Chapter 3. Parametrizations

In regular exponential families maximum likelihood estimation is closely related to the so-called mean value parametrization. This parametrization will be described after some brief preliminaries. The relation to maximum likelihood is pursued in Chapter 5.

3.1 Notation

For $v \in R^k$, $\alpha \in R$ let $H(v, \alpha)$ denote the hyperplane

$$H(v, \alpha) = \{x \in R^k : v \cdot x = \alpha\} \quad .$$

Let $H^+(a, \alpha)$ and $H^-(a, \alpha)$ be the open half spaces

$$H^+(v, \alpha) = \{x \in R^k : v \cdot x > \alpha\}$$

$$H^-(v, \alpha) = \{x \in R^k : v \cdot x < \alpha\} \quad .$$

When (v, α) are clear from the context they will be omitted from the notation Note that the closure of $H^{\pm}$ is written $\overline{H}^{\pm}$ and, of course, satisfies $\overline{H}^{\pm} = H \cup H^{\pm}$.

STEEP FAMILIES

Most exponential families occurring in practice are regular (i.e. N is open). However, for technical reasons which will become clear in Chapter 6, it is very useful to prove the parametrization Theorem 3.6 for steep families as well.

3.2 Definition

Let $\phi: R^k \to (-\infty, \infty]$ be convex. Let $N = \{\theta \in R^k: \phi(\theta) < \infty\}$. Assume ϕ is continuously differentiable on $N°$. Let $\theta_1 \in N - N°$, $\theta_0 \in N°$, and let $\theta_\rho = \theta_0 + \rho(\theta_1 - \theta_0)$, $0 < \rho < 1$, denote points on the line joining θ_0 to θ_1. Then, ϕ is called *steep* if for all $\theta_1 \in N - N°$, $\theta_0 \in N°$,

(1) $$\lim_{\rho \uparrow 1} (\theta_1 - \theta_0) \cdot \nabla\phi(\theta_\rho) = \infty \quad .$$

Note that (1) is the same as

(1') $$\lim_{\rho \uparrow 1} \frac{\partial}{\partial \rho} \phi(\theta_\rho) = \infty$$

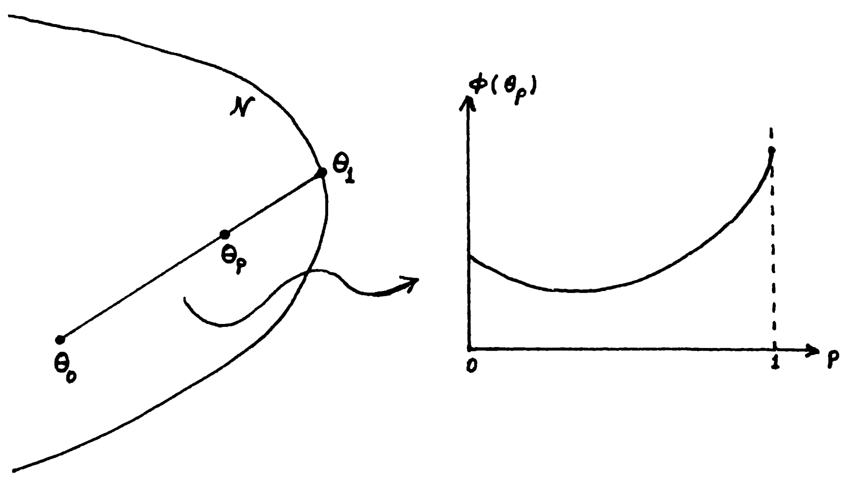

Figure 3.2(1): An illustration of the definition of steepness

A standard exponential family is called steep if its cumulant generating function, ψ, is steep. (A steep convex function is sometimes referred to as an "essentially smooth" convex function.) Note that if the exponential family is regular then it is *a fortiori* steep since $N - N° = \phi$.

Here is a convenient necessary and sufficient condition for steepness.

3.3 Proposition

A minimal standard exponential family is steep if and only if

$$(1) \qquad E_\theta(||x||) = \infty \quad \text{for all} \quad \theta \in N - N^\circ \quad .$$

Proof. Suppose the family is steep. Then

$$(\theta_1 - \theta_0) \cdot \nabla\psi(\theta_\rho) = (\theta_1 - \theta_0) \cdot \xi(\theta_\rho) \to \infty \quad \text{as} \quad \rho \uparrow 1 \quad .$$

This implies $E_{\theta_\rho}((\theta_1 - \theta_0) \cdot X) \to \infty$, which implies (1).

