

On the entropy flows to disorder

C.T.J. Dodson

School of Mathematics
University of Manchester
Manchester M13 9PL, UK
(e-mail: ctdodson@manchester.ac.uk)

Abstract. Gamma distributions, which contain the exponential as a special case, have a distinguished place in the representation of near-Poisson randomness for statistical processes; typically, they represent distributions of spacings between events or voids among objects. Here we look at the properties of the Shannon entropy function and calculate its corresponding flow curves, relating them to examples of constrained degeneration from ordered processes. We consider univariate and bivariate gamma, and Weibull distributions since these also include exponential distributions.

Keywords: Shannon entropy, integral curves, gamma distribution, bivariate gamma, McKay distribution, Weibull distribution, randomness, information geometry. MSC classes: 82-08; 15A52.

1 Introduction

The smooth family of gamma probability density functions is given by

$$f : [0, \infty) \rightarrow [0, \infty) : x \mapsto \frac{e^{-\frac{x\kappa}{\mu}} x^{\kappa-1} \left(\frac{\kappa}{\mu}\right)^{\kappa}}{\Gamma(\kappa)} \quad \mu, \kappa > 0. \quad (1)$$

Here μ is the mean, and the standard deviation σ , given by $\kappa = \left(\frac{\mu}{\sigma}\right)^2$, is proportional to the mean. Hence the coefficient of variation $\frac{1}{\sqrt{\kappa}}$ is unity in the case that (1) reduces to the exponential distribution. Thus, $\kappa = 1$ corresponds to an underlying Poisson random process complementary to the exponential distribution. When $\kappa < 1$ the random variable X represents spacings between events that are more clustered than for a Poisson process and when $\kappa > 1$ the spacings X are more uniformly distributed than for Poisson. The case when $\mu = n$ is a positive integer and $\kappa = 2$ gives the Chi-Squared distribution with $n - 1$ degrees of freedom; this is the distribution of $\frac{(n-1)s^2}{\sigma_G^2}$ for variances s^2 of samples of size n taken from a Gaussian population with variance σ_G^2 .

The gamma distribution has a conveniently tractable information geometry [1,2], and the Riemannian metric in the 2-dimensional manifold of gamma distributions (1) is

$$[g_{ij}] (\mu, \kappa) = = \begin{bmatrix} \frac{\kappa}{\mu^2} & 0 \\ 0 & \frac{d^2}{d\kappa^2} \log(\Gamma) - \frac{1}{\kappa} \end{bmatrix}. \quad (2)$$

So the coordinates (μ, κ) yield an orthogonal basis of tangent vectors, which is useful in calculations because then the arc length function is simply

$$ds^2 = \frac{\kappa}{\mu^2} d\gamma^2 + \left(\left(\frac{\Gamma'(\kappa)}{\Gamma(\kappa)} \right)' - \frac{1}{\kappa} \right) d\kappa^2.$$

The system of geodesic equations is difficult to solve analytically but numerical solutions using the *Mathematica* programs of Gray [7] were obtained in [2]. Figure 1 shows a spray of some maximally extended geodesics emanating from the point $(\mu, \kappa) = (1, 1)$. Geodesic curves have tangent vectors that are parallel along them and yield minimal information arc length distances in the gamma 2-manifold.

