Chapter 2

The class of analytic models

1 Introduction

Let $\{P_\beta : \beta \in B \subseteq V\}$ be a statistical model, i.e., a family of probability measures on some space E, parametrized by β which takes values in (a subset of) a vector space V of finite dimension, $\dim V = p$, say. Let $f(y;\beta), y \in E$, denote (a version of) the densities with respect to some measure ν on E.

For asymptotic considerations it is often convenient to be able to expand the densities as

$$f(y;\beta) \sim f(y;\beta_0)\exp\left\{D_1(\beta-\beta_0)+\frac{1}{2}D_2(\beta-\beta_0)^2+\cdots\right\}, \qquad (1.1)$$

where $D_k = D_\beta^k \log f(y;\beta_0)$ is the kth differential of the log-likelihood function evaluated at β_0. For asymptotic calculations based on the distribution P_{β_0} the natural procedure would be to base the calculations on a truncated version of this series. If this is 'legitimate' the model is simplified to a curved exponential family. The problem is what magnitude of error is the result of such a truncation, partly for the approximation of the density itself when it is used, e.g., in the construction of the maximum likelihood estimator, and partly for the approximation of the measure P_β for the purpose of investigating properties of estimators and test statistics.

It is the purpose here to define a class of models, the analytic models, which is large enough to cover a wide range of statistical models, and which at the same time is sufficiently well-behaved to allow expansions like the one above. It will be shown, as a goal in itself, that a model which is analytic satisfies most of the commonly used conditions of asymptotic likelihood based inference — except that it may be discrete as well as continuous. To some extent this is the content of the lemmas in Section 3, but in particular this claim will be justified in Chapter 5.

One way of defining the class of analytic models is to take as the starting point the requirement that the infinite series in the exponent in (1.1) is absolutely convergent, not for all β in the parameter space, but for all β in some neighbourhood of β_0. This neighbourhood may not depend on the observation y. In combination with a uniform moment condition on the D_k's this condition guarantees that the model can be embedded locally into an infinite-dimensional exponential family, and it turns out that many of the properties from curved exponential families are preserved. That the embedding is only defined locally is a major difference compared to theories of infinite-dimensional exponential families as defined by various authors, since it is a much weaker requirement than the convergence of the expansion throughout the sample space. This is demonstrated by a number of examples of analytic models in the next chapter.

The requirement of the absolute convergence of the Taylor series expansion of the log-likelihood function around a fixed point is formulated below as the requirement that the log-likelihood function is analytic at that point. One of the regularity conditions of asymptotic results which is most difficult to verify is often an integrability condition on the remainder of a Taylor series expansion of the log-likelihood or a related function, usually formulated as an integrability condition on the supremum of a certain derivative within a neighbourhood of the fixed parameter point. This supremum stems from the 'intermediate point' evaluation of the remainder of the Taylor series expansion in question. Thus, if $g(z)$ is some real (or complex) function, the remainder

$$R_K = g(z) - g(z_0) - \sum_{k=1}^{K} \frac{1}{k!} g^{(k)}(z_0)(z - z_0)^k$$

is usually bounded by $\{(K+1)!\}^{-1}|g^{K+1}(z^*)|\,|z - z_0|^{K+1}$ when g is $K+1$ times continuously differentiable. In statistics, when g is random, the moments of this bound will generally be in question and they are often hard to estimate. Now, if g is *analytic*, i.e., its infinite Taylor series expansion is absolutely convergent and represents the function in a neighbourhood of the chosen point, then the remainder above may be written as

$$R_K = \sum_{K+1}^{\infty} \frac{1}{k!} g^{(k)}(z_0)(z - z_0)^k,$$

which only involves derivatives at the point z_0. Despite the fact that this sum is infinite it is often much easier to handle than the supremum of a certain derivative over some neighbourhood. Therefore it may be easier to verify regularity conditions if it can be assumed that the functions involved are analytic.

For statistical models, however, it is not sufficient to assume that the likelihood function is analytic; for most results moment conditions on the derivatives are also required. Hence the definition of analytic models involves a moment condition on the log-likelihood differentials as well as the condition that the likelihoods are analytic, as the two major requirements.

A quite different approach towards the definition of this class of models starts from the orders of magnitude of the cumulants of the log-likelihood differentials, usually in terms of powers of $n^{-1/2}$ where n is the sample size. Anyone who has been, deriving higher order asymptotic expansions in relation to likelihood based inference will have noticed that the second order terms, i.e., the terms of order $n^{-1/2}$, involve the third cumulant of the score statistic D_1, the covariance between D_1 and D_2, and the mean of D_3, while the third order terms, i.e., the terms of order n^{-1}, involve the fourth cumulant of D_1, the covariance between D_1 and D_3, the third (mixed) cumulant of D_1, D_1 and D_2, and so on. The rule is that an increase by one of either the order of the cumulant or of the order of one of the differentials involved, is accompanied by a decrease of the order of magnitude by the factor $n^{-1/2}$, not in the magnitude of the cumulant itself, but in the order of magnitude of the term in which it appears. On further exploration, cf. Section 4, this fact may be used to define not only the class of analytic models, but also the 'asymptotic rate' $n^{-1/2}$. The idea is to postulate bounds on the infinite number of cumulants of log-likelihood differentials, in which powers of the quantity corresponding to $n^{-1/2}$ appear as factors. The existence of such bounds is the essential requirement for the model to be analytic, and the factor itself is the quantity defined as the 'index' of the model in Section 5. In several ways the index takes over the role of $n^{-1/2}$ for general sequences of models, and in Chapter 4 it is shown that several asymptotic results can be derived from the condition that the index tends to zero for a sequence of analytic models.

In Section 2 below we define the class of analytic models by the first approach, i.e., by the requirement that the log-likelihood function is analytic combined with a moment condition. The equivalence with the other definition, in terms of bounds on the cumulants of the log-likelihood differentials, is proved in Section 4. Some basic auxiliary results for analytic models are derived in Section 3. As mentioned above the index of an analytic model is defined in Section 5 where its behaviour in connection with independent observations is also investigated. Some possibilities of obtaining analytic models from other analytic models are explored in Section 6. These include analytic reparametrizations, and reductions by sufficiency and ancillarity.

While the material until Section 6 is being used in later chapters, the last two sections give some independent results for potential use in asymptotic theory, but not developed further here. Thus Sections 7 and 8 can be read independently of each other, or omitted at first reading. In Section 7 we justify the type of approximation of the model by a curved exponential family as discussed in the beginning of this section. The relation to the asymptotic sufficiency of a finite number of the log-likelihood differentials should be noticed here. In Section 8 we explore the alternative approach to the truncation of the Taylor series expansion of the exponent in (1.1), namely to keep the infinite series and work within the infinite-dimensional exponential family. This approach is not carried far here, but some elementary properties of the infinite-dimensional family are derived.

2 Definition of an analytic model

Let $\{ P_\beta : \beta \in B \subseteq V \}$ be a family of probability measures dominated by some measure ν on a measurable space E, where B is a subset of the finite dimensional real vector space V with $\dim V = p$ and equipped with an inner product norm denoted $\|\cdot\|$. We let $f(y;\beta)$ denote the density of P_β with respect to ν at $y \in E$. Since these densities are not generally uniquely determined by the measures and, furthermore, smoothness properties as well as some statistical concepts depend on the choice of densities, we shall think of these as part of the model. Hence we refer to the model as $\{ (E,\nu), f(\cdot;\beta) : \beta \in B \subseteq V \}$ or often as $\{ f(y;\beta) : \beta \in B \subseteq V \}$ or even $f(y;\beta)$ in somewhat more imprecise notations when the rest is understood. In examples, when the distributions of the model are stated, we shall usually assume that the obvious choice of densities has been made.

When $\log f(y;\beta)$ exists and is k times differentiable with respect to β at some point β_0 we let the k-linear symmetric form

$$D_k(\beta_0) = D^k_\beta \log f(y;\beta_0) : V^k \to \mathbf{R} \tag{2.1}$$

denote the kth differential of $\log f$ with respect to β, i.e.,

$$D_k(\beta)(v^k) = \frac{d^k}{dh^k} \log f(y;\beta + hv), \qquad h \in \mathbf{R}, \tag{2.2}$$

evaluated at $h = 0$, is the kth derivative in the direction v, where we have used the abbreviation $v^k = (v, \dots, v) \in V^k$.

Recall that two measures P and Q, say, are said to be *equivalent*, or *mutually absolutely continuous*, if for any measurable set A, $P(A) = 0$ if and only if $Q(A) = 0$.

Definition 2.1. *A statistical model $\{(E,\nu), f(\cdot,\beta) : \beta \in B \subseteq V\}$ is said to be analytic at a point $\beta_0 \in \mathrm{int}(B)$ if there exists a neighbourhood $U(\beta_0) \subseteq B$ and a measurable set $E_1 \subseteq E$ such that the following conditions hold:*

(i) The measures $\{P_\beta : \beta \in U(\beta_0)\}$ are mutually absolutely continuous.

(ii) The set E_1 satisfies $P_{\beta_0}(E_1) = 1$, and $f(y;\beta_0) > 0$ for all $y \in E_1$.

(iii) For all $y \in E_1$ the function $\beta \mapsto f(y;\beta)$ is measurable with respect to the Borel σ-algebra on V and analytic as a function of β in $U(\beta_0)$.

(iv) There exist a constant $\rho(\beta_0) \geq 0$ and a function $M(\cdot;\beta_0) : E \to \mathbf{R}$ such that $M(Y;\beta_0)$ has finite exponential moments with respect to P_{β_0} and

$$|D_k(\beta_0)(v^k)| \leq k!\, M(y;\beta_0)\rho(\beta_0)^{k-1}\|v\|^k \tag{2.3}$$

for all $v \in V, k \in \mathbf{N}$ and $y \in E_1$.

Definition 2.2. *A statistical model is said to be analytic in $B_0 \subseteq B$ if B_0 is a subset of $B \subseteq V$ and the model is analytic at every point $\beta_0 \in B_0$.*

REMARK 2.3. Some comments to Definition 2.1:

(1) It should be noted that although the name 'analytic' for brevity is used for these models one should not think of the definition as a requirement that the densities are analytic plus some regularity conditions. There is a strong moment condition (iv) on the derivatives of the log-likelihood function also, whereas the conditions (i) and (ii) are more like regularity conditions.

(2) Notice that the measurability condition on the density function and the requirement that it is positive, together imply the measurability of the derivatives of the log-density. In fact, the measurability is also implicit in the assumption of integrability, cf. the moment condition (iv). In the sequel we shall usually omit comments on measurability considerations.

(3) The concept of analyticity of the model at β_0 does not depend on the choice of norm on V. A change of norm is simply compensated by a corresponding change of $M(y;\beta_0)$ and $\rho(\beta_0)$.

(4) The definition depends on the model also through the choice of versions of densities. This may not be ideal, but, as indicated above, as long as likelihood inference and other inferential procedures are depending on this choice such a dependence seems unavoidable if statistical properties are to be shown to follow from the definition. Also, in this way, it is guaranteed that we are working with densities chosen in a sensible manner. When (i) is satisfied it is, of course, always possible to take $\nu = P_{\beta_0}$ and $f(y;\beta_0) = 1$ in which case (ii) is automatically satisfied.

(5) To clarify the meaning of the condition (iv) consider the log likelihood function, $g(h) = \log f(y;\beta_0 + hv), h \in \mathbf{R}$, in a particular direction $v \in V$ from β_0. As a function of $h \in \mathbf{R}$ it is analytic at zero, because of (iii), with derivatives

$$g^{(k)}(h) = D_k(\beta_0)(v^k), \qquad k \in \mathbf{N}$$

satisfying inequalities of the form $|g^{(k)}(h)| \leq k!MC^k$ where M depends on $y \in E_1$ and C on v, C being proportional to $\|v\|$. Since the radius of convergence is at least C^{-1} we see, e.g., by taking $\|v\| = 1$, that (iv) implies the radius of convergence at β_0 of $\log f(y;\beta)$ to be bounded below by some positive constant that does not depend on $y \in E_1$, but, possibly, on the model and on β_0. As a second implication of (iv), notice that the factor $M(Y;\beta_0)$, having exponential moments, ensures the same property for all derivatives of the log-likelihood derivatives, even in some uniform manner. In particular all moments of these derivatives exist for an analytic model.

3 Basic lemmas for analytic models

In this section we shall establish some properties of analytic models. Thus, it will frequently be assumed that we have to do with an analytic model, i.e., that the conditions of Definition 2.1 hold. To avoid repetitious statements of trivial kinds we shall then refer to the functions $\rho(\beta_0), M(y;\beta_0)$ and the set $U(\beta_0)$ as any given such for which the conditions hold, even though these objects are not uniquely determined by the model. For similar reasons we shall avoid the statement for all $y \in E_1$ by assuming without mentioning that the model is restricted to the set E_1, or equivalently simply by convention. Also, unless otherwise stated, the norm $\|\cdot\|$ refers to a given inner product norm on V, as in Definition 2.1, but often results are extended to cover the case when a semi-norm fulfils condition (iv) in Definition 2.1. We shall frequently use abbreviations like $M(y)$ and ρ for $M(y;\beta)$ and $\rho(\beta)$, respectively, evaluated at the obvious point, usually β_0, but for ease of reference such notations will be mentioned in connection with their application.

The results in this section are all of somewhat technical nature and mainly intended for use in later proofs, although they also illustrate the mathematical nature of the models considered.

Lemma 3.1. *Assume that the model $\{ f(y;\beta) : \beta \in B \subseteq V\}$ is analytic at β_0 and that $\|\cdot\|$ is a semi-norm on V satisfying the condition (iv) of Definition 2.1. Then there is some positive constant $a < \rho(\beta_0)^{-1}$ and an $s > 0$ such that $E_\beta \exp\{sM(Y;\beta_0)\}$ is bounded uniformly in β in the set*

$$U_a(\beta_0) = U(\beta_0) \cap \{\beta \in B : \|\beta - \beta_0\| < a\}. \tag{3.1}$$

Proof. We use the abbreviated notations $M(y) = M(y;\beta_0), \rho = \rho(\beta_0)$ and $D_k = D_k(\beta_0)$. Notice that $M(y)$ is non-negative. For $\beta \in U_a(\beta_0)$ and $s > 0$ we have

$$\begin{aligned} E_\beta \exp\{sM(Y)\} &= \int \exp\{sM(y)\}\,(f(y;\beta)/f(y;\beta_0))\,dP_{\beta_0}(y) \\ &= \int \exp\left\{sM(y) + \sum_{k=1}^{\infty} \frac{1}{k!} D_k(w^k)\right\} dP_{\beta_0}(y), \end{aligned}$$

where $w = \beta - \beta_0$, because $f(y;\beta)$ and hence $\log f(y;\beta)$ is analytic in β in $U_a(\beta_0)$ for sufficiently small a. To see this, notice that $f(y;\beta_0) > 0$ and that the radius of convergence for the Taylor series expansion of $\log f(y;\beta)$ is at least ρ^{-1} independently of y. Thus,

$$\begin{aligned} E_\beta \exp\{sM(Y)\} &\le \int \exp\{sM(y) + \sum_{k=1}^{\infty} M(y)\rho^{k-1}\|w\|^k\}\,dP_{\beta_0}(y) \\ &\le E_{\beta_0} \exp\{M(y)[s + a/(1 - a\rho)]\}, \end{aligned}$$

which is finite for sufficiently small a and s. ∎

In the sequel we shall be using the norm of a symmetric k-linear form B on V which should be recalled to be given by

$$\|B\| = \sup\{\,|B(v^k)| : \|v\| \le 1\,\}, \tag{3.2}$$

cf. (1.1.19), a definition which is extended to cover semi-norms also.

Lemma 3.2. *Assume that the conditions (i)–(iii) of Definition 2.1 hold. Then, for any $K \in \mathbb{N}$, condition (iv) is equivalent to the condition that both of (1) and (2) below hold:*

(1) $(D_1(\beta_0), \ldots, D_K(\beta_0))$ has exponential moments with respect to P_{β_0}.

(2) The condition (iv) in Definition 2.1 holds for $k \ge K + 1$.

Proof. As above, the abbreviations $M(y)$, ρ and D_k implies evaluation at $\beta = \beta_0$. Assume first that (iv) in Definition 2.1 holds. Then (2) above is trivial and for any $k \in \mathbb{N}$, (iv) immediately shows that $D_k(v^k)$ has exponential moments for any fixed $v \in V$. Hence the vector $(D_1, \ldots, D_K)$, being of finite dimension, also has exponential moments.

Next, assume that (1) and (2) above hold for some fixed K, with a function M that we denote M_K, and ρ_K in place of ρ. Let

$$M(y) = \sup\left\{ M_K(y), \frac{1}{k!}|D_k(v^k)|/\left(\rho^{k-1}\|v\|^k\right) : k = 1,\dots,K, v \in V\setminus\{0\}\right\}$$
$$\le \max\left\{ M_K(y), \frac{1}{k!}\|D_k\|/\rho^{k-1} : k = 1,\dots,K\right\},$$

where $\rho = \rho_K$, if this is different from zero, otherwise we may take $\rho = 1$, say. Now, by construction, the inequality in (iv) in Definition 2.1 is fulfilled for all $k \in \mathbf{N}$ and all $v \in V$. Furthermore $M(y)$ has finite exponential moments because it is bounded by a maximum of a finite number of variables with finite exponential moments. ∎

Lemma 3.3. *Assume that the model* $\{ f(y;\beta) : \beta \in B \subseteq V\}$ *is analytic at* β_0 *and that* $\|\cdot\|$ *is a semi-norm satisfying condition (iv) in Definition 2.1. Let* $(A_k : k \in \mathbf{N})$ *be a sequence of linear mappings from* $\mathrm{Sym}_k(V;\mathbf{R})$ *to* $\mathbf{R}$ *or* $\mathbf{C}$, *i.e., mappings from the spaces of* D_k*'s. Assume that for some* $c \ge 0, a < \rho(\beta_0)^{-1}$ *and* $m \ge 0$ *the* A_k*'s are bounded by*

$$\|A_k\| \le \frac{c}{k!}a^k k^m, \tag{3.3}$$

where the norm is induced by the norm in (3.2) by the definition

$$\|A_k\| = \sup\{\,|A_k(B)| : B \in \mathrm{Sym}_k(V;\mathbf{R}), \|B\| \le 1\,\}, \tag{3.4}$$

cf. (1.1.21). Then the infinite sum

$$S = \sum_{k=1}^{\infty} A_k\left(D_k(\beta_0)\right) \tag{3.5}$$

is absolutely convergent almost surely (ν) *and, given* a *and* m, *for sufficiently small* c *it has uniformly bounded exponential moments with respect to* P_β *for* β *in some neighbourhood of* β_0. *The same conclusion holds for sufficiently small* a *when* m *and* c *are given.*

Proof. From (iv) in Definition 2.1 it follows that $\|D_k(\beta_0)\|$ is finite for all $k \in \mathbf{N}$ and hence that

$$|S| \le \sum_{k=1}^{\infty} c a^k k^m M(y)\rho^{k-1}$$
$$= caM(y)\sum_{k=1}^{\infty}(a\rho)^{k-1}k^m,$$

where the argument β_0 has been omitted from M and ρ. The sum in the last expression is convergent and since the sum is also non-decreasing in a the assertions concerning exponential moments follow from Lemma 3.1. ∎

The inclusion of the factor k^m in (3.3) facilitates application, but it is easy to see that the result is equally strong without this factor.

The result in Lemma 3.3 is used first of all to demonstrate the existence of the infinite-dimensional exponential family investigated in Section 8.

The main point of the following lemma is that, in analogy with analytic functions, a model that is analytic at some point is also analytic in a neighbourhood of this point.

Lemma 3.4. *Assume that the model* $\{ f(y;\beta) : \beta \in B \subseteq V \}$ *is analytic at* β_0 *and that* $\|\cdot\|$ *is a semi-norm satisfying condition (iv) in Definition 2.1. Let* $U_a(\beta_0)$ *denote the set defined in (3.1). Then for any* $a < \rho(\beta_0)^{-1}$ *the following two assertions hold:*

(1) $f(y;\beta) > 0$ *for all* $\beta \in U_a(\beta_0)$ *and* $y \in E_1$.

(2) There exists a constant $\tilde{\rho} > 0$ *and a function* $\tilde{M} : E_1 \to \mathbf{R}$ *that has exponential moments with respect to* P_{β_0}*, such that*

$$|D_k(\beta)(v^k)| \leq k!\, \tilde{M}(y)\tilde{\rho}^{k-1}\|v\|^k \tag{3.6}$$

for all $\beta \in U_a(\beta_0), v \in V$ *and* $y \in E_1$. *For some sufficiently small* $a > 0$ *the following two assertions hold:*

(3) The random variable $\tilde{M}(Y)$ *in (3.6) has uniformly bounded exponential moments with respect to all* $P_\beta, \beta \in U_a(\beta_0)$.

(4) The model is analytic in $U_a(\beta_0)$ *and the set* E_1*, the (semi)-norm used in (iv), the constant* ρ*, and the function* $M(y)$ *in Definition 2.1 may be chosen independently of* $\beta \in U_a(\beta_0)$.

Proof. From the conditions in Definition 2.1 we know that $f(y;\beta_0) > 0$ and that $f(y;\beta)$ is analytic for any $y \in E_1$. Hence, for any fixed $y \in E_1$, $\log f(y;\beta)$ is analytic in the subset of $U(\beta_0)$ on which $f(y;\beta) > 0$. Condition (iv) in Definition 2.1 implies that within any of the sets $U_a(\beta_0)$ with $a\rho(\beta_0) < 1$, the function $\log f(y;\beta)$ has an absolutely convergent power series expansion around β_0, implying that $f(y;\beta)$ is positive on this set. To prove (2), consider a fixed $\beta \in U_a(\beta_0)$. Let $w = \beta - \beta_0$ and use the usual convention that an omitted argument implies evaluation at β_0. Then

$$\begin{aligned} |D_k(\beta)(v^k)| &= |\sum_{j=k}^{\infty} \frac{1}{(j-k)!} D_j(v^k, w^{j-k})| \\ &\leq \sum_{j=k}^{\infty} \frac{j!}{(j-k)!} M(y)\rho^{j-1}\|v\|^k\|w\|^{j-k} \\ &\leq k!\, M(y)\rho^{k-1}\|v\|^k \sum_{j=k}^{\infty} \binom{j}{k} (a\rho)^{j-k} \end{aligned}$$

$$= k!\, M(y)\rho^{k-1}\|v\|^k(1 - a\rho)^{-k-1}$$

which proves (3.6) if we take $\tilde{\rho} = \rho/(1 - a\rho)$ and $\tilde{M}(y) = M(y)/(1 - a\rho)^2$.

