Information geometry and entropy in a stochastic epidemic rate process

Dodson, CTJ

2009

MIMS EPrint: **2009.39**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

Information geometry and entropy in a stochastic epidemic rate process

C.T.J. Dodson
School of Mathematics, University of Manchester, Manchester M13 9PL, UK
ctdodson@manchester.ac.uk

May 17, 2009

1 Introduction

A commonly recurring approximation to real rate processes is of the form:

$$\dot{N} = -m\,N$$

where m is some positive rate constant and $N(t)$ measures the current value of some property relevant to the process—radioactive decay is our typical student example. The simplest stochastic version addresses the situation where $N(t)$ is the size of the current population and the rate constant depends on the distribution of properties in the population—so different sections decay at different rates. Then the interest lies in the evolution of the distribution of properties and of the related statistical features like entropy, mean and variance, for given initial distribution. We show that there is a simple closed solution for an example of an epidemic in which the latency and infectivity are distributed properties controlled by a bivariate gamma distribution.

We consider a class of simple stochastic rate processes where a population N is classified by a smooth family of probability density functions $\{P_t, t \geq 0\}$ with random variable $a \geq 0$, having mean $E_t(a)$ and variance $\sigma^2(t)$. Karev [4] formulated this situation in the form

$$N(t) = \int_0^\infty l_t(a)\,da \quad \text{and} \quad P_t(a) = \frac{l_t(a)}{N(t)} \tag{1}$$

$$\dot{l}_t(a) = -a l_t(a) \quad \text{so} \quad l_t(a) = l_0(a)e^{-at} \tag{2}$$

from which he obtained a large range of general solutions. We summarise as follows

$$N(t) = N(0)L_0(t) \quad \text{where } L_0(t) = \int_0^\infty P_0(a)e^{-at}\,da \tag{3}$$

$$\dot{N} = -E_t(a)\,N \quad \text{where } E_t(a) = \int_0^\infty a\,l_t(a)\,da = -\frac{d\log L_0}{dt} \tag{4}$$

$$\dot{E}_t(a) = -\sigma^2(t) = (E_t(a))^2 - E_t(a^2) \tag{5}$$

$$P_t(a) = e^{-at}\frac{P_0(a)}{L_0(t)} \quad \text{and } l_t(a) = e^{-at}L_0(t) \tag{6}$$

$$\dot{P}_t(a) = P_t(a)(E_t(a) - a). \tag{7}$$

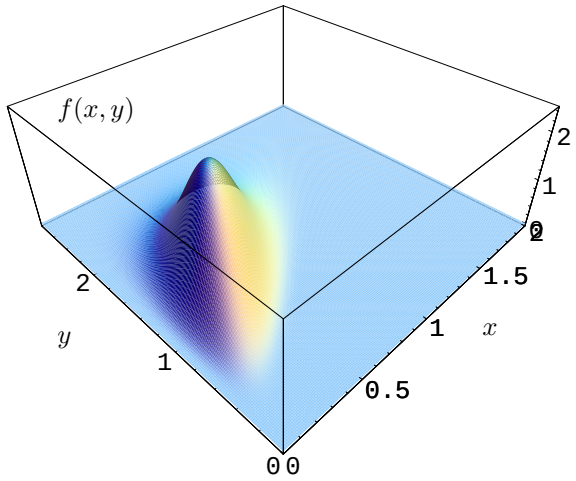

Figure 1: *Part of the family of McKay bivariate gamma probability density functions* $f(x, y)$ *with correlation coefficient* $\rho = 0.6$ *and* $\alpha_1 = 5$.

The elegance of Karev's approach lies in the fact that $L_0(t)$ 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 $\frac{N(t)}{N(0)}$. He gave the particular solutions for the cases of initial distributions that were Poisson, gamma or uniform.

2 Entropy dynamics

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

$$S_t = -E_t \left(\log P_t(a) \right) = -E_t \left(\log \frac{P_0(a) e^{-at}}{L_0(t)} \right) \tag{8}$$

which reduces to

$$S_t = S_0 + \log L_0(t) + E_t(a)\, t. \tag{9}$$

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

$$\frac{dS_t}{dt} = -\sigma^2(t)\, t. \tag{10}$$

This result shows how the variance controls the entropy change during quite general inhomogeneous population processes. In fact equation (10) and further related results were given also in a subsequent paper Karev [5]. We note that the reverse process of population growth may have applications in constrained disordering type situations [3].