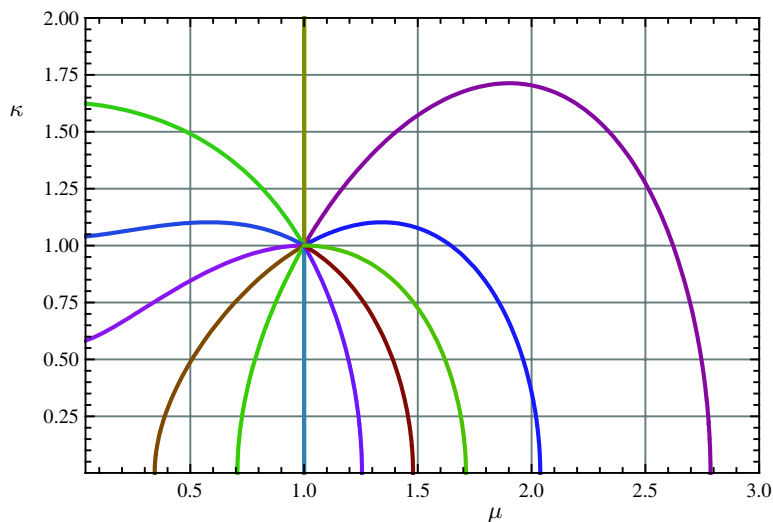


Fig. 1. Some examples of maximally extended geodesics passing through $(\mu, \kappa) = (1, 1)$ in the gamma 2-manifold.

We note the following important uniqueness property:

Theorem 1 (Hwang and Hu [8]). *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 property is one of the main reasons for the large number of applications of gamma distributions: many near-random natural processes have standard deviation approximately proportional to the mean [2].

2 Gamma entropy flows

Papoulis [13] Chapter 15 gives an account of the role of the Shannon entropy function in probability theory, stochastic processes and coding theory. The entropy of (1) is shown in Figure 2 using

$$S_f = - \int_0^\infty f \log f \, dx : \mathbb{R}^{2+} \rightarrow \mathbb{R}$$

$$(\mu, \kappa) \mapsto \kappa - \log\left(\frac{\kappa}{\mu}\right) + \log(\Gamma(\kappa)) - (\kappa - 1)\psi(\kappa) \quad (3)$$

with gradient

$$\nabla S_f(\mu, \kappa) = \left(\frac{1}{\mu}, -\frac{(\kappa - 1)(\kappa\psi'(\kappa) - 1)}{\kappa} \right). \quad (4)$$

where $\psi = \frac{\Gamma'}{\Gamma}$ is the digamma function. At fixed κ , the entropy increases like $\log \mu$. At fixed mean μ , the maximum entropy is given by $\kappa = 1$, the exponential distribution case of maximal disorder or chaos.

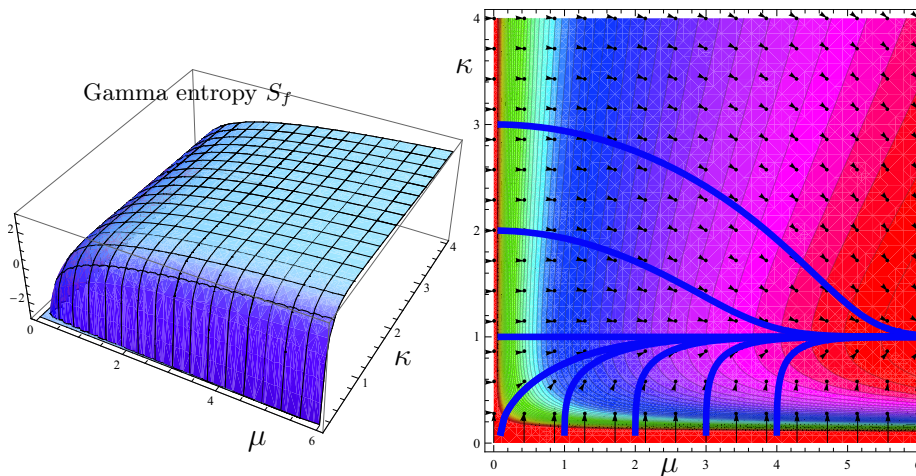


Fig. 2. Shannon entropy function S_f for the gamma family as a surface with respect to mean μ and κ (left) and as a contour plot with entropy gradient flow and integral curves (right). The asymptote is $\kappa = 1$, the exponential case of maximum disorder.

