MONOGRAFIAS DE MATEMÁTICA Nº 48

EXPONENTIAL DISPERSION MODELS

Bent Jørgensen

I S B N
85-244-0052-8

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO
INSTITUTO DE MATEMÁTICA PURA E APLICADA
Estrada Dona Castorina, 110
22.460 – Rio de Janeiro – RJ

MONOGRAFIAS DE MATEMÁTICA

1) Alberto Azevedo, Renzo Piccinini - INTRODUÇÃO À TEORIA DOS GRUPOS (reprod)
2) Nathan M. Santos - VETORES E MATRIZES (esgotada)
3) Manfredo P. do Carmo - INTRODUÇÃO À GEOMETRIA DIFERENCIAL GLOBAL (esgot.)
4) Jacob Palis Jr. - SISTEMAS DINÂMICOS (esgotada)
5) João Pitombeira de Carvalho - INTRODUÇÃO À ÁLGEBRA LINEAR (esgotada)
6) Pedro J. Fernandez - INTRODUÇÃO À TEORIA DAS PROBABILIDADES (esgotada)
7) R. C. Robinson - LECTURES ON HAMILTONIAN SYSTEMS (esgotada)
8) Manfredo P. do Carmo - NOTAS DE GEOMETRIA RIEMANNIANA (esgotada)
9) Chaim S. Hönig - ANÁLISE FUNCIONAL E O PROBLEMA DE STURM-LIOUVILLE (esgot)
10) Welington de Melo - ESTABILIDADE ESTRUTURAL EM VARIEDADES DE DIMENSÃO 2 ""
11) Jaime Lesmes - TEORIA DAS DISTRIBUIÇÕES E EQUAÇÕES DIFERENCIAIS (esgotada)
12) Clóvis Vilanova - ELEMENTOS DA TEORIA DOS GRUPOS E DA TEORIA DOS ANÉIS ""
13) Jean Claude Douai - COHOMOLOGIE DES GROUPES (esgotada)
14) H. Blaine Lawson Jr. - LECTURES ON MINIMAL SUBMANIFOLDS, Vol. I (esgotada)
15) Elon L. Lima - VARIEDADES DIFERENCIÁVEIS (esgotada)
16) Pedro Mendes - TEOREMAS DE Ω-ESTABILIDADE E ESTABILIDADE ESTRUTURAL EM VARIEDADES ABERTAS (esgotada)
17) Herbert Amann - LECTURES ON SOME FIXED POINT THEOREMS (esgotada)
18) - EXERCÍCIOS DE MATEMÁTICA / IMPA (esgotada)
19) Djairo G. de Figueiredo - NÚMEROS IRRACIONAIS E TRANSCENDENTES (esgotada)
20) C. E. Zeeman - UMA INTRODUÇÃO INFORMAL À TOPOLOGIA DAS SUPERFÍCIES (esgot)
21) Manfredo P. do Carmo - NOTAS DE UM CURSO DE GRUPOS DE LIE (esgotada)
22) Alexander Prestel - LECTURES ON FORMALLY REAL FIELDS (esgotada)
23) Aron Simis - INTRODUÇÃO À ÁLGEBRA (esgotada)
24) Jaime Lesmes - SEMINÁRIO DE ANÁLISE FUNCIONAL (esgotada)
25) Fred Brauer - SOME STABILITY AND PERTURBATION PROBLEM FOR DIFFERENTIAL AND INTEGRAL EQUATIONS (esgotada)
26) Lúcio Rodriguez - GEOMETRIA DAS SUBVARIEDADES (esgotada)
27) Mário Miranda - FRONTIÈRE MINIME
28) Fernando Cardoso - RESOLUBILIDADE LOCAL DE EQUAÇÕES DIFERENCIAIS PARCIAIS (esgotada)
29) Eberhard Becker - HEREDITARILY-PYTHAGOREAN FIELDS AND ORDERINGS OF HIGHER LEVEL
30) Hyman Bass - PROJECTIVE MODULES AND SYMMETRIC ALGEBRAS
31) J. Neyman - PROBABILIDADE FREQÜENTISTA E ESTATÍSTICA FREQÜENTISTA
32) Freddy Dumortier - SINGULARITIES OF VECTOR FIELDS (esgotada)
33) T. M. Viswanathan - INTRODUÇÃO À ÁLGEBRA E ARITMÉTICA (esgotada)
34) F. Javier Thayer - NOTES ON PARTIAL DIFFERENTIAL EQUATIONS
35) Edward Bierstone - THE STRUCTURE OF ORBIT SPACES AND THE SINGULARITIES OF EQUIVARIANT MAPPINGS
36) F. Javier Thayer - THÉORIE SPECTRALE
37) Manfredo P. do Carmo - FORMAS DIFERENCIAIS E APLICAÇÕES
38) Alexander Prestel, Peter Roquette - LECTURES ON FORMALLY p-ADIC FIELDS
39) Yves Lequain, Arnaldo Garcia - ÁLGEBRA: UMA INTRODUÇÃO (esgotada)
40) J. Lucas Barbosa, A. Gervásio Colares - MINIMAL SURFACES IN R^3
41) Pierre H. Bérard - SPECTRAL GEOMETRY: DIRECT AND INVERSE PROBLEMS
42) Pierre H. Bérard - ANALYSIS ON RIEMANNIAN MANIFOLDS AND GEOMETRIC APPLICATIONS: AN INTRODUCTION
43) Felipe Cano Torres - DESINGULARIZATION STRATEGIES FOR THREE-DIMENSIONAL VECTOR FIELDS
44) Otto Endler - TEORIA DOS CORPOS
45) Winfried Bruns, Udo Vetter - DETERMINANTAL RINGS
46) Abramo Hefez - INTRODUÇÃO À GEOMETRIA PROJETIVA
47) Fernando Quadros Gouvêa - FORMAS MODULARES: UMA INTRODUÇÃO
48) Bent Jørgensen - EXPONENTIAL DISPERSION MODELS

Contents

1. INTRODUCTION

Exponential families, together with several other fundamental concepts in statistics date back to the paper by Fisher (1934). Other early contributions to the study of exponential families were Darmois (1935), Koopman (1936) and Pitman (1936). The exponential family continues to play an important role in statistical theory and practice, and many papers devoted to the theory of exponential families are published every year.

A fundamental account of exponential families may be found in the book by Lehman (1959), which influences the subject even today. Almost twenty years later, the book by Barndorff-Nielsen (1978) was published, providing a definitive account of the theory of exponential families, and exploring conditional inference for full exponential families, and using convexity in the estimation theory. Johansen (1979) gives a brief account of the theory of regular exponential families, and complements Barndorff-Nielsen's book in areas such as test theory and asymptotic theory. A more recent account is that of Brown (1986), developing decision theory aspects of exponential families.

Recent developments have provided generalizations of exponential family models, such as reproductive exponential families (Barndorff-Nielsen and Blæsild, 1983, 1988), and exponential dispersion models (Jørgensen, 1986, 1987a). The class of dispersion models introduced by Jørgensen (1983, 1987b) generalizes exponential dispersion models, and includes the class of proper dispersion models, studied by Barndorff-Nielsen and Jørgensen (1989) which is in a sense a dual to the class of exponential dispersion models. While most of the ideas mentioned above are fairly recent, the exponential dispersion model form dates back to Tweedie (1947). However, Tweedie's ideas seemed to be ahead of his own time, and did not have any immediate impact. About twenty-five years later, Nelder and Wedderburn (1972) proposed the same form of distribution, independently of Tweedie, as the error distribution for their class of generalized linear models. The ideas of Nelder and Wedderburn (1972) has had a considerable impact on current statistical methodology, as indicated by the number of papers published on generalized linear models, and a number of books, such as McCullagh and Nelder (1989) and Aitkin, Anderson, Francis and Hinde (1989).

In the present notes, we develop the mathematical theory of exponential dispersion models, including relevant aspects of characteristic function theory. In Chapter 2 we give a brief, but self-contained introduction to the theory of moment generating functions and characteristic functions. Although this material is standard, there is no easily accessible single source that suits our needs. Lukacs (1970) treats characteristic functions, and Kawata (1972) treat Fourier transforms, both in the univariate case.

In Chapter 3 and 4 we introduce exponential families and exponential dispersion models via equivalence relations on the space of (probability) measures on $\mathbf{R}^k$. This leads to a simple mathematical treatment of basic properties of the models. The inspiration for this approach came from the paper Letac and Mora (1990) and other papers by these authors, whose treatment of natural exponential families is closely related to our approach.

2. MOMENT GENERATING FUNCTIONS

The moment generating function is an indispensable tool for handling exponential families and exponential dispersion models. The present chapter summarizes the requisite theory for moment generating functions, characteristic functions and Fourier-Laplace transforms for multivariate distributions.

2.1 Definition and properties of moment generating functions

Let $\mathcal{M}_k$ denote the set of all propability distributions on $\mathbf{R}^k$. For $P \in \mathcal{M}_k$, we define the *moment generating function* of P by

$$M_P(s) = \int e^{s \cdot x} P(dx), \qquad s \in \mathbf{R}^k,$$

and we define the *cumulant generating function* of P by

$$\mathcal{K}_P(s) = \log M_P(s), \qquad s \in \mathbf{R}^k.$$

The *effective domain* of M_P, respectively $\mathcal{K}_P$, is defined by

$$\Theta_P = \{s \in \mathbf{R}^k \colon\ M_P(s) < \infty\}.$$

If X is a random vector with distribution P we write M_X instead of M_P, and similarly $\mathcal{K}_X$ and Θ_X instead of $\mathcal{K}_P$ and Θ_P, respectively. When no confusion arises we sometimes omit the subscript P, respectively X, etc.

Theorem 2.1.1. *Assume that $X \sim P$, where $P \in \mathcal{M}_k$. Then*

(i) $0 < M_P(s) \le \infty$ *for* $s \in \mathbf{R}^k$.

(ii) $M_P(0) = 1$ *and* $\mathcal{K}_P(0) = 0$.

(iii) If B is an $\ell \times k$ matrix and c an $\ell \times 1$ vector then

$$\Theta_{BX+c} = \{s \in \mathbf{R}^\ell \colon\ B^T s \in \Theta_X\}$$

$$M_{BX+c}(s) = M_X(B^T s) e^{s \cdot c}, \quad s \in \mathbf{R}^\ell\ ,$$

and

$$\mathcal{K}_{BX+c}(s) = \mathcal{K}_X(B^T s) + s \cdot c, \quad s \in \mathbf{R}^\ell.$$

(iv) Let $X = (X_1^T, X_2^T)^T$, where X_1 is d-dimensional and X_2 is $(k-d)$-dimensional, and let $s = (s_1^T, s_2^T)^T$ be a similar partition of s. If X_1 and X_2 are independent, then

$$M_X(s) = M_{X_1}(s_1) M_{X_2}(s_2), \quad s \in \mathbf{R}^k$$

and

$$\mathcal{K}_X(s) = \mathcal{K}_{X_1}(s_1) + \mathcal{K}_{X_2}(s_2), \quad s \in \mathbf{R}^k$$

Proof: See Exercise 2.1.

