Chapter 1

Mathematical prerequisites

This chapter contains some basic material from mathematics and probability theory. To some extent the purpose is to provide the notation, in particular in relation to multilinear forms, differentials, moments and cumulants. Another purpose is to list some results to be used later on. Most of these results are well known, but a few are more advanced, especially those in Section 5. Proofs are omitted for most of the results, and references are given only to less familiar results. In some cases proofs are given even for well-known results, mainly because the technique used is in line with the theory developed in later chapters and thereby gives some useful background knowledge.

Notations and concepts used in subsequent chapters are described in Sections 1–4. Sections 1 and 3 consist entirely of such notations although the induced norms described in the paragraphs containing equations (1.19) to (1.25) are rarely used and may be skipped at first reading. In Section 2 the notations are described in the first part, ending at expression (2.12), the remaining part of the section contains results that may be looked up in connection with their applications. Similarly, in Section 4 the notation part ends with Definition 4.8. The rest of the section and the lemmas 4.4 and 4.5 are used for reference purposes only. Section 5 contains some results that, although they are essential for the theory developed, are of a technical nature and may be consulted when needed.

1 Multilinear mappings between vector spaces

In this section we introduce some notations for multilinear mappings from one finite-dimensional vector space to another, together with some basic definitions and results.

Throughout this section let $V, W, V_1, V_2, \ldots$ denote finite-dimensional real vector spaces. Occasionally, in later sections, we shall be dealing with complex vector spaces, but the modifications required for these cases are quite trivial and will not be commented upon.

Definition 1.1. *A mapping $A : V_1 \times \cdots \times V_k \to W$ is said to be k-linear if, for any $j = 1, \ldots, k$, the mapping*

$$v_j \mapsto A(v_1, \ldots, v_j, \ldots, v_k), \qquad v_j \in V, \tag{1.1}$$

is linear in v_j for any fixed set of v_i's, $i \neq j$. The space of k-linear mappings of $V_1 \times \cdots \times V_k$ into W is denoted $\mathrm{Lin}(V_1, \ldots, V_k; W)$.

Notice that the space of k-linear mappings between finite-dimensional real vector spaces is also a real vector space of finite dimension. For the special case $V_1 = \cdots =$

V_k we use the abbreviation $\mathrm{Lin}_k(V;W)$ to denote the space of k-linear mappings of V^k into W.

Definition 1.2. *A k-linear mapping $A \in \mathrm{Lin}_k(V;W)$ is said to be symmetric if, for all $v_1,\ldots,v_k \in V$ and any permutation σ on the set $\{1,\ldots,k\}$,*

$$A(v_1,\ldots,v_k) = A\left(v_{\sigma(1)},\ldots,v_{\sigma(k)}\right). \tag{1.2}$$

The space of symmetric k-linear mappings of V^k into W is denoted $\mathrm{Sym}_k(V;W)$.

In the case $W = \mathbf{R}$ we speak of k-linear forms and of k-linear symmetric forms. The space $V^* = \mathrm{Lin}(V;\mathbf{R})$ denotes the *dual* space of V, and it is a vector space of the same dimension as V.

We generally use superscripts on vectors to denote repetitions, i.e.,

$$v^k = (v,\ldots,v) \in V^k \tag{1.3}$$

for $v \in V, k \in \mathbf{N}$. Thus, in particular, if $A \in \mathrm{Lin}_k(V;W)$,

$$A(v^k) = A(v,\ldots,v) \in W. \tag{1.4}$$

In coordinates the objects above may be written as follows. Let the vector spaces be equipped with bases and use square brackets to denote coordinates, i.e., $[v_1]_i$ denotes the ith coordinate of $v_1 \in V_1$, etc. Then a k-linear mapping $A \in \mathrm{Lin}_k(V;W)$ has a coordinate representation of the form

$$\begin{aligned}[A(v_1,\ldots,v_k)]_j &= \sum_{i_1}\cdots\sum_{i_k}[A]_j^{i_1\cdots i_k}[v_1]_{i_1}\cdots[v_k]_{i_k} \\ &= [A]_j^{i_1\cdots i_k}[v_1]_{i_1}\cdots[v_k]_{i_k}\end{aligned} \tag{1.5}$$

which expresses the jth coordinate of $A(v_1,\ldots,v_k) \in W$, and we have used some generally adopted tensor notations, cf. McCullagh (1987), namely that indices may appear as superscripts as well as subscripts and that summation over any index that appears twice in a term is understood. In the case $W = \mathbf{R}$ the coordinates of A make a k-dimensional array of numbers. If $k = 2$ it reduces to the familiar matrix representation of a bilinear form.