Figure 2 on the right shows entropy as a contour plot with superimposed also some examples of integral curves of the entropy gradient flow field, namely curves c satisfying

$$c : [0, \infty) \rightarrow \mathbb{R}^2 : \dot{c}(t) = \nabla S_f|_{(c(t))}. \quad (5)$$

By inspection, we can see that the entropy gradient components are each in one variable only and in particular the first component has solution

$$\mu(t) = \mu_0 e^t$$

so the mean increases exponentially with time. Such curves represent typical trajectories for processes subordinate to gamma distributions; the processes become increasingly disordered as $\kappa \rightarrow 1$. The entropy gradient curves correspond to systems with external input—the mean increases as disorder increases. The asymptote is $\kappa = 1$, the exponential case of maximum disorder. Conversely, the reverse direction of the curves corresponds to evolution from total disorder to other states (clustered for $\kappa < 1$, and smoothed out, ‘more crystal-like’, for $\kappa > 1$) while the mean is allowed to reduce—somewhat like the situation after the Big Bang, see Dodson [4].

3 Constrained degeneration of order

Lucarini [11] effectively illustrated the degeneration of order in his perturbations of the simple 3D cubic crystal lattices (SC, BCC, FCC) by an increasing spatial Gaussian noise. Physically, the perturbing spatial noise intensity corresponds somewhat to a lattice temperature in the structural symmetry degeneration. With rather moderate levels of noise, quite quickly the three tessellations became indistinguishable. In the presence of intense noise they all converged to the 3D Poisson-Voronoi tessellations, for which exact analytic results are known [6]. Moreover, in all cases the gamma distribution was an excellent model for the observed probability density functions of all metric and topological properties. See also Jarai-Szabo and Nédá [9] for some analytic approximations using gamma distributions for two and three dimensional Poisson-Voronoi cell size statistics. Lucarini provided plots showing the evolution of the mean and standard deviation of these properties as they converge asymptotically towards the Poisson-Voronoi case, illustrating the degeneration of crystallinity from $\kappa \sim \infty$ to lower values.

Of course, the constraint of remaining tessellations, albeit highly disordered ones, precludes convergence down to the maximum entropy limit $\kappa = 1$. In fact the 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 is not appropriate. However, for the positive real random variables, polyhedron volume in the limit has $\kappa \approx 5.6$ and polygon face area $\kappa \approx 16$. Lucarini [10] 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.

We can give another perspective on the level of constraint persistent in the limits for these disordered tessellations: for infinite random matrices the best fitting gamma distributions for eigenvalue spacings have $\kappa = 2.420, 4.247, 9.606$ respectively for orthogonal, unitary and symplectic Gaussian ensembles [2]. These values in a sense indicate the increasing statistical constraints of algebraic relations as the ensembles change through orthogonality, unitarity and symplectivity.

In the context of analytic number theory, gamma distributions give approximate distributions for the the reported data on spacings between zeros of the Riemann zeta function [12]. The best fit gamma distributions to spacings between the first two million zeros of the Riemann zeta function has $\kappa \approx 5.6$ [2].

For the SARS disease epidemic outbreak [3,14] the gamma distribution gave a good model for the infection process with $\kappa \approx 3$, cf. [5].

4 Bivariate gamma processes

Next we consider the bivariate case of a pair of coupled gamma processes. The McKay family can be thought of as giving the probability density for the two random variables X and $Y = X + Z$ where X and Z both have gamma distributions. This smooth bivariate family M of density functions is defined in the positive octant of random variables $0 < x < y < \infty$ with parameters $\alpha_1, c, \alpha_2 \in \mathbb{R}^+$ and probability density functions

$$m(x, y) = \frac{c^{(\alpha_1 + \alpha_2)} x^{\alpha_1 - 1} (y - x)^{\alpha_2 - 1} e^{-cy}}{\Gamma(\alpha_1) \Gamma(\alpha_2)}. \quad (6)$$

The marginal density functions, of X and Y are:

$$m_X(x) = \frac{c^{\alpha_1} x^{\alpha_1 - 1} e^{-cx}}{\Gamma(\alpha_1)}, \quad x > 0 \quad (7)$$

$$m_Y(y) = \frac{c^{(\alpha_1 + \alpha_2)} y^{(\alpha_1 + \alpha_2) - 1} e^{-cy}}{\Gamma(\alpha_1 + \alpha_2)}, \quad y > 0. \quad (8)$$