Example 2.1.2: Let us derive the moment generating function of the multivariate normal distribution. Consider first the case where $X \sim N_k(0, I_k)$. Then for $s \in \mathbf{R}^k$,

$$M_X(s) = (2\pi)^{-k/2} \int \exp(-\frac{1}{2} x \cdot x + s \cdot x) dx$$
$$= (2\pi)^{-k/2} \exp(\frac{1}{2} s \cdot s) \int \exp\{-\frac{1}{2}(x \cdot x - 2s \cdot x + s \cdot s)\} dx$$
$$= \exp(\frac{1}{2} s \cdot s) \int (2\pi)^{-k/2} \exp\{-\frac{1}{2}(x - s) \cdot (x - s)\} dx$$
$$= \exp(\frac{1}{2} s \cdot s).$$

If B is an $\ell \times k$ matrix an μ an $\ell \times 1$ vector then $BX + \mu \sim N_\ell(\mu, \sum)$, where $\sum = BB^T$. By Theorem 2.1.1 (iii) we have

$$M_{BX+\mu}(s) = \exp\left\{\frac{1}{2}(B^T s) \cdot B^T s + s \cdot \mu\right\}$$
$$= \exp\left\{\frac{1}{2} s^T \Sigma s + s \cdot \mu\right\}, \quad s \in \mathbf{R}^\ell,$$

which is hence the moment generating function of the ℓ-variate normal distribution $N_\ell(\mu, \sum)$. Since the matrix B was arbitrary, the result holds even if $\sum$ is singular. ∎

The next theorem concerns convexity properties of M_P and $\mathcal{K}_P$. But first we introduce *Hölder's inequality.*

Proposition 2.1.3. *Let $P \in \mathcal{M}_k$. Then for any $s_1, s_2 \in \mathbf{R}^k$ and $0 \leq \alpha \leq 1$ we have*

$$\int \exp\{\alpha s_1 \cdot x + (1-\alpha) s_2 \cdot x\} P(dx) \leq \left\{ \int e^{s_1 \cdot x} P(dx) \right\}^{\alpha} \left\{ \int e^{s_2 \cdot x} P(dx) \right\}^{1-\alpha}. \tag{1.1}$$

If $0 < \alpha < 1$ and $s_1 \neq s_2$ the inequality is strict if and only if P is not concentrated on an affine subspace of $\mathbf{R}^k$. ∎

Proof: The logarithm is a concave function, and hence for any $a_1 > 0$, $a_2 > 0$ and $0 \leq \alpha \leq 1$ we have

$$\alpha \log a_1 + (1-\alpha) \log a_2 \leq \log(\alpha a_1 + (1-\alpha) a_2). \tag{1.2}$$

Let $c_i = \int e^{s_i \cdot x} P(dx)$, $i = 1, 2$. If either $c_1 = \infty$ or $c_2 = \infty$ the inequality (1.1) is trivial. If both c_1 and c_2 are finite let

$$a_i = e^{s_1 \cdot x} / c_i, \qquad i = 1, 2. \tag{1.3}$$

Inserting this in (1.2) and taking the exponential function on both sides we get

$$\begin{aligned} &\exp\{\alpha s_1 \cdot x + (1-\alpha) s_2 \cdot x\} / (c_1^{\alpha} c_2^{1-\alpha}) \\ &\leq \alpha e^{s_1 \cdot x} / c_1 + (1-\alpha) e^{s_2 \cdot x} / c_2. \end{aligned} \tag{1.4}$$

By integrating both sides of (1.4) with respect to x we obtain (1.1).

Since log is a strictly concave function we have strict inequality in (1.2) if $0 < \alpha < 1$ and $a_1 \neq a_2$. Hence, if $0 < \alpha < 1$ and $s_1 \neq s_2$, then equality in (1.1) is obtained if and only if in (1.3) $a_1 = a_2$ with probability 1 with respect to P, which is equivalent to

$$(s_1 - s_2) \cdot x = \log(c_1 / c_2) \tag{1.5}$$

with probability 1 with respect to P. Since $s_1 - s_2 \neq 0$, the set of xs that satisfy (1.5) is an affine subspace of $\mathbf{R}^k$, and hence the condition for strict inequality in (1.1) follows. ∎

Using the result of Proposition 2.1.3, we may now obtain the convexity properties of M_P and $\mathcal{K}_P$.

Theorem 2.1.4. *Let $P \in \mathcal{M}_k$. Then*

(i) The set Θ_P is convex.

(ii) M_P is a convex function on Θ_P.

(iii) $\mathcal{K}_P$ is a convex function on Θ_P, and strictly convex if and only if P is not concentrated on an affine subspace of $\mathbf{R}^k$. ∎

Proof: Assume that $s_1, s_2 \in \Theta_P$ and $0 \le \alpha \le 1$. Then by Hölder's inequality

$$\begin{aligned} M_P(\alpha s_1 + (1-\alpha)s_2) &= \int \exp\{\alpha s_1 \cdot x + (1-\alpha)s_2 \cdot x\} P(dx) \\ &\le M_P(s_1)^\alpha M_P(s_2)^{1-\alpha} \end{aligned} \tag{1.6}$$

By the definition of Θ_P we have $M_P(s_i) < \infty$ for $i = 1, 2$, and hence by (1.6) $M_P(\alpha s_1 + (1-\alpha)s_2) < \infty$. This implies that $\alpha s_1 + (1-\alpha)s_2 \in \Theta_P$, and hence Θ_P is convex.

The convexity of $\mathcal{K}_P$ follows from (1.1) by taking logs on both sides. The condition for strict convexity follows from the condition for strict inequality in Hölder's inequality. Finally, the convexity of M_P follows because M_P is the composition of exp and $\mathcal{K}_P$, where exp is convex and increasing and $\mathcal{K}_P$ is convex, see Exercise 2.11. ∎

Example 2.1.5: Consider a multinomial random vector $X = (X_1, \ldots, X_k)^T$ with probability function

$$f(x_1, \ldots, x_k) = \binom{n}{x_1 \ldots x_k} p_1^{x_1} \ldots p_k^{x_k},$$

where $p_i \ge 0$, $x_i \ge 0$, $i = 1, \ldots, k$, $p_1 + \cdots + p_k = 1$ and $x_1 + \cdots + x_k = n$. The moment generating function of X is, for $s = (s_1, \ldots, s_k)^T \in \mathbf{R}^k$,

$$\begin{aligned} M_X(s) & \sum_{\substack{x_1+\cdots+x_k=n \\ x_i \ge 0}} \binom{n}{x_1 \ldots x_k} p_1^{x_1} \cdot \ldots \cdot p_k^{x_k} \exp(s_1 x_1 + \cdots + s_k x_k) \\ &= \sum_{x_1+\cdots+x_k=n} \binom{n}{x_1 \ldots x_k} \prod_{j=1}^{k} \left(\frac{p_j e^{s_i}}{\sum_{i=1}^k p_i e^{s_i}} \right)^{x_j} \left(\sum_{i=1}^{k} p_i e^{s_i} \right)^{x_1+\cdots+x_k} \\ &= \left(\sum_{i=1}^{k} p_i e^{s_i} \right)^n \end{aligned}$$

Since $X_1 + \cdots + X_k = n$ the distribution is concentrated on an affine subspace of $\mathbf{R}^k$, and it is not difficult to verify directly (Exercise 2.2) that the cumulant generating function of X is not strictly convex, in agreement with Theorem 2.1.4 (iii).

Let us consider the linear transformation $Y = BX = (X_1, \ldots, X_{k-1})^T$, where

$$B = \begin{pmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & & \\ 0 & & \cdots & 1 & 0 \end{pmatrix}$$

By Theorem 2.1.1 (iii) the moment generating function of Y is, for $s = (s_1, \ldots, s_{k-1}) \in \mathbf{R}^{k-1}$,

$$M_Y(s) = \left(\sum_{i=1}^{k-1} p_i e^{s_i} + p_k \right)$$

2.2 The characteristic function and the Fourier-Laplace transform

A very important property of the moment generating function is that it is analytic, making the powerful tools of complex function theory available to us.

For this purpose we need to extend the definition of the moment generating function to complex arguments. For $P \in \mathcal{M}_k$ we define the *Fourier-Laplace* transform of P by

$$\tilde{M}_P(z) = \int e^{z \cdot x} P(dx), \qquad z \in \mathbf{C}^k \tag{2.1}$$

More explicitly, let $z = (z_1, \ldots, z_k)^T$ and write $z_j = s_j + it_j$ where i is the imaginary unit. Then

$$\tilde{M}(s_1 + it_1, \ldots, s_k + it_k) = \int e^{s \cdot x} \cos(t \cdot x) P(dx)$$
$$+ \int e^{s \cdot x} \sin(t \cdot x) P(dx),$$

where $s = (s_1, \ldots, s_k)^T$ and $t = (t_1, \ldots, t_k)^T$.

A special case of the Fourier-Laplace transform is the *characteristic function*, defined by $\varphi_P(t) = \tilde{M}_P(it)$ for $t \in \mathbf{R}^k$. Hence

$$\varphi_P(t) = \int e^{it\cdot x} P(dx) = \int \cos(t \cdot x) P(dx)$$
$$+ i \int \sin(t \cdot x) P(dx), \qquad t \in \mathbf{R}^k.$$

We use the same conventions regarding random variables as in the case of moment generating functions, and write $\tilde{M}_X$ instead of $\tilde{M}_P$ if $X \sim P$ etc.

The next theorem summarizes the elementary properties of $\tilde{M}_P$ and φ_P.

Theorem 2.2.1. *Let $P \in \mathcal{M}_k$. Then*

(i) The integral (2.1) is absolutely convergent if and only if $z \in \tilde{\Theta}_P$, where

$$\tilde{\Theta}_P = \Theta_P + i\mathbf{R}^k = \{s + it\colon\ s \in \Theta_P, \quad t \in \mathbf{R}^k\}.$$

(ii) $\tilde{M}_P(s) = M_P(s) \quad$ *for* $\quad s \in \Theta_P$.

(iii) $\tilde{M}_P(it) = \varphi_P(t) \quad$ *for* $\quad t \in \mathbf{R}^k$.

(iv) $\tilde{M}_P(0) = \varphi_P(0) = 1$.

(v) $\left|\tilde{M}_P(z)\right| \le M_P(\mathcal{R}e\ z) \quad$ *for* $\quad z \in \tilde{\Theta}_P$, where $\mathcal{R}e\ z = (\mathcal{R}e\, z_1, \ldots, \mathcal{R}e\ z_k)^T$.

(vi) $|\varphi_P(t)| \le 1 \quad$ *for* $\quad t \in \mathbf{R}^k$.

(vii) Let B be an $\ell \times k$ matrix and c an $\ell \times 1$ vector, and let X have distribution P. Then

$$\tilde{M}_{BX+c}(z) = \tilde{M}_X(B^T z) e^{z\cdot c}, \ B^T Z \in \tilde{\Theta}_P$$

and

$$\varphi_{BX+c}(t) = \varphi_X(B^T t) e^{it\cdot c}, \qquad t \in \mathbf{R}^k$$

viii) Let $X = (X_1^T, X_2^T)^T$ be a partition of X with components of dimension d and $k - d$, respectively, and let $z = (z_1^T, z_2^T)^T$ and $t = (t_1^T, t_2^T)^T$ be similar partitions of $z \in \mathbf{C}^k$ and $t \in \mathbf{R}^k$. If X_1 and X_2 are independent then

$$\tilde{M}_X(z) = \tilde{M}_{X_1}(z_1)\tilde{M}_{X_2}(z), \qquad z \in \tilde{\Theta}_P$$

and

$$\varphi_X(t) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2), \qquad t \in \mathbf{R}^k. \quad ▮$$

Proof: For $z = s + it \in \mathbf{C}^k$ we have

$$|e^{z \cdot x}| = |e^{s \cdot x} e^{it \cdot x}| = e^{s \cdot x}. \tag{2.2}$$

Hence, by majorization we have that the integral (2.1) is convergent if and only if $s = \mathcal{R}e\ z \in \Theta_P$, which shows (i). By (2.2)

$$\left|\tilde{M}_P(z)\right| \leq \int |e^{z \cdot x}| P(dx) = M_P(s),$$

which shows (v) and, for $s = 0$, (vi). The remaining parts of the theorem are trivial, and the proofs are left for the reader. ▮

2.3 Analytic properties of univariate moment generating functions

To facilitate the discussion of the analytic properties we begin with the univariate case, which allows us to rely on the familiar theory for analytic functions in one variable. The results in the multivariate case, treated in Section 2.5, may often be derived from the corresponding univariate result.

Let $P \in \mathcal{M}_1$ and let $\tilde{M}_P$ be the Fourier-Laplace transform of P and M_P the moment generating function of P. The effective domain for M_P is then an interval on the real line, and $\tilde{\Theta}_P$ is the corresponding vertical strip in the complex plane. The following theorem shows that $\tilde{M}_P$ is an analytic function.