Since the bound (3.6) holds for β_0, in particular, it follows directly from Lemma 3.1 that $\tilde{M}(Y)$ has uniformly bounded exponential moments with respect to all β in $U_a(\beta_0)$ for sufficiently small $a > 0$. Thus, (3) and hence (4) is proved. ∎

Lemma 3.5. *Assume that the model* $\{ f(y;\beta) : \beta \in B \subseteq V \}$ *is analytic at* β_0. *Then any integral of the form*

$$\int \left(D_{k_1}(\beta)(v_1^{k_1})\right) \cdots \left(D_{k_m}(\beta)(v_m^{k_m})\right) f(y;\beta)\, d\nu(y) \tag{3.7}$$

is analytic at β_0 *and its derivatives at* β_0 *can be found by successive differentiation under the integral sign.*

Proof. Let $G(y;\beta)$ denote the integrand and $D^kG(y)$ its kth differential with respect to β at β_0. As usual we frequently omit the argument β_0 from any function. We know that for sufficiently small $a > 0$ the integrand is analytic in the set $U_a(\beta)$ defined in (3.1) for any fixed y, and therefore it equals the sum of its Taylor series expansion at β_0. For each fixed y the differentials satisfy inequalities of the form

$$|D^kG(y)\left(w^k\right)| \le c(y)k!\, a(y)^k\|w\|^k, \qquad k \in \mathbb{N} \tag{3.8}$$

for some $a(y)$ and $c(y)$, and any $w \in V$. We shall prove that the integrals of the differentials satisfy similar inequalities except that a and c, may not, of course, depend on y. Then it will follow from the theorem of majorized convergence that integration and summation can be interchanged and hence that the integral has an absolutely convergent power series expansion around β_0 with terms bounded by a geometric series. Hence the integral is analytic and the sum of the integrals of the individual terms is its Taylor series expansion around β_0, proving that the derivatives of the integral are identical to the integrals of the derivatives.

Each of the components $D_{k_j}(\beta)(v_j^{k_j})$ in $G(y;\beta)$ has a Taylor series expansion with terms satisfying (3.8) with $c(y) = M(y)$, and $a(y) = \rho$ which is independent of y. Hence, the product of these m components has a Taylor series expansion in which the kth term is bounded by

$$cM(y)^m a^k\|\beta - \beta_0\|^k, \tag{3.9}$$

for some constants c and a that do not depend on y but, possibly, on $k_1, \ldots, k_m$ and on $v_1, \ldots, v_m$. In fact, any $a > \rho$ will do here with a proper choice of c. An expansion of $f(y;\beta)/f(y;\beta_0)$ as in the proof of Lemma 3.1 yields

$$f(y;\beta)/f(y;\beta_0) \le \exp\left\{\sum_{k=1}^{\infty} M(y)\rho^{k-1}\|\beta - \beta_0\|^k\right\}, \tag{3.10}$$

where the individual terms in the sum are bounds for the corresponding terms in the Taylor series expansion of the exponent. Therefore, if we multiply together the bounds from (3.9) and (3.10), and expand in powers of $\|\beta-\beta_0\|$, then the individual terms of the expansion will also be upper bounds for the corresponding terms of the expansion of the entire integrand. By use of (1.2.22) and (1.2.27) we get the following expression for the Kth term of the product of the bounds in (3.9) and (3.10),

$$\|\beta-\beta_0\|^K \left\{ cM(y)^m a^K + \sum_{k=1}^{K} cM(y)^m a^{K-k} \frac{1}{k!} \sum_{\alpha} k! \prod_{j=1}^{k} \left(\{\rho^{j-1} M(y)\}^{\alpha_j} / \alpha_j! \right) \right\}$$

$$= \|\beta-\beta_0\|^K \left\{ cM(y)^m a^K + \sum_{k=1}^{K} cM(y)^m a^{K-k} \sum_{\alpha} M(y)^{\Sigma\alpha_j} \rho^{k-\Sigma\alpha_j} / \prod \alpha_j! \right\}$$

$$= \|\beta-\beta_0\|^K \left\{ cM(y)^m a^K + c\sum_{k=1}^{K} a^{K-k} \sum_{n=1}^{k} \frac{1}{n!} M(y)^{n+m} \rho^{k-n} \binom{k-1}{n-1} \right\}, \quad (3.11)$$

where $\sum_\alpha$ denotes the sum over $T(k)$, i.e., over all sequences $(\alpha_1, \ldots, \alpha_k) \in \mathbb{N}_0^k$ with $\sum j\alpha_j = k$. If $s > 0$ is chosen such that $\mathrm{E}\exp\{sM(y)\} = C(s) < \infty$, then

$$\mathrm{E}\{M(y)^n\} \leq n!\, C(s) s^{-n}$$

for all $n \in \mathbb{N}$ and therefore the integral of (3.11) is bounded by

$$\begin{aligned} &\|\beta-\beta_0\|^K \Bigg\{ cm!\, C(s) s^{-m} a^K \\ &\quad + c\sum_{k=1}^{K} a^{K-k} \sum_{n=1}^{k} C(s) s^{-(n+m)} \frac{(n+m)!}{n!} \rho^{k-n} \binom{k-1}{n-1} \Bigg\} \\ &\leq \|\beta-\beta_0\|^K \Bigg\{ cm!\, C(s) s^{-m} a^K \\ &\quad + c\sum_{k=1}^{K} a^{K-k} C(s) s^{-m} (k+m)^m \sum_{n=1}^{k} s^{-n} \rho^{k-n} \binom{k}{n} \Bigg\} \\ &< \|\beta-\beta_0\|^K \left\{ cm!\, C(s) s^{-m} a^K + c\sum_{k=1}^{K} a^{K-k} C(s) s^{-m} (k+m)^m \left(\frac{1}{s} + \rho \right)^k \right\} \end{aligned}$$

which by simple considerations shows that the integral of the Taylor series expansion is bounded by a geometric series. Hence the lemma follows from the argument given above.

Note: It would not have been difficult to show directly that the differential of G in a neighbourhood of β_0 is bounded by an integrable function, and therefore integration and differentiation can be interchanged. Also, it is easy to see that the derivative will be a linear combination of terms of the same form, showing that repeated differentiation of the integral is permissible by differentiation of the integrand. However, to show that the function is analytic an argument like the one above is required, or a direct evaluation of the integral of the derivatives, which is more difficult. ∎

Lemma 3.5 implies that the equation

$$\int f(y;\beta)\,d\nu(y) = \int \exp\{\log f(y;\beta)\}\,d\nu(y) = 1 \tag{3.12}$$

may be differentiated repeatedly under the integral sign. Thus, the lemma implies the validity of the usual relations between moments of the log-likelihood derivatives. These moments are defined as the multilinear forms

$$\begin{gathered}\mu_{k_1\cdots k_m}(\beta) \in \operatorname{Lin}(V^{k_1+\cdots+k_m};\mathbf{R})\\ \{\mu_{k_1\cdots k_m}(\beta)\}\left(v_1^{k_1},\ldots,v_m^{k_m}\right) = E_\beta\left\{\left(D_{k_1}(\beta)(v_1^{k_1})\right)\cdots\left(D_{k_m}(\beta)(v_m^{k_m})\right)\right\}\end{gathered} \tag{3.13}$$

for $v_1,\ldots,v_m \in V$. Then the first three of the relations between these, evaluated at any β where the model is analytic, are

$$\begin{aligned}&\mu_1 = 0,\\ &\mu_2 + \mu_{11} = 0,\\ &\mu_3 + 3\,\mathrm{sym}\{\mu_{12}\} + \mu_{111} = 0 \in \operatorname{Lin}_3(V;\mathbf{R})\end{aligned} \tag{3.14}$$

where $\mathrm{sym}\{\cdots\}$ denotes the symmetrized version of the multilinear function in question, obtained by averaging over all permutations of the arguments.

It follows from Skovgaard (1986a), see also Lemma 6.6, that the same results hold if moments are replaced by cumulants throughout. These cumulants of the log-likelihood derivatives are defined as the multilinear forms

$$\begin{gathered}\chi_{k_1\cdots k_m}(\beta) = \mathrm{cum}_\beta\left(D_{k_1}(\beta),\ldots,D_{k_m}(\beta)\right) : V^{k_1}\times\cdots\times V^{k_m} \to \mathbf{R},\\ \{\chi_{k_1\cdots k_m}(\beta)\}\left(v_1^{k_1},\ldots,v_m^{k_m}\right) = \mathrm{cum}_\beta\left\{D_{k_1}(\beta)(v_1^{k_1}),\ldots,D_{k_m}(\beta)(v_m^{k_m})\right\}\end{gathered} \tag{3.15}$$

where it should be noticed that the variables $D_{k_j}(\beta)(v_j^{k_j})$ are one-dimensional, and that the multilinear form $\chi_{k_1\cdots k_m}(\beta)$ is symmetric under permutations within each of the components V^{k_j}, but not in general for other permutations. The same is true for the moments defined in (3.13). The conclusion of the previous lemma is that the relations, of which the first three are given in (3.14), hold for the moments and for the cumulants of the log-likelihood derivatives.

In particular, the *Fisher information*, $I(\beta)$, which by definition equals the symmetric positive semi-definite bilinear form $\chi_{11}(\beta)$ satisfies the equality

$$I(\beta) = \chi_{11}(\beta) = -\chi_2(\beta) \tag{3.16}$$

if the family is analytic at β.

There are a few conditions to check before the result in Skovgaard (1986a) can be applied, but these follow from Lemma 3.5. It is a trivial consequence of Lemma 3.5 that all the cumulants in (3.15) are analytic at β_0, because they are polynomial functions of the moments in (3.13).

Corollary 3.6. *Assume that the model* $\{ f(y;\beta) : \beta \in B \subseteq V \}$ *is analytic at* β_0 *and that* $\rho(\beta_0) = 0$ *in Definition 2.1. Then* P_β *is constant as a function of* β *in a neighbourhood of* β_0.

Proof. From (iv) in Definition 2.1 it is seen that $D_2(\beta_0) = 0$ for all $y \in E_1$. By Lemma 3.5, this implies that $I(\beta_0) = -\chi_2(\beta_0) = 0$. Thus $D_1(\beta_0)$ is seen from (3.14) and (3.16) to be degenerate at zero, just as $D_k(\beta_0)$ for any $k \geq 2$. Therefore the Taylor series expansion of $\log f(y;\beta)$ around β_0 is constantly zero. Since $\log f(y;\beta)$ is analytic in some neighbourhood of β_0 independently of $y \in E_1$, cf. (4) of Lemma 3.4, it follows that $f(x;\beta)$ and hence P_β is constant in this neighbourhood. ∎

The following lemma gives one of the few available global properties of analytic models and shows that the class of analytic models does not comprise models for which the support of the distribution depends on the parameter.

Lemma 3.7. *If a model parametrized by* $\beta \in B \subseteq V$ *is analytic in an open, connected set* $B_0 \in B$*, then any two measures* P_{β_1} *and* P_{β_2}*,* $\beta_1, \beta_2 \in B_0$*, are mutually absolutely continuous.*

Proof. Let $\beta_1 \in B_0$ be an arbitrary fixed point and consider the subset, S say, of B_0 on which the measures are dominated by P_{β_1}. By the assumption that the model is analytic in B_0 it follows that each measure in B_0 dominates all measures in a neighbourhood. Since dominance is a transitive relation it follows that the set S is open. Now, consider a sequence of β's in S, that converges to a point β_0 in B_0. Because the model is analytic at β_0, the measure at this point is dominated by all measures in a neighbourhood and hence eventually by one of the measures in the sequence considered. Thus, S is also closed in B_0 and since it is not empty it equals the entire set B_0. ∎

4 Equivalent definitions

The main result in this section is Theorem 4.2 below that gives three conditions that each may replace condition (iv) in Definition 2.1. The implication of the theorem is that the bound in (iv) may be proved on the basis of certain inequalities for the cumulants of the log-likelihood derivatives rather than from bounds on the random variables themselves. The key to this result is the lemma given below which, in turn, relies heavily on the result in Lemma 1.5.3.

Lemma 4.1. *Assume that the model $\{f(y;\beta) : \beta \in B \subseteq V\}$ satisfies the condition (ii) of Definition 2.1 and that $f(y;\beta)$ is infinitely often differentiable at β_0 for all $y \in E_1$. Assume also that two constants $c \geq 0$ and $\tau \geq 0$ exist such that the cumulants of the log-likelihood derivatives at β_0, defined in (3.15), satisfy the condition*

$$|\chi_{\underbrace{k\cdots k}_{m}}(\beta_0)\,(v^{km})\,| \leq c^2(m-1)!\,k!^m\tau^{km-2}\|v\|^{km} \tag{4.1}$$

for all $v \in V$, $k \in \mathbf{N}$, and $m \geq 2$, where $\|\cdot\|$ is some inner product semi-norm on V. Then there exists a function $H(\cdot;\beta_0)$ of E into $\mathbf{R}$ such that on a set of β_0-probability 1,

$$|\{D_k(\beta_0) - \chi_k(\beta_0)\}(v^k)| \leq ck!\,H(y;\beta_0)\rho(\beta_0)^{k-1}\|v\|^k \tag{4.2}$$

for all $v \in V$, where $\rho = (2e\sqrt{p})\tau$, $p = \dim V$, and moreover

$$\mathrm{E}_{\beta_0}(\exp\{\delta H(Y;\beta_0)\}) < \gamma(p)\exp\left\{2p(e\delta)^2/(1-\delta\rho/c)\right\} \tag{4.3}$$

for all $\delta < c/\rho$, where $\gamma(p)$ is some constant depending on the model only through the dimension p.

Proof. As in previous proofs we omit the argument β_0 from the various functions. Notice first that if $\tau = 0$ then all D_k are degenerate, except possibly D_1 which has a normal distribution because all cumulants after the first two vanish, and the result is easy. Thus, we assume that $\tau > 0$. We also exclude the trivial case $c = 0$ from the proof. Let $\rho = (2e\sqrt{p})\tau$ and define for each fixed $v \in V$, $k \in \mathbf{N}$ and $h \geq 0$ the event

$$A_k(h;v) = \left\{\, y \in E : \left|(D_k - \chi_k)(v^k)\right| \leq ck!\,h\rho^{k-1}\|v\|^k \,\right\} \tag{4.4}$$

and let

$$\begin{aligned} A_k(h) &= \cap_{v\in V} A_k(h;v), \\ A(h) &= \cap_{k=1}^{\infty} A_k(h), \\ H(y) &= \inf\{h : y \in A(h)\} \in [0,\infty]. \end{aligned} \tag{4.5}$$

The mth cumulant of $D_k - \chi_k$ satisfies (4.1) for all $m \in \mathbf{N}$ while the first cumulant is zero. Hence, the Taylor series expansion of the cumulant generating function in terms of the cumulants is bounded by a geometric series in powers of $k!\,\tau^k\|v\|^k$, and consequently the cumulants identify the distribution and for any positive $t < \left(k!\,\tau^k\|v\|^k\right)^{-1}$ we have

$$\begin{aligned}
&\mathrm{E}\exp\left\{t|\{D_k-\chi_k\}(v^k)|\right\}\\
&\quad< \mathrm{E}\exp\left\{t\{D_k-\chi_k\}(v^k)\right\}+\mathrm{E}\exp\left\{-t\{D_k-\chi_k\}(v^k)\right\}\\
&\quad= \exp\sum_{m=2}^{\infty}\frac{c^2}{m!}t^m(m-1)!\,k!^m\tau^{km-2}\|v\|^{km}\\
&\qquad\quad+\exp\sum_{m=2}^{\infty}\frac{c^2}{m!}(-t)^m(m-1)!\,k!^m\tau^{km-2}\|v\|^{km}\\
&\quad\le 2\exp\sum_{m=2}^{\infty}\frac{c^2}{m!}t^m(m-1)!\,k!^m\tau^{km-2}\|v\|^{km}\\
&\quad< 2\exp\left\{\frac{c^2}{2}t^2k!^2\tau^{2k-2}\|v\|^{2k}/\left(1-tk!\,\tau^k\|v\|^k\right)\right\}.
\end{aligned}$$

The event $A_k(h)$ involves the simultaneous occurrence of the events $A_k(h;v)$ for all v. To obtain a bound for the probability of this event we use the result in Lemma 1.5.3. Assume for the moment that $\|\cdot\|$ is a norm, i.e., the inner product is positive definite. Let $(v_1,\ldots,v_p)$ be a basis on V that is orthonormal with respect to this inner product. Consider the finite set of vectors

$$V_k=\left\{\sum_{j=1}^{k}\alpha_jv_{i_j}:\alpha_j\in\{-1,1\};i_j\in\{1,\ldots,p\}\right\}.$$

For the number of elements $|V_k|$ in V_k we have the estimate (1.5.8)

$$|V_k|<\frac{1}{p!}2^p(k+p)^p$$

and the result in Lemma 1.5.3 tells us that for all $v\in V$ with $\|v\|\le 1$,

$$|(D_k-\chi_k)(v^k)|<(e\sqrt{p})^k\sup\left\{\,|(D_k-\chi_k)\left(w^k\right)|:w\in V_k\,\right\}.$$

Therefore, for fixed $v\in V$ with $\|v\|>0$, consider the estimate obtained by use of Chebychev's inequality in the form of (1.4.36) with $t=\delta\rho\left(ck!\,\tau^k\|v\|^k\right)^{-1}$, where $\delta<c/\rho$,

$$\begin{aligned}
&1-\mathrm{P}\left\{A_k\left(h/(e\sqrt{p})^k;v\right)\right\}\\
&\quad\le \exp\left\{-tck!\,h\rho^{k-1}\|v\|^k/(e\sqrt{p})^k\right\}\\
&\qquad\quad\times 2\exp\left\{\frac{c^2}{2}t^2k!^2\tau^{2k-2}\|v\|^{2k}/\left(1-tk!\,\tau^k\|v\|^k\right)\right\}
\end{aligned}$$

$$\leq 2\exp\left\{-\delta 2^k h + \frac{1}{2}\delta^2(2e\sqrt{p})^2/(1-\delta\rho/c)\right\}. \tag{4.6}$$

Thus, it follows that

$$\begin{aligned}
\mathrm{P}\{H(y) > h\} &\leq 1 - \mathrm{P}(A(h)) \\
&\leq \sum_{k=1}^{\infty}\sum_{v\in V_k}\left\{1 - \mathrm{P}\left[A_k\left(h/(e\sqrt{p})^k; v\right)\right]\right\} \\
&\leq \sum_{k=1}^{\infty}|V_k| 2\exp\left\{-\delta 2^k h + 2p(\delta e)^2/(1-\delta\rho/c)\right\} \\
&\leq \frac{1}{p!}2^{p+1}\exp\left\{2p(\delta e)^2/(1-\delta\rho/c)\right\}\sum_{k=1}^{\infty}(k+p)^p\exp\left\{-\delta 2^k h\right\}.
\end{aligned} \tag{4.7}$$

As the final step we obtain

$$\begin{aligned}
\mathrm{E}(\exp\{\delta H(Y)\}) &= \int_0^{\infty}\mathrm{P}\{H(Y) > \delta^{-1}\log z\}\,dz \\
&< \frac{1}{p!}2^{p+1}\exp\left\{2p(\delta e)^2/(1-\delta\rho/c)\right\}\left\{1 + \int_1^{\infty}\sum_{k=1}^{\infty}(k+p)^p z^{-2^k}\,dz\right\} \\
&= \frac{1}{p!}2^{p+1}\exp\left\{2p(\delta e)^2/(1-\delta\rho/c)\right\}\left\{1 + \sum_{k=1}^{\infty}(k+p)^p/\left(2^k-1\right)\right\},
\end{aligned}$$

from which the inequality (4.3) follows because the sum is convergent for any p. Notice that it is permissible to interchange integration and summation, leading to the last equality above, because the functions involved are positive. Evidently, the inequality implies that the function $H(Y)$ is finite with probability one.

Now consider the modifications required for the general case of a semi-norm $\|\cdot\|$ generated by a pseudo inner product $\langle\cdot,\cdot\rangle$. Let $(v_1,\ldots,v_p)$ be a basis on V which is orthogonal with respect to this pseudo inner product and has either $\|v_j\| = 1$ or $\|v_j\| = 0$ for all j. For each $n \in \mathbb{N}$ consider the inner product that is obtained from $\langle\cdot,\cdot\rangle$ by defining $(v_1,\ldots,v_p)$ to be orthonormal, except that any base vector v_j with $\|v_j\| = 0$ is replaced by nv_j, which consequently is defined to have unit length with respect to this new n-norm, denoted $\|\cdot\|_n$. Notice that, as n tends to infinity, the n-norm decreases monotonically towards the original semi-norm. Now define the events in (4.4) and (4.5) as before, except with the norm $\|\cdot\|_n$ in (4.4). Denote the events defined in this way by $A_{k,n}(h;v)$, $A_{k,n}(h)$ and $A_n(h)$, respectively. The event $A_{k,n}\left(h/(e\sqrt{p})^k;v\right)$ contains $A_k\left(h/(e\sqrt{p})^k;v\right)$ for which we still have the last estimate in (4.6), which trivially is valid for any $\delta > 0$ if $\|v\| = 0$. Therefore the estimate in (4.6) applies to $A_{k,n}\left(h/(e\sqrt{p})^k;v\right)$ also, and Lemma 1.5.3 provides the same bound for $1 - \mathrm{P}\{A_{k,n}(h)\}$ as is obtained by summing only over $v \in V_k$ in (4.7). For $n \in \mathbb{N}$, the events $A_{k,n}(h)$ form a decreasing sequence of events, the probability of which are bounded below by the same number. Hence, this bound applies also to the intersection which is the event $A_k(h)$. Thus, the bound in (4.7) still applies and the argument is completed as before. ∎

We are now in position to prove the main theorem in relation to the definition of analytic families. It implies that instead of obtaining uniform bounds for the random log-likelihood derivatives as in condition (iv) in Definition 2.1, it is sufficient to obtain uniform bounds for their cumulants. This may not always be easier in concrete examples, but for theoretical considerations in relation, e.g., to convolutions of experiments, it turns out to be useful.

