Chapter 3

Examples and special types of models

1 Introduction

The purpose of the present chapter is to indicate the range of applications of analytic models by showing some examples of models and classes of models that are analytic. Some considerations relating to the indices of the models, as defined in Section 2.5, are included because these quantities are of primary importance in the asymptotic theory, in particular as developed in Chapter 4. We also include a few examples of non-analytic models to emphasize the limitations of the theory.

Let us recall some notations from Chapter 2. We consider models of the form

$$\{(E,\nu); f(y;\beta); \beta \in B \subseteq V\}, \tag{1.1}$$

where E is a space equipped with a σ-algebra, ν is a measure on E, $f(y;\beta)$ is the density at $y \in E$ of a probability measure denoted P_β, indexed by a parameter β ranging over a subset of a finite-dimensional real vector space V. Unless otherwise stated the notation $\|v\|$, for $v \in V$, refers to an arbitrary pre-given norm on V. Since we shall consider sub-models of other models in some cases it may occasionally be convenient to denote the parameter by a symbol different from β.

The differentials of the log-likelihood function are denoted

$$D_k(\beta) = D_\beta \log f(y;\beta) \tag{1.2}$$

for $k \in \mathbf{N}$, $\beta \in B$, and $y \in E$, whenever they exist, and their joint cumulants are denoted

$$\chi_{k_1\cdots k_m}(\beta)(v_1^{k_1},\ldots,v_m^{k_m}) = \mathrm{cum}_\beta\{D_{k_1}(\beta)(v^{k_1}),\ldots,D_{k_m}(\beta)(v^{k_m})\} \tag{1.3}$$

for $k_1,\ldots,k_m \in \mathbf{N}$, $v_1,\ldots,v_m \in V$, and $\beta \in B$, again provided their existence.

2 Linear exponential families

Consider a model with densities

$$f(y;\theta) = a(y)\exp\{\langle\,\theta\,,\,t(y)\,\rangle - \kappa(\theta)\}, \tag{2.1}$$

where $t : E \to W$ is a measurable mapping to a finite-dimensional real vector space W, $a : E \to \mathbf{R}$ is a non-negative measurable function, θ is the parameter ranging over the space

$$\Theta = \left\{\, \theta \in W^* : \mu(\theta) = \int a(y)\exp\{\langle\,\theta\,,\,t(y)\,\rangle\}\,d\nu(y) < \infty \,\right\} \tag{2.2}$$

where W^* is the dual of W and we have used the notation

$$\langle\, w^*\,,\,w\,\rangle = w^*(w), \qquad w \in W, w^* \in W^*, \tag{2.3}$$

and finally

$$\kappa(\theta) = \log\mu(\theta), \qquad \theta \in \Theta. \tag{2.4}$$

A model of this form is called a full exponential family, $\theta \in \Theta$ is called the canonical parameter, and $t(y)$ the canonical sufficient statistic. Such families, and sub-families of these, play an important role in the statistical theory and comprehensive accounts may be found, e.g., in Barndorff-Nielsen (1978) and Brown (1986).

The differential at θ of the function κ is $D\kappa(\theta) \in \mathrm{Lin}(W^*;\mathbf{R})$ which may be interpreted as an element in W, in which case it equals the mean value mapping τ, say, at θ, i.e.,

$$\tau(\theta) = D\kappa(\theta) = \mathrm{E}_\theta\{t(Y)\}. \tag{2.5}$$

The differential of the log-likelihood may then be written as

$$\{D_\theta \log f(y;\theta)\}(w^*) = \langle\, w^*\,,\,t(y) - \tau(\theta)\,\rangle \tag{2.6}$$

for $\theta \in \mathrm{int}(\Theta)$ and $w^* \in W^*$. For any $\theta \in \mathrm{int}(\Theta)$ the statistic $t(Y)$ has finite exponential moments and its cumulant generating function is given by

$$\log \mathrm{E}_\theta \exp\{\langle\, s\,,\,t(Y)\,\rangle\} = \kappa(\theta + s) - \kappa(\theta) \tag{2.7}$$

which exists whenever $\theta + s \in \Theta$. The cumulants of $t(Y)$ are therefore seen to be

$$\mathrm{cum}_k\{t(Y)\} = D^k\kappa(\theta) \tag{2.8}$$

for $k \in \mathbf{N}$, in the distribution corresponding to $\theta \in \mathrm{int}(\Theta)$.