The converse seems not to be easy to prove without further preparation. We postpone the proof to Chapter 6. It appears after the proof of Lemma 6.8. ||

3.4 Example

There is one classic example of a steep non-regular family which occurs in a variety of applications. It is the family of densities defined by

$$(1) \qquad (\pi)^{-1/2}\, z^{-3/2} \exp(\theta_1 z + \theta_2(1/z) - (-2(\theta_1\theta_2)^{1/2} - (1/2)\ln(-2\theta_2)))$$

relative to Lebesque measure on $z \in (0, \infty)$. The canonical statistics are $(x_1, x_2) = (z, 1/z)$ and the natural parameter space is

$$(2) \qquad N = (-\infty, 0] \times (-\infty, 0) \quad .$$

Thus the family is not regular but is steep since $E_{(0,\theta_2)}(x_1) = \infty$ for all $\theta_2 \in (-\infty, 0)$. These densities are referred to as *inverse Gaussian*. They arise, for example, as the distribution of the first time (x_1) that a standard Brownian motion crosses the line $\ell(t) = \sqrt{-2\theta_2} - \sqrt{-2\theta_1}\, t$. Note that these densities with $\theta_1 = 0$ are the scale family of stable densities on $(0, \infty)$ with index ½. See Feller (1966). For some other steep non-regular families see Bar-Lev and Enis (1984).

MEAN VALUE PARAMETRIZATION

We begin with a useful lemma which involves a natural relation between parameter space (Θ) and sample space (X). Similar relations will reoccur several times and we have found it useful to draw pictures to illustrate the geometric relationships involved. Figure 3.5.1, below, is a simple example of such a picture which illustrates the hypotheses of Lemma 3.5.

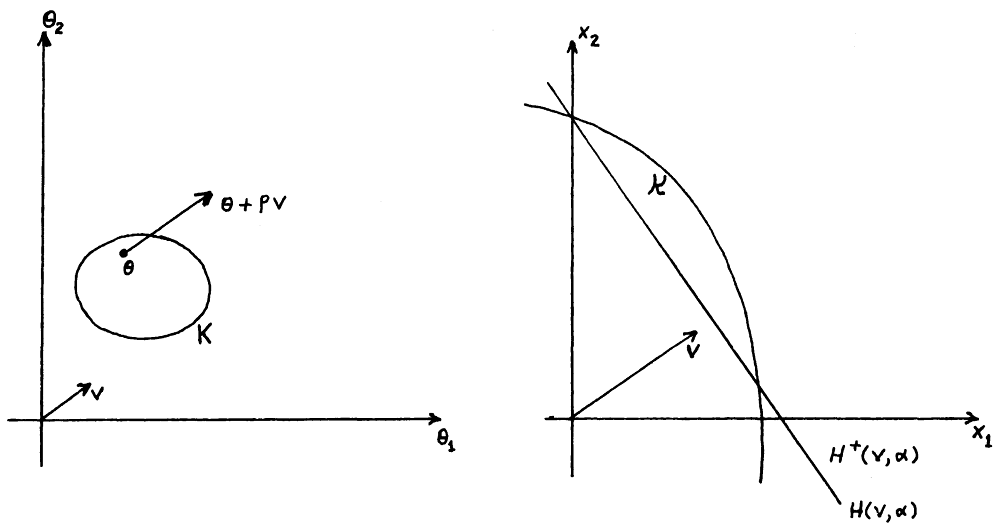

Figure 3.5.1: Illustrating the hypotheses of Lemma 3.5 when k = 2.

3.5 Lemma

Let $v \in R^k$, $\alpha \in R$. Let $K \subset R^k$ be compact. Suppose $\nu(\bar{H}^+(v, \alpha)) > 0$. Then there exists a constant $c > 0$ such that

(1) $$\lambda(\theta + \rho v) \geq ce^{\rho\alpha} \qquad \forall\, \theta \in K, \qquad \rho \geq 0$$

(Note that (1) is equivalent to

(1') $$\psi(\theta + \rho v) \geq \rho\alpha + \log c \quad \forall\, \theta \in K, \qquad \rho \geq 0$$

If $\theta + \rho v \notin N$ then $\lambda(\theta + \rho v) = \infty$ so that (1) is trivial.)

Proof.

$$(2) \qquad \lambda(\theta + \rho v) = \int e^{(\theta+\rho v)\cdot x}\nu(dx) \geq e^{\rho\alpha} \int_{\bar{H}^+} e^{\theta\cdot x}\nu(dx) \geq ce^{\rho\alpha}$$

where

$$(3) \qquad c = \inf_{\theta\in K} \int_{\bar{H}^+} e^{\theta\cdot x}\nu(dx) > 0 \quad .$$

((2) shows that if $c = \infty$ here then $\lambda(\theta + \rho v) = \infty$ for all $\theta \in K$ and all $\rho \geq 0$.) ||

Note that (3) provides an explicit formula for the constant c appearing in formula (1). Exercise 3.5.1 contains a converse to this lemma.

Here is the main result.

3.6 Theorem

Let $\{p_\theta\}$ be a minimal steep standard exponential family. Then $\xi(\theta) = E_\theta(X)$ defines a homeomorphism of N° and K° (i.e., $\xi\colon N^\circ \to K^\circ$ is continuous, 1-1, and onto. Of course, if $\{p_\theta\}$ is regular then $\xi\colon N \to K^\circ$ since $N = N^\circ$).