A key feature of symmetric k-linear forms is that they are determined by their values on the 'diagonal', i.e., by the values of the form in (1.4). Explicitly, this follows from the *polarization identity*, cf. Federer (1969, Section 1.9.3),

$$A(v_1,\ldots,v_k) = \left(k!\,2^k\right)^{-1}\sum_{\alpha}\left(\prod\alpha_j\right)A\left\{\left(\sum\alpha_j v_j\right)^k\right\}, \tag{1.6}$$

for any $A \in \mathrm{Sym}_k(V;W), v_1,\ldots,v_k \in V$, where the first sum is over all sequences $(\alpha_1,\ldots,\alpha_k) \in \{-1,1\}^k$. Notice that we use the word 'diagonal' in a different sense than a matrix diagonal. The relation does not imply that, e.g., a bilinear symmetric form is determined by the diagonal values in a particular matrix representation.

Any function, $f : V \mapsto W$, of the form

$$f(v) = w_0 + A_1(v) + A_2(v^2) + \cdots + A_k(v^k), \tag{1.7}$$

where $w_0 \in W$ and $A_j \in \mathrm{Sym}_j(V;W); j = 1,\ldots,k$, is called a *polynomial* on V with values in W and the A_j's are called the coefficients. The largest j for which A_j is non-vanishing is called the *degree* of the polynomial.

We shall frequently need mappings obtained from, e.g., $A \in \mathrm{Lin}(V_1,\ldots,V_k;W)$ by 'plugging in' some of the arguments. Thus, for fixed $v_1 \in V,\ldots,v_m \in V, m < k$, we let

$$A(v_1,\ldots,v_m) \in \mathrm{Lin}(V_{m+1},\ldots,V_k;W) \tag{1.8}$$

denote the mapping

$$(v_{m+1},\ldots,v_k) \mapsto A(v_1,\ldots,v_k) \tag{1.9}$$

although this is, strictly speaking, misuse of notation. In coordinates $A(v_1,\ldots,v_m)$ would be the $k-m+1$ dimensional array with entries

$$[A(v_1,\ldots,v_m)]_j^{i_{m+1}\cdots i_k} = [A]_j^{i_1\cdots i_k}[v_1]_{i_1}\cdots[v_m]_{i_m}. \tag{1.10}$$

In particular, any bilinear form, $A \in \mathrm{Lin}_2(V;\mathbf{R})$ may in this way be identified with a linear function of V into V^*, by the identification

$$\begin{aligned} v_1 &\mapsto A(v_1) \in \mathrm{Lin}(V;\mathbf{R}) = V^* \\ \{A(v_1)\}(v_2) &= A(v_1,v_2) \end{aligned} \tag{1.11}$$

Thus, notationally we do not distinguish between the two mappings $A \in \mathrm{Lin}_2(V;\mathbf{R})$ and $A \in \mathrm{Lin}(V;V^*)$.

A bilinear form $A \in \mathrm{Lin}_2(V;\mathbf{R})$ is said to be positive semi-definite if, for all $v \in V$,

$$A(v,v) \geq 0 \tag{1.12}$$

and positive definite if, for all $v \neq 0$,

$$A(v,v) > 0. \tag{1.13}$$

A symmetric positive definite bilinear mapping, $A \in \mathrm{Sym}_2(V;\mathbf{R})$ defines an inner product on V, denoted

$$\langle v_1,v_2\rangle_A = A(v_1,v_2), \qquad v_1,v_2 \in V, \tag{1.14}$$

or just $\langle v_1,v_2\rangle$ if A is understood. This inner product induces a norm on V,

$$\|v\|_A = (\langle v,v\rangle_A)^{1/2}, \qquad v \in V. \tag{1.15}$$

A real vector space equipped with an inner product is called a Euclidean space. If it is only assumed that the symmetric bilinear form A above is positive semi-definite then we refer to (1.14) and (1.15) as a pseudo inner product and a semi-norm, respectively. An inner product norm on V is a norm that is induced by an inner product by (1.15). Similarly we shall speak, slightly incorrectly, of an inner product semi-norm for a semi-norm induced by a pseudo inner product.