Consider now a model parametrized by $\beta \in B \subseteq V$ in which θ in (2.1) is replaced by the image $\theta(\beta)$ of a mapping

$$\theta : B \to \Theta. \tag{2.9}$$

Then the densities become

$$f(y;\beta) = a(y)\exp\{\langle\, \theta(\beta)\, ,\, t(y)\, \rangle - \kappa[\theta(\beta)]\}. \tag{2.10}$$

In this model the first differential of the log-likelihood is given by

$$D_1(\beta)(v) = \langle\, D\theta(\beta)(v)\, ,\, t(y) - \tau[\theta(\beta)]\, \rangle \tag{2.11}$$

for $\beta \in \mathrm{int}(B)$ and $v \in V$, if θ is differentiable at β and $\theta(\beta)$ belongs to $\mathrm{int}(\Theta)$. Notice that $D\theta(\beta)$ is a mapping in $\mathrm{Lin}(V; W^*)$.

When the mapping θ in (2.9) is linear the model (2.10) is called a *linear exponential family*. In that case $D\theta(\beta)$ does not depend on β and we denote it $D\theta$. Its transpose, denoted $D\theta^T$ is a mapping in $\mathrm{Lin}(W; V^*)$ defined by the relation

$$\langle\, D\theta(v)\, ,\, w\, \rangle = \langle\, v\, ,\, D\theta^T(w)\, \rangle \tag{2.12}$$

for all $v \in V$ and $w \in W$, cf. (1.1.26). Then we may rewrite the model (2.10) as

$$f(y;\beta) = a(y)\exp\{\langle\, \beta\, ,\, D\theta^T(t(y))\, \rangle - \kappa[\theta(\beta)]\}. \tag{2.13}$$

Thus the model is reduced to the form (2.1) with canonical parameter β, canonical sufficient statistic $D\theta^T(t(y))$ and κ replaced by $\kappa \circ \theta$. In the sequel we therefore assume, without loss of generality, that the linear exponential family is of the form (2.1), except that the parameter space is a subset $B \subseteq \Theta$.

The first differential of the log-likelihood is given in (2.6) from which it is seen that for $k \geq 2$ the differential

$$D_k(\beta) = -D^k\kappa(\beta) \tag{2.14}$$

is non-random, where $\beta \in \mathrm{int}(\Theta)$. Thus, at any such point the cumulants of the log-likelihood differentials are given by

$$\chi_{\underbrace{1\cdots 1}_{k}}(\beta) = -\chi_k(\beta) = D^k\kappa(\beta) \tag{2.15}$$

for $k \geq 2$, and

$$\chi_{k_1\cdots k_m}(\beta) = 0 \tag{2.16}$$

for $m \geq 2$ if $k_j \geq 2$ for any j.

The following result shows that the class of analytic models includes the class of linear exponential families.

Lemma 2.1. *Any linear exponential family is analytic at any point $\beta \in \mathrm{int}(B)$, where the interior is interpreted as the interior relative to the smallest affine space spanned by B. Furthermore the index of the model is finite at any such point.*

Proof. We shall verify the conditions (i)–(iv) in Definition 2.2.1. As the set E_1 we take the set of y's with $a(y) > 0$. On this set all densities are positive and consequently the measures are mutually absolutely continuous.

There is no loss of generality in assuming that B is not concentrated on any affine subspace of dimension strictly less than the dimension of W^*, because we might otherwise embed the parameter space B in a vector space of this dimension and then use the device (2.13) to bring the model back into the form (2.1) with a parameter space that would now be of the same dimension as Θ.

Hence, assuming that B is of the same dimension as W^*, any interior point of B is also an interior point of $\Theta \subseteq W^*$. Since κ is known to be analytic at any such point it follows that the conditions (i)–(iii) in Definition 2.2.1 hold, and that there exist two constants $c(\beta) \geq 0$ and $\lambda(\beta) \geq 0$ such that

$$\left|D^k\kappa(\beta)(w^*)^k\right| \leq k!\, c(\beta)^2\lambda(\beta)^{k-2}\|w^*\|^k$$

for all $k \geq 2$ and $w^* \in W^*$. Thus, it is seen from (2.15) and (2.16) that any of the cumulant conditions (v)–(vii) in Theorem 2.4.2 hold, and hence that the model is analytic at β.

