Topics in Information Geometry

Dodson, CTJ

2005

MIMS EPrint: **2005.49**

ISSN 1749-9097

Topics in Information Geometry

C.T.J.Dodson ©2004
School of Mathematics, Manchester University, Manchester M60 1QD UK
Print date: July 28, 2005

Contents

Preface

The original draft of these lecture notes was prepared for a short course at Centro de Investigaciones de Matematica (CIMAT), Guanajuato, Mexico in May 2004. Many of the calculations arise from joint publications with Khadiga Arwini and from her PhD thesis [6]. The course was given also in the Departamento de Xeometra e Topoloxa, Facultade de Matemticas, Universidade de Santiago de Compostela, Spain in February 2005.

Chapter 1

Mathematical statistics and information theory

There are many easily found good books on probability theory and mathematical statistics (eg [31, 32, 33, 42, 44, 45, 69]), stochastic processes (eg [11, 60]) and information theory (eg [62]); here we just outline some topics to help make the sequel more self contained. For those who have access to the computer algebra package *Mathematica*[73], the approach to mathematical statistics and accompanying software in Rose and Smith [63] will be particularly helpful.

The word stochastic comes from the Greek *stochastikos*, meaning skillful in aiming and *stochazesthai* to aim at or guess at, and *stochos* means target or aim. In our context, stochastic means involving chance variations around some event—rather like the variation in positions of strikes aimed at a target. In its turn, the later word statistics comes through eighteenth century German from the Latin root *status* meaning state; originally it meant the study of political facts and figures. The noun *random* was used in the sixteenth century to mean a haphazard course, from the Germanic *randir* to run, and as an adjective to mean without a definite aim, rule or method, the opposite of purposive. From the middle of the last century, the concept of a *random variable* has been used to describe a variable that is a function of the result of a well-defined statistical experiment in which each possible outcome has a definite probability of occurrence. The organization of probabilities of outcomes is achieved by means of a *probability function* for discrete random variables and by means of a *probability density function* for continuous random variables. The result of throwing two fair dice and summing what they show is a discrete random variable.

Mainly, we are concerned with *continuous random variables* (here measurable functions defined on some $\mathbb{R}^n$) with differentiable probability density measure functions, but we do need also to mention the Poisson distribution for the discrete case. However, since the Poisson is a limiting approximation to the Binomial distribution which arises from the Bernoulli distribution (which everyone encountered in school!) we mention also those examples.

1.1 Probability functions

Here we are concerned with discrete random variables so we take the domain set to be $\mathbb{N} \cup \{0\}$. We may view a *probability function* as a subadditive measure function of unit weight on $\mathbb{N} \cup \{0\}$

$$p : \mathbb{N} \cup \{0\} \to [0, \infty) \quad \text{(nonnegativity)} \tag{1.1}$$

$$\sum_{k=0}^{\infty} p(k) = 1 \quad \text{(unit weight)} \tag{1.2}$$

$$p(A \cup B) \leq p(A) + p(B), \ \forall A, B \subset \mathbb{N} \cup \{0\}, \quad \text{(subadditivity)} \tag{1.3}$$
$$\text{with equality} \iff A \cap B = \emptyset.$$

Formally, we have a discrete measure space of total measure 1 with σ-algebra the power set and measure function induced by p

$$sub(\mathbb{N}) \to [0, \infty) : A \mapsto \sum_{k \in A} p(k)$$

and as we have anticipated above, we usually abbreviate $\sum_{k \in A} p(k) = p(A)$.

We have the following *expected values* of the random variable and its square

$$\mathcal{E}(k) = \overline{k} = \sum_{k=0}^{\infty} k p(k) \tag{1.4}$$

$$\mathcal{E}(k^2) = \overline{k^2} = \sum_{k=0}^{\infty} k^2 p(k). \tag{1.5}$$

With slight but common abuse of notation, we call $\overline{k}$ the mean, $\overline{k^2} - (\overline{k})^2$ the variance, $\sigma_k = +\sqrt{\overline{k^2} - (\overline{k})^2}$ the standard deviation and $\sigma_k / \overline{k}$ the coefficient of variation, respectively, of the random variable k. The variance is the square of the standard deviation.

The *moment generating function* $\Psi(t) = \mathcal{E}(e^{tX}),\ t \in \mathbb{R}$ of a distribution generates the r^{th} moment as the value of the r^{th} derivative of Ψ evaluated at $t = 0$. Hence, in particular, the mean and variance are given by:

$$\mathcal{E}(X) = \Psi'(0) \tag{1.6}$$

$$Var(X) = \Psi''(0) - (\Psi'(0))^2, \tag{1.7}$$

which can provide an easier method for their computation in some cases.

1.1.1 Example: Bernoulli distribution

It is said that a random variable X has a **Bernoulli distribution** with parameter p $(0 \leq p \leq 1)$ if X can take only the values 0 and 1 and the probabilities are

$$P_r(X = 1) = p \tag{1.8}$$

$$P_r(X = 0) = 1 - p \tag{1.9}$$

Then the probability function of X can be written as follows:

$$f(x|p) = \begin{cases} p^x (1-p)^{1-x} & \text{if } x = 0, 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.10}$$

If X has a Bernoulli distribution with parameter p, then we can find its expectation or mean value $\mathcal{E}(X)$ and variance $Var(X)$ as follows.

$$\mathcal{E}(X) = 1 \cdot p + 0 \cdot (1 - p) = p \tag{1.11}$$

$$Var(X) = \mathcal{E}(X^2) - (\mathcal{E}(X))^2 = p - p^2 \tag{1.12}$$

The moment generating function of X is the expectation of e^{tX},

$$\Psi(t) = \mathcal{E}(e^{tX}) = pe^t + q \tag{1.13}$$

which is finite for all real t.

1.1.2 Example: Binomial distribution

If n random variables $X_1, X_2, \ldots, X_n$ are independently identically distributed, and each has a Bernoulli distribution with parameter p, then it is said that the variables $X_1, X_2, \ldots, X_n$ form n Bernoulli trials with parameter p.

If the random variables $X_1, X_2, \ldots, X_n$ form n Bernoulli trials with parameter p and if $X = X_1 + X_2 + \ldots + X_n$, then X has a **binomial distribution** with parameters n and p.

The binomial distribution is of fundamental importance in probability and statistics because of the following result for any experiment which can result in only either success or failure. The experiment is performed n times independently and the probability of the success of any given performance is p. If X denotes the total number of successes in the n performances, then X has a binomial distribution with parameters n and p. The probability function of X is:

$$P(X = r) = P(\sum_{i=1}^{n} X_i = r) = \binom{n}{r} p^r (1-p)^{n-r} \tag{1.14}$$

where $r = 0, 1, 2, \ldots, n$.

We write

$$f(r|p) = \begin{cases} \binom{n}{r} p^r (1-p)^{n-r} & \text{if r=0, 1, 2, \ldots, n} \\ 0 & \text{otherwise} \end{cases} \tag{1.15}$$

In this distribution n must be a positive integer and p must lie in the interval $0 \leq p \leq 1$. If X is represented by the sum of n Bernoulli trials, then it is easy to get its expectation, variance and moment generating function by using the properties of sums of random variables.

$$\mathcal{E}(X) = \sum_{i=1}^{n} \mathcal{E}(X_i) = np \tag{1.16}$$

$$Var(X) = \sum_{i=1}^{n} Var(X_i) = np(1-p) \tag{1.17}$$

$$\Psi(t) = \mathcal{E}(e^{tX}) = \prod_{i=1}^{n} \mathcal{E}(e^{tX_i}) = (pe^t + q)^n. \tag{1.18}$$

1.1.3 Example: The Poisson probability function

The Poisson distribution is widely discussed in the statistical literature; one monograph devoted to it and its applications is Haight [38]. Take $t, \tau \in (0, \infty)$

$$p : \mathbb{N} \cup \{0\} \to [0, \infty) : k \mapsto \left(\frac{t}{\tau}\right)^k \frac{1}{k!} e^{-t/\tau} \tag{1.19}$$

$$\overline{k} = t/\tau \tag{1.20}$$

$$\sigma_k = t/\tau. \tag{1.21}$$

This probability function is used to model the number k of random events in a region of measure t when the mean number of events per unit region is τ. Importantly, the Poisson distribution can give a good approximation to the binomial distribution when n is large and p is close to 0. This is easy to see by making the correspondences:

$$e^{-pn} \longrightarrow (1-(n-r)p) \tag{1.22}$$

$$n!/(n-r)! \longrightarrow n^r. \tag{1.23}$$

1.2 Probability density functions

We are concerned with the case of continuous random variables defined on some $\Omega \subseteq \mathbb{R}^m$. For our present purposes we may view a *probability density function* (pdf) on $\Omega \subseteq \mathbb{R}^m$ as a subadditive measure function of unit weight, namely, a nonnegative map on Ω

$$f \quad : \quad \Omega \to [0,\infty) \qquad \text{(nonnegativity)} \tag{1.24}$$

$$\int_\Omega f \quad = \quad f(\Omega) = 1 \qquad \text{(unit weight)} \tag{1.25}$$

$$f(A \cup B) \quad \leq \quad f(A) + f(B),\ \forall A, B \subset \Omega, \qquad \text{(subadditivity)} \tag{1.26}$$

$$\text{with equality} \iff A \cap B = \emptyset.$$

Formally, we have a measure space of total measure 1 with σ-algebra typically the Borel sets or the power set and the measure function induced by f

$$sub(\Omega) \to [0,\infty) : A \mapsto \int_A f$$

and as we have anticipated above, we usually abbreviate $\int_A f = f(A)$. Given an integrable (ie measurable in the σ-algebra) function $u : \Omega \to \mathbb{R}$, the expectation or mean value of u is defined to be

$$\mathcal{E}(u) = \overline{u} = \int_\Omega uf.$$

We say that f is the joint pdf for the random variables $x_1, x_2, \ldots, x_m$, being the coordinates of points in Ω, or that these random variables have the joint probability distribution f. If x is one of these random variables, and in particular for the important case of a single random variable x, we have the following

$$\mathcal{E}(x) = \overline{x} \quad = \quad \int_\Omega xf \tag{1.27}$$

$$\mathcal{E}(x^2) = \overline{x^2} \quad = \quad \int_\Omega x^2 f. \tag{1.28}$$

Again with slight abuse of notation, we call $\overline{x}$ the mean, $\overline{x^2} - (\overline{x})^2$ the variance, $\sigma_x = +\sqrt{\overline{x^2} - (\overline{x})^2}$ the standard deviation and $\sigma_x/\overline{x}$ the coefficient of variation, respectively, of the random variable x. The variance is the square of the standard deviation. Usually, a probability density function depends on a set of parameters, $\theta_1, \theta_2, \ldots, \theta_n$ and we say that we have an n-dimensional family of pdfs.

1.2.1 Example: The exponential pdf

Take $\lambda \in \mathbb{R}^+$; this is called the parameter of the exponential pdf

$$f \quad : \quad [0,\infty) \to [0,\infty) : [a,b] \mapsto \int_{[a,b]} \frac{1}{\lambda} e^{-x/\lambda} \tag{1.29}$$

$$\overline{x} \quad = \quad \lambda \tag{1.30}$$

$$\sigma_x \quad = \quad \lambda. \tag{1.31}$$

The parameter space of the exponential distribution is $\mathbb{R}^+$, so exponential distributions form a 1-parameter family. In the sequel we shall see that quite generally we may provide a Riemannian structure to the parameter space of a family of distributions. Sometimes we call a family of pdfs a parametric statistical model.

Observe that, in the Poisson probability function (1.19), the value at zero is

$$p(0) = e^{-t/\tau}$$

and it is not difficult to show that the probability density function for the distance t between random events on $[0, \infty)$ is an exponential pdf 1.29 given by

$$f : [0, \infty) \to [0, \infty) : t \mapsto \frac{1}{\tau} e^{-t/\tau}$$

where τ is the mean number of events per unit interval.

1.3 Joint probability density functions

Let f be a pdf, defined on $\mathbb{R}^2$ (or some subset thereof). This is an important case since here we have two variables, X, Y, say, and we can extract certain features of how they interact. In particular, we define their respective mean values and their covariance, σ_{xy} :

$$\overline{x} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \ f(x, y) \, dxdy \tag{1.32}$$

$$\overline{y} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \ f(x, y) \, dxdy \tag{1.33}$$

$$\sigma_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \ f(x, y) \, dxdy \ - \overline{x}\overline{y}. \tag{1.34}$$

The *marginal* pdf of X is f_X, obtained by integrating f over all y,

$$f_X(x) = \int_{v=-\infty}^{\infty} f_{X,Y}(x, v) \, dv \tag{1.35}$$

and similarly the marginal pdf of Y is

$$f_Y(y) = \int_{u=-\infty}^{\infty} f_{X,Y}(u, y) \, du \tag{1.36}$$

The jointly distributed random variables X and Y are called independent if their marginal density functions satisfy

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \quad for\, all \quad x, y \in R \tag{1.37}$$

It is easily shown that if the variables are independent then their covariance (1.34) is zero but the converse is not true. Feller [31] gives a simple counterexample: Let X take values $-1, +1, -2, +2$, each with probability $\frac{1}{4}$ Let $Y = X^2$; then the covariance is zero but there is evidently a (nonlinear) dependence.

The extent of dependence between two random variables can be measured in a normalised way by means of the correlation coefficient: the ratio of the covariance to the product of marginal standard deviations:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \tag{1.38}$$

Note that $-1 \leq \rho_{xy} \leq 1$, whenever it exists, the limiting values corresponding to the case of linear dependence between the variables. Intuitively, $\rho_{xy} < 0$ if y tends to increase as x decreases, and $\rho_{xy} > 0$ if x and y tend to increase together.

1.3.1 Example: Bivariate Gaussian distributions

The probability density function of the two-dimensional Gaussian distribution has the form:

$$(1.39)\quad f(x,y) = \frac{1}{2\,\pi\,\sqrt{\sigma_1\,\sigma_2 - {\sigma_{12}}^2}}\, e^{-\frac{1}{2\,(\sigma_1\,\sigma_2 - {\sigma_{12}}^2)}\left(\sigma_2 (x-\mu_1)^2 - 2\;\sigma_{12}\,(x-\mu_1)\,(y-\mu_2) + \sigma_1 (y-\mu_2)^2\right)},$$

where

$$-\infty < x_1 < x_2 < \infty, \quad -\infty < \mu_1 < \mu_2 < \infty, \quad 0 < \sigma_1, \sigma_2 < \infty.$$

This contains the five parameters $(\mu_1, \mu_2, \sigma_1, \sigma_{12}, \sigma_2) = (\xi^1, \xi^2, \xi^3, \xi^4, \xi^5) \in \Theta$. So we have a five-dimensional parameter space Θ.

1.4 Information theory

Jaynes [41] provided the foundation for information theoretic methods in, among other things, Bayes hypothesis testing—as described by Tribus et al. [70, 71]; see also Shannon [64], Slepian [65] and Roman [62].

Given a set of observed values $< g_\alpha(t) >$ for functions g_α of the random variable x, we seek a 'least prejudiced' set of probability values p_i satisfying

$$(1.40)\qquad < g_\alpha(t) > \;=\; \sum_{i=1}^{i=n} p_i\, g_\alpha(x_i) \quad \text{for } \alpha = 1, 2, \ldots, N$$

$$(1.41)\qquad 1 \;=\; \sum_{i=1}^{i=n} p_i$$

Jaynes showed that this occurs if we choose those p_i that maximize Shannon's entropy function

$$(1.42)\qquad S = -\sum_{i=1}^{i=n} p_i \log(p_i)$$

It turns out [71] that if we have no data on observed functions of x, (so the set of equations (1.40) is empty) then the maximum entropy choice gives the exponential distribution. If we have estimates of the first two moments of the distribution of x, then we obtain the (truncated) Gaussian. If we have estimates of the mean and mean logarithm of x, then the maximum entropy choice is the gamma distribution.

In the sequel we shall consider the particular case of the gamma distribution for several reasons:

- the exponential distributions form a subclass of gamma distributions and exponential distributions represent intervals between events in a Poisson (ie. 'random') process
- the sum of n independent identical exponential random variables follows a gamma distribution
- lognormal distributions may be well-approximated by gamma distributions
- products of gamma distributions are well-approximated by gamma distributions
- stochastic porous media have been modeled using gamma distributions [26].

Other parametric statistical models based on different distributions may be treated in a similar way.

Let Θ be the parameter space of a parametric statistical model, that is an n-dimensional smooth family of probability density functions defined on some fixed event space Ω of unit measure,

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

For each sequence $X = \{X_1, X_2, \ldots, X_n\}$, of independent identically distributed observed values, the *likelihood function* lik_X on Θ which measures the likelihood of the sequence arising from different $p_\theta \in S$ is defined by

$$lik_X : \Theta \to [0,1] : \theta \mapsto \prod_{i=1}^{n} p_\theta(X_i).$$

Statisticians use the likelihood function, or *log-likelihood* its logarithm $l = \log lik$, in the evaluation of goodness of fit of statistical models. The so-called 'method of maximum likelihood' is used to obtain optimal fitting of the parameters in a distribution to observed data.

1.4.1 Example: Maximum likelihood gamma pdf

The family of gamma distributions has event space $\Omega = \mathbb{R}^+$ and probability density functions given by

$$\Theta \equiv \{f(x;\gamma,\mu) | \gamma, \mu \in \mathbb{R}^+\}$$

so here $\Theta = \mathbb{R}^+ \times \mathbb{R}^+$ and the random variable is $x \in \Omega = \mathbb{R}^+$ with

$$f(x;\gamma,\mu) = \left(\frac{\mu}{\gamma}\right)^{\mu} \frac{x^{\mu-1}}{\Gamma(\mu)} e^{-x\mu/\gamma} \tag{1.43}$$

Then $\bar{x} = \gamma$ and $Var(x) = \gamma^2/\mu$ and we see that γ controls the mean of the distribution while μ controls its variance and hence the shape. Indeed, the property that the variance is proportional to the square of the mean actually characterizes gamma distributions as shown recently by Hwang and Hu [39]. They proved, for $n \geq 3$ independent positive random variables $x_1, x_2, \ldots, x_n$ with a common continuous probability density function h, that having independence of the sample mean and sample coefficient of variation is equivalent to h being a gamma distribution. The special case $\mu = 1$ corresponds to the situation of the random or Poisson process with mean inter-event interval γ. In fact, for *integer* $\mu = 1, 2, \ldots$, (**??**) models a process that is Poisson but with intermediate events removed to leave only every ν^{th}. Formally, the gamma distribution is the ν-fold convolution of the exponential distribution, called also the Pearson Type III distribution. Figure 1.1 shows a family of gamma distributions, all of unit mean, with $\mu = 0.5,\ 1,\ 2,\ 7$.

Shannon's information theoretic entropy or 'uncertainty' is given, up to a factor, by the negative of the expectation of the logarithm of the probability density function (**??**), that is

$$S_f(\gamma,\mu) = -\int_0^\infty \log(f(x;\gamma,\mu))\, f(x;\gamma,\mu)\, dx = \mu + (1-\mu)\frac{\Gamma'(\mu)}{\Gamma(\mu)} + \log\frac{\gamma\,\Gamma(\mu)}{\mu} \tag{1.44}$$

At unit mean, the maximum entropy (or maximum uncertainty) occurs at $\mu = 1$, which is the random case, and then $S_f(\gamma, 1) = 1 + \log\gamma$. So, a Poisson process of points on a line is such that the points are as disorderly as possible and among all homogeneous point processes with a given density, the Poisson process has maximum entropy. Figure 1.2 shows a plot of $S_f(\mu,\beta)$, for the case of unit mean $\mu = 1$.