Any two norms $\|\cdot\|_1$ and $\|\cdot\|_2$, say, on a finite-dimensional real vector space V are equivalent in the sense that there exist two positive constants a and b, say, such that

$$a\|v\|_1 \le \|v\|_2 \le b\|v\|_1$$

for all $v \in V$.

To any inner product corresponds a uniquely determined Lebesgue measure, namely the one that assigns unit volume to the unit cube spanned by an orthonormal basis. The Lebesgue measure induced by $A \in \mathrm{Sym}_2(V;\mathbf{R})$ is denoted λ_A.

If $A \in \mathrm{Sym}_2(V;\mathbf{R})$ is positive definite, it has an inverse $A^{-1} \in \mathrm{Sym}_2(V^*;\mathbf{R})$ which is also positive definite and is defined as follows. Consider A as a mapping in $\mathrm{Lin}(V;V^*)$, cf. (1.11). This mapping can be shown to be one-to-one and onto V^* such that it has an inverse $A^{-1} \in \mathrm{Lin}(V^*;V)$. By the reverse construction of (1.11), A^{-1}, defined in this way, can be identified with a mapping in $\mathrm{Lin}_2(V^*;\mathbf{R})$, defined by

$$A^{-1}(v_1^*, v_2^*) = v_2^*\left(A^{-1}(v_1^*)\right) \in \mathbf{R}, \tag{1.16}$$

where $v_1^*, v_2^* \in V^*$. This mapping, $A^{-1} \in \mathrm{Lin}(V^*;\mathbf{R})$, can be shown also to be symmetric and positive definite. Hence, it induces an inner product on V^*, denoted

$$\langle v_1^*, v_2^*\rangle_A = A^{-1}(v_1^*, v_2^*) \in \mathbf{R}, \qquad v_1^*, v_2^* \in V^*, \tag{1.17}$$

although the notation with subscript A^{-1} would be correct, according to (1.14). Because A and A^{-1} are mappings in different spaces there is no risk of ambiguity by using A as subscript, which we prefer when the inner product is constructed as above, starting from A. In coordinates, A^{-1} is the usual matrix inverse of A, i.e.,

$$[A]^{ij}[A^{-1}]_{jk} = \delta^i_k, \tag{1.18}$$

where δ^i_k is the Kronecker delta that equals 1 if $i = k$, 0 otherwise.

Any vector $v \in V$ may also be considered as an element of V^{**}, the dual of V^*, by the identification $v(v^*) = v^*(v)$, for $v^* \in V^*$, and, in fact, V may be identified with V^{**}. Hence, the construction above that leads from an inner product on V to an inner product on V^* may also go the opposite way. Finally, an inner product A on V may also be used to identify V and V^*, simply by identifying $v \in V$ with $A(v) \in V^*$. In this way any orthonormal basis on V is mapped into an orthonormal basis on V^* with respect to the inner products induced by A, and the coordinates of v and $A(v)$ would be the same with respect to two such corresponding bases. The matrix of A with respect to such two bases would, of course, be the identity matrix, and any mapping defined on either of V or V^* would be identified uniquely with one defined on the other, with the same coordinate representation.