Proof. ξ is continuous on N° by Theorem 2.2 and Corollary 2.3. It is 1-1 by Corollary 2.5. It remains to prove that $\xi(N^\circ) = K^\circ$, that is, to show

$$(1) \qquad x \in K^\circ \Rightarrow x \in \xi(N) \quad .$$

It suffices to prove (1) for $x = 0$, for then the desired result for arbitrary $x \in K^\circ$ follows upon translating the origin, which is justified by Proposition 1.6. So, assume $0 \in K^\circ$.

Let $S_1 = \{v \in R^k\colon \|v\| = 1\}$. Since $0 \in K^\circ$ there is an $\varepsilon > 0$ such that

$$(2) \qquad \nu(\bar{H}^+(v, \varepsilon)) > c > 0$$

for all $v \in S_1$. (If not, there would be sequences $v_i \in S_1$ with

$v_i \to v \in S_1$ and $\varepsilon_i \to 0$ for which $\nu(\bar{H}^+(v_i, \varepsilon_i)) \to 0$. This would imply $\nu(\bar{H}^+(v, 0)) = 0$ which contradicts $0 \in K°$.) Now apply Lemma 3.2 (with $v = \theta/||\theta||$ and $\rho = ||\theta||$) including the expression 3.2(3) for the constant appearing in the lemma to get

(3) $$\psi(\theta) \geq ||\theta||\varepsilon + \log c$$

with c as in (2). Thus

(4) $$\lim_{||\theta||\to\infty} \psi(\theta) = \infty \quad .$$

(See Exercise 3.6.2 and Lemma 5.3(3) for restatements of (3), (4).)

Any lower semi-continuous function (such as ψ) defined on a closed set and which also satisfies (4) must assume its minimum. To see this, let $\psi(\theta_i) = \inf\ \{\psi(\theta) : \theta \in R^k\}$. $||\theta_i|| \to \infty$ is impossible by (4). So, there is a convergent subsequence, $\theta_{i'} \to \theta^*$, and $\psi(\theta^*) = \inf\ \{\psi(\theta) : \theta \in R^k\}$ by lower semi-continuity.) This minimum is assumed at a point $\theta^* \in N$.

Suppose $\theta^* \in N - N°$. Then, for some $0 < \rho' < 1$, $\psi(\theta_{\rho'}) < \psi(\theta^*) = \lim_{\rho\uparrow 1} \psi(\theta_0 + \rho(\theta^* - \theta_0))$ by virtue of 3.2(1') of the definition of steepness. Hence no $\theta^* \in N - N°$ can be the minimum point for ψ. It follows that $\theta^* \in N°$.

Hence

$$\xi(\theta^*) = \nabla\psi(\theta^*) = 0$$

since ψ is differentiable on a neighborhood of θ^*. (Here we use Theorem 2.2, Corollary 2.3, and the fact that $\theta^* \in N°$ an open set.) This proves (1) for $x = 0$ and, as noted, completes the proof of the theorem. ||

3.7 Interpretation

Theorem 3.6 shows that a minimal, steep family with parameter space $N°$ can be parametrized by $\xi = \xi(\theta)$, and the range of this parameter is $K°$. This is the *mean value parametrization*. In this parametrization the resulting family is an exponential family, but of course is no longer a standard exponential family (except when $\xi(\cdot)$ is affine). Write

(1) $$\theta(x) = \xi^{-1}(x) = (\theta : \xi(\theta) = x) \quad .$$

The exponential family parametrized by ξ then has densities $\hat{p}_{\xi}(x) = \exp(\theta(\xi) \cdot x - \psi(\theta(\xi)))$. For a number of applications this parametrization is more convenient than the "natural" parametrization described by the canonical parameter θ. If $\{p_\theta\}$ is regular then $N = N^\circ$ and the mean value parametrization reparametrizes the full family.

Minimality was used in Theorem 3.6 only to guarantee that the map is 1-1. Even without minimality the map ξ discriminates between different distributions in $\{P_\theta : \theta \in N\}$. Hence one can still use the mean-value parametrization to conveniently index $\{P_\theta : \theta \in N^\circ\}$, and the range of the mean value parameter is the relative interior of K. (Equivalently, one may reduce to a minimal family by Theorem 1.9 and then apply Theorem 3.3.)

If the family is not steep then $\xi(N^\circ) \subset K^\circ$. We leave this fact -- relatively unimportant for statistical application -- as an exercise. In this case it is even possible to have $\xi(N^\circ)$ not convex. See Exercise 3.7.1 for an example due to Efron (1978).

3.8 Example (Fisher-VonMises Distribution)

For a number of common exponential families the mean value parametrization is the familiar parametrization, or nearly so. For example, for the Binomial (N, π) family the expectation parameter is $N\pi$, for the Poisson (λ) family the expectation parameter is λ, and for the exponential distributions (gamma distributions with index $\alpha = 1$ and unknown scale, σ) the expectation parameter is σ. For the multivariate normal (μ, Σ) family the expectation parameters are μ and $\mu\mu' + \Sigma$ (corresponding to the canonical statistics of 1.14). The mean value parameters are not always so convenient. Nevertheless it is necessary to consider this parametrization in order to construct maximum likelihood estimators. See especially Theorem 5.5.

