Chapter 2

Explicit Examples

This chapter presents a few examples of usual statistical models (normal, lognormal, beta, gamma, Bernoulli, and geometric) for which we provide the Fisher metric explicitly and, if possible, the geodesics and α-autoparallel curves. Some Fisher metrics will involve the use of non-elementary functions, such as the digamma and trigamma functions.

A distinguished role is dedicated to the normal distribution, which is associated with a manifold of negative constant curvature (hyperbolic space) and to the multinomial geometry, which corresponds to a space with positive constant curvature (spherical space).

2.1   The Normal Distribution

In this section we shall determine the geodesics with respect to the Fisher information metric of a family of normal distributions. Given two distributions of the same family, the geodesics are curves of minimum information joining the distributions. We shall see that such a curve always exists between any two distributions on a normal family. This is equivalent with the possibility of deforming one distribution into the other by keeping the change of information to a minimum.

2.1.1   The Fisher Metric

Recall the formula for the density of a normal family

$$p(x, \xi) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \;\; x \in \mathcal{X} = \mathbb{R},$$

with parameters $(\xi^1, \xi^2) = (\mu, \sigma) \in \mathbb{R} \times (0, \infty)$. Using Proposition 1.6.3 we obtain the following components for the Fisher–Riemann metric.

Proposition 2.1.1 *The Fisher information matrix for the normal distribution is given by*

$$g_{ij} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}. \tag{2.1.1}$$

For the computation details see Problem 2.1. It is worth noting that the metric does not depend on μ, i.e., it is translation invariant. This metric is also very similar to the upper-half plane metric.

2.1.2   The Geodesics

A straightforward computation shows that the nonzero Christoffel symbols of first and second kind are:

$$\Gamma_{11,2} = \frac{1}{\sigma^3}, \quad \Gamma_{12,1} = -\frac{1}{\sigma^3}, \quad \Gamma_{22,2} = -\frac{2}{\sigma^3}$$

$$\Gamma_{ij}^1 = \begin{pmatrix} 0 & -\frac{1}{\sigma} \\ -\frac{1}{\sigma} & 0 \end{pmatrix}, \qquad \Gamma_{ij}^2 = \begin{pmatrix} \frac{1}{2\sigma} & 0 \\ 0 & -\frac{1}{\sigma} \end{pmatrix}.$$

Consequently, the geodesics equations (1.13.43) are solutions of a Riccati ODE system

$$\ddot{\mu} - \frac{2}{\sigma}\dot{\mu}\dot{\sigma} = 0 \tag{2.1.2}$$

$$\ddot{\sigma} + \frac{1}{2\sigma}(\dot{\mu})^2 - \frac{1}{\sigma}(\dot{\sigma})^2 = 0. \tag{2.1.3}$$

Separating and integrating in the first equation yields

$$\frac{\ddot{\mu}}{\dot{\mu}} = \frac{2\dot{\sigma}}{\sigma} \iff \frac{d}{ds}\ln\dot{\mu} = 2\frac{d}{ds}\ln\sigma \iff \dot{\mu} = c\sigma^2,$$

with c constant. We solve the equation in the following two cases:

   1. The case $c = 0$. It follows that $\mu = $ constant, which corresponds to vertical half lines. Then σ satisfies the equation $\ddot{\sigma} = \frac{1}{\sigma}\dot{\sigma}^2$. Writing the equation as

$$\frac{\ddot{\sigma}}{\dot{\sigma}} = \frac{\dot{\sigma}}{\sigma}$$

and integrating yields $\ln \dot{\sigma} = \ln(C\sigma)$, with C constant. Integrating again, we find $\sigma(s) = Ke^{Cs}$. Hence, the geodesics in this case have the following explicit equations

$$\mu \;\; = \;\; c \tag{2.1.4}$$

$$\sigma(s) \;\; = \;\; Ke^{Cs}, \tag{2.1.5}$$

with $c, C \in \mathbb{R}$, $K > 0$ constants.

2. *The case $c \neq 0$.* Substituting $\dot{\mu} = x\sigma^2$ in Eq. (2.1.3), we obtain the following equation in σ

$$\sigma\ddot{\sigma} + \frac{c^2}{2}\sigma^4 - (\dot{\sigma})^2 = 0. \tag{2.1.6}$$

Let $\dot{\sigma} = u$. Then $\ddot{\sigma} = \dfrac{du}{d\sigma}u$ and (2.1.6) becomes

$$\sigma\frac{du}{d\sigma}u + \frac{c^2}{2}\sigma^4 - u^2 = 0.$$

Multiplying by the integrant factor $\dfrac{1}{\sigma^3}$ leads to the exact equation

$$\underbrace{\frac{u}{\sigma^2}}_{=M}\,du + \underbrace{\left(\frac{c^2}{2}\sigma - \frac{u^2}{\sigma^3}\right)}_{N}d\sigma = 0,$$

since

$$\frac{\partial M}{\partial \sigma} = \frac{\partial N}{\partial u} = -2u\sigma^{-3}.$$

Then there is a function $f(\sigma, u)$ such that $df = 0$, with

$$\frac{\partial f}{\partial u} = M, \qquad \frac{\partial f}{\partial \sigma} = N.$$

Integrating in the first equation yields

$$f(\sigma, u) = \frac{u^2}{2\sigma^2} + h(\sigma),$$

with function h to be determined in the following. Differentiating with respect to σ in the above equation,

$$\frac{\partial f}{\partial \sigma} = -\frac{u^2}{\sigma^3} + h'(\sigma),$$

and comparing with

$$\frac{\partial f}{\partial \sigma} = N = \frac{c^2}{2}\sigma - \frac{u^2}{\sigma^3},$$

we get

$$h'(\sigma) = \frac{c^2}{2}\sigma \implies h(\sigma) = \frac{c^2\sigma^2}{4} + c_0,$$

with c_0 constant. Hence, a first integral for the system is

$$f(\sigma, u) = \frac{u^2}{2\sigma^2} + \frac{c^2\sigma^2}{4} = \frac{E}{2},$$

with E positive constant. Solving for u, we obtain

$$\frac{u^2}{\sigma^2} + \frac{c^2\sigma^2}{2} = E \iff$$
$$\frac{\dot\sigma}{\sigma} = \frac{c}{\sqrt{2}}\sqrt{C^2 - \sigma^2},$$

where $C^2 = 2E/c^2$. Separating and integrating, we find

$$\int \frac{d\sigma}{\sigma\sqrt{C^2 - \sigma^2}} = (s + s_0)\frac{c}{\sqrt{2}}.$$

Using the value of the integral

$$\int \frac{dx}{x\sqrt{C^2 - x^2}} = -\frac{1}{\sqrt{C}}\tanh^{-1}\sqrt{1 - \left(\frac{x}{C}\right)^2},$$

we obtain

$$-\frac{1}{\sqrt{C}}\tanh^{-1}\sqrt{1 - \left(\frac{\sigma}{C}\right)^2} = (s + s_0)\frac{c}{\sqrt{2}}.$$

Solving for σ, we get

$$\sigma = c\sqrt{1 - \tanh^2\left(\sqrt{E}(s + s_0)\right)} = \frac{c}{\cosh\left(\sqrt{E}(s + s_0)\right)}.$$

In order to find μ we integrate in $\dot\mu = c\sigma^2$ and obtain

$$\mu(s) = \int \frac{c^3}{\cosh^2\left(\sqrt{E}(s + s_0)\right)}\, ds = \frac{c^3}{\sqrt{E}}\tanh\left(\sqrt{E}(s + s_0)\right) + K.$$

Since we have

$$\sigma(s)^2 + (\mu(s) - K)^2 \frac{E}{c^4} = c^2,$$

the geodesics will be half-ellipses, with $\sigma > 0$.