Note that we cannot have both marginal distributions exponential. The covariance and correlation coefficient of X and Y are:

$$\sigma_{12} = \frac{\alpha_1}{c^2} \quad \text{and} \quad \rho(X, Y) = \sqrt{\frac{\alpha_1}{\alpha_1 + \alpha_2}}.$$

Unlike other bivariate gamma families, the McKay information geometry is surprisingly tractable and there are a number of applications discussed in Arwini and Dodson [2]. In fact the parameters $\alpha_1, c, \alpha_2 \in \mathbb{R}^+$ are natural coordinates for the 3-dimensional manifold M of this family. The Riemannian

information metric is

$$[g_{ij}(\alpha_1, c, \alpha_2)] = \begin{bmatrix} \psi'(\alpha_1) & -\frac{1}{c} & 0 \\ -\frac{1}{c} & \frac{\alpha_1 + \alpha_2}{c^2} & -\frac{1}{c} \\ 0 & -\frac{1}{c} & \psi'(\alpha_2) \end{bmatrix}. \quad (9)$$

It is difficult to present graphics of curves in the McKay 3-manifold M , but it has an interesting 2-dimensional submanifold $M_1 \subset M$: $\alpha_1 = 1$. The density functions are of form:

$$h(x, y; 1, c, \alpha_2) = \frac{c^{1+\alpha_2}(y-x)^{\alpha_2-1}e^{-cy}}{\Gamma(\alpha_2)}, \quad (10)$$

defined on $0 < x < y < \infty$ with parameters $c, \alpha_2 \in \mathbb{R}^+$. The correlation coefficient and marginal functions of X and Y are given by:

$$\rho(X, Y) = \frac{1}{\sqrt{1 + \alpha_2}} \quad (11)$$

$$h_X(x) = ce^{-cx}, \quad x > 0 \quad (12)$$

$$h_Y(y) = \frac{c^{(1+\alpha_2)}y^{\alpha_2}e^{-cy}}{\alpha_2\Gamma(\alpha_2)}, \quad y > 0 \quad (13)$$

In fact $\alpha_2 = \frac{1-\rho^2}{\rho^2}$, which in applications would give a measure of the variability not due to the correlation. The matrix of metric components $[g_{ij}]$ on M_1 is

$$[g_{ij}] = \begin{bmatrix} \frac{1+\alpha_2}{c^2} & -\frac{1}{c} \\ -\frac{1}{c} & \psi'(\alpha_2) \end{bmatrix}. \quad (14)$$

Some geodesics emanating from $(\alpha_2, c) = (1, 1) \in M_1$ are shown on the left of Figure 3. The density functions can be presented also in terms of the positive parameters $(\alpha_1, \sigma_{12}, \alpha_2)$ where σ_{12} is the covariance of X and Y

$$m(x, y; \alpha_1, \sigma_{12}, \alpha_2) = \frac{\left(\frac{\alpha_1}{\sigma_{12}}\right)^{\frac{(\alpha_1+\alpha_2)}{2}} x^{\alpha_1-1} (y-x)^{\alpha_2-1} e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}}y}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \quad (15)$$