Assume now that V is equipped with an inner product $\langle\cdot,\cdot\rangle_A$ and consequently with the norm $\|\cdot\|_A$. If W is also equipped with a norm, $|\cdot|$ say, not necessarily induced by an inner product, then a norm, also denoted $\|\cdot\|_A$ (when the norm on W is understood, e.g., if $W = \mathbf{R}$), is induced on $\mathrm{Sym}_k(V;W)$ through the definition

$$\|B\|_A = \sup\left\{\left|B\left(v^k\right)\right| : v \in V, \|v\|_A \leq 1\right\}, \tag{1.19}$$

when $B \in \mathrm{Sym}_k(V;W)$. An important inequality, due to Hörmander (1954), cf. Federer (1969, Section 1.10.5), that we shall use frequently, is

$$|B(v_1,\ldots,v_k)| \leq \|B\|_A\|v_1\|_A\cdots\|v_k\|_A, \tag{1.20}$$

for all $v_1,\ldots,v_k \in V$, but it depends on the fact that the norm on V is induced by an inner product. For later use we record that the process can be continued to the space dual to $\mathrm{Sym}_k(V;W)$. Thus, if $D \in \mathrm{Lin}\left(\mathrm{Sym}_k(V;W);\mathbf{R}\right)$, then

$$\|D\|_A = \sup\{|D(B)| : \|B\|_A \leq 1\}. \tag{1.21}$$

Notice that a set $\{v_1,\ldots,v_k\}$ of vectors in V define such a mapping D, namely by

$$D(B) = B(v_1,\ldots,v_k), \tag{1.22}$$

in which case

$$\|D\|_A \leq \|v_1\|_A\cdots\|v_k\|_A \tag{1.23}$$

follows immediately from (1.20).

We shall extend the definitions in (1.19) and (1.21) to cover the case when $\|\cdot\|_A$ is a semi-norm on V. Then the definition (1.19) implies that $\|B\|_A$ may be infinite, namely if $B(v^k)$ is non-zero for some $v \neq 0$ with $\|v\|_A = 0$. We shall then refer to (1.19) as an extended norm on $\mathrm{Sym}_k(V;W)$. Notice that the inequality

$$|B(v^k)| \leq \|B\|_A\,\|v\|_A^k \tag{1.24}$$

is still valid whenever $\|B\|_A$ is finite. In fact, also (1.20) holds for an extended norm whenever $\|B\|_A$ is finite. This is seen by the following continuity argument. Choose a sequence of inner product norms, $\|\cdot\|_n$, indexed by $n \in \mathbf{N}$, such that $\|v\|_n$ is decreasing and converges to $\|v\|_A$ for any $v \in V$. For any B for which $\|B\|_A$ is finite (1.24) holds, also if $\|v\|_A$ is replaced by $\|v\|_n$. Thus

$$\|B\|_n \leq \|B\|_A$$

and consequently (1.20) holds if we replace $\|v_j\|_A$ by $\|v_j\|_n$ for all j. A simple continuity argument now shows that (1.20) holds for the semi-norm $\|\cdot\|_A$ also.

Next, we admit the extended norm $||B||_A$ to be used in the definition (1.21). It is easy to see that this defines a semi-norm, i.e., $||D||_A$ may be zero for non-zero D, but infinite values cannot occur. The inequality

$$|D(B)| \leq ||D||_A \, ||B||'_A \tag{1.25}$$

holds whenever $||B||_A$ is finite, and the inequality (1.23) is therefore still valid.

Finally we recall that the *transpose* (or the *dual*) of a linear mapping, $f \in \mathrm{Lin}(V;W)$, is the unique linear mapping, $f^T \in \mathrm{Lin}(W^*;V^*)$, satisfying

$$w^*\{f(v)\} = \{f^T(w^*)\}(v) \tag{1.26}$$

for all $v \in V$ and $w^* \in W^*$. The matrix representing f^T is the transpose of the matrix representing f, when dual coordinate systems are used on (V, V^*) and on (W, W^*).

2 Differentiability of functions between vector spaces

Throughout this section U, V,W, and $V_1, V_2, \ldots$ denote finite dimensional real (or complex) vector spaces. We assume that all vector spaces are equipped with norms, denoted $||\cdot||$. It does not matter for the concepts of differentiability that are introduced below, which norms are chosen because all norms on finite dimensional vector spaces are equivalent.