Accordingly, we now discuss the mean value parametrization for the Fisher-VonMises distribution.

Let ν be uniform measure on the sphere of radius one in R^k. Consider the exponential family generated by ν. When $k = 2$ this is the *VonMises family*. When $k = 3$ it is the *Fisher family* of distributions. These distributions appear often in applications, with a variety of parametrizations, to model angular data in R^k. Consult Mardia (1972) for an extended treatment of these families; see also Beran (1979). (Frequently one considers a sample of n observations from one of these distributions. The sample mean, $\bar{X}_n$, is then also said to have a VonMises or Fisher distribution. The mean value parametrization for the family of distributions of $\bar{X}_n$ is, of course, identical to that below since $E_\theta(\bar{X}_n) = E_\theta(X)$. See also 5.5(3).)

The Laplace transform of ν is

(1) $$\lambda_\nu(\theta) = (2\pi)^{k/2} I_{k/2-1}(||\theta||)/||\theta||^{k/2-1}$$

where $I_s(\cdot)$ denotes the modified Bessel function of order s. When k is odd these functions have a convenient representation in terms of hyperbolic functions; for example

(2) $$I_{1/2}(r) = (2/\pi r)^{1/2} \sinh r$$

$$I_{3/2}(r) = (2/\pi r)^{1/2}(\cosh r - (\sinh r)/r) \quad .$$

(See, for example, Courant and Hilbert (1953).) These functions also have nice recurrence relations; in particular

(3) $$I_s'(r) = I_{s+1}(r) + sI_s(r)/r \ , \qquad s \geq 0, \qquad r > 0 \quad .$$

By symmetry, or by calculation, it follows that $\xi(\theta)$ lies in the same direction as θ, that is

(4) $$\xi(\theta)/||\xi(\theta)|| = \theta/||\theta|| \ , \qquad \theta \neq 0, \text{ and } \xi(0) = 0 \quad .$$

It remains therefore to give a formula for $||\xi(\theta)||$. For this purpose it suffices to consider the case where $\theta_r = (r,0,\ldots,0)$, and to calculate $\frac{d}{dr}\ell n\,\lambda_\nu(\theta_r)$. For the Fisher distribution (k = 3) one gets from (1) - (3) that

(5) $$||\xi(\theta)|| \;=\; \coth ||\theta|| - ||\theta||^{-1} \qquad .$$

For the Von Mises distribution (k = 2) one gets only the less convenient expression

(6) $$||\xi(\theta)|| \;=\; I_1(||\theta||)/I_0(||\theta||) \qquad .$$

Although (6) is less convenient that (5), it can be used in conjunction with series expansions or tables of the modified Bessel function to provide numerical values for $||\xi(\theta)||$, and other information about $||\xi(\theta)||$.

MIXED PARAMETRIZATION

We refer to the type of situation discussed in 1.7. $M = \binom{M_1}{M_2}$ is a partitioned k×k non-singular matrix with $M_1M_2' = 0$. Write

(1) $$\begin{aligned} M_i x &= z_i & i = 1,\,2 \\ (M_i^-)'\theta &= \phi_i & i = 1,\,2 \end{aligned} \qquad .$$

(Thus $\binom{\phi_1}{\phi_2} = (M^{-1})'\theta$.) Where convenient we write $\phi_i = \phi_i(\theta)$ to emphasize the dependence on θ, etc.)

Note that

(2) $$M_i\xi(\theta) \;=\; E_\theta(M_iX) \;=\; E_\theta(Z_i) \;=\; \zeta_i(\theta) \quad \text{(say)} \quad i = 1,\,2 \quad .$$

Recall also that one may without loss of generality visualize only the case where $M = I$. In this case $\phi_1' = (\theta_1,\ldots,\theta_m)$, $z_2' = (x_{m+1},\ldots,x_k)$, $\zeta_2' = (\xi_{m+1},\ldots,\xi_k)$, etc.

The following result is valid for steep families but for simplicity we state and prove it here only for regular families. See Exercise 3.9.1.

3.9 Theorem

Let $\{p_\theta\}$ be minimal and regular. Then the map

$$(3) \qquad \theta \to \begin{pmatrix} \zeta_1(\theta) \\ \phi_2(\theta) \end{pmatrix}$$

is 1 - 1 and continuous on N° $(=N)$ with range

$$(4) \qquad \zeta_1(N^\circ) \times \phi_2(N^\circ) \;=\; K^\circ_{(1)} \times \phi_2(N^\circ) \quad .$$

Proof. Fix $\phi_2^0 \in \phi_2(N)$ and refer to Theorem 1.7. The distributions of Z_1 given $\phi_2(\theta) = \phi_2^0$ form the minimal regular standard exponential family generated by $\nu_{\phi_2^0}$. According to Theorem 3.6 this family can be parametrized (in a 1 - 1 manner) by $\zeta_1(\theta) = E_\theta(Z_1)$. The range of this map is

$$\text{int (conhull (supp } \nu_{\phi_2^0})) \;=\; K^\circ_{\phi_2^0} \qquad \text{(say)} \quad .$$