We can see the role of the log-likelihood function in the case of a set $X = \{X_1, X_2, \ldots, X_n\}$ of measurements, drawn from independent identically distributed random variables, to which we wish to fit the maximum likelihood gamma distribution. The procedure to optimize the choice of γ, μ is as follows. For independent events X_i, with identical distribution $p(x;\gamma,\mu)$, their joint probability density is the product of the marginal densities so a measure of the 'likelihood' of finding such a set of events is

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

We seek a choice of γ, μ to maximize this product and since the log function is monotonic increasing it is simpler to maximize the logarithm

$$l_X(\gamma,\mu) = \log lik_X(\gamma,\mu) = \log[\prod_{i=1}^{n} f(X_i;\gamma,\mu)].$$

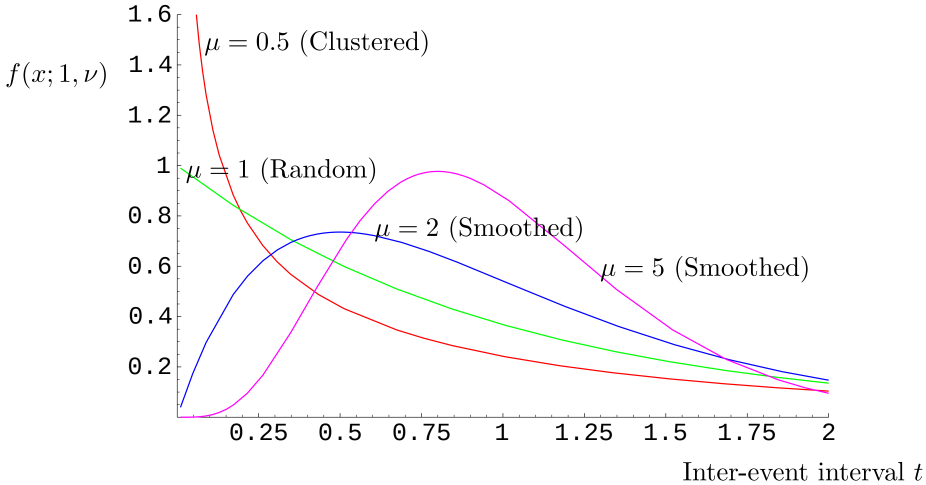

Figure 1.1: *Probability density functions, $f(x; \gamma, \mu)$, for gamma distributions of inter-event intervals t with unit mean $\gamma = 1$, and $\mu = 0.5,\ 1,\ 2,\ 5$. The case $\mu = 1$ corresponds to an exponential distribution from an underlying Poisson process; $\nu \neq 1$ represents some organization—clustering ($\mu < 1$) or smoothing ($\mu > 1$).*

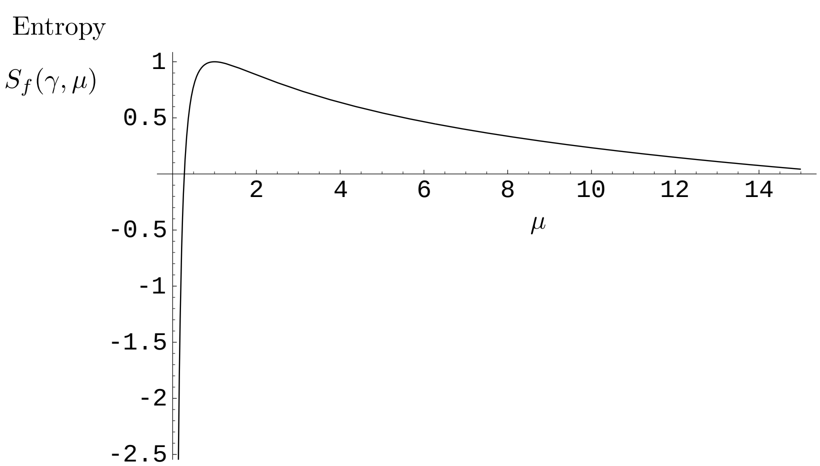

Figure 1.2: *Information theoretic entropy $S_f(\gamma, \mu)$, for gamma distributions of inter-event intervals t with unit mean $\gamma = 1$. The maximum at $\mu = 1$ corresponds to an exponential distribution from an underlying Poisson process. The regime $\mu < 1$ corresponds to clustering of events and $\nu > 1$ corresponds to smoothing out of events, relative to a Poisson process. Note that, at constant mean, the variance of x decays like $1/\mu$.*

Substitution gives us

$$\begin{aligned} l_X(\gamma,\mu) &= \sum_{i=1}^{n}[\mu(\log\mu-\log\gamma)+(\mu-1)\log X_i-\frac{\mu}{\gamma}X_i-\log\Gamma(\mu)] \\ &= n\mu(\log\mu-\log\gamma)+(\mu-1)\sum_{i=1}^{n}\log X_i-\frac{\mu}{\gamma}\sum_{i=1}^{n}X_i-n\log\Gamma(\mu). \end{aligned}$$

Then, solving for $\partial_\gamma l_X(\gamma,\mu)=\partial_\mu l_X(\gamma,\mu)=0$ in terms of properties of the X_i, we obtain the maximum likelihood estimates $\hat{\gamma},\hat{\mu}$ of γ,μ in terms of the mean and mean logarithm of the X_i

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

where $\overline{\log X}=\frac{1}{n}\sum_{i=1}^{n}\log X_i$.

Chapter 2

Information geometry

Amari [3] and Amari and Nagaoka [4] provide modern accounts of the differential geometry that arises from the Fisher information metric.

2.1 Fisher information metric

Let Θ be the parameter space of a parametric statistical model, that is an n-dimensional smooth family of probability density functions defined on some fixed event space Ω of unit measure,

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

Denote by $\mathcal{E}$ the expectation operator (measure) for functions defined on Ω; in particular, the mean $\mathcal{E}(x) = \bar{x}$, and variance $\mathcal{E}(x^2) - \bar{x}^2 = Var(x)$, which will be functions of θ. Then, at each point $\theta \in \Theta$, the covariance of partial derivatives of the log-likelihood function, $l = \log p_\theta$, is a matrix with entries the expectations

$$g_{ij} = \mathcal{E}\left(\frac{\partial l}{\partial \theta^i}\frac{\partial l}{\partial \theta^j}\right) = -\mathcal{E}\left(\frac{\partial^2 l}{\partial \theta^i \partial \theta^j}\right) \quad \text{(for coordinates } (\theta^i) \text{ about } \theta \in \Theta). \tag{2.1}$$

This gives rise to a positive definite matrix, so inducing a Riemannian metric g on Θ using for coordinates the parameters (θ^i); this metric is called the expected information metric for the family of probability density functions; the original ideas are due to Fisher and Rao [34, 61]. Of course, the second equality in equation (2.1) depends on certain regularity conditions but when it holds it can be particularly convenient to use. Amari [3] and Amari and Nagaoka [4] provide modern accounts of the differential geometry that arises from the Fisher information metric.

2.2 Exponential family

An n-dimensional parametric statistical model $\Theta \equiv \{p_\theta | \theta \in \Theta\}$is said to be an **exponential family** or of **exponential type**, when the density function can be expressed in terms of functions $\{C, F_1, ..., F_n\}$ on Λ and a function ψ on Θ as:

$$p(x;\theta) = e^{\{C(x) + \sum_i \theta_i F_i(x) - \psi(\theta)\}}\,, \tag{2.2}$$

then we say that (θ_i) are its **natural** or **canonical parameters**, and ψ is the **potential function**. From the normalization condition $\int p(x;\theta)\,dx = 1$ we obtain:

$$\psi(\theta) = \log \int e^{\{C(x) + \sum_i \theta_i F_i(x)\}}\,dx\,. \tag{2.3}$$

This potential function is therefore a distinguished function of the coordinates alone and in the sequel we make use of it for the presentation of the manifold as an immersion in $\mathbb{R}^{n+1}$.

2.3 α-connections

Let $\Theta \equiv \{p_\xi\}$ be an n-dimensional model, and consider the function $\Gamma^{(\alpha)}_{ij,k}$ which maps each point ξ to the following value:

$$\left(\Gamma^{(\alpha)}_{ij,k}\right)_\xi = E\left[\left(\frac{\partial^2 \log f}{\partial\xi^i\partial\xi^j} + \frac{1-\alpha}{2}\frac{\partial \log f}{\partial\xi^i}\frac{\partial \log f}{\partial\xi^j}\right)\frac{\partial \log f}{\partial\xi^k}\right], \tag{2.4}$$

where α is some arbitrary real number. So we have an affine connection $\nabla^{(\alpha)}$ on Θ defined by

$$\langle \nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k\rangle = \left(\Gamma^{(\alpha)}_{ij,k}\right)_\xi, \tag{2.5}$$

where $g = \langle\,,\,\rangle$ is the Fisher metric. We call this $\nabla^{(\alpha)}$ the α-connection. The α-connection is clearly a symmetric connection. We also have

$$\begin{aligned}\nabla^{(\alpha)} &= (1-\alpha)\,\nabla^{(0)} + \alpha\,\nabla^{(1)}, \\ &= \frac{1+\alpha}{2}\nabla^{(1)} + \frac{1-\alpha}{2}\nabla^{(-1)}.\end{aligned} \tag{2.6}$$

Proposition 2.3.1 *The 0-connection is the Riemannian connection with respect to the Fisher metric.*

In general, when $\alpha \neq 0$, $\nabla^{(\alpha)}$ is not metric.

The notion of exponential family has a close relation to $\nabla^{(1)}$. From the definition of an exponential family given in Equation (2.2), with $\partial_i = \frac{\partial}{\partial\theta_i}$, we obtain

$$\partial_i \ell(x;\theta) = F_i(x) - \partial_i\psi(\theta) \tag{2.7}$$

and

$$\partial_i\partial_j \ell(x;\theta) = -\partial_i\partial_j\psi(\theta). \tag{2.8}$$

where $\ell(x;\theta) = \log f(x;\theta)$.
Hence we have $\Gamma^{(1)}_{ij,k} = -\partial_i\partial_j\psi\, E_\theta[\partial_k\ell_\theta]$, which is 0. In other words, we see that (θ_i) is a 1-affine coordinate system, and Θ is 1-flat.

In particular, the 1-connection is said to be an *exponential connection*, and the (-1)-connection is said to be a *mixture connection*. We say that an α-connection and the $(-\alpha)$-connection are *mutually dual* with respect to the Fisher metric g since the following formula holds:

$$Xg(Y,Z) = g(\nabla^{(\alpha)}_X Y, Z) + g(Y, \nabla^{(-\alpha)}_X Z),$$

where X, Y and Z are arbitrary vector fields on M.

Now, Θ is an exponential family, so a mixture coordinate system is given by a potential function, that is,

$$\eta_i = \frac{\partial\psi}{\partial\theta_i}. \tag{2.9}$$

Since (θ_i) is a 1-affine coordinate system, (η_i) is a (-1)-affine coordinate system, and they are mutually dual with respect to the Fisher metric. Therefore the statistical manifold has dually orthogonal foliations (Section 3.7 in [4]).

The coordinates in (η_i) admit a potential function given by:

$$\lambda = \theta_i\,\eta_i - \psi(\theta). \tag{2.10}$$

2.4 Affine immersions

Let M be an m-dimensional manifold, f an immersion from M to $\mathbb{R}^{m+1}$, and ξ a vector field along f. We can $\forall x \in R^{m+1}$, identify $T_x R^{m+1} \equiv \mathbb{R}^{m+1}$. The pair $\{f, \xi\}$ is said to be an *affine immersion* from M to $\mathbb{R}^{m+1}$ if, for each point $p \in M$, the following formula holds:

$$T_{f(p)} R^{m+1} = f_*(T_p M) \oplus Span\{\xi_p\} \, .$$

We call ξ a transversal vector field.

We denote by D the standard flat affine connection of $\mathbb{R}^{m+1}$. Identifying the covariant derivative along f with D, we have the following decompositions:

$$\begin{aligned} D_X f_* Y &= f_*(\nabla_X Y) + h(X,Y)\xi, \\ D_X \xi &= -f_*(Sh(X)) + \mu(X)\xi. \end{aligned}$$

The induced objects ∇, h, Sh and μ are the induced connection, the affine fundamental form, the affine shape operator and the transversal connection form, respectively. If the affine fundamental form h is positive definite everywhere on M, the immersion f is said to be strictly convex. And if $\mu = 0$, the affine immersion $\{f, \xi\}$ is said to be equiaffine. It is known that a strictly convex equiaffine immersion induces a statistical manifold. Conversely, the condition when a statistical manifold can be realized in an affine space has been studied. We say that an affine immersion $\{f, \xi\} : \Theta \to \mathbb{R}^{m+1}$ is a graph immersion if the hypersurface is a graph of ψ in $\mathbb{R}^{m+1}$:

$$f : M \to \mathbb{R}^{m+1} : \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \theta_m \end{bmatrix} \mapsto \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \theta_m \\ \psi(\theta) \end{bmatrix}, \quad \xi = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

Set $\partial_i = \frac{\partial}{\partial \theta_i}$, $\psi_{ij} = \frac{\partial^2 \psi}{\partial \theta_i \, \partial \theta_j}$. Then we have

$$D_{\partial_i} f_* \partial_j = \psi_{ij} \, \xi.$$

This implies that the induced connection ∇ is flat and (θ_i) is a ∇-affine coordinate system.

Proposition 2.4.1 *Let (M, h, ∇, ∇^*) be a simply connected dually flat space with a global coordinate system and (θ) an affine coordinate system of ∇. Suppose that ψ is a θ-potential function. Then (M, h, ∇) can be realized in $\mathbb{R}^{m+1}$ by a graph immersion whose potential is ψ.*

2.5 Example: Gamma manifold

The family of gamma distributions has event space $\Omega = \mathbb{R}^+$ and probability density functions given by

$$S = \{f(x; \gamma, \mu) | \gamma, \mu \in \mathbb{R}^+\}$$

so here $M \equiv \mathbb{R}^+ \times \mathbb{R}^+$ and the random variable is $x \in \Omega = \mathbb{R}^+$ with

$$f(x; \gamma, \mu) = \left(\frac{\mu}{\gamma}\right)^{\mu} \frac{x^{\mu-1}}{\Gamma(\mu)} \, e^{-x\mu/\gamma} \tag{2.11}$$

Proposition 2.5.1 *Let $\mathcal{G}$ be the gamma manifold. Set $\beta = \frac{\mu}{\gamma}$, Then (β, μ) is a natural coordinate system of the 1-connection and*

$$\psi(\theta) = \log \Gamma(\mu) - \mu \, \log \beta \tag{2.12}$$

is the corresponding potential function.

Proof: Using $\beta = \frac{\mu}{\gamma}$, the logarithm of gamma distributions can be written as

$$\begin{aligned} \log p(x;\beta,\mu) &= \log \beta^{\mu} \frac{x^{\mu-1}}{\Gamma(\mu)} e^{-\beta x} \\ &= -\log x + (\mu \log x - \beta x) - (\log \Gamma(\mu) - \mu \log \beta) \end{aligned} \tag{2.13}$$

Hence the set of all gamma distributions is an exponential family. The coordinates $(\theta_1, \theta_2) = (\beta, \mu)$ is a natural coordinate system, and $\psi(\theta) = \log \Gamma(\mu) - \mu \log \beta$ is its potential function. □

Corollary 2.5.2 *Since $\psi(\theta)$ is a potential function, the Fisher metric is given by the Hessian of ψ, that is, with respect to natural coordinates:*

$$[g_{ij}] = \left[\frac{\partial^2 \psi(\theta)}{\partial \theta_i \partial \theta_j}\right] = \begin{bmatrix} \frac{\mu}{\beta^2} & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \psi''(\mu) \end{bmatrix} \tag{2.14}$$

2.5.1 α-Connection

For each $\alpha \in \mathbb{R}$, the α *(or $\nabla^{(\alpha)}$)-connection* is the torsion-free affine connection with components:

$$\Gamma^{(\alpha)}_{ij,k} = \frac{1-\alpha}{2} \partial_i \, \partial_j \, \partial_k \psi(\theta) \,,$$

where $\psi(\theta)$ is the potential function, and $\partial_i = \frac{\partial}{\partial \theta_i}$.

Since the set of gamma distributions is an exponential family, the connection $\nabla^{(1)}$ is flat. In this case, (β, μ) is a 1-affine coordinate system.

So the 1 and (-1)-connections on the gamma manifold are flat.

Proposition 2.5.3 *The functions $\Gamma^{(\alpha)}_{ij,k}$ are given by*

$$\begin{aligned} \Gamma^{(\alpha)}_{11,1} &= -\frac{(1-\alpha)\,\mu}{\beta^3} \,, \\ \Gamma^{(\alpha)}_{12,1} &= \Gamma^{(\alpha)}_{12,2} = \frac{1-\alpha}{2\,\beta^2} \,, \\ \Gamma^{(\alpha)}_{22,2} &= \frac{(1-\alpha)\,\psi''(\mu)}{2} \end{aligned} \tag{2.15}$$

while the other independent components are zero. □

We have an affine connection $\nabla^{(\alpha)}$ defined by

$$\langle \nabla^{(\alpha)}_{\partial_i} \partial_j, \partial_k \rangle = \Gamma^{(\alpha)}_{ij,k} \,,$$

So by solving the equations

$$\Gamma^{(\alpha)}_{ij,k} = \sum_{h=1}^{2} g_{kh} \, \Gamma^{h(\alpha)}_{ij} \quad , (k = 1, 2).$$

we obtain the components of $\nabla^{(\alpha)}$:

Proposition 2.5.4 *The components $\Gamma^{i(\alpha)}_{jk}$ of the $\nabla^{(\alpha)}$-connection are given by*

$$\begin{aligned}
\Gamma^{(\alpha)1}_{11} &= \frac{(\alpha-1)\,(-1+2\,\mu\,\psi'(1,\mu))}{2\,\beta\,(-1+\mu\,\psi'(\mu))}\,,\\
\Gamma^{(\alpha)1}_{12} &= -\frac{(\alpha-1)\,\psi'(1,\mu)}{-2+2\,\mu\,\psi'(\mu)}\,,\\
\Gamma^{(\alpha)1}_{22} &= -\frac{(\alpha-1)\,\beta\,\psi''(\mu)}{-2+2\,\mu\,\psi'(\mu)}\,,\\
\Gamma^{(\alpha)2}_{11} &= \frac{(\alpha-1)\,\mu}{2\,\beta^2\,(-1+\mu\,\psi'(\mu))}\,,\\
\Gamma^{(\alpha)2}_{12} &= \frac{1-\alpha}{-2\,\beta+2\,\beta\,\mu\,\psi'(\mu)}\,,\\
\Gamma^{(\alpha)2}_{22} &= -\frac{(\alpha-1)\,\mu\,\psi''(\mu)}{-2+2\,\mu\,\psi'(\mu)}\,.
\end{aligned} \tag{2.16}$$

while the other independent components are zero. □

2.5.2 α-Curvatures

Proposition 2.5.5 *Direct calculation gives the α-curvature tensor of $\mathcal{G}$*

$$R^{\alpha}_{1212} = \frac{\left(\alpha^2-1\right)\,(\psi'(\mu)+\mu\,\psi''(\mu))}{4\,\beta^2\,(-1+\mu\,\psi'(\mu))}\,, \tag{2.17}$$

while the other independent components are zero.