Assume now that $I(\beta)(w^*)^2 = 0$ for some $w^* \neq 0$, $w^* \in W^*$, and $\beta \in \text{int}(B)$. Since

$$I(\beta)(w^*)^2 = \text{var}_\beta\{\langle\, w^*\,,\, t(Y)\,\rangle\}$$

all cumulants of $\langle\, w^*\,,\, t(Y)\,\rangle$ must be zero if the Fisher information is zero. It follows from (2.8), (2.14), (2.15) and (2.16) that in that case all the $D^k(\beta)$'s are zero and hence that the model is constant in the direction w^* from β in some neighbourhood of zero. We now use Lemma 2.5.9 to deduce that the index is finite at β. ∎

Note that it follows from Theorem 2.6.1 that for a linear exponential family the index cannot be reduced by any analytic one-to-one reparametrization. In a general sense this suggests that the canonical parameter is the 'best' parametrization in terms of the results that can be obtained by the present theory, cf. Section 4.3, concerning the approximation of the distribution of the maximum likelihood estimator by a normal distribution. However, other parametrizations may provide better approximations in particular cases, because the results only provide bounds on the error, not an expression for the error itself.

3 Curved exponential families

Consider again a family of the form (2.1) and a parametrization

$$\theta : B \to \Theta, \tag{3.1}$$

of the model

$$f(y;\beta) = a(y)\exp\{\langle \theta(\beta), t(y)\rangle - \kappa[\theta(\beta)]\}, \tag{3.2}$$

where $B \subseteq V$ and κ is given in (2.4). Such a model is called a *curved exponential family* although it may, as a special case, be linear. Sometimes this name is understood to imply certain smoothness conditions, such as differentiability of the function θ. From the previous section and from Section 2.5 it is easy to see when such a family is analytic. It essentially requires that the mapping θ is analytic.

Lemma 3.1. *A curved exponential family is analytic at any point $\beta \in \operatorname{int}(B)$ for which the mapping θ is analytic at β and $\theta(\beta) \in \operatorname{int}(\Theta)$. If the representation of the canonical sufficient statistic is minimal, i.e., if $\operatorname{var}_{\theta_0}\{t(Y)\}$ is positive definite for any, and hence all, $\theta_0 \in \Theta$, then the condition that θ is analytic at a point $\beta \in \operatorname{int}(B)$ with $\theta(\beta) \in \operatorname{int}(\Theta)$ is also necessary for the model to be analytic at β. Furthermore, under the same condition of minimality of the representation of $t(y)$, the index of the model is finite at any such point β if and only if $D^k\theta(\beta)(v^k) = 0$ for any $k \in \mathbb{N}$ and $v \in V$ for which $D\theta(\beta)(v) = 0$.*

Proof. Throughout the proof consider any fixed point $\beta \in \operatorname{int}(B)$ for which $\theta(\beta) \in \operatorname{int}(\Theta)$. It follows directly from Lemma 2.1 and Lemma 2.6.1 that the curved exponential family is analytic at β if the mapping θ is analytic at β. The Fisher information at β is given by

$$I(\beta)(v^2) = \left(\operatorname{var}_{\theta(\beta)}\{t(Y)\}\right)\{D\theta(\beta)(v)\}^2$$

which, if the representation of $t(Y)$ is minimal, is zero if and only if $D\theta(\beta)(v) = 0$. If this implies that $D^k\theta(\beta)(v^k) = 0$ for all $k \in \mathbb{N}$ then the analyticity of θ guarantees that $\theta(\beta)$, and hence the model, is constant in the direction v, and from Lemma 2.5.9 it then follows that the index at β is finite. Conversely, if there exists a $v \in V$ and a $k \in \mathbb{N}$ with $D^k\theta(\beta)(v^k) \neq 0$ and $D\theta(\beta)(v) = 0$, then $I(\beta)(v^2) = 0$ but the model is not constant in the direction v from β and it then follows from Corollary 2.5.8 that the index is infinite.

Finally, assume again that the representation of $t(y)$ is minimal, and that the model is analytic at β. Then it follows from condition (iii) in Definition 2.2.1 that the function

$$\langle \theta(\beta), t(y)\rangle - \kappa[\theta(\beta)] \tag{3.3}$$

is analytic at β for all y in a set of probability one. Since $t(y)$ is not concentrated with probability one on any proper subspace of W it is possible to find a collection $y_1, \ldots, y_n$ of points in E such that the function in (3.3) is analytic at any of these points and such that the linear space spanned by $t(y_1), \ldots, t(y_n)$ is equal to W. Hence, by considering pairwise differences between the functions in (3.3) at these points, we deduce that $\langle\, \theta(\beta)\,,\, w\,\rangle$ is analytic at β for any $w \in W$. This implies that θ itself is analytic at β. ∎