Theorem 4.2. *Assume that the conditions (ii) and (iii) of Definition 2.1 hold for the model* $\{ f(y;\beta) : \beta \in B \subseteq V \}$ *, and let* $\langle \cdot, \cdot \rangle$ *be a given pseudo inner product on* V *with corresponding semi-norm* $\|\cdot\|$*. Then (iv) in Definition 2.1 is equivalent to each of the conditions (v), (vi) and (vii) below. We include the condition (iv) in the list to facilitate comparisons.*

(iv) *There exist a constant* $\rho(\beta_0) \geq 0$ *and a function* $M(\cdot;\beta_0) : E \to \mathbf{R}$ *such that* $M(Y,\beta_0)$ *has finite exponential moments with respect to* P_{β_0} *and*

$$|D_k(\beta_0)(v^k)| \leq k!\, M(y;\beta_0)\rho(\beta_0)^{k-1}\|v\|^k \tag{4.8}$$

for all $v \in V, k \in \mathbf{N}$ *and* $y \in E_1$*.*

(v) *'Mixed cumulant condition': There exist two constants* $\lambda(\beta_0) \geq 0$ *and* $c(\beta_0) \geq 0$ *such that*

$$\begin{aligned} &\left|\chi_{k_1\cdots k_m}(\beta_0)\left(v_1^{k_1},\ldots,v_m^{k_m}\right)\right| \\ &\qquad \leq c(\beta_0)^2(m-1)!\left(\prod_{j=1}^{m} k_j!\,\|v_j\|^{k_j}\right)\lambda(\beta_0)^{k_1+\cdots+k_m-2} \end{aligned} \tag{4.9}$$

for all $m \geq 2, k_j \in \mathbf{N}, v_j \in V; j = 1,\ldots,m$*.*

(vi) *'Directional mixed cumulant condition': There exist two constants* $\lambda(\beta_0) \geq 0$ *and* $c(\beta_0) \geq 0$ *such that*

$$\begin{aligned} &\left|\chi_{k_1\cdots k_m}(\beta_0)\left(v^{k_1+\cdots+k_m}\right)\right| \\ &\qquad \leq c(\beta_0)^2(m-1)!\left(\prod_{j=1}^{m} k_j!\right)\|v\|^{k_1+\cdots+k_m}\lambda(\beta_0)^{k_1+\cdots+k_m-2} \end{aligned} \tag{4.10}$$

for all $v \in V, m \geq 2, k_j \in \mathbf{N}; j = 1,\ldots,m$*.*

(vii) *'Pure cumulant condition': There exist two constants* $\tau(\beta_0) \geq 0$ *and* $c(\beta_0) \geq 0$ *such that*

$$|\chi_{\underbrace{k\cdots k}_{m}}(\beta_0)\left(v^{km}\right)| \leq c(\beta_0)^2(m-1)!\,k!^m\|v\|^{km}\tau(\beta_0)^{km-2} \tag{4.11}$$

for all $v \in V, m \in \mathbf{N}, k \in \mathbf{N}, km \geq 2$*.*

Moreover, if (iv) holds and

$$C(s) = \mathrm{E}\exp\{sM(Y;\beta_0)\} < \infty$$

then (v) and (vi) hold with $c(\beta_0) = 2C(s)/s$ *and* $\lambda(\beta_0) = \max\{\rho(\beta_0), 2C(s)/s\}$, *and if (vi) holds then (vii) holds with the same value of* $c(\beta_0)$ *as in (vi) and* $\tau(\beta_0) = 2\lambda(\beta_0)$.

Proof. Throughout the proof the value β_0 is fixed and if the argument or subscript β is omitted the value β_0 is understood. We shall prove the theorem by showing that (iv) implies (v), (vi) implies (vii), and (vii) implies (iv). That (v) implies (vi) is trivial.

Assume first that (iv) holds and let $v_1, \ldots, v_m$ be arbitrary fixed vectors in V. Notice that if $s > 0$ is chosen such that $\mathrm{E}\exp\{sM(Y)\} = C(s) < \infty$, then

$$\begin{aligned} \mathrm{E}\left\{M(Y)^m\right\} &\le m!\, s^{-m} \sum_{j=0}^{\infty} \frac{1}{j!} s^j \mathrm{E}\left\{M(Y)^j\right\} \\ &= m!\, s^{-m} C(s) \end{aligned} \tag{4.12}$$

for any $m \in \mathbb{N}$. Now, consider the mixed cumulant of the m real random variables

$$X_j = D_{k_j}\left(v_j^{k_j}\right)$$

for $k_j \in \mathbb{N}, v_j \in V; j = 1, \ldots, m \ge 2$. In terms of the moments of these random variables, the mixed cumulant may be written

$$\mathrm{cum}(X_1, \ldots, X_m) = \sum_S (-1)^A (A-1)!\, \mu(S_1) \cdots \mu(S_A), \tag{4.13}$$

where the sum is over all partitions $S = (S_1, \ldots, S_A), A = 1, \ldots, m$, of $\{1, \ldots, m\}$ into A non-empty subsets, and

$$\mu(S_a) = \mathrm{E}\left\{\prod_{j \in S_a} X_j\right\} \tag{4.14}$$

for $a = 1, \ldots, A$, cf. (1.4.16). From (iv) and (4.12) we obtain the estimate

$$\begin{aligned} \left|\prod_{a=1}^{A} \mu(S_a)\right| &= \prod_{a=1}^{A} \left|\mathrm{E}\left\{\prod_{j \in S_a} D_{k_j}\left(v_j^{k_j}\right)\right\}\right| \\ &\le \rho^{\Sigma(k_j - 1)} \left(\prod_{j=1}^{m} \left\{k_j!\, \|v_j\|^{k_j}\right\}\right) \prod_{a=1}^{A} \left\{|S_a|!\, s^{-|S_a|} C(s)\right\} \end{aligned}$$

$$= \rho^{\Sigma(k_j-1)} \left(\prod_{j=1}^{m} \{k_j! \, \|v_j\|^{k_j}\} \right) s^{-m} C(s)^A \prod_{a=1}^{A} |S_a|! \,, \qquad (4.15)$$

where $|S_a|$ is the number of elements in S_a. On combination of (4.13) and (4.15) we obtain

$$\operatorname{cum}(X_1, \ldots, X_m) \leq \rho^{(\Sigma k_j)-m} \left(\prod_{j=1}^{m} \{k_j! \, \|v_j\|^{k_j}\} \right) s^{-m} \sum_{S} C(s)^A (A-1)! \prod_{a=1}^{A} |S_a|! . \qquad (4.16)$$

It is fairly straightforward to see that

$$\sum_{S} (A-1)! \prod_{a=1}^{A} |S_a|! < (m-1)! \, 2^m \qquad (4.17)$$

and since $C(s) > 1$, the right side in (4.16) is bounded by

$$\rho^{(\Sigma k_j)-m} \left(\prod_{j=1}^{m} \{k_j! \, \|v_j\|^{k_j}\} \right) (C(s)/s)^m (m-1)! \, 2^m$$

which shows that the condition (v), and hence also (vi), holds with $c = 2C(s)/s$ and $\lambda = \max\{\rho, 2C(s)/s\}$.

Assume now that (vi) holds. We need to show (vii) for $m = 1$, $k \in \mathbb{N}$, only. Suppose for the moment that we can rely on the relation

$$\int D^k f(y; \beta) \, d\nu(y) = 0 \qquad (4.18)$$

at $\beta = \beta_0$ for any $k \in \mathbb{N}$. Then $\chi_1(\beta_0) = 0$, and for $k \geq 2$ it follows from (1.2.26) that for any $v \in V$,

$$0 = \int \left(D^k \exp\{\log f(y; \beta)\} \right) (v^k) \, d\nu(y)$$
$$= \int \sum_{m=1}^{k} \sum_{\alpha \in S_m(k)} \frac{k!}{m!} \left\{ \prod_{j=1}^{m} \alpha_j! \right\}^{-1} \left\{ \prod_{j=1}^{m} D_{\alpha_j} (v^{\alpha_j}) \right\} f(y; \beta_0) \, d\nu(y)$$

where the derivative is evaluated at β_0 and $S_m(k)$ is the set of sequences defined in (1.2.24). Because the same identity holds for cumulants of the D_k's as for the moments, cf. Skovgaard (1986a) and Lemma 6.6, we get

$$|\chi_k(v^k)| = \left| \sum_{m=2}^{k} \sum_{\alpha \in S_m(k)} \frac{k!}{m!} \left\{ \prod \alpha_j! \right\}^{-1} \chi_{\alpha_1 \cdots \alpha_m}(v^k) \right|.$$

Thus, from (vi) and (1.2.27) we obtain

$$\begin{aligned} |\chi_k(v^k)| &\leq \sum_{m=2}^{k} \sum_{\alpha \in S_m(k)} \frac{k!}{m!} \left\{ \prod \alpha_j! \right\}^{-1} c^2 (m-1)! \left(\prod \alpha_j! \right) \lambda^{k-2} \|v\|^k \\ &= c^2 k!\, \lambda^{k-2} \|v\|^k \sum_{m=2}^{k} \frac{1}{m} \binom{k-1}{m-1} \\ &= c^2 k!\, \lambda^{k-2} \|v\|^k k^{-1} \left(2^k - k - 1\right) \end{aligned}$$

from which it follows that (vii) holds for $m = 1$ with $\tau = 2\lambda$ and c unchanged.

The only missing point is to show the validity of (4.18) for any $k \in \mathbb{N}$. We shall prove that in some neighbourhood of β_0 the derivatives of $\exp\{\log f(y;\beta)\}$ are bounded by integrable functions. Any such derivative is of the form of a polynomial in the D_k's multiplied by $f(y;\beta)$. As an intermediate step we shall provide bounds for the D_k's. Let $y_0 \in E$ be an arbitrary point, except that we might need to avoid a null-set, and notice that, by assumption, $f(y_0;\beta)$ is analytic in some neighbourhood of β_0. Since $f(y_0;\beta_0) > 0$, the function $\log f(y_0;\beta)$ is also analytic in some neighbourhood of β_0 and therefore there exists an $R > 0$ such that

$$\left|\left\{D^k \log f(y_0;\beta_0)\right\}(v^k)\right| \leq k!\, R^k \|v\|^k$$

for all $k \in \mathbb{N}$ and $v \in V$. According to Lemma 4.1 we then have, for (almost) all $y \in E$,

$$\begin{aligned} \left|\left\{D^k \log f(y;\beta_0)\right\}(v^k)\right| &\leq \left|\left\{D^k \log f(y;\beta_0) - \chi_k\right\}(v^k)\right| \\ &\quad + \left|\left\{D^k \log f(y_0;\beta_0) - \chi_k\right\}(v^k)\right| + \left|\left\{D^k \log f(y_0;\beta_0)\right\}(v^k)\right| \\ &\leq ck!\, \|v\|^k \rho^{k-1} [H(y) + H(y_0)] + k!\, \|v\|^k R^k \end{aligned} \tag{4.19}$$

where $\rho = (2e\sqrt{p})\lambda$ and it should be noticed that $H(Y)$ from Lemma 4.1 has exponential moments. It follows that the radius of convergence of the Taylor series expansion of $\log f(y;\beta)$ around β_0 is at least $(\rho + R)^{-1}$. Thus the expansion of $\log f(y;\beta)$ is valid and absolutely convergent in some neighbourhood of β_0 that does not depend on y and therefore the same is true for the derivatives of this function. Hence, from (4.19), in some fixed neighbourhood of β_0 with $\|\beta - \beta_0\| \leq \eta$ for any $\eta < (\rho + R)^{-1}$,

$$\left|\left\{D^k \log f(y;\beta)\right\}(v^k)\right| \leq \sum_{j=0}^{\infty} \frac{(k+j)!}{j!} \|v\|^k \eta^j \left\{c\rho^{k+j-1}\left[H(y) + H(y_0)\right] + R^{k+j}\right\}$$

$$= k!\,\|v\|^k \left\{c\rho^{k-1}\left[H(y)+H(y_0)\right]/(1-\eta\rho)^{k+1} + R^k/(1-\eta R)^{k+1}\right\} \tag{4.20}$$

for any $k \in \mathbb{N}$, and

$$\begin{aligned}
&\exp|\log f(y;\beta) - \log f(y;\beta_0)| \\
&\quad \le \exp\sum_{k=1}^{\infty}\frac{1}{k!}k!\,\eta^k\left\{c\rho^{k-1}\left[H(y)+H(y_0)\right]+R^k\right\} \\
&\quad = \exp\left\{c\eta\left[H(y)+H(y_0)\right]/(1-\eta\rho)+\eta R/(1-\eta R)\right\} \\
&\quad \le \exp\left\{\delta\left[H(y)+H(y_0)+1\right]\right\} \qquad (4.21)
\end{aligned}$$

for any fixed δ if η is chosen accordingly. On combination of (4.20) and (4.21) we see that any derivative of $\exp\{\log f(y;\beta)\}$ is bounded in a neighbourhood of β_0 by a function of the form

$$q\{H(y)\}\exp\{\delta H(y)\}f(y;\beta_0)$$

for some polynomial q. Since $H(Y)$ has exponential moments it follows that any such function is integrable with respect to ν. Hence, it is permissible to interchange repeated differentiation at β_0 and integration of $f(y;\beta)$, and the proof that (vi) implies (vii) with c unchanged and $\tau = 2\lambda$ is completed.

Assume now that (vii) holds. Then, in particular, for any $k \ge 2$ and $v \in V$,

$$\left|\chi_k(v^k)\right| \le c^2 k!\,\tau^{k-2}\|v\|^k$$

while $\chi_1(v)$ is finite for any $v \in V$ because cumulants of higher order, i.a., the variance, exist. Since V is of finite dimension it follows that $\|\chi_1\| < \infty$, i.e., there exists a constant $\alpha \ge 0$, say, such that

$$|\chi_1(v)| \le \alpha\|v\|$$

for all $v \in V$. By use of Lemma 4.1 we then obtain

$$\begin{aligned}
\left|D_k(v^k)\right| &\le \left|\{D_k - \chi_k\}(v^k)\right| + \left|\chi_k(v^k)\right| \\
&\le ck!\,\rho^{k-1}H(y)\,\|v\|^k + k!\,\left(c^2\tau^{k-2}I_{\{k>1\}} + \alpha\tau^{k-1}\right)\|v\|^k
\end{aligned}$$

for all $k \in \mathbb{N}$, where $\rho = (2e\sqrt{p})\tau$. Since $H(Y)$ is known from Lemma 4.1 to have finite exponential moments it is easy to see that (iv) is satisfied if $M(y)$ is chosen appropriately and ρ possibly modified compared to above if $\tau = 0$. ∎

5 The index of a model

The classical first order asymptotic theory of likelihood based inference is based on a local approximation of the model with a normal linear model. More precisely, a typical proof of asymptotic normality of a local maximum likelihood estimator, or of the asymptotic chi-squared distribution of the likelihood ratio test statistic, requires a normal approximation to the score statistic, $D_1(\beta_0)$, and a proof of the asymptotic equivalence between the maximum likelihood estimator and an affine function of the score statistic. The latter approximation requires that higher order derivatives of the log-likelihood function are bounded in certain ways. Higher order expansions of the Edgeworth type on which, e.g., the Bartlett adjustment of the likelihood ratio statistic is based, require the derivatives of the log-likelihood, suitably scaled, to decay at a certain rate and, moreover, that also their standardized cumulants decrease in powers of this rate, typically $\sqrt{n}^{-1}$ for the case of n independent replications. In more general settings the rate may be different, e.g., in stochastic processes.

In this section we shall introduce a quantity, referred to as the index of a model at a given parameter value, which will play the role of $\sqrt{n}^{-1}$ for asymptotic likelihood based theory. When there is any risk of confusion with other quantities referred to as the index it will be called the index of linear normality, because it may be thought of as a measure of the deviation from a normal linear model in an absolute sense. Thus, it provides a bound for this deviation contrary to curvature measures that provide bounds for or approximations to second order terms only, see, e.g., Beale (1960), Efron (1975), and Bates and Watts (1980). As a consequence it is possible to prove, e.g., that the standardized distribution of a local maximum likelihood estimator converges to a normal distribution for any sequence of models for which the index tends to zero. This and other related results concerning asymptotic properties will be treated separately. We adopt the notation from sections 2 and 3, in particular, the D_k's from (2.1), the Fisher information $I(\beta)$ from (3.16), and the cumulants of the log-likelihood derivatives from (3.15), i.e., the χ's. In this section we shall frequently be working with the (semi)-norm on V defined in terms of the Fisher information instead of a pre-given (semi)-norm. Thus, consider the (pseudo) inner product

$$\langle v_1, v_2 \rangle_{I(\beta)} = I(\beta)(v_1, v_2), \qquad v_1, v_2 \in V, \tag{5.1}$$

and the corresponding (semi)-norm

$$\|v\|_{I(\beta)} = \langle v, v \rangle_{I(\beta)}, \tag{5.2}$$

cf. (1.1.14) and (1.1.15). Thus, lengths in V are measured in terms of units defined via the Fisher information, the effect of which is the same as a kind of standardization. E.g., the variance matrix of the score statistic, $D_1(\beta)$, is the identity matrix when the Fisher information is used to define the orthonormal bases on V and V^*.

Definition 5.1. *For any model $\{ f(y;\beta) : \beta \in B \subseteq V \}$ that is analytic at a point $\beta_0 \in \mathrm{int}(B)$ we define the index, $\lambda(\beta_0)$, of the model at β_0, or more precisely* *the index of linear normality,* as

$$\lambda(\beta_0) = \inf\Big\{ \lambda \geq 0 : \left|\chi_{k_1\cdots k_m}(\beta_0)\left(v_1^{k_1},\ldots,v_m^{k_m}\right)\right| \leq (m-1)!\left\{\prod_{j=1}^{m}\left(k_j!\,\|v_j\|^{k_j}_{I(\beta_0)}\right)\right\}\lambda^{k_1+\cdots+k_m-2}; \quad k_1,\ldots,k_m \in \mathbb{N}; v_1,\ldots,v_m \in V; m \geq 2\Big\}. \tag{5.3}$$

The quantity may take the value $+\infty$.

Notice the similarity with the condition (v) in Theorem 4.2. The difference is that we use the Fisher information (semi)-norm instead of a pre-defined norm, and consequently the factor c^2 in (v) is not needed because the inequality (5.3) is automatically satisfied for $\sum k_j = 2$. It follows from Theorem 4.2 that the index is finite whenever the Fisher information is positive definite because of the equivalence of any two norms on a finite dimensional vector space.

As mentioned above the use of the Fisher information norm has the effect of a standardization. However, the effect is not the same as would have been obtained if we had instead considered the standardized cumulants of the D_k's in (5.3). This latter approach would have implied a standardization in terms of the variances, $(\chi_{kk}(\beta_0))$, of the D_k's, which only for the score statistic, $D_1(\beta_0)$, corresponds to our approach.

In asymptotic theory we shall typically consider a sequence of models for which the index at a given parameter point tends to zero. To see why this limiting value is of importance assume that the index of a model equals zero at a point β_0. Then the only non-vanishing cumulants at β_0 are $\chi_{11}(\beta_0) = -\chi_2(\beta_0)$. Thus, only the first differential of the log-likelihood at β_0 is stochastic while the third and higher order differentials vanish. This implies that we have the expansion

$$\log f(y;\beta) = \log f(y;\beta_0) + D_1(\beta_0)(\beta-\beta_0) + \frac{1}{2}\chi_2(\beta_0)(\beta-\beta_0)^2 \tag{5.4}$$

which is known from Lemma 3.4 to be valid in a neighbourhood of β_0 that does not depend on y. Moreover, the β_0-distribution of the score function, $D_1(\beta_0)$, is exactly normal with mean zero because all of its cumulants vanish, except the second. Thus, the model agrees in a neighbourhood of β_0 with a Gaussian shift experiment, cf. LeCam (1986, Section 9.3). In terms of the minimal sufficient statistic, $D_1(\beta_0)$, the model is a linear normal model on a finite dimensional vector space. Any such model is of the form (5.4) and hence characterized by the property that the index is zero. Therefore, if the index tends to zero for a sequence of models a certain, quite strong, form of convergence to a linear normal model, or more precisely to a Gaussian shift experiment, is guaranteed.

If the model is analytic in the entire parameter space $B \subseteq V$, then $\lambda(\beta_0) = 0$ at some fixed point $\beta_0 \in B$ is seen to imply that $\lambda(\beta) = 0$ for all $\beta \in B$ because the Taylor series expansion (5.4) is of finite length and hence of infinite radius of convergence.

Less drastic demands than $\lambda(\beta_0) = 0$ may also imply simplifications of the model. To see this assume that instead of the inequality (5.3) we have

$$\left|\chi_{k_1\cdots k_m}(\beta_0)\left(v_1^{k_1},\ldots,v_m^{k_m}\right)\right| \\ \leq (m-1)!\left\{\prod_{j=1}^{m}\left(k_j!\,\|v_j\|_{I(\beta_0)}^{k_j}\right)\right\}\lambda_N(\beta_0)^{m-2}\lambda_L(\beta_0)^{k_1+\cdots+k_m-m} \tag{5.5}$$

for two constants $\lambda_N(\beta_0)$ and $\lambda_L(\beta_0)$, still considering only $m \geq 2$. The two possibilities $\lambda_N(\beta_0) = 0$ and $\lambda_L(\beta_0) = 0$ are of special interest. First notice that the expansion

$$\log f(y;\beta) = \log f(y;\beta_0) + D_1(\beta_0)(\beta - \beta_0) + \frac{1}{2}D_2(\beta_0)(\beta - \beta_0)^2 + \cdots \tag{5.6}$$

is known to be valid (and absolutely convergent) in some neighbourhood of β_0 that does not depend on y. Now, if $\lambda_N(\beta_0) = 0$ all cumulants of order three and higher vanish and consequently the β_0-distribution of $(D_1(\beta_0), D_2(\beta_0), \ldots)$ is exactly normal. Thus, the model agrees locally with a non-linear normal model which reduces to a linear normal model if and only $\lambda_L(\beta_0) = 0$, also. Usually, in the case $\lambda_N(\beta_0) = 0$, the statistic $(D_1(\beta_0), D_2(\beta_0), \ldots)$ will be concentrated on a finite dimensional linear subspace and then (5.6) takes the form

$$\log f(y;\beta) = \log f(y;\beta_0) + g(\beta)(X), \tag{5.7}$$

where $X \in W$, $g(\beta) \in W^*$, and W is a finite dimensional real vector space. In terms of the sufficient statistic X this model is a non-linear normal regression model with known variance. However, (5.6) allows also for the possibility of infinite-dimensional observations.

In the second case, $\lambda_L(\beta_0) = 0$, all differentials $D_k(\beta_0)$ are degenerate, except the first differential $D_1(\beta_0)$ which is then sufficient. Then (5.6) reduces to the form

$$\log f(y;\beta) = \log f(y;\beta_0) + D_1(\beta_0)(\beta - \beta_0) + h(\beta), \tag{5.8}$$

for some analytic function $h : U(\beta_0) \to \mathbf{R}$, defined on some neighbourhood of β_0. This is the form of an exponential family with canonical parameter β and canonical sufficient statistic $D_1(\beta_0)$. We shall refer to such a model as a *linear exponential family* as opposed to a curved exponential family for which the canonical parameter might be a non-linear function of β.