By contraction we obtain:

α-Ricci tensor:

$$[R^{(\alpha)}_{ij}] = \left(\alpha^2-1\right)\begin{bmatrix} \frac{-\mu\left(\psi'(\mu)+\mu\,\psi''(\mu)\right)}{4\,\beta^2\,(-1+\mu\,\psi'(\mu))^2} & \frac{\left(\psi'(\mu)+\mu\,\psi''(\mu)\right)}{4\,\beta\,(-1+\mu\,\psi'(\mu))^2} \\ \frac{\left(\psi'(\mu)+\mu\,\psi''(\mu)\right)}{4\,\beta\,(-1+\mu\,\psi'(\mu))^2} & \frac{-\psi'(\mu)\left(\psi'(\mu)+\mu\,\psi''(\mu)\right)}{4\,(-1+\mu\,\psi'(\mu))^2} \end{bmatrix} \tag{2.18}$$

In addition, the eigenvalues and the eigenvectors for the α-Ricci tensor are given by

$$\left(1-\alpha^2\right)\begin{pmatrix} \frac{\left(\mu+\beta^2\,\psi'(\mu)+\sqrt{4\,\beta^2+\mu^2-2\,\beta^2\,\mu\,\psi'(\mu)+\beta^4\,\psi'(\mu)^2}\right)\left(\psi'(\mu)+\mu\,\psi''(\mu)\right)}{8\,\beta^2\,(-1+\mu\,\psi'(\mu))^2} \\ \frac{\left(\mu+\beta^2\,\psi'(\mu)-\sqrt{4\,\beta^2+\mu^2-2\,\beta^2\,\mu\,\psi'(\mu)+\beta^4\,\psi'(\mu)^2}\right)\left(\psi'(\mu)+\mu\,\psi''(\mu)\right)}{8\,\beta^2\,(-1+\mu\,\psi'(\mu))^2} \end{pmatrix} \tag{2.19}$$

$$\begin{pmatrix} \frac{-\left(\mu-\beta^2\,\psi'(\mu)+\sqrt{4\,\beta^2+\mu^2-2\,\beta^2\,\mu\,\psi'(\mu)+\beta^4\,\psi'(\mu)^2}\right)}{2\,\beta} & 1 \\ \frac{-\mu+\beta^2\,\psi'(\mu)+\sqrt{4\,\beta^2+\mu^2-2\,\beta^2\,\mu\,\psi'(\mu)+\beta^4\,\psi'(\mu)^2}}{2\,\beta} & 1 \end{pmatrix} \tag{2.20}$$

α-Scalar curvature:

$$R^{(\alpha)} = \frac{\left(1-\alpha^2\right)\,(\psi'(\mu)+\mu\,\psi''(\mu))}{2\,(-1+\mu\,\psi'(\mu))^2}\,. \tag{2.21}$$

We note that $R^{(\alpha)} \to \frac{\left(1-\alpha^2\right)}{2}$ as $\mu \to 0$. □

2.5.3 Gamma manifold geodesics

The Fisher information metric for the gamma manifold is given in (γ,μ) coordinates by the arc length function

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

The Levi-Civita connection is that given by setting $\alpha = 0$ in the α-connections of the previous section; for this case we give in Figure 2.1 some examples of geodesic sprays in the vicinities of the points

$$(\gamma,\mu) = (1, 0.5),\ (1,1),\ (1,2).$$

2.5.4 Mutually dual foliations

Now, $\mathcal{G}$ represents an exponential family of pdfs, so a mixture coordinate system is given by a potential function. Since (β,μ) is a 1-affine coordinate system, (η_1,η_2) given by

$$\begin{aligned} \eta_1 &= \frac{\partial\psi}{\partial\beta} = -\frac{\mu}{\beta}, \\ \eta_2 &= \frac{\partial\psi}{\partial\mu} = \phi(\mu) - \log\beta. \end{aligned} \tag{2.22}$$

is a (-1)-affine coordinate system, and they are mutually dual with respect to the Fisher metric. Therefore the gamma manifold has dually orthogonal foliations and potential function

$$\lambda = -\mu + \psi(\mu) - \log\Gamma(\mu). \tag{2.23}$$

2.5.5 Affine Immersions for Gamma Manifold

The gamma manifold has an affine immersion in $\mathbb{R}^3$.

Proposition 2.5.6 *Let $\mathcal{G}$ be the gamma manifold with the Fisher metric g and the exponential connection $\nabla^{(1)}$. Denote by (β,μ) a natural coordinate system. Then $\mathcal{G}$ can be realized in $\mathbb{R}^3$ by the graph of a potential function*

$$f : \mathcal{G} \to \mathbb{R}^3 \begin{pmatrix} \beta \\ \mu \end{pmatrix} \mapsto \begin{pmatrix} \beta \\ \mu \\ \log\Gamma(\mu) - \mu\log\beta \end{pmatrix}, \quad \xi = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The submanifold of exponential distributions is represented by the curve

$$(0,\infty) \to \mathbb{R}^3 : \beta \mapsto \{\beta, 1, \log\frac{1}{\beta}\}$$

and a tubular neighbourhood of this curve will contain all immersions for small enough perturbations of exponential distributions. □

2.6 Example: Log-gamma distributions

The family of probability density functions for random variable $N \in [0,1]$ given by

$$g(N,\mu,\beta) = \frac{\frac{1}{N}^{1-\frac{\beta}{\mu}} (\frac{\beta}{\mu})^{\beta} (\log\frac{1}{N})^{\beta-1}}{\Gamma(\beta)} \quad \text{for } \mu > 0 \text{ and } \beta > 0\ . \tag{2.24}$$

Some of these pdfs with central mean are shown in Figure 2.2.

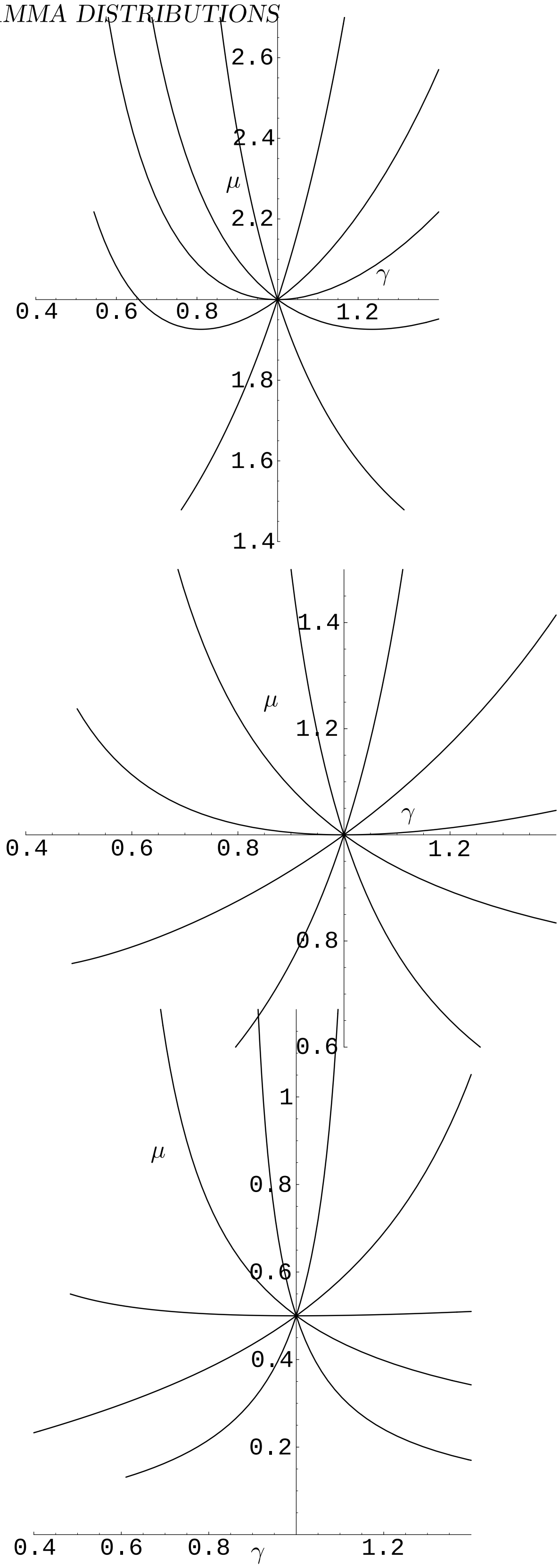

Figure 2.1: *Geodesic sprays in the gamma manifold, radiating from the points with unit mean* $\gamma = 1$, *and* $\mu = 0.5,\ 1,\ 2$ *increasing vertically. The case* $\mu = 1$ *corresponds to an exponential distribution from an underlying Poisson process;* $\mu < 1$ *corresponds to clustering and* μ *increasing above* 1 *corresponds to the opposite process, dispersion leading to greater uniformity.*

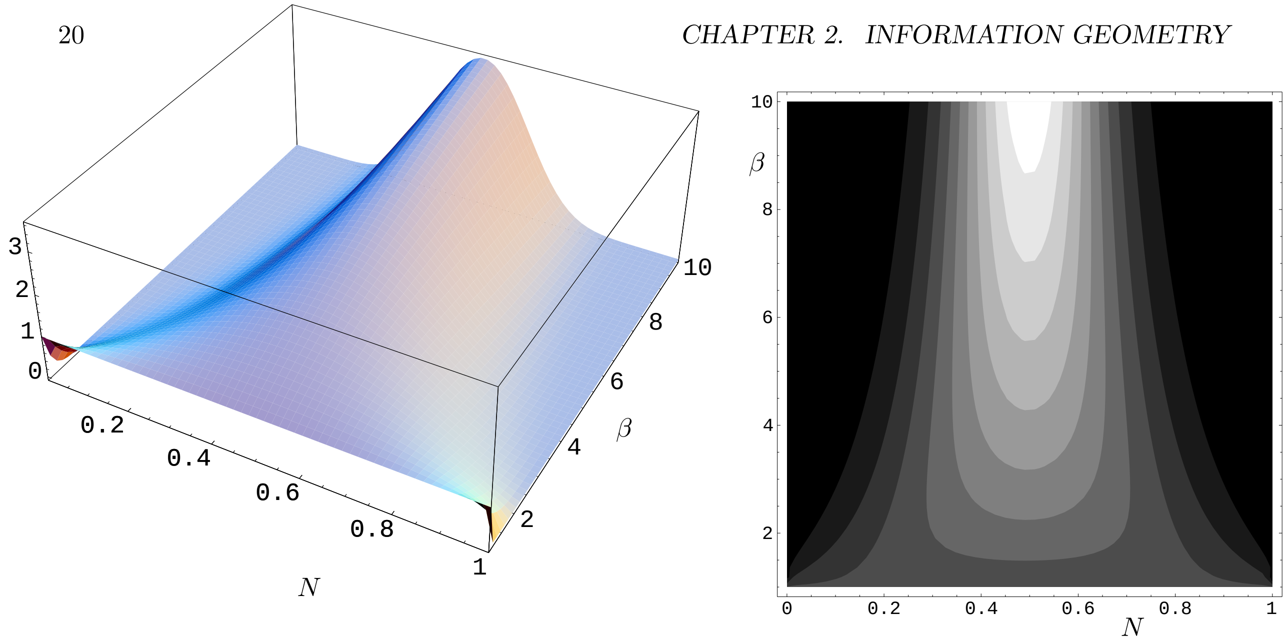

Figure 2.2: *The log-gamma family of densities with central mean* $< N >= \frac{1}{2}$ *as a surface and as a contour plot for* $\beta \geq 1$.

Proposition 2.6.1 *The log-gamma family (2.24) with information metric determines a Riemannian 2-manifold* $\mathcal{L}$ *with the following properties*
- *it contains the uniform distribution*
- *it contains approximations to truncated Gaussian distributions*
- *it is an isometric isomorph of a the manifold* $\mathcal{G}$ *of gamma distributions.*

Proof
By integration, it is easily checked that the family given by equation (2.24) consists of probability density functions for the random variable $N \in [0,1]$. The limiting densities are given by

$$\lim_{\beta \to 1^+} g(N,\mu,\beta) = g(N,\mu,1) = \frac{1}{\mu}\left(\frac{1}{N}\right)^{1-\frac{1}{\mu}} \tag{2.25}$$

$$\lim_{\mu \to 1} g(N,\mu,1) = g(N,1,1) = 1\,. \tag{2.26}$$

The mean, $< N >$, standard deviation σ_N, and coefficient of variation cv_N, of N are given by

$$< N > = \left(\frac{\beta}{\beta+\mu}\right)^{\beta} \tag{2.27}$$

$$\sigma_N = \sqrt{\left(\frac{\beta}{\beta+2\,\mu}\right)^{\beta} - \left(\frac{\beta}{\beta+\mu}\right)^{2\,\beta}} \tag{2.28}$$

$$cv_N = \frac{\sigma_N}{< N >} = \sqrt{\left(\frac{\beta}{\beta+2\,\mu}\right)^{\beta}\left(\frac{\beta+\mu}{\beta}\right)^{2\,\beta} - 1}. \tag{2.29}$$

We can obtain the family of densities having central mean in $[0,1]$, by solving $< N >= \frac{1}{2}$, which corresponds to the locus $\mu = \beta(2^{1/\beta} - 1)$; some of these are shown in Figure 2.2 and Figure 2.3. Evidently, the distributions with central mean and large β provide approximations to Gaussian distributions truncated on $[0,1]$.

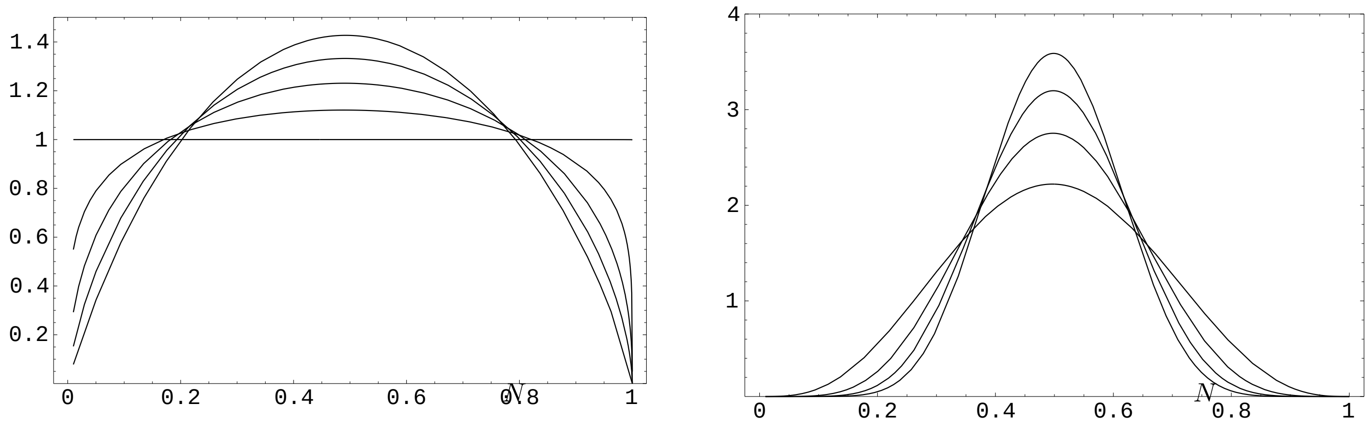

Figure 2.3: *Examples from the log-gamma family of probability densities with central mean* $< N >= \frac{1}{2}$. *Left:* $\beta = 1, 1.2, 1.4, 1.6, 1.8$. *Right:* $\beta = 4, 6, 8, 10$.

For the log-gamma densities [22, 28], the Fisher information metric [3] on the parameter space $\Theta = \{(\mu, \beta) \in (0, \infty) \times [1, \infty)\}$ is given by its arc length function

$$ds^2_{\mathcal{L}} = \sum_{ij} g_{ij}\, dx^i dx^j = \frac{\beta}{\mu^2}\, d\mu^2 + \left((\frac{\Gamma'(\beta)}{\Gamma(\beta)})' - \frac{1}{\beta} \right) d\beta^2, \tag{2.30}$$

In fact, (2.24) arises from the gamma family

$$f(x, \mu, \beta) = \frac{x^{\beta-1}\, (\frac{\beta}{\mu})^{\beta}}{\Gamma(\beta)}\, e^{-\frac{x\,\beta}{\mu}} \tag{2.31}$$

for the non-negative random variable $x = \log \frac{1}{N}$. It is known that the gamma family (3.11) has also the information metric (2.30) so the identity map on the space of coordinates (μ, β) is an isometry of Riemannian manifolds. □

2.7 Example: Gaussian distributions

The family of univariate normal or Gaussian distributions has event space $\Omega = \mathbb{R}$ and probability density functions given by

$$\mathcal{N} \equiv \{N(\mu, \sigma^2)\} = \{p(x; \mu, \sigma) |\, \mu \in \mathbb{R},\, \sigma \in \mathbb{R}^+\}$$

with mean μ and variance σ^2. So here $\mathcal{N} = \mathbb{R} \times \mathbb{R}^+$ is the upper half -plane, and the random variable is $x \in \Omega = \mathbb{R}$ with

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\,\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\,\sigma^2}} \tag{2.32}$$

The mean μ and standard deviation σ are frequently used as a local coordinate system $\xi = (\xi_1, \xi_2) = (\mu, \sigma)$.

Shannon's information theoretic entropy is given by:

$$S_{\mathcal{N}}(\mu,\sigma) = -\int_{-\infty}^{\infty} \log(p(t;\mu,\sigma))\, p(t;\mu,\sigma)\, dt = \frac{1}{2}\,(1+\log(2\,\pi)) + \log(\sigma) \tag{2.33}$$

At unit variance the entropy is $S_{\mathcal{N}} = \frac{1}{2}\,(1+\log(2\,\pi))$.

2.7.1 Natural coordinate system and potential function

Proposition 2.7.1 *Let $\mathcal{N}$ be the normal manifold. Set $\theta_1 = \frac{\mu}{\sigma^2}$ and $\theta_2 = -\frac{1}{2\,\sigma^2}$. Then $(\frac{\mu}{\sigma^2}, -\frac{1}{2\,\sigma^2})$ is a natural coordinate system and*

$$\psi = -\frac{{\theta_1}^2}{4\,\theta_2} + \frac{1}{2}\,\log(-\frac{\pi}{\theta_2}) = \frac{\mu^2}{2\,\sigma^2} + \log(\sqrt{2\,\pi}\,\sigma) \tag{2.34}$$

is the corresponding potential function.