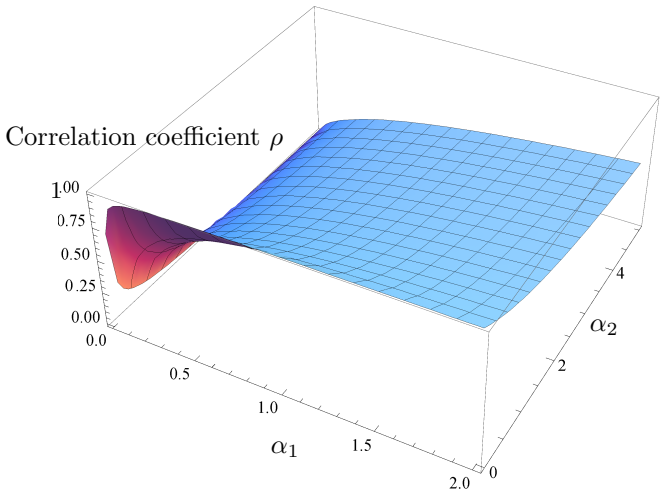

Figure 2: *Correlation coefficient ρ from equation (15) for McKay probablity density functions, in terms of α_1, α_2. This represents correlation between latency and infectivity periods.*

3 Epidemic model

This example adds to the work of Britton and Lindenstrand [2] in which they use independent univariate gamma distributions for the periods of latency and infectivity in an epidemic model that they illustrated with data from the SARS outbreak [6]. Our contribution is to use a bivariate gamma distribution which allows possible correlation between the random variables representing the periods of latency x and infectivity y. We obtain a closed analytic solution and show that the same qualitative features persist in the presence of correlation.

The family of McKay bivariate gamma density functions M, is defined on $0 < x < y < \infty$ with parameters $\alpha_1, \sigma_{12}, \alpha_2 > 0$ and probability density functions, Figure 1,

$$f(x,y;\alpha_1,\sigma_{12},\alpha_2) = \frac{(\frac{\alpha_1}{\sigma_{12}})^{\frac{(\alpha_1+\alpha_2)}{2}} x^{\alpha_1-1}(y-x)^{\alpha_2-1}e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}}y}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \ . \tag{11}$$

Here σ_{12} 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. Thus z is the period during which infection occurs and it controls the rate process as we see in the sequel. We obtain the means, standard deviations

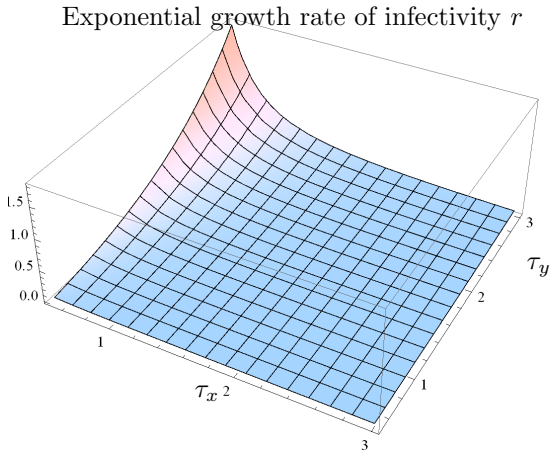

Figure 3: *Plot of the Malthusian parameter r against the coefficients of variation of latency τ_x and infectivity τ_y with means $\mu_x = 3$, $\mu_y = 2$ and $R_0 = 2.2$ from the SARS data. So the exponential growth rate of infectivity decreases with variability in latency (τ_x) but increases with variability in infectivity (τ_y).*

and coefficients of variation by direct integration:

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

$$\text{SD}: \quad \sigma_x = \sqrt{\sigma_{12}}, \quad \sigma_y = \sqrt{\frac{\sigma_{12}(\alpha_1+\alpha_2)}{\alpha_1}}, \quad \sigma_z = \sqrt{\frac{\alpha_2\sigma_{12}}{\alpha_1}} \tag{13}$$

$$\text{CV}: \quad \tau_x = \frac{1}{\sqrt{\alpha_1}}, \quad \tau_y = \frac{1}{\sqrt{\alpha_1+\alpha_2}}, \quad \tau_z = \frac{1}{\sqrt{\alpha_2}} \tag{14}$$