In the case $c = 0$, the unknown σ satisfies

$$\frac{\dot{\sigma}}{\sigma} = \sqrt{E} \iff \frac{d}{ds} \ln \sigma = \sqrt{E}$$

with solution

$$\sigma(s) = \sigma(0) e^{\sqrt{E}s},$$

while μ is constant, $\mu = K$. The geodesics in this case are vertical half-lines.

Proposition 2.1.2 *Consider two normal distributions with equal means, $\mu_0 = \mu_1$, and distinct standard deviations σ_0 and σ_1. Then the smallest information transform, which deforms the first distribution into the second one, is a normal distribution with constant mean and standard deviation*

$$\sigma(s) = \sigma_1^{s/\tau} \sigma_0^{1-s/\tau}, \quad s \in [0, \tau].$$

Proof: The geodesic in this case is a vertical half-line with constant mean and $\sigma(s) = \sigma(0)e^{\sqrt{E}s}$. The amount $\sqrt{E}$ can be found from the boundary condition $\sigma(\tau) = \sigma_1$. ∎

Let $x_0 = \ln \sigma_0$, $x_1 = \ln \sigma_1$, and $x(s) = \ln \sigma(s)$. Then $x(s) = \frac{s}{\tau}x_0 + \left(1 - \frac{s}{\tau}\right)x_1$, which corresponds to a line segment. The minimal information loss during the deformation occurs when the log-likelihood function describes a line segment.

2.1.3 α-Autoparallel Curves

A straightforward computation, using (1.11.34), yields the following Christoffel coefficients of first kind

$$\Gamma_{11,1}^{(\alpha)} = \Gamma_{21,2}^{(\alpha)} = \Gamma_{12,2}^{(\alpha)} = \Gamma_{22,1}^{(\alpha)} = 0$$

$$\Gamma_{11,2}^{(\alpha)} = \frac{1-\alpha}{\sigma^3}, \quad \Gamma_{12,1}^{(\alpha)} = \Gamma_{21,1}^{(\alpha)} = -\frac{1+\alpha}{\sigma^3}, \quad \Gamma_{22,2}^{(\alpha)} = -\frac{2(1+2\alpha)}{\sigma^3}.$$

The Christoffel symbols of second kind are obtained by rising indices

$$\Gamma_{ij}^{1\,(\alpha)} = g^{11}\Gamma_{ij,1}{}^{(\alpha)} + g^{12}\Gamma_{ij,2}{}^{(\alpha)} = \sigma^2\Gamma_{ij,1}{}^{(\alpha)}$$

$$= \sigma^2 \begin{pmatrix} 0 & -\frac{1+\alpha}{\sigma^3} \\ -\frac{1+\alpha}{\sigma^3} & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1+\alpha}{\sigma} \\ -\frac{1+\alpha}{\sigma} & 0 \end{pmatrix}.$$

$$\Gamma_{ij}^{2\,(\alpha)} = g^{21}\Gamma_{ij,1}{}^{(\alpha)} + g^{22}\Gamma_{ij,2}{}^{(\alpha)}$$

$$= \frac{\sigma^2}{2} \begin{pmatrix} \frac{1-\alpha}{\sigma^3} & 0 \\ 0 & -2\frac{1+2\alpha}{\sigma^3} \end{pmatrix} = \begin{pmatrix} \frac{1-\alpha}{2\sigma} & 0 \\ 0 & -\frac{1+2\alpha}{\sigma} \end{pmatrix}.$$

The Riccati equations (1.13.43) for the α-autoparallel curves are

$$\ddot{\mu} - \frac{2(1+\alpha)}{\sigma}\dot{\sigma}\dot{\mu} = 0$$

$$\ddot{\sigma} + \frac{1-\alpha}{2\sigma}\dot{\mu}^2 - \frac{1+2\alpha}{\sigma}\dot{\sigma}^2 = 0.$$

The first equation can be transformed as in the following

$$\frac{\ddot{\mu}}{\dot{\mu}} = 2(1+\alpha)\frac{\dot{\sigma}}{\sigma} \Longleftrightarrow$$

$$\frac{d}{ds}\ln\dot{\mu} = 2(1+\alpha)\frac{d}{ds}\ln\sigma \Longleftrightarrow$$

$$\ln\dot{\mu} = 2(1+\alpha)\ln\sigma + c_0 \Longleftrightarrow$$

$$\dot{\mu} = c\,\sigma^{2(1+\alpha)},$$

with c constant. Substituting in the second equation yields

$$\ddot{\sigma} + \frac{1-\alpha}{2\sigma}c^2\sigma^{4(1+\alpha)} - \frac{1+2\alpha}{\sigma}\dot{\sigma}^2 = 0,$$

which after the new substitution $u = \dot{\sigma}$ writes as

$$\frac{du}{d\sigma}u + \frac{1-\alpha}{2\sigma}c^2\sigma^{4(1+\alpha)} - \frac{1+2\alpha}{\sigma}u^2 = 0.$$

Multiplying the equation by an integral factor of the form σ^{k+1}, we obtain

$$\underbrace{\sigma^{k+1}u}_{=M}\,du + \underbrace{\left(\frac{1-\alpha}{2}c^2\sigma^{4(\alpha+1)+k} - (1+2\alpha)\sigma^k u^2\right)}_{=N}d\sigma = 0.$$

From the closeness condition

$$\frac{\partial M}{\partial \sigma} = \frac{\partial N}{\partial u},$$

we determine $k + 1 = -(4\alpha + 2)$. The exact equation we need to solve is

$$u\sigma^{-(4\alpha+2)} \, du + \left(\frac{1 - \alpha}{2}c^2\sigma - (1 + 2\alpha)u^2\sigma^{-(4\alpha+3)}\right) d\sigma = 0.$$

We need to determine a function f that satisfies the system

$$\frac{\partial f}{\partial u} = u\sigma^{-(4\alpha+2)}$$

$$\frac{\partial f}{\partial \sigma} = \frac{1 - \alpha}{2}c^2\sigma - (1 + 2\alpha)u^2\sigma^{-(2\alpha+3)}.$$

From the first equation, we have

$$f = \frac{u^2}{2}\sigma^{-(4\alpha+2)} + h(\sigma) \Longrightarrow \frac{\partial f}{\partial \sigma} = -(1 + 2\alpha)u^2\sigma^{-(4\alpha+3)} + h'(\sigma)$$

and comparing with the second equation yields

$$h'(\sigma) = \frac{1 - \alpha}{2}c^2\sigma \Longrightarrow h(\sigma) = \frac{(1 - \alpha)c^2}{4}\sigma^2 + C.$$

