feller, W. (1971). *An Introduction to Probability Theory and its Applications.* Volume II 2nd Edition, Wiley, New York.

[Finch, (2003)] Finch, S.R. (2003). *Mathematical Constants.* Cambridge University Press, Cambridge.

[Fisher, 1958] Fisher, R.A. (1958). *The Genetical Theory of Natural Selection.* 2nd Edition, Dover, New York.

[Frank, 2009a] Frank, S.A. (2009). *Natural selection maximizes Fisher information.* Journal of Evolutionary Biology 22(2) 231-244.

[Frank, 2009b] Frank, S.A. (2009). *The common patterns of nature.* Journal of Evolutionary Biology 22(8) 1563-1585.

[Freund, 1961] Freund, R.J. (1961). *A bivariate extension of the exponential distribution.* Journal of the American Statistical, 56, 971-977.

[Gray, 1998] Gray, A. (1998). *Modern Differential Geometry of Curves and Surfaces.* 2nd Edition, Boca Raton, CRC Press.

[Grzegorzewski & Wieczorkowski, 1999] Grzegorzewski, P. & Wieczorkowski, R. (1999). *Entropy-based goodness-of-fit test for exponentiality.* Communications Statististical Theory and Methodology 28(5) 1183-1202.

[Hamburger, 1975] Hamburger, V. *Cell-death in the development of the lateral motor column of the chick embryo.* J. Comparative Neurology 160, 535-546.

[He et al, 2009] He, D. Ionides, E. L. & King, A. A. (2009). *Plug-and-play inference for disease dynamics: measles in large and small populations as a case study.* Journal Royal Society Interface doi: doi:10.1098/rsif.2009.0151

[Huttenlocher, 1984] Huttenlocher, P.R. *Synapse elimination and plasticity in developing human cerebral cortex.* American J. Mental Deficiency 88, 5 488-496.

[Hwang & Hu, 1999] Hwang, T-Y & Hu, C-Y. (1999). *On a characterization of the gamma distribution: The independence of the sample mean and the sample coefficient of variation.* Annals Institute Statistical Mathematics 51(4) 749-753.

[IGA Conference, 2010] *Information Geometry and its Applications III.* Max-Planck-Institut fr Mathematik in den Naturwissenschaften, Leipzig, 2-6 August 2010.
http://www.mis.mpg.de/calendar/conferences/2010/infgeo.html 2 August 2010.

[Izhikevich, 2003] Izhikevich, E.M. *Simple model of spiking neurons.* IEEE Transactions on Neural Networks 14, 6 1569-1572.

[Karev, 2003] Karev, G.P. (2003). *Inhomogeneous models of tree stand self-thinning.* Ecological Modelling 160 23-37.

[Karev, 2010a] Karev, G.P. (2010). *Replicator equations and the principle of minimal production of information.* Bulletin Mathematical Biology 72 (5) 1124-1142. DOI: 10.1007/s11538-009-9484-9 5 21 March 2009.

[Karev, 2010b] Karev, G.P. (2010). *On mathematical theory of selection: continuous time population dynamics.* Journal Mathematical Biology 60 107-129.

[Laha, 1954] Laha, R.G. (1954). *On a characterization of the gamma distribution.* Annals of Mathematical Statistics 25 784-787.

[Lloyd, 2001] Lloyd, A.L. (2001). *Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods.* Proceedings Royal Society London B 268 985-993.

[Lucarini, 2008] Lucarini, V. (2008). *From Symmetry Breaking to Poisson Point Processes in 2D Voronoi Tessellations: the Generic Nature of Hexagons.* Journal Statistical Physics, 130 1047-1062.

[Lucarini, 2009] Lucarini, Valerio (2009). *Three-dimensional Random Voronoi Tessellations: From Cubic Crystal Lattices to Poisson Point Processes.* Journal Statistical Physics, 134(1) 1047-1062.

[McKay, 1934] McKay A.T. (1934). *Sampling from batches.* J. Royal Statist. Soc. 2 207-216.

[Miller et al, 2009] Miller, Joel C., Davoudi, B., Meza, R., Slim, A & Pourbohloul, B (2009). *Epidemics with general generation interval distributions.* http://arxiv.org/abs/0905.2174v2.pdf

[Nishiura et al, 2009] Nishiura, H. Chowell, G. Heesterbeek H. and Wallinga, J. (2009). *The ideal reporting interval for an epidemic to objectively interpret the epidemiological time course.* Journal of Royal Society Interface. 10.1098/rsif.2009.0153

[Rao, 1945] Rao, C.R. (1945). *Information and accuracy attainable in the estimation of statistical parameters.* Bull. Calcutta Math. Soc. 37, 81-91.

[Roff, 2010] Roff, D.A. (2010). *Modeling evolution. An introduction to numerical techniques.* Oxford, Oxford University Press.

[Rose & Smith, 2002] Rose, C. & Smith, M.D. (2002). *Mathematical Statistics with Mathematica.* Berlin, Springer texts in statistics, Springer-Verlag.

[Rushkin, Soto et al, 2001] Rushkin, A. Soto, J. et al. (2001). *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications.* National Institute of Standards & Technology, Gaithersburg, MD USA.

[Ryabko & Monarev, 2005] Ryabko, B.Ya. & Monarev, V.A. (2005). *Using information theory approach to randomness testing.* arXiv:CS.IT/0504006 v1, 3 April 2005.

[Wallinga & Lipsitch, 2007] Wallinga J. & Lipsitch, M. (2007). *How generation intervals shape the relationship between growth rates and reproductive numbers.* Proceedings Royal Society London B, 274 599-604.

[WHO, 2003] WHO (2003). *Cumulative Number of Reported Probable Cases of Severe Acute Respiratory Syndrome (SARS).* http://www.who.int/csr/sars/country/en/ 26 September 2003.

[Wolfram, 1996] Wolfram, S. (1996). *The Mathematica Book.* 3rd edition, Cambridge University Press, Cambridge.

9 Key Terms and Definitions

Information metric: Riemannian distance structure and hence arc length function, defined by the covariance matrix function of a smooth family of probability density functions.

Information entropy: The negative of the expectation of the logarithm of a probability density function.

Inhomogeneous rate process: A first order differential process defined on a population where the rate of change of a cohort is proportional to the local density of that cohort.

Constrained disordering: A structure defined by a probability distribution of its features may be perturbed but the structural rules control the degree to which total disorder may be approached.

Pseudorandom number generator: An algorithm that generates numbers from a given set with each interval of the set having a probability of occurrence in proportion to its total length, so approximating a uniform distribution.

Random variable: A variable that follows a well-defined discrete probability distribution, or continuous probability density distribution.

Integral curve of a gradient field: A curve for which the rate of change with time at a point is equal to the gradient vector of a field at that point.

Voronoi cells: Given a set of nodes in the plane $\mathbb{R}^2$, for each point in the plane there is one node closest to it or at most two closest nodes equidistant from it. The Voronoi cell for a node is the interior of the convex polygon of nearest points. This definition extends to higher dimensions—leading to convex polyhedra in $\mathbb{R}^3$.