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

$$\rho = \sqrt{\frac{\alpha_1}{\alpha_1+\alpha_2}} > 0 \tag{15}$$

$$f_1(x) = \frac{(\frac{\alpha_1}{\sigma_{12}})^{\frac{\alpha_1}{2}} x^{\alpha_1-1} e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}}x}}{\Gamma(\alpha_1)}, \quad x>0 \tag{16}$$

$$f_2(y) = \frac{(\frac{\alpha_1}{\sigma_{12}})^{\frac{(\alpha_1+\alpha_2)}{2}} y^{(\alpha_1+\alpha_2)-1} e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}}y}}{\Gamma(\alpha_1+\alpha_2)}, \quad y>0 \tag{17}$$

Figure 2 shows a plot of the correlation coefficient from equation (15). The marginal probability density functions of latency x and infectivity 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.

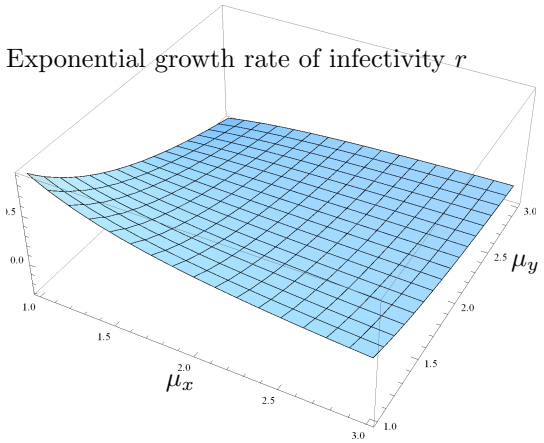

Figure 4: *Plot of the Malthusian parameter r against the means of latency μ_x and infectivity μ_y with coefficients of variation $\tau_x = \tau_y = 4/7$ and $R_0 = 2.2$ from the SARS data. So the exponential growth rate of infectivity decreases with mean latency (μ_x) and with mean infectivity (μ_y).*

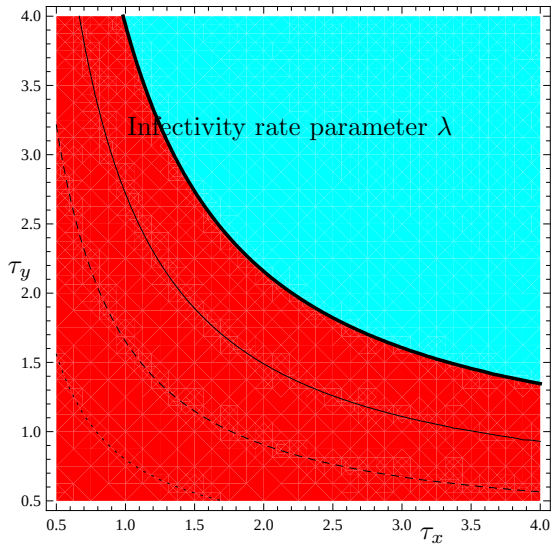

Figure 5: *Contour plot of the infectivity rate λ against the coefficients of variation of latency τ_x and infectivity τ_y with mean latency $\mu_x = 3$ and $r = 0.053$ from the SARS data. The levels are $\lambda = 1.1$ dotted, $\lambda = 1.5$ dashed, $\lambda = 3$ thin and $\lambda = 10$ thick.*

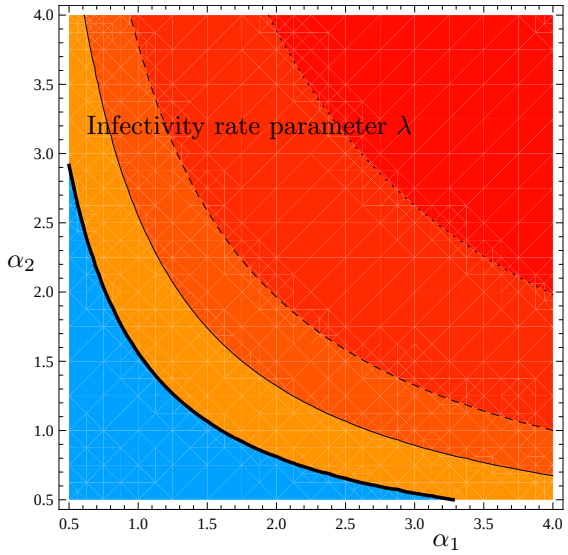

Figure 6: *Contour plot of the infectivity rate* λ *against the McKay parameters* α_1 *and* α_2 *with mean latency* $\mu_x = 3$ *and* $r = 0.053$ *from the SARS data. The levels are* $\lambda = 1.02$ *dotted,* $\lambda = 1.04$ *dashed,* $\lambda = 1.06$ *thin and* $\lambda = 1.1$ *thick.*