The formula for $\nu_{\phi_2^0}$ is given in 1.7(5), but all that needs to be noted is that $K^\circ_{\phi_2^0} \equiv K^\circ_{(1)}$. The map in (3) is therefore 1 - 1 with range as in (4). Continuity of the map in (3) is immediate from continuity of ξ. ||

3.10 Interpretation

The above theorem has an interpretation like that of Theorem 3.6. Any minimal regular exponential family can be parametrized by parameters of the form 3.9(3), above. This parametrization is called the *mixed parametrization.*

Consider a mixed parametrization with parameter $\binom{\zeta_1}{\phi_2}$, as above. Then the family of densities corresponding to the parameters $\{\binom{\zeta_1}{\phi_2} : \phi_2 = \phi_2^0\}$

forms a full standard exponential family of order m. (See Theorem 1.7.) However, if one fixes the expectation coordinate and looks at the family corresponding to the parameters $\{\binom{\zeta_1}{\phi_2} : \zeta_1 = \zeta_1^0\}$ then one gets in general only some non-full standard family of dimension and order k, whose parameter space is a (k - m) dimensional manifold in N. Here is an example.

Consider the parametrization of the three dimensional multinomial (N, π) family discussed following 1.8(6). A mixed parametrization for this family involves

$$\binom{\zeta_1}{\zeta_2} = E\binom{Z_1}{Z_2} = \binom{2\pi_1 + \pi_2}{\pi_2 + 2\pi_3}N$$

and

$$\phi_3 = (\tfrac{1}{2}) \log (\pi_2^2/4\pi_1 \pi_3) \quad .$$

Note that the range of $\binom{\zeta_1}{\zeta_2}$ is

$$\{\binom{\zeta_1}{2N - \zeta_1} : 0 < \zeta_1 < 2N\}$$

independent of the value of $\phi_3 \in (-\infty, \infty)$, as claimed by Theorem 3.9. For fixed $\phi_3 = \phi_3^0$ the distributions of $\binom{Z_1}{Z_2}$ form a 2 dimensional exponential family (of order 1) having expectation parameter $\binom{\zeta_1}{\zeta_2}$. (In the genetic interpretation for this parametrization the parameter ϕ_3 measures the strength of selection in favor of the heterozygote character Gg.)

On the other hand the family of distributions corresponding to fixed $\binom{\zeta_1}{\zeta_2}$ is not so convenient. It is the non-linear subfamily of the usual full standard family described by

$$(1) \qquad \Theta = \{\theta : 2e^{\theta_1} + e^{\theta_2} = (\zeta_1/N)\Sigma e^{\theta_i}\} \quad .$$

(If one reduces the usual standard exponential family to a minimal family of dimension 2, then the parameter set becomes a smooth one-dimensional curve within R^2. This provides an example of a curved exponential family, as defined below. See Exercise 3.11.2.)

DIFFERENTIABLE SUBFAMILIES

3.11 Description

A *differentiable subfamily* is a standard exponential family with parameter space Θ an m-dimensional differentiable manifold in N. An especially convenient situation occurs when Θ is a one-dimensional manifold -- i.e. a differentiable curve. Such a family is called a *curved exponential family*. (A technical point: it is often convenient to assume that the parameter space is smoother than being merely differentiable -- for example, to assume it possesses second derivatives. Whenever convenient we consider such an assumption *implicit* in the definition of a differentiable subfamily; writing formulae for relevant second or higher derivatives (as in (3) below) carries with it the assumption that these derivatives exist.)

In a differentiable subfamily the parameter space can be written locally as $\{\theta(t) : t \in N\}$ where N is a neighborhood in R^m and $\theta(\cdot)$ is differentiable and one to one. Properties of such a family around some $\theta_0 \in \Theta$ can often be most conveniently studied after invoking Proposition 1.6 to rewrite the family in a more convenient form. For example in a curved exponential family m = 1 and the proper choice of ϕ_0, z_0 and M in that proposition transforms the problem into one in which

$$\theta_0 = 0 = \theta(t_0)$$

(1) $$\xi(\theta_0) = E_{\theta_0}(X) = 0$$

$$\Sigma(\theta_0) = I$$

(2)
$$\dot{\theta}(t_0) = \frac{d}{dt}\theta(t_0) = \begin{pmatrix} a \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\ddot{\theta}(t_0) = \frac{d^2}{dt^2}\theta(t_0) = \begin{pmatrix} a^2 b \\ a^2/\rho \\ 0 \\ \vdots \\ 0 \end{pmatrix} .$$

(The value $\rho = \infty$ is possible.) Furthermore, one can linearly reparametrize the curve so that $\theta_0 = \theta(0)$ (i.e. so that $t_0 = 0$) and so that $a = 1$ and (2) becomes