For the record we note that

$$D_k(\beta)(v^k) = \langle\, D^k\theta(\beta)(v^k)\,,\, t(y) - \tau[\theta(\beta)]\,\rangle$$
$$-\sum_{m=2}^{k} \sum_{a \in S_m(k)} \frac{k!}{m!} \left\{\prod a_j!\right\}^{-1}$$
$$\times\, \kappa^{(m)}[\theta(\beta)]\,\{D^{a_1}\theta(\beta)(v^{a_1}), \ldots, D^{a_m}\theta(\beta)(v^{a_m})\} \quad (3.4)$$

for $k \in \mathbf{N}$ and $v \in V$, where τ is defined in (2.5), $S_m(k)$ is the set of sequences from (1.2.24), $\kappa^{(m)}$ is the function $\theta \mapsto D^m\kappa(\theta)$, and the sum should be read as zero if $k = 1$. From this equation we see that the cumulants of the log-likelihood differentials are given by

$$\chi_{k_1\cdots k_m}(\beta)(v_1^{k_1}, \ldots, v_m^{k_m}) = \kappa^{(m)}[\theta(\beta)]\left\{D^{k_1}\theta(\beta)(v_1^{k_1}), \ldots, D^{k_m}\theta(\beta)(v_m^{k_m})\right\} \quad (3.5)$$

for $m \geq 2$, $k_j \in \mathbf{N}$, and $v_j \in V$, while $\chi_k(\beta)(v^k)$ equals the second term on the right hand side in (3.4).

4 One-dimensional location models

Let

$$f(y;\beta) = g(y - \beta), \qquad \beta \in \mathbf{R}, \quad (4.1)$$

where $g : \mathbf{R} \to \mathbf{R}_+$ is some density function which is assumed to be positive and analytic throughout $\mathbf{R}$. Because the β-distribution of $Y - \beta$ does not depend on β we see that the β-moments of

$$D_\beta^k \log f(y;\beta) = (-1)^k D_y^k \log g(y - \beta) \quad (4.2)$$

do not depend on $\beta \in \mathbf{R}$. Therefore the model is analytic at all points β if it is analytic at any one point. Because g is assumed to be positive and analytic, the criterion for the model to be analytic is that there exist a constant $\rho \geq 0$ and a function $M : \mathbf{R} \to [0, \infty)$ such that

$$\left|D^k \log g(y)\right| \leq k!\, M(y)\rho^{k-1} \quad (4.3)$$

for all $k \in \mathbf{N}$, and

$$\int \exp\{sM(y)\}g(y)\,dy \leq \infty \quad (4.4)$$

for some $s > 0$. From (4.3) it follows that a necessary condition for the model to be analytic is that the radius of convergence of the analytic function $y \to \log g(y)$

is bounded below by some positive constant R, say, uniformly in $y \in \mathbb{R}$. Assuming this to be the case we may extend the function $\log g(y)$ to be defined and analytic throughout the strip

$$\mathbb{C}_R = \{z \in \mathbb{C} : |\operatorname{Im} z| < R\} \tag{4.5}$$

where $\operatorname{Im} z$ denotes the imaginary part of z. Then Cauchy's inequalities tell us that

$$\left|D^k \log g(y)\right| \leq k!\, r^{-k} M_r(y) \tag{4.6}$$

for any $k \in \mathbb{N}$ and r satisfying $0 < r < R$, where

$$M_r(y) = \sup\{\, |\log g(y + rz) - \log g(y)| : z \in \mathbb{C}, |z| = 1\,\}. \tag{4.7}$$

The function $\rho M_r(y)$ then satisfies (4.3) with $\rho = r^{-1}$ and the problem is whether there exists an $r > 0$ such that $M_r(Y)$ has finite exponential moments. Notice that whenever $y \in \mathbb{R}$, $s > 0$, $0 < r < R$, and $|z| = 1$ satisfy the condition

$$|g(y + rz) - g(y)| < g(y)$$

we have

$$\begin{aligned}
&\exp\{s|\log g(y + rz) - \log g(y)|\} \\
&\quad \leq \exp\left\{ s \sum_{k=1}^{\infty} \frac{1}{k} \left[|g(y + rz) - g(y)|/g(y)\right]^k \right\} \\
&\quad \leq \exp\{s \log[1 - |g(y + rz) - g(y)|/g(y)]\} \\
&\quad = \{1 - |g(y + rz) - g(y)|/g(y)\}^{-s}.
\end{aligned} \tag{4.8}$$

Hence we have proved the following lemma.