Theorem 2.3.1. *Let $P \in \mathcal{M}_1$ and assume that $0 \in$ int Θ_P. Then the Fourier-Laplace transform $\tilde{M}_P$ is analytic on int $\tilde{\Theta}_P$. The Taylor expansion of $\tilde{M}_P$ around 0 is*

$$\tilde{M}_P(z) = \sum_{j=0}^{\infty} \frac{\mu_j(P)}{j!} z^j, \tag{3.1}$$

where

$$\mu_j(P) = \int x^j P(dx)$$

is the jth moment for P, which exists for any $j = 0, 1, 2, \ldots$. The cumulant generating function $\mathcal{K}_P$ is analytic with Taylor expansion

$$\mathcal{K}_P(s) = \sum_{j=0}^{\infty} \frac{\mathcal{K}_j(P)}{j!} s^j \tag{3.2}$$

around 0, where $\mathcal{K}_j(P)$ is the jth cumulant of P, which exists for any $j = 0, 1, 2, \ldots$. In particular, P has mean and variance (writing $X \sim P$)

$$E(X) = \mathcal{K}'_P(0) \quad \mathrm{Var}(X) = \mathcal{K}''_P(0),$$

respectively. If P is not degenerate, then $\mathrm{Var}(X) > 0$. ∎

Proof: Let $z_0 = s_0 + it_0 \in$ int $\tilde{\Theta}_P$ and write $\tilde{M}_P$ as follows for $z - z_0 < \varepsilon$

$$\begin{aligned}\tilde{M}_P(z) &= \int \exp\{(z - z_0)x + z_0 x\} P(dx) \\ &= \int \sum_{j=0}^{\infty} \frac{(z - z_0)^j x^j}{j!} \, e^{z_0 x} P(dx),\end{aligned} \tag{3.3}$$

where $\varepsilon > 0$ is such that $|z - z_0| < \varepsilon$ implies $z \in$ int $\tilde{\Theta}_P$. For any $n \geq 0$ we have

$$\begin{aligned}\left| \sum_{j=0}^{n} \frac{(z - z_0)^j x^j}{j!} \, e^{z_0 x} \right| &\leq \sum_{j=0}^{\infty} \frac{|z - z_0|^j \, |x|^j}{j!} \, |e^{z_0 x}| \\ &= \exp(|z - z_0| \, |x| + s_0 x) \\ &\leq (e^{\varepsilon x} + e^{-\varepsilon x}) e^{s_0 x}\end{aligned} \tag{3.4}$$

We have $s_0 \pm \varepsilon \in \Theta_P$, so the integral of (3.4) with respect to P is finite. By Lebesgue's dominated convergence theorem we may hence interchange integration and summation in (3.3). Thus, for $|z - z_0| < \varepsilon$

$$\tilde{M}_P(z) = \sum_{j=0}^{\infty} \frac{(z-z_0)^j}{j!} \int x^j e^{z_0 x} P(dx),$$

which shows that $\tilde{M}_P$ is analytic on int $\tilde{\Theta}_P$. For $z_0 = 0$ we obtain (3.1), and that $\mu_j(P)$ exists for any $j \geq 0$.

Let Log denote the principal branch of the complex logarithm. Since $M_P(s) > 0$ for $s \in \Theta_P$, $\operatorname{Log} \tilde{M}_P(z)$ exists in the domain

$$R = \{z \in \text{ int } \tilde{\Theta}_P: \tilde{M}_P(z) \notin (-\infty, 0]\} \tag{3.5}$$

and is analytic in R. Hence $\mathcal{K}_P(s) = \log M_P(s)$ is analytic with Taylor series (3.2). Finally, if P is not degenerate, then $\mathcal{K}_P''(0) =$ Var $(X) > 0$, cf. Exercise 2.20. ∎

Example 2.3.2: Let us find the characteristic function of the normal distribution. From Example 2.1.2 we know that the moment generating function of the univariate normal distribution $N(0,1)$ is $\exp\left(\frac{1}{2}s^2\right)$. By analytic continuation, the corresponding characteristic function is $\exp\left\{\frac{1}{2}(it)^2\right\} = \exp\left(-\frac{1}{2}t^2\right)$. By Theorem 2.2.1 (viii) the characteristic function of $N_k(0, \boldsymbol{I}_k)$ is hence

$$\exp(-\frac{1}{2}t_1^2) \cdot \ldots \cdot \exp(-\frac{1}{2}t_k^2) = \exp(-\frac{1}{2}t \cdot t),$$

where $\boldsymbol{t} = (t_1, \ldots, t_k)$. Finally, by Theorem 2.2.1 (vii) we conclude that the characteristic function of the normal distribution $N_k(\mu, \Sigma)$ is

$$\exp(-\frac{1}{2}t^T \Sigma t + it\mu),$$

where we have used the transformation $X \to BX + \mu$, $\quad \Sigma = BB^T$, as in Example 2.1.2. ∎

By Theorem 2.3.1, $M_P(s)$ is continuous and differentiable on int Θ_P, and in the next theorem we show that M_P is continuous at the boundary of Θ_P in the one-dimensional case. This is not the case in general in the multivariate case, cf. Barndorff-Nielsen (1978, p. 105).

Theorem 2.3.3. *Let $P \in \mathcal{M}_1$, and assume that Θ_P has a finite endpoint θ_0. Let $\lim_{\theta\to\theta_0}$ denote either $\lim_{\theta\downarrow\theta_0}$ or $\lim_{\theta\uparrow\theta_0}$, depending on whether θ_0 is the upper or lower endpoint of Θ_0. Then the following two statments are equivalent*

(i) $\theta_0 \in \Theta_P$

(ii) $\lim_{\theta\to\theta_0} M_P(\theta)$ *exists and is finite.*

If (i) or (ii) holds, then $M_P(\theta_0) = \lim_{\theta\to\theta_0} M_P(\theta)$. ∎

Proof: We assume that θ_0 is the lower endpoint of Θ_P, the proof in the opposite case being similar. If $\theta_0 \in \Theta_P$, then $M_P(\theta_0) < \infty$. For $\theta_0 \le \theta \le \theta_0 + \varepsilon$, we have

$$e^{\theta x} \le e^{\theta_0 x} + e^{(\theta_0+\varepsilon)x}. \tag{3.6}$$

For $\theta_0 + \varepsilon \in \Theta_P$ we have $M_P(\theta_0 + \varepsilon) < \infty$, which together with (3.6) and Lebesgue's Dominated Convergence Theorem implies that $\lim_{\theta\downarrow\theta_0} M_P(\theta)$ exists and is equal to $M_P(\theta_0)$. This shows the implication $(i) \Rightarrow (ii)$. Now assume that (ii) holds. By Fatou's Lemma, applied to the sequence of positve functions $e^{\theta x}$ for a sequence of θs, we find that $M_P(\theta_0) \le \lim_{\theta\to\theta_0} M_P(\theta)$. Hence $\theta_0 \in \Theta_P$, concluding the proof. ∎

2.4 The uniqueness theorem for characteristic functions

We now show that a distribution is characterized by its characteristic function. The Fourier-Laplace transform provides a link between the moment generating function M_P and the characteristic function φ_P. This allows us to use the fact that φ_P characterizes P to show that M_P also characterizes P.

Theorem 2.4.1. *Let $P_1, P_2 \in \mathcal{M}_k$ be two distributions such that*

$$\varphi_{P_1}(t) = \varphi_{P_2}(t) \quad for \quad t \in \mathbf{R}^k.$$

Then $P_1 = P_2$. ∎

Proof: We show how a distribution $P \in \mathcal{M}_K$ may be recovered from its characteristic function φ. The starting point of the proof is the fact that

$$e^{-it\cdot s}\varphi(t) = \int e^{it\cdot(x-s)}P(dx).$$

By integrating both sides of this equation with respect to the density function of the normal distribution $N_k(0, a^{-1}I_k)$, we get

$$\begin{aligned}
&\left(\frac{2\pi}{a}\right)^{-k/2} \int \varphi(t) \exp(-it \cdot s - \frac{1}{2}at \cdot t)dt \\
&= \iint e^{it\cdot(x-s)} P(dx) \left(\frac{2\pi}{a}\right)^{-k/2} e^{-1/2at\cdot t} dt \\
&= (\frac{2\pi}{a})^{-k/2} \iint \exp\left\{it \cdot (x-s) - \frac{1}{2}at \cdot t\right\} dtP(dx) \\
&= \int \exp\{-\frac{1}{2a}(s-x) \cdot (s-x)\} P(dx),
\end{aligned}$$

where we have use the result of Example 2.3.2. Hence we have the relation

$$\left(\frac{2\pi}{a}\right)^{-\frac{k}{2}} \int \exp(-it \cdot s - \frac{1}{2}at \cdot t)\varphi(t)dt = \int f_a(s-x)P(dx), \tag{4.1}$$

where f_a is the density function of the normal distribution $N_k(0, a\, I_k)$. The right-hand side of equation (4.1) is the density function of the convolution of P with the normal distribution $N_k(0, a\, I_k)$, whereas the left-hand side depends on P only through φ. Since the convolution on the right-hand side converges in distribution to P as a tends to 0, we may hence recover P from φ, which proves the theorem. ∎

Corollary 2.4.2 (Cramér-Wold). *Let X have distribution $P \in \mathcal{M}_k$. Then P is uniquely determined by the set of marginal distributions of $\theta \cdot X$ for $\theta \in \mathbf{R}^k$.* ∎

Proof: The characteristic function of $\theta \cdot X$ is for $s \in \mathbf{R}$

$$\varphi_{\theta\cdot X}(s) = \int \exp(is\theta \cdot x)P(dx) = \varphi_X(s\theta). \tag{4.2}$$

If the distribution of $\theta \cdot X$ is known for any $\theta \in \mathbf{R}^k$, then by (4.2) the characteristic function $\varphi_X(u)$ is known for any $u = s\theta \in \mathbf{R}^k$. Hence, by the uniqueness theorem, the distribution of X is known. ∎

Using Theorem 2.4.1 we may now show that a univariate analytic moment generating function characterizes its distribution. The corresponding result for the multivariate case is shown in the next section.

Theorem 2.4.3. *Let P_1 and P_2 belong to $\mathcal{M}_1$. If there exists an open set $S \subseteq \Theta_{P_1} \cap \Theta_{P_2}$ such that*

$$M_{P_1}(s) = M_{P_2}(s) \quad \text{for} \quad s \in S, \tag{4.3}$$

then $P_1 = P_2$. ∎

Proof: Let $\theta_0 \in S$ and define

$$Q_i(dx) = \{e^{\theta_0 x}/M_{P_i}(\theta_0)\}P_i(dx), \quad i = 1, 2. \tag{4.4}$$

Then Q_1 and Q_2 are distributions in $\mathcal{M}_1$, and actually (4.4) is an example of a linear exponential family which we study in Chapter 3. The Fourier-Laplace transform of Q_i is

$$\tilde{M}_{Q_i}(z) = \tilde{M}_{P_i}(z + \theta_0)/M_{P_i}(\theta_0), \mathcal{R}e\, z \in \Theta_{P_i} - \theta_0.$$

Defining $R = \operatorname{int} \tilde{\Theta}_{P_1} \cap \operatorname{int} \tilde{\Theta}_{P_2} - \theta_0$, we have $S - \theta_0 \subseteq R$, and by (4.3) $\tilde{M}_{Q_1}$ and $\tilde{M}_{Q_2}$ are identical on $S - \theta_0$. By analytic continuation, $\tilde{M}_{Q_1}$ and $\tilde{M}_{Q_2}$ are hence identical on R, and since R includes the imaginary axis we conclude that Q_1 and Q_2 have the same characteristic function. Hence $Q_1 = Q_2$, and by (4.4) this implies $P_1 = P_2$. ∎

2.4.5 Analytic properties of multivariate moment generating functions

We now generalize the results of Section 2.3 to the multivariate case. We first show that an analytic moment generating function characterizes its distribution.