(3)
$$\frac{d}{dt}\theta(0) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} , \qquad \frac{d^2}{dt^2}\theta(0) = \begin{pmatrix} b \\ 1/\rho \\ 0 \\ \vdots \\ 0 \end{pmatrix} .$$

In this form ρ is the radius of curvature of the curve $\theta(t)$ at $t = 0$. The value of $1/\rho$ is sometimes referred to as the *statistical curvature* of the family at θ_0. Its magnitude is uniquely determined by the above reduction process. Alternately, in an arbitrary curved exponential family it has the formula

(4)
$$\rho^{-1}(t_0) = \left(\frac{|M|}{m_{11}^3}\right)^{\frac{1}{2}}$$

where

$$M = \begin{pmatrix} \dot{\theta}'\Sigma\dot{\theta} & \dot{\theta}'\Sigma\ddot{\theta} \\ \ddot{\theta}'\Sigma\dot{\theta} & \ddot{\theta}'\Sigma\ddot{\theta} \end{pmatrix}$$

with $\dot{\theta} = \dot{\theta}(t_0)$, $\ddot{\theta} = \ddot{\theta}(t_0)$, $\Sigma = \Sigma(\theta(t_0))$. See Efron (1975).

Remark on Notation. The general functional notation $\theta(\cdot)$ was introduced in 3.7(1) as $\theta(x) = \xi^{-1}(x)$. We will continue to use this general notation in contexts not involving specific differentiable subfamilies. In contexts involving differentiable subfamilies the notation $\theta(\cdot)$ will usually refer to a (local) parametrization of the subfamily; if so, this fact will be explicitly noted. Although this means that the very convenient notation $\theta(\cdot)$ can henceforth have either of two meanings we hope there will be no confusion -- simply remember that $\theta(\cdot)$ is defined by 3.7(1) except where explicitly stated otherwise.

3.12 Example

Let Z have exponential density, $f_\lambda(z) = e^{-\lambda z}\chi_{(0,\infty)}(z)$, relative to Lebesgue measure. Let $T > 0$ be a fixed constant. Let Y be the truncated variable $Y = \min(Z, T)$ and $X(y) \in R^2$ be

$$X(y) = \begin{cases} (y, 0) & \text{if} \quad y < T \\ (y, 1) & \text{if} \quad y = T \end{cases} \quad .$$

For $\lambda \in (0, \infty)$ the distribution of X form a standard curved exponential family. The dominating measure ν is composed of linear Lebesgue measure on the line $((0, T) \times 0)$ plus a point mass on $(T, 1)$. The parameter space for this family is

(1) $$\Theta = \{\theta \in R^2 : \theta_1 = -\lambda, \ \theta_2 = -\ln \lambda, \ \lambda \in (0, \infty)\} \quad ,$$

and

(2) $$\psi(\theta) = \log \left[\frac{1}{\theta_1}(e^{\theta_1 T} - 1) + e^{\theta_1 T+\theta_2}\right] \quad .$$

(The natural parameter space is R^2, since ν has bounded support.) Figure 1 displays both Θ and K on a single plot.

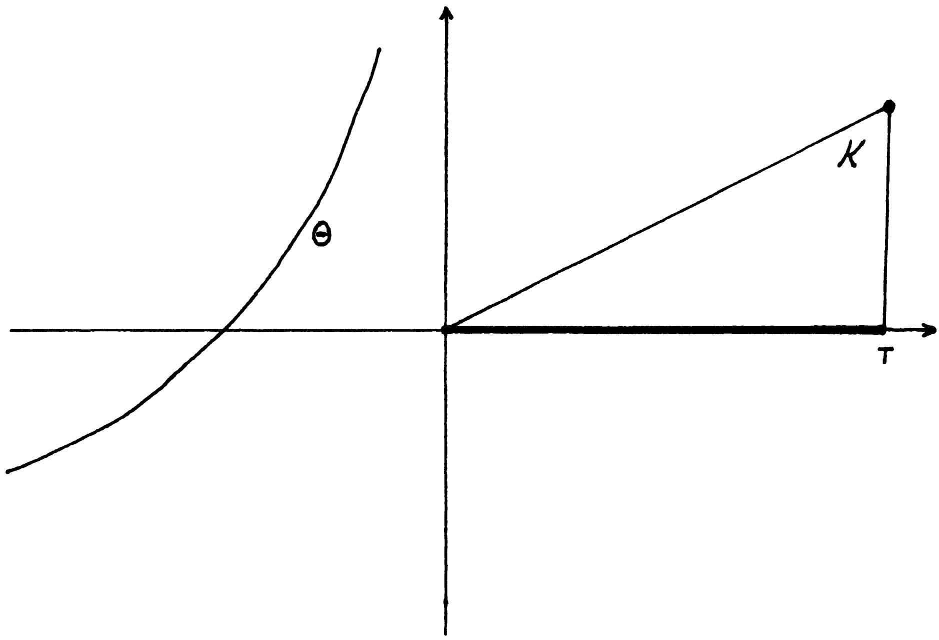

Figure 3.12(1):

Θ and K for Example 3.12.