Lemma 4.1. *A location model of the form (4.1) is analytic at any point $\beta \in \mathbb{R}$ if the density g is positive and analytic throughout $\mathbb{R}$, the function $z \mapsto \log g(z)$ may be extended to an analytic function without singularities on a complex strip of the form $\mathbb{C}_R$ defined in (4.5), for some $R > 0$, and*

$$\int \sup\left\{ [1 - |g(y + rz) - g(y)|/g(y)]^{-s} : z \in \mathbb{C}, |z| = 1 \right\} g(y)\, dy < \infty \tag{4.9}$$

for some $s > 0$ and $r \in (0, R)$.

The condition that the convergence radius of $\log g(y)$ is bounded below by R uniformly in y has here been replaced by the equivalent condition that the function is analytic without singularities in the strip $\mathbb{C}_R$. The condition (4.9) may also be replaced by another condition which is simpler, although not implied by (4.9).

Lemma 4.2. *A location model of the form (4.1) is analytic at any point $\beta \in \mathbf{R}$ if the density g is positive and analytic throughout $\mathbf{R}$, the function $z \mapsto \log g(z)$ may be extended to an analytic function without singularities on a complex strip of the form $\mathbf{C}_R$ defined in (4.5), for some $R > 0$, and the derivative of the function $z \to \log g(z)$ is bounded on some strip $\mathbf{C}_r$ with $0 < r < R$.*

Proof. Let

$$|D \log g(z)| \leq K < \infty$$

for $z \in \mathbf{C}_r$. Then

$$\exp\{s|\log g(y + rz) - \log g(y)|\} \leq \exp\{srK\}$$

for all $s > 0$, $y \in \mathbf{R}$, and $z \in \mathbf{C}$ satisfying $|z| = 1$. Thus, the function $M_r(Y)$ defined in (4.7) is seen to have finite exponential moments. ∎

Concerning the index we have the following result.

Lemma 4.3. *If a location model of the form (4.1) is analytic, then its Fisher information is positive, its index is finite, and both of these quantities are independent of the parameter $\beta \in \mathbf{R}$.*

Proof. From Lemma 2.5.9 we know that the index is finite if the Fisher information is positive. But

$$\begin{aligned} I(\beta) &= \operatorname{var}_\beta\{D_\beta \log g(y - \beta)\} \\ &= \int \{D_y \log g(y - \beta)\}^2 g(y - \beta)\, dy \\ &= \int \{D \log g(y)\}^2 g(y)\, dy \end{aligned}$$

which is constant and cannot be zero because that would imply $g(y)$ to be constant. Hence the model has finite index at any point β. The invariance argument underlying the computation above is easily extended to show that the index is independent of β. ∎

The results in the three lemmas above generalize directly to the case of multidimensional location models, but we shall not pursue that subject here.

5 Cauchy location model

The following example is included to show that the moment condition (iv) in Definition 2.1 is not as restrictive as it may seem at first. Let

$$f(y;\beta) = g(y-\beta) = \frac{1}{\pi\{1+(y-\beta)^2\}} \tag{5.1}$$

for $y \in \mathbf{R}$ and $\beta \in \mathbf{R}$. Thus P_β is a Cauchy distribution centered at β and the model is a location model as discussed in the previous section. For any $z \in \mathbf{C}$ with $|z| \le r < 1$ we have

$$\log g(y+z) = -\log \pi - \log\{1+(y+z)^2\} \tag{5.2}$$

which is clearly analytic without singularities throughout the strip $\mathbf{C}_r$ from (4.5). Furthermore, if $z = y + i\theta$ where $|\theta| \le r$, we have

$$|D \log g(z)| = \left|\frac{2z}{1+z^2}\right| \tag{5.3}$$

which is bounded throughout the strip $\mathbf{C}_r$ when $r < 1$. It now follows from Lemma 4.2 that the model (5.1) is analytic, and Lemma 4.3 shows that the index is finite.

6 Location and scale models

Let

$$f(y;\beta) = e^\gamma g(\alpha + e^\gamma y), \tag{6.1}$$

where $\beta = (\alpha, \gamma) \in \mathbf{R}^2$, $y \in \mathbf{R}$, and $g : \mathbf{R} \to \mathbf{R}_+$ is a density function which is assumed to be positive and analytic. The (α,γ)-distribution of

$$U = \alpha + e^\gamma Y \tag{6.2}$$

is the standardized distribution with density g corresponding to $\alpha = 0$ and $\gamma = 0$ in (6.1). Thus the model for Y is obtained from the density g by location and scale changes.