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

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

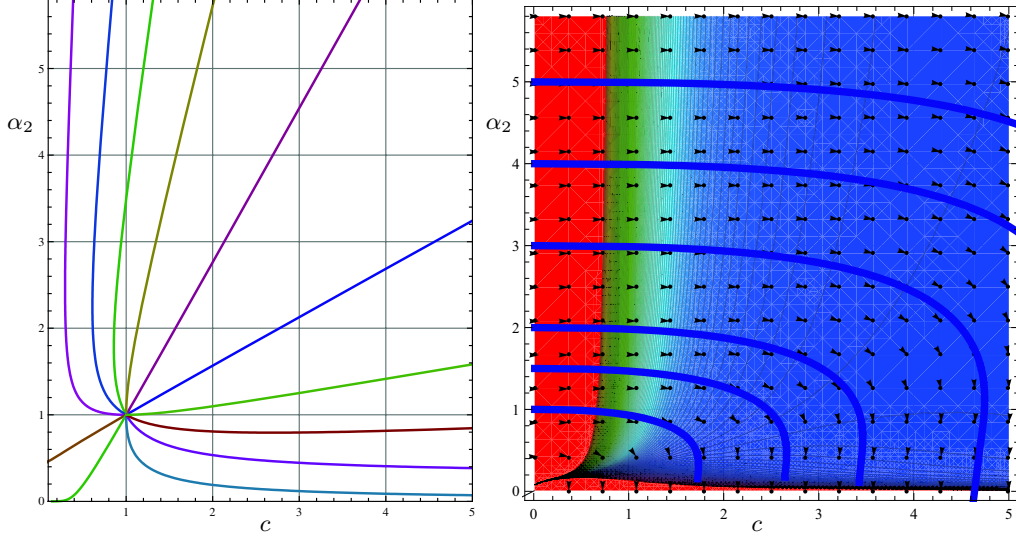


Fig. 3. Geodesics passing through $(c, \alpha_2) = (1, 1)$ (left) and a contour plot of the entropy with some integral gradient curves (right), in the McKay submanifold M_1 which has $\alpha_1 = 1$.

The entropy function is

$$S_m : \mathbb{R}^{3+} \rightarrow \mathbb{R} : (\alpha_1, c, \alpha_2) \mapsto \sqrt{\alpha_1} c^{-\alpha_1-1} K \quad (18)$$

$$K = \log \frac{c^2}{\Gamma(\alpha_1)\Gamma(\alpha_2)} + (\alpha_1 - 1)\psi(\alpha_1) + (\alpha_2 - 1)\psi(\alpha_2) - (\alpha_1 + \alpha_2).$$

On M_1

$$S_{m|M_1} = \frac{1}{c^2} \left(\log \frac{c^2}{\Gamma(\alpha_2)} + (\alpha_2 - 1)\psi(\alpha_2) - (1 + \alpha_2) \right) \quad (19)$$

$$\nabla S_{m|M_1} = \left(\frac{2}{c^3} \left(\log \left(\frac{\Gamma(\alpha_2)}{c^2} \right) - \psi(\alpha_2)(\alpha_2 - 1) + \alpha_2 + 2 \right), \frac{1}{c^2} (\psi'(\alpha_2)(\alpha_2 - 1) - 1) \right).$$

On the right of Figure 3 is a contour plot of the entropy showing its gradient field and some integral curves. It may be helpful to express the entropy in terms of α_2 and ρ , which gives

$$S_{m|M_1} = \frac{4\rho^4}{(\rho^2 + 1)^2} \left(\log \frac{(\rho^2 + 1)^2}{4\rho^4 \Gamma(\alpha_2)} + (\alpha_2 - 1)\psi(\alpha_2) - (1 + \alpha_2) \right). \quad (20)$$

The McKay entropy in M_1 with respect to α_2, ρ is shown in Figure 4 (left) with the gradient flow on a contour plot (right) together with the approximate locus curve of maximum entropy. The two thinner curves show the loci of $\alpha_1 = 1$ (upper curve) and $\alpha_1 + \alpha_2 = 1$ (lower curve), which correspond, respectively, to Poisson random processes for the X and Y variables.