Theorem 2.5.1. *Let $P \in \mathcal{M}_k$ and let M_P be the moment generating function of P with effective domain Θ_P. If $\operatorname{int} \Theta_P \neq \emptyset$ then M_P characterizes P.* ∎

Proof: Let $\theta_0 \in \operatorname{int} \Theta_P$, and define the distribution Q by

$$Q(dx) = \{e^{\theta_0 \cdot x}/M_P(\theta_0)\}P(dx). \tag{5.1}$$

with moment generating function

$$M_Q(s) = M_P(\theta_0 + s)/M_P(\theta_0), \quad s \in \Theta_P - \theta_0.$$

If $X \sim Q$ and $\theta \in \mathbf{R}^k$, then $\theta \cdot X$ has moment generating function

$$M_{\theta \cdot X}(s) = M_P(\theta_0 + s\theta)/M_P(\theta_0),$$

and since $\operatorname{int} \Theta_P \neq \phi$ we have $\operatorname{int} \Theta_{\theta \cdot X} \neq \phi$ for any $\theta \in \mathbf{R}^k$. By Theorem 2.4.3 the distribution of $\theta \cdot X$ may be recovered from M_Q, which in turn is defined in terms of M_P. Hence, by Corollary 2.4.2, the distribution of Q may be recovered from M_P. By (5.1) P is uniquely determined by Q, and hence the conclusion of the theorem follows. ∎

Corollary 2.5.2. *Let $P \in \mathcal{M}_k$ and assume that $\operatorname{int} \Theta_P \neq \phi$. The the function $s \to M_P(\theta_0 + s\theta)$ is analytic for any $\theta_0 \in \operatorname{int} \Theta_P$ and $\theta \in \mathbf{R}^k$. In particular, M_P is analytic separately in each coordinate.*

Proof: Follows immediately from the proof of Theorem 2.5.1. ∎

As a prologue to the multivariate version of Theorem 2.3.1 we look at the Taylor expansion of the exponential function. The reason is that in the proof of Theorem 2.3.1 the Taylor expansion of the Fourier-Lapace transform was obtained by integrating the Taylor expansion of the exponential function term by term. In the multivariate case we need the expansion of $\exp(z_1 + \cdots + z_k)$, which may be obtained from the Taylor expansion of exp. Thus

$$\begin{aligned}
&\exp(z_1 + \cdots + z_k) \\
&= \sum_{i=0}^{\infty} \frac{(z_1 + \cdots + z_k)^i}{i!} \\
&= \sum_{i=0}^{\infty} \frac{1}{i!} \sum_{\substack{i_1,\ldots,i_k \geq 0 \\ i_1+\cdots+i_k=i}} \binom{i}{i_1 \ldots i_k} z_1^{i_1} \ldots z_k^{i_k} \\
&= \sum_{i_1=\cdots=i_k=0}^{\infty} \frac{z_1^{i_1} \ldots z_k^{i_k}}{i_1 \ldots i_k}.
\end{aligned}$$

This shows that $\exp(z_1 + \cdots + z_k)$ is analytic as a function of the complex arguments $z_1, \ldots, z_k$, by displaying it as the sum of its Taylor series. In general, a function of several complex variables, f, is said to be analytic in a region S if for every $z_0 \in S$, f is given by its Taylor expansion around z_0 in some neighbourhood of z_0,

$$f(z) = \sum_{i_1=\cdots=i_k=0}^{\infty} \frac{\partial^{i_1+\cdots+i_k} f(z_0)}{\partial z_1^{i_1} \ldots \partial z_k^{i_k}} \cdot \frac{(z_1 - z_{01})^{i_1} \cdot \ \ldots \ \cdot (z_k - z_{0k})^{i_k}}{i_1! \ldots i_k!}.$$

The multivariate version of Theorem 2.3.1 is

Theorem 2.5.3. *Let $P \in \mathcal{M}_k$ and assume that $0 \in \operatorname{int} \Theta_P$. Then the Fourier-Laplace transform $\tilde{M}_P$ is analytic on $\operatorname{int} \tilde{\Theta}_P$. The Taylor expansion of $\tilde{M}_P$ is*

$$\tilde{M}_P(z) = \sum_{i_1=\cdots=i_k=0} \frac{\mu_{i_1 \ldots i_k}(P)}{i_1! \cdot \ \ldots \ \cdot i_k!} z_1^{i_1} \cdot \ \ldots \ \cdot z_k^{i_k} \tag{5.2}$$

where

$$\mu_{i_1 \ldots i_k}(P) = \int x_1^{i_1} \ldots x_k^{i_k} P(dx)$$

is the $(i_1, \ldots, i_k)th$ moment of P and $z = (z_1, \ldots, z_k)^T$, $x = (x_1, \ldots, x_k)^T$. The cumulant generating function $\mathcal{K}_P$ is analytic with Taylor expansion around 0.

$$\mathcal{K}_p(s) = \sum_{i_1=\cdots=i_k=0}^{\infty} \frac{\mathcal{K}_{i_1 \ldots i_k}(P)}{i_1! \ldots ! i_k!} s_1^{i_1} \cdot \ \ldots \ \cdot s_k^{i_k}, \tag{5.3}$$

where $\mathcal{K}_{i_1 \ldots i_k}(P)$ is the $(i_1, \ldots, i_k)th$ cumulant of P and $s = (s_1, \ldots, s_k)^T$. In particular, P has moments and cumulants of arbitrary order. ∎

Proof: Let $z_0 = s_0 + it_0 \in \operatorname{int} \tilde{\Theta}_P$ be given, and let $\varepsilon > 0$ be such that $|z - z_0| < \varepsilon$ implies $z \in \operatorname{int} \tilde{\Theta}_P$. Here $|z|$ denotes the Euclidean norm on $\mathbf{C}^k$, obtained by identifying $\mathbf{C}^k$ with the Euclidean space $\mathbf{R}^{2k}$. Using the univariate Taylor expansion of the exponential function we obtain

$$\tilde{M}_P(z) = \int \exp\{(z - z_0) \cdot x + z_0 \cdot x\} P(dx)$$

$$= \int \sum_{i=0}^{\infty} \frac{\{(z - z_0) \cdot x\}^i}{i!} e^{z_0 \cdot x} P(x). \tag{5.4}$$

The following inequality justifies the use of Lebesgue's theorem of dominated convergence, writing $s_0 = (s_{01}, \ldots, s_{0k})^T$ etc,

$$\begin{aligned}
&\left|\sum_{i=0}^{n} \frac{\{(z-z_0)\cdot x\}^i}{i!}. e^{z_0\cdot x}\right| \le \sum_{i=0}^{\infty} \frac{|(z-z_0)\cdot x|^i}{i!}\; e^{s_0\cdot x} \\
&= \exp\{|(z-z_0)\cdot x| + s_0\cdot x\} \\
&\le \exp\{|(z_1-z_{01})|x_1 + \cdots + |(z_k-z_{0k})x_k| + s_0\cdot x\} \\
&\le \prod_{i=1}^{k}\{(e^{\varepsilon x_i} + e^{-\varepsilon x_i})e^{s_{0i}x_i}\}.
\end{aligned} \tag{5.5}$$

We may choose ε such that the rectangle with vertices $s_{0i} \pm \varepsilon$ is contained in $\operatorname{int}\Theta_P$. Then the function (5.5) is integrable with respect to P, and interchanging summation and integration in (5.4) we obtain

$$\begin{aligned}
\tilde{M}_P(z) &= \sum_{i=0}^{\infty} \int \frac{\{(z-z_0)\cdot x\}^i}{i!}\; e^{z_0\cdot x} P(dx) \\
&= \sum_{i=0}^{\infty} \int \frac{1}{i} \sum_{\substack{i_1,\ldots,i_k\ge 0 \\ i_1+\cdots+i_k=i}} \binom{i}{i_1 \ldots i_k} (z_1-z_{01})^{i_1} x_1^{i_1} \cdot \;\ldots\; \cdot (z_k-z_{0k})^{i_k} x_k^{i_k} e^{z_0\cdot x} P(dx) \\
&= \sum_{i_1=\cdots=i_k=0}^{\infty} \frac{(z_1-z_{01})^{i_1}\ldots(z_k-z_{0k})^{i_k}}{i_1!\cdot\ldots\cdot i_k!} \int x_1^{i_1}\cdot\ldots\cdot x_k^{i_k} e^{z_0\cdot x} P(dx),
\end{aligned}$$

which shows that $\tilde{M}_P$ is analytic in $\operatorname{int}\tilde{\Theta}_P$. For $z_0 = 0$ we obtain (5.2). Since $M_P(s) > 0$ for $s \in \Theta_P$, $\operatorname{Log}\tilde{M}_P$ is defined and analytic in a region of $\mathbf{C}^k$ containing $\tilde{\Theta}_P$, and hence $\mathcal{K}_P = \log M_P$ is analytic, with Taylor expansion (5.3). ∎

Corollary 2.5.4. *Let $P \in \mathcal{M}_k$, let $X \sim P$, and assume that $0 \in \operatorname{int}\Theta_P$. Then X has mean vector $\mu = (\mu_1, \ldots, \mu_k)^T$*

$$\mu_i = \frac{\partial \mathcal{K}_P(s_1, \ldots, s_k)}{\partial s_i}\Big|_{s=0}$$

and variance matrix $V = \{v_{ij},\quad i,j = 1, \ldots, k\}$ with

$$v_{ij} = \frac{\partial^2 \mathcal{K}_P(s_1, \ldots, s_k)}{\partial s_i \partial s_j}\Big|_{s=0}.$$

The variance matrix V is positive-definite if and only if P is not concentrated on an affine subspace of $\mathbf{R}^k$. ▮

Proof: The expressions for μ_i and v_{ij} follow from the results of Theorem 5.3. The condition for positive-definiteness of V follows from general properties of variance matrixes, cf. Exercise 2.20. ▮

Exercises

Exercise 2.1: Prove Theorem 2.1.1.

Exercise 2.2: Show that the cumulant generating function of the multinomial distribution is not strictly convex.

Exercise 2.3: Show, without using Theorem 2.1.4, that the cumulant generating function of the multivariable normal distribution $N_k(\mu, \Sigma)$ is strictly convex if and only if Σ is positive-definite.

Exercise 2.4: Use the convexity of the exponential function to show, without using Theorem 2.1.4, that the moment generating function M_P is convex.

Exercise 2.5: Find the moment and cumulant generating functions of the following distributions: degenerate, uniform, exponential, gamma, normal, inverse Gaussian, Poisson, binomial and negative binomial.

Exercise 2.6: Let P be the measure with probability density $\frac{1}{2}\, e^{-|x|}$. Find M_P and Θ_P for this measure.

Exercise 2.7: Let P be the Cauchy distribution with density function $f(x) = 1/\{\pi(1+x^2)\}$. Show that $\Theta_P = \{0\}$.

Exercise 2.8: Show that the moment generating function of a univariate distribution P is strictly convex if and only if P is not concentrated in zero.

Exercise 2.9: Show that the only univariate distributions whose cumulant generating function is not strictly convex are the degenerate distributions.

Exercise 2.10: Let $U = \{a + bt: \ t \in \mathbf{R}\}$ with $a, b \in \mathbf{R}$, $b \neq 0$, denote an affine subspace of $\mathbf{R}^2$. Give an example of a distribution P concentrated on U, but not degenerate, and show that $\mathcal{K}_P$ is not strictly convex.

Exercise 2.11: Show that if f: $A \to B$, and g: $B \to \mathbf{R}$, where $A \subseteq \mathbf{R}^k$ is convex and $B \subseteq \mathbf{R}$ is an interval, and f and g are convex functions and g increasing, then $g \circ f$ is convex. Give conditions under which $g \circ f$ is strictly convex. Show that $\exp\{f(x)\}$ is a convex function for f convex.

Exercise 2.12: Show that if $P \in \mathcal{M}_k$ has bounded support, then $\Theta_P = \mathbf{R}^k$.

Exercise 2.13: Consider the moment generating function M_X in Example 2.1.5 for $k = 2$. Derive the relation between M_X and the moment generating function of the binomial distribution.

Exercise 2.14: Consider the moment generating function M_Y in Example 2.1.5. Give necessary and sufficient conditions for $\mathcal{K}_Y = \log M_Y$ to be strictly convex.

Exercise 2.15: Make a plot of $\tilde{\Theta}_P$ for the distributions mentioned in Exercise 2.5.

Exercise 2.16: Verify the details of the proof of Theorem 2.4.1 in the case $k = 1$.

Exercise 2.17: Show that the cumulants of the standard normal distribution $N(0, 1)$ are $\mathcal{K}_1 = 0$, $\mathcal{K}_2 = 1$ and $\mathcal{K}_j = 0$ for $j \geq 2$.

Exercise 2.18: Write the Taylor expansions (3.1) and (3.2) for each of the distributions of Exercise 2.5. In particular, find mean and variance in each case.

Exercise 2.19: Let P and Q be distributions in $\mathcal{M}_k$ such that there exists an open set S with $0 \in S \subseteq \Theta_P \cap \Theta_Q$. With $M_P(s) = M_Q(s)$ for $s \in S$. Show that $P = Q$. Can yon suggest any improvements to this result?

Exercise 2.20: Show that $\mathrm{Var}(X) \geq 0$ for any random variable X, and that $\mathrm{Var}(X) > 0$, unless X is degenerate. Use this result to show that the variance-covariance matrix V for a random vector X is nonnegative definite, and that V is positive definite, unless X is concentrated on an affine subspace of $\mathbf{R}^k$.

Exercise 2.21: Prove that $X + aY \xrightarrow{d} X$ for $a \to 0$ for any two random vectors X and Y.

3. NATURAL EXPONENTIAL FAMILIES

This chapter contains an account of the theory of exponenital families, providing the relevant background for the study of exponential dispersion models in Chapter 4.