The technique used for the location models in Section 4 can be used again to obtain quite similar results.

Lemma 6.1. *A location and scale model of the form (6.1) is analytic at any parameter point $(\alpha,\gamma) \in \mathbf{R}^2$ if the density g is positive and analytic throughout $\mathbf{R}$ and the function $z \mapsto \log g(z)$ may be extended to an analytic function without singularities on a set of the form*

$$A_R = \{\, z \in \mathbf{C} : |\mathrm{Im}\, z| < R(1 + |\mathrm{Re}\, z|)\,\} \tag{6.3}$$

for some $R > 0$, *and*

$$\int_{-\infty}^{\infty} \sup\{\,[1 - |(1+z_2)g(u(1+z_2)+z_1) - g(u)|/g(u)]^{-s} : \\ z_1, z_2 \in \mathbb{C}; |z_1| \le r, |z_2| \le r\,\}g(u)\,du < \infty \tag{6.4}$$

for some $s > 0$ *and* $r \in (0, R)$.

Proof. As in Section 4 we see that the condition we need to prove to show that the model is analytic is that for all $k \in \mathbb{N}$,

$$\|D_k(\alpha,\gamma)(a,b)^k\| \le k!\rho^{k-1}M(y)(a^2+b^2)^{k/2} \tag{6.5}$$

for some $\rho \ge 0$ and $M : \mathbf{R} \to [0,\infty)$ satisfying

$$\mathrm{E}_{(\alpha,\gamma)} \exp\{sM(Y)\} < \infty \tag{6.6}$$

for some $s > 0$. If we let

$$M_r(y) = \sup\{\,|zb + \log g(\alpha + za + e^{\gamma+zb}y) - \log g(\alpha + e^{\gamma}y)| : \\ a, b \in \mathbf{R}, z \in \mathbb{C}; a^2 + b^2 = r^2, |z| = 1\,\} \tag{6.7}$$

for $0 < r < R$, then Cauchy's inequalities show that $\rho M_r(y)$ satisfies (6.5) with $\rho = r^{-1}$. It remains to be shown that $M_r(Y)$ has finite exponential moments for some $r \in (0, R)$.

Let $a^2 + b^2 = r^2$ and $|z| = 1$, $z \in \mathbb{C}$. Then, as in (4.8) we obtain

$$\begin{aligned} &\exp\{s\,|zb + \log g(\alpha + za + e^{\gamma+zb}Y) - \log g(\alpha + e^{\gamma}Y)|\} \\ &\quad \le \{1 - |e^{zb}g(\alpha + za + e^{\gamma+zb}Y) - g(\alpha + e^{\gamma}Y)|/g(\alpha + e^{\gamma}Y)\}^{-s} \\ &\quad = \{1 - |(1+z_2)g(z_1 + (1+z_2)U) - g(U)|/g(U)\}^{-s} \end{aligned}$$

where

$$|z_1| = |\alpha(1 - e^{zb}) + za| \le \alpha r e^{r} + r$$

and

$$|z_2| = |e^{zb} - 1| \le re^{r}.$$

It is now clear that if we choose r sufficiently small then, for any fixed (α,γ), the expectation of the last expression is bounded by an integral of the form (6.4) with another (larger) value of r. ▮

Also as in Section 4 we have the following simpler sufficient condition for the model to be analytic.

Lemma 6.2. *A location and scale model of the form (6.1) is analytic at any parameter point* $(\alpha, \gamma) \in \mathbf{R}^2$ *if the density* g *is positive and analytic throughout* $\mathbf{R}$, *the function* $z \mapsto \log g(z)$ *may be extended to an analytic function without singularities on a set of the form (6.3) for some* $R > 0$, *and*

$$zD \log g(z) \tag{6.8}$$

is bounded on this set.