It appears that in (5.5), $\lambda_N(\beta_0)$ is a measure of deviance from a normal model while $\lambda_L(\beta_0)$ is a measure of deviance from a linear exponential family, or equivalently a measure of insufficiency of the score statistic. However, the two quantities, $\lambda_N(\beta_0)$ and $\lambda_L(\beta_0)$, are not well defined from (5.5) unless one of them is zero. One might be decreased on behalf of the other, still fulfilling the inequality (5.5).

Another justification for the consideration of the index of a model is provided by the case of independent replications. Assume that a model for $Y \in E$ is analytic at $\beta_0 \in B \subseteq V$ and let $\chi_{k_1\cdots k_m}(\beta_0)$ denote the cumulants and $\lambda(\beta_0)$ the index at β_0. Let $Y_1,\ldots,Y_n$ be independent random variables from this model. Then the differentials of the log-density at β_0 for the model for $(Y_1,\ldots,Y_n)$ will be given by

$$D_k^{(n)}(\beta_0) = \sum_{i=1}^{n} D_{k,i}(\beta_0) \tag{5.9}$$

for $k \in \mathbb{N}$, where $D_{k,i}(\beta_0)$ denotes the kth differential corresponding to Y_i. Also, in obvious notation,

$$\chi^{(n)}_{k_1\cdots k_m}(\beta_0) = n\chi_{k_1\cdots k_m}(\beta_0) \tag{5.10}$$

for all $k_1,\ldots,k_m \in \mathbb{N}$ and $m \in \mathbb{N}$. In particular, $I^{(n)}(\beta_0) = nI(\beta_0)$ denotes the Fisher information based on the n observations.

Theorem 5.2. *Consider n independent replications from a model $\{ f(y;\beta) : \beta \in B \subseteq V \}$ which is analytic at $\beta_0 \in \text{int}(B)$. The model based on the n replications is analytic at β_0 and the index, $\lambda^{(n)}(\beta_0)$, at β_0 from this model satisfies the relation*

$$\lambda^{(n)}(\beta_0) = \lambda(\beta_0)/\sqrt{n}. \tag{5.11}$$

Proof. It is trivial to check that the conditions (i)-(iii) of Definition 2.1 hold for the model of $(Y_1,\ldots,Y_n)$ if we take E_1^n as the set of probability one and the product of the densities as the density with respect to the n-fold product of the underlying measure ν with itself. The relation (5.10) immediately shows that if any of the cumulant conditions (v), (vi) or (vii) holds for the model for a single observation then it also holds for the model for n observations, thus proving that the latter model is also analytic at β_0. A simple substitution of (5.10) and the relation between the Fisher informations into (5.3) shows that the indices satisfy the relation (5.11) whether they are both finite or infinite. ∎

An obvious extension is to consider independent but not necessarily identically distributed observations from analytic models. Let $Y_1 \in E_1,\ldots,Y_n \in E_n$ be independent random variables, the model for each of which being parametrized by $\beta \in B \subseteq V$ and analytic at β_0. Denote the cumulants of the log-likelihood derivatives from the entire experiment by

$$\chi^{(n)}_{k_1\cdots k_m}(\beta_0) = \sum_{i=1}^{n} \kappa^{(i)}_{k_1\cdots k_m}(\beta_0), \tag{5.12}$$

expressed here in terms of the corresponding cumulants from the individual models. Let $I^{(n)}(\beta_0)$ denote the Fisher information for the entire experiment and let $I_i(\beta_0)$ be the Fisher informations for the individual models such that

$$I^{(n)}(\beta_0) = \sum_{i=1}^{n} I_i(\beta_0). \tag{5.13}$$

Theorem 5.3. *Let $Y = (Y_1, \ldots, Y_n)$ be such that $Y_1, \ldots, Y_n$ are independent and that the model for each of the Y_i's is analytic at β_0. Then the model for Y is analytic at β_0 and, with notation from above, the index $\lambda^{(n)}(\beta_0)$ of the model for Y at β_0 satisfies*

$$\lambda^{(n)}(\beta_0) \le a_n = \sup\{ a_n(v) : v \in V \}, \tag{5.14}$$

where $a_n(v) \ge 0$ is given by

$$a_n(v)^2 = \sup\left\{ \lambda_i(\beta_0)^2[I_i(\beta_0)(v^2)] / \sum_{j=1}^{n} I_j(\beta_0)(v^2) : i = 1, \ldots, n \right\} \tag{5.15}$$

in terms of the indices and the Fisher informations from the models for the Y_i's.

Proof. The conditions (i)-(iii) of Definition 2.1 are easily verified for the model for Y with the obvious choices of density and underlying measure on the product space. As the set E_1 in the definition we take the product of the corresponding sets for the models for each of the Y_i's. Since any of the cumulants in (5.12) for the model for Y is the sum of the corresponding cumulants from the individual models it is trivial to see from any of the cumulant conditions (v)-(vii) in Theorem 4.2 that it holds for the model for Y because it holds for the models for the Y_i's. To obtain the inequality for the index consider the following computation in which we notationally suppress any dependence on β_0,

$$\begin{aligned}
&\left|\chi_{k_1\cdots k_m}(v_1^{k_1}, \ldots, v_m^{k_m})\right| \\
&\quad = \left|\sum_{i=1}^{n} \kappa^{(i)}_{k_1\cdots k_m}(v_1^{k_1}, \ldots, v_m^{k_m})\right| \\
&\quad \le (m-1)! \left(\prod k_j!\right) \sum_{i=1}^{n} \lambda_i^{k_1+\cdots+k_m-2} I_i(v_1^2)^{k_1/2} \cdots I_i(v_m^2)^{k_m/2}.
\end{aligned} \tag{5.16}$$

For each factor of the form $\lambda_i\sqrt{I_i(v_j^2)}$ we have the upper bound

$$\lambda_i\sqrt{I_i(v_j^2)} \le a_n(v_j)\sqrt{I^{(n)}(v_j^2)} \le a_n\sqrt{I^{(n)}(v_j^2)}.$$

If we use this estimate for $\sum k_j - 2$ of the factors in the last sum in (5.16) and leave $\sqrt{\{I_i(v_1^2)I_i(v_2^2)\}}$ in the sum we obtain

$$\begin{aligned}
&\left|\chi_{k_1\cdots k_m}(v_1^{k_1}, \ldots, v_m^{k_m})\right| \\
&\quad \le (m-1)! \left(\prod k_j!\right) a_n^{k_1+\cdots+k_m-2} \{I^{(n)}(v_1^2)\}^{(k_1-1)/2} \{I^{(n)}(v_2^2)\}^{(k_2-1)/2} \\
&\quad\quad \times \left(\prod_{j=3}^{m} \{I^{(n)}(v_j^2)\}^{k_j/2}\right) \sum_{i=1}^{n} \{I_i(v_1^2) I_i(v_2^2)\}^{1/2},
\end{aligned}$$

where the product should be read as 1 if $m = 2$. This computation shows that (5.3) holds with $\lambda = a_n$ because of the Cauchy-Schwartz inequality

$$\left|\sum b_i c_i\right| \leq \left(\sum b_i^2\right)^{1/2} \left(\sum c_i^2\right)^{1/2}$$

applied to $b_i = \sqrt{I_i(v_1^2)}$ and $c_i = \sqrt{I_i(v_2^2)}$. ∎

Theorem 5.2 is, of course, a special case of Theorem 5.3. The bound (5.14) for the index in Theorem 5.3 may not be optimal but turns out to be good enough in many situations, provided that the indices for the models for the Y_i's are all finite. The reason is that if the bound a_n does not tend to zero for a sequence of models, then the contribution from one of the n random variables will be non-negligible, asymptotically speaking. In that case expansions of the usual type will hardly be valid. There are, of course, cases where, e.g., a first order expansion may still be valid if it is the behaviour of higher order cumulants that prevents the bound in (5.14) from converging to zero.

As argued above it is of interest what happens as the index tends to zero for a sequence of models. In this connection we shall be needing bounds for the log-likelihood derivatives similar to those in (iv) in Definition 2.1 and (4.1) in Lemma 4.1, but in a slightly different form and, as (4.1), expressible in terms of the index. These bounds are given in the following corollary.

Corollary 5.4. *Assume that the model* $\{ f(y;\beta) : \beta \in B \subseteq V \}$ *is analytic at* β_0 *and that the constants* $\lambda \geq 0$ *and* $c \geq 0$ *satisfy the directional mixed cumulant condition (vi) in Theorem 4.2, where* $\|\cdot\|$ *denotes a given semi-norm. Then for all* $y \in E$, $k \in \mathbf{N}$, *and* $v \in V$ we have

$$|D_k(\beta_0)(v^k)| \leq k!\|v\|^k \left\{c^2 \rho^{k-2} I_{\{k>1\}} + c\rho^{k-1} H(y;\beta_0)\right\}, \tag{5.17}$$

where $\rho = 2(e\sqrt{p})\lambda$, $I_{\{\cdot\}}$ *is the indicator function of the set in question, and the function* $H(\cdot;\beta_0)$ *of* E *into* $\mathbf{R}$ *is finite with probability one and satisfies the exponential moment inequality (4.3). In particular, when the Fisher information semi-norm is used, (5.17) holds with* $c = 1$ *and* λ *equal to the index* $\lambda(\beta_0)$.

Proof. We split $D_k(\beta_0)$ into a deterministic and a random part,

$$D_k(\beta_0) = \chi_k(\beta_0) + \{D_k(\beta_0) - \chi_k(\beta_0)\}.$$

The random part is known from Lemma 4.1 to be bounded by the second term in (5.17) and for $k \geq 2$ the first term is the bound for $\chi_k(\beta_0)(v^k)$ proved in Theorem 4.2, in fact with $\rho = 2\lambda$. Finally we know from Lemma 3.5 that $\chi_1(\beta_0) = 0$. ∎

It is natural for an analytic model to define the index $\lambda(\beta_0; v)$ in direction $v \in V$ from β_0, for $v \neq 0$, as the index at β_0 for the submodel with parameter space

$$\{ \beta_0 + hv : h \in \mathbf{R} \}.$$

Then the inequality (5.3) holds for all $v = v_1 = \cdots = v_m$ if $\lambda(\beta_0)$ is replaced by $\lambda(\beta_0; v)$. While this concept is obviously of relevance to properties of the submodel, it is not very useful for the full model. The reason is that although the inequality (5.17) then holds with $\rho(\beta_0)$ replaced by $\rho(\beta_0; v) = 2(e\sqrt{p})\lambda(\beta_0; v)$, the function H will also depend on the direction and we have no uniform control on the magnitude of H, as we have from (4.3) in Lemma 4.1.

We already know from Lemma 3.4 that if a model is analytic at a point then it is also analytic in a neighbourhood of the point. It is a trivial matter to show that if the index is finite at such a point then it is finite in a neighbourhood and, in fact, bounded in such a neighbourhood, but this property turns out to be too weak for our purposes. As seen in Corollary 5.7 below, the bound for the index in a neighbourhood of a point can be given purely in terms of the index at the point. The following lemma provides the tools for the derivation of that result.

Lemma 5.5. *Let the model $\{ f(y;\beta) : \beta \in B \subseteq V \}$ be analytic at β_0 and assume that the constants $c \geq 0$ and $\lambda \geq 0$ satisfy the mixed cumulant condition (v) in Theorem 4.2 for some semi-norm $\|\cdot\|$. Then for any β in the set $U_a(\beta_0)$, cf. (3.1), with $a = (2\lambda)^{-1}$, the bound*

$$\left|\chi_{k_1\cdots k_m}(\beta)(v_1^{k_1}, \dots v_m^{k_m})\right| \leq c^2(m-1)! \left(\prod_{j=1}^{m} k_j! \, \|v_j\|^{k_j}\right) \lambda^{\Sigma k_j - 2} (1 - \lambda\|\beta - \beta_0\|)^{-\Sigma k_j} (1 - 2\lambda\|\beta - \beta_0\|)^{-m} \tag{5.18}$$

holds for all $k_j \in \mathbb{N}$, $v_j \in V$, $j = 1, \dots, m \geq 2$. In particular, (5.18) holds with $c = 1$ and λ equal to the index of the model at β_0 when the Fisher information semi-norm is used.

Proof. To expand $\chi_{k_1\cdots k_m}(\beta)$ around β_0 we need an expression for any derivative of $\chi_{k_1\cdots k_m}(\beta)$ at β_0. From Skovgaard (1986a), see also Lemma 6.6, it follows that these derivatives are obtained by replacing $\chi_{k_1\cdots k_m}$ by the corresponding moment $\mu_{k_1\cdots k_m}$ from (3.13), for which we have the expression (3.7), differentiating this repeatedly under the integral sign, and then substitute χ's for μ's throughout the resulting expression. To obtain the derivatives of (3.7) we write $f(y;\beta)$ as $\exp\{\log f(y;\beta)\}$ and use Leibnitz' rule (1.2.16) generalized to a product of several functions combined with the rule (1.2.26) for higher order differentials of the composite function $\exp\{\log f(y;\beta)\}$. We use γ to denote the number of differentiations of this function, and obtain

$$\left|\{D^s \chi_{k_1\cdots k_m}(\beta_0)(v_1^{k_1}, \dots, v_m^{k_m})\}(w^s)\right|$$
$$= \Bigg| \sum_{a_1+\cdots+a_m+\gamma=s} \binom{s}{a_1 \cdots a_m \, \gamma} \sum_{r=1}^{\gamma} \sum_{b \in S_r(\gamma)} \frac{\gamma!}{r!} \left\{\prod_{j=1}^{r} b_j!\right\}^{-1}$$

$$
\begin{aligned}
&\times \chi_{(k_1+a_1)\cdots(k_m+a_m)b_1\cdots b_r}(v_1^{k_1}, w^{a_1}, \ldots, v_m^{k_m}, w^{a_m}, w^{\gamma})\Big| \\
&\le \sum_{a_1+\cdots+a_m+\gamma=s} s!\left(\gamma!\prod a_j!\right)^{-1} \sum_{r=1}^{\gamma}\sum_{b\in S_r(\gamma)} \frac{\gamma!}{r!}\left\{\prod b_j!\right\}^{-1} c^2(m+r-1)! \\
&\quad\times\left\{\prod_{j=1}^{m}\left[(k_j+a_j)!\,\|v_j\|^{k_j}\right]\right\}\left\{\prod_{j=1}^{r} b_j!\right\}\|w\|^s\lambda^{\Sigma k_j+s-2} \\
&= c^2 s!\,\lambda^{\Sigma k_j+s-2}\left\{\prod\|v_j\|^{k_j}\right\}\|w\|^s \sum_{a_1+\cdots+a_m+\gamma=s}\left\{\prod_{j=1}^{m}[(k_j+a_j)!/a_j!]\right\} \\
&\quad\times\sum_{r=1}^{\gamma}\sum_{b\in S_r(\gamma)}\frac{(m+r-1)!}{r!} \\
&= c^2 s!\,\lambda^{\Sigma k_j+s-2}\left\{\prod\|v_j\|^{k_j}\right\}\|w\|^s\sum_{r=0}^{s}\frac{(m+r-1)!}{r!}\sum_{(a,b)}^{(s,r)}\prod_{j=1}^{m}\{(k_j+a_j)!/a_j!\},
\end{aligned}
\tag{5.19}
$$

where $S_r(\gamma)$ is the set of sequences defined in (1.2.24) and the last sum is over all sequences $(a_1, \ldots, a_m) \in \mathbb{N}_0^m$ and $(b_1, \ldots, b_r) \in \mathbb{N}^r$ with $\sum a_j + \sum b_j = s$. Obvious modifications of the expressions are required for the terms with $\gamma = 0$ but the last expression may be seen to be valid without such reservations. This expression may now be used to obtain a bound for the Taylor series expansion for $\chi_{k_1\cdots k_m}(\beta)$. For $w = \beta - \beta_0$ sufficiently small we get

$$
\begin{aligned}
&\left|\chi_{k_1\cdots k_m}(\beta)(v_1^{k_1}, \ldots, v_m^{k_m})\right| \\
&\le \sum_{s=0}^{\infty}\frac{1}{s!}\left|\{D^s\chi_{k_1\cdots k_m}(\beta_0)(v_1^{k_1}, \ldots, v_m^{k_m})\}(w^s)\right| \\
&\le c^2\lambda^{\Sigma k_j-2}\left\{\prod\|v_j\|^{k_j}\right\}\sum_{r=0}^{\infty}\frac{(m+r-1)!}{r!} \\
&\quad\times\sum_{a_1=0}^{\infty}\cdots\sum_{a_m=0}^{\infty}\sum_{b_1=1}^{\infty}\cdots\sum_{b_r=1}^{\infty}\frac{(k_1+a_1)!}{a_1!}\cdots\frac{(k_1+a_1)!}{a_1!}(\lambda\|w\|)^{\Sigma a_j+\Sigma b_j} \\
&= c^2(m-1)!\left\{\prod\left(k_j!\,\|v_j\|^{k_j}\right)\right\}\lambda^{\Sigma k_j-2}\sum_{r=0}^{\infty}\binom{m+r-1}{r} \\
&\quad\times\left\{\prod_{j=1}^{m}\sum_{a_j=0}^{\infty}\binom{k_j+a_j}{a_j}(\lambda\|w\|)^{a_j}\right\}\left(\frac{\lambda\|w\|}{1-\lambda\|w\|}\right)^r \\
&= c^2(m-1)!\left\{\prod\left(k_j!\,\|v_j\|^{k_j}\right)\right\}\lambda^{\Sigma k_j-2}\sum_{r=0}^{\infty}\binom{m+r-1}{r}\left(\frac{\lambda\|w\|}{1-\lambda\|w\|}\right)^r \\
&\quad\times(1-\lambda\|w\|)^{-\Sigma k_j-m}
\end{aligned}
$$

$$= c^2(m-1)! \left\{\prod \left(k_j! \|v_j\|^{k_j}\right)\right\} \lambda^{\Sigma k_j - 2} \left(\frac{1-\lambda\|w\|}{1-2\lambda\|w\|}\right)^m (1-\lambda\|w\|)^{-\Sigma k_j - m}$$

which equals the right hand side of (5.18). These computations are valid whenever $\lambda\|w\| < \frac{1}{2}$, provided that β is within the set $U(\beta_0)$ on which $\log f(y;\beta)$ is known to be analytic. ∎

Lemma 5.6. *Let the model $\{ f(y;\beta) : \beta \in B \subseteq V \}$ be analytic at β_0 with finite index $\lambda(\beta_0)$. Then for any β in the set $U_a(\beta_0)$ from (3.1) defined in terms of the Fisher information semi-norm, with $a = \frac{1}{2}\lambda(\beta_0)^{-1}$, and any $v \in V$, we have*

$$|I(\beta)(v^2) - I(\beta_0)(v^2)| \\ \leq I(\beta_0)(v^2)\left\{\left(1-\lambda(\beta_0)\|\beta-\beta_0\|_{I(\beta_0)}\right)^{-2}\left(1-2\lambda(\beta_0)\|\beta-\beta_0\|_{I(\beta_0)}\right)^{-2} - 1\right\}. \tag{5.20}$$

Proof. The quantity $I(\beta) = \chi_{11}(\beta)$ is, of course, one particular of the cumulants considered in Lemma 5.5, and the proof of (5.20) is obtained by mimicking the proof of that lemma, except that $c = 1$, that the index $\lambda(\beta_0)$ replaces λ, that the Fisher information semi-norm is used, and that we subtract the zero-order term $I(\beta_0)(v^2)$ from the Taylor series expansion used to obtain the bound. From (5.19) it is seen that the bound for this term, corresponding to $s = 0$, equals the term itself. Therefore the bound (5.20) is obtained from the bound (5.18) applied to this particular cumulant, simply by subtraction of $I(\beta_0)(v^2)$. ∎

Corollary 5.7. *Assume that the model $\{ f(y;\beta) : \beta \in B \subseteq V \}$ is analytic at β_0 with finite index $\lambda(\beta_0)$. Then it is analytic with finite index at any β in the set $U_a(\beta_0)$ from (3.1) defined in terms of the Fisher information semi-norm, with $a = a_0/\lambda(\beta_0)$, where a_0 is a universal constant, and for any such β we have the inequality*

$$\lambda(\beta) \leq \lambda(\beta_0)\, h\left(\lambda(\beta_0)\|\beta-\beta_0\|_{I(\beta_0)}\right), \tag{5.21}$$

where $h : [0, a_0) \to \mathbf{R}_+$ is a continuous function satisfying $h(0) = 1$. In fact, possible choices of a_0 and h are

$$a_0 = \frac{3}{4} - \frac{1}{4}\sqrt{1+4\sqrt{2}} \tag{5.22}$$

and

$$h(x) = z^3(2-z^2)^{-3/2} \tag{5.23}$$

where $z = \{(1-x)(1-2x)\}^{-1}$.