Proof: Set $\theta_1 = \frac{\mu}{\sigma^2}$ and $\theta_2 = -\frac{1}{2\,\sigma^2}$. Then the logarithm of univariate normal distributions can be written as

$$\begin{aligned} \log p(x;\theta_1,\theta_2) &= \log e^{\left(\frac{\mu}{\sigma^2}\right)x + \left(\frac{-1}{2\,\sigma^2}\right)x^2 - \left(\frac{\mu^2}{2\,\sigma^2} + \log(\sqrt{2\,\pi}\,\sigma)\right)} \\ &= \theta_1\,x + \theta_2\,x^2 - \left(-\frac{{\theta_1}^2}{4\,\theta_2} + \frac{1}{2}\,\log(-\frac{\pi}{\theta_2})\right) \end{aligned} \tag{2.35}$$

Hence the set of all univariate normal distributions is an exponential family. The coordinates (θ_1,θ_2) is a natural coordinate system, and $\psi = -\frac{{\theta_1}^2}{4\,\theta_2} + \frac{1}{2}\,\log(-\frac{\pi}{\theta_2}) = \frac{\mu^2}{2\,\sigma^2} + \log(\sqrt{2\,\pi}\,\sigma)$ is its potential function. □

2.7.2 Fisher information metric

Proposition 2.7.2 *The Fisher metric with respect to natural coordinates (θ_1,θ_2) is given by:*

$$[g_{ij}] = \begin{bmatrix} \frac{-1}{2\,\theta_2} & \frac{\theta_1}{2\,{\theta_2}^2} \\ \frac{\theta_1}{2\,{\theta_2}^2} & \frac{\theta_2 - {\theta_1}^2}{2\,{\theta_2}^3} \end{bmatrix} = \begin{bmatrix} \sigma^2 & 2\,\mu\,\sigma^2 \\ 2\,\mu\,\sigma^2 & 2\,\sigma^2\,\left(2\,\mu^2+\sigma^2\right) \end{bmatrix} \tag{2.36}$$

Proof: Since ψ is a potential function, the Fisher metric is given by the Hessian of ψ, that is,

$$g_{ij} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}. \tag{2.37}$$

Then, we have the Fisher metric by a straightforward calculation. □

2.7.3 Mutually dual foliations

Since $\mathcal{N}$ represents an exponential family, a mixture coordinate system is given by a potential function. We have

$$\begin{aligned} \eta_1 &= \frac{\partial\psi}{\partial\theta^1} = \frac{-\theta_1}{2\,\theta_2} = \mu, \\ \eta_2 &= \frac{\partial\psi}{\partial\theta^2} = \frac{{\theta_1}^2 - 2\,\theta_2}{4\,{\theta_2}^2} = \mu^2 + \sigma^2\,. \end{aligned} \tag{2.38}$$

Since (θ_1, θ_2) is a 1-affine coordinate system, (η_1, η_2) is a (-1)-affine coordinate system, and they are mutually dual with respect to the Fisher metric. Therefore the normal manifold has dually orthogonal foliations. The coordinates in (2.38) admit the potential function

$$\lambda = -\frac{1}{2}\left(1 + \log(-\frac{\pi}{\theta_2})\right) = \frac{-1}{2}\left(1 + \log(2\,\pi) + 2\,\log(\sigma)\right)\,. \tag{2.39}$$

2.7.4 Affine immersions for Gaussian manifold

We show that the normal manifold can be realized in Euclidean $\mathbb{R}^3$ by an affine immersion.

Proposition 2.7.3 *Let $\mathcal{N}$ be the normal manifold with the Fisher metric g and the exponential connection $\nabla^{(1)}$. Denote by (θ_1, θ_2) a natural coordinate system. Then M can be realized in $\mathbb{R}^3$ by the graph of a potential function, namely, $\mathcal{G}$ can be realized by the affine immersion $\{f, \xi\}$:*

$$f : \mathcal{N} \to \mathbb{R}^3 : \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \mapsto \begin{pmatrix} \theta_1 \\ \theta_2 \\ \psi \end{pmatrix}, \quad \xi = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

where ψ is the potential function $\psi = -\frac{{\theta_1}^2}{4\,\theta_2} + \frac{1}{2}\,\log(-\frac{\pi}{\theta_2})\,.$

The submanifold of univariate normal distributions with zero mean (i.e. $\theta_1 = 0$) is represented by the curve

$$(-\infty, 0) \to \mathbb{R}^3 : \theta_2 \mapsto \{0, \theta_2, \frac{1}{2}\,\log(-\frac{\pi}{\theta_2})\}\,,$$

In addition, the submanifold of univariate normal distributions with unit variance (i.e. $\theta_2 = -\frac{1}{2}$) is represented by the curve

$$\mathbb{R} \to \mathbb{R}^3 : \theta_1 \mapsto \{\theta_1, -\frac{1}{2}, \frac{{\theta_1}^2}{2} + \frac{1}{2}\,\log(2\,\pi)\}.$$

□

2.8 Example: Bivariate Gaussian 5-manifold

Here we collect the results from Yoshiharu Sato, Kazuaki Sugawa and Michiaki Kawaguchi [66], for comparison with bivariate gamma manifolds.

The probability density function for the two-dimensional Gaussian distribution has the form:

$$f(x, y) = \frac{1}{2\,\pi\,\sqrt{\sigma_1\,\sigma_2 - {\sigma_{12}}^2}}\, e^{-\frac{1}{2\,(\sigma_1\,\sigma_2 - {\sigma_{12}}^2)}\left(\sigma_2 (x-\mu_1)^2 - 2\,\sigma_{12}\,(x-\mu_1)\,(y-\mu_2) + \sigma_1 (y-\mu_2)^2\right)} \tag{2.40}$$

where

$$-\infty < x_1, x_2 < \infty, \quad -\infty < \mu_1, \mu_2 < \infty, \quad 0 < \sigma_1, \sigma_2 < \infty.$$

This contains the five parameters $(\mu_1, \mu_2, \sigma_1, \sigma_{12}, \sigma_2) \in \mathbb{R}^5$. It is easy to show that the marginal distributions are Gaussian with parameters (μ_1, σ_1) and (μ_2, σ_2); the covariance is σ_{12}.

2.8.1 Fisher information metric

The information geometry of 2.40 has been studied by Sato et al. [66]; the metric tensor takes the following form:

$$(2.41)\qquad G = [g_{ij}] = \begin{bmatrix} \frac{\sigma_2}{\triangle} & -\frac{\sigma_{12}}{\triangle} & 0 & 0 & 0 \\ -\frac{\sigma_{12}}{\triangle} & \frac{\sigma_1}{\triangle} & 0 & 0 & 0 \\ 0 & 0 & \frac{(\sigma_2)^2}{2\triangle^2} & -\frac{\sigma_{12}\,\sigma_2}{\triangle^2} & \frac{(\sigma_{12})^2}{2\triangle^2} \\ 0 & 0 & -\frac{\sigma_{12}\,\sigma_2}{\triangle^2} & \frac{\sigma_1\,\sigma_2+(\sigma_{12})^2}{\triangle^2} & -\frac{\sigma_1\,\sigma_{12}}{\triangle^2} \\ 0 & 0 & \frac{(\sigma_{12})^2}{2\triangle^2} & -\frac{\sigma_1\,\sigma_{12}}{\triangle^2} & \frac{(\sigma_1)^2}{2\triangle^2} \end{bmatrix},$$

where $\triangle$ is the determinant

$$\triangle = \sigma_1\,\sigma_2 - (\sigma_{12})^2$$

The inverse $[g^{ij}]$ of the fundamental tensor $[g_{ij}]$ defined by the relation

$$g_{ij}g^{ik} = \delta_j^k$$

is given by

$$(2.42)\qquad G^{-1} = [g^{ij}] = \begin{bmatrix} \sigma_1 & \sigma_{12} & 0 & 0 & 0 \\ \sigma_{12} & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & 2\,(\sigma_1)^2 & 2\,\sigma_1\,\sigma_{12} & 2\,(\sigma_{12})^2 \\ 0 & 0 & 2\,\sigma_1\,\sigma_{12} & \sigma_1\,\sigma_2+(\sigma_{12})^2 & 2\,\sigma_{12}\,\sigma_2 \\ 0 & 0 & 2\,(\sigma_{12})^2 & 2\,\sigma_{12}\,\sigma_2 & 2\,(\sigma_2)^2 \end{bmatrix}.$$

2.9 Example: Freund distributions

Freund [35] introduced a bivariate exponential mixture distribution arising from the following reliability considerations. Suppose that an instrument has two components A and B with lifetimes X and Y respectively having density functions (when both components are in operation)

$$\begin{aligned} f_X(x) &= \alpha_1\, e^{-\alpha_1 x} \\ f_Y(y) &= \alpha_2\, e^{-\alpha_2 y} \end{aligned}$$

for $(\alpha_1, \alpha_2 > 0; x, y > 0)$. Then X and Y are dependent in that a failure of either component changes the parameter of the life distribution of the other component. Thus when A fails, the parameter for Y becomes β_2; when B fails, the parameter for X becomes β_1. There is no other dependence. Hence the joint density function of X and Y is

$$(2.43)\qquad f(x,y) = \begin{cases} \alpha_1\beta_2 e^{-\beta_2 y-(\alpha_1+\alpha_2-\beta_2)x} & \text{for } 0 < x < y, \\ \alpha_2\beta_1 e^{-\beta_1 x-(\alpha_1+\alpha_2-\beta_1)y} & \text{for } 0 < y < x \end{cases}$$

where $\alpha_i, \beta_i > 0 \quad (i = 1, 2)$.

Provided that $\alpha_1 + \alpha_2 \neq \beta_1$, the marginal density function of X is

$$(2.44)\quad f_X(x) = \left(\frac{\alpha_2}{\alpha_1+\alpha_2-\beta_1}\right)\beta_1\, e^{-\beta_1 x} + \left(\frac{\alpha_1-\beta_1}{\alpha_1+\alpha_2-\beta_1}\right)(\alpha_1+\alpha_2)\, e^{-(\alpha_1+\alpha_2)x},\ x \geq 0$$

and provided that $\alpha_1 + \alpha_2 \neq \beta_2$, The marginal density function of Y is

$$(2.45)\quad f_Y(y) = \left(\frac{\alpha_1}{\alpha_1+\alpha_2-\beta_2}\right)\beta_2\, e^{-\beta_2 y} + \left(\frac{\alpha_2-\beta_2}{\alpha_1+\alpha_2-\beta_2}\right)(\alpha_1+\alpha_2)\, e^{-(\alpha_1+\alpha_2)y},\ y \geq 0$$

We can see that the marginal density functions are not exponential but rather mixtures of exponential distributions if $\alpha_i > \beta_i$; otherwise, they are weighted averages. For this reason, this system of distributions should be termed bivariate mixture exponential distributions rather than simply bivariate exponential distributions. The marginal density functions $f_X(x)$ and $f_Y(y)$ are exponential distributions only in the special case $\alpha_i = \beta_i \quad (i = 1, 2)$.

Freund discussed the statistics of the special case when $\alpha_1 + \alpha_2 = \beta_1 = \beta_2$, and obtained the joint density function as:

$$f(x, y) = \begin{cases} \alpha_1(\alpha_1 + \alpha_2)e^{-(\alpha_1+\alpha_2)y} & \text{for } 0 < x < y, \\ \alpha_2(\alpha_1 + \alpha_2)e^{-(\alpha_1+\alpha_2)x} & \text{for } 0 < y < x \end{cases} \tag{2.46}$$

with marginal density functions:

$$f_X(x) = (\alpha_1 + \alpha_2(\alpha_1 + \alpha_2)x)\ e^{-(\alpha_1+\alpha_2)x}\ x \geq 0\,, \tag{2.47}$$

$$f_Y(y) = (\alpha_2 + \alpha_1(\alpha_1 + \alpha_2)y)\ e^{-(\alpha_1+\alpha_2)y}\ y \geq 0 \tag{2.48}$$

The covariance and correlation coefficient of X and Y are

$$Cov(X, Y) = \frac{\beta_1\,\beta_2 - \alpha_1\,\alpha_2}{\beta_1\,\beta_2\,(\alpha_1 + \alpha_2)^2}\,, \tag{2.49}$$

$$\rho(X, Y) = \frac{\beta_1\,\beta_2 - \alpha_1\,\alpha_2}{\sqrt{{\alpha_2}^2 + 2\,\alpha_1\,\alpha_2 + {\beta_1}^2}\,\sqrt{{\alpha_1}^2 + 2\,\alpha_1\,\alpha_2 + {\beta_2}^2}} \tag{2.50}$$

Note that $-\frac{1}{3} < \rho(X, Y) < 1$. The correlation coefficient $\rho(X, Y) \to 1$ when $\beta_1, \beta_2 \to \infty$, and $\rho(X, Y) \to -\frac{1}{3}$ when $\alpha_1 = \alpha_2$ and $\beta_1, \beta_2 \to 0$. In many applications, $\beta_i > \alpha_i \quad (i = 1, 2)$ (i.e., lifetime tends to be shorter when the other component is out of action) ; in such cases the correlation is positive.

2.9.1 Fisher information metric

Proposition 2.9.1 *Let $\mathcal{F}$ be the set of Freund bivariate mixture exponential distributions, (2.43). Then it becomes a 4-manifold with Fisher information metric*

$$g_{ij} = \int_0^\infty \int_0^\infty \frac{\partial^2 \log f(x, y)}{\partial x_i \partial x_j}\ f(x, y)\ dx\ dy$$

and $(x_1, x_2, x_3, x_4) = (\alpha_1, \beta_1, \alpha_2, \beta_2)$.
is given by

$$[g_{ij}] = \int_0^\infty \int_0^\infty \frac{\partial^2 \log f(x, y)}{\partial x_i \partial x_j}\ f(x, y)\ dx\ dy$$

$$\begin{bmatrix} \frac{1}{{\alpha_1}^2+\alpha_1\,\alpha_2} & 0 & 0 & 0 \\ 0 & \frac{\alpha_2}{{\beta_1}^2\,(\alpha_1+\alpha_2)} & 0 & 0 \\ 0 & 0 & \frac{1}{{\alpha_2}^2+\alpha_1\,\alpha_2} & 0 \\ 0 & 0 & 0 & \frac{\alpha_1}{{\beta_2}^2\,(\alpha_1+\alpha_2)} \end{bmatrix} \tag{2.51}$$

The inverse $[g^{ij}]$ of $[g_{ij}]$ is given by

$$[g^{ij}] = \begin{bmatrix} {\alpha_1}^2 + \alpha_1\,\alpha_2 & 0 & 0 & 0 \\ 0 & \frac{{\beta_1}^2\,(\alpha_1+\alpha_2)}{\alpha_2} & 0 & 0 \\ 0 & 0 & {\alpha_2}^2 + \alpha_1\,\alpha_2 & 0 \\ 0 & 0 & 0 & \frac{{\beta_2}^2\,(\alpha_1+\alpha_2)}{\alpha_1} \end{bmatrix}. \tag{2.52}$$

□

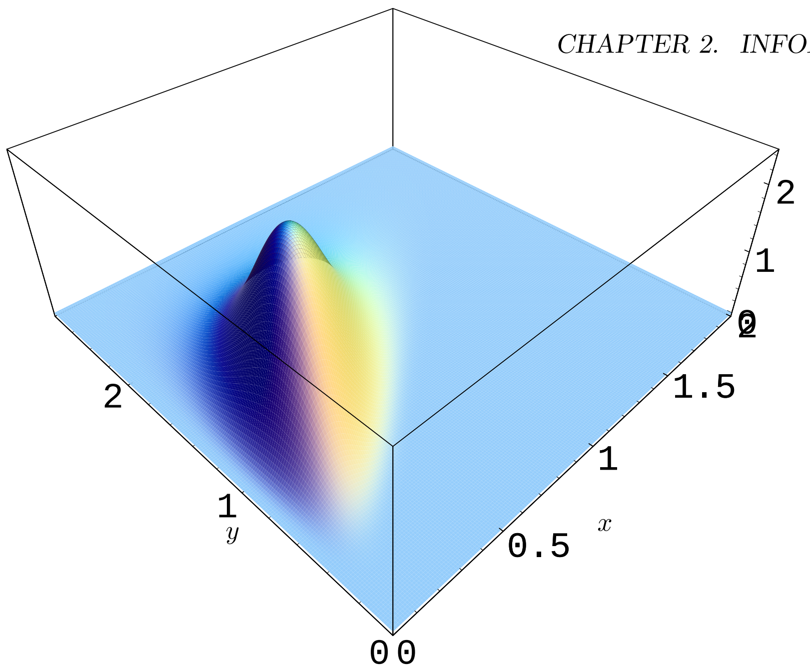

Figure 2.4: *Part of the family of McKay bivariate gamma pdfs; Observe that the McKay pdf is zero outside the octant* $0 < x < y < \infty$*. Here the correlation coefficient has been set to* $\rho_{xy} = 0.6$ *and* $\alpha_1 = 5$.

2.10 Example: Mckay bivariate gamma 3-manifold

The information geometry of the 3-manifold of McKay bivariate gamma distributions can provide a metrization of departures from randomness and departures from independence for bivariate processes. The curvature objects are derived, including those on three submanifolds. As in the case of bivariate normal manifolds, we have negative scalar curvature but here it is not constant and we show how it depends on correlation. These results have potential applications, for example, in the characterization of stochastic materials.

Fisher information metric

The classical family of Mckay bivariate gamma distributions M, is defined on $0 < x < y < \infty$ with parameters $\alpha_1, \sigma_{12}, \alpha_2 > 0$ and pdfs

$$f(x, y; \alpha_1, \sigma_{12}, \alpha_2) = \frac{\left(\frac{\alpha_1}{\sigma_{12}}\right)^{\frac{(\alpha_1+\alpha_2)}{2}} x^{\alpha_1-1}(y-x)^{\alpha_2-1} e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}}y}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} . \tag{2.53}$$

Here σ_{12} is the covariance of X and Y. One way to view this is that $f(x, y)$ is the probability density for the two random variables X and $Y = X + Z$ where X and Z both have gamma distributions.

The correlation coefficient, and marginal functions, of X and Y are given by

$$\rho(X, Y) = \sqrt{\frac{\alpha_1}{\alpha_1 + \alpha_2}} > 0 \tag{2.54}$$

$$f_X(x) = \frac{\left(\frac{\alpha_1}{\sigma_{12}}\right)^{\frac{\alpha_1}{2}} x^{\alpha_1-1} e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}}x}}{\Gamma(\alpha_1)}, \quad x > 0 \tag{2.55}$$

$$f_Y(y) = \frac{\left(\frac{\alpha_1}{\sigma_{12}}\right)^{\frac{(\alpha_1+\alpha_2)}{2}} y^{(\alpha_1+\alpha_2)-1} e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}}y}}{\Gamma(\alpha_1+\alpha_2)}, \quad y > 0 \tag{2.56}$$

The marginal distributions of X and Y are gamma with shape parameters α_1 and $\alpha_1+\alpha_2$, respectively; note that it is not possible to choose parameters such that both marginal functions are exponential.

Proposition 2.10.1 *Let M be the family of Mckay bivariate gamma distributions, then $(\alpha_1, \sigma_{12}, \alpha_2)$ is a local coordinate system, and M becomes a 3-manifold. Fisher information metric*

$$[g_{ij}] = \begin{bmatrix} \frac{-3\,\alpha_1+\alpha_2}{4\,\alpha_1{}^2} + (\frac{\Gamma'(\alpha_1)}{\Gamma(\alpha_1)})' & \frac{\alpha_1-\alpha_2}{4\,\alpha_1\,\sigma_{12}} & -\frac{1}{2\,\alpha_1} \\ \frac{\alpha_1-\alpha_2}{4\,\alpha_1\,\sigma_{12}} & \frac{\alpha_1+\alpha_2}{4\,\sigma_{12}{}^2} & \frac{1}{2\,\sigma_{12}} \\ -\frac{1}{2\,\alpha_1} & \frac{1}{2\,\sigma_{12}} & (\frac{\Gamma'(\alpha_2)}{\Gamma(\alpha_2)})' \end{bmatrix} \tag{2.57}$$

The inverse $[g^{ij}]$ of $[g_{ij}]$ is given by:

$$\begin{aligned} g^{11} &= -\left(\frac{-1+(\alpha_1+\alpha_2)\;\psi'(\alpha_2)}{\psi'(\alpha_2)+\psi'(\alpha_1)\;(1-(\alpha_1+\alpha_2)\;\psi'(\alpha_2))}\right), \\ g^{12} &= g^{21} = \frac{\sigma_{12}\;(1+(\alpha_1-\alpha_2)\;\psi'(\alpha_2))}{\alpha_1\;(\psi'(\alpha_2)+\psi'(\alpha_1)\;(1-(\alpha_1+\alpha_2)\;\psi'(\alpha_2)))}, \\ g^{13} &= g^{31} = \frac{1}{-\psi'(\alpha_2)+\psi'(\alpha_1)\;(-1+(\alpha_1+\alpha_2)\;\psi'(\alpha_2))}, \\ g^{22} &= \frac{\sigma_{12}{}^2\;\left(-1+\left(-3\,\alpha_1+\alpha_2+4\,\alpha_1{}^2\,\psi'(\alpha_1)\right)\;\psi'(\alpha_2)\right)}{\alpha_1{}^2\;(-\psi'(\alpha_2)+\psi'(\alpha_1)\;(-1+(\alpha_1+\alpha_2)\;\psi'(\alpha_2)))}, \\ g^{23} &= g^{32} = \frac{\sigma_{12}\;(-1+2\,\alpha_1\,\psi'(\alpha_1))}{\alpha_1\;(\psi'(\alpha_2)+\psi'(\alpha_1)\;(1-(\alpha_1+\alpha_2)\;\psi'(\alpha_2)))}, \\ g^{33} &= -\left(\frac{-1+(\alpha_1+\alpha_2)\;\psi'(\alpha_1)}{\psi'(\alpha_2)+\psi'(\alpha_1)\;(1-(\alpha_1+\alpha_2)\;\psi'(\alpha_2))}\right), \end{aligned} \tag{2.58}$$

where we have abbreviated $\psi'(\alpha_1) = (\frac{\Gamma'(\alpha_1)}{\Gamma(\alpha_1)})'$. □

Chapter 3

Universal connections and curvature

It is common to have to consider a number of linear connections on a given statistical manifold and so it is important to know the corresponding universal connection and curvature; then all linear connections and their curvatures are pullbacks. An important class of statistical manifolds is that arising from the so-called exponential families [3, 4] and one particular family is that of gamma distributions, which we showed recently [7] to have important uniqueness properties in stochastic processes. Here we describe the system of all linear connections on the manifold of exponential families, using the tangent bundle or the frame bundle to give the system space. Moreover, we provide formulae for the universal connections and curvatures and give an explicit example for the manifold of gamma distributions.