3.1 Linear exponential families

To define a linear exponential family, we consider an equivalence relation among measures on $\mathbf{R}^k$. Let $\tilde{\mathcal{M}}_k$ denote the set of all σ-finite measures on $\mathcal{B}_k$, the class of Borel sets in $\mathbf{R}^k$. Recall that $\mathcal{M}_k \subseteq \tilde{\mathcal{M}}_k$, where $\mathcal{M}_k$ denotes the class of probability measures on $\mathcal{B}_k$. Define the relation $\leftrightarrow$ on $\tilde{\mathcal{M}}_k$ by $\nu \leftrightarrow \nu'$ if and only if there exists a pair $(a, b) \in \mathbf{R}^k \times \mathbf{R}$ such that

$$\nu'(dx) = exp(a \cdot x + b)\nu(dx) \tag{1.1}$$

Proposition 3.1.1. *The relation $\leftrightarrow$ is an equivalence relation in $\tilde{\mathcal{M}}_k$, and the restriction of $\leftrightarrow$ g to $\mathcal{M}_k$ is also an equivalence relation.* ∎

Proof: Taking $\nu = \nu'$ and $a = 0$, $b = 0$ in (1.1) we see that the relation is reflexive. Solving (1.1) for $\nu(dx)$ we obtain

$$\nu(dx) = \exp\{(-a) \cdot x - b\}\nu'(dx),$$

which shows that the relation is symmetric. Finally, if $\nu \leftrightarrow \nu'$ and $\nu' \leftrightarrow \nu''$ and (1.1) holds, then there exists a $(c, d) \in \mathbf{R}^k \times \mathbf{R}$

$$\begin{aligned}\nu''(dx) &= \exp(c \cdot x + d)v'(dx) \\ &= \exp\{(a + c) \cdot x + b + d\}\nu(dx).\end{aligned}$$

Hence $\nu \leftrightarrow \nu''$, so that the relation is transitive, and we have shown that $\leftrightarrow$ is an equivalence relation. The restriction of an equivalence relation to a subset of its domain is again an equivalence relation, and hence the restriction of $\leftrightarrow$ to $\mathcal{M}_k$ is an equivalence relation.

Definition 3.1.2: If $\mathcal{P} \subseteq \mathcal{M}_k$ is an equivalence class with respect to $\leftrightarrow$ we say that $\mathcal{P}$ is a *linear exponential family*. If $\mathcal{A} \subseteq \tilde{\mathcal{M}}_k$ is the equivalence class in $\tilde{\mathcal{M}}_k$ containing $\mathcal{P}$, then for any $\nu \in \mathcal{A}$ we say that $\mathcal{P}$ is the *linear exponential family generated by* ν, which we indicate by writing $\mathcal{P}_\nu$ instead of $\mathcal{P}$. The set $\mathcal{A}$ is called the *basis* of $\mathcal{P}$ and is denoted $\mathcal{B}(P)$. ∎

Let ν be a measure in $\tilde{\mathcal{M}}_k$ and define

$$\mathcal{K}_\nu(s) = \log \int e^{s \cdot x} \nu(dx), \quad s \in \mathbf{R}^k.$$

By analogy with the cumulant generating function of a probability distribution we call $\mathcal{K}_\nu$ the *cumulant transform* of ν, even when ν is not a probability measure. The set

$$\Theta_\nu = \{s \in \mathbf{R}^k \colon \mathcal{K}_\nu(s) < \infty\}$$

is called the *effective domain* for $\mathcal{K}_\nu$.

We now consider the representation of a linear exponential family in terms of probability density functions. Let

$$P(dx) = \exp(a \cdot x + b)\nu(dx),$$

where $(a, b) \in \mathbf{R}^k \times \mathbf{R}$ and $\nu \in \tilde{\mathcal{M}}_k$. If $P \in \mathcal{M}_k$ we have

$$\int e^{a \cdot x + b} \nu(dx) = 1,$$

and hence $b = -\mathcal{K}_\nu(a)$ and $a \in \Theta_\nu$. Hence, the linear exponential family generated by ν is the class of probability measures defined by

$$P_\theta(dx) = \exp\{\theta \cdot x - \mathcal{K}_\nu(\theta)\}\nu(dx), \quad \theta \in \Theta_\nu. \tag{1.2}$$

The parameter θ is called the *canonical parameter* corresponding to the representation (1.2), and Θ_ν is called the *canonical parameter domain.*

It may be convenient to express ν in terms of some kind of standard measure, for example Lebesgue measure or counting measure. If $\nu(dx) = f(x)\rho(dx)$ for some measure $\rho \in \tilde{\mathcal{M}}_k$, then

$$P_\theta(dx) = f(x)\exp\{\theta \cdot x - \mathcal{K}_\nu(\theta)\}\rho(dx). \tag{1.3}$$

Conversely, if a and $\mathcal{K}$ are functions on $\mathbf{R}^k$ such that f is non-negative, and $\rho \in \tilde{\mathcal{M}}_k$ is such that

$$P_\theta(dx) = f(x)e^{\theta \cdot x - \mathcal{K}(\theta)}\rho(dx) \tag{1.4}$$

is a probability measure for any $\theta \in \Theta$, where $\Theta \subseteq \mathbf{R}^k$, then for every $\theta \in \Theta$

$$\mathcal{K}(\theta) = \log \int f(x)e^{\theta \cdot x}\rho(dx)$$

is the cumulant transform of the measure $\nu(dx) = f(x)\rho(dx)$, the family $\mathcal{P} = \{P_\theta\colon \theta \in \Theta\}$ is a subset of the linear exponential family $\mathcal{P}_\nu$, and $\Theta \subseteq \Theta_\nu$.

Example 3.1.3: By Example 2.1.2 the standard normal distribution in $\mathbf{R}^k, N_k(0, I_k)$, has moment generating function $M(s) = \exp(\frac{1}{2}s \cdot s)$. Hence the linear exponential family generated by $N_k(0, I_k)$ is given by

$$\begin{aligned} P_\theta(dx) &= (2\pi)^{-k/2}\exp\{-\frac{1}{2}x \cdot x + \theta \cdot x - \frac{1}{2}\theta \cdot \theta\}dx \\ &= (2\pi)^{-k/2}\exp\{-\frac{1}{2}(x-\theta)\cdot(x-\theta)\}dx, \end{aligned}$$

which is the class of normal distributions $N_k(\theta, I_k)$ where $\theta \in \mathbf{R}^k$. ∎

Example 3.1.4: The Poisson distribution $Po(m)$ has probability function

$$\begin{aligned} P(X = x) &= \frac{m^x}{x!}e^{-m} \\ &= \frac{1}{x!}\exp(\theta x - e^\theta), \quad x \in \mathbf{N}_0, \end{aligned}$$

where $\theta = \log m$. Hence we have a linear exponential family of the form (1.3) with $f(x) = 1/x!$, $\mathcal{K}(\theta) = e^\theta$ and ρ the counting measure on $\mathbf{N}_0$. ∎

The next theorem shows how two representations of the form (1.3) for the same family are related.

Theorem 3.1.5. *Consider a linear exponential family $\mathcal{P}$ with representation (1.3) and let*

$$\tilde{P}_\psi(dx) = g(x)\exp\{\psi \cdot x - \mathcal{K}_\delta(\psi)\}\gamma(dx)$$

be a representation of $\mathcal{P}$ of the same form with $\delta(dx) = g(x)\gamma(dx)$.

Then there exist constants $a \in \mathbf{R}^k$ and $b \in \mathbf{R}$ such that

(i) $\Theta_\delta = \Theta_\nu - a$

(ii) $\mathcal{K}_\delta(\psi) = \mathcal{K}_\nu(\psi + a) + b, \quad \psi \in \Theta_\delta$ ∎

Proof: Since the measures δ and ν generate the same exponential family they are equivalent in the sense of Proposition 3.1.1. Hence there exist constants $a \in \mathbf{R}^k$ and $b \in \mathbf{R}$ such that $\delta(dx) = \exp(a \cdot x + b)\nu(dx)$, and

$$\tilde{P}_\psi(dx) = \exp\{\psi \cdot x - \mathcal{K}_\delta(\psi) + a \cdot x + b\}\nu(dx).$$

Integrating both sides with respect to x we get

$$\mathcal{K}_\delta(\psi) - b = \mathcal{K}_\nu(\psi + a),$$

which implies (i) and (ii). ∎

Corollary 3.1.6. *Any canonical parameter domain for a linear exponential family is convex.* ∎

Proof: Choose ν to be a probability measure in (1.2), then by Theorem 1.1.4, Θ_ν is convex. By Theorem 3.1.5 (i) this implies that any canonical parameter domain is convex. ∎

Having examined the representation of a linear exponential family in terms of probability density functions, we now consider the representation in terms of cumulant generating functions. Let (1.2) represent the linear exponential family generated by ν, and let $\mathcal{K}_\theta$ denote the cumulant generating function of P_θ. Then

$$\mathcal{K}_\theta(s) = \mathcal{K}_\nu(s + \theta) - \mathcal{K}_\nu(\theta), \quad s \in \Theta_\nu - \theta. \tag{1.5}$$

The next theorem shows that any class of cumulant generating functions of the form (1.5) correspond to a linear exponential family.

Theorem 3.1.7. *Let $\mathcal{K}$: $\Theta \to \mathbf{R}$ with $\Theta \subseteq \mathbf{R}^k$ be a given function, and define*

$$\tilde{\mathcal{K}}_\theta(s) = \mathcal{K}(\theta + s) - \mathcal{K}(\theta), \quad s \in \Theta - \theta. \tag{1.6}$$

Suppose there exists a θ_0 in Θ such that on $\Theta - \theta_0$ $\tilde{\mathcal{K}}_{\theta_0}$ is identical to the cumulant generating function of a probability distribution P_0, say. Then for every $\theta \in \Theta, \mathcal{K}_\theta$ is identical on $\Theta - \theta$ to the cumulant generating function of a probability distribution in the linear exponential family generated by P_0. ∎

Proof: Le $\mathcal{P}$ be the linear exponential family generated by P_0. By (1.5), the cumulant generating function of an arbitrary member of $\mathcal{P}$ is, for $\theta_0 + \theta \in \Theta$ and $s \in \Theta - (\theta_0 + \theta)$, given by

$$\begin{aligned}\tilde{\mathcal{K}}_{\theta_0}(\theta + s) - \tilde{\mathcal{K}}_{\theta_0}(\theta) &\doteq \mathcal{K}(\theta_0 + \theta + s) - \mathcal{K}(\theta_0) - \mathcal{K}(\theta_0 + \theta) + \mathcal{K}(\theta_0) \\ &= \tilde{\mathcal{K}}_{\theta_0 + \theta}(s),\end{aligned} \tag{1.7}$$

which shows the theorem. ∎

In Theorem 3.1.7, if $\operatorname{int}\Theta \neq \emptyset$, then by Theorem 2.5.1, the cumulant generating functions (1.6) uniquely determine a linear exponential family, whereas if $\operatorname{int}\Theta = \emptyset$ this not the case, and there may be many linear exponential families corresponding to (1.6). The following theorem, which is similar in contents to Theorem 3.1.7, avoids this problem by using the Fourier-Laplace transform.

Theorem 3.1.8. *Let $\mathcal{P}$ be a family of distributions on $\mathbf{R}^k$. Then $\mathcal{P}$ is a linear exponential family if and only if there exists a $\nu \in \tilde{\mathcal{M}}_k$ such that the distributions in $\mathcal{P}$ have Fourier-Laplace transforms*

$$\tilde{M}_\theta(z) = \tilde{M}(\theta + z)/\tilde{M}(\theta), \quad \theta \in \Theta_\nu, \quad \theta + \mathcal{R}ez \in \Theta_\nu, \tag{1.8}$$

where $\tilde{M}(z) = \int e^{z \cdot x}\nu(dx)$. ∎

Proof: If $\mathcal{P}$ is a linear exponential family, then (1.8) easily follows from (1.2). Conversely, if $\theta_0 \in \Theta_\nu$ is given, the linear exponential family generated by the distribution corresponding to $\tilde{M}_{\theta_0}$ is given by the class of Fourier-Laplace transforms (1.8). ∎

3.2 Support, convex support and affine support

In this and the next section we study the support and the canonical parameter domain of a linear exponential family. Specifically, we consider the case where either of these two sets is contained in an affine subspace of $\mathbf{R}^k$.

Let $\nu \in \tilde{M}_k$ be given. The *support* S_ν of ν is the set of points x in $\mathbf{R}^k$ for which any neighbourhood of x has positive ν-measure. By C_ν we denote the *convex support* of ν, which is the convex hull of S_ν, the smallest convex set which contains S_ν. Similarly, the *affine support* $A_\nu =$ aff S_ν is the smallest affine space containing S_ν. These three sets are nested as follows: $S_\nu \subseteq C_\nu \subseteq A_\nu$.