Hence, a first integral of motion is given by

$$f = \frac{u^2}{2}\sigma^{-(4\alpha+2)} + \frac{1 - \alpha}{4}c^2\sigma^2 = \frac{E}{2},$$

with E constant. Next we shall solve for σ. Using that $u = \dot{\sigma}$, we have

$$\frac{u^2}{\sigma^{2(2\alpha+1)}} + \frac{1 - \alpha}{2}c^2\sigma^2 = E \Longleftrightarrow$$

$$\left(\frac{\dot{\sigma}}{\sigma^{2\alpha+1}}\right)^2 + \frac{1 - \alpha}{2}c^2\sigma^2 = E \Longleftrightarrow$$

$$\left(\frac{\dot{\sigma}}{\sigma^{2\alpha+1}}\right)^2 = E - \frac{1 - \alpha}{2}c^2\sigma^2 \Longleftrightarrow$$

$$\int \frac{d\sigma}{\sigma^{2\alpha+1}\sqrt{E - \frac{1-\alpha}{2}c^2\sigma^2}} = \pm s + s_0 \Longleftrightarrow$$

$$\int \frac{d\sigma}{\sigma^{2\alpha+1}\sqrt{C^2 - \sigma^2}} = (\pm s + s_0)\sqrt{\frac{1 - \alpha}{2}}c, \qquad (2.1.7)$$

where

$$C = C_\alpha = \frac{2E}{c}\frac{1}{1 - \alpha}.$$

The left side integral can be transformed using the substitutions $t = \sigma^2$, $v = \sqrt{C^2 - t}$ as follows

$$\int \frac{d\sigma}{\sigma^{2\alpha+1}\sqrt{C^2 - \sigma^2}} = \int \frac{dt}{2\sigma^{2(\alpha+1)}\sqrt{C^2 - \sigma^2}} = \int \frac{dt}{2t^{\alpha+1}\sqrt{C^2 - t}}$$

$$= \int \frac{-2v\,dv}{2t^{\alpha+1}v} = -\int \frac{dv}{(C^2 - v^2)^{\alpha+1}},$$

and hence (2.1.7) becomes

$$-\int \frac{dv}{(C^2 - v^2)^{\alpha+1}} = (\pm s + s_0)\sqrt{\frac{1 - \alpha}{2}}\, c. \qquad (2.1.8)$$

The μ-component is given by

$$\mu = c \int \sigma^{2(1+\alpha)}(s)ds. \qquad (2.1.9)$$

There are a few particular values of α for which this equation can be solved explicitly.

Case $\alpha = -1$

Equation (2.1.8) becomes

$$-v - K = (\pm s + s_0)\sqrt{\frac{1 - \alpha}{2}}\, c,$$

with solution

$$\sigma^2(s) = C^2 - \left((\pm s + s_0)\sqrt{\frac{1 - \alpha}{2}}\, c + K\right)^2,$$

for K constant. Equation (2.1.3) easily yields

$$\mu(s) = cs + \mu(0).$$

Case $\alpha = 1/2$

Since

$$\int \frac{dv}{\left(C^2 - v^2\right)^{3/2}} = \frac{v}{C^2\sqrt{C^2 - v^2}},$$

we solve

$$-\frac{v}{C^2\sqrt{C^2 - v^2}} = (\pm s + s_0)\frac{c}{2} + K$$

and obtain

$$\sigma(s) = \frac{C}{\sqrt{1 + C^4 \left((\pm s + s_0)\frac{c}{2} + K \right)^2}}.$$

The μ-component is given by the integral

$$\mu(s) = c \int \sigma^3(s) \, ds.$$

2.2 Jeffrey's Prior

In the following we shall compute the prior on the statistical model

$$\mathcal{S}_\mu = \{p_\xi; E[p_\xi = \mu], Var[p_\xi] > 1\} = \{p_{(\mu,\sigma)}; \sigma > 1\}$$

which represents a vertical half line in the upper-half plane. The determinant is

$$G(\xi) = \det g_{ij}(\xi) = \det \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix} = \frac{2}{\sigma^4}.$$

Then the volume is computed as

$$Vol(\mathcal{S}_\mu) = \int_1^\infty \sqrt{G(\xi)} \, d\sigma = \int_1^\infty \frac{\sqrt{2}}{\sigma^2} \, d\sigma = \sqrt{2} < \infty.$$

Therefore the prior on $\mathcal{S}_\mu$ is given by

$$Q(\sigma) = \frac{\sqrt{G(\sigma)}}{Vol(\mathcal{S}_\mu)} = \frac{1}{\sigma^2}.$$

2.3 Lognormal Distribution

In the case of lognormal distribution

$$p_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \qquad x > 0,$$

the Fisher information matrix (Fisher–Riemann metric) is given by

$$g = \begin{pmatrix} g_{\mu\mu} & g_{\mu\sigma} \\ g_{\sigma\mu} & g_{\sigma\sigma} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}.$$

The computation details are left for the reader and are the subject of Problem 2.2. It is worth noting that this coincides with the Fisher metric of a normal distribution model. Hence, the associated geodesics are vertical half lines or halfs of ellipses.

2.4   Gamma Distribution

In this case the statistical model is defined by the following family of densities

$$p_\xi(x) = p_{\alpha,\beta}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \, x^{\alpha-1} e^{-x/\beta},$$

with $(\alpha, \beta) \in (0, \infty) \times (0, \infty)$, $x \in (0, \infty)$. In the study of this model we need some special functions. Let

$$\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}, \qquad \psi_1(\alpha) = \psi'(\alpha) \qquad\qquad (2.4.10)$$

be the *digamma* and the *trigamma* functions, respectively. Differentiating in the Dirichlet's integral representation (see Erdélyi [42] vol. I, p. 17)

$$\psi(\alpha) = \int_0^\infty [e^{-t} - (1+t)^{-\alpha}] t^{-1} \, dt, \qquad\qquad \alpha > 0$$

yields the following integral expression for the trigamma function

$$\psi_1(\alpha) = \psi'(\alpha) = \int_0^\infty \frac{\ln(1+t)}{t(1+t)^\alpha} \, dt. \qquad\qquad (2.4.11)$$

Another interesting formula is the expression of the trigamma function as a *Hurwitz zeta function*

$$\psi_1(\alpha) = \zeta(2, \alpha) = \sum_{n \geq 0} \frac{1}{(\alpha + n)^2}, \qquad\qquad (2.4.12)$$

which holds for $\alpha \notin \{0, -1, -2, -3, \dots\}$, relation obviously satisfied in our case since $\alpha > 0$.