We follow the method of Britton and Lindenstrand [2] §3.2, to compute the initial exponential growth rate of the epidemic from their equation (3.4) which we write for bivariate x, y in the form

$$\int_0^\infty \int_y^x e^{-r(y-x)}\, \lambda\, f(x,y)\; dy\, dx = 1 \tag{18}$$

Here, from [2], $R_0 = \lambda\mu_x$ for the average number of infections per infective, so λ is the contact rate; this gives the Malthusian parameter as

$$r = \frac{1}{\mu_x \tau_x^2}\left(\left(\frac{R_0}{\mu_y}\right)^{\tau_y^2} - 1\right) \tag{19}$$

Thus, r is monotonically decreasing with μ_x, μ_y and τ_x but increasing with τ_y; this is illustrated in Figure 3 and Figure 4 using typical values from the SARS epidemic [6]. Figure 5 and Figure 6 show contour plots of the infectivity rate λ using typical values from the SARS epidemic [6].

We turn next to the evolution of the distribution of $z = y - x$, the interval of actual infection as the population $N(t)$ of uninfected individuals declines with time t. From §1 above, this corresponds to the case when $P_0(z)$ is a gamma distribution and by the result of Karev [4]

$$P_t(z) = \frac{P_0(z)}{L_0(t)} e^{-zt} = \frac{(s+t)^k\, z^{k-1}}{\Gamma(k)}\, e^{-z(s+t)}. \tag{20}$$

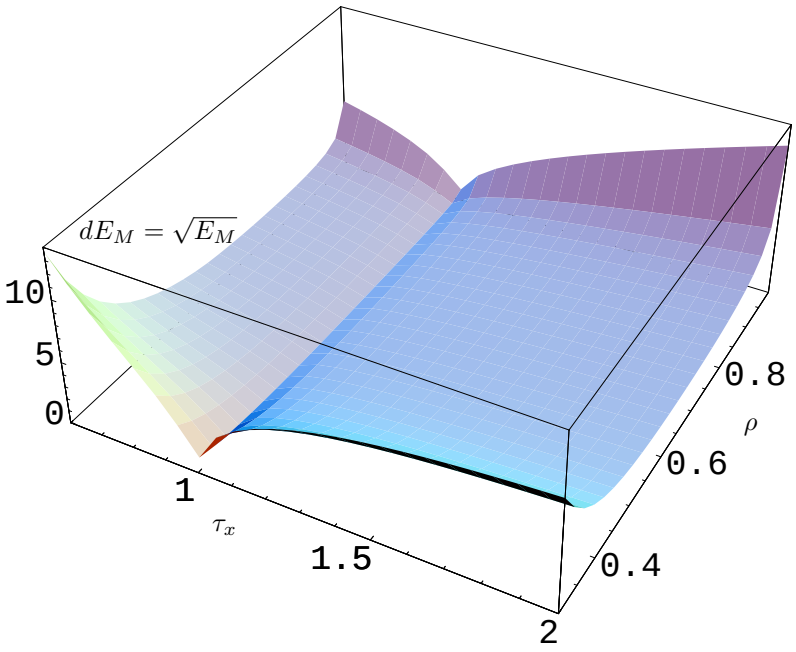

Figure 7: *Approximate information distances* $dE_M = \sqrt{E_M}$ *(equation (24)) in the McKay manifold, measured from distributions* T_0 *with exponential marginal distribution for* x *so* $\alpha_1 = 1$ *and* $\tau_x = 1$*. So the surface represents distances from a Poisson process for latency.*

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

$$\mu_z(t) = \frac{k}{s+t} \tag{21}$$

$$\sigma_z(t) = \frac{\sqrt{k}}{s+t} \tag{22}$$

$$\tau_z(t) = \frac{1}{\sqrt{k}}. \tag{23}$$

4 Information geometry

Information geometry of the smooth family M of McKay bivariate gamma probability density functions has been studied in detail in Arwini and Dodson [1]. This provides a Riemannian metric on M, so yielding a curved 3-manifold, and moreover the metric is subordinate to maximum likelihood/maximum entropy processes. Here we illustrate how the geometry may be used to provide a natural distance structure on the space of McKay distributions. First we measure distances from distributions with exponential marginal distributions—those for which latency is controlled by a Poisson event process with $\alpha_1 = 1$, $\tau_x = 1$. The derivation of a distance from distribution T_0 is given in [1], and yields in terms