First we need to define the o and O symbols. Let g and h be two functions defined on a subset of V in which x_0 is an inner point. Then we write $g(x) = o(h(x))$ as $x \to x_0$ if for any $\epsilon > 0$ a neighbourhood, U say, of x_0 exists such that $||g(x)|| \leq \epsilon ||h(x)||$ for all $x \in U$, where $||h(x)|| = |h(x)|$ if $h(x)$ is real. The notation $g(x) = O(h(x))$ as $x \to x_0$ implies the existence of a neighbourhood U of x_0 and a constant $c > 0$ such that $||g(x)|| \leq c||h(x)||$ for all $x \in U$.

Definition 2.1. *A function $f : B \to W$, $B \subseteq V$, is said to be differentiable at $x_0 \in \mathrm{int}(B)$ if there exists a linear function, $Df(x_0)$ say, of V into W, such that*

$$f(x) = f(x_0) + \{Df(x_0)\}(x - x_0) + o(||x - x_0||) \tag{2.1}$$

as $x \to x_0$. The function $Df(x_0) \in \mathrm{Lin}(V;W)$ is called the (first) differential of f at x_0.

Notice that because any function in $\mathrm{Lin}(V;W)$ is continuous, any function that is differentiable at x_0 is also continuous at x_0.

Definition 2.2. *A function $f : B \to W$, $B \subseteq V$, is said to be differentiable in B_0 if $B_0 \subseteq B$ is an open subset of V and f is differentiable at any point in B_0. If, furthermore, the function*

$$\begin{aligned} Df : B_0 &\to \mathrm{Lin}(V;W) \\ x &\mapsto Df(x) \end{aligned} \tag{2.2}$$

is continuous at $x_0 \in B_0$ or in B_0, then f is said to be continuously differentiable at x_0 or in B_0, respectively.

If $f : B \to W$ is differentiable in $B_0 \subseteq B \subseteq V$ and the function in (2.2) is differentiable at $x_0 \in B_0$, then f is said to be twice differentiable at x_0. In that case the function

$$D(Df)(x_0) \in \mathrm{Lin}(V; \mathrm{Lin}(V; W)) \tag{2.3}$$

may be identified with a function, denoted $D^2 f(x_0)$, in $\mathrm{Lin}_2(V; W)$ by the identification

$$\{D^2 f(x_0)\}(v_1, v_2) = \{(D(Df)(x_0))(v_2)\}(v_1), \tag{2.4}$$

where $v_1, v_2 \in V$. The function $D^2 f(x_0)$ is called the second differential of f at x_0.This process may be repeated to the following inductive definition of higher order differentials.

Definition 2.3. *A function $f : B \to W$, $B \subseteq V$, is said to be k times (continuously) differentiable at $x_0 \in \mathrm{int}(B)$ if it is $k - 1$ times differentiable in a neighbourhood B_0 of x_0, and the function*

$$D^{k-1} f : B_0 \to \mathrm{Lin}_{k-1}(V; W) \tag{2.5}$$

is (continuously) differentiable at x_0. In that case we define the kth differential of f at x_0 as the k-linear function

$$\begin{aligned} &D^k f(x_0) \in \mathrm{Lin}_k(V; W), \\ &\{D^k f(x_0)\}(v_1, \ldots, v_k) = \{(D(D^{k-1} f)(x_0))(v_k)\}(v_1, \ldots, v_{k-1}), \end{aligned} \tag{2.6}$$

where $v_1, \ldots, v_k \in V$.

For a function of several variables we shall use the notation $D_x^k f$ for the kth differential of f with respect to the variable x.