3.1 Systems of connections and universal objects

The concept of system (or structure) of connections was introduced by Mangiarotti and Modugno [50, 57], they were concerned with finite-dimensional bundle representations of the space of all connections on a fibred manifold. On each system of connections there exists a unique universal connection of which every connection in the family of connections is a pullback. A similar relation holds between the corresponding universal curvature and the curvatures of the connections of the system.

Definition 3.1.1 ***A system of connections*** *on a fibred manifold* $p : E \longrightarrow M$ *is a fibred manifold* $p_c : C \longrightarrow M$ *together with a first jet-valued fibred morphism*

$$\xi : C \times_M E \longrightarrow JE$$

over M*, such that each section* $\tilde{\Gamma} : M \longrightarrow C$ *determines a unique connection* $\Gamma = \xi \circ (\tilde{\Gamma} \circ p, I_E)$ *on* E*. Then* C *is the* ***space of connections*** *of the system.*

In the sequel we are interested in the system of linear connections on a Riemannian manifold. The system of all linear connections is the subject of studies in eg. [14, 15, 17, 24].

Theorem 3.1.2 ([50, 57]) *Let* (C, ξ) *be a system of connections on a fibred manifold* $p : E \longrightarrow M$*. Then there is a unique connection form* $\Lambda : C \times_M E \to J(C \times_M E)$ *On the fibred manifold* $\pi_1 : C \times_M E \to C$ *which has the coordinate expression*

$$\Lambda = dx^\lambda \otimes \partial_\lambda + dc^a \otimes \partial_a + \xi^i_\lambda \, dx^\lambda \otimes \partial_i.$$

*This Λ is called the **universal connection** because it describes all the connections of the system.*

Explicitly, each $\tilde{\Gamma} \in Sec(C/M)$ gives an injection $(\tilde{\Gamma} \circ p, I_E)$, of E into $C \times E$, which is a section of π_1 and Γ coincides with the restriction of Λ to this section:

$$\Lambda_{|(\tilde{\Gamma} \circ p, I_E)E} = \Gamma.$$

*A similar relation holds between its curvature Ω, called **universal curvature**, and the curvatures of the connections of the system.*

$$\Omega = \frac{1}{2}\,[\Lambda, \Lambda] = d_\Lambda \Lambda : C \times_M E \to \wedge^2(T^*C) \otimes_E V(E).$$

So the universal curvature Ω has the coordinate expression:

$$\Omega = \frac{1}{2}\left(\xi^j_\lambda\,\partial_j \xi^i_\mu\,dx^\lambda \wedge dx^\mu + 2\,\partial_a \xi^i_\mu\,dx^a \wedge dx^\mu\right) \otimes \partial_i\,.$$

3.2 Exponential family of probability density functions on $\mathbb{R}$

From the definition of an exponential family, and putting $\partial_i = \frac{\partial}{\partial x_i}$, we may obtain

$$\partial_i \ell(y;x) \quad = \quad F_i(y) - \partial_i \psi(x) \tag{3.1}$$

and

$$\partial_i \partial_j \ell(y;x) \quad = \quad -\partial_i \partial_j \psi(x)\,. \tag{3.2}$$

where $\ell(y;x) = \log p(y;x)$.
Hence the Fisher information metric [3, 4] on the n-dimensional space of parameters $\Theta \subset \mathbb{R}^n$, has coordinates:

$$g = [g_{ij}] = -E[\partial_i \partial_j \ell(y;x)] = \partial_i \partial_j \psi(x) = \psi_{ij}(x)\,, \tag{3.3}$$

The Levi-Civita connection with respect to g is given by:

$$\Gamma^k_{ij}(x) = \sum_{h=1}^{n} \frac{1}{2}\,g^{kh}\,\left(\partial_i g_{jh} + \partial_j g_{ih} - \partial_h g_{ij}\right) = \sum_{h=1}^{n} \frac{1}{2}\,g^{kh}\,\partial_i \partial_j \partial_h \psi(x) = \sum_{h=1}^{n} \frac{1}{2}\,\psi^{kh}(x)\,\psi_{ijh}(x)\,. \tag{3.4}$$

where $(\psi^{hk}(x))$ represents the dual metric to $(\psi_{hk}(x))$.

3.3 Universal connection and curvature for exponential families

In this section we provide explicit formulae for universal connection and curvature for an arbitrary statistical n-dimensional exponential manifold. We do this in the two available ways: using the tangent bundle and using the frame bundle systems of all linear connections.

3.3.1 Tangent bundle system: $C_T \times TM \to JTM$

The system of all linear connections on a manifold M has a representation on the tangent bundle

$$E = TM \to M$$

with system space

$$C_T = \{\alpha \otimes j\gamma \in T^*M \otimes_M JTM \mid j\gamma : TM \to TTM \text{ projects onto } I_{TM}\}$$

Here we view I_{TM} as a section of $T^*M \otimes TM$, which is a subbundle of $T^*M \otimes TTM$, with local expression $dx^\lambda \otimes \partial_\lambda$.

The fibred morphism for this system is given by

$$\begin{aligned} \xi_T : C_T \times_M TM &\to JTM \subset T^*M \otimes_{TM} TTM\,, && (3.5)\\ (\alpha \otimes j\gamma, \nu) &\longmapsto \alpha(\nu) j\gamma. && (3.6) \end{aligned}$$

In coordinates

$$\begin{aligned} \xi_T &= dx^\lambda \otimes (\partial_\lambda - \gamma^i_\lambda\, \partial_i)\\ &= dx^\lambda \otimes (\partial_\lambda - y^j\, \Gamma^i_{j\lambda}\, \partial_i)\\ &= dx^\lambda \otimes (\partial_\lambda - y^j\, (\sum_{h=1}^{n} \frac{1}{2}\, \psi^{ih}\, \psi_{j\lambda h})\, \partial_i) && (3.7) \end{aligned}$$

Each section of $C_T \to M$, such as $\tilde{\Gamma} : M \to C_T : (x^\lambda) \to (x^\lambda, \gamma_{\mu\vartheta})$; determines the unique linear connection $\Gamma = \xi_T \circ (\tilde{\Gamma} \circ \pi_T, I_{TM})$ with Christoffel symbols $\Gamma^\lambda_{\mu\vartheta}$.

On the fibred manifold $\pi_1 : C_T \times_M TM \to C_T$; the universal connection is given by:

$$\begin{aligned} \Lambda_T : C_T \times_M TM &\to J(C_T \times_M TM) \subset T^*C_T \otimes T(C_T \times_M TM)\,,\\ (x^\lambda, v^\lambda_{\mu\nu}, y^\lambda) &\longmapsto [(X^\lambda, V^\lambda_{\mu\nu}) \to (X^\lambda, V^\lambda_{\mu\nu}, Y^\mu V^\lambda_{\mu\nu} X^\nu)]. \end{aligned}$$

briefly,

$$\begin{aligned} \Lambda_T &= dx^\lambda \otimes \partial_\lambda + dv^a \otimes \partial_a + y^\mu v^i_{\mu\nu}\, dx^\nu \otimes \partial_i\\ &= dx^\lambda \otimes \partial_\lambda + dv^a \otimes \partial_a + y^\mu\, (\sum_{h=1}^{n} \frac{1}{2}\, \psi^{ih}\, \psi_{\mu\nu h})\, dx^\nu \otimes \partial_i. && (3.8) \end{aligned}$$

Explicitly, each $\tilde{\Gamma} \in Sec(C_T/M)$ gives an injection $(\tilde{\Gamma} \circ \pi_T, I_{TM})$, of TM into $C_T \times TM$, which is a section of π_1 and Γ coincides with the restriction of Λ_T to this section:

$$\Lambda_{T|(\tilde{\Gamma}\circ\pi_T, I_{TM})TM} = \Gamma.$$

and the universal curvature of the connection Λ is given by:

$$\Omega_T = d_{\Lambda_T}\Lambda_T : C_T \times_M TM \to \wedge^2(T^*C_T) \otimes_{TM} V(TM).$$

So here the universal curvature Ω_T has the coordinate expression:

$$\begin{aligned}\Omega_T &= \frac{1}{2}\left(y^k v^j_{k\lambda}\,\partial_j y^m v^i_{m\mu}\,dx^\lambda \wedge dx^\mu + 2\,\partial_a y^m v^i_{m\mu}\,dx^a \wedge dx^\mu\right)\otimes \partial_i\\ &= \frac{1}{2}\left(y^k(\sum_{h=1}^n \frac{1}{2}\,\psi^{jh}\,\psi_{k\lambda h})\partial_j y^m(\sum_{h=1}^n \frac{1}{2}\,\psi^{ih}\,\psi_{m\mu h})dx^\lambda \wedge dx^\mu + 2\,\partial_a y^m(\sum_{h=1}^n \frac{1}{2}\,\psi^{ih}\,\psi_{m\mu h})dx^a \wedge dx^\mu\right)\otimes \partial_i\,.\end{aligned}$$

3.3.2 Frame bundle system: $C_F \times FM \to JFM$

A linear connection is also a principal (i.e. group invariant) connection on the principal bundle of frames FM with:

$$E = FM \to M = FM/G$$

consisting of linear frames (ordered bases for tangent spaces) with structure group the general linear group, $G = G\ell(n)$. Here the system space is

$$C_F = JFM/G \subset T^*M \otimes_{TM} TFM/G,$$

consisting of G-invariant jets. The system morphism is

$$\begin{aligned}\xi_F : C_F \times FM &\to JFM \subset T^*M \otimes_{TM} TFM\,,\\ ([js_x], b) &\longmapsto [T_xM \longmapsto T_bFM].\end{aligned}$$

In coordinates

$$\begin{aligned}\xi_F &= dx^\lambda \otimes (\partial_\lambda - X^\mu\,\partial_\mu s^\lambda_\nu)\,\tilde{\partial}_{\nu^\lambda}\\ &= dx^\lambda \otimes (\partial_\lambda - X^\mu\,\Gamma^\lambda_{\mu\nu})\,\tilde{\partial}_{\nu^\lambda}\\ &= dx^\lambda \otimes (\partial_\lambda - X^\mu \sum_{h=1}^n \frac{1}{2}\,\psi^{\lambda h}\,\psi_{\mu\nu h})\,\tilde{\partial}_{\nu^\lambda}\end{aligned} \tag{3.9}$$

where $\tilde{\partial}_{\nu^\lambda} = \frac{\partial}{\partial b^\lambda_\nu}$ is the natural base on the vertical fibre of T_bFM induced by coordinates (b^λ_ν) on FM.

Each section of $C_F \to M$ that is projectable onto I_{TM}, such as, $\hat{\Gamma} : M \to C_F : (x^\lambda) \to (x^\lambda, [j\gamma_x])$ with $\Gamma^\lambda_{\mu\nu} = \partial_\mu s^\lambda_\nu$; determines the unique linear connection $\Gamma = \xi_F \circ (\hat{\Gamma} \circ \pi_F, I_{FM})$ with Christoffel symbols $\Gamma^\lambda_{\mu\nu}$.

On the principal G-bundle $\pi_1 : C_F \times_M FM \to C_F$; the universal connection is given by:

$$\begin{aligned}\Lambda_F : C_F \times_M FM &\to J(C_F \times_M FM) \subset T^*C_F \otimes_{FM} T(C_F \times_{FM} FM)\,,\\ (x^\lambda, v^\lambda_{\mu\nu}, b^\mu_\nu) &\longmapsto [(X^\lambda, Y^\lambda_{\mu\nu}) \to (X^\lambda, Y^\lambda_{\mu\nu}, b^\mu_\nu v^\lambda_{\mu\theta} X^\theta)].\end{aligned}$$

briefly,

$$\begin{aligned}\Lambda_F &= dx^\lambda \otimes \partial_\lambda + dv^a \otimes \partial_a + b^\mu_\nu v^\lambda_{\mu\theta}\,dx^\theta \otimes \tilde{\partial}_{\nu^\lambda}\\ &= dx^\lambda \otimes \partial_\lambda + dv^a \otimes \partial_a + b^\mu_\nu\,(\sum_{h=1}^n \frac{1}{2}\,\psi^{\lambda h}\,\psi_{\mu\theta h})\,dx^\theta \otimes \tilde{\partial}_{\nu^\lambda}.\end{aligned} \tag{3.10}$$

Explicitly, each $\tilde{\Gamma} \in Sec(C_F/M)$ gives an injection $(\tilde{\Gamma} \circ \pi_F, I_{FM})$, of FM into $C_F \times FM$, which is a section of π_1 and Γ coincides with the restriction of Λ_F to this section:

$$\Lambda_{F|(\tilde{\Gamma}\circ\pi_F, I_{FM})FM} = \Gamma.$$

and the universal curvature of the connection Λ is given by:

$$\Omega = d_{\Lambda_F}\Lambda_F : C_F \times_M FM \to \wedge^2(T^*C_F) \otimes_{FM} V(FM).$$

So here the universal curvature form Ω_F has the coordinate expression:

$$\begin{aligned}\Omega_F &= \frac{1}{2}\left(b^k_\nu v^\beta_{k\lambda}\,\tilde{\partial}_{\nu^\beta} b^m_\omega v^\alpha_{m\mu}\,dx^\lambda \wedge dx^\mu + 2\,\partial_a b^m_\omega v^\alpha_{m\mu}\,dx^a \wedge dx^\mu\right) \otimes \tilde{\partial}_{\omega^\alpha} \\ &= \frac{1}{2}\left(b^k_\nu(\sum_{h=1}^n \frac{1}{2}\,\psi^{\beta h}\,\psi_{k\lambda h})\tilde{\partial}_{\nu^\beta} b^m_\omega(\sum_{h=1}^n \frac{1}{2}\,\psi^{\alpha h}\,\psi_{m\mu h})dx^\lambda \wedge dx^\mu + 2\partial_a b^m_\omega(\sum_{h=1}^n \frac{1}{2}\,\psi^{\alpha h}\,\psi_{m\mu h})dx^a \wedge dx^\mu\right) \otimes \tilde{\partial}_{\omega^\alpha}\,.\end{aligned}$$

3.4 Universal connection and curvature on the gamma manifold

Here we give explicit forms for the system space and its universal connection and curvature on the 2-dimensional statistical manifold of gamma distributions.

Let $\mathcal{G}$ be the family of gamma probability density functions given by

$$\{p(x;\beta,\mu) = \beta^\mu\,\frac{x^{\mu-1}}{\Gamma(\mu)}\,e^{-\beta\,x}|\beta,\mu \in \mathbb{R}^+\}, \quad x \in \mathbb{R}^+. \tag{3.11}$$

So the space of parameters is topologically $\mathbb{R}^+ \times \mathbb{R}^+$. It is an exponential family and it includes as a special case ($\mu = 1$) the exponential distribution itself, which complements the Poisson process on a line. Since the set of all gamma distributions is an exponential family with natural coordinate system (β,μ) and potential function $\psi = \log\Gamma(\mu) - \mu\,\log\beta$, [4] the Fisher metric (3.3) is given by:

$$[g_{ij}] = \begin{bmatrix} \frac{\mu}{\beta^2} & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \phi'(\mu) \end{bmatrix} \tag{3.12}$$

where $\phi(\mu) = \frac{\Gamma'(\mu)}{\Gamma(\mu)}$ is the logarithmic derivative of the gamma function.

The Levi-Civita connection components (3.4) are given by:

$$\begin{aligned}\Gamma^1_{11} &= \frac{(1 - 2\,\mu\,\phi'(\mu))}{2\,\beta\,(-1 + \mu\,\phi'(\mu))}\,, \\ \Gamma^1_{12} &= \frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\,, \\ \Gamma^1_{22} &= \frac{\beta\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\,, \\ \Gamma^2_{11} &= \frac{\mu}{2\,\beta^2\,(1 - \mu\,\phi'(\mu))}\,, \\ \Gamma^2_{12} &= \frac{1}{2\,\beta\,(\mu\,\phi'(\mu) - 1)}\,, \\ \Gamma^2_{22} &= \frac{\mu\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\end{aligned} \tag{3.13}$$

while the other independent components are zero.