We return to this example in Chapter 5.

EXERCISES

3.4.1

Let $X_1, X_2, \ldots, X_n$ be a sample from a population with the inverse Gaussian distribution 3.4(1). (i) Show that $S = \sum_{i=1}^{n} X_i$ also has an inverse Gaussian distribution with parameters $\theta_1, n^2\theta_2$. [Examine $E(e^{tS})$.] (ii) Show that S and $(X_i^{-1} - \bar{X}^{-1})$ are independent. [(i) shows that $(S, \frac{n^2}{S}) \sim \text{Expf}\ (\theta_1, \theta_2)$. Now use Theorem 2.14.]

3.5.1 (Converse to Lemma 3.5)

Let $v \in R^k$, $\alpha \in R$. Let $K \subset N$ be compact. If $\nu(H^{-+}(v, \alpha)) = 0$ then

(1) $$\limsup_{\rho\to\infty} \sup_{\theta\in K} \lambda(\theta + \rho v)/e^{\rho\alpha} = 0 \quad .$$

Also, if $\nu(H^{+}(v, \alpha)) = 0$ then

(2) $$\limsup_{\rho\to\infty} \sup_{\theta\in K} \lambda(\theta + \rho v)/e^{\rho\alpha} < \infty \quad .$$

(Be careful, these results may be false if $K \not\subset N$.)

In particular, for $\theta \in N$

(3) $$\psi(\theta + \rho v) \to -\infty \quad \text{as} \quad \rho \to \infty$$

if and only if $\nu(\bar{H}^{+}(v, 0)) = 0$.

3.5.2

Let $Z \in K^{\circ}$. Let $\varepsilon' = \inf\{||x - Z||: x \notin K\} > 0$. Show

(1) $$\lim_{||\theta||\to\infty} \left(\frac{\psi(\theta) - \theta\cdot Z}{||\theta||}\right) = \varepsilon' \quad .$$

[Translate to the case where $Z = 0$, using 1.6(3) with $\phi_0 = 0$, $Z_0 = Z$. Then this result is a minor variation of 3.6(3), and could also have been used to establish 3.6(4).]

3.6.1

Is the following assertion a valid converse to Theorem 3.6: Let $\{p_\theta\}$ be a minimal standard exponential family. Then $\xi : N^\circ \to K^\circ$ is a homeomorphism if and only if $\{p_\theta\}$ is steep.(?) [If k = 1 this is easy to prove.]

3.7.1

Define the measure ν on $\{(x_1, x_2) : -\infty < x_1 < \infty\ ,\ x_2 = 0$ or $x_1 = 0,\ x_2 > 0\}$ by

$$\nu((A, 0)) = \int_A c_0 \frac{e^{-|t|}}{1+t^4}\, dt\ , \quad A \subset (-\infty, \infty), \quad \nu((R, 0)) = 1\ ,$$

(1)

$$\nu((0, A)) = \int_A e^{-t}\, dt \qquad A \subset (0, \infty) \quad .$$

(i) Show the exponential family generated by ν has $N = \{\theta: -1 \le \theta_1 \le 1,\ \theta_2 < 1\}$ and is *not* steep. (ii) Show that $\xi(N^\circ) \subsetneq K^\circ = \{x : x_2 \ge 0\}$ and furthermore that $\xi(N^\circ)$ is not even convex. [Show

$$(2) \qquad \xi(N^\circ) = \{\xi : |\xi_1| < c\left[1 - \frac{(\xi_2^2 + 4\xi_2 k)^{\frac{1}{2}} - \xi_2}{2k}\right]$$

for appropriate c, k.] See Efron (1978).

3.9.1

Prove the conclusion of Theorem 3.9 if $\{p_\theta\}$ is minimal and steep. [In the proof of Theorem 3.9 let $\phi_2^0 \in N^\circ$ and show (using Definition 3.2) that $\nu_{\phi_2^0}$ is steep. For ease of proof assume (w.l.o.g.) that M = I.]

3.11.1

Verify the formula 3.11(4) for the statistical curvature of a curved exponential family.

3.11.2

(i) Verify 3.10(1). (ii) Reduce the three-dimensional multinomial family to a two-dimensional minimal family and show that 3.10(1) now corresponds to a curved exponential family. (iii) Fix ζ_1 and calculate the statistical curvature of the resulting family as a function of the remaining parameter, ϕ_3. (iv) For what value(s) of ζ_1, ϕ_3 is the curvature zero? Why?

3.11.3

Consider an m-dimensional differentiable subfamily inside a k parameter exponential family. Write a canonical form for this family analogous to that in 4.14(1) - (3). [The case m = 1 required two canonical parameters -- b,ρ -- in 4.14(3). The general case requires $m + m(m + 1)/2$ parameters.]