Proof. Let $x = \lambda(\beta_0)\|\beta-\beta_0\|_{I(\beta_0)} < \frac{1}{2}$. For any $v \in V$ with $\|v\|_{I(\beta_0)} > 0$ simple manipulations of (5.20) yield

$$\begin{aligned} \|v\|_{I(\beta_0)} / \|v\|_{I(\beta)} &= \left\{I(\beta_0)(v^2)/I(\beta)(v^2)\right\}^{1/2} \\ &\leq \left\{2 - (1-x)^{-2}(1-2x)^{-2}\right\}^{-1/2} \\ &= (2-z^2)^{-1/2}, \end{aligned}$$

if $z^2 < 2$, where z is defined below (5.23). Notice that this inequality for z is guaranteed by the choice of a_0 in (5.22). Now consider the inequalities defining the index in (5.3) at the point β in $U_a(\beta_0)$. From Lemma 5.6 we know that if $\|v_j\|_{I(\beta_0)} = 0$ for any of the v_j's then also $\|v_j\|_{I(\beta)} = 0$ and the inequality (5.3) will be satisfied for any choice of $\lambda(\beta)$. Hence we exclude such v_j's in the sequel. Also notice that if $k_1 + \cdots + k_m \geq 3$ and $m \geq 2$ then

$$m \leq \sum k_j \leq 3\left(\sum k_j - 2\right).$$

Hence, with x and z as above, it follows from Lemma 5.5 that

$$\begin{aligned}
&\left|\chi_{k_1\cdots k_m}(\beta)(v_1^{k_1},\ldots,v_m^{k_m})\right| \\
&\quad\leq (m-1)!\left\{\prod\left(k_j!\|v_j\|_{I(\beta)}^{k_j}\right)\right\}\lambda(\beta_0)^{\Sigma k_j-2}(1-x)^{-\Sigma k_j}(1-2x)^{-m} \\
&\qquad\times\left\{\prod\left(\|v_j\|_{I(\beta_0)}/\|v_j\|_{I(\beta)}\right)^{k_j}\right\} \\
&\quad\leq (m-1)!\left\{\prod\left(k_j!\|v_j\|_{I(\beta)}^{k_j}\right)\right\}\lambda(\beta_0)^{\Sigma k_j-2}\left\{z^3(2-z^2)^{-3/2}\right\}^{\Sigma k_j-2}
\end{aligned}$$

as was to be proved. The case $\sum k_j = 2$ in (5.3) is satisfied by definition of the Fisher information semi-norm. Thus we have proved that the mixed cumulant condition (v) in Theorem 4.2 holds. It is easy to see that the conditions (i)-(iii) are satisfied at β in $U_a(\beta_0)$ and hence that the model is analytic throughout this neighbourhood. The computation above shows that the index is finite in this set and satisfies the inequality (5.21). ∎

The importance of Corollary 5.7 is not so much the explicit expressions for a_0 and h in (5.21), as their existence. In particular, the result implies the continuity of the index at any point β_0 where it is finite. Given $\epsilon > 0$, it follows immediately from the corollary that $\lambda(\beta) < \lambda(\beta_0)+\epsilon$ in some neighbourhood of β_0, proving the upper semi-continuity of the index. The lower semi-continuity follows directly from the definition of the index in (5.3) because it is a supremum of continuous functions. To formulate the definition of the index precisely as a supremum, attention should be restricted to vectors v_j with $\|v_j\|_{I(\beta_0)} > 0$. In fact, the lower semi-continuity also follows from Corollary 5.7 used "the other way around", i.e., to give a bound for $\lambda(\beta_0)$ in terms of $\lambda(\beta)$, re-expressed in terms of the $I(\beta_0)$-norm by use of Lemma 5.6. A reasonable conjecture is that the index is analytic at any point where it is finite.

Incidentally, Lemma 5.6 also leads to the following result.

Corollary 5.8. *Assume that the model* $\{ f(y;\beta) : \beta \in B \subseteq V \}$ *is analytic in an open connected subset* $B_0 \subseteq B$. *Suppose that there is a point* $\beta_0 \in B_0$ *at which the index is finite and that* $\|v\|_{I(\beta_0)} = 0$ *for some fixed vector* $v \in V$. *Then* $\|v\|_{I(\beta)} = 0$ *for all* $\beta \in B_0$.

Proof. Let

$$A = \{ \beta \in B_0 : \|v\|_{I(\beta)} = 0 \}.$$

The set A is non-empty because it contains β_0. From Lemma 3.5 we know that the mapping $\beta \mapsto I(\beta)$ is continuous, and hence the same is true for the mapping $v \mapsto \|v\|_\beta$. Thus, A is closed relative to B_0 because it is the intersection of B_0 with a closed set. From (5.20) it follows directly that A is open. Since B_0 is connected and open we conclude that $A = B_0$. ∎

The importance of Corollary 5.8 is that it shows that for any analytic model with finite index it is always possible to find a linear reparametrization of the model such that the Fisher information becomes positive definite, without changing the model viewed as a family of probability measures. To see this notice that if $I(\beta_0)(v^2) = 0$ for some $v \neq 0$ then the model is constant along any line of the form $\beta = \beta_1 + hv \in B_0$ where $h \in \mathbf{R}$, because $I(\beta)(v^2) = 0$ for all $\beta \in B_0$, and hence Corollary 5.4 shows that all derivatives of $\log f(y;\beta)$ vanish in the direction v. Thus a new parametrization of the model may be obtained by projecting the parameter space onto a subspace of dimension one less than V, along the direction v. This process may be repeated until the Fisher information is strictly positive for all vectors $v \neq 0$.

The following lemma shows that, conversely, the index is finite if the model is constant in any direction in which the Fisher information is zero.

Lemma 5.9. *Assume that the model* $\{ f(y;\beta) : \beta \in B \subseteq V \}$ *is analytic at the point* $\beta_0 \in \operatorname{int}(B)$. *Then the index,* $\lambda(\beta_0)$, *of the model at* β_0 *is finite if and only if, for any* $v \neq 0$, $v \in V$, *the condition* $\|v\|_{I(\beta_0)} = 0$ *implies that the model is constant on a line segment of the form*

$$\{ \beta_h = \beta_0 + hv : h \in (a,b) \} \tag{5.24}$$

with $a < 0$ *and* $b > 0$.

Proof. Consider the linear subspace

$$N = \{ v \in V : \|v\|_{I(\beta_0)} = 0 \}$$

and let L be a complementary linear space, i.e., a subspace of V such that any vector $v \in V$ can be uniquely represented as a sum

$$v = v_N + v_L$$

where $v_N \in N$ and $v_L \in L$. Then, if the model is constant on a line segment of the form (5.24) for any $v \in N$, it follows that

$$D_k(\beta_0)(v^k) = D_k(\beta_0)(v_L^k)$$

for all $k \in \mathbf{N}$ and $v \in V$, and consequently that

$$\chi_{k_1\cdots k_m}(\beta_0)(v_1^{k_1},\ldots,v_m^{k_m}) = \chi_{k_1\cdots k_m}(\beta_0)\left((v_1)_L^{k_1},\ldots,(v_m)_L^{k_m}\right)$$

for all $k_1, \dots, k_m \in \mathbb{N}$ and $v_1, \dots, v_m \in V$. Let $\|\cdot\|$ be any given norm on V and consider the mixed cumulant condition (4.9) in terms of this norm. Because the Fisher information norm $\|v\|_{I(\beta_0)}$ is positive definite on the subspace L there exists a constant $a > 0$ such that

$$\|v\|_{I(\beta_0)} = \|v_L\|_{I(\beta_0)} \geq a\|v_L\|$$

for all $v \in V$. Hence a suitable modification of the constants c and λ in the mixed cumulant condition shows that this condition also holds with the Fisher information semi-norm in place of $\|\cdot\|$ and it is concluded that the index at β_0 is finite. The converse statement is contained in Corollary 5.8. ∎

6 Invariance properties

The title of this section refers to operations on analytic models that lead to other analytic models – not to invariance of the models in any stronger sense. Such operations include analytic reparametrization, reductions by sufficiency and ancillarity, and products of analytic models in the sense of independent observations as we saw in the previous section.

Consider a model $\{ f(y;\beta) : \beta \in B \subseteq V \}$ assumed to be analytic at $\beta_0 \in \mathrm{int}(B)$. Let us first investigate the problem of reparametrizations. As in the previous sections V denotes a finite-dimensional real vector space. Suppose that we have a mapping, conveniently, although somewhat incorrectly, denoted β,

$$\beta : A \to B, \tag{6.1}$$

where A is a subset of a finite-dimensional real vector space W. Assume also that there is a point $\alpha_0 \in \mathrm{int}(A)$ with $\beta(\alpha_0) = \beta_0$ and consider the model

$$\{ \tilde{f}(y;\alpha) : \alpha \in A \subseteq W \}, \qquad \tilde{f}(y;\alpha) = f\big(y;\beta(\alpha)\big), \tag{6.2}$$

parametrized by $\alpha \in A$. This setup includes smoothly parametrized sub-models as well as one-to-one reparametrizations. Recall the notation for the log-likelihood differentials, (D_k), and their cumulants, $(\chi_{k_1\cdots k_m})$, from (2.1) and (3.15). The corresponding quantities defined in terms of the parameter α will be equipped with a tilde, e.g.,

$$\tilde{D}_k(\alpha) = D^k \log \tilde{f}(y;\alpha). \tag{6.3}$$

The log-likelihood differentials in the reparametrized (sub)-model may be expressed in terms of the D_k's from the original parametrization as

$$\begin{aligned}\tilde{D}_k&(\alpha)(w^k)\\ &= \sum_{m=1}^{k} \sum_{b \in S_m(k)} \frac{k!}{m!} \left\{ \prod b_j! \right\}^{-1} D_m(\beta(\alpha)) \left\{ D^{b_1}\beta(\alpha)(w^{b_1}), \dots, D^{b_m}\beta(\alpha)(w^{b_m}) \right\}\end{aligned} \tag{6.4}$$

with $w \in W$ and $k \in \mathbb{N}$, which is obtained by differentiation of the logarithm of both sides of (6.2) by use of (1.2.26). The set $S_m(k)$ is defined in (1.2.24). The relation is noticed to be linear in the original D_k's.

Theorem 6.1. *The model $\{\tilde{f}(y;\alpha) : \alpha \in A \subseteq W\}$ defined in (6.2) above is analytic at the point $\alpha_0 \in \operatorname{int}(A)$ if the mapping $\beta : A \to B$ is analytic at α_0 and the model $\{f(y;\beta) : \beta \in B \subseteq V\}$ is analytic at $\beta_0 = \beta(\alpha_0)$. Moreover, if the semi-norm $\|\cdot\|$ and the quantities $c \geq 0$, $\lambda_N \geq 0$ and $\lambda_L \geq 0$ satisfy the cumulant condition*

$$\left|\chi_{k_1\cdots k_m}(\beta_0)(v_1^{k_1},\ldots,v_m^{k_m})\right| \leq c^2(m-1)!\left\{\prod_{j=1}^{m}\left(k_j!\,\|v_j\|^{k_j}\right)\right\}\lambda_N^{m-2}\lambda_L^{\Sigma k_j - m} \tag{6.5}$$

for all $k_j \in \mathbb{N}$, $v_j \in V$ and $m \geq 2$, and if for some $a \geq 0$ and $R \geq 0$ the inequality

$$\|D^k\beta(\alpha_0)(w^k)\| \leq ak!\,R^{k-1}\|w\|^k \tag{6.6}$$

holds for some semi-norm $\|\cdot\|$ on W and all $k \in \mathbb{N}$ and $w \in W$, then

$$\left|\tilde{\chi}_{k_1\cdots k_m}(\alpha_0)(w_1^{k_1},\ldots,w_m^{k_m})\right| \leq (ac)^2(m-1)!\left\{\prod_{j=1}^{m}\left(k_j!\,\|w_j\|^{k_j}\right)\right\}(a\lambda_N)^{m-2}(a\lambda_L+R)^{\Sigma k_j - m}. \tag{6.7}$$

Finally, if for some $R \geq 0$ the inequality

$$\|D^k\beta(\alpha_0)(w^k)\|_{I(\beta_0)} \leq k!\,R^{k-1}\|D\beta(\alpha_0)(w)\|_{I(\beta_0)}^k \tag{6.8}$$

holds for all $k \in \mathbb{N}$ and $w \in W$, and the index of the model parametrized by β is $\lambda(\beta_0)$, then the index at α_0 of the model (6.2) satisfies the inequality

$$\tilde{\lambda}(\alpha_0) \leq \lambda(\beta_0) + R. \tag{6.9}$$

Proof. Assume that (6.5) holds and that the model parametrized by β is analytic at β_0. Consider the conditions (i)–(iv) in Definition 2.1. Choose a neighbourhood $\tilde{U}(\alpha_0)$ such that the mapping $\alpha \mapsto \beta(\alpha)$ is analytic in $\tilde{U}(\alpha_0)$ and $\beta(\alpha) \in U(\beta_0)$ for all $\alpha \in \tilde{U}(\alpha_0)$, where $U(\beta_0)$ is the neighbourhood from Definition 2.1. Then it follows immediately from (i) that the measures $\{\tilde{P}_\alpha = P_{\beta(\alpha)}\}$ are mutually absolutely continuous on $\tilde{U}(\alpha_0)$ and that $\tilde{f}(y;\alpha)$ is strictly positive for all y in the set E_1 from Definition 2.1. That the condition (iii) is satisfied for the new model (6.2) for all $y \in E_1$ follows from the fact that a composition of analytic functions yields an analytic functions. It would be fairly easy to show directly that also (iv) holds for the new model but we shall continue to establish (6.7) which, by use of Theorem 4.2, shows that (iv) holds and hence that the model (6.2) is analytic at α_0.

We use the identity (6.4) to obtain bounds for the cumulants of the log-likelihood differentials in the new model through the following computation involving the sequences $b_s = (b_{s1}, \ldots, b_{sn})$ in $S_n(k_s)$ from (1.2.24) for which we use the identity (1.2.27). Thus, notationally discarding the arguments α_0 and β_0, we get

$$
\begin{aligned}
&\left|\tilde{\chi}_{k_1\cdots k_m}(w_1^{k_1}, \ldots, w_m^{k_m})\right| \\
&\leq \sum_{n_1=1}^{k_1} \sum_{b_1 \in S_{n_1}(k_1)} \cdots \sum_{n_m=1}^{k_m} \sum_{b_m \in S_{n_m}(k_m)} \left\{ \prod_{s=1}^{m} \left(\frac{k_s!}{n_s!} \prod_{j=1}^{n_s} b_{sj}!^{-1} \right) \right\} \\
&\quad \times \chi_{n_1\cdots n_m} \left\{ (D^{b_{11}}\beta(w^{b_{11}}), \ldots, (D^{b_{1n_1}}\beta(w^{b_{1n_1}}), \ldots, (D^{b_{mn_m}}\beta(w^{b_{mn_m}}) \right\} \\
&\leq \sum_{n_1=1}^{k_1} \sum_{b_1 \in S_{n_1}(k_1)} \cdots \sum_{n_m=1}^{k_m} \sum_{b_m \in S_{n_m}(k_m)} \left\{ \prod_{s=1}^{m} \left(\frac{k_s!}{n_s!} \prod_{j=1}^{n_s} b_{sj}!^{-1} \right) \right\} \\
&\quad \times c^2(m-1)! \lambda_N^{m-2} \lambda_L^{\Sigma n_s - m} \prod_{s=1}^{m} \left\{ n_s! \prod_{j=1}^{n_s} \left(a b_{sj}! R^{b_{sj}-1} \|w\|^{b_{sj}} \right) \right\} \\
&= (ac)^2 (m-1)! \left\{ \prod_{s=1}^{m} \left(k_s! \|w\|^{k_s} \right) \right\} (a\lambda_N)^{m-2} \\
&\quad \times \prod_{s=1}^{m} \sum_{n=1}^{k_s} \binom{k_s - 1}{n-1} (a\lambda_L)^{n-1} R^{k_s - n} \\
&= (ac)^2 (m-1)! \left\{ \prod_{s=1}^{m} \left(k_s! \|w\|^{k_s} \right) \right\} (a\lambda_N)^{m-2} \prod_{s=1}^{m} (a\lambda_L + R)^{k_s - 1}
\end{aligned}
$$

which equals the right hand side of (6.7) and hence proves that the model (6.2) is analytic at α_0.

Assume now that (6.8) holds and that $\lambda(\beta_0)$ is the index at β_0. To investigate the index, $\tilde{\lambda}(\alpha_0)$, of the new model we need to consider the Fisher information at α_0. This is given by the equality

$$
\begin{aligned}
\tilde{I}(\alpha_0)(w^2) &= \mathrm{var}_{\beta_0} \left\{ \left[D_\alpha \log \tilde{f}(y; \beta(\alpha_0)) \right] (w) \right\} \\
&= \mathrm{var}_{\beta_0} \left\{ D_1(\beta_0) \left(D\beta(\alpha_0)(w) \right) \right\} \\
&= I(\beta_0) \left(D\beta(\alpha_0)(w) \right)^2 .
\end{aligned}
$$

Thus,

$$\|w\|_{\tilde{I}(\alpha_0)} = \|D\beta(\alpha_0)(w)\|_{I(\beta_0)}.$$

Therefore the inequality (6.8) is a recast of (6.6) with $a = 1$ and with the Fisher information semi-norms on V and W. Since (6.5) holds with $c = 1$ and $\lambda_L = \lambda_N = \lambda(\beta_0)$ the result (6.9) follows from (6.7). ∎

We could, of course, have formulated the result given in (6.9) in terms of two quantities λ_N and λ_L as in (6.7). The result in that case is seen directly from (6.5) and (6.7) if we let $a = c = 1$ and replace the semi-norms with the Fisher information semi-norm. In the light of the discussion in Section 5, this shows that the quantity λ_N, in some sense bounding the deviation from a normal model, is not increased by a reparametrization, whereas λ_L may well be affected, e.g., if a linear model is reparametrized by a non-linear mapping. The quantities λ_N and λ_L are, however, only certain kinds of bounds and should not be taken as absolute measures, although the fact remains true that an analytic reparametrization of a normal model, i.e., a model with $\lambda_N = 0$ in (6.5), remains normal.

The result of the theorem, that an analytically parametrized sub-model of an analytic model is itself analytic, is a useful tool for proving that a model is analytic. For example, in Chapter 3 we shall use this result to establish the (rather trivial) result that a curved exponential family is analytic whenever the canonical parameter is an analytic function of the parameter of the model.

As a special case it follows from (6.9) that if the parametrization $\beta : A \rightarrow B$ is linear, the index of the model parametrized by α is bounded by the index of the original model parametrized by β. In particular, a one-to-one linear reparametrization does not change the index of an analytic model.

Next, let us turn to changes in the space of observations. We shall consider changes of the underlying measure, by marginalization, and by conditioning. For any such operation there will always be a choice of versions of densities in the derived model, and it is usually possible, in non-discrete models, to choose versions such that the conditions in Definition 2.1 for analyticity of the model break down. However, with sensible choices of densities several operations of the kind mentioned above lead to analytic model when the original model is analytic.

Lemma 6.2. *Assume that the model $\{(E,\nu), f(\cdot;\beta); \beta \in B \subseteq V\}$ is analytic at $\beta_0 \in \text{int}(B)$ and let $\tilde{\nu}$ be a measure dominating P_β for all β in the neighbourhood $U(\beta_0)$ from Definition 2.1. Then there exist densities $\tilde{f}(y;\beta)$ of P_β with respect to $\tilde{\nu}$ for $\beta \in U(\beta_0)$ and a measurable function $h : E \rightarrow \mathbf{R}_+$ such that*

$$\tilde{f}(y;\beta) = f(y;\beta)h(y) \qquad (6.10)$$

for all $\beta \in U(\beta_0)$ and all $y \in \tilde{E}_1$, where $\tilde{E}_1$ is a set with $P_{\beta_0}(\tilde{E}_1) = 1$. The model $\{(\tilde{E},\nu), \tilde{f}(\cdot;\beta); \beta \in B \subseteq V\}$ is then analytic at β_0.

Proof. It follows from the Lebesgue decomposition theorem and the assumptions that it is possible to choose a set $\tilde{E}_1$ such that the measures ν and $\tilde{\nu}$ are mutually absolutely continuous on this set and such that $P_{\beta_0}(\tilde{E}_1) = 1$. As the function h we then take the Radon-Nikodym derivative $h(y) = (d\nu/d\tilde{\nu})(y)$. The set of points on which $h(y) = 0$ is a null-set which therefore can be discarded from $\tilde{E}_1$, thus proving (6.10). For $y \in \tilde{E}_1$ the density $\tilde{f}(y;\beta)$ is now positive whenever $f(y;\beta)$ is positive. Thus, the conditions (i)–(iii) of Definition 2.1 are satisfied for $\tilde{f}(y;\beta)$.

Also, the log-likelihood derivatives are unchanged as well as their distributions. Therefore (iv) also holds, and the lemma is proved. ∎

It follows from the lemma that the choice of underlying measure is completely immaterial for the theory outlined so far, and, in fact, also for that in the sequel, since the log-likelihood derivatives and their distributional properties are unaffected by this choice. A convenient choice, at least for the development of theoretical results, will often be to take $\nu = P_{\beta_0}$ and $f(y;\beta_0) = 1$ for all y, in which case condition (ii) in Definition 2.1 is automatically satisfied. Also, this choice simplifies considerations related to marginal and conditional distributions as will be seen later in this section.

Now, let

$$t : E \to \tilde{E} \tag{6.11}$$

be a measurable mapping from the measure space (E,ν) to another space $\tilde{E}$ equipped with the measure $\tilde{\nu}$. The next lemma takes care of the case when t is sufficient.

Lemma 6.3. *If the model $\{(E,\nu), f(\cdot;\beta); \beta \in B \subseteq V\}$ is analytic at $\beta_0 \in \mathrm{int}(B)$ and the mapping t in (6.11) is sufficient, then the model $\{(\tilde{E},\tilde{\nu}), \tilde{f}(\cdot;\beta); \beta \in B \subseteq V\}$ is analytic at β_0 if the densities $\tilde{f}(t;\beta)$ of $t(Y)$, of the distributions induced by the distributions (P_β) of Y, are chosen appropriately. Furthermore, there exists a ν-measurable function $h : E \to \mathbf{R}_+$ such that*

$$\tilde{f}(y;\beta) = h(y) f(y;\beta) \tag{6.12}$$

for all y in a set of P_{β_0}-probability one.

Proof. The relation (6.12) is Neyman's factorization criterion, except for the strict positivity of h which is obtained simply by restricting attention to the set of y's on which h is positive. The log-likelihood derivatives are seen from (6.12) to be unchanged in the sense that

$$D^k \log f(y;\beta_0) = D^k \log \tilde{f}(t(y);\beta_0)$$

for all $y \in E$. Since they depend on Y only through $t(Y)$ their distributions are unchanged by sufficient reductions. ∎

Incidentally, the proof of Lemma 6.3 included a proof of the minimal sufficiency of the log-likelihood derivatives $(D_1(\beta_0), D_2(\beta_0), \ldots)$, provided that they are sufficient, because it was shown that they are functions of any sufficient statistic. While these derivatives at a given fixed point are not, in general, sufficient for the entire model, it is easily seen from Lemma 3.3 that they are sufficient, and hence minimally sufficient, for the model restricted to a neighbourhood of any fixed point at which the model is analytic. Notice that the minimal sufficiency does not imply that the representation is minimal, e.g., some of the $D_k(\beta_0)$'s might be degenerate. The minimal sufficiency only implies that the statistic contains no excess randomness.

Equally simple as a sufficient is the case of conditioning on an ancillary statistic. Recall that an ancillary statistic is a mapping

$$a : E \to \tilde{E},$$

say, such that the distribution of $a(Y)$, induced by the distribution P_β of Y, does not depend on β.