Then the components of the Fisher–Riemann metric are obtained from Proposition 1.6.3, using the relations

$$\int_0^\infty p_\xi(x) \, dx = 1, \qquad \int_0^\infty x p_\xi(x) \, dx = \alpha\beta$$

and the derivatives of the log-likelihood function that are asked to be computed in Problem 2.2:

$$
\begin{aligned}
g_{\alpha\alpha} &= -E[\partial_\alpha^2 \ell_x(\xi)] = \int_0^\infty \psi'(\alpha) p_\xi(x)\, dx = \psi'(\alpha) = \psi_1(\alpha), \\
g_{\beta\beta} &= -E[\partial_\beta^2 \ell_x(\xi)] = -\int_0^\infty \left(\frac{\alpha}{\beta^2} - \frac{2x}{\beta^3} \right) p_\xi(x)\, dx \\
&= -\frac{\alpha}{\beta^2} + \frac{2}{\beta^3} \int_0^\infty x p_\xi(x)\, dx = \frac{\alpha}{\beta^2}, \\
g_{\alpha\beta} &= -E[\partial_{\alpha\beta} \ell_x(\xi)] = \int_0^\infty \frac{1}{\beta} p_\xi(x)\, dx = \frac{1}{\beta}.
\end{aligned}
$$

Proposition 2.4.1 *The Fisher information matrix (Fisher–Riemann metric) for the gamma distribution is*

$$
g = \begin{pmatrix} \psi_1(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{pmatrix} = \begin{pmatrix} \displaystyle\sum_{n\geq 0} \frac{1}{(\alpha+n)^2} & \frac{1}{\beta} \\ \frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{pmatrix}.
$$

It is worth noting that here α is the parameter for the gamma distribution and it has nothing to do with α-connections.

2.5 Beta Distribution

The Fisher information metric for the beta distribution

$$
p_{a,b} = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \qquad a, b > 0, x \in [0,1]
$$

will be worked in terms of trigamma functions. Since the beta function

$$
B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx
$$

can be expressed in terms of gamma functions as

$$
B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},
$$

then its partial derivatives can be written in terms of digamma functions, using relation (2.11.17), see Problem 2.4, part (*a*).

The log-likelihood function and its partial derivatives are left for the reader as an exercise in Problem 2.4, parts (*b*) and (*c*). Since the

second partial derivatives of $\ell(a, b)$ do not depend on x, they will coincide with their own expected values. It follows the next components for the Fisher–Riemann metric:

$$
\begin{aligned}
g_{aa} &= -E[\partial_a^2 \ell(a, b)] = \psi_1(a) - \psi_1(a + b) \\
g_{bb} &= -E[\partial_b^2 \ell_x(a, b)] = \psi_1(b) - \psi_1(a + b) \\
g_{ab} &= g_{ba} = -E[\partial_a \partial_b \ell_x(a, b)] = -\psi_1(a + b).
\end{aligned}
$$

Proposition 2.5.1 *The Fisher information matrix (Fisher–Riemann metric) for the beta distribution is given by*

$$
g = \begin{pmatrix} \psi_1(a) - \psi_1(a + b) & -\psi_1(a + b) \\ -\psi_1(a + b) & \psi_1(b) - \psi_1(a + b) \end{pmatrix},
$$

where ψ_1 stands for the trigamma function.

2.6   Bernoulli Distribution

Consider the sample space $\mathcal{X} = \{0, 1, \ldots, n\}$ and parameter space $\mathbb{E} = [0, 1]$. The Bernoulli, or binomial distribution, is given by

$$
p(k; \xi) = \binom{n}{k} \xi^k (1 - \xi)^{n-k},
$$

where the parameter ξ denotes the success probability. Then $\mathcal{S} = \{p_\xi; \xi \in [0, 1]\}$ becomes a one-dimensional statistical model. The derivatives of the log-likelihood function $\ell_k(\xi) = \ln p(k; \xi)$ are proposed as an exercise in Problem 2.5. Then the Fisher information is given by the function

$$
\begin{aligned}
g_{11}(\xi) &= -E_\xi[\partial_\xi^2 \ell(\xi)] = \sum_{k=0}^n p(k; \xi) \partial_\xi^2 \ell_k(\xi) \\
&= \sum_{k=0}^n \frac{k}{\xi^2} p(k; \xi) + \sum \frac{(n - k)}{(1 - \xi)^2} p(k; \xi) \\
&= \frac{n}{\xi} + \frac{n(1 - \xi)}{(1 - \xi)^2} = \frac{n}{\xi(1 - \xi)},
\end{aligned}
$$

where we used that the mean of a Bernoulli distribution is $n\xi$. Using that the variance is $n\xi(1 - \xi)$, it follows that

$$
g_{11}(\xi) = \frac{n^2}{Var(p_\xi)},
$$

which is a Cramér–Rao type identity corresponding to an efficient estimator.

2.7 Geometric Probability Distribution

Let $\mathcal{X} = \{1, 2, 3, \dots\}$, $\mathbb{E} = [0, 1]$ and consider $p(k; \xi) = (1 - \xi)^{k-1}\xi$, $k \in \mathcal{X}$, $\xi \in \mathbb{E}$. The formulas for the partial derivatives of the log-likelihood function are left as an exercise for the reader in Problem 2.6. Then the Fisher information becomes

$$
\begin{aligned}
g_{11}(\xi) &= -E_\xi[\partial_\xi^2 \ell(\xi)] \\
&= \sum_{k \geq 1} \frac{(k-1)p(k; \xi)}{(\xi - 1)^2} + \sum_{k \geq 1} \frac{1}{\xi^2} p(k; \xi) \\
&= \frac{1}{\xi^2(1 - \xi)},
\end{aligned}
$$

where we used the expression for the mean $\sum_{k \geq 1} k\, p(k; \xi) = \frac{1}{\xi}$.

2.8 Multinomial Geometry

In this section we investigate the geometry associated with the multinomial probability distribution. The computation performed here is inspired from Kass and Vos [49]. Consider m independent, identical trials with n possible outcomes. The probability that a single trial falls into class i is p_i, $i = 1, 2, \dots, n$, and remains the same from trial to trial. Since $p_1 + \cdots + p_n = 1$, the parameter space is given by the $(n-1)$-dimensional simplex

$$
\mathbb{E} = \{(p_1, \dots, p_{n-1}); 0 \leq p_i \leq 1, \sum_{i=1}^{n-1} p_i = 1\}.
$$

It is advantageous to consider the new parameterization

$$
z_i = 2\sqrt{p_i}, \qquad i = 1, \dots, n.
$$

Then $\sum_{i=1}^{n} z_i^2 = 4$, and hence

$$
z \in \mathbb{S}_{2,+}^{n-1} = \{z \in \mathbb{R}^n; \|z\|^2 = 4, z_i \geq 0\}.
$$

Therefore, the statistical manifold of multinomial probability distributions can be identified with $\mathbb{S}_{2,+}^{n-1}$, the positive portion of the $(n-1)$-dimensional sphere of radius 2. The Fisher information matrix with respect to a local coordinate system (ξ^i) is

$$g_{rs}(\xi) = 4\sum_{i=1}^{n} \partial_r \sqrt{p_i(\xi)}\partial_s \sqrt{p_i(\xi)}$$

$$= \sum_{i=1}^{n} \partial_r z_i(\xi)\, \partial_s z_i(\xi)$$

$$= \langle \partial_r z, \partial_s z\rangle,$$

where $\partial_s = \partial_{\xi^s}$. Therefore, the Fisher metric is the natural metric induced from the Euclidean metric of $\mathbb{R}^n$ on the sphere $\mathbb{S}_{2,+}^{n-1}$. We note that $\partial_r z$ is a tangent vector to the sphere in the direction of ξ^r.