3.12.1

Let $\{p_\theta\}$ be a canonical k parameter exponential family.
Let $\inf\{\psi(\theta)\colon \theta \in N\} < C < \sup\{\psi(\theta)\colon \theta \in N\}$
and let $\Theta = \{\theta \in N^\circ\colon \psi(\theta) = C\}$. $\{p_\theta\colon \theta \in \Theta\}$ can be called a *stratum* of $\{p_\theta\colon \theta \in N\}$. (i) Show that $\{p_\theta\colon \theta \in \Theta\}$ is a $(k - 1)$ dimensional differentiable subfamily of $\{p_\theta\colon \theta \in N\}$. (ii) Let $\theta' = (\theta_{(1)}, \theta_{(2)})'$ where $\theta_{(1)}$ is $(k - 1) \times 1$ and $\theta_{(2)}$ is 1×1. Let $\theta(t)$ be any (local) parametrization of $\{p_\theta\colon \theta \in N\}$ with $t \in T \subset R^{k-1}$. Then

$$(1) \qquad \xi_{(1)}(\theta(t)) \cdot \frac{\partial\theta_{(1)}(t)}{\partial t_j} + \xi_{(2)}(\theta(t)) \frac{\partial\theta_2(t)}{\partial t_j} = 0 \quad .$$

(iii) Let $\theta^\circ \in \Theta$ be any point with $\xi_{(2)}(\theta^\circ) \neq 0$. Then on a neighborhood of θ° in Θ one may write $\theta_{(2)}$ as a function of $\theta_{(1)}$ -- i.e. $\theta_{(2)} = \theta_{(2)}(\theta_{(1)})$ -- and

$$(2) \qquad \nabla\theta_{(2)}(\theta_{(1)}) = -\frac{\xi_{(1)}(\theta)}{\xi_{(2)}(\theta)} \quad .$$

3.12.2

Show that the distributions of X described below can be represented as strata of canonical exponential families (See 3.12.1 for definition.)

(i) $X \sim N(\theta, I)$, $||\theta||^2 = C$.

(ii) The distributions of X-(0,1) with X defined in Example 3.12.

(iii) Let $Y_1, Y_2, \ldots$ be i.i.d. from a canonical regular exponential family, $\{p_\phi\}$. Let N be any Markov stopping time (i.e. $\{y: N(y) = n\}$ is measurable with respect to $Y_1,\ldots,Y_n$). Let $S_n = \sum_{i=1}^{n} Y_i$. Let $X = (S_N, N) = (X_{(1)}, X_{(2)})$, and consider only values of ϕ such that $P_\phi(N < \infty) = 1$. [Let $\theta = (\phi \; -\psi(\theta))$ where $\psi(\phi)$ is the cumulant generating function for the original family $\{p_\phi\}$.]

3.12.3

In 3.12.2 (iii) show that 3.12.1(2) is identical to the following conclusion also derivable from the martingale stopping theorem:

$$(1) \qquad E(S_N) = E(Y)\, E(N) .$$

($(S_n - n\,E(Y)$ is a martingale and so (1) also follows from the stopping theorem applied to this martingale.)

3.12.4

(i) For the family in 3.12.1(i) show the statistical curvature is the constant $1/\sqrt{C}$. (ii) Calculate the statistical curvature for the families described in Example 3.12 and Exercise 3.12.1(ii).

3.12.5

A Poisson process on [0, 1] with intensity function $\rho(t) \geq 0$ may be characterized by the property that the number of observations in any interval $(a, b) \subset [0, 1]$ has $P(\int_a^b \rho(t)dt)$ distribution, and the number of observations in disjoint intervals are independent random variables. Let

$T_1 < \ldots < T_Y$ denote the observations from a Poisson process on [0, 1]. Suppose

$$(1) \qquad \rho(t) = \prod_{i=1}^{m} \rho_i^{\alpha_i}(t)$$

where $\rho_i > 0$ are known (measurable) functions on [0, 1] and α_i are unknown parameters. Show that the distributions of $(T_1, \ldots, T_Y, Y)$ form a differentiable subfamily of dimension m in an (m + 1) parameter exponential family. Identify the canonical statistics and observations for this family. Is this family a stratum of the original family? [The conditional distribution of $T_1, \ldots T_Y$ given Y is that of an ordered sample of Y independent observations from a distribution on [0, 1] with density proportional to $\prod_{i=1}^{m} \rho_i^{\alpha_i}(t)$.]

3.12.6

Let Z_{ij} be independent identically distributed variables with a power series distribution:

$$(1) \qquad P(Z_{ij} = z) = C(\lambda)\, h(z)\lambda^z, \quad z=0,1,\ldots, \quad \lambda > 0.$$

Let $Y_0 = 1$ and define $Y_1, \ldots$ inductively as $Y_i = \sum_{j=1}^{Y_i - 1} Z_{ij}$. $Y_0, Y_1, \ldots$ is called the *Galton-Watson* process with generating distribution (1). Fix $2 \leq n < \infty$. Show that the distributions of $Y_0, Y_1, \ldots, Y_n$ form a curved exponential family with natural statistics $(\sum_0^{n-1} Y_j, \sum_0^n Y_j)$ and this curved exponential family is a stratum of the corresponding full exponential family.