Lemma 6.4. *Let the model* $\{ f(y;\beta) : \beta \in B \subseteq V \}$ *be analytic at* $\beta_0 \in \operatorname{int}(B)$ *and let* $a(Y)$ *denote an ancillary statistic. Then there are versions of conditional densities,* $\tilde{f}(y \mid a_0;\beta)$ *say, of* Y *given* $a(Y) = a_0 \in \tilde{E}$ *with respect to some underlying measure on the level surface*

$$E_{a_0} = \{ y \in E : a(y) = a_0 \},$$

such that

$$f(y;\beta) = \tilde{f}(y \mid a(y);\beta)h(y), \tag{6.13}$$

where $h : E \to \mathbf{R}_+$ *is some measurable function, and such that the model for the conditional distribution is analytic for any* a_0 *in a set of* P_{β_0}*-probability one.*

Proof. The representation (6.13) is easily established, e.g., by taking conditional P_{β_0}-distributions given $a(Y)$ as the underlying measures, in which case

$$\tilde{f}(y \mid a(y);\beta) = c\{a(y)\}f(y;\beta)/f(y;\beta_0)$$

and $h(y) = f(y;\beta_0)/c\{a(y)\}$ would be proper choices, where $c(a)$ is a normalizing constant that does not depend on β because $a(Y)$ is ancillary. Then h is positive almost surely for β in some neighbourhood of β_0 and may therefore, if necessary, be modified to be positive for all y. The possibility of obtaining (6.13) with a function h that depends only on $a(y)$ is not of relevance here. Based on (6.13) it is trivial to verify that the conditions (i)–(iii) in Definition 2.1 hold for all $a_0 = a(y)$ where y belongs to a set of probability one, and that the log-likelihood derivatives, $D_k(\beta_0)$, are the same as in the original model for Y and therefore still satisfy the bound in (iv) in Definition 2.1. It only remains to be shown that $M(Y;\beta_0)$ has finite exponential moments in the conditional distributions given $a(Y)$. But for any $s > 0$,

$$\mathrm{E}\exp\{sM(Y;\beta_0)\} = \mathrm{E}\{\mathrm{E}[\exp\{sM(Y;\beta_0)\} \mid a(Y)]\}.$$

Since the expectation on the left is finite for some $s > 0$ the same must be true for the conditional expectations on the right, for almost all $a(Y)$. ∎

As a final trivial case we consider a mapping which is a cut with respect to two components of the parameter, cf. Barndorff-Nielsen (1978). Thus, let $V = V_1 \times V_2$ be the product space of two finite-dimensional real vector spaces and let

$$\beta = (\phi,\psi) \in B = B_1 \times B_2,$$

where $B_1 \subseteq V_1$ and $B_2 \subseteq V_2$. A *cut* in the model $\{f(y;\beta) : \beta \in B \subseteq V\}$ is a measurable mapping $t : E \to \tilde{E}$ such that the marginal distributions of $t(Y)$, induced by $\{P_\beta; \beta \in B\}$, depends on β only through $\phi \in B_1$ while the conditional distribution of Y given $t(Y)$ depends only on $\psi \in B_2$.

Assume that the model $\{f(y;\beta) : \beta \in B \subseteq V\}$ is analytic at $\beta_0 = (\phi_0, \psi_0)$. It is obvious, except for some measure theoretical considerations, that we may write the density $f(y;\beta)$ as a product

$$f(y;\beta) = f_1(t(y);\phi) f_2(y;\psi), \tag{6.14}$$

where $f_1(t;\phi), \phi \in B_1$, are the marginal densities of $t(Y)$ and $f_2(y;\psi), \psi \in B_2$, are the conditional densities of Y given $t(Y)$. Thus, it is clear that the log-likelihood derivatives with respect to ϕ in the marginal models, and with respect to ψ in the conditional models, are the same as in the full model. The derivatives with respect to ϕ depend on y only through $t(y)$ and consequently their distributions are unaffected by marginalization from Y to $t(Y)$. Therefore the model based on $t(Y)$ is analytic at the point ϕ_0, or at β_0 if β is considered as the parameter. The conditional distributions of the log-likelihood derivatives with respect to ψ do, however, differ form their original distributions, but it is seen as in the proof of Lemma 6.4 that these conditional models are also analytic at ψ_0.

For the general case of an (arbitrary) function $t(Y)$ it will appear from the sequel that the conditional distribution almost surely is analytic at β_0 if the original model is analytic at this point. This result is, however, not very useful as long as it is not coupled with some kind of uniformity in the conditioning variable of the bound (iv) in Definition 2.1 or of any of the bounds in Theorem 4.2. Whether the marginal distribution of $t(Y)$ is always analytic is an open question. We can show that the conditions (i)–(iii) of Definition 2.1 hold; the problem is the condition (iv). It turns out, however, that there is a fairly simple representation of the log-likelihood derivatives from the model for $t(Y)$ in terms of the conditional cumulants of the original log-likelihood derivatives given $t(Y)$. Since this representation may be of some independent interest in connection with particular cases, the remaining part of this section will be concerned with its derivation and immediate consequences.

Thus, let

$$t : E \to \tilde{E}$$

be a measurable mapping from E to another measurable space and consider the model for $t(Y)$ induced from the original model

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

which is assumed to be analytic at $\beta_0 \in \text{int}(B)$. Let Q_β denote the distribution of $t(Y)$ induced by the P_β-distribution on Y, and let $P_\beta^{\cdot t}$ denote (versions of) the conditional distributions of Y given $t(Y) = t \in \tilde{E}$. The distribution $P_\beta^{\cdot t}$ is concentrated on the level surface

$$E_t = \{ y \in E : t(y) = t \}. \tag{6.15}$$

As the underlying measure for the model for $t(Y)$ we now choose Q_{β_0} and note that as densities in the model for $t(Y)$ we may choose the functions $\tilde{f}(\cdot;\beta)$ defined by

$$\begin{aligned}(dQ_\beta/dQ_{\beta_0})(t) &= \tilde{f}(t;\beta)\\ &= \mathrm{E}_{\beta_0}\{\, f(y;\beta)/f(y;\beta_0) \mid t(Y)=t\,\}\\ &= \int_{E_t} \exp\{D_1(\beta_0)(\beta-\beta_0) + \frac{1}{2}D_2(\beta_0)(\beta-\beta_0)^2 + \cdots\}\, dP^{\cdot t}_{\beta_0}(y), \qquad (6.16)\end{aligned}$$

where the last equation is known to be valid for β in some neighbourhood of β_0. We shall not discuss the proof of this (quite trivial) result.

Lemma 6.5. *With notation and setup from above, the log-likelihood derivatives*

$$\tilde{D}_k(\beta_0) = D^k \log \tilde{f}(t;\beta_0), \qquad k \in \mathbb{N}, \qquad (6.17)$$

are given by the equations

$$\begin{aligned}\tilde{D}_k(\beta_0)(v^k) &= \sum_{m=1}^{k} \sum_{a\in S_m(k)} \frac{k!}{m!}\left\{\prod_{j=1}^{m} a_j!\right\}^{-1}\\ &\quad\times \mathrm{cum}_{\beta_0}\{D_{a_1}(\beta_0)(v^{a_1}),\ldots,D_{a_m}(\beta_0)(v^{a_m}) \mid t(Y)=t\} \qquad (6.18)\end{aligned}$$

for $v \in V$, where $S_m(k)$ is the set of sequences defined in (1.2.24) and the joint cumulant is the cumulant in the conditional P_{β_0}-distribution given $t(Y)=t$.

Before proving the lemma let us examine the equation (6.18). Notice its formal similarity with the expression for a moment in terms of its cumulants. Thus, if $D_j(\beta_0)(v^j)$ were the jth cumulant of a random variable, such as $\langle v, X\rangle$, say, $\tilde{D}^k(\beta_0)(v^k)$ its kth mean, and we forgot about the conditional cumulant on the right hand side, then the same relation would hold. Furthermore, as one extreme, suppose that $t(y)$ is constant. Then the left hand side of the equation is zero and the conditional cumulants equal the unconditional cumulants. In this case the well-known relations between the cumulants of the log-likelihood derivatives are recovered, cf. the discussion relating to equations (3.14)–(3.16) in Section 3. At the other extreme, if $t(y) = y$, the relation (6.18) reduces to the trivial identity between the D_k's and the $\tilde{D}_k$'s, because the conditional distribution of any of the D_k's is degenerate. The same identity arises, for the same reason, whenever $t(Y)$ is sufficient.

Thus, it is not surprising that the proof of the identity (6.18) is related to the relations between the cumulants of the log-likelihood derivatives. To prove it we need a generalization of the result in Skovgaard (1986a) on which the proof of these relations may be based. In that paper the rule for differentiation of the log-likelihood derivatives was given, but it turns out that this rule is not confined to log-likelihood derivatives but applies more generally to statistics depending on the parameter.

Before we state this rule of differentiation, let us consider some formal differentiations of moments. Thus, let $g : E \times B \to \mathbf{R}$ and consider

$$\begin{aligned} D_\beta \mathrm{E}_\beta\{g(Y;\beta)\} &= D_\beta \int g(y;\beta) f(y;\beta)\, d\nu(y) \\ &= \mathrm{E}_\beta\big\{D_\beta g(Y;\beta)\big\} + \mathrm{E}_\beta\big\{g(Y;\beta) D_1(\beta)\big\}. \end{aligned} \tag{6.19}$$

In particular, if

$$g(y;\beta) = \prod_{j=1}^{m} g_j(y;\beta)$$

where $g_j : E \times B \to \mathbf{R}$, then Leibnitz' rule for differentiation of a product, together with (6.19), shows that

$$\begin{aligned} D_\beta \mathrm{E}_\beta \left\{ \prod_{j=1}^{m} g_j(Y;\beta) \right\}(v) &= \sum_{i=1}^{m} \mathrm{E}_\beta \left\{ [D_\beta g_i(Y;\beta)(v)] \prod_{j\neq i} g_j(Y;\beta) \right\} \\ &\quad + \mathrm{E}_\beta \left\{ D_1(\beta)(v) \prod_{j=1}^{m} g_j(Y;\beta) \right\}. \end{aligned} \tag{6.20}$$

The, somewhat surprising, result is now that the same relation holds if expectations of products are replaced by joint cumulants of the factors throughout (6.20). Of course, regularity conditions are needed to assure that differentiation and integration can be interchanged.

Lemma 6.6. *Consider a statistical model $\{(E,\nu); f(y;\beta); \beta \in B \subseteq V\}$ for which there is some neighbourhood $U(\beta_0)$ of a point $\beta_0 \in \mathrm{int}(B)$, such that $f(y;\beta)$ is positive and differentiable as a function of β in $U(\beta_0)$ for all $y \in E$. let $D_1(\beta) = D_\beta \log f(y;\beta)$, and for each $j = 1,\ldots,m$, let $g_j : E \times B \to \mathbf{R}$ be mappings such that the mapping*

$$y \mapsto g_j(y;\beta)$$

is measurable for any $\beta \in U(\beta_0)$, and the mapping

$$\beta \mapsto g_j(y;\beta)$$

is differentiable in $\beta \in U(\beta_0)$ for any $y \in E$. Furthermore, assume that there exist functions $M_0, M_1, \ldots, M_m$ from E to $\mathbf{R}$ such that each of the inequalities

$$\|D_\beta g_j(y;\beta)\| \leq M_j(y), \qquad j = 1,\ldots,m, \tag{6.21}$$

and

$$\|D_\beta f(y;\beta)\| / f(y;\beta_0) \leq M_0(y) \tag{6.22}$$

hold for all $\beta \in U(\beta_0)$ and $y \in E$, and such that

$$\mathrm{E}_{\beta_0}\left\{(1+M_0(Y))\prod_{j\in J}\left[|g_j(Y;\beta_0)|+M_j(Y)\right]\right\}<\infty \tag{6.23}$$

for any subset J of $\{1,\ldots,m\}$, including the empty set for which the product should be read as 1.

Then for any $v \in V$, we have

$$\begin{aligned}&\{D_\beta \mathrm{cum}_\beta(g_1(Y;\beta),\ldots,g_m(Y;\beta))\}(v)\\ &\quad=\mathrm{cum}_{\beta_0}\{D_1(\beta_0)(v),g_1(Y;\beta_0),\ldots,g_m(Y;\beta_0)\}\\ &\qquad+\sum_{i=1}^{m}\mathrm{cum}_{\beta_0}\{g_1(Y;\beta_0),\ldots,D_\beta g_i(Y;\beta_0)(v),\ldots,g_m(Y;\beta_0)\},\end{aligned} \tag{6.24}$$

where the differential of the function on the left is evaluated at $\beta=\beta_0$.

Proof. We shall give a somewhat superficial proof of this lemma since its full proof is almost identical to any of the two proofs given in Skovgaard (1986a). The conditions given here are sufficient to assure that the expectation of the product of any subset of the g_j's can be differentiated by differentiation under the integral sign as shown in (6.19) and (6.20). Since the cumulant on the left in (6.24) is a polynomial in these expectations, we may differentiate this cumulant by differentiation of the polynomial, i.e., by differentiation of moments of the g_j's. Therefore the conditions in the lemma are sufficient to validate the formal calculations below. Alternatively, we might have used the analogue of the proof in Section 3 in Skovgaard (1986a) in which the result is shown directly by differentiation of the moments.

Let $h=(h_1,\ldots,h_m)$ and consider the moment generating function

$$\phi(h;\beta)=\int\exp\left\{\sum_{j=1}^{m}h_j g_j(Y;\beta)\right\}f(y;\beta)\,d\nu(y)$$

of the g_j's. Then, since the cumulants of the g_j's are the derivatives of $\log\phi(h;\beta)$ we obtain the following relations in which all derivatives are to be evaluated at $h_1=\cdots=h_m=0$ and $\beta=\beta_0$,

$$\begin{aligned}&\{D_\beta \mathrm{cum}_\beta(g_1(Y;\beta),\ldots,g_m(Y;\beta))\}(v)\\ &\quad=\left\{D_\beta\frac{\partial}{\partial h_1}\cdots\frac{\partial}{\partial h_m}\log\phi(h;\beta)\right\}(v)\\ &\quad=\frac{\partial}{\partial h_1}\cdots\frac{\partial}{\partial h_m}\int\left\{D_1(\beta_0)(v)+\sum h_j D_\beta g_j(y;\beta_0)(v)\right\}\\ &\qquad\qquad\times\exp\left\{\sum h_j g_j(y;\beta)\right\}f(y;\beta_0)\,d\nu(y)\Big/\phi(h;\beta_0)\end{aligned}$$

$$= \frac{\partial}{\partial h_1} \cdots \frac{\partial}{\partial h_m} \frac{\partial}{\partial h_0} \log \int \exp \left\{ h_0 D_1(\beta_0)(v) + \sum h_j g_j(y;\beta_0) \right\} f(y;\beta_0)\, d\nu(y)$$

$$+ \sum_{i=1}^{m} \frac{\partial}{\partial h_1} \cdots \frac{\partial}{\partial h_{i-1}} \frac{\partial}{\partial h_{i+1}} \cdots \frac{\partial}{\partial h_m}$$

$$\int \{D_\beta g_i(y;\beta_0)(v)\} \exp \left\{ \sum_{j=1}^{m} h_j g_j(y;\beta_0) \right\} f(y;\beta_0)\, d\nu(y) \Big/ \phi(h;\beta_0)$$

$$= \operatorname{cum}_{\beta_0} \{D_1(\beta_0)(v), g_1(Y;\beta_0), \ldots, g_m(Y;\beta_0)\}$$

$$+ \sum_{i=1}^{m} \frac{\partial}{\partial h_1} \cdots \frac{\partial}{\partial h_m}$$

$$\log \int \exp \left\{ [h_i D_\beta g_i(y;\beta_0)(v)] + \sum_{j \neq i} h_j g_j(y;\beta_0) \right\} f(y;\beta_0)\, d\nu(y),$$

from which the result follows. In the third equality we have used the fact that in the jth term of the first sum on the left hand side, differentiation of the exponential or of the denominator $\phi(h;\beta_0)$ with respect to h_j results in zero when evaluated at $h_j = 0$. The proof is identical to the one in Section 2 in Skovgaard (1986a) where, unfortunately, the denominator $\phi(h;\beta)$ was missing in the expression just preceding the last equality sign above. ∎

We may now return to the proof of Lemma 6.5.

Proof of Lemma 6.5. We apply Lemma 6.6 to differentiate cumulants in the conditional distribution given $t(Y) = t$. From (6.16) we see that the density of the conditional P_β-distribution given $t(Y) = t$ with respect to P_{β_0}-distribution may be chosen as

$$\left(dP_\beta^{\cdot t} / dP_{\beta_0}^{\cdot t}\right)(y) = \{f(y;\beta)/f(y;\beta_0)\}/\tilde{f}(t;\beta)$$

Provided that differentiation and integration can be interchanged it follows from (6.16) that the differential $D^{\cdot t}(\beta)$, say, of the logarithm of the conditional density is

$$D^{\cdot t}(\beta) = D_\beta \log\{f(y;\beta)/\tilde{f}(t;\beta)\} = D_1(\beta) - \mathrm{E}_\beta\{ D_1(\beta) \mid t(Y) = t \}. \qquad (6.25)$$

To apply Lemma 6.6 with each of the functions $g_j(y;\beta)$ being of the form $D_k(\beta)(v^k)$ we need to provide bounds for these functions satisfying the conditions of the lemma. Instead of doing this we shall show directly that differentiation and integration can be interchanged for any integral of the form

$$\int D_{k_1}(\beta)(v^{k_1}) \cdots D_{k_m}(\beta)(v^{k_m}) \{f(y;\beta)/f(y;\beta_0)\}\, dP_{\beta_0}^{\cdot t}(y)$$

from which it will also follow that the representation given above for the differential of the log-density of the conditional distribution is valid. The omission of the normalizing constant $\tilde{f}(t;\beta)$ from the conditional density is of no importance. From Lemma 3.4 we know that we have bounds for the D_k's of the form

$$|D_k(\beta)(v^k)| \leq k!\,\rho^{k-1}M(y)\|v\|^k,$$

for all $k \in \mathbb{N}$ where $M(Y)$ has finite (unconditional) exponential moments and the bounds apply to a neighbourhood of β_0. Furthermore, as in the proof of Lemma 3.1 we see that also the bound

$$\log\{f(y;\beta)/f(y;\beta_0)\} \leq M(y)\|\beta-\beta_0\|\,/(1-\rho\|\beta-\beta_0\|)$$

holds in some neighbourhood of β_0. From the equality

$$\mathrm{E}_\beta \exp\{sM(Y)\} = \mathrm{E}_\beta\left[\mathrm{E}_\beta\left\{\exp\left(sM(Y)\right) \mid t(Y)\right\}\right]$$

it is seen that $M(Y)$ has exponential moments in the conditional distribution given $t(Y) = t$ for almost all t and hence that we have provided the bounds required to prove that differentiation and integration can be interchanged for any of the integrals of the type considered including the one with $m = 0$, i.e., without any of the factors $D_{k_j}(\beta)(v^{k_j})$.

From (6.25) it follows immediately that the relation (6.18) holds for $k = 1$, or, more correctly, that was seen from (6.16) to derive (6.25). Thus, we have

$$\tilde{D}(\beta)(v) = \mathrm{E}_\beta\{\, D_1(\beta) \mid t(Y) = t\,\}$$

for all β in a neighbourhood of β_0. Now we may differentiate this equation repeatedly with respect to β by use of Lemma 6.6 because the expectation is identical to the first cumulant. The next equation becomes

$$\tilde{D}_2(\beta)(v^2) = \mathrm{E}_\beta\{\, D_2(\beta)(v^2) \mid t(Y) = t\,\} + \mathrm{var}_\beta\{\, D_1(\beta)(v) \mid t(Y) = t\,\},$$

and the process continues, formally in precisely the same way as for the unconditional cumulants for which repeated differentiation leads to the well-known relations between the (unconditional) cumulants of the log-likelihood derivatives. Since these relations are identical to the ones in (6.18), except that the left hand side is zero and that the cumulants are unconditional, it follows that (6.18) holds. ∎

From the representation (6.18) of the derivatives of $\log \tilde{f}(t;\beta)$ we can now obtain bounds for these to show that the function $\tilde{f}(t;\beta)$ is analytic at β_0 for almost all t. To see this we use the condition (iv) from Definition 2.1. Let $M(y;\beta_0)$ denote the function from this condition and choose $s > 0$ such that

$$\mathrm{E}_{\beta_0}\{\exp\left(sM(Y;\beta_0)\right) \mid t(Y) = t\,\} = C_t(s) < \infty.$$

Then, for any $k_1, \dots, k_m \in \mathbb{N}$ and $v \in V$, condition (iv) in Definition 2.1 shows that

$$\begin{aligned}
&\mathrm{E}_{\beta_0}\{ D_{k_1}(\beta_0)(v^{k_1}) \cdots D_{k_m}(\beta_0)(v^{k_m}) \mid t(Y) = t \} \\
&\quad \le \left(\prod k_j!\right) \rho^{k_1+\cdots+k_m-m} \mathrm{E}\{ M(Y;\beta_0)^m \mid t(Y) = t \} \|v\|^{k_1+\cdots+k_m} \\
&\quad \le \left(\prod k_j!\right) \rho^{k_1+\cdots+k_m-m} m!\, s^{-m} C_t(s) \|v\|^{k_1+\cdots+k_m}.
\end{aligned}$$

By a fairly simple extension of Lemma 1.4.4, based on (1.4.16), it now follows that

$$\begin{aligned}
&\left|\mathrm{cum}_{\beta_0}\{ D_{k_1}(\beta_0)(v^{k_1}), \cdots, D_{k_m}(\beta_0)(v^{k_m}) \mid t(Y) = t \}\right| \\
&\quad \le (m-1)! \left(\prod k_j!\right) \|v\|^{k_1+\cdots+k_m} \gamma^{k_1+\cdots+k_m},
\end{aligned} \tag{6.26}$$

where $\gamma = (1 + C_t(s)) \max\{\rho, s^{-1}\}$. But then, from (6.18) and (1.2.27) it follows that for any $k \in \mathbb{N}$ we have

$$\begin{aligned}
&|\tilde{D}_k(\beta_0)(v^k)| \\
&\quad \le \sum_{m=1}^{k} \sum_{a \in S_m(k)} \frac{k!}{m!} \left(\prod a_j!\right) (m-1)!\, (\|v\|\gamma)^{\Sigma a_j} \\
&\quad = k!\, (\|v\|\gamma)^k \sum_{m=1}^{k} \frac{1}{m} \binom{k-1}{m-1} \\
&\quad < (k-1)!\, (2\|v\|\gamma)^k,
\end{aligned} \tag{6.27}$$

which proves that the Taylor series expansion of $\log \tilde{f}(t;\beta)$ around β_0 is bounded by a geometric series. We shall not go through a detailed proof that $\tilde{f}(t;\beta)$ agrees with its Taylor series expansion in some neighbourhood of β_0 and hence is analytic. It is a simple consequence of the fact, derived from Lemma 3.4, that the bounds in (6.27) hold uniformly in such a neighbourhood if ρ and M are chosen appropriately.