3.4.1 Tangent bundle system

The system space is

$$C_T = \{\alpha \otimes j\gamma \in T^*\mathcal{G} \otimes_{\mathcal{G}} JT\mathcal{G} \mid j\gamma : T\mathcal{G} \to TT\mathcal{G} \ \text{projects onto} \ I_{T\mathcal{G}}\}$$

and the system morphism is

$$\begin{aligned}
\xi_T &= dx^\lambda \otimes (\partial_\lambda - y^j \Gamma^i_{j\lambda}\, \partial_i) \\
&= dx^1 \otimes \left(\partial_1 - (\frac{(1-2\,\mu\,\phi'(\mu))}{2\,\beta\ (-1+\mu\,\phi'(\mu))}\, y^1 + \frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\, y^2)\, \partial_1\right) \\
&+ dx^1 \otimes \left(\partial_1 - (\frac{\mu}{2\,\beta^2\ (1-\mu\,\phi'(\mu))}\, y^1 + \frac{1}{2\,\beta\,(\mu\,\phi'(\mu)-1)}\, y^2)\, \partial_2\right) \\
&+ dx^2 \otimes \left(\frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\, y^1 + \frac{\beta\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\, y^2)\, \partial_1\right) \\
&+ dx^2 \otimes \left(\partial_2 - (\frac{1}{2\,\beta\,(\mu\,\phi'(\mu)-1)}\, y^1 + \frac{\mu\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\, y^2)\, \partial_2\right).
\end{aligned}$$

The universal connection on the gamma manifold is given here by:

$$\begin{aligned}
\Lambda_T &= dx^\lambda \otimes \partial_\lambda + d^i_{\lambda j} \otimes \partial^i_{\lambda j} + y^\tau\, \Gamma^i_{\tau\nu} dx^\nu \otimes \partial_i \\
&= dx^\lambda \otimes \partial_\lambda + d^i_{\lambda j} \otimes \partial^i_{\lambda j} \\
&+ \left(\frac{(1-2\,\mu\,\phi'(\mu))}{2\,\beta\ (-1+\mu\,\phi'(\mu))}\, y^1 + \frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\, y^2\right) dx^1 \otimes \partial_1 \\
&+ \left(\frac{\mu}{2\,\beta^2\ (1-\mu\,\phi'(\mu))}\, y^1 + \frac{1}{2\,\beta\,(\mu\,\phi'(\mu)-1)}\, y^2\right) dx^1 \otimes \partial_2 \\
&+ \left(\frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\, y^1 + \frac{\beta\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\, y^2\right) dx^2 \otimes \partial_1 \\
&+ \left(\frac{1}{2\,\beta\,(\mu\,\phi'(\mu)-1)}\, y^1 + \frac{\mu\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\, y^2\right) dx^2 \otimes \partial_2\,.
\end{aligned}$$

The universal curvature on the gamma manifold is:

$$
\begin{aligned}
\Omega_T &= \frac{1}{2}\left(y^k\Gamma^j_{k\lambda}\,\partial_j y^m\Gamma^i_{m\mu}\,dx^\lambda\wedge dx^\mu + 2\,\partial_a y^m\Gamma^i_{m\mu}\,dx^a\wedge dx^\mu\right)\otimes\partial_i\\
&= \frac{1}{2}[(y^1\Gamma^1_{11}\,\partial_1 y^1\Gamma^i_{11} + y^1\Gamma^2_{11}\,\partial_2 y^1\Gamma^i_{11} + y^2\Gamma^1_{21}\,\partial_1 y^1\Gamma^i_{11} + y^2\Gamma^2_{21}\,\partial_2 y^1\Gamma^i_{11}\\
&+ y^1\Gamma^1_{11}\,\partial_1 y^2\Gamma^i_{21} + y^1\Gamma^2_{11}\,\partial_2 y^2\Gamma^i_{21} + y^2\Gamma^1_{21}\,\partial_1 y^2\Gamma^i_{21} + y^2\Gamma^2_{21}\,\partial_2 y^2\Gamma^i_{21})\,dx^1\wedge dx^1\\
&+ 2\,\partial_1(y^1\Gamma^i_{11} + y^2\Gamma^i_{21})\,dx^1\wedge dx^1]\otimes\partial_1\\
&+ \frac{1}{2}[(y^1\Gamma^1_{11}\,\partial_1 y^1\Gamma^i_{11} + y^1\Gamma^2_{11}\,\partial_2 y^1\Gamma^i_{11} + y^2\Gamma^1_{21}\,\partial_1 y^1\Gamma^i_{11} + y^2\Gamma^2_{21}\,\partial_2 y^1\Gamma^i_{11}\\
&+ y^1\Gamma^1_{11}\,\partial_1 y^2\Gamma^i_{21} + y^1\Gamma^2_{11}\,\partial_2 y^2\Gamma^i_{21} + y^2\Gamma^1_{21}\,\partial_1 y^2\Gamma^i_{21} + y^2\Gamma^2_{21}\,\partial_2 y^2\Gamma^i_{21})\,dx^1\wedge dx^1\\
&+ 2\,\partial_2(y^1\Gamma^i_{11} + y^2\Gamma^i_{21})\,dx^2\wedge dx^1]\otimes\partial_i\\
&+ \frac{1}{2}[(y^1\Gamma^1_{12}\,\partial_1 y^1\Gamma^i_{11} + y^1\Gamma^2_{12}\,\partial_2 y^1\Gamma^i_{11} + y^2\Gamma^1_{22}\,\partial_1 y^1\Gamma^i_{11} + y^2\Gamma^2_{22}\,\partial_2 y^1\Gamma^i_{11}\\
&+ y^1\Gamma^1_{12}\,\partial_1 y^2\Gamma^i_{21} + y^1\Gamma^2_{12}\,\partial_2 y^2\Gamma^i_{21} + y^2\Gamma^1_{22}\,\partial_1 y^2\Gamma^i_{21} + y^2\Gamma^2_{22}\,\partial_2 y^2\Gamma^i_{21})\,dx^2\wedge dx^1\\
&+ 2\,\partial_1(y^1\Gamma^i_{11} + y^2\Gamma^i_{21})\,dx^1\wedge dx^1]\otimes\partial_i\\
&+ \frac{1}{2}[(y^1\Gamma^1_{12}\,\partial_1 y^1\Gamma^i_{11} + y^1\Gamma^2_{12}\,\partial_2 y^1\Gamma^i_{11} + y^2\Gamma^1_{22}\,\partial_1 y^1\Gamma^i_{11} + y^2\Gamma^2_{22}\,\partial_2 y^1\Gamma^i_{11}\\
&+ y^1\Gamma^1_{12}\,\partial_1 y^2\Gamma^i_{21} + y^1\Gamma^2_{12}\,\partial_2 y^2\Gamma^i_{21} + y^2\Gamma^1_{22}\,\partial_1 y^2\Gamma^i_{21} + y^2\Gamma^2_{22}\,\partial_2 y^2\Gamma^i_{21})\,dx^2\wedge dx^1\\
&+ 2\,\partial_2(y^1\Gamma^i_{11} + y^2\Gamma^i_{21})\,dx^2\wedge dx^1]\otimes\partial_i\\
&+ \frac{1}{2}[(y^1\Gamma^1_{11}\,\partial_1 y^1\Gamma^i_{12} + y^1\Gamma^2_{11}\,\partial_2 y^1\Gamma^i_{12} + y^2\Gamma^1_{21}\,\partial_1 y^1\Gamma^i_{12} + y^2\Gamma^2_{21}\,\partial_2 y^1\Gamma^i_{12}\\
&+ y^1\Gamma^1_{11}\,\partial_1 y^2\Gamma^i_{22} + y^1\Gamma^2_{11}\,\partial_2 y^2\Gamma^i_{22} + y^2\Gamma^1_{21}\,\partial_1 y^2\Gamma^i_{22} + y^2\Gamma^2_{21}\,\partial_2 y^2\Gamma^i_{22})\,dx^1\wedge dx^2\\
&+ 2\,\partial_1(y^1\Gamma^i_{12} + y^2\Gamma^i_{22})\,dx^1\wedge dx^2]\otimes\partial_1\\
&+ \frac{1}{2}[(y^1\Gamma^1_{11}\,\partial_1 y^1\Gamma^i_{12} + y^1\Gamma^2_{11}\,\partial_2 y^1\Gamma^i_{12} + y^2\Gamma^1_{21}\,\partial_1 y^1\Gamma^i_{12} + y^2\Gamma^2_{21}\,\partial_2 y^1\Gamma^i_{12}\\
&+ y^1\Gamma^1_{11}\,\partial_1 y^2\Gamma^i_{22} + y^1\Gamma^2_{11}\,\partial_2 y^2\Gamma^i_{22} + y^2\Gamma^1_{21}\,\partial_1 y^2\Gamma^i_{22} + y^2\Gamma^2_{21}\,\partial_2 y^2\Gamma^i_{22})\,dx^1\wedge dx^2\\
&+ 2\,\partial_2(y^1\Gamma^i_{12} + y^2\Gamma^i_{22})\,dx^2\wedge dx^2]\otimes\partial_i\\
&+ \frac{1}{2}[(y^1\Gamma^1_{12}\,\partial_1 y^1\Gamma^i_{12} + y^1\Gamma^2_{12}\,\partial_2 y^1\Gamma^i_{12} + y^2\Gamma^1_{22}\,\partial_1 y^1\Gamma^i_{12} + y^2\Gamma^2_{22}\,\partial_2 y^1\Gamma^i_{12}\\
&+ y^1\Gamma^1_{12}\,\partial_1 y^2\Gamma^i_{22} + y^1\Gamma^2_{12}\,\partial_2 y^2\Gamma^i_{22} + y^2\Gamma^1_{22}\,\partial_1 y^2\Gamma^i_{22} + y^2\Gamma^2_{22}\,\partial_2 y^2\Gamma^i_{22})\,dx^2\wedge dx^2\\
&+ 2\,\partial_1(y^1\Gamma^i_{12} + y^2\Gamma^i_{22})\,dx^1\wedge dx^2]\otimes\partial_i\\
&+ \frac{1}{2}[(y^1\Gamma^1_{12}\,\partial_1 y^1\Gamma^i_{12} + y^1\Gamma^2_{12}\,\partial_2 y^1\Gamma^i_{12} + y^2\Gamma^1_{22}\,\partial_1 y^1\Gamma^i_{12} + y^2\Gamma^2_{22}\,\partial_2 y^1\Gamma^i_{12}\\
&+ y^1\Gamma^1_{12}\,\partial_1 y^2\Gamma^i_{22} + y^1\Gamma^2_{12}\,\partial_2 y^2\Gamma^i_{22} + y^2\Gamma^1_{22}\,\partial_1 y^2\Gamma^i_{22} + y^2\Gamma^2_{22}\,\partial_2 y^2\Gamma^i_{22})\,dx^2\wedge dx^2\\
&+ 2\,\partial_2(y^1\Gamma^i_{12} + y^2\Gamma^i_{22})\,dx^2\wedge dx^2]\otimes\partial_i \quad (i=1,2).
\end{aligned}
$$

3.4.2 Frame bundle system

The system space is $C_F = JF\mathcal{G}/G$ and the system morphism is

$$
\begin{aligned}
\xi_F &= dx^\lambda\otimes(\partial_\lambda - X^\tau\,\Gamma^\lambda_{\tau\nu})\,\tilde{\partial}_{\nu^\lambda}\\
&= dx^1\otimes\left(\partial_1 - (\frac{(1-2\,\mu\,\phi'(\mu))}{2\,\beta\,(-1+\mu\,\phi'(\mu))}\,X^1 + \frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\,X^2)\right)\tilde{\partial}_{1^1}\\
&+ dx^2\otimes\left(\partial_2 - (\frac{\mu}{2\,\beta^2\,(1-\mu\,\phi'(\mu))}\,X^1 + \frac{1}{2\,\beta\,(\mu\,\phi'(\mu)-1)}\,X^2)\right)\tilde{\partial}_{1^2}\\
&+ dx^1\otimes\left(\partial_1 - (\frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\,X^1 + \frac{\beta\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\,X^2)\right)\tilde{\partial}_{2^1}\\
&+ dx^2\otimes\left(\partial_2 - (\frac{1}{2\,\beta\,(\mu\,\phi'(\mu)-1)}\,X^1 + \frac{\mu\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu)-1)}\,X^2)\right)\tilde{\partial}_{2^2}.
\end{aligned}
$$

The universal connection on the gamma manifold is:

$$\begin{aligned}
\Lambda_F &= dx^\lambda \otimes \partial_\lambda + d^i_{\lambda j} \otimes \partial^i_{\lambda j} + b^\nu_\tau \, \Gamma^\lambda_{\tau\theta} dx^\theta \otimes \tilde{\partial}_{\nu^\lambda} \\
&= dx^\lambda \otimes \partial_\lambda + d^i_{\lambda j} \otimes \partial^i_{\lambda j} \\
&+ \left(\frac{(1 - 2\,\mu\,\phi'(\mu))}{2\,\beta\,(-1 + \mu\,\phi'(\mu))}\, b^1_1 + \frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\, b^2_1 \right) dx^1 \otimes \tilde{\partial}_{1^1} \\
&+ \left(\frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\, b^1_1 + \frac{\beta\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\, b^2_1 \right) dx^2 \otimes \tilde{\partial}_{1^1} \\
&+ \left(\frac{(1 - 2\,\mu\,\phi'(\mu))}{2\,\beta\,(-1 + \mu\,\phi'(\mu))}\, b^1_2 + \frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\, b^2_2 \right) dx^1 \otimes \tilde{\partial}_{2^1} \\
&+ \left(\frac{\phi'(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\, b^1_2 + \frac{\beta\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\, b^2_2 \right) dx^2 \otimes \tilde{\partial}_{2^1} \\
&+ \left(\frac{\mu}{2\,\beta^2\,(1 - \mu\,\phi'(\mu))}\, b^1_1 + \frac{1}{2\,\beta\,(\mu\,\phi'(\mu) - 1)}\, b^2_1 \right) dx^1 \otimes \tilde{\partial}_{1^2} \\
&+ \left(\frac{1}{2\,\beta\,(\mu\,\phi'(\mu) - 1)}\, b^1_1 + \frac{\mu\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\, b^2_1 \right) dx^2 \otimes \tilde{\partial}_{1^2} \\
&+ \left(\frac{\mu}{2\,\beta^2\,(1 - \mu\,\phi'(\mu))}\, b^1_2 + \frac{1}{2\,\beta\,(\mu\,\phi'(\mu) - 1)}\, b^2_2 \right) dx^1 \otimes \tilde{\partial}_{2^2} \\
&+ \left(\frac{1}{2\,\beta\,(\mu\,\phi'(\mu) - 1)}\, b^1_2 + \frac{\mu\,\phi''(\mu)}{2\,(\mu\,\phi'(\mu) - 1)}\, b^2_2 \right) dx^2 \otimes \tilde{\partial}_{2^2} \,.
\end{aligned}$$

The universal curvature on the gamma manifold is:

$$\Omega_F = \frac{1}{2} \left(b^k_\nu \Gamma^\beta_{k\lambda}\, \tilde{\partial}_{\nu^\beta} b^m_\omega \Gamma^\alpha_{m\mu}\, dx^\lambda \wedge dx^\mu + 2\, \partial_a b^m_\omega \Gamma^\alpha_{m\mu}\, dx^a \wedge dx^\mu \right) \otimes \tilde{\partial}_{\omega^\alpha}$$

The analytic form of this is known but is rather cumbersome and omitted here.

Chapter 4

Neighbourhoods of randomness, independence, and uniformity

We obtain some results that augment our information geometric measures for distances in spaces of distributions, by providing explicit geometric representations of neighbourhoods for each of these important states for stochastic processes:

- randomness,
- independence,
- uniformity.

Such results are significant theoretically because they are very general, and practically because they are topological and so therefore stable under perturbations.

4.1 Gamma manifold $\mathcal{G}$ and neighbourhoods of randomness

The univariate gamma distribution is widely used to model processes involving a continuous positive random variable. Its information geometry is known and has been applied recently to represent and metrize departures from randomness of, for example, the processes that allocate gaps between occurrences of each amino acid along a protein chain within the *Saccharomyces cerevisiae* genome, see Cai et al [12], clustering of galaxies and communications, Dodson [21, 22, 19]. In fact, we have made precise the statement that around every random process there is a neighbourhood of stochastic processes subordinate to the gamma distribution, so gamma distributions can approximate any small enough departure from randomness.

Proposition 4.1.1 *Every neighbourhood of a random process contains a neighbourhood of stochastic processes subordinate to gamma distributions.*

Proof
Dodson and Matsuzoe [23] have provided an affine immersion in Euclidean $\mathbb{R}^3$ for $\mathcal{G}$, the manifold of gamma distributions with Fisher information metric. The coordinates ($\mu = \alpha/\beta$, α) form a natural coordinate system (cf Amari and Nagaoka [4]) for the gamma manifold $\mathcal{G}$. Then $\mathcal{G}$ can be realized in Euclidean $\mathbb{R}^3$ as the graph of the affine immersion $\{h, \xi\}$ where ξ is a transversal vector field along h [4, 23]:

$$h : \mathcal{G} \to \mathbb{R}^3 : \begin{pmatrix} \mu \\ \alpha \end{pmatrix} \mapsto \begin{pmatrix} \mu \\ \alpha \\ \log\Gamma(\alpha) - \alpha \log \mu \end{pmatrix}, \quad \xi = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

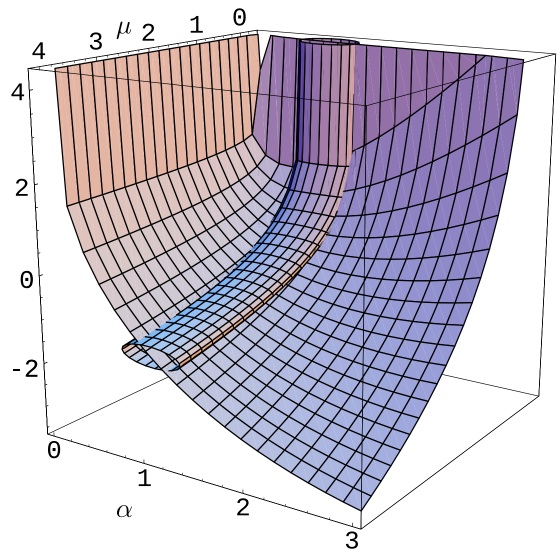

Figure 4.1: *Tubular neighbourhood of all random processes. Affine immersion in natural coordinates $\mu = \alpha/\beta, \alpha$ as a surface in $\mathbb{R}^3$ for the gamma manifold $\mathcal{G}$; the tubular neighbourhood surrounds all exponential distributions—these lie on the curve $\alpha = 1$ in the surface. Since the log-gamma manifold $\mathcal{L}$ is an isometric isomorph of $\mathcal{G}$, this figure represents also a tubular neighbourhood in $\mathbb{R}^3$ of the uniform distribution from the log-gamma manifold.*

The submanifold of exponential distributions is represented by the curve

$$(0, \infty) \to \mathbb{R}^3 : \mu \mapsto \{\mu, 1, \log\frac{1}{\mu}\}$$

and for this curve, a tubular neighbourhood in $\mathbb{R}^3$ such as that bounded by the surface

$$\left\{ \{\mu - \frac{0.6\,\cos\theta}{\sqrt{1+\mu^2}}, 1 - 0.6\,\sin\theta, \frac{-0.6\,\mu\,\cos\theta}{\sqrt{1+\mu^2}} - \log\mu\}\ \theta \in [0, 2\pi) \right\} \tag{4.1}$$

will contain all immersions for small enough perturbations of exponential distributions. In Figure 4.1 this is depicted in natural coordinates μ, α. The tubular neighbourhood (4.1) intersects with the gamma manifold immersion to yield the required neighbourhood of gamma distributions, which completes our proof. □

4.2 Log-gamma manifold $\mathcal{L}$ and neighbourhoods of uniformity

The family of log-gamma distributions discussed in 2.6 has probability density functions for random variable $N \in [0, 1]$ given by

$$g(N, \mu, \beta) = \frac{\frac{1}{N}^{1-\frac{\beta}{\mu}}\,(\frac{\beta}{\mu})^\beta\,(\log\frac{1}{N})^{\beta-1}}{\Gamma(\beta)} \quad \text{for } \mu > 0 \text{ and } \beta > 0. \tag{4.2}$$

It has the uniform distribution as a limit

$$\lim_{\mu\to 1} g(N, \mu, 1) = g(N, 1, 1) = 1\ .$$

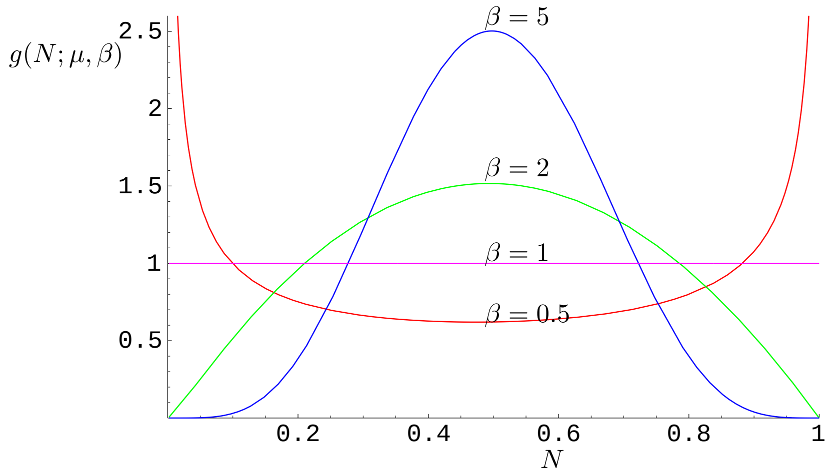

Figure 4.2: *Log-gamma probability density functions* $g(N;\mu,\beta)$, $N \in [0,1]$, *with central mean* $< N >= 0.5$, *and* $\beta = 0.5, 1,\ 2,\ 5$. *The cases* $\beta < 1$ *correspond in gamma distributions to clustering in an underlying spatial process; conversely,* $\beta > 1$ *corresponds to dispersion and greater evenness than random.*

Figure 4.2 shows some log-gamma pdfs around the uniform distribution and Figure 2.2 shows the continuous family from which these are drawn.