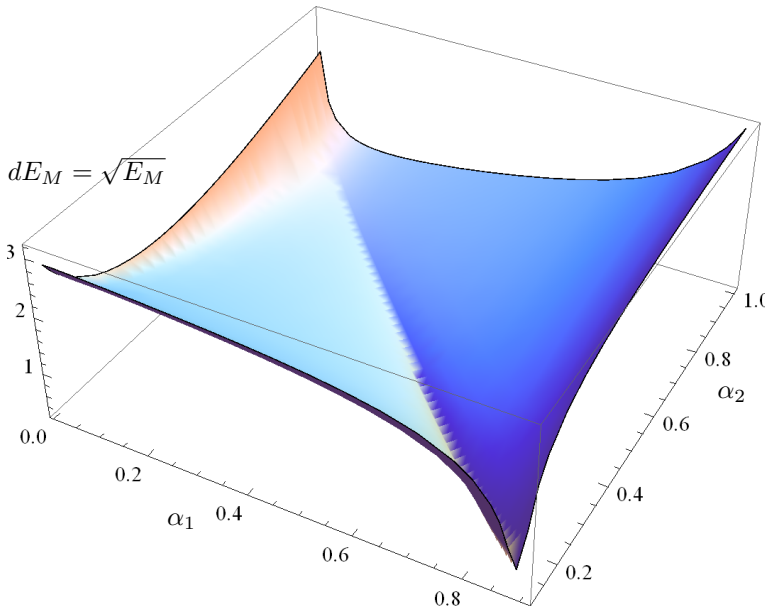

Figure 8: *Approximate information distances* $dE_M = \sqrt{E_M}$ *(equation (25)) in the McKay manifold, measured from distributions* T_0 *with exponential marginal distribution for* y*, so* $\alpha_1 + \alpha_2 = 1$ *and* $\tau_y = 1$*. So the surface represents distances from a Poisson process for infectivity.*

of τ_x and ρ

$$\begin{aligned} E_M(\tau_x, \rho)_{|[T_0:\alpha_1=1]} &= \frac{\left(\rho^2+1\right)^2}{16\rho^6}\left|\frac{1}{\tau_x^2}-1\right| \\ &+ \frac{1}{4}\left|\left(1-\frac{1}{\tau_x^2}\right)\left(1-\frac{1}{\rho^2}\right)+3\log\left(\tau_x^2\right)\right| \\ &+ \left|\psi\left(\frac{1}{\tau_x^2}\frac{1}{\rho^2}-1\right)-\psi\left(\frac{1}{\rho^2}-1\right)\right| \\ &+ \left|\psi\left(\frac{1}{\tau_x^2}\right)+\gamma\right| \end{aligned} \tag{24}$$

where $\psi(u) = \frac{d\log\Gamma(u)}{du}$ is the digamma function and γ is the Euler gamma constant—of numerical value about 0.577. Figure 7 shows a plot of $dE_M = \sqrt{E_M(\tau_x, \rho)}$ From equation (24). This is an approximation to the Riemannian distance but it represents the main features of the information distance of arbitrary latency 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 infectivity distribution (and a Poisson process of

infections) we obtain

$$\begin{aligned} E_M(\alpha_1,\alpha_2)_{|[T_0:\alpha_1+\alpha_2=1]} &= \left|\psi(\alpha_2)-\psi(1-\alpha_1)\right| \\ &+ \frac{1}{4}\left|\frac{(2\alpha_1+\alpha_2)^2}{4\alpha_1}-\frac{1}{2}(\alpha_1+1)\right|. \end{aligned} \tag{25}$$

This is plotted in Figure 8.

Geodesic curves in Riemannian manifolds give minimal arc length and examples are given in [3] for manifolds of Weibull, gamma and bigamma 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 the book [1].

References

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

[2] T. Britton and D. Lindenstrand. Epidemic modelling: aspects where stochasticity matters. Preprint 2008. http://arxiv.org/abs/0812.3505

[3] C.T.J. Dodson. On the entropy flows to disorder. Preprint, 2008. http://arxiv.org/abs/0811.4318

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

[5] G.P. Karev. Replicator equations and the principle of minimal production of information. Preprint, 2009. http://arxiv.org/abs/0901.2378

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