Proof. Notice first that it follows from the assumptions that also the function

$$D \log g(z)$$

is bounded on the set A_R because the subset with $|z| \leq \epsilon$, say, is compact and the function is continuous. As in the proof of Lemma 6.1 we see that if $a^2 + b^2 = r^2$ and $|z| = 1$ we may write

$$\alpha + za + e^{\gamma + zb} Y = z_1 + (1 + z_2)U$$

where $|z_1|, |z_2| < \delta$ for some $\delta > 0$ that may be made arbitrarily small by an appropriate choice of r. Furthermore we then have

$$(1 + z_2)U = e^{z_3} U \in A_{2\delta}, \qquad |z_3| \leq \log(1 - \delta),$$

and record that

$$D_x \log g(Ue^x) = Ue^x D \log g(z)$$

is bounded for $z = e^x U \in A_{2\delta}$, $x \in \mathbf{C}$. Hence, for any fixed parameter point $(\alpha, \gamma) \in \mathbf{R}^2$ and $s > 0$ we obtain

$$\begin{aligned}\exp\{s\,|zb + \log g(\alpha + za + e^{\gamma + zb}Y) - \log g(\alpha + e^{\gamma}Y)|\} \\ \leq \exp\Big(sr + s\delta \sup\{\,|D \log g(z)| : z \in A_{2\delta}\,\} \\ + s \log(1 - \delta) \sup\{\,|zD \log g(z)| : z \in A_{2\delta}\,\}\Big)\end{aligned}$$

which is uniformly bounded in a, b and z. Therefore $M_r(Y)$ defined in (6.7) has finite exponential moments. ∎

Also the result concerning the index generalizes from the location model although the proof is a bit more complicated.

Lemma 6.3. *If a location and scale model of the form (6.1) is analytic, then its Fisher information is positive definite and its index is finite.*

Proof. Let the parameter be fixed and observe from (6.1) and (6.2) that

$$\begin{aligned}D_1(\alpha, \gamma)(a, b) &= b + (a + be^{\gamma}Y) D \log g(\alpha + e^{\gamma}Y) \\ &= b + (a - b\alpha + bU) D \log g(U)\end{aligned} \tag{6.9}$$

for any $(a,b) \in \mathbf{R}^2$, where $D\log g$ refers to differentiation with respect to the argument of the function g. Thus, the Fisher information may be written

$$I(\alpha,\gamma)(a,b)^2 = \int_{-\infty}^{\infty} \{b + (a - b\alpha + bU)D\log g(u)\}^2 \, du \tag{6.10}$$

which is positive unless

$$D\log g(u) = -b/(a - b\alpha + bu)$$

almost everywhere. This is clearly impossible if $b = 0$ and otherwise it implies that $g(u)$ is proportional to $|a - b\alpha + bu|^{-1}$ which is not integrable. This proves that the Fisher information is positive definite and hence by Lemma 2.5.9 that the index is finite at any parameter value. ∎

The parametrization of the model (6.1) is somewhat unusual but it follows from Theorem 2.6.1 that a reparametrization of the model to, e.g., the usual form

$$f(y;\alpha,\gamma) = \frac{1}{\gamma} g\left(\frac{y-\alpha}{\gamma}\right), \quad \gamma > 0, \tag{6.11}$$

does not affect the conclusion that the model is analytic and the index is finite whenever the conditions of Lemma 6.1 or Lemma 6.2 hold.

For the parametrization in (6.11) we can furthermore show that the index is constant throughout the parameter space.

Lemma 6.4. *Consider the model (6.11), where g is some fixed positive and analytic density function on $\mathbf{R}$. If this model is analytic then its index is finite and constant as a function of the parameter (α,γ).*

Proof. We already know that the index is finite. To show that it is constant we write the log-density as

$$\log f(y;\alpha,\gamma) = -\log\gamma + \log g(u), \tag{6.12}$$

where

$$u = (y-\alpha)/\gamma,$$

and notice that

$$\frac{\partial u}{\partial \alpha} = \frac{-1}{\gamma}, \quad \frac{\partial u}{\partial \gamma} = \frac{-u}{\gamma}.$$

Hence, it follows that

$$D^k \log f(y;\alpha,\gamma) = \gamma^{-k} D^k \log f(u;0,1)$$

where the differentiation in both cases are with respect to the two parameters (u being regarded as fixed in the second case), but in the second case evaluated at the 'standardized' distribution with $\alpha = 0$ and $\gamma = 1$. Since the distribution of U does not depend on the parameters we deduce that

$$\chi_{k_1 \cdots k_m}(\alpha, \gamma) = \gamma^{-(k_1 + \cdots + k_m)} \chi_{k_1 \cdots k_m}(0, 1).$$

It is now clear that the γ's cancel from the two sides of the inequality defining the index of the model in (2.5.3), because γ^{-1} appears as a factor in the Fisher information norm. ∎