If f is k times differentiable at x_0 then (2.1) generalizes to the kth order Taylor series expansion

$$f(x) = f(x_0) + \{Df(x_0)\}(x - x_0) + \cdots + \frac{1}{k!}\{D^k f(x_0)\}\{(x - x_0)^k\} + o\left(\|x - x_0\|^k\right) \tag{2.7}$$

as $x \to x_0$, where $(x - x_0)^k = (x - x_0, \ldots, x - x_0) \in V^k$ by convention, cf. (1.3). To avoid the many parentheses required for such expansions we shall prefer the simpler, though less precise, notation

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \cdots + \frac{1}{k!} D^k f(x_0)(x - x_0)^k + o\left(\|x - x_0\|^k\right) \tag{2.8}$$

as $x \to x_0$, instead of (2.7) when there is no risk of ambiguity. Notice that the expansion (2.8) looks exactly the same as for one-dimensional functions. This is no coincidence since (2.8) is really an expansion of the function $f(x_0 + h(x - x_0))$ as a function of $h \in \mathbf{R}$. More precisely, this follows from the identity

$$D^k f(x_0)(x - x_0)^k = \left(\frac{\partial}{\partial h}\right)^k f(x_0 + h(x - x_0)) \tag{2.9}$$

evaluated at $h = 0$. This identity holds whenever f is k times differentiable at x_0. More generally we have

$$D^k f(x_0)(v_1, \dots, v_k) = \frac{\partial}{\partial h_k} \cdots \frac{\partial}{\partial h_1} f(x_0 + h_1 v_1 + \cdots + h_k v_k), \tag{2.10}$$

evaluated at $h_1 = 0, \dots, h_k = 0$, where $h_j \in \mathbf{R}$ and $v_j \in V$ for $j = 1, \dots, k$.

An important, well-known result is the following. The proof may be found, e.g., in Federer (1969, Section 3.1.11).

Lemma 2.4. *If $f : B \to W$, $B \subseteq V$, is k times differentiable at $x_0 \in B$, then $D^k f(x_0) \in \mathrm{Sym}_k(V; W)$, i.e., the kth differential at x_0 is symmetric.*

Thus, from the fact that a symmetric k-linear function is determined by its values on the diagonal it follows that the kth differential is determined by the kth derivatives of the one-dimensional 'directional' functions in (2.9).

It may be seen from (2.10) that the coordinate representation of $D^k f(x_0)$ is the $(k+1)$-dimensional array of partial derivatives

$$\left[D^k f(x_0)\right]_j^{i_1 \cdots i_k} = \frac{\partial}{\partial x_{i_k}} \cdots \frac{\partial}{\partial x_{i_1}} f_j(x) \tag{2.11}$$

evaluated at $x = x_0$, where $f_j(x)$ is the jth coordinate of $f(x) \in W$ and x_i is the ith coordinate of $x \in V$. Thus, except for the error term, the coordinate version of (2.8) is

$$f_j(x) \sim f_j(x_0) + \sum_{m=1}^{k} \frac{1}{m!} [D^m f]_j^{i_1 \cdots i_m} [x - x_0]_{i_1} \cdots [x - x_0]_{i_m}. \tag{2.12}$$

It is often necessary to provide a bound for the remainder in Taylor's formula (2.8). If we let

$$R_k(x) = f(x) - f(x_0) - \sum_{m=1}^{k} \frac{1}{m!} D^m f(x_0)(x - x_0)^m, \tag{2.13}$$

and assume that f is $k+1$ times continuously differentiable at any point $x_0 + h(x - x_0)$ with $0 \le h \le 1$, then

$$R_k(x) = \int_0^1 \frac{1}{k!} D^{k+1} f(x_0 + h(x - x_0))(x - x_0)^{k+1} (1 - h)^k \, dh. \tag{2.14}$$

Hence, for any such function we deduce that

$$
\begin{aligned}
|R_k(x)| &\le \frac{1}{(k+1)!} \sup\left\{ \left|D^{k+1} f(x_0 + h(x - x_0))(x - x_0)^{k+1}\right| : 0 \le h \le 1 \right\} \\
&\le \frac{1}{(k+1)!} \|x - x_0\|^{k+1} \sup\left\{ \left\|D^{k+1} f(x_0 + h(x - x_0))\right\| : 0 \le h \le 1 \right\}.
\end{aligned}
\tag{2.15}
$$

Of course, the result (2.14) and the first inequality in (2.15), identified through (2.9), hold under the weaker condition that the function $h \to f(x_0 + h(x - x_0))$ is $(k+1)$ times continuously differentiable on $0 \le h \le 1$.