Thus $\tilde{f}(y;\beta)$ is analytic at β_0 and it is then easy to see that the conditions (i)–(iii) in Definition 2.1 hold for the model $\{\tilde{f}(t;\beta); \beta \in B \subseteq V\}$. An investigation of whether (iv) holds in general requires an investigation of the cumulants of the $\tilde{D}_k$'s from (6.18). For apparent reasons we shall not attempt such an investigation.

Incidentally, the considerations above are sufficient to show that the models for the conditional distributions given $t(Y) = t$ are analytic because the conditional P_β-density of Y given $t(Y) = t$ may be written

$$f(y;\beta)/\{f(y;\beta_0)\tilde{f}(t;\beta)\} \tag{6.28}$$

with respect to $P^{\cdot t}_{\beta_0}$. Hence the kth differential of the conditional log-likelihood is

$$D_k(\beta) - \tilde{D}_k(\beta)$$

which deviates from $D_k(\beta)$ only by a non-random analytic function. The verification of (i)–(iv) in Definition 2.1 is then straightforward for (P_{β_0}-almost) any fixed $t \in \tilde{E}$. However, as mentioned earlier, this result is not very useful in itself without some uniformity in $t \in \tilde{E}$ of any of the bounds in Theorem 4.2. To provide such uniform bounds amounts to the same problem as proving that the model for $t(Y)$ is analytic.

7 Some approximation results

Assume that the model $\{ f(y;\beta) : \beta \in B \subseteq V \}$ is analytic at $\beta_0 \in \text{int}(B)$, and consider for any fixed $K \in \mathbb{N}_0$ the approximation

$$f_K(y;\beta) = f(y;\beta_0)\exp\{D_1(\beta_0)(\beta-\beta_0) + \cdots + \frac{1}{K!}D_K(\beta_0)(\beta-\beta_0)^K\} \quad (7.1)$$

to the density $f(y;\beta)$ for β in some neighbourhood of β_0. In the case $K = 0$ the right hand side is to be read as $f(y;\beta_0)$. In this section we shall obtain bounds for the errors of various approximations based on (7.1). In particular we shall see what happens with the derivatives of the log-likelihood function when $f(y;\beta)$ is replaced by a normalized version of $f_K(y;\beta)$, and derive a bound for the total variation between the two measures corresponding to these two densities. This latter measure is closely related to measures of asymptotic sufficiency, cf. Michel (1978), and to the deficiency as defined by LeCam, cf. LeCam (1986).

All results in this section will be expressed in terms of the directional mixed cumulant condition (vi) in Theorem 4.2, i.e., in terms of a pre-given semi-norm $\|\cdot\|$ and constants $c \geq 0$ and $\lambda \geq 0$. The results are then easily reformulated in terms of the index $\lambda(\beta_0)$ at β_0 by substituting this for λ, 1 for c, and the Fisher information semi-norm for $\|\cdot\|$, throughout.

Observe that the integral of $f_K(y;\beta)$ may not be 1. Therefore we may consider the normalized version

$$\bar{f}_K(y;\beta) = f_K(y;\beta) / \int f_K(y;\beta)\, d\nu(y). \quad (7.2)$$

For any fixed K the family of measures $\bar{f}(y;\beta)$ constitutes a curved exponential family of distributions with canonical sufficient statistic $(D_1(\beta_0), \ldots, D_K(\beta_0))$ and canonical parameter

$$\left(\beta - \beta_0, \ldots, \frac{1}{K!}(\beta-\beta_0)^K\right).$$

Since it may be difficult to work out the normalization constant involved in the computation of $\bar{f}(y;\beta)$ in (7.2) we shall consider an approximate version instead. Let

$$\xi_K(\beta) = \exp\left\{\frac{1}{(K+1)!}\chi_{K+1}(\beta_0)(\beta-\beta_0)^{K+1}\right\} \quad (7.3)$$

and define

$$\tilde{f}_K(y;\beta) = f(y;\beta)\xi_K(\beta) \tag{7.4}$$

which is to be viewed as an approximation to $\bar{f}(y;\beta)$. The following lemma provides a bound for the error of that approximation.

Lemma 7.1. *Let the model $\{f(y;\beta) : \beta \in B \subseteq V\}$ be analytic at β_0 and assume that the constants $c \geq 0$ and $\lambda \geq 0$ satisfy the directional mixed cumulant condition (vi) in Theorem 4.2. Then, for any $K \in \mathbb{N}_0$ and β in the set $U_a(\beta_0)$ from (3.1) with $a < \frac{1}{2}\rho^{-1}$, we have*

$$\begin{aligned}\left|\int \tilde{f}_K(y;\beta)\,d\nu(y) - 1\right| \\ &\leq c\rho^K a^{K+1}(1-a\rho)^{-1}\exp\{(ca)^2/(1-a\rho)\}C(\delta)\left(ca + (\delta - s)^{-1}\right) \\ &= O(\lambda^K)\end{aligned} \tag{7.5}$$

as $\lambda \to 0$ with a and c bounded, where $\rho = (2e\sqrt{p})\lambda$, $s = ca/(1-a\rho) < \delta < c/\rho$ and

$$C(\delta) = \mathrm{E}\exp\{\delta H(Y)\},$$

$H(y)$ being the function from Lemma 4.1 for which it is known that $C(\delta)$ is bounded by a function depending on δ, c, p, and λ, only.

Proof. Let

$$\begin{aligned}R_K(\beta) &= \log f(y;\beta) - \log f_K(y;\beta) - \log \xi_K(\beta) \\ &= \sum_{k=K+2}^{\infty} \frac{1}{k!}\chi_k(\beta_0)(\beta-\beta_0)^k \\ &\qquad + \sum_{k=K+1}^{\infty} \frac{1}{k!}\{D_k(\beta_0) - \chi_k(\beta_0)\}(\beta-\beta_0)^k\end{aligned} \tag{7.6}$$

and notice that for $\beta \in U_a(\beta_0)$ it follows from Lemma 4.1 and Theorem 4.2 that

$$\begin{aligned}|R_K(\beta)| &\leq \sum_{k=K+2}^{\infty} c^2\rho^{k-2}a^k + \sum_{k=K+1}^{\infty} cH(y)\rho^{k-1}a^k \\ &= \{ca + H(y)\}c\rho^K a^{K+1}/(1-a\rho)\end{aligned} \tag{7.7}$$

where $\rho = (2e\sqrt{p})\lambda$ and $H(y)$ is the function from Lemma 4.1. By use of the inequality $|\exp z - 1| \leq |z|\exp|z|$ we obtain

$$\begin{aligned}\left|\int\{\tilde{f}_K(y;\beta) - 1\}\,d\nu(y)\right| &= \left|\int\{f_K(y;\beta)\xi_K(\beta) - f(y;\beta)\}\,d\nu(y)\right| \\ &\leq \int f_K(y;\beta)\xi_K(\beta)\,|1 - \exp R_K(\beta)|\,d\nu(y)\end{aligned}$$

$$\leq \int |R_K(\beta)| \exp\left\{\sum_{k=1}^{\infty} \frac{1}{k!} \left|D_k(\beta_0)(\beta-\beta_0)^k\right|\right\} dP_{\beta_0}(y)$$
$$\leq \mathrm{E}_{\beta_0}\left\{|R_K(\beta)| \exp\left([(ca)^2 + caH(Y)]/(1-a\rho)\right)\right\}.$$

The result (7.5) now follows by insertion of (7.7) and use of the inequality

$$\mathrm{E}\{H(Y)\exp(sH(Y)\} \leq \mathrm{E}\exp\{\delta H(Y)\}/(\delta - s) = C(\delta)/(\delta - s)$$

for $s < \delta$. ∎

For a sequence of statistical models we will typically be concerned primarily with β's that, in terms of the Fisher information semi-norm, are within a fixed (or slowly increasing) distance from β_0, such that the bound in (7.5) will be $O(\lambda(\beta_0)^K)$ where $\lambda(\beta_0)$ is the index at β_0. This type of asymptotics where the constant a is fixed and λ varies, may be worth having in mind throughout this section. In the case of n independent replications $\lambda(\beta_0)$ will be proportional to $1/\sqrt{n}$, as shown in Theorem 5.2.

The point of Lemma 7.1 is to show that $\xi_K(\beta)$ may be used as an approximate normalizing constant when it is difficult to obtain the exact constant. The result shows that in the type of asymptotics mentioned above the exactly normalized and the approximately normalized densities will deviate by a factor of order $1 + O\{\lambda(\beta_0)^K\}$, or in the case of n independent replications by $1 + O(n^{-K/2})$. This is the typical order of magnitude obtained in approximation results based on the truncated densities $\tilde{f}_K(y;\beta)$.

For the proof of Theorem 7.3 below we need to be able to estimate sums of the form given in the following lemma.

Lemma 7.2. *For any $a < 1$, $s \in \mathbb{N}_0$ and $m \in \mathbb{N}_0$, we have*

$$\sum_{k=s}^{\infty} \frac{(k+m)!}{k!} a^k < a^s \left\{(s+m)^m + 2^m\sqrt{a}\left[\frac{(m+s+1)^m}{-\log a} + \frac{m!}{(-\log a)^{m+1}}\right]\right\}$$
$$= O(a^s) \tag{7.8}$$

as $a \to 0$.

Proof. The convexity of the decreasing function $x \to a^x$ shows that

$$a^k \leq \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} a^x \, dx$$

for any $k \in \mathbb{N}$. By use of this inequality we obtain

$$\sum_{k=s}^{\infty} \frac{(k+m)!}{k!} a^k \leq \sum_{k=s}^{\infty} (k+m)^m a^k$$

$$\leq (s+m)^m a^s + \int_{s+1}^{\infty} (x+m)^m a^{x-\frac{1}{2}}\, dx$$
$$\leq (s+m)^m a^s + a^{s+\frac{1}{2}} \int_0^{\infty} (s+m+1+u)^m a^u\, du$$
$$\leq (s+m)^m a^s + a^{s+\frac{1}{2}} \int_0^{\infty} \{[2(s+m+1)]^m + (2u)^m\} a^u\, du$$

from which the result follows by integration. ∎

We are now in position to establish the degree of approximation to the log-likelihood function and its derivatives, involved in the replacement of $f(y;\beta)$ by $\tilde{f}(y;\beta)$.

Theorem 7.3. *Let the model $\{ f(y;\beta) : \beta \in B \subseteq V\}$ be analytic at β_0 and assume that the constants $c \geq 0$ and $\lambda \geq 0$ satisfy the directional mixed cumulant condition (vi) in Theorem 4.2. Then for any $m \in \mathbb{N}_0$, $K \in \mathbb{N}_0$ and β in the set $U_a(\beta_0)$ from (3.1) with $a < \rho^{-1}$, we have*

$$\begin{aligned} &\|D^m \log f(y;\beta) - D^m \log \tilde{f}_K(y;\beta)\| \\ &\quad \leq \begin{cases} c\rho^K a^{K+1-m}\{ca + H(y)\}\gamma(a\rho, K, m), & \text{if } m \leq K, \\ c\rho^K \left\{m!\, H(y)(1-a\rho)^{-m-1} + ca\gamma(a\rho, K, m)\right\}, & \text{if } m = K+1, \\ cm!\, \rho^{m-2}(1-a\rho)^{-m-1}\{c + \rho H(y)\}, & \text{if } m > K+1 \end{cases} \\ &\quad = O(\lambda^K) \end{aligned} \tag{7.9}$$

as $\lambda \to 0$, where

$$\gamma(a\rho, K, m) = \\ (K+2)^m + 2^m (a\rho)^{1/2}\left\{(K+3)^m\{-\log(a\rho)\}^{-1} + m!\{-\log(a\rho)\}^{-m-1}\right\},$$

$\rho = (2e\sqrt{p})\lambda$, $H(y)$ is the function from Lemma 4.1, and the bound should be observed to depend on the model only through the quantities involved, i.e., through c, a, λ, m, K, and the function H.

Proof. Notice first that

$$\begin{aligned} &\log f(y;\beta) - \log \tilde{f}(y;\beta) \\ &\quad = \sum_{k=K+2}^{\infty} \frac{1}{k!}\chi_k(\beta_0)(\beta-\beta_0)^k + \sum_{k=K+1}^{\infty} \frac{1}{k!}\{D_k(\beta_0) - \chi_k(\beta_0)\}(\beta-\beta_0)^k. \end{aligned}$$

We differentiate this equation m times, use the bound from Lemma 4.1 for $D_k(\beta_0) - \chi_k(\beta_0)$ and the bound from Theorem 4.2 for $\chi_k(\beta_0)$, and obtain

$$\begin{aligned} &\|D^m \log f(y;\beta) - D^m \log \tilde{f}(y;\beta)\| \\ &\quad \leq \sum_{k=\max\{m,K+2\}}^{\infty} \frac{k!}{(k-m)!} c^2 (2\lambda)^{k-2} a^{k-m} \end{aligned}$$

$$+ \sum_{k=\max\{m,K+1\}}^{\infty} \frac{k!}{(k-m)!} cH(y)\rho^{k-1} a^{k-m}$$

$$\leq c^2\rho^K \sum_{k=\max\{0,K+2-m\}}^{\infty} \frac{(k+m)!}{k!} \rho^{k-(K+2-m)} a^k$$

$$+ cH(y)\rho^K \sum_{k=\max\{0,K+1-m\}}^{\infty} \frac{(k+m)!}{k!} \rho^{k-(K+1-m)} a^k,$$

where $\rho = (2e\sqrt{p})\lambda$ and $H(y)$ is the function from Lemma 4.1. When the first sum in the last expression ranges from zero it equals $m!\rho^{m-K-2}/(1-a\rho)^{m+1}$, otherwise Lemma 7.2 is used to obtain a bound for the sum. The second sum is treated similarly, and after a few trivial majorizations the bounds in (7.9) are obtained. ∎

The consequences of Theorem 7.3 is that statistical quantities that depend on the likelihood function, such as maximum likelihood or Bayes estimators, can usually be shown to deviate by $O(\lambda^K)$ only, when calculated from $\tilde{f}_K(y;\beta)$ instead as from $f(y;\beta)$. This does not, however, imply that their distributions, based on $\tilde{f}_K(y;\beta)$ or $f(y;\beta)$, are not far apart. For the purpose of such distributional approximations we need other types of bounds, such as the one derived in the following theorem which in a certain sense generalizes the result of Lemma 7.1.

Theorem 7.4. *Consider the approximation $\tilde{f}_K(y;\beta)$ from (7.4) to a model $\{f(y;\beta) : \beta \in B \subseteq V\}$ that is analytic at β_0. Suppose that the constants $c \geq 0$ and $\lambda \geq 0$ satisfy the directional mixed cumulant condition (vi) in Theorem 4.2, and consider any measurable function $g : E \to \mathbf{R}$. Then, for $\beta \in U_a(\beta_0)$ we have*

$$\begin{aligned}
&\left| \int_A g(y) f(y;\beta)\, d\nu(y) - \int_A g(y)\tilde{f}_K(y;\beta)\, d\nu(y) \right| \\
&\quad \leq \int_A |g(y)|\, |f(y;\beta) - \tilde{f}_K(y;\beta)|\, d\nu(y) \\
&\quad \leq \left\{ \int_A |g(y)|^{1/\alpha}\, dP_\beta(y) \right\}^\alpha 2c\rho^K a^{K+1}(1-a\rho)^{-1} \\
&\qquad \times \exp\left\{ (ca)^2 \left(1 + (a\rho)^K\right)(1-a\rho)^{-1} \right\} \\
&\qquad \times \{\mathrm{E}\exp[\delta H(Y)]\}^{1-\alpha} \left\{ ca + (\delta - s)^{-1}\{(1-\alpha)^{-1}\}!^{1-\alpha} \right\} \\
&\quad = \left\{ \int_A |g(y)|^{1/\alpha}\, dP_\beta(y) \right\}^\alpha O(\lambda^K)
\end{aligned} \tag{7.10}$$

as $\lambda \to 0$ with other constants being held fixed, where $0 \leq \alpha < 1$, $\rho = (2e\sqrt{p})\lambda$,

$$s = \{1 + (a\rho)^K/(1-\alpha)\}ca/(1-a\rho) < \delta < c/\rho,$$

and, by convention,

$$\left\{ \int_A |g(y)|^{1/\alpha}\, dP_\beta(y) \right\}^\alpha$$

equals $\sup\{\,|g(y)| : y \in A\}$ *for* $\alpha = 0$.

Proof. The first inequality in (7.10) is trivial and by use of Hölder's inequality we get

$$\int_A |g(y)|\,|f(y;\beta) - \tilde{f}_K(y;\beta)|\,d\nu(y)$$
$$\leq \left\{\int_A |g(y)|^{1/\alpha}\,dP_\beta(y)\right\}^\alpha \left\{\int_A |1 - \exp\{-R_K(\beta)\}|^{1/(1-\alpha)}\,dP_\beta(y)\right\}^{1-\alpha},$$

where $R_K(\beta)$ is defined in (7.6). For the second integral on the right we use the bound from (7.7) for $R_K(\beta)$ to obtain

$$\left\{\int_A |1 - \exp\{-R_K(\beta)\}|^{1/(1-\alpha)} \exp\left(\sum_{k=1}^{\infty} \frac{1}{k!} D_k(\beta_0)(\beta - \beta_0)^k\right) dP_{\beta_0}(y)\right\}^{1-\alpha}$$
$$\leq \left\{\int_A [|R_K(\beta)| \exp|R_K(\beta)|]^{1/(1-\alpha)} \right.$$
$$\left. \times \exp\{[(ca)^2 + caH(y)]/(1 - a\rho)\}\,dP_{\beta_0}(y)\right\}^{1-\alpha}$$
$$\leq \frac{c\rho^K a^{K+1}}{1 - a\rho} \exp\left\{\frac{(ca)^2(1 - \alpha) + c^2\rho^K a^{K+2}}{1 - a\rho}\right\}$$
$$\times \left\{\int_A [ca + H(y)]^{1/(1-\alpha)} \exp\left[\left(\frac{c\rho^K a^{K+1}}{1 - \alpha} + \frac{ca}{1 - a\rho}\right) H(y)\right] dP_{\beta_0}(y)\right\}^{1-\alpha}$$
$$\leq \frac{c\rho^K a^{K+1}}{1 - a\rho} \exp\left\{\frac{(ca)^2\left(1 + (a\rho)^K\right)}{1 - a\rho}\right\}$$
$$\times \left\{\int_A \left[(2ca)^{1/(1-\alpha)} + \{2H(y)\}^{1/(1-\alpha)}\right] \exp\{sH(y)\}\,dP_{\beta_0}(y)\right\}^{1-\alpha}.$$

The second inequality in (7.10) now follows by an application of the inequality

$$H(y)^r \leq r!\,\exp\{(\delta - s)H(y)\}/(\delta - s)^r, \qquad r \geq 1, \tag{7.11}$$

with $r = 1/(1 - \alpha)$. The result in (7.11), which is trivial when $r \in \mathbf{N}$, may be proved by use of Stirling's formula in the general case. ∎

As a special case, with $g(y) = 1$ for $y \in E$, Theorem 7.4 provides a bound for the total variation of the difference between the two measures with densities $f(y;\beta)$ and $\tilde{f}(y;\beta)$. Since the statistic $(D_1(\beta_0), \ldots, D_K(\beta_0))$ is sufficient for the family of measures with densities $\tilde{f}_K(y;\beta)$, or rather for the exactly normalized measures $\bar{f}_K(y;\beta)$ from (7.2), the result may be interpreted as a result of approximate sufficiency. For a sequence of models with indices $\lambda_n(\beta_0)$, say, we see that the statistic $(D_1(\beta_0), \ldots, D_K(\beta_0))$ is an asymptotic higher-order (local) sufficient statistic with error term $O(\lambda_n(\beta_0)^K)$ as $\lambda_n(\beta_0) \to 0$, in terms of total variation. This result relates closely to Michel (1978) where an error of order $o(n^{-(K-2)/2})$ was obtained for the case of n independent replications when $D_1(\beta_0)$ was replaced by a sequence of asymptotic maximum likelihood estimators.

Theorem 7.4 has been formulated more generally to cover also unbounded functions g, e.g., polynomial functions in the $D_k(\beta_0)$'s that arise frequently as stochastic expansions of estimators.

It is seen from Theorem 7.4 that for the measure of a set A, for which we take g as the indicator function of A, we obtain a 'relative' error of order $O(\lambda^K)$, relative to $P_\beta(A)^\alpha$ for any $\alpha < 1$, but $\alpha = 1$ is not included.

8 The generated infinite-dimensional exponential family

In this section we shall explore some properties of the exponential family of distributions generated by the log-likelihood differentials at a certain parameter value $\beta_0 \in \text{int}(B)$ from the model $\{f(y;\beta) : \beta \in B \subseteq V\}$. Throughout the section the model is assumed to be analytic at the fixed point β_0. Around β_0 we then have the expansion

$$f(y;\beta) = f(y;\beta_0)\exp\{D_1(\beta_0)(\beta-\beta_0) + \frac{1}{2}D_2(\beta_0)(\beta-\beta_0)^2 + \cdots\} \tag{8.1}$$

as noted in the introduction to the present chapter. Because the model is assumed to be analytic at β_0 it is known that this expansion is valid and absolutely convergent in some neighbourhood of β_0. The kth term in the exponent may be considered as a linear function of the statistic $D_k(\beta_0)$. This leads us to consider the extended family of densities of the form

$$\tilde{f}(y;\boldsymbol{\theta}) = f(y;\beta_0)a(\boldsymbol{\theta})\exp\{\theta_1[D_1(\beta_0)] + \theta_2[D_2(\beta_0)] + \cdots\}, \tag{8.2}$$

where

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots), \qquad \theta_k \in \text{Lin}(\text{Sym}_k(V;\mathbf{R});\mathbf{R}), \tag{8.3}$$

such that θ_k maps $D_k(\beta_0)$ linearly into $\mathbf{R}$ and $a(\boldsymbol{\theta})$ is a normalizing constant. Convergence and integrability must be assumed for the definition in (8.2) to make sense. Obviously (8.1) is a special case of (8.2) with $\theta_k = (\beta-\beta_0)^k/k!$ viewed as a linear mapping on the space of $D_k(\beta_0)$. The family of measures with densities of the form (8.2) has the appearance of an exponential family except that the sum in the exponent is infinite. Thus, the original model, or rather the model in a neighbourhood of β_0, may be viewed as a curved sub-model of the infinite-dimensional exponential family (8.2), with

$$\boldsymbol{\theta}(\beta) = \left((\beta-\beta_0), \frac{1}{2}(\beta-\beta_0)^2, \ldots\right) \tag{8.4}$$

representing the canonical parameter as a function of the parameter β. In special cases the 'canonical sufficient statistic' $(D_1(\beta_0), D_2(\beta_0), \ldots)$ may be concentrated on a finite-dimensional vector space, in which case (8.2) may be represented as a finite-dimensional exponential family. Exponential families of infinite dimension have been considered by, i.a., Soler (1977) and Johansen (1977), but in more general settings. The type of families considered here are of a special structure that allow us to show more specific results in certain connections.