The log-gamma manifold $\mathcal{L}$ has information metric (2.30), isometric with the gamma manifold, $\mathcal{G}$ by Proposition 2.6.1. Hence, from the result of Dodson and Matsuzoe [23] the immersion of $\mathcal{G}$ in $\mathbb{R}^3$, Figure 4.1, represents also the log-gamma manifold $\mathcal{L}$. Then, since the isometry sends the exponential distribution to the uniform distribution on $[0,1]$, we obtain a general deduction

Proposition 4.2.1 *Every neighbourhood of the uniform distribution contains a neighbourhood of log-gamma distributions.* $\square$

Equivalently,

Proposition 4.2.2 *Every neighbourhood of a uniform stochastic process contains a neighbourhood of stochastic processes subordinate to log-gamma distributions.* $\square$

4.3 Freund manifold $\mathcal{F}$ and neighbourhoods of independence

Let $\mathcal{F}$ be the manifold of Freund bivariate mixture exponential distributions 2.9, so with positive parameters α_i, β_i,

$$\mathcal{F} \equiv \{f|f(x,y;\alpha_1,\beta_1,\alpha_2,\beta_2) = \left\{ \begin{array}{ll} \alpha_1\beta_2 e^{-\beta_2 y-(\alpha_1+\alpha_2-\beta_2)x} & \text{for } 0 \le x < y \\ \alpha_2\beta_1 e^{-\beta_1 x-(\alpha_1+\alpha_2-\beta_1)y} & \text{for } 0 \le y \le x \end{array} \right\} . \tag{4.3}$$

4.3.1 Submanifold $F_2 \subset \mathcal{F}$: $\alpha_1 = \alpha_2,\ \beta_1 = \beta_2$

The distributions are of form :

$$f(x,y;\alpha_1,\beta_1) = \left\{ \begin{array}{ll} \alpha_1\beta_1\, e^{-\beta_1 y-(2\,\alpha_1-\beta_1)x} & \text{for } 0 < x < y \\ \alpha_1\beta_1\, e^{-\beta_1 x-(2\,\alpha_1-\beta_1)y} & \text{for } 0 < y < x \end{array} \right. \tag{4.4}$$

with parameters $\alpha_1, \beta_1 > 0$. The covariance, correlation coefficient and marginal density functions, of X and Y are given by :

$$Cov(X,Y) = \frac{1}{4}\left(\frac{1}{{\alpha_1}^2} - \frac{1}{{\beta_1}^2}\right), \tag{4.5}$$

$$\rho(X,Y) = 1 - \frac{4\,{\alpha_1}^2}{3\,{\alpha_1}^2 + {\beta_1}^2}, \tag{4.6}$$

$$f_X(x) = \left(\frac{\alpha_1}{2\,\alpha_1 - \beta_1}\right)\beta_1\, e^{-\beta_1 x} + \left(\frac{\alpha_1 - \beta_1}{2\,\alpha_1 - \beta_1}\right)(2\,\alpha_1)\, e^{-2\,\alpha_1 x}\,,\ x \geq 0\,, \tag{4.7}$$

$$f_Y(y) = \left(\frac{\alpha_1}{2\,\alpha_1 - \beta_1}\right)\beta_1\, e^{-\beta_1 y} + \left(\frac{\alpha_1 - \beta_1}{2\,\alpha_1 - \beta_1}\right)(2\,\alpha_1)\, e^{-2\,\alpha_1 y}\,,\ y \geq 0\,. \tag{4.8}$$

Proposition 4.3.1 *F_2 forms an exponential family, with parameters (α_1, β_1) and potential function*

$$\psi = -\log(\alpha_1\,\beta_1) \tag{4.9}$$

Proposition 4.3.2 *$Cov(X,Y) = \rho(X,Y) = 0$ if and only if $\alpha_1 = \beta_1$ and in this case the density functions are of form*

$$f(x,y;\alpha_1,\alpha_1) = \alpha_1^2\, e^{\alpha_1|y-x|} = f_X(x) f_Y(y) \tag{4.10}$$

so that here we do have independence of these exponentials if and only if the covariance is zero.

Neighbourhoods of independence in F_2

An important practical application of the Freund submanifold F_2 is the representation of a bivariate stochastic proces for which the marginals are identical exponentials. The next result is important because it provides topological neighbourhoods of that subspace W in F_2 consisting of the bivariate processes that have zero covariance: we obtain neighbourhoods of independence for random (ie exponentially distributed) processes.

Proposition 4.3.3 *Every neighbourhood of an independent pair of identical random processes contains a neighbourhood of bivariate stochastic processes subordinate to Freund distributions.*

Proof
Let $\{F_2, g, \nabla^{(1)}, \nabla^{(-1)}\}$ be the manifold F_2 with Fisher metric g and exponential connection $\nabla^{(1)}$. Then F_2 can be realized in Euclidean $\mathbb{R}^3$ by the graph of a potential function, via the affine immersion

$$h : \mathcal{G} \to \mathbb{R}^3 : \begin{pmatrix}\alpha_1\\ \beta_1\end{pmatrix} \mapsto \begin{pmatrix}\alpha_1\\ \beta_1\\ -\log(\alpha_1\,\beta_1)\end{pmatrix}.$$

In F_2, the submanifold W consisting of the independent case ($\alpha_1 = \beta_1$) is represented by the curve

$$W : (0,\infty) \to \mathbb{R}^3 : (\alpha_1) \mapsto (\alpha_1, \alpha_1, -2\log\alpha_1). \tag{4.11}$$

This is illustrated in Figure 4.3 which shows an affine embedding of F_2 as a surface in $\mathbb{R}^3$, and an $\mathbb{R}^3$-tubular neighbourhood of W, the curve $\alpha_1 = \beta_1$ in the surface. This curve W represents all bivariate distributions having identical exponential marginals and zero covariance; its tubular neighourhood represents all small enough departures from independence. □

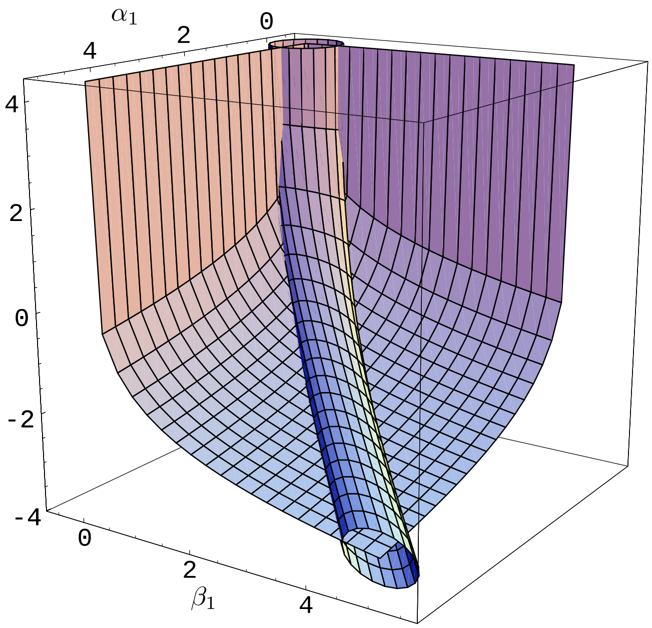

Figure 4.3: *Tubular neighbourhood of independent random processes. An affine immersion in natural coordinates (α_1, β_1) as a surface in $\mathbb{R}^3$ for the Freund submanifold F_2; the tubular neighbourhood surrounds the curve ($\alpha_1 = \beta_1$ in the surface) consisting of all bivariate distributions having identical exponential marginals and zero covariance.*

4.4 Neighbourhoods of independence for Gaussian processes [8]

The bivariate Gaussian distribution has the form:

$$f(x,y) = \frac{1}{2\pi\sqrt{\sigma_1\,\sigma_2 - {\sigma_{12}}^2}}\, e^{-\frac{1}{2(\sigma_1\,\sigma_2-{\sigma_{12}}^2)}\left(\sigma_2(x-\mu_1)^2-2\sigma_{12}(x-\mu_1)(y-\mu_2)+\sigma_1(y-\mu_2)^2\right)}, \tag{4.12}$$

defined on $-\infty < x\,, y < \infty$ with parameters $(\mu_1,\, \mu_2,\, \sigma_1,\, \sigma_{12},\, \sigma_2)$; where $-\infty < \mu_1\,, \mu_2 < \infty$, $0 < \sigma_1\,, \sigma_2 < \infty$ and σ_{12} is the covariance of X and Y.

The marginal functions, of X and Y are univariate Gaussian distributions:

$$f_X(x, \mu_1, \sigma_1) = \frac{1}{\sqrt{2\,\pi\,\sigma_1}}\, e^{-\frac{(x-\mu_1)^2}{2\,\sigma_1}}\,, \tag{4.13}$$

$$f_Y(y, \mu_2, \sigma_2) = \frac{1}{\sqrt{2\,\pi\,\sigma_2}}\, e^{-\frac{(y-\mu_2)^2}{2\,\sigma_2}}\,. \tag{4.14}$$

The correlation coefficient is:

$$\rho(X,Y) = \frac{\sigma_{12}}{\sqrt{\sigma_1\,\sigma_2}}$$

Since ${\sigma_{12}}^2 < \sigma_1\,\sigma_2$ then $-1 < \rho(X,Y) < 1$; so we do not have the case when Y is a linearly increasing (or decreasing) function of X. The space of bivariate Gaussians becomes a Riemannian 5-manifold N with Fisher information metric.

Independence submanifold: $N_1 \subset N$**:** $\sigma_{12} = 0$

The distributions are of form:

$$f(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2) = f_X(x, \mu_1, \sigma_1).f_Y(y, \mu_2, \sigma_2) \tag{4.15}$$

This is the case for statistical independence of X and Y, so the space N_1 is the direct product of two Riemannian spaces

$$\{f_X(x, \mu_1, \sigma_1),\, \mu_1 \in \mathbb{R},\, \sigma_1 \in \mathbb{R}^+\} \text{ and } \{f_Y(y, \mu_2, \sigma_2),\, \mu_2 \in \mathbb{R},\, \sigma_2 \in \mathbb{R}^+\}.$$

Identical marginal Gaussian submanifold: $N_2 \subset N$**:** $\sigma_1 = \sigma_2 = \sigma$ **and** $\mu_1 = \mu_2 = \mu$

The distributions are of form:

$$f(x, y; \mu, \sigma, \sigma_{12}) = \frac{1}{2\pi\sqrt{\sigma^2 - {\sigma_{12}}^2}}\, e^{-\frac{1}{2(\sigma^2 - {\sigma_{12}}^2)}\left(\sigma(x-\mu)^2 - 2\sigma_{12}(x-\mu)(y-\mu) + \sigma(y-\mu)^2\right)} \tag{4.16}$$

The marginal functions are $f_X = f_Y \equiv N(\mu, \sigma)$, with correlation coefficient $\rho(X, Y) = \frac{\sigma_{12}}{\sigma}$.

Central mean submanifold: $N_3 \subset N$**:** $\mu_1 = \mu_2 = 0$

The distributions are of form:

$$f(x, y; \sigma_1, \sigma_2, \sigma_{12}) = \frac{1}{2\,\pi\,\sqrt{\sigma_1\,\sigma_2 - {\sigma_{12}}^2}}\, e^{-\frac{1}{2\,(\sigma_1\,\sigma_2 - {\sigma_{12}}^2)}\left(\sigma_2 x^2 - 2\,\sigma_{12}\,x\,y + \sigma_1 y^2\right)} \tag{4.17}$$

The marginal functions are $f_X(x, 0, \sigma_1)$ and $f_Y(y, 0, \sigma_2)$, with correlation coefficient $\rho(X, Y) = \frac{\sigma_{12}}{\sqrt{\sigma_1\,\sigma_2}}$.

By similar methods to that used for Freund distributions, the following results are obtained [8] for the case of Gaussian marginal distributions

Proposition 4.4.1 *The bivariate Gaussian 5-manifold admits a 2-dimensional submanifold through which can be provided a neighbourhood of independence for bivariate Gaussian processes.* □

Corollary 4.4.2 *Via the Central Limit Theorem, by continuity the tubular neighbourhoods of the curve of zero covariance will contain all immersions of limiting bivariate processes sufficiently close to the independence case for all processes with marginals that converge in distribution to Gaussians.* □

Figure 4.4 shows explicitly a tubular neighbourhood for the curve of zero covariance processes ($\sigma_{12} = 0$,) in the submanifold of bivariate Gaussian distributions with zero means and identical standard deviation σ.

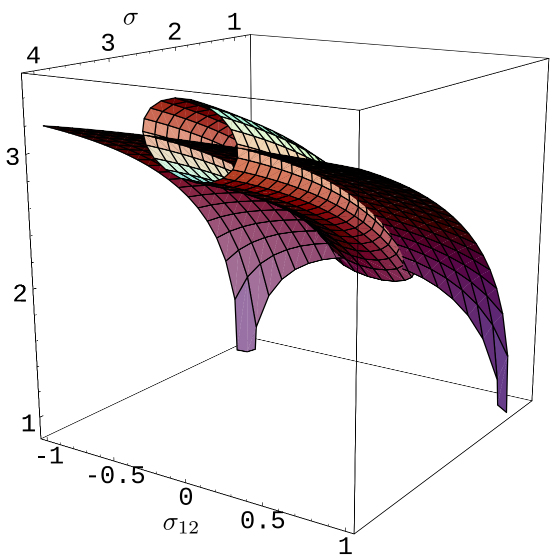

Figure 4.4: *Continuous image, as a surface in* $\mathbb{R}^3$ *using standard coordinates, of an affine immersion for the bivariate Gaussian distributions with zero means and identical standard deviation* σ*. The tubular neighbourhood surrounds the curve of independence cases* ($\sigma_{12} = 0$) *in the surface.*

Chapter 5

Applications

Listed here with their abstracts are some recent applications.

- Khadiga Arwini, L. Del Riego and C.T.J. Dodson. Universal connection and curvature for statistical manifold geometry. Preliminary Report to **American Mathematical Society and Mexican Mathematical Society Joint Meeting** Houston, 12-15 May 2004. *Abstracts American Mathematical Society* March 2004. Final version to appear in **Houston Journal of Mathematics**.

  Statistical manifolds are representations of smooth families of probability density functions that allow differential geometric methods to be applied to problems in stochastic processes, mathematical statistics and information theory. It is common to have to consider a number of linear connections on a given statistical manifold and so it is important to know the corresponding universal connection and curvature; then all linear connections and their curvatures are pullbacks. An important class of statistical manifolds is that arising from the exponential families and one particular family is that of gamma distributions, which we showed recently to have important uniqueness properties in stochastic processes. Here we provide formulae for universal connections and curvatures on exponential families and give an explicit example for the manifold of gamma distributions.

- Khadiga Arwini, C.T.J. Dodson and Hiroshi Matsuzoe. Alpha connections and affine embedding of McKay bivariate gamma 3-manifold. **Internat. J. Pure Appl. Math.** 9, 2 (2003) 253-262.

  The McKay bivariate gamma distribution has marginal gamma densities with positive covariance and recently its information geometry as a 3-manifold has been provided. Here we derive: natural coordinates, explicit expressions for the α-connections, mutually dual foliations and an affine embedding in Euclidean $\mathbb{R}^4$. We compute also the Kullback-Leibler divergence and compare it with the canonical divergence for two McKay densities.

- Khadiga Arwini, C.T.J. Dodson, S. Felipussi and J. Scharcanski. Comparison of distance measures between bivariate gamma processes. Preprint (2003).

  Yue et al.[Yue, S., Ouarda, T.B.M.J. and Bobée, B. 2001. A review of bivariate gamma distributions for hydrological application. *Journal of Hydrology*, 246, 1-4, 1-18] recently reviewed various bivariate gamma distribution models and concluded that they will be useful in hydrology. Here we contribute a detailed study of the McKay bivariate gamma distribution and demonstrate its applicability to the joint probability distribution of void and capillary sizes obtained from soil tomography. The information geometry of the space of McKay bivariate gamma distributions provides a useful mechanism for discriminating between bivariate stochastic processes with positive covariance and gamma marginal distributions. In most cases we found that the information-theoretic metric is more sensitive than the classical Bhattacharyya distance or the Kullback-Leibler divergence; this finding persisted also for data from model porous media, and for data from simulations.

- Khadiga Arwini and C.T.J. Dodson. Information geometry of the Freund bivariate exponential 4-manifold, (2003). *Mathematica* Notebook version in press **The Mathematica Journal**.

  The Freund family of distributions becomes a Riemannian 4-manifold with Fisher information as metric; we derive the induced α-geometry, i.e., the α-curvature, α-Ricci curvature with its eigenvales and eigenvectors, the α-scalar curvature etc. We show that the Freund manifold has a positive constant 0-scalar curvature, so geometrically it constitutes part of a sphere. We consider special cases as submanifolds and discuss their geometrical structures; one submanifold yields examples of neighbourhoods of the independent case for bivariate distributions having identical exponential marginals. Thus, since exponential distributions complement Poisson point processes, we obtain a means to discuss the neighbourhood of independence for random processes.

- Khadiga Arwini and C.T.J. Dodson. Information geometric neighbourhoods of randomness and geometry of the McKay bivariate gamma 3-manifold. **Sankhya: Indian Journal of Statistics**, 66, 2 (2004) 211-231.

  We show that gamma distributions provide models for departures from randomness since every neighbourhood of an exponential distribution contains a neighbourhood of gamma distributions, using an information theoretic metric topology. We derive also the information geometry of the 3-manifold of McKay bivariate gamma distributions, which can provide a metrization of departures from randomness and departures from independence for bivariate processes. The curvature objects are derived, including those on three submanifolds. As in the case of bivariate normal manifolds, we have negative scalar curvature but here it is not constant and we show how it depends on correlation. These results have potential applications, for example, in the characterization of stochastic materials.

- C.T.J. Dodson and Hiroshi Matsuzoe. An affine embedding of the gamma manifold. **Applied Sciences** 5, 1 (2003) 1-6.

  For the space of gamma distributions with Fisher metric and exponential connections, natural coordinate systems, potential functions and an affine immersion in $\mathbb{R}^3$ are provided.

- C.T.J. Dodson and H. Wang. Information geometry for bivariate distribution control. **International Conference on Computer, Communication and Control Technologies (CCCT '03)** Orlando, USA, July 31 to August 2, 2003.

  The optimal control of stochastic processes through sensor estimation of probability density functions has a geometric setting via information theory and the information metric. Information theory identifies the exponential distribution as the maximum entropy distribution if only the mean is known and the gamma distribution if also the mean logarithm is known. Previously, we used the surface representing gamma models to provide an appropriate structure on which to represent the dynamics of a univariate process and algorithms to control it. In this paper we extend these procedures to gamma models with positive correlation, for which the information theoretic 3-manifold geometry has recently been formulated. For comparison we summarize also the case for bivariate Gaussian processes with arbitrary correlation.