To find the information distance between two multinomial distributions p and q, we need to find the length of the shortest curve on the sphere $\mathbb{S}_{2,+}^{n-1}$, joining p and q. The curve that achieves the minimum is an arc of great circle passing through p and q, and this curve is unique.

Let z_p and z_q denote the points on the sphere corresponding to the aforementioned distributions. The angle α made by the unit vectors $z_p/2$ and $z_q/2$ satisfies $\cos\alpha = \langle z_p/2, z_q/2\rangle$. Since the distance on the sphere is the product between the radius and the central angle, we have

$$d(p,q) = 2\alpha = 2\arccos\Big(\sum_{i=1}^{n}\frac{z_p^i}{2}\frac{z_q^i}{2}\Big)$$

$$= 2\arccos\Big(\sum_{i=1}^{n}(p_i q_i)^{1/2}\Big).$$

It is worthy to note that the Euclidean distance between p and q can be written as

$$\|z_p - z_q\|^2 = \Big(\sum_{i=1}^{n}(z_p^i - z_q^i)^2\Big)^{1/2} = 2\Big(\sum_{i=1}^{n}\Big(\frac{z_p^i}{2} - \frac{z_q^i}{2}\Big)^2\Big)^{1/2}$$

$$= 2\Big(\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2\Big)^{1/2} = d_H(p,q),$$

which is called the *Hellinger distance* between p and q. We shall discuss about this distance in more detail later.

The foregoing computation of the Fisher metric was exploiting geometric properties. In the following we shall provide a direct computation. We write the statistical model of multinomial distributions by $\mathcal{S} = \{p(k;\xi)\}$, with

$$p(k;\xi) = \frac{n!}{k_1!\dots k_m!}p_1^{k_1}\dots p_{m-1}^{k_{m-1}}p_m^{k_m},$$

where
$$\mathcal{X} = \{k = (k_1, \ldots k_m) \in \mathbb{N}^m; k_1 + \ldots k_m = n\},$$

and $\xi = (\xi^1, \ldots, \xi^{m-1}) \in \mathbb{E} = [0,1]^{m-1}$, with $\xi^i = p_i$, $i = 1, \ldots m-1$, and $p_m = 1 - p_1 - \cdots - p_{m-1}$. Then a straightforward computation shows

$$\partial_i \ell(k; \xi) = \frac{k_i}{p_i} - \frac{k_m}{p_m}$$

$$\partial_j \partial_i \ell(k; \xi) = -\left[\frac{k_i \delta_{ij}}{p_i^2} + \frac{k_m}{p_m^2}\right].$$

Using the formula for the marginal probability

$$\sum_k k_i p(k; \xi) = n p_i,$$

we have

$$
\begin{aligned}
g_{ij}(\xi) &= -E_\xi[\partial_i \partial_j \ell(k; \xi)] = E\left[\frac{k_i \delta_{ij}}{p_i^2} + \frac{k_m}{p_m^2}\right] \\
&= \frac{\delta_{ij}}{p_i^2} \sum_k k_i p(k; \xi) + \frac{1}{p_m^2} \sum_k k_m p(k; \xi) \\
&= n\left[\frac{\delta_{ij}}{p_i} + \frac{1}{p_m}\right] = n\left[\frac{\delta_{ij}}{\xi^i} + \frac{1}{1 - \xi^1 - \cdots - \xi^{m-1}}\right].
\end{aligned}
$$

2.9 Poisson Geometry

Consider m independent Poisson distributions with parameters λ_i, $i = 1, \ldots, m$. The joint probability function is given by the product

$$p(x; \lambda) = \prod_{i=1}^m p_{\lambda_i}(x_i) = \prod_{i=1}^m e^{-\lambda_i} \frac{\lambda_i^{x_i}}{x_i!},$$

with $\lambda = (\lambda_1, \ldots, \lambda_{m+1}) \in \mathbb{E} = (0, \infty)^m$, and $x = (x_1, \ldots, x_m) \in \mathcal{X} = (\mathbb{N} \cup \{0\})^m$. The log-likelihood function and its derivatives with respect to $\partial_j = \partial_{\lambda_j}$ are

$$
\begin{aligned}
\ell(x; \lambda) &= -\lambda_i + x_i \ln \lambda_i - \ln(x_i!) \\
\partial_j \ell(x; \lambda) &= -1 + \frac{x_j}{\lambda_j} \\
\partial_k \partial_j \ell(x; \lambda) &= -\frac{x_j}{\lambda_j^2} \delta_{kj}.
\end{aligned}
$$

Then the Fisher information is obtained as

$$
\begin{aligned}
g_{jk}(\lambda) &= E\left[\frac{x_j}{\lambda_j^2}\delta_{kj}\right] = \frac{1}{\lambda_j^2}\delta_{kj}E[x_j] \\
&= \frac{1}{\lambda_j^2}\delta_{kj}\sum_x x_j p(x;\lambda) = \frac{1}{\lambda_j}\delta_{kj}.
\end{aligned}
$$

Therefore the Fisher matrix has a diagonal form with positive entries.

2.10   The Space $\mathcal{P}(\mathcal{X})$

Let $\mathcal{X} = \{x_1, \ldots, x_n\}$ and consider the statistical model $\mathcal{P}(\mathcal{X})$ of all discrete probability densities on $\mathcal{X}$. The space $\mathcal{P}(\mathcal{X})$ can be imbedded into the function space $\mathbb{R}^{\mathcal{X}} = \{f; f : \mathcal{X} \to \mathbb{R}\}$ in several ways, as we shall describe shortly. This study can be found in Nagaoka and Amari [61].

For any $\alpha \in \mathbb{R}$ consider the function $\varphi_\alpha : (0, \infty) \to \mathbb{R}$

$$
\varphi_\alpha(u) = \begin{cases} \dfrac{2}{1-\alpha}u^{\frac{1-\alpha}{2}}, & \text{if } \alpha \neq 1 \\[2mm] \ln u, & \text{if } \alpha = 1. \end{cases}
$$

The imbedding

$$
\mathcal{P}(\mathcal{X}) \ni p(x;\xi) \to \varphi_\alpha\big(p(x;\xi)\big) \in \mathbb{R}^{\mathcal{X}}
$$

is called the α-*representation* of $\mathcal{P}(\mathcal{X})$. A distinguished role will be played by the α-*likelihood functions*

$$
\ell^{(\alpha)}(x;\xi) = \varphi_\alpha\big(p(x;\xi)\big).
$$

The coordinate tangent vectors in this representation are given by

$$
\partial_i \ell^{(\alpha)}(x;\xi) = \partial_{\xi^i}\varphi_\alpha\big(p(x;\xi)\big).
$$

The α-representation can be used to define the Fisher metric and the $\nabla^{(\alpha)}$-connection on $\mathcal{P}(\mathcal{X})$.