7 Cauchy location and scale model

Let $\beta = (\alpha, \gamma) \in \mathbf{R}^2$ and

$$f(y; \beta) = e^{\gamma} g(\alpha + e^{\gamma} y) = \frac{1}{\pi\{1 + (\alpha + e^{\gamma} y)^2\}}, \tag{7.1}$$

where g is the density function of the Cauchy distribution as in Section 5. The model is a location and scale model of the form (6.1). We shall use Lemma 6.2 to show that it is analytic throughout the parameter space. Thus, consider $z = z_1 + iz_2 \in A_R$ from (6.3), with $z_1, z_2 \in \mathbf{R}$. Then, for $R \leq \frac{1}{2}$ we have

$$\begin{aligned} \mathrm{Re}\,(1 + z^2) &= 1 + z_1^2 - z_2^2 \\ &> 1 + z_1^2 - R^2(1 + |z_1|)^2 \\ &> (1 + z_1^2)(1 - 4R^2) \end{aligned}$$

which shows that the function

$$\log g(z) = -\log \pi - \log(1 + z^2)$$

is analytic on the set A_R with $R = \frac{1}{2}$. Next, we see that

$$\begin{aligned} |zD \log g(z)| &= |2z^2| \,/\, |1 + z^2| \\ &\leq 2(z_1^2 + z_2^2) / \,|\mathrm{Re}\,(1 + z^2)| \\ &\leq \left(2z_1^2 + R^2(1 + |z_1|)^2\right) / \{(1 + z_1^2)(1 - 4R^2)\} \end{aligned}$$

which is clearly bounded on A_R if $R < \frac{1}{2}$. It then follows from Lemma 6.2 that the model is analytic. It is furthermore known from Lemma 6.3 that the model has positive definite Fisher information and finite index throughout the parameter space.

8 Uniform distributions

As an obvious example of failure of a model to be analytic consider the following example. Let

$$f(y;\beta) = \begin{cases} \beta^{-1} & \text{if } 0 < y < \beta, \\ 0 & \text{otherwise} \end{cases} \tag{8.1}$$

for $\beta > 0$, such that the β-distribution of Y is uniform on $(0, \beta)$. Since the distributions all have different supports the model is not analytic at any point.

9 Piecewise linear regression

Let $Y_1, \ldots, Y_n$ be independent normally distributed random variables with common variance σ^2 and expectations given by

$$\mathrm{E}Y_i = \begin{cases} \alpha + \gamma_1(x_i - \theta) & \text{if } x_i \leq \theta, \\ \alpha + \gamma_2(x_i - \theta) & \text{if } x_i \geq \theta \end{cases} \tag{9.1}$$

where $x_1, \ldots, x_n$ are (known) covariates and the parameters are α, γ_1, γ_2, θ in $\mathbf{R}$, and $\sigma^2 > 0$. In a neighbourhood of any fixed parameter point, except of those with $\theta = x_i$ for some i, the model agrees with a linear normal model, and therefore it is easily verified that the model is analytic at any point, except at the singularities with $\theta = x_i$. However, if we let n tend to infinity the set of singularities may become more dense and consequently the neighbourhoods $U_0(\beta)$ from Definition 2.2.1 may shrink. The consequence is that although the model is analytic at any parameter point, except on a null-set, the asymptotic theory based on the concept of analytic models may not be of much use.

10 The Weibull distribution

As a less obvious example of a non-analytic model consider the one-parameter family of Weibull distributions with densities

$$f(y;\beta) = \beta y^{\beta-1} \exp\{-y^\beta\} \tag{10.1}$$

for $y > 0$ and $\beta > 0$. The distribution of Y^β is an exponential distribution with expectation 1. Direct computations yield

$$\log f(y;\beta) = \log \beta + (\beta - 1) \log y - y^\beta \tag{10.2}$$

and

$$D_k(\beta) = (-1)^{k-1}(k-1)!\,\beta^{-k} - (\log y)^k y^\beta \tag{10.3}$$

for $k \geq 2$. For $D_k(\beta)$ to have finite exponential moments the integral

$$\int_0^\infty \exp\left\{s\left[x\beta^{-k}(\log x)^k\right] - x\right\} dx \tag{10.4}$$

must be finite for some $s > 0$. Since this is not the case we conclude from Lemma 2.3.2 that the model is not analytic at any parameter point. In fact, all moments of the D_k's exist, but as seen above they fail to have exponential moments.

Notice that the model may be transformed to a scale family based on the density g of $\log X$ where X is exponentially distributed with mean 1. The inclusion of a location parameter would lead to the two-parameter family of Weibull distributions. The conditions in Lemma 6.1 and Lemma 6.2 may be seen directly to fail for this family, in accordance with the result above that the scale model is not analytic.