- C.T.J. Dodson and J. Scharcanski. Information Geometric Similarity Measurement for Near-Random Stochastic Processes. **IEEE Transactions SMC** A, 33, 4 (2003) 435-440.

- Y. Cai, C.T.J. Dodson, O. Wolkenhauer and A.J. Doig. Information Theoretic Analysis of Protein Sequences shows that Amino Acids Self Cluster. **J. Theoretical Biology** 218, 4 (2002) 409-418.

  We analyse for each of 20 amino acids X the statistics of spacings between consecutive occurrences of X within the well-characterised *Saccharomyces cerevisiae* genome. The occurrences of amino acids may exhibit near random, clustered or smoothed out behaviour, like 1-dimensional stochastic processes along the protein chain. If amino acids are distributed randomly within a sequence then they follow a Poisson process and a histogram of the number of observations of each gap size would asymptotically follow a negative exponential distribution. The novelty of the present approach lies in the use of differential geometric methods to quantify information on sequencing of amino acids and groups of amino acids, via the sequences of intervals between their occurrences. The differential geometry arises from an information-theoretic distance function on the 2-dimensional space of stochastic processes subordinate to gamma distributions—which latter include the random process as a special case. We find that maximum-likelihood estimates of parametric statistics show that all 20 amino acids tend to cluster, some substantially. In other words, the frequencies of short gap lengths tends to be higher and the variance of the gap lengths is greater than expected by chance. This may be because localising amino acids with the same properties may favour secondary structure formation or transmembrane domains. Gap sizes of 1 or 2 are generally disfavoured, 1 strongly so. The only exceptions to this are Gln and Ser, as a result of poly(Gln) or poly(Ser) sequences. There are preferences for gaps of 4 and 7 that can be attributed to α-helices. In particular, a favoured gap of 7 for Leu is found in coiled-coils. Our method contributes to the characterisation of whole sequences by extracting and quantifying stable stochastic features.

- C.T.J. Dodson and H. Wang. Iterative approximation of statistical distributions and relation to information geometry. **J. Statistical Inference for Stochastic Processes** 147, (2001) 307-318.

  The optimal control of stochastic processes through sensor estimation of probability density functions is given a geometric setting via information theory and the information metric. Information theory identifies the exponential distribution as the maximum entropy distribution if only the mean is known and the gamma distribution if also the mean logarithm is known. The surface representing gamma models has a natural Riemannian information metric and the exponential distributions form a 1-dimensional subspace of the 2-dimensional space of all gamma distributions, so we have an isometric embedding of the random model as a subspace of the gamma models. This geometry provides an appropriate structure on which to represent the dynamics of a process and algorithms to control it. We illustrate by showing a minimal information-distance path between an initial and target distribution.

- C.T.J. Dodson. Geometry for stochastically inhomogeneous spacetimes. **Nonlinear Analysis** 47, 5 (2001) 2951-2958.

  Manifolds of gamma probability density functions, which carry a Riemannian information metric, have been used to model the evolution of the intergalactic void statistics and corresponding galactic density statistics. This allowed representation of galaxy clustering using a derived manifold of log-gamma distribution functions to model stochastic spatial variations and provided a metric that measured departures from the random or chaotic state. We show that this log-gamma manifold is an isometric diffeomorph of the gamma manifold and propose a method to approximate matter density statistics in spacetime through a coupling of spacetime and statistical manifold geometries.

- C.T.J. Dodson and S.M. Thompson. A metric space of test distributions for DPA and SZK proofs. *Poster Session*, **Eurocrypt 2000**, Bruges, 14-19 May 2000.

  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 some 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 with compact support on $[0, 1]$, ranging from the uniform distribution to symmetric unimodular distributions of arbitrarily small variance. Illustrative calculations are provided.

- C.T.J. Dodson. Spatial statistics and information geometry for parametric statistical models of galaxy clustering. **Int. J. Theor. Phys.** 38, 10 (1999) 2585-2597.

  Poisson spatial processes of points and of extended objects representing smoothed clusters of galaxies are considered; some results are obtained for planar representations of random filaments, which may help interpret the findings of the Las Campanas Redshift Survey. Based on a model for the void probability function, a family of gamma-related distributions is investigated as a three-dimensional model for the clustering of galaxies. The unclustered models in this family correspond to the random case and to maximum information theoretic entropy. The Riemannian information metric and Gaussian curvature are derived for the parameter space of the family of models, which provides a background on which to write dynamics for cluster evolution.

- C.T.J. Dodson. A geometrical representation for departures from randomness of the inter-galactic void probablity function. **Workshop on Statistics of Cosmological Data Sets** NATO-ASI Isaac Newton Institute 8-13 August 1999.

  A number of recent studies have estimated the inter-galactic void probability function and investigated its departure from various random models. We study a family of parametric statistical models based on gamma distributions, which do give realistic descriptions for other stochastic porous media. Gamma distributions contain as a special case the exponential distributions, which correspond to the 'random' void size probability arising from Poisson processes. The random case corresponds to the information-theoretic maximum entropy or maximum uncertainty model. Lower entropy models correspond on the one hand to more 'clustered' structures or 'more dispersed' structures than expected at random. The space of parameters is a surface with a natural Riemannian structure, the Fisher information metric. This surface contains the Poisson processes as an isometric embedding and provides the geometric setting for quantifying departures from randomness and perhaps on which may be written evolutionary dynamics for the void size distribution. Estimates are obtained for the two parameters of the void diameter distribution for an illustrative example of data published by Fairall.

- C.T.J. Dodson. Information geodesics for communication clustering. **J. Statistical Computation and Simulation** 65, (2000) 133-146.

  Recently, Akyildiz [1, 2] called for further work on non-Poisson models for communication arrivals in distributed networks such as cellular phone systems. The basic 'random' model for stochastic events is the Poisson process; for events on a line this results in an exponential distribution of intervals between events. Network designers and managers need to monitor and quantify call clustering in order to optimize resource usage; the natural reference state from which to measure departures is that arising from a Poisson process of calls. Here we consider gamma distributions, which contain exponential distributions as a special case. The surface representing gamma models has a natural Riemannian information metric and we obtain some geodesic sprays for this metric. The exponential distributions form a 1-dimensional subspace of the

2-dimensional space of all gamma distributions, so we have an isometric embedding of the random model as a subspace of the gamma models. This geometry may provide an appropriate structure on which to represent clustering as quantifiable departures from randomness and on which to impose dynamic control algorithms to optimize traffic at receiving nodes in distributed communication networks. In practice, we may expect correlation between call arrival times and call duration, reflecting for example peaks of different users of internet services. This would give rise to a twisted product of two surfaces with the twisting controlled by the correlation. Though bivariate gamma models do exist, such as Kibble's [43], none has tractable information geometry nor sufficiently general marginal gammas, but a simulation method of approach is suggested.

Appendix: Mathematica Notebooks

1. C.T.J. Dodson. TRYMMA.nb

   Introduction to using Mathematica.

2. C.T.J. Dodson. StatsIntro.nb

   Elementary statistics with Mathematica.

3. Khadiga Arwini and C.T.J. Dodson. Neighbourhoods of randomness and information geometry of the McKay bivariate gamma 3-manifold. In Proc. **International Mathematica Symposium 2003** Imperial College Press, London 2003 pp. 247-254.

   This contains the code necessary to present neighbourhoods of randomness in an embedding of the gamma 2-manifold, and for computing the metric, connection and curvature for the McKay bivariate gamma 3-manifold[7].

4. Khadiga Arwini and C.T.J. Dodson. Information geometry of the Freund bivariate exponential 4-manifold and neighbourhoods of independence, (2003). *Mathematica* Notebook to appear in **The Mathematica Journal**.

   This contains the code necessary to compute the metric, connection and curvature for the Freund 4-manifold and to present neighbourhood of independence in an embedding of a 2-dimensional submanifold of coupled random processes [35].

5. C.T.J. Dodson. Distances in the gamma manifold (2004).

   Code for presenting experimental data for gamma parameters on a distance surface [27, 12].

6. C.T.J. Dodson and W.W. Sampson. McKay pdf (2004).

   Plotting code with arbitrary choice of parameters for the McKay bivariate gamma family.

7. C.T.J. Dodson and W.W. Sampson. Simulator for bivariate gamma processes (2004).

   Code for simulating bivariate gamma processes and for fitting maximum likelihood parameters for McKay distribution.

Bibliography

[1] I. Akyildiz. Mobility management in current and future communication networks. *IEEE Network Mag.* 12, 4 (1998) 39-49.

[2] I. Akyildiz. Performance modeling of next generation wireless systems, Keynote Address, Conference on Simulation Methods and Applications, 1-3 November 1998, Orlando, Florida.

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

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

[5] Flavio S. Anselmetti, Stefan Luthi and Gregor P. Eberli. Quantitative Characterization of Carbonate Pore Systems by Digital Image Analysis. *AAPG Bulletin*, 82, 10, (1998) 1815-1836.

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

[7] Khadiga Arwini and C.T.J. Dodson. Information geometric neighbourhoods of randomness and geometry of the McKay bivariate gamma 3-manifold. *Sankhya: Indian Journal of Statistics* 66, 2 (2004) 211-231.

[8] Khadiga Arwini and C.T.J. Dodson. Neighbourhoods of independence and associated geometry. http://www.ma.umist.ac.uk/kd/PREPRINTS/nhdindep.pdf Preprint (2004).

[9] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their distributions. *Bull. Calcutta Math. Soc.* 35 (1943) 99-110.

[10] Marcelo Biassusi. **Estudo da Deformao de um Vertissolo Atravs da Tomografia Computadorizada de Dupla Energia Simultnea**. *Phd Thesis*, UFRGS - Federal University of Rio Grande do Sul, Porto Alegre, Brazil. February 1996.

[11] K. Borovkov. **Elements of Stochastic Modelling**, World Scientific and Imperial College Press, Singapore and London, 2003.

[12] Y. Cai, C.T.J. Dodson, O. Wolkenhauer and A.J. Doig. Gamma Distribution Analysis of Protein Sequences shows that Amino Acids Self Cluster. *J. Theoretical Biology* 218, 4 (2002) 409-418. http://www.ma.umist.ac.uk/kd/PREPRINTS/amino.pdf

[13] M. Calvo and J.M. Oller. An explicit solution of information geodesic equations for the multivariate normal model. *Statistics & Decisions* 9, (1990) 119-138.

[14] L.A. Cordero, C.T.J. Dodson and M. deLeon. **Differential Geometry of Frame Bundles.** Kluwer, Dordrecht, 1989.

[15] L. Del Riego and C.T.J. Dodson. Sprays, universality and stability. *Math. Proc. Camb. Phil. Soc.* 103(1988), 515-534.

[16] M. Deng. **Differential Geometry in Statistical Inference** PhD thesis, Department of Statistics, Pennsylvania State University, 1990.

[17] C.T.J. Dodson. Systems of connections for parametric models. In **Proc. Workshop on Geometrization of Statistical Theory** 28-31 October 1987. Ed. C.T.J. Dodson, ULDM Publications, University of Lancaster, 1987, pp 153-170.

[18] C.T.J. Dodson. Gamma manifolds and stochastic geometry. In: **Proceedings of the Workshop on Recent Topics in Differential Geometry**, Santiago de Compostela 16-19 July 1997. *Public. Depto. Geometría y Topología* 89 (1998) 85-92.

[19] C.T.J. Dodson. Information geodesics for communication clustering. *J. Statistical Computation and Simulation* 65, (2000) 133-146.

[20] C.T.J. Dodson. Evolution of the void probability function. Presented at **Workshop on Statistics of Cosmological Data Sets**, 8-13 August 1999, Isaac Newton Institute, Cambridge. http://www.ma.umist.ac.uk/kd/PREPRINTS/vpf.ps.

[21] C.T.J. Dodson. Spatial statistics and information geometry for parametric statistical models of galaxy clustering. *Int. J. Theor. Phys.*, 38, 10, (1999) 2585-2597.

[22] C.T.J. Dodson. Geometry for stochastically inhomogeneous spacetimes. *Nonlinear Analysis*, 47 (2001) 2951-2958.
http://www.ma.umist.ac.uk/kd/PREPRINTS/gsis.ps.

[23] C.T.J. Dodson and Hiroshi Matsuzoe. An affine embedding of the gamma manifold. *InterStat*, January 2002, 2 (2002) 1-6. http://www.ma.umist.ac.uk/kd/PREPRINTS/affimm.pdf.

[24] C.T.J. Dodson and M. Modugno. Connections over connections and universal calculus. In Proc. **VI Convegno Nazionale di Relativita General a Fisic Della Gravitazione** Florence, 10-13 Octobar 1984, Eds. R. Fabbri and M. Modugno, pp. 89-97, Pitagora Editrice, Bologna, 1986.

[25] C.T.J. Dodson and T. Poston. **Tensor Geometry** Graduate Texts in Mathematics 130, Second edition, Springer-Verlag, New York, 1991.

[26] C.T.J. Dodson and W.W. Sampson. Modeling a class of stochastic porous media. *App. Math. Lett.* 10, 2 (1997) 87-89.

[27] C.T.J. Dodson and J. Scharcanski. Information Geometric Similarity Measurement for Near-Random Stochastic Processes. *IEEE Transactions on Systems, Man and Cybernetics - Part A* 33, 4 (2003) 435-440.

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

[29] F. Downton. Bivariate exponential distributions in reliability theory. *J. Royal Statist. Soc. Series B* 32 (1970) 408-417.

[30] A.P. Fairall. **Large-scale structure in the universe.** Wiley-Praxis, Chichester 1998.

[31] W. Feller. **An Introduction to Probability Theory and its Applications.** Volume 1, John Wiley, Chichester 1968.

[32] W. Feller. **An Introduction to Probability Theory and its Applications.** Volume 2, John Wiley, Chichester 1971.

[33] M. Fisz. **Probability Theory and Mathematical Statistics.** 3^{rd} edition, John Wiley, Chichester 1963.

[34] R.A. Fisher. Theory of statistical estimation. *Proc. Camb. Phil. Soc.* 122 (1925) 700-725.

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

[36] A. Gray **Modern Differential Geometry of Curves and Surfaces** 2^{nd} Edition, CRC Press, Boca Raton 1998.

[37] R.C. Griffiths. The canonical correlation coefficients of bivariate gamma distributions. *Annals Math. Statist.* 40, 4 (1969) 1401-1408.

[38] F.A. Haight. **Handbook of the Poisson Distribution** J. Wiley, New York, 1967.

[39] 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 (1999) 749-753.

[40] T.P. Hutchinson and C.D. Lai. **Continuous Multivariate Distributions, Emphasising Applications**, Rumsby Scientific Publishing, Adelaide 1990.

[41] E.T. Jaynes. Information theory and statistical inference. *The Physical Review* 106 (1957) 620-630 and 108 (1957) 171-190.

[42] M. Kendall and A. Stuart. **The Advanced Theory of Statistics, Volume 2 Inference and Relationship** 4^{th} Edition. Charles Griffin, London, 1979.

[43] W.F. Kibble. A two variate gamma-type distribution. *Sankhyā* 5 (1941) 137-150.

[44] S. Kokoska and C. Nevison. **Statistical Tables and Formulae** Springer Texts in Statistics, Springer-Verlag, New York 1989.

[45] Samuel Kotz, N.Balakrishnan and N.Johnson. **Continuous Multivariate Distributions** 2^{nd} Edition, Volume 1 (2000).

[46] S. Kullback. **Information and Statistics**, J. Wiley, New York, 1959.

[47] T. Kurose. On the divergences of 1-conformally flat statistical manifolds. *Tôhoku Math. J.*, 46 (1994) 427-433.

[48] S.L. Lauritzen. Statistical Manifolds. In **Differential Geometry in Statistical Inference**, Institute of Mathematical Statistics Lecture Notes, Volume 10, Berkeley 1987, pp 163-218.

[49] C.D. Lai. Constructions of bivariate distributions by a generalized trivariate reduction technique. *Statistics and Probability Letters* 25, 3 (1995) 265-270.

[50] L.Mangiarotti and M.Modugno. Fibred spaces, jet spaces and connections for field theories. In Proc. International Meeting on **Geometry and Physics,** Florence, 12-15 October 1982, ed. M.Modugno, Pitagora Editrice, Bologna, 1983 pp 135-165.

[51] K.V. Mardia. **Families of Bivariate Distributions**. Griffin, London 1970.

[52] A.W. Marshall and I. Olkin. A generalized bivariate exponential distribution. *J. Appl. Prob.* 4 (1967) 291-302.

[53] H. Matsuzoe. On realization of conformally-projectively flat statistical manifolds and the divergences. *Hokkaido Math. J.*, 27 (1998) 409-421

[54] H. Matsuzoe. Geometry of contrast functions and conformal geometry. *Hiroshima Math. J.*, 29 (1999) 175-191.

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

[56] G.K. Miller and U.N. Bhat. Estimation for renewal processes with unobservable gamma or Erlang interarrival times. *J. Statistical Planning and Inference* 61, 2 (1997) 355-372.

[57] M. Modugno. Systems of vector valued forms on a fibred manifold and applications to gauge theories. In Proc. Conference **Differential Geomeometric Methods in Mathematical Physics**, Salamanca 1985, Lecture Notes in Mathematics 1251, Springer-Verlag, Berlin 1987, pp. 238-264.

[58] K. Nomizu and T. Sasaki. **Affine differential geometry: Geometry of Affine Immersions**. Cambridge University Press, Cambridge, 1994.

[59] S.H. Ong. Computation of bivariate-gamma and inverted-beta distribution functions. *J. Statistal Computation and Simulation* 51, 2-4 (1995) 153-163.

[60] A. Papoulis. **Probability, Random Variables and Stochastic Processes** 3^{rd} edition, McGraw-Hill, New York 1991.

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

[62] S. Roman. **Coding and Information Theory**, Graduate Texts in Mathematics, 134 Springer-Verlag, New York, 1992, cf. Part 1.

[63] Colin Rose and Murray D.Smith. **Mathematical Statistics with Mathematica** Springer texts in statistics, Springer-Verlag, Berlin 2002.

[64] C.E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.* 27, (1948) 379-423 and 623-656.

[65] D. Slepian, ed. **Key papers in the development of information theory**, IEEE Press, New York, 1974.

[66] Y.Sato, K.Sugawa and M.Kawaguchi. The geometrical structure of the parameter space of the two-dimensional normal distribution. Division of information engineering, Hokkaido University, Sapporo, Japan (1977).

[67] D. Shi and C.D. Lai. Fisher information for Downton's bivariate exponential distribution. *J. Statistical Computation and Simulation* 60, 2 (1998) 123-127.

[68] L.T. Skovgaard. A Riemannian geometry of the multivariate normal model. *Scand. J. Statist.* 11 (1984) 211-223.

[69] D. Stoyan, W.S. Kendall and J. Mecke. **Stochastic Geometry and its Applications** 2^{nd} Edition, John Wiley, Chichester, 1995.

[70] M. Tribus. **Thermostatics and Thermodynamics** D. Van Nostrand and Co., Princeton N.J., 1961.

[71] M. Tribus, R. Evans and G. Crellin. The use of entropy in hypothesis testing. In Proc. **Tenth National Symposium on Reliability and Quality Control** 7-9 January 1964.

[72] H. J. Vogel, A. Kretzchmar. **Topological Characterization of Pore Space in Soil - Sample Preparation and Digital Image-Processing**. *Geoderma*, v. 73, pp. 23-18. 1996.

[73] S. Wolfram. **The Mathematica Book** 3^{rd} edition, Cambridge University Press, Cambridge, 1996.