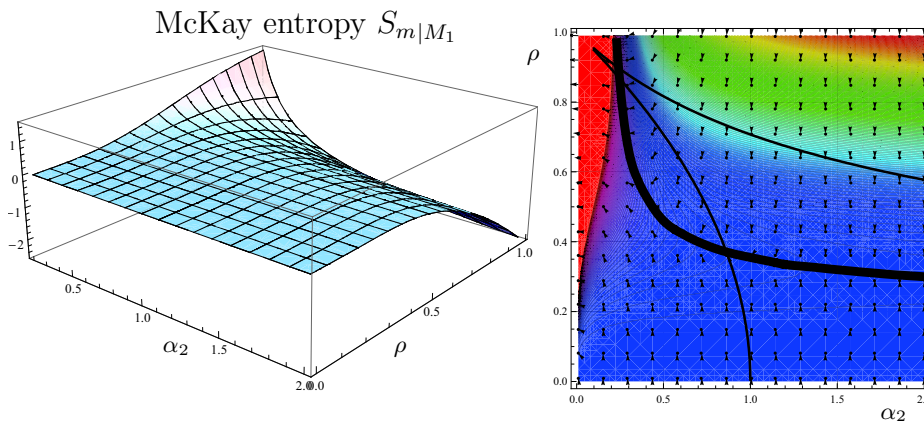


Fig. 4. Surface representation of the Shannon entropy function S_m for the submanifold M_1 of the McKay family with respect to parameter α_2 and correlation coefficient ρ (left), and contour plot (right). Superimposed also are gradient flow arrows and the thick curve is the locus of maximum entropy. The two thinner curves show the loci of $\alpha_1 = 1$ (upper curve) and $\alpha_1 + \alpha_2 = 1$ (lower curve), which correspond, respectively, to Poisson random processes for the X and Y variables.

Qualitatively, what we may see in Figure 4 is that for correlated random variables X and Y subordinate to the bivariate gamma density (10), the maximum entropy locus is roughly hyperbolic. The maximum entropy curve is rather insensitive to the correlation coefficient ρ when $\alpha_2 > 1$ and the difference $Y - X$ is dispersed more evenly than Poisson. When $Y - X$ is actually Poisson random, with $\alpha_2 = 1$, the critical value is at $\rho \approx 0.355$. However, as α_2 reduces further—corresponding to clustering of $Y - X$ values—so the locus turns rapidly to increasing correlation. If we take the situation of a bivariate gamma process with constant marginal mean values, then the McKay probability density has constant correlation coefficient ρ ; in this case the gradient flow lies along lines of constant α_2 .

Dodson [5] gives an application to an epidemic model in which the latency and infectivity for individuals in a population are jointly distributed properties controlled by a bivariate gamma distribution.

5 Weibull processes

Like the gamma family, the Weibull family of distributions contains the exponential distribution as a special case; it has wide application in models for reliability and lifetime statistics for random variable $t > 0$. The probability density function can be presented in terms of positive parameters ξ, β

$$w : \mathbb{R}^+ \rightarrow \mathbb{R} : t \mapsto \frac{\beta}{\xi} \left(\frac{t}{\xi} \right)^{\beta-1} e^{-(t/\xi)^\beta}. \quad (21)$$

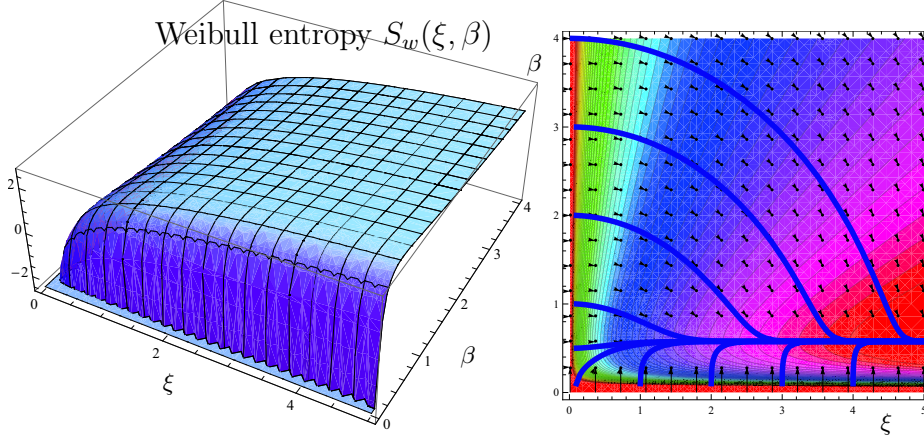


Fig. 5. Surface representation of the Shannon entropy function S_w for the Weibull family (left), and contour plot (right) with gradient flow and integral curves.