Proposition 2.10.1 *The Fisher metric can be written in terms of the α-likelihood functions as in the following*

(*i*) $g_{ij}(\xi) = \displaystyle\sum_{k=1}^{n} \partial_i \ell^{(\alpha)}(x_k; \xi) \partial_j \ell^{(-\alpha)}(x_k; \xi);$

(*ii*) $g_{ij}(\xi) = -\dfrac{2}{1+\alpha} \displaystyle\sum_{k=1}^{n} p(x_k; \xi)^{\frac{1+\alpha}{2}} \partial_i \partial_j \ell^{(\alpha)}(x_k; \xi).$

Proof: Differentiating yields

$$
\begin{aligned}
\partial_i \ell^{(\alpha)} &= p^{\frac{1-\alpha}{2}} \partial_i \ell; & (2.10.13)\\
\partial_i \ell^{(-\alpha)} &= p^{\frac{1+\alpha}{2}} \partial_i \ell; & (2.10.14)\\
\partial_i \partial_j \ell^{(\alpha)} &= p^{\frac{1-\alpha}{2}} \left(\partial_i \partial_j \ell + \frac{1-\alpha}{2} \partial_i \ell \partial_j \ell \right), & (2.10.15)
\end{aligned}
$$

where $\ell(x; \xi) = \ln p(x; \xi)$.

(*i*) The previous computations and formula (1.6.16) provide

$$
\begin{aligned}
\sum_{k=1}^{n} \partial_i \ell^{(\alpha)}(x_k; \xi) \partial_j \ell^{(-\alpha)}(x_k; \xi) &= \sum_{k=1}^{n} p^{\frac{1-\alpha}{2}} \partial_i \ell(x_k) p^{\frac{1+\alpha}{2}} \partial_j \ell(x_k)\\
&= \sum_{k=1}^{n} p(x_k; \xi) \partial_i \ell(x_k) \partial_j \ell(x_k)\\
&= E_\xi[\partial_i \ell \, \partial_j \ell] = g_{ij}(\xi).
\end{aligned}
$$

(*ii*) Relation (2.10.15) implies

$$
p^{\frac{1+\alpha}{2}} \partial_i \partial_j \ell^{(\alpha)}(x; \xi) = p(x; \xi) \partial_i \partial_j \ell(x; \xi) + \frac{1-\alpha}{2} p(x; \xi) \partial_i \ell \partial_j \ell(x; \xi).
$$

Summing and using (1.6.16) and (1.6.18), we have

$$
\begin{aligned}
\sum_{k=1}^{n} p(x_k; \xi)^{\frac{1+\alpha}{2}} \partial_i \partial_j \ell^{(\alpha)}(x_k; \xi) &= E_\xi[\partial_i \partial_j \ell] + \frac{1-\alpha}{2} E_\xi[\partial_i \ell \, \partial_j \ell]\\
&= -g_{ij}(\xi) + \frac{1-\alpha}{2} g_{ij}(\xi)\\
&= -\frac{1+\alpha}{2} g_{ij}(\xi).
\end{aligned}
$$

∎

The symmetry of relation (*i*) implies that the Fisher metric induced by both α and $-\alpha$-representations are the same.

Proposition 2.10.2 *The components of the α-connection are given in terms of the α-representation as*

$$\Gamma_{ij,k}^{(\alpha)} = \sum_{r=1}^{n} \partial_i \partial_j \ell^{(\alpha)}(x_r; \xi) \, \partial_k \ell^{(-\alpha)}(x_r; \xi). \qquad (2.10.16)$$

Proof: Combining relations (2.10.14) and (2.10.15)

$$
\begin{aligned}
\sum_{r=1}^{n} \partial_i \partial_j \ell^{(\alpha)} \, \partial_k \ell^{(-\alpha)} &= \sum_{r=1}^{n} p(x_r; \xi) \left(\partial_i \partial_j \ell + \frac{1-\alpha}{2} \partial_i \ell \partial_j \ell \right) \partial_k \ell(x_r; \xi) \\
&= E_\xi \left[\left(\partial_i \partial_j \ell + \frac{1-\alpha}{2} \partial_i \ell \partial_j \ell \right) \partial_k \ell \right] \\
&= \Gamma_{ij,k}^{(\alpha)},
\end{aligned}
$$

by (1.11.34).   ∎

   The particular values $\alpha = -1, 0, 1$ provide distinguished important cases of representations of $\mathcal{P}(\mathcal{X})$.

2.10.1 -1-Representation

If $\alpha = -1$, then $\varphi_{-1}(u) = u$, and $\ell^{(-1)}\big(p(x; \xi)\big) = p(x; \xi)$ is the identical imbedding of $\mathcal{P}(\mathcal{X})$ into $\mathbb{R}^{\mathcal{X}}$. Thus $\mathcal{P}(\mathcal{X})$ is an open set of the affine space $\mathcal{A}_1 = \{f : \mathcal{X} \to \mathbb{R}; \sum_{k=1}^{n} f(x_k) = 1\}$. Therefore, the tangent space at any point p_ξ can be identified with the following affine variety

$$T_\xi^{(-1)}(\mathcal{P}(\mathcal{X})) = \mathcal{A}_0 = \{f : \mathcal{X} \to \mathbb{R}; \sum_{k=1}^{n} f(x_k) = 0\}.$$

The coordinate vector fields in this representation are given by

$$(\partial_i^{-1})_\xi = \partial_i p_\xi.$$

We can easily check that

$$\sum_{k=1}^{n} (\partial_i^{-1})_\xi(x_k) = \sum_{k=1}^{n} \partial_i p_\xi(x_k) = \partial_i(1) = 0,$$

so $(\partial_i^{-1})_\xi \in T_\xi(\mathcal{P})$, for any ξ.

2.10.2 0-Representation

This is also called the *square root representation*. In this case $\varphi_0(u) = 2\sqrt{u}$, and the imbedding $\varphi_0 : \mathcal{P}(\mathcal{X}) \to \mathbb{R}^{\mathcal{X}}$ is

$$p(x;\xi) \to \varphi_0(p(x;\xi)) = \ell^{(0)}(x;\xi) = 2\sqrt{p(x;\xi)} = \theta(x) \in \mathbb{R}^{\mathcal{X}}.$$

Since $\sum_{k=1}^{n} \theta(x_k)^2 = 4$, the image of the imbedding φ_0 is an open subset of the sphere of radius 2,

$$\varphi_0(\mathcal{P}(\mathcal{X})) \subset \{\theta; \theta : \mathcal{X} \to \mathbb{R}; \sum_{k} \theta(x_k)^2 = 4\}.$$

The induced metric from the natural Euclidean metric of $\mathbb{R}^{\mathcal{X}}$ on this sphere is

$$\begin{aligned}
\langle \partial_i \theta, \partial_j \theta \rangle &= \sum_{k=1}^{n} \partial_i \theta(x_k) \partial_j \theta(x_k) \\
&= 4 \sum_{k=1}^{n} \partial_i \sqrt{p(x_k;\xi)} \partial_j \sqrt{p(x_k;\xi)} \\
&= g_{ij}(\xi),
\end{aligned}$$

i.e., the Fisher metric on the statistical model $\mathcal{P}(\mathcal{X})$.