Theorem 3.2.1. *If two measures ν and ν' in $\tilde{\mathcal{M}}_k$ are equivalent in the sense of Proposition 3.1.1, then $S_\nu = S_{\nu'}$, $C_\nu = C_{\nu'}$ and $A_\nu = A_{\nu'}$. In particular, the probability measures of a linear exponential family all have the same support, the same convex support and the same affine support.* ∎

Proof: The fact that $S_\nu = S_{\nu'}$, follows from equation (1.1), and since the support of a measure ν determines its convex support and affine support, the assertions of the theorem follows. ∎

Theorem 3.2.1 allows us to refer to the common support of the members of a linear exponential family $\mathcal{P}$ as the support of $\mathcal{P}$, which we denote $S_\mathcal{P}$. In a similar way we speak of the convex suppor $C_\mathcal{P}$ of $\mathcal{P}$ and the affine support $A_\mathcal{P}$ of $\mathcal{P}$.

Let δ_a denote the generate distribution with support $\{a\} \subseteq \mathbf{R}^k$. It is easy to see (Exercise 3.2) that the linear exponential family generated by δ_a consists of δ_a alone, whereas the canonical parameter space is $\mathbf{R}^k$ for any representation of the family. Hence, the canonical parameter does not yield a parametrization of the family. In dimension two and higher, it is easy to construct non-trivial examples of this phenomenon. The following example illustrates this.

Example 3.2.2: Let X_1 and X_2 be random variables such that $X_1 = c$, where c is a constant, and $X_2 \sim N(0,1)$. The joint distribution of X_1 and X_2 is hence concentrated on the affine subspace $\{(x_1,x_2)^T \in \mathbf{R}^2\colon x_1 = c\}$ of $\mathbf{R}^2$. The linear exponential family $\mathcal{P}$

generated by this distribution is given by

$$P_{(\theta_1,\theta_2)}(dx) = (2\pi)^{-1/2} \exp\{-\frac{1}{2}(x_2 - \theta_2)^2\}(\delta_c \times \gamma)(dx),$$

where γ is the Lebesgue measure on $\mathbf{R}$. Since (θ_1, θ_2) varies in $\mathbf{R}^2$ and $P_{(\theta_1,\theta_2)} = P_{(\theta_1,\theta_2)}$ for any $\theta_1, \theta_1' \in \mathbf{R}$, (θ_1, θ_2) does not parametrize $\mathcal{P}$. Note that by applying an affine transformation, we may obtain a distribution concentrated on any given affine subspace of $\mathbf{R}^2$. ∎

In the next theorem we identify the cases where a canonical parameter does not parametrize the family.

Theorem 3.2.3. *Let $\mathcal{P}$ be a linear exponential family and let θ be a canonical parameter for $\mathcal{P}$ with domain Θ. Then the mapping $\theta \to P_\theta$ is a parametrization of $\mathcal{P}$ if and only if $A_{\mathcal{P}} = \mathbf{R}^k$.* ∎

Proof: Consider the representation with respect to $\nu \in \mathcal{B}(\mathcal{P})$

$$P_\theta(dx) = \exp\{\theta \cdot x - \mathcal{K}_\nu(\theta)\}\nu(dx)$$

for $\mathcal{P}$, and let $\theta_1, \theta_2 \in \Theta = \Theta_\nu$. Then

$$P_{\theta_1}(B) = P_{\theta_2}(B) \quad \forall B \in \mathcal{B}_k \tag{2.1}$$

if and only if

$$\exp\{\theta_1 \cdot x - \mathcal{K}_\nu(\theta_1)\} = \exp\{\theta_2 \cdot x - \mathcal{K}_\nu(\theta_2)\} \ \ [\nu]$$

which in turn is equivalent to

$$(\theta_2 - \theta_1) \cdot x = \mathcal{K}_\nu(\theta_2) - \mathcal{K}_\nu(\theta_1) \ \ [\nu]. \tag{2.2}$$

If $A_\nu = \mathbf{R}^k$, then (2.2) implies $\theta_1 = \theta_2$, and hence θ parametrizes $\mathcal{P}$, which shows the "if" part of the theorem. If $A_\nu \neq \mathbf{R}^k$, there exists a vector $a \neq 0$ in $\mathbf{R}^k$ such that for every

$x \in A_\nu$ we have $a \cdot x = c$, where c is a constant. Let $\theta_1 \in \Theta_\nu$ be given and let $\theta_2 = \theta_1 + a$. Then

$$\int e^{\theta_2 \cdot x} \nu(dx) = \int \exp\{(\theta_2 - \theta_1) \cdot x + \theta_1 \cdot x\} \nu(dx) =$$

$$= e^c \int e^{\theta_1 \cdot x} \nu(dx) < \infty,$$

which shows that $\theta_2 \in \Theta_\nu$ and $c = \mathcal{K}_\nu(\theta_2) - \mathcal{K}_\nu(\theta_1)$. Hence we have found a pair θ_1, θ_2 in Θ_ν with $\theta_1 \neq \theta_2$ which satisfy (2.2). This implies that θ does not parametrize $\mathcal{P}$. ∎

It $A_\mathcal{P} = \mathbf{R}^k$ for a linear exponential family $\mathcal{P}$, we say that $\mathcal{P}$ has *full affine support*. This means that $\mathcal{P}$ is represented in a space of the smallest possible dimension. We shall elaborate on this point later (Theorem 3.4.2).

3.3 The canonical parameter domain

In the previous section we saw that in case a linear exponential family $\mathcal{P}$ does not have full affine support, the canonical parameter does not parametrize $\mathcal{P}$. In an analogous way, if the canonical parameter space is contained in an affine subspace of $\mathbf{R}^k$, then the members of $\mathcal{P}$ are not characterized by their cumulant generating function.

Example 3.3.1: Let X_1 and X_2 be independent random variables with $X_1 \sim N(\theta, 1)$ and X_2 having a Cauchy distribution with probability density $f(x) = \{\pi(1+x^2)\}^{-1}$. The joint distribution of X_1 and X_2 is hence

$$P_{(\theta_1,\theta_2)}(dx) = (2\pi)^{-1/2} \exp\{-\frac{1}{2}(x_1 - \theta_1)^2\} f(x_2) dx,$$

for $(\theta_1, \theta_2)^T \in \mathbf{R} \times \{0\}$. This shows that we have a linear exponential family with canonical parameter space contained in an affine subspace of $\mathbf{R}^k$. Correspondingly the cumulant generating function for X_2, $\mathcal{K}(s)$, say, exists for $s = 0$ only. ∎

The distinctive feature of Example 3.3.1 is that the canonical parameter space is contained in an affine subspace of $\mathbf{R}^k$. We say that a linear exponential family $\mathcal{P}$ has *open* *kernel* if there exists a canonical parameter space Θ for $\mathcal{P}$ such that $\text{int}\,\Theta \neq \emptyset$. If $\mathcal{P}$ has open kernel, then by Theorem 3.1.5(i) any canonical parameter space for $\mathcal{P}$ has $\text{int}\,\Theta \neq \emptyset$.

To illustrate the concept of open kernel, consider a linear exponential family $\mathcal{P}$, and let $Y = X \cdot a$, where $a \neq 0$ is a vector in $\mathbf{R}^k$, and $X \sim P_\theta \in \mathcal{P}$, where P_θ is given by (1.2), say. Using (1.5), we see that Y has cumulant generating function

$$\mathcal{K}_\theta(sa) = \mathcal{K}_\nu(sa + \theta) - \mathcal{K}_\nu(\theta), \tag{3.1}$$

as a function of s, where $sa \in \Theta_\nu - \theta$. If $\mathcal{P}$ has open kernel, then the domain of variation for s in (3.1) is an interval with non-expty interior, which by Theorem 2.4.3 implies that the marginal distribution of Y is determined by the cumulant generating function (3.1).

3.4 Natural exponential families and affine transformations

The discussion in the previous two sections indicate that desirable properties of a linear exponential family are lost if either its support or its canonical parameter domain are contained in an affine subspace of $\mathbf{R}^k$. This motivates the following definition.

Definition 3.4.1: A linear exponential family with full affine support and open kernel is called a *natural exponential family*. ∎

Based on what has already been said, it seems reasonable to concentrate on natural exponential families, as we shall do in the following. There is little lost in doing so, as the next theorem shows, because any linear exponential family may be transformed into a natural exponential family by an affine transformation. In the following, the notation span(B) denotes the column space of a matrix B.

Theorem 3.4.2. *Let $\mathcal{P}$ be a linear exponential family, let the random vector X have distribution $P \in \mathcal{P}$, and let*

$$Y = BX + c,$$

where B is an $\ell \times k$ matrix and c is an $\ell \times 1$ vector. Let θ with domain Θ denote a canonical parameter for $\mathcal{P}$, and consider the family of distributions of Y. Then *(i) The family of distribution of Y for θ varying in a set of the form*

$$\Theta \cap (\text{ span } (B^T) + d)$$

for some $d \in \mathbf{R}^k$, is a linear exponential family, which we denote $\tilde{\mathcal{P}}_d$.

(ii) The family $\tilde{\mathcal{P}}_d$ has full affine support if and only if rank $(B) = \ell$ and span $(B^T) \subseteq A_{\mathcal{P}} - a$ for some $a \in A_{\mathcal{P}}$.

(iii) The family $\tilde{\mathcal{P}}_d$ has open kernel if and only if span $(B^T) \subseteq$ aff $\Theta - \theta_0$ for some $\theta_0 \in \Theta$.

(iv) The family $\tilde{\mathcal{P}}_d$ is a natural exponential family if and only if rank $(B) = \ell \le k$ and span $(B^T) \subseteq (A_{\mathcal{P}} - a) \cap (\text{ Aff } \Theta - \theta_0)$ for some $\theta_0 \in \Theta$ and $a \in A_{\mathcal{P}}$. If furthermore $(A_{\mathcal{P}} - a) \cap (\text{ Aff } \Theta - \theta_0) \subseteq$ span (B^T), then the family of distributions of Y for P varying in $\mathcal{P}$ is a natural exponential family.

Proof: Following Theorem 3.1.8 we use the Fourier-Laplace transform . Let

$$P_\theta(dx) = \exp\{\theta \cdot x - \mathcal{K}(\theta)\}\nu(dx)$$

be the representation of $\mathcal{P}$ corresponding to the canonical parameter θ. The Fourier-Laplace transform of P_θ is

$$\tilde{M}_\theta(z) = \tilde{M}(\theta + z)/\tilde{M}(\theta), \quad \mathcal{R}e\ z \in \Theta - \theta,$$

where $\tilde{M}(z) = \int \exp(z \cdot x)\nu(dx)$. Hence the Fourier-Laplace transform of Y is

$$\tilde{M}_Y(z) = \tilde{M}(\theta + B^T z)\exp(c \cdot z)/\tilde{M}(\theta), \quad \mathcal{R}e(B^T z) \in \Theta - \theta,$$

where z is now a complex vector in ℓ variables. If $\theta \in \Theta \cap (\text{ span } (B^T) + d)$ we may write $\theta = B^T\psi + d$ for some $\psi \in \mathbf{R}^\ell$, and the Fourier-Laplace transform of Y becomes

$$\tilde{M}_Y(z) = \overline{M}(\psi + z)/\overline{M}(\psi),$$

where $\overline{M}(\psi) = \tilde{M}(B^T\psi + d)\exp(c \cdot \psi)$. By Theorem 3.1.8 the family $\mathcal{P}_d$ is hence a linear exponential family. This shows (i).

To show (ii), we consider a vector u in $\mathbf{R}^{\ell}$. If $\mathcal{P}_d$ does not have full affine support, there exists a $u \neq 0$ such that

$$u \cdot Y = u^T B X + u^T c$$

has a degenerate distribution. This may happen if either the rows of B are linearly dependent, or since $X \in A_{\mathcal{P}}[\mathcal{P}]$, if span (B^T) is not a subspace of $A_{\mathcal{P}} - a$, because then u may be chosen such that $u^T B X$ is constant.

To show (iii) we must look at the domain for the canonical parameter ψ, given by

$$B^T \psi \in \Theta - d$$

in the case of the family $\mathcal{P}_d$. Hence we must have span $(B^T) \subseteq$ Aff $\Theta - d$, because otherwise there would be a linear contraint on ψ.