The study of these generated exponential families is not carried far here and there may well be prospects for further developments. For asymptotic analyses in relation to sequences of statistical models it may not make much difference for the results whether calculations are done on the infinite-dimensional exponential family as an entity, or a truncation is applied before further calculations are carried out, as in Section 7, but once the theory for the infinite-dimensional case is developed it may turn out to clarify calculations and results.

Notice that it is not in any ways attempted to restrict the original models considered by some requirement that (8.1) should converge for all β. This model can be any analytic model which can then, in a neighbourhood of some fixed point only, be embedded into a model of the form (8.2) spanned by the original model. This approach corresponds closely to the patching of local representations of a differentiable manifold, and an analytic model could be viewed as an analytic manifold with local embeddings into infinite-dimensional exponential families. Since we are going to study the vicinity of one fixed point only, we shall not develop that approach.

To define the family (8.2) more precisely consider the space

$$\mathcal{D}_\infty = \mathop{\times}_{k=1}^{\infty} \mathrm{Sym}_k(V;\mathbf{R}) \tag{8.5}$$

in which the sufficient statistic $(D_1(\beta_0), D_2(\beta_0), \ldots)$ takes values, and the space

$$\Theta_\infty = \mathop{\times}_{k=1}^{\infty} \mathrm{Lin}(\mathrm{Sym}_k(V;\mathbf{R});\mathbf{R}) \tag{8.6}$$

of sequences $(\theta_1, \theta_2, \ldots)$ of linear forms on the spaces of the D_k's. Instead of defining the exponential family as in (8.2) we modify the expression slightly by subtracting the mean from each of the D_k's. Thus, let

$$T_k = D_k(\beta_0) - \chi_k(\beta_0) \in \mathrm{Sym}_k(V;\mathbf{R}) \tag{8.7}$$

and define the canonical parameter space

$$\Theta = \{\, \theta \in \Theta_\infty : \mu(\theta) = \int \exp\{\theta(\mathbf{T})\}\, dP_{\beta_0}(y) < \infty \,\} \tag{8.8}$$

where

$$\theta(\mathbf{T}) = \sum_{k=1}^{\infty} \theta_k(T_k) \tag{8.9}$$

is implied to converge for P_{β_0}-almost all $y \in E$ if $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The generated exponential family is now defined as the family of probability measures on E with densities

$$\tilde{f}(y;\boldsymbol{\theta}) = \exp\{\boldsymbol{\theta}(\mathbf{T}) - \kappa(\boldsymbol{\theta})\}, \qquad \boldsymbol{\theta} \in \boldsymbol{\Theta}, \tag{8.10}$$

with respect to P_{β_0}, where

$$\kappa(\boldsymbol{\theta}) = \log \mu(\boldsymbol{\theta}). \tag{8.11}$$

The probability measure with density $\tilde{f}(y;\boldsymbol{\theta})$ is denoted $P_{\boldsymbol{\theta}}$ and for the measure P_{β_0} we use the abbreviation P_0.

Notice that we have limited the canonical parameter space to linear forms of the form (8.6) on the space of the sufficient statistic, rather than admitting arbitrary linear forms.

The subtraction of the means from the D_k's in the definition (8.10) compared to (8.2) is a technical adjustment that admits us to prove the convergence of $\boldsymbol{\theta}(\mathbf{T})$ for a larger set of $\boldsymbol{\theta}$'s, but if $\sum \theta_k\,(\chi_k(\beta_0))$ is convergent it is seen that this adjustment is absorbed into the normalizing constant, and the two definitions (8.2) and (8.10) coincide.

Lemma 8.1. *In the exponential family*

$$\{(E, P_0); \tilde{f}(y;\boldsymbol{\theta}); \boldsymbol{\theta} \in \boldsymbol{\Theta}\} \tag{8.12}$$

from (8.10), generated by the model $\{f(y;\beta) : \beta \in B \subseteq V\}$ *which is assumed to be analytic at* β_0*, any two measures are mutually absolutely continuous. The canonical statistic*

$$\mathbf{D} = (D_1(\beta_0), D_2(\beta_0), \ldots) \in \mathfrak{D}_\infty$$

is a minimal sufficient statistic for the family, i.e., it is sufficient and is a function of any sufficient statistic.

Proof. Since any measure in the family has a density with respect to P_0 that is positive and finite on a set of P_0-probability one, any such measure and P_0, and hence any two measures in the family, are mutually absolutely continuous.

The statistic $\mathbf{T}$, and hence $\mathbf{D}$, is obviously sufficient for the model (8.12). We shall show that it is minimally sufficient. Since the original model $\{f(y;\beta) : \beta \in B \subseteq V\}$, when restricted to some neighbourhood of β_0, is a sub-model of the model considered here, according to Lemma 3.3, it suffices to show that $\mathbf{D}$ is minimally sufficient for this sub-model. This follows, as noted in Section 6, from Neyman's factorization criterion, because the log-likelihood derivatives at β_0, based on the derived model for any sufficient statistic $g(Y)$, say, are identical to the log-likelihood derivatives for the original model. Hence the $D_k(\beta_0)$'s depend on Y only through $g(Y)$. ∎

We shall see later, in Lemma 8.5, that the minimal sufficient statistic is, in fact, also complete. The precise statement of this result requires an accurate specification of the σ-algebra considered and is furthermore conveniently postponed until we have examined the parameter space a little closer.

Suppose that the semi-norm $||\cdot||$ on V is given and that the two constants $c \geq 0$ and $\lambda \geq 0$ satisfy the mixed cumulant condition (v) in Theorem 4.2. Define the semi-norm

$$||\boldsymbol{\theta}||_\lambda = \sum_{k=1}^{\infty} ||\theta_k|| k!\, \lambda^{k-1}, \qquad \boldsymbol{\theta} \in \Theta_\lambda, \tag{8.13}$$

where

$$\Theta_\lambda = \{\, \boldsymbol{\theta} \in \Theta_\infty : \sum_{k=1}^{\infty} ||\theta_k|| k!\, \lambda^{k-1} < \infty \,\} \tag{8.14}$$

and $||\theta_k||$ is the semi-norm of θ_k induced by the semi-norm on V as shown in (1.1.21). It is easily verified that $||\cdot||_\lambda$ is a semi-norm on the real vector space Θ_λ where addition is defined as component-wise addition and scalar-multiplication as the multiplication of all the components, θ_k, by the scalar. If the semi-norm on V is, in fact, a norm, then so is that on Θ_λ which is then seen to be a Banach space.

Lemma 8.2. *In the setup from above any* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots) \in \Theta_\lambda$ *satisfies*

$$\sum_{k=1}^{\infty} |\theta_k(T_k)| < \infty \tag{8.15}$$

for P_0*-almost all* $y \in E$, *and*

$$\mathrm{E}_0 \exp\left\{ h \sum_{k=1}^{\infty} |\theta_k(T_k)| \right\} < 2 \exp\left\{ \frac{1}{2}(hc||\boldsymbol{\theta}||_\lambda)/(1 - h\lambda||\boldsymbol{\theta}||_\lambda) \right\} \tag{8.16}$$

for all $h < (\lambda||\boldsymbol{\theta}||_\lambda)^{-1}$, *where* E_0 *refers to the* P_0*-expectation. Furthermore, for any* $\boldsymbol{\theta} \in \Theta_\lambda$, *we have*

$$|\mathrm{cum}_m\{\boldsymbol{\theta}(\mathbf{T})\}| \leq c^2(m-1)!\, \lambda^{m-2} ||\boldsymbol{\theta}||_\lambda^m \tag{8.17}$$

for any $m \geq 2$, *where the cumulant is from the* P_0*-distribution, while* $\mathrm{E}_0\{\boldsymbol{\theta}(\mathbf{T})\} = 0$.

Proof. For any $k \in \mathbf{N}$, let

$$a_k = ||\theta_k|| k!\, \lambda^{k-1},$$

such that $||\boldsymbol{\theta}||_\lambda = \sum a_k$. From the mixed cumulant condition (4.9) it follows that for any $k \in \mathbf{N}$ and $m \geq 2$, we have

$$\begin{aligned} |\mathrm{cum}_m\{\theta_k(T_k)\}| &\leq c^2(m-1)!\, k!^m \lambda^{km-2} ||\theta_k||^m \\ &\leq c^2(m-1)!\, \lambda^{m-2} a_k^m, \end{aligned}$$

while $\mathrm{E}\{\theta_k(T_k)\} = 0$, where the distribution considered is understood to be P_0 here as in the sequel. Thus, from Lemma 1.4.11 we have

$$\begin{aligned} |\log \mathrm{E}\exp\{h\theta_k(T_k)\}| &\le \sum_{m=2}^{\infty} \frac{c^2}{m}\lambda^{m-2}(|h|\, a_k)^m \\ &\le \frac{1}{2}(c\,|h|\, a_k)^2/(1-|h|\, a_k\lambda) \end{aligned}$$

for any $|h| < (a_k\lambda)^{-1}$, and hence

$$\begin{aligned} \mathrm{E}\exp\{h\,|\theta_k(T_k)|\} &< \mathrm{E}\exp\{h\theta_k(T_k)\} + \mathrm{E}\exp\{-h\theta_k(T_k)\} \\ &\le 2\exp\{\frac{1}{2}(cha_k)^2/(1-ha_k\lambda)\} \end{aligned}$$

for $0 \le h < (a_k\lambda)^{-1}$. Now, for each $N \in \mathbb{N}$ consider

$$M_N = \sum_{k=1}^{N} a_k$$

and use Hölder's inequality with exponents a_k/M_N for $k = 1, \ldots, N$ to show that

$$\begin{aligned} \mathrm{E}\exp\left\{h\sum_{k=1}^{N}|\theta_k(T_k)|\right\} &\le \prod_{k=1}^{N}(\mathrm{E}\exp\{(hM_N/a_k)\,|\theta_k(T_k)|\})^{a_k/M_N} \\ &< \prod_{k=1}^{N}\left(2\exp\{\frac{1}{2}(chM_N)^2/(1-hM_N\lambda)\}\right)^{a_k/M_N} \\ &= 2\exp\{\frac{1}{2}(chM_N)^2/(1-hM_N\lambda)\} \\ &\le 2\exp\{\frac{1}{2}(ch\|\,\theta\,\|_\lambda)^2/(1-h\|\,\theta\,\|_\lambda\lambda)\} \end{aligned}$$

for $0 < h < (\lambda\|\,\theta\,\|_\lambda)^{-1}$. It now follows from the theorem of monotone convergence that the last bound also holds for the expectation of the limit $\exp\{h\sum|\theta_k(T_k)|\}$, thus proving (8.16). Since this expectation is finite for some positive h it follows that the sum (8.15) is finite almost surely.

For the cumulants of $\boldsymbol{\theta}(\mathbf{T})$ we use the mixed cumulant condition (4.9) to show that

$$\begin{aligned} &|\mathrm{cum}_m\{\boldsymbol{\theta}(\mathbf{T})\}| \\ &= \left|\sum_{k_1=1}^{\infty}\cdots\sum_{k_m=1}^{\infty}\mathrm{cum}\,\{\theta_{k_1}(T_{k_1}),\ldots,\theta_{k_m}(T_{k_m})\}\right| \end{aligned}$$

$$\leq \sum_{k_1=1}^{\infty} \cdots \sum_{k_m=1}^{\infty} c^2(m-1)!\,\lambda^{m-2} \prod_{j=1}^{m} \{k_j!\,\lambda^{k_j-1}\|\theta_{k_j}\|\}$$

$$= c^2(m-1)!\,\lambda^{m-2} \left\{ \sum_{k=1}^{\infty} \|\theta_k\| k!\,\lambda^{k-1} \right\}^m$$

for all $m \geq 2$, as claimed in (8.17). That the expectation is zero is evident. Notice that all considerations in the proof are valid also for the cases $\lambda = 0$ or $\|\theta\|_\lambda = 0$. ∎

In the mixed cumulant condition (4.9) there is a choice of semi-norm which is, to some extent, arbitrary. It is natural to wonder how this choice affects the result concerning the existence of exponential moments in (8.16). A proportional change of semi-norm leaves the result in (8.16) unaffected if the constants c and λ in (4.9) are scaled proportionally to compensate the change of semi-norm. More precisely, if $\|v\|$ is multiplied by a certain constant factor, and c and λ are divided by the same constant, then the resulting effect on $\|\theta\|_\lambda$ is a multiplication by this constant and therefore $c\|\theta\|_\lambda$ and $\lambda\|\theta\|_\lambda$ are unaffected. However, more generally, some choices of semi-norms in (4.9) may provide more effective bounds than others and will thereby provide more effective results in Lemma 8.2 also.

The following two lemmas show that, just as for the finite-dimensional exponential families, the cumulant generating function is convex and analytic on an open set contained in its domain. The proofs of these assertions are quite similar to the ones from the finite-dimensional case, see, e.g., Barndorff-Nielsen (1978).

Lemma 8.3. *The cumulant generating function κ, defined in (8.11), is convex on Θ. If $\theta_1, \theta_2 \in \Theta$ are not identical and satisfy*

$$\kappa(\theta_1) = \kappa(\alpha\theta_1 + (1-\alpha)\theta_2) = \kappa(\theta_2)$$

for some α with $0 < \alpha < 1$, then P_θ is constant on the line

$$\theta_h = \theta + h(\theta_2 - \theta_1), \qquad h \in \mathbf{R},$$

which is then contained in Θ for any $\theta \in \Theta$.

Proof. Let θ_1 and θ_2 be fixed points in Θ and let $0 < \alpha < 1$. Since $\theta_1(\mathbf{T})$ and $\theta_2(\mathbf{T})$ are both (almost surely) convergent series, cf. (8.8), then

$$\{\alpha\theta_1 + (1-\alpha)\theta_2\}(\mathbf{T}) = \alpha\theta_1(\mathbf{T}) + (1-\alpha)\theta_2(\mathbf{T})$$

is also convergent, and it follows from Hölder's inequality that

$$\begin{aligned}
&\mu\{\alpha\theta_1 + (1-\alpha)\theta_2\} \\
&\quad= \int \exp\{\alpha\theta_1(\mathbf{T}) + (1-\alpha)\theta_2(\mathbf{T})\}\, dP_0(y) \\
&\quad\leq \left\{\int \exp\{\theta_1(\mathbf{T})\}\, dP_0(y)\right\}^\alpha \left\{\int \exp\{\theta_2(\mathbf{T})\}\, dP_0(y)\right\}^{1-\alpha}
\end{aligned}$$

$$= \mu(\theta_1)^{\alpha}\mu(\theta_2)^{1-\alpha},$$

implying that

$$\kappa\{\alpha\theta_1 + (1-\alpha)\theta_2\} \le \alpha\kappa(\theta_1) + (1-\alpha)\kappa(\theta_2)$$

with equality if and only if $\theta_2(\mathrm{T}) - \theta_1(\mathrm{T}) = a$ for some constant a on a set of P_0-probability one. In that case we know that for any $\theta \in \Theta$ and $h \in \mathbf{R}$, we have

$$\theta_h(\mathrm{T}) = \{\theta + h(\theta_2 - \theta_1)\}(\mathrm{T}) = \theta(\mathrm{T}) + ha$$

which shows that the two probability measures P_θ and P_{θ_h} are identical. ∎

To show that the cumulant generating function is analytic we need to extend the Definition 1.3.1 of analytic functions to cover infinite-dimensional vector spaces. There are only two changes compared to the finite-dimensional case. The first change is that the differential at a given point, cf. Definition 1.2.1, is required to be a *continuous*, or equivalently a *bounded*, linear mapping. The second 'change' is that the choice of norm on an infinite-dimensional space is of importance for the concepts of differentiability and analyticity. In our case we choose the norm $\|\cdot\|_\lambda$ and show that κ is analytic with respect to this norm.

Lemma 8.4. *The cumulant generating function κ, defined in (8.11), is analytic on the subset*

$$\{\,\theta \in \Theta_\lambda : \|\theta\|_\lambda < \lambda^{-1}\,\} \tag{8.18}$$

of Θ_λ equipped with the norm $\|\cdot\|_\lambda$. Also the moment generating function μ, defined in (8.8) is analytic on this set.

Proof. It follows directly from Lemma 8.2 and Lemma 1.4.11 that for any $\theta \in \Theta_\lambda$ with $\|\theta\|_\lambda < \lambda^{-1}$ we have the infinite series expansion

$$\kappa(\theta) = \sum_{m=1}^{\infty} \frac{1}{m!}\mathrm{cum}_m\{\theta(\mathrm{T})\}$$

where the sum is absolutely convergent. It is easily verified, along the lines of the proof of Lemma 8.2, that the mapping

$$(\theta_1, \ldots, \theta_m) \mapsto \mathrm{cum}\{\theta_1(\mathrm{T}), \ldots, \theta_m(\mathrm{T})\}$$

is a continuous m-linear mapping with respect to the norm $\|\cdot\|_\lambda$. The inequality (8.17), already used to verify the series expansion, shows that these m-linear mappings satisfy inequalities of the form (1.3.1). It now follows, e.g., from Federer (1969, Section 3.1.24), that the existence of an infinite series expansion of the function κ on the set in question, with terms that are continuous m-linear functions satisfying inequalities of this form, implies that κ is analytic on the set given in (8.18). Since $\mu(\theta) = \exp\{\kappa(\theta)\}$, this function is also analytic on the same set. ∎

It is a consequence of the results above, in particular of the inequality (8.16), that the moment generating function μ may be differentiated infinitely often at any point $\theta \in \Theta_\lambda$ with $\|\theta\|_\lambda < \lambda^{-1}$ by differentiation of

$$\int \exp\{\theta(\mathbf{T})\}\, dP_0(y)$$

under the integral sign. The proof of this assertion follows the line of proof from the finite-dimensional case.

Our next result states that the statistic $\mathbf{T}$, or equivalently $\mathbf{D}$, is a complete sufficient statistic in the generated exponential family. Recall the definition that a sufficient statistic $t(Y)$, say, is *complete* if, for any measurable real function $g\{t(Y)\}$ the identity

$$\mathrm{E}_P g\{t(Y)\} = 0 \tag{8.19}$$

for all measures P in the model considered, implies that

$$g\{t(Y)\} = 0 \tag{8.20}$$

almost surely, i.e., on a set of P-probability one for all P in the model. In our case, when considering the statistic $t(Y) = \mathbf{T} \in \mathcal{D}_\infty$, cf. (8.5), we need to define the σ-algebra on $\mathcal{D}_\infty$ with respect to which the function g considered in (8.19) is required to be measurable. The natural choice might seem to be the σ-algebra that contains all sets of the form $(g \circ t)^{-1}(A)$, where A is a Borel subset of $\mathbf{R}$. Since all our considerations relating to the sequence $\mathbf{T}$ are, however, based on limits of functions of the finite-dimensional projections $(T_1, \ldots, T_N)$ we need to restrict attention to the Borel-σ-algebra generated by these finite-dimensional projections. Thus, let σ_N denote the smallest σ-algebra on $\mathcal{D}_\infty$ containing any set of the form $\{(T_1, \ldots, T_N) \in A\}$ where A is a Borel subset of $\mathbf{R}^N$. Any such set A is called a cylindrical subset of $\mathcal{D}_\infty$, and it follows from Definition 2.1 that $\mathbf{T}(y)$ is measurable with respect to this σ-algebra. The Borel product σ-algebra, or just the Borel σ-algebra, on $\mathcal{D}_\infty$ is now defined as the smallest σ-algebra containing σ_N for all $N \in \mathbf{N}$.

Lemma 8.5. *Consider the generated exponential family (8.11), possibly restricted to any parameter space $\Theta_0 \subseteq \Theta$, say, which contains a neighbourhood with respect to the $\|\cdot\|_\lambda$-norm of a point $\theta_0 \in \Theta_\lambda$ with $\|\theta_0\|_\lambda < \lambda^{-1}$. For any function $g : \mathcal{D}_\infty \to \mathbf{R}$ which is measurable with respect to the Borel product σ-algebra on $\mathcal{D}_\infty$ and the Borel σ-algebra on $\mathbf{R}$, the relations*

$$\mathrm{E}_\theta\, |g(\mathbf{T})| < \infty, \quad \text{and} \quad \mathrm{E}_\theta\, g(\mathbf{T}) = 0 \tag{8.21}$$

for all $\theta \in \Theta_0$ imply that $g(\mathbf{T}) = 0$, P_0-almost surely. Thus, $\mathbf{T}$ is a complete sufficient statistic in any such sub-family of the generated exponential family.

Proof. The proof is essentially the same as in the finite-dimensional case, cf., e.g., Johansen (1979, Theorem 2.6). Consider a function g satisfying (8.21). Define

$$g^+(y) = \max\{0, g(\mathbf{T})\}, \qquad g^-(y) = \max\{0, -g(\mathbf{T})\}.$$

Then (8.21) implies that

$$\int g^{+}(y)\, dP_{\theta}(y) = \int g^{-}(y)\, dP_{\theta}(y)$$

for all $\theta \in \Theta_0$ and hence that

$$\int \exp\{(\theta - \theta_0)(\mathbf{T})\} g^{+}(y)\, dP_{\theta_0}(y) = \int \exp\{(\theta - \theta_0)(\mathbf{T})\} g^{-}(y)\, dP_{\theta_0}(y)$$

for all $\theta \in \Theta_0$, where θ_0 is the point from the statement of the lemma. Consider the two measures with densities $g^{+}(y)$ and $g^{-}(y)$, respectively, with respect to P_{θ_0} and notice that these are defined on the product Borel σ-algebra. It follows that the two moment generating functions for any finite-dimensional projection $(T_1, \ldots, T_N)$ of $\mathbf{T}$ induced by these two measures exist and are identical on a set containing zero as an interior point. Hence the finite-dimensional projections of the two measures agree and consequently they are identical. Thus $g^{+}(y) = g^{-}(y)$ on a set of P_{θ_0}-probability one. Since P_0 and P_{θ_0} are mutually absolutely continuous, the completeness of the statistic is proved. The sufficiency is trivial. ∎