The coordinate vector fields are given by

$$(\partial_i^0)_\xi = \partial_i \ell^{(0)}(x;\xi) = \frac{1}{\sqrt{p(x;\xi)}} \partial_i p(x;\xi).$$

The next computation deals with the tangent space generated by $(\partial_i^0)_\xi$. We have

$$\begin{aligned}
\langle \theta, (\partial_i^0)_\xi \rangle &= \sum_{k=1}^{n} \theta(x_k) \frac{1}{p(x_k;\xi)} \partial_i p(x_k;\xi) \\
&= \sum_{k=1}^{n} 2\sqrt{p(x_k;\xi)} \frac{1}{p(x_k;\xi)} \partial_i p(x_k;\xi) \\
&= 2\partial_i \sum_{k=1}^{n} p(x_k;\xi) = 0,
\end{aligned}$$

so that the vector $(\partial_i^0)_\xi$ is perpendicular on the vector θ, and hence belongs to the tangent plane to the sphere at θ. This can be identified with the tangent space $T_\xi^{(0)}\mathcal{P}(\mathcal{X})$ in the 0-representation.

2.10.3   1-Representation

This is also called the *exponential (or the logarithmic) representation,* because each distribution $p(x;\xi) \in \mathcal{P}(\mathcal{X})$ is identified with $\ln p(x;\xi) \in \mathbb{R}^{\mathcal{X}}$. In this case the 1-likelihood function becomes $\ell^{(1)}(x;\xi) = \ell(x;\xi) = \ln p(x;\xi)$, i.e., the usual likelihood function.

The coordinate vector fields are given by

$$(\partial_i^1)_\xi = \partial_i \ell^{(1)}(x;\xi) = \frac{1}{p(x;\xi)}\partial_i p(x;\xi).$$

In the virtue of the computation

$$E_p[(\partial_i^1)_\xi] = E_p[\partial_i \ell^{(1)}(x;\xi)] = \sum_{k=1}^n \partial_i p(x_k;\xi) = \partial_i(1) = 0,$$

it follows that the tangent space in this representation is given by

$$T_p^{(1)}(\mathcal{P}(\mathcal{X})) = \{f; f \in \mathbb{R}^{\mathcal{X}}\}, E_p[f] = 0\}.$$

It is worth noting that tangent spaces are invariant objects, that do not depend on any representation. However, when considering different system of parameters, tangent vectors can be described by some particular relations, like in the cases of ± 1 and 0 representations.

2.10.4   Fisher Metric

Let $\xi^i = p(x_i;\xi)$, $i = 1, \ldots, n-1$, be the coordinates on $\mathcal{P}(\mathcal{X})$. Since $p(x_n;\xi) = 1 - \sum_{j=1}^{n-1}\xi^j$, then the partial derivatives with respect to ξ^j are

$$\partial_i p(x_k;\xi) = \begin{cases} \delta_{ik}, & \text{if } k = 1, \ldots, n-1 \\ -1, & \text{if } k = n. \end{cases}$$

Then the Fisher metric is given by

$$\begin{aligned}
g_{ij}(\xi) &= E_p[\partial_i \ell\, \partial_j \ell] = \sum_{k=1}^n p(x_k;\xi)\partial_i \ln p(x_k;\xi)\partial_j \ln p(x_k;\xi) \\
&= \sum_{k=1}^n \frac{\partial_i p(x_k;\xi)\partial_j p(x_k;\xi)}{p(x_k;\xi)} \\
&= \sum_{k=1}^{n-1} \frac{\delta_{ik}\delta_{jk}}{\xi^k} + \frac{1}{1 - \sum_{j=1}^{n-1}\xi^j} \\
&= \frac{\delta_{ij}}{\xi^j} + \frac{1}{1 - \sum_{j=1}^{n-1}\xi^j}.
\end{aligned}$$

2.11 Problems

2.1. Consider the statistical model given by the densities of a normal family

$$p(x, \xi) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathcal{X} = \mathbb{R},$$

with parameters $(\xi^1, \xi^2) = (\mu, \sigma) \in \mathbb{R} \times (0, \infty)$.

(a) Show that the log-likelihood function and its derivatives are given by

$$\ell_x(\xi) = \ln p(x, \xi) = -\frac{1}{2}\ln(2\pi) - \ln\sigma - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\partial_\sigma \ell_x(\xi) = \partial_\sigma \ln p(x, \xi) = -\frac{1}{\sigma} + \frac{1}{\sigma^3}(x-\mu)^2$$

$$\partial_\sigma \partial_\sigma \ell_x(\xi) = \partial_\sigma \partial_\sigma \ln p(x, \xi) = \frac{1}{\sigma^2} - \frac{3}{\sigma^4}(x-\mu)^2$$

$$\partial_\mu \ell_x(\xi) = \partial_\mu \ln p(x, \xi) = \frac{1}{\sigma^2}(x-\mu)$$

$$\partial_\mu \partial_\mu \ell_x(\xi) = \partial_\mu \partial_\mu \ln p(x, \xi) = -\frac{1}{\sigma^2}$$

$$\partial_\sigma \partial_\mu \ell_x(\xi) = \partial_\sigma \partial_\mu \ln p(x, \xi) = -\frac{2}{\sigma^3}(x-\mu).$$

(b) Show that the Fisher–Riemann metric components are given by

$$g_{11} = \frac{1}{\sigma^2}, \qquad g_{12} = g_{21} = 0, \qquad g_{22} = \frac{2}{\sigma^2}.$$

2.2. Consider the statistical model defined by the lognormal distribution

$$p_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \qquad x > 0.$$

(a) Show that the log-likelihood function and its derivatives are given by

$$\ell(\mu, \sigma) = -\ln\sqrt{2\pi} - \ln\sigma - \ln x - \frac{1}{2\sigma^2}(\ln x - \mu)^2$$

$$\partial_\mu^2 \ell(\mu, \sigma) = -\frac{1}{\sigma^2}$$

$$\partial_\sigma^2 \ell(\mu, \sigma) = \frac{1}{\sigma^2} - \frac{3}{\sigma^4}(\ln x - \mu)^2$$

$$\partial_\mu \partial_\sigma \ell(\mu, \sigma) = -\frac{2}{\sigma^3}(\ln x - \mu).$$

(b) Using the substitution $y = \ln x - \mu$, show that the components of the Fisher–Riemann metric are given by

$$g_{\sigma\sigma} = \frac{2}{\sigma^2}, \qquad g_{\mu\mu} = \frac{1}{\sigma^2}, \qquad g_{\mu\sigma} = g_{\sigma\mu} = 0.$$

2.3. Let

$$p_\xi(x) = p_{\alpha,\beta}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \, x^{\alpha-1} e^{-x/\beta},$$

with $(\alpha, \beta) \in (0, \infty) \times (0, \infty)$, $x \in (0, \infty)$ be the statistical model defined by the gamma distribution.