Finally, the first part of (iv) follows from (ii) and (iii). If $(A_{\mathcal{P}} - a) \cap ($ Aff $\Theta - \theta_0) =$ span (B^T), then $\mathcal{P}_d$ is the whole family of distributions for Y, which implies the second part of (iv). ∎

Example 3.4.3: We illustrate the results of Theorem 3.4.2 by a simple example. Let X_1, X_2 and X_3 be independent random variables with X_1 normally distributed $N(\theta_1, 1)$, $\theta_1 \in \mathbf{R}$, X_2 Cauchy distributed as in Examp 3.3.1 and $X_3 = c$, where c is a constant. The family of joint distributions for $(X_1, X_2, X_3), \mathcal{P}$, is clearly a linear exponential family with support $\mathbf{R}^2 \times \{c\}$ and canonical parameter domain $\mathbf{R} \times \{0\} \times \mathbf{R}$. Hence $\mathcal{P}$ does not have full affine support, nor does it have open kernel. Because of the independence, each of the three variables follow a linear exponential family. Hence, the condition in Theorem 3.4.2 (i) is sufficient, but not necessary for an affine transformation to yield a linear exponential family. As an illustration of (ii), note that the distribution of X_1 has full affine support, but not the distribution of X_3. The distribution of X_1 has open kernel, but not the distribution of X_2. Finally, note that X_1 follows a natural exponential family, and that this corresponds to the condition in (iv). The reader may write down for himself the matrix B in each case and check the conditions of the theorem. ∎

After considering affine transformations of linear exponential families, we now take a brief look at the conditional distribution given an affine transformation. The next theorem treats the simplest case of a partitioned random variable. By combining this with Theorem 4.2 it is possible to treat the general case of conditioning on an arbitrary affine transformation.

Theorem 3.4.4. *Let $\mathcal{P}$ be a linear exponential family, and suppose $\mathcal{P}$ has a representation of the form*

$$P_\theta(dx) = a(x_1, x_2)\exp\{\theta_1 \cdot x_1 + \theta_2 \cdot x_2 - \mathcal{K}(\theta_1, \theta_2)\}(\nu_1 \times \nu_2)(dx), \tag{4.1}$$

where $x = (x_1^T, x_2^T)^T$, $\theta = (\theta_1^T, \theta_2^T)^T \in \mathbf{R}^{d_1} \times \mathbf{R}^{d_2}$ are partitions of x and θ, and where ν_1 and ν_2 are measures on $\mathbf{R}^{d_1}$ and $\mathbf{R}^{d_2}$, respectively. Then for every $\theta \in \Theta$, the conditional distribution of x_1 given x_2 has probability density

$$P_{\theta_1}(dx_1 \mid x_2) = \frac{a(x_1, x_2)\exp(\theta_1 \cdot x_1)}{\int a(x_1, x_2)\exp(\theta_1 \cdot x_1)\nu_1(dx_1)}\, \nu_1(dx_1) \tag{4.2}$$

In particular this conditional distribution is a subset of a linear exponential family with canonical parameter θ_1, independently of the value of θ_2. ∎

Proof: The marginal distribution of x_2 has probability density function

$$P_\theta^{x_2}(dx_2) = \exp\{\theta_2 \cdot x_2 - \mathcal{K}(\theta_1, \theta_2)\} \int a(x_1, x_2)\exp(\theta_1 \cdot x_1)\nu_1(dx_1)\nu_2(dx_2) \tag{4.3}$$

By dividing (4.3) into (4.1) we obtain (4.2), from which the remaining conclusions easily follow. ∎

Exercises

Exercise 3.1: Let P be the uniform distribution on $[0, 1]$. Derive the linear exponential family generated by P.

Exercise 3.2: Find the linear exponential family generated by δ_a, the degenerate distribution in $a \in \mathbf{R}^k$, and show that it consists of δ_a alone. Show that the canonical parameter space of the family is $\mathbf{R}^k$.

Exercise 3.3: Show that the following families of distributions are natural exponential families: exponential, Poisson, binomial, normal $N(\mu, 1)$, gamma (shape parameter know), invess Gaussian (shape parameter known).

Exercise 3.4: Find the linear exponential family generated by the normal distribution $N_k(0, \Sigma)$ for a given positive-definite matrix Σ.

Exercise 3.5: Consider a family of distributions of the form

$$p(x; \theta) = f(x) e^{\theta \cdot x - \mathcal{K}_\nu(\theta)} \rho(dx),$$

where $\nu(dx) = f(x)\rho(dx)$, where $f(x) \geq 0$, $\quad x \in \mathbf{R}^k$, and θ varies in a set Θ. Show that this is a subset of a linear exponential family.

Exercise 3.6: Let $M > 0$ be a function defined on a set $\Theta \subseteq \mathbf{R}^k$ with $\operatorname{int} \Theta \neq \emptyset$, such that for every $\theta \in \Theta$, the function

$$M_\theta(s) = \frac{M(\theta + s)}{M(\theta)}, \quad s \in \Theta - \theta$$

coincides with a moment generating function of some distribution P_θ. Show that the family of distributions $\{P_\theta \colon \theta \in \Theta\}$ is a subset of a linear exponential family.

Exercise 3.7: Let $P \in \mathcal{M}_1$ be a distribution with support $S_P \leq (0, \infty)$, and define

$$\mu_\delta(\theta) = \int x^\delta e^{\theta x} dP \quad \delta \geq 0.$$

Show that $\mu_j(\theta) < \infty$ if and only if $\theta \in \Theta_P$. Show that, for each $\delta \geq 0$, the family of distributions

$$p_\delta(x; \theta)^{dx} = \frac{x^\delta e^{\theta x}}{\mu_\delta(\theta)} dP$$

is a linear exponetial family.

Exercise 3.8: Let $X_1, \ldots, X_n$ be i.i.d random variables with distribution $p_\delta(\cdot;\theta)$, as defined in Exercise 3.7. Show that the distribution of (U, V), where

$$U = \sum_{i=1}^{n} X_i$$

$$V = \sum_{i=1}^{n} \log X_i,$$

is a linear exponential family on $\mathbf{R}^2$.

4. EXPONENTIAL DISPERSION MODELS

This chapter deals with the definition and properties of exponential dispersion models. We introduce a new equivalence relation, which together with the definition of a natural exponential family leads to the definition of what we call a convolution family, from which, in turn, we define exponential dispersion models.

4.1 Convolution families

In Chapter 3 we defined linear exponential families via an equivalence relation in $\tilde{M}_k$, the space of all σ-finite measures on $\mathbf{R}^k$. To define an exponential dispersion model, we introduce a second equivalence relation and study the corresponding equivalence classes.

Let $\overline{M}_k$ denote the subset of $\tilde{M}_k$ consisting of measures ν for which $\operatorname{int}\Theta_\nu \neq \emptyset$ and ν has full affine support, that is $A_\nu = \mathbf{R}^k$. Recall that natural exponential families consist of measures in the intersection $\overline{\mathcal{M}}_k \cap \mathcal{M}_k$, where $\mathcal{M}_k$ is the class of probability measures on $\mathbf{R}^k$.

Now, define a relation $\nleftrightarrow$ in $\overline{\mathcal{M}}_k$ by $\nu \nleftrightarrow \nu'$ if and only if there exists a real number $\lambda \neq 0$ and an open set $O \subset \Theta_\nu \cap \Theta_{\nu'}$ such that

$$\mathcal{K}_{\nu'}(\theta) = \lambda \mathcal{K}_\nu(\theta), \quad \theta \in O. \tag{1.1}$$

If (1.1) holds, we write $\nu' = \nu^{(\lambda)}$.

Lemma 4.1.1. *If $\nu' = \nu^{(\lambda)}$, then $\Theta_\nu = \Theta_{\nu'}$, and $\mathcal{K}_{\nu'}(\theta) = \lambda\mathcal{K}_\nu(\theta)$ for every θ in Θ_ν.* ∎

Proof: Let $\theta_0 \in O$, and define two probability measures P and P' by

$$P(dx) = e^{\theta_0 \cdot x - \mathcal{K}_\nu(\theta_0)}\nu(dx)$$

$$P'(dx) = e^{\theta_0 \cdot x - \mathcal{K}_{\nu'}(\theta_0)}\nu'(dx).$$

The corresponding cumulant generating function is, respectively,

$$\mathcal{K}_P(\theta) = \mathcal{K}_\nu(\theta_0 + \theta) - \mathcal{K}_\nu(\theta_0)$$

and

$$\mathcal{K}_{P'}(\theta) = \mathcal{K}_{\nu'}(\theta_0 + \theta) - \mathcal{K}_{\nu'}(\theta_0).$$

Hence, by Corollary 2.5.2, the functions

$$f_1(s) = \mathcal{K}_{\nu}(\theta_0 + \theta s), \quad s \in \mathbf{R}$$

$$f_2(s) = \mathcal{K}_{\nu'}(\theta_0 + \theta s), \quad s \in \mathbf{R}$$

are analytic in a neighbourhood of zero for any given $\theta \in \mathbf{R}^k$. By (1.1) we have

$$f_2(s) = \lambda f_1(s) \tag{1.2}$$

for $\theta_0 + s\theta \in O$. By analytic continuation, (1.2) holds for any s in the interval defined by $\theta_0 + s\theta \in \operatorname{int}\Theta_\nu$. Similarly, (1.2) holds for any s such that $\theta_0 + s\theta \in \operatorname{int}\Theta_{\nu'}$. By continuity, (Theorem 2.3.3), (1.2) also holds on the boundary of the interval. Since $\theta \in \mathbf{R}^k$ was arbitrary, we conclude that $\Theta_\nu = \Theta_{\nu'}$ and $\mathcal{K}_{\nu'}(\theta) = \lambda\mathcal{K}_\nu(\theta)$ for any θ in Θ_ν. ∎

Proposition 4.1.2. *The relation* $\nleftrightarrow$ *is an equivalence relation in* $\overline{\mathcal{M}}_k$, *and the restriction of* $\nleftrightarrow$ *to* $\mathcal{M}_k \cap \overline{\mathcal{M}}_k$ *is also an equivalence relation.* ∎

Proof: The symmetry and reflexiveness of the relation easily follows from (1.1). Now, assume that $\nu_1 \nleftrightarrow \nu_2$ and $\nu_2 \nleftrightarrow \nu_3$. By Lemma 4.1.1 we have

$$\Theta_{\nu_1} = \Theta_{\nu_2} = \Theta_{\nu_3}.$$

Letting $\nu_2 = \nu_1^{(\lambda_1)}$ and $\nu_3 = \nu_2^{(\lambda_2)}$, we obtain for any θ in Θ_{ν_1}

$$\begin{aligned}\mathcal{K}_{\nu_3}(\theta) &= \lambda_2\mathcal{K}_{\nu_2}(\theta)\\ &= \lambda_2\lambda_1\mathcal{K}_{\nu_1}(\theta),\end{aligned}$$

so that $\nu_3 = \nu_1^{(\lambda_1\lambda_2)}$, and $\nu_3 \nleftrightarrow \nu_1$, showing transitiveness. The relation $\nleftrightarrow$ is hence an equivalence relation in $\overline{\mathcal{M}}_k$, and the same is the case for the restriction of $\nleftrightarrow$ to $\mathcal{M}_k \cap \overline{\mathcal{M}}_k$. ∎

Let $\mathcal{N}_k$ denote the class of natural exponential families on $\mathbf{R}^k$. The equivalence relation $\nleftrightarrow$ gives rise to a relation in $\mathcal{N}_k$, also denoted $\nleftrightarrow$, defined as follows. If $\mathcal{P}_1, \mathcal{P}_2 \in \mathcal{N}_k$ we write $\mathcal{P}_1 \nleftrightarrow \mathcal{P}_2$ if and only if there exists $P_1 \in \mathcal{P}_1$ and $P_2 \in \mathcal{P}_2$ such that $P_1 \nleftrightarrow P_2$. In the following $\mathcal{B}(\mathcal{P})$ denotes the basis for the linear exponential family $\mathcal{P}$, as defined in Section 3.1.