In applications of (21), reliability $R(t)$ is the probability of survival to time t and it is related to the failure rate $Z(t)$ at time t through

$$R(t) = \int_t^\infty w(t) dt = e^{-(t/\xi)^\beta} \text{ and } Z(t) = \frac{w(t)}{R(t)} = \beta \left(\frac{1}{\xi}\right)^\beta t^{\beta-1}. \quad (22)$$

The Weibull mean, standard deviation and entropy are

$$\mu_w = \xi \Gamma\left(1 + \frac{1}{\beta}\right) \quad (23)$$

$$\sigma_w = \xi \sqrt{\Gamma\left(\frac{\beta+2}{\beta}\right) - \Gamma\left(1 + \frac{1}{\beta}\right)^2} \quad (24)$$

$$S_w(\xi, \beta) = -\log(\beta) - \log\left(\frac{1}{\xi}\right) - \frac{\gamma}{\beta} + \gamma + 1 \quad (25)$$

$$\nabla S_w(\xi, \beta) = \left(\frac{1}{\xi}, \frac{\gamma - \beta}{\beta^2}\right). \quad (26)$$

In (25) γ is the Euler constant, approximately 0.577. In case $\beta = \frac{1}{n}$ for positive integer n , then the coefficient of variation is $\frac{\sigma_w}{\mu_w} = \sqrt{\frac{(2n)!}{(n!)^2} - 1}$, and we see that $\beta = 1$ reduces (21) to the exponential density with $\mu_w = \sigma_w = \xi$ and unit coefficient of variation. Figure 5 shows a surface plot of the Weibull entropy S_w and a corresponding contour plot with gradient flow and some integral curves. There is a resemblance to the gamma distribution entropy in Figure 2 and from (26) again here we have $\mu_w(t) = \mu_w(0)e^t$, but in the Weibull case the asymptotic curve has $\beta = \gamma \approx 0.577$, which corresponds to a process with coefficient of variation ≈ 4.65 compared with unity for the exponential distribution and Poisson randomness.

Acknowledgement

The author is grateful to V. Lucarini for discussions concerning the interpretation of his simulations.

References

- 1.S-I. Amari and H. Nagaoka. *Methods of Information Geometry*. American Mathematical Society, Oxford University Press, Oxford, 2000.
- 2.Khadiga Arwini and C. T. J. Dodson. *Information Geometry: Near Randomness and Near Independence*. Springer-Verlag, New York, Berlin, 2008.
- 3.T. Britton and D. Lindenstrand. Epidemic modelling: aspects where stochasticity matters. <http://arxiv.org/abs/0812.3505>, pages 1–16, 2008.
- 4.C. T. J. Dodson. Quantifying galactic clustering and departures from randomness of the inter-galactic void probability function using information geometry. <http://arxiv.org/abs/astro-ph/0608511>, 2006.
- 5.C. T. J. Dodson. Information geometry and entropy in a stochastic epidemic rate process. <http://arxiv.org/abs/0903.2997>, 2009.
- 6.S.R. Finch. *Mathematical Constants*. Cambridge University Press, Cambridge, 2003.
- 7.A. Gray. *Modern Differential Geometry of Curves and Surfaces 2nd Edition*. CRC Press, Boca Raton, 1998.
- 8.T-Y. Hwang and C-Y. Hu. On a characterization of the gamma distribution: The independence of the sample mean and the sample coefficient of variation. *Annals Inst. Statist. Math.*, 51(4):749–753, 1999.
- 9.F. Jarai-Szabo and Z. Neda. On the size distribution of poisson voronoi cells. <http://arxiv.org/abs/cond-mat/0406116v2>, 2008.
- 10.Valerio Lucarini. From symmetry breaking to poisson point processes in 2d voronoi tessellations: the generic nature of hexagons. *J. Statist. Phys.*, 130:1047–1062, 2008.
- 11.Valerio Lucarini. Three-dimensional random voronoi tessellations: From cubic crystal lattices to poisson point processes. *J. Statist. Phys.*, page In press, 2008.
- 12.A. Odlyzko. Tables of zeros of the riemann zeta function . http://www.dtc.umn.edu/~odlyzko/zeta_tables/index.html, 2009.
- 13.A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York.
- 14.WHO. Cumulative number of reported probable cases of severe acute respiratory syndrome (sars). <http://www.who.int/csr/sars/country/en/>, 2003.