(a) Show that the log-likelihood function is

$$\ell_x(\xi) = \ln p_\xi = -\alpha \ln \beta - \ln \Gamma(\alpha) + (\alpha - 1) \ln x - \frac{x}{\beta}.$$

(b) Verify the relations

$$
\begin{aligned}
\partial_\beta \ell_x(\xi) &= -\frac{\alpha}{\beta} + \frac{x}{\beta^2} \\
\partial_{\alpha\beta} \ell_x(\xi) &= -\frac{1}{\beta} \\
\partial_\beta^2 \ell_x(\xi) &= \frac{\alpha}{\beta^2} - \frac{2x}{\beta^3} \\
\partial_\alpha \ell_x(\xi) &= -\ln \beta - \psi(\alpha) + \ln x \\
\partial_\alpha^2 \ell_x(\xi) &= -\psi_1(\alpha),
\end{aligned}
$$

where

$$\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}, \qquad \psi_1(\alpha) = \psi'(\alpha) \qquad (2.11.17)$$

are the *digamma* and the *trigamma* functions, respectively.

(c) Prove that for $\alpha > 0$, we have

$$\sum_{n \geq 0} \frac{\alpha}{(\alpha + n)^2} > 1.$$

2.4. Consider the beta distribution

$$p_{a,b} = \frac{1}{B(a, b)} \, x^{a-1}(1 - x)^{b-1}, \qquad a, b > 0, x \in [0, 1].$$

(*a*) Using that the beta function

$$B(a, b) = \int_0^1 x^{a-1}(1 - x)^{b-1}\, dx$$

can be expressed in terms of gamma function as

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)},$$

show that its partial derivatives can be written in terms of digamma functions, as

$$\begin{align}
\partial_a \ln B(a, b) &= \psi(a) - \psi(a + b) & (2.11.18)\\
\partial_b \ln B(a, b) &= \psi(b) - \psi(a + b). & (2.11.19)
\end{align}$$

(*b*) Show that the log-likelihood function is given by

$$\ell(a, b) = \ln p_{a,b} = -\ln B(a, b) + (a - 1)\ln x + (b - 1)\ln(1 - x).$$

(*c*) Take partial derivatives and use formulas (2.11.18) and (2.11.19) to verify relations

$$\begin{align}
\partial_a \ell(a, b) &= -\partial_a \ln B(a, b) + \ln x = \psi(a + b) - \psi(a) + \ln x\\
\partial_b \ell(a, b) &= \psi(a + b) - \psi(b) + \ln(1 - x)\\
\partial_a^2 \ell(a, b) &= \psi'(a + b) - \psi'(a) = \psi_1(a + b) - \psi_1(a)\\
\partial_b^2 \ell(a, b) &= \psi'(a + b) - \psi'(b) = \psi_1(a + b) - \psi_1(b)\\
\partial_a \partial_b \ell(a, b) &= \psi'(a + b) = \psi_1(a + b).
\end{align}$$

(*c*) Using the expression of trigamma functions as a Hurwitz zeta function, show that the Fisher information matrix can be written as a series $g = \sum_{n \geq 0} g_n$, where

$$g_n = \begin{pmatrix} \frac{1}{(a+n)^2} - \frac{1}{(a+b+n)^2} & -\frac{1}{(a+b+n)^2} \\[2ex] -\frac{1}{(a+b+n)^2} & \frac{1}{(b+n)^2} - \frac{1}{(a+b+n)^2} \end{pmatrix}.$$

2.5. Let $\mathcal{S} = \{p_\xi; \xi \in [0, 1]\}$ be a one-dimensional statistical model, where

$$p(k; \xi) = \binom{n}{k}\xi^k(1 - \xi)^{n-k}$$

is the Bernoulli distribution, with $k \in \{0, 1, \ldots, n\}$ and $\xi \in [0, 1]$. Show that the derivatives of the log-likelihood function $\ell_k(\xi) = \ln p(k; \xi)$ are

$$\partial_\xi \ell_k(\xi) = \frac{k}{\xi} - (n - k)\frac{1}{1 - \xi}$$

$$\partial_\xi^2 \ell_k(\xi) = -\frac{k}{\xi^2} - (n - k)\frac{1}{(1 - \xi)^2}.$$

2.6. Consider the geometric probability distribution $p(k; \xi) = (1 - \xi)^{k-1}\xi$, $k \in \{1, 2, 3, \ldots\}$, $\xi \in [0, 1]$. Show that

$$\partial_\xi \ell_k(\xi) = \frac{k - 1}{\xi - 1} + \frac{1}{\xi}$$

$$\partial_\xi^2 \ell_k(\xi) = -\frac{(k - 1)}{(\xi - 1)^2} - \frac{1}{\xi^2}.$$

2.7. Let f be a density function on $\mathbb{R}$ and define the statistical model

$$\mathcal{S}_f = \left\{ p(x; \mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right); \mu \in \mathbb{R}, \sigma > 0 \right\}.$$

(a) Show that $\int_\mathbb{R} p(x; \mu, \sigma) \, dx = 1$.

(b) Verify the following formulas involving the log-likelihood function $\ell = \ln p(\cdot; \mu, \sigma)$:

$$\partial_\mu \ell = -\frac{1}{\sigma}\frac{f'}{f}, \qquad \partial_\sigma \ell = -\frac{1}{\sigma} - \frac{(x - \mu)}{\sigma^2}\frac{f'}{f}$$

$$\partial_\mu \partial_\sigma \ell = \frac{1}{f^2}\left[\left(\frac{f'}{\sigma^2} + \frac{1}{\sigma}\frac{x - \mu}{\sigma^2}f''\right)f - \frac{1}{\sigma}\frac{x - \mu}{\sigma^2}(f')^2\right].$$

(b) Show that for any continuous function h we have

$$E_{(\mu, \sigma)}\left[h\left(\frac{x - \mu}{\sigma}\right)\right] = E_{(0, 1)}[h(x)].$$

(c) Assume that f is an even function (i.e., $f(-x) = f(x)$). Show that the Fisher–Riemann metric, g, has a diagonal form (i.e., $g_{12} = 0$).

(d) Prove that the Riemannian space $(\mathcal{S}_f, g)$ has a negative, constant curvature.

(e) Consider $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Use the aforementioned points to deduct the formula for g_{ij} and to show that the curvature $K = -\frac{1}{2}$.

2.8. Search the movement of the curve

$$(\mu, \sigma) \to p_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \mu^2 + \sigma^2 = 1$$

with $(\mu, \sigma, p) \in \mathbb{R} \times (0, \infty) \times (0, \infty)$, $x \in \mathbb{R}$, fixed, in the direction of the binormal vector field.

2.9. The graph of the normal density of probability

$$x \to p_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

is called Gauss bell. Find the equation of the surface obtained by revolving the Gauss bell about:

(a) Ox axis;
(b) Op axis.

2.10. Inspect the movement of the trajectories of the vector field (y, z, x) after the direction of the vector field

$$\left(1, 1, \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right),$$

where μ and σ are fixed.

2.11. The normal surface

$$(\mu, \sigma) \to p_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$(\mu, \sigma) \in \mathbb{R} \times (0, \infty); \; x \in \mathbb{R}$$

is deformed into $p_{\mu,\sigma}(tx)$, $t \in \mathbb{R}$. What happens with the Gauss curvature?

2.12. The gamma surface

$$(\alpha, \beta) \to p_{\alpha,\beta}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

$$(\alpha, \beta) \in (0, \infty) \times (0, \infty); \; x \in (0, \infty)$$

is deformed into $p_{t\alpha,\beta}(x)$, $t \in (0, \infty)$. What happens with the mean curvature?