Lemma 4.1.3. *If $\mathcal{P}_1$ and $\mathcal{P}_2$ are natural exponential families such that $\mathcal{P}_2 \nrightarrow \mathcal{P}_1$, then for every $\nu_1 \in \mathcal{B}(\mathcal{P}_1)$ there exists a unique $\nu_2 \in \mathcal{B}(\mathcal{P}_2)$ such that $\nu_1 \nrightarrow \nu_2$ and the value of λ in the relation $\nu_2 = \nu_1^{(\lambda)}$ is independent of ν_1 for $\nu_1 \in \mathcal{B}(\mathcal{P}_1)$. If furthermore $\nu_1 \in \mathcal{P}_1$ then $\nu_2 \in \mathcal{P}_2$.* ∎

Proof: By the definition of $\nrightarrow$ and Lemma 4.1.1, there exist $P_1 \in \mathcal{P}_1$ and $P_2 \in \mathcal{P}_2$, such that

$$\mathcal{K}_{P_2}(\theta) = \lambda \mathcal{K}_{P_1}(\theta), \qquad \theta \in \Theta \tag{1.3}$$

for some $\lambda \neq 0$, where $\Theta = \Theta_{P_1} = \Theta_{P_2}$. For any $\nu_1 \in \mathcal{B}(\mathcal{P}_1)$ and $\nu_2 \in \mathcal{B}(\mathcal{P}_2)$ there exist $a_1, a_2 \in \mathbf{R}^k$ and $b_1, b_2 \in \mathbf{R}$ such that

$$\nu_1(dx) = e^{a_1 \cdot x + b_1} P_1 \cdot (dx)$$

and

$$\nu_2(dx) = e^{a_2 \cdot x + b_2} P_2 \cdot (dx).$$

Hence

$$\mathcal{K}_{\nu_1}(\theta) = \mathcal{K}_{P_1}(\theta + a_1) + b_1 \qquad (\theta \in \Theta - a_1) \tag{1.4}$$

and

$$\mathcal{K}_{\nu_2}(\theta) = \mathcal{K}_{P_2}(\theta + a_2) + b_2 \qquad (\theta \in \Theta - a_2). \tag{1.5}$$

Introducing (1.3) and (1.4) in (1.5) we obtain

$$\begin{aligned}\mathcal{K}_{\nu_2}(\theta) &= \lambda \mathcal{K}_{P_1}(\theta + a_2) + b_2 \\ &= \lambda \mathcal{K}_{\nu_1}(\theta - a_1 + a_2) - \lambda b_1 + b_2 \quad (\theta \in \Theta - a_2).\end{aligned} \tag{1.6}$$

Hence, for

$$a_2 = a_1, \qquad b_2 = \lambda b_1 \tag{1.7}$$

we find that

$$\nu_2 = \nu_1^{(\lambda)}, \tag{1.8}$$

so that for $v_1 \in \mathcal{B}(\mathcal{P}_1)$ given, there exists a $\nu_2 \in \mathcal{B}(\mathcal{P}_2)$ such that $\nu_1 \nrightarrow \nu_2$. With ν_2 given by (1.6) and (1.7) and furthermore $\nu_1 \in \mathcal{P}_1$, then $b_1 = -\mathcal{K}_{P_1}(a_1)$, so

$$\begin{aligned}\nu_2(dx) &= e^{a_1 \cdot x - \lambda \mathcal{K}_{P_1}(a_1)} P_2(dx) \\ &= e^{a_1 \cdot x - \mathcal{K}_{P_2}(a_1)} P_2(dx),\end{aligned}$$

which shows that $\nu_2 \in \mathcal{P}_2$.

We now show that for $\nu_1 \in \mathcal{B}(\mathcal{P}_1)$ given, ν_2 satisfying $\nu_1 \nleftrightarrow \nu_2$ is unique. Assuming $\nu_2 = \nu_1^{(\lambda_1)}$, we hence need to show that a_2 and b_2 satisfy (1.7). By Lemma 4.1.1, we obtain $\Theta_{\nu_1} = \Theta_{\nu_2}$, and since $\Theta_{\nu_1} = \Theta - a_1$ and $\Theta_{\nu_2} = \Theta - a_2$ we find $a_1 = a_2$. Inserting $a_1 = a_2$ and $\mathcal{K}_{\nu_2}(\theta) = \lambda_1 \mathcal{K}_{\nu_1}(\theta)$ in (1.6) we obtain

$$\mathcal{K}_{\nu_1}(\theta)(\lambda - \lambda_1) = \lambda b_1 - b_2 . \tag{1.9}$$

If $\lambda \neq \lambda_1$, $\mathcal{K}_{\nu_1}$ is constant and ν_1 hence degenerate, which contradicts $\nu_1 \in \overline{\mathcal{M}}_k$. Hence $\lambda = \lambda_1$ and $b_2 = \lambda b_1$, so we have shown that b_2 and a_2 must satisfy (1.7). Hence ν_2 is unique.

In particular, for any $Q_1 \in \mathcal{P}_1$ and $Q_2 \in \mathcal{P}_2$ satisfying $Q_1 \nleftrightarrow Q_2$ we have $Q_2 = Q_1^{(\lambda)}$. Hence, the value of λ in (1.8) does not depend on the choice of P_1 and P_2. This concludes the proof. ∎

Proposition 4.1.4. *The relation* $\nleftrightarrow$ *is an equivalence relation in the class of natural exponential families.* ∎

Proof: It is obvious that $\nleftrightarrow$ is reflexive and symmetric, because $\nleftrightarrow$ has these properties in $\overline{\mathcal{M}}_k$. To show the transitiveness, assume that $\mathcal{P}_1 \nleftrightarrow \mathcal{P}_2$ and $\mathcal{P}_2 \nleftrightarrow \mathcal{P}_3$. By the definition of $\nleftrightarrow$ there exist $P_1 \in \mathcal{P}_1$ and $P_2 \in \mathcal{P}_2$ such that $P_1 \nleftrightarrow P_2$. By Lemma 4.1.2 there exists $P_3 \in \mathcal{P}_3$ such that $P_2 \nleftrightarrow P_3$. By the transitiveness of $\nleftrightarrow$ in $\overline{\mathcal{M}}_k$ we have $P_1 \nleftrightarrow P_3$, and hence $\mathcal{P}_1 \nleftrightarrow \mathcal{P}_3$, which shows the transitiveness of $\nleftrightarrow$ in $\mathcal{N}_k$. ∎

Definition 4.1.5. *If* $\overline{\mathcal{P}}$ *is an equivalence class in* $\mathcal{N}_k$ *with respect to* $\nleftrightarrow$ *we say that* $\overline{\mathcal{P}}$ *is a **convolution family**. The **basis** $B(\overline{\mathcal{P}})$ is defined as the union* $\cup B(\mathcal{P})$ *for* $\mathcal{P} \in \overline{\mathcal{P}}$. ∎

A convolution family $\overline{\mathcal{P}}$ is, by definition, a class of natural exponential families. Alternatively, we may view $\overline{\mathcal{P}}$ as the corresponding subset of $\mathcal{M}_k$, in which case we think of $\overline{\mathcal{P}}$ as a statistical model, as it were. We use either point of view in the following, depending on the circumstances.

If $\mathcal{P}$ is a natural exponential family, the equivalence class to which $\mathcal{P}$ belongs is called the ***convolution family generated by*** $\mathcal{P}$. If P is a distribution in $\overline{\mathcal{M}}_k \cap \mathcal{M}_k$, we speak of the convolution family that contains P as the *convolution family generated by* P. Similary, if $\nu \in \mathcal{B}(\overline{\mathcal{P}})$ for some convolution family $\overline{\mathcal{P}}$, where ν is a measure in $\overline{\mathcal{M}}_k$, then $\overline{\mathcal{P}}$ is called the *convolution family generated* ν, and is denoted $\overline{\mathcal{P}}_\nu$.

4.2 Representations of convolution families

We now turn to representations of a convolution family in terns of probability density functions. Let $\nu \in \overline{\mathcal{M}}_k$ be given, and let $\overline{\mathcal{P}_\nu}$ be the convolution family generated by ν. The natural exponential family generated by ν is given by

$$P_\theta(dx) = \exp\{\theta \cdot x - \mathcal{K}_\nu(\theta)\}\nu(dx), \quad \theta \in \Theta_\nu . \tag{2.1}$$

Similarly, the natural exponential family generated by $\nu^{(\lambda)}$ is given by

$$P_\theta^{(\lambda)}(dx) = \exp\{\theta \cdot x - \lambda\mathcal{K}_\nu(\theta)\}\nu^{(\lambda)}(dx), \quad \theta \in \Theta_\nu . \tag{2.2}$$

Note that we may arrive at (2.2) by two routes. First, it is the representation of the natural exponential family generated by $\nu^{(\lambda)}$. Second, the cumulant generating function of (2.2) is $\lambda\{\mathcal{K}_\nu(\theta + s) - \mathcal{K}(\theta)\}$, showing that $P_\theta \twoheadrightarrow P_\theta^{(\lambda)}$, justifying the notation in (2.2). We call (2.2) the *canonical representation of* $\overline{\mathcal{P}}$ *with respect to* ν. By Lemma 4.1.3, the domain of variation for the parameter (θ, λ) in (1.4) is $\Theta_\nu \times \Lambda_\nu$, where Λ_ν is the set of $\lambda \neq 0$ such that the measure $\nu^{(\lambda)}$ exists.

Exercises

Exercise 4.1 Show that each of the following families of distributions are exponential dispersion models: the normal, the gamma, and the inverse Gaussian distributions.

Exercise 4.2 Let $\nu \in \tilde{\mathcal{M}}_k$. Show that

$$A_\nu = A_{\nu(\lambda)} \quad \forall\, \lambda \in \Lambda_\nu .$$

Exercise 4.3 Show that the set of scale transformations of $\mathbf{R}^k$, given by

$$x \longmapsto cx, \quad x \in \mathbf{R}^h,$$

where $c > 0$ is a constant, defines an equivalence relation on $\tilde{\mathcal{M}}_k$. Show that the equivalence classes for this relation on $\mathcal{M}_k$ is the class of scale-models on $\mathbf{R}^k$.

Exercise 4.4 Let $\rho \in \tilde{\mathcal{M}}_k$ be a given measure, and consider the family of distributions

$$P_{\lambda,\theta}(dx) = \exp\{\theta \cdot x - \lambda\mathcal{K}(\theta)\}a(\lambda, x)\rho(dx),$$

for suitable functions $\mathcal{K}$ and a. Show that this is a subset of a convolution family.

Exercise 4.5 Let $\mathcal{P}$ be the natural exponential family generated by the measure

$$\nu(dx) = (x!)^{-1}\rho(dx), \quad x = 0, 1, \ldots,$$

where ρ is the counting measure. Show that $\mathcal{P}$ is the family of Poisson distributions. Show that $\mathcal{P}$ is the convolution family generated by ν. Can this phenomenon be used to characterise the Poisson distribution?

Acknowledgements

I am grateful to Ole Barndorff-Nielsen, Barry Reese James, Rodrigo Salgado Labouriau and José Raúl Martínez for their useful comments on earlier versions of this manuscript.

References

Aitkin, M., Anderson, D., Francis, B., and Hinde, I. (1989). *Statistical Modelling in GLIM* Oxford: Clarendon Press.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory*. Chichester: Wiley.

Barndorff-Nielsen, O. E. and Blæsild, P. (1983). Reproductive exponential families. *Ann. Statist.* **11**, 770-782.

Barndorff-Nielsen, O. E. and Blæsild, P. (1988). Reproductive models. In *Encyclopedia of Statistical Sciences* vol.8 (S. Kotz and N.L. Johnson eds.), p. 86-89.

Barndorff-Nielsen, O.E. and Jørgensen, B. (1989). Proper dispersion models. Research Report 000, Dept. Theor. Statistics, Aarhus University.

Brown, L.D. (1986). *Fundamentals of Statistical Exponential Families with Application in Statistical Decision Theory*. Lecture Notes - Monograph Series Vol. 9 , Hayward, California: Institute of Mathematical Statistics.

Darmois, G. (1935). Sur les lois de probabilité a estimation exhaustive. *C.R. Acad. Sci. Paris*, **260**, 1265-1266.

Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc.* A **144**, 285-307.

Johansen, S. (1979). *Introduction to the Theory of Regular Exponential Families*. Lecture Notes 3, Institute of Mathematical Statistics, University of Copenhagen.

Jørgensen, B. (1983). Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models. *Biometrika* **70**, 19-28.

Jørgensen, B. (1986). Some properties of exponential dispersion models. *Scand. J. Statist.* **13**, 187-198.

Jørgensen, B. (1987a). Exponential dispersion models (with Discussion). *J.R. Statist. Soc.* **B 49**, 127-162.

Jørgensen, B. (1987b). Small-dispersion asymptotics. *Rev. Bras. Prob. Estat.* **1**, 59-90.

Kawata, T. (1972). *Fourier Analysis in Probability Theory*. New York: Academic Press.

Koopman, L.H. (1936). On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.* **39**, 399-409.

Lehman, E.L. (1959). *Testing Statistical Hypotheses*. New York: Wiley.

Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variances. *Ann. Statist.* **18** (To appear).

Lukacs, E. E. (1970). *Characteristic Functions*. Second Edition. London, Griffin.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. Second Edition. London: Chapman and Hall.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *J.R. Statist. Soc.* **A 135**, 370-384.

Pitman, E.J.G. (1936). Sufficient Statistics and intrinsic accuracy. *Proc. Camb. Phil. Soc.* **32**, 567-579.

Tweedie, M.C.K. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proc. Camb. Phil. Soc.* **49**, 41-49.

Impresso na Gráfica do
impa