Beside the elementary rules of differentiation we shall need a few results on higher order differentials that we record below for later reference.

For a product, $(fg)(x) = f(x)g(x)$, of two functions $f, g : B \to \mathbf{R}$ where $B \subseteq V$, Leibnitz' rule, cf. Federer (1969, Section 3.1.11), states that

$$
D^k(fg)(x)(v^k) = \sum_{j=0}^{k} \binom{k}{j} \left\{D^j f(x)(v^j)\right\} \left\{D^{k-j} g(x)(v^{k-j})\right\} \tag{2.16}
$$

for $v \in V$, whenever f and g are k times differentiable at $x \in \operatorname{int}(B)$. We shall need this formula in a slightly more general setting, in particular as a formula for the differentiation of a composite linear function with respect to a parameter. For this purpose we only need to generalize the product from above to a more general kind of product, namely any mapping $\pi \in \operatorname{Lin}(V_1, V_2; W)$ for which a constant $c > 0$ exists, such that

$$
\|\pi(v_1, v_2)\| \le c\|v_1\|\,\|v_2\| \tag{2.17}
$$

for all $v_1 \in V_1$ and $v_2 \in V_2$. Then, if $f_1 : B \to V_1$ and $f_2 : B \to V_2$, where $B \subseteq V$, are both k times differentiable at $x_0 \in \operatorname{int}(B)$, Leibnitz' rule applies to the product mapping

$$
\pi(f_1, f_2) : x \mapsto \pi\left(f_1(x), f_2(x)\right) \in W
$$

in the form

$$
D^k \left\{\pi(f_1, f_2)\right\}(x_0)(v^k) = \sum_{j=0}^{k} \binom{k}{j} \pi\left(D^j f_1(x_0)(v^j), D^{k-j} f_2(x_0)(v^{k-j})\right), \tag{2.18}
$$

for all $v \in V$. As mentioned above, the main application of (2.18) is to compositions of linear functions depending on a parameter. Thus, if $f_\beta : V_1 \to V_2$ and $g_\beta : V_2 \to V_3$ are linear functions indexed by a parameter $\beta \in B \subseteq V$, say, then differentiation of the composition $g_\beta \circ f_\beta$ with respect to β obeys (2.18) in which π denotes composition of the two linear functions.

Related to the products is the differentiation of reciprocals. We shall give this result only in a special case. Consider a function $f \in \operatorname{Lin}(V; W)$, where $\dim(V) =$

$\dim(W)$. Let R denote the mapping

$$R(f) = f^{-1} \in \mathrm{Lin}(W;V) \tag{2.19}$$

defined on the subspace of one-to-one linear mappings of V into W. Then R is indefinitely often differentiable at any point f_0 in this space and

$$D^k R(f_0)(h^k) = (-1)^k k!\, f_0^{-1} \circ h \circ f_0^{-1} \circ \cdots \circ h \circ f_0^{-1} \in \mathrm{Lin}(W;V) \tag{2.20}$$

for any $k \in \mathbb{N}$ and $h \in \mathrm{Lin}(V;W)$, where h appears k times and f_0 appears $k+1$ times on the right. A more familiar expression of this result is in terms of the Taylor series expansion

$$(f_0 + h)^{-1} \sim f_0^{-1} - f_0^{-1} h f_0^{-1} + f_0^{-1} h f_0^{-1} h f_0^{-1} - \cdots \in \mathrm{Lin}(W;V). \tag{2.21}$$

Also the chain rule generalizes to higher order differentials, cf. Federer (1969, Section 3.1.11). Thus, consider two mappings $f : B_1 \to B_2$ and $g : B_2 \to W$, where $B_1 \subseteq V_1$ and $B_2 \subseteq V_2$, and assume that they are k times differentiable at $x_1 \in \mathrm{int}(B_1)$ and $x_2 = f(x_1) \in \mathrm{int}(B_2)$, respectively. Then the composite function $g \circ f : B_1 \to W$ is k times differentiable at x_1 and

$$\begin{aligned} &D^k(g \circ f)(x_1)(v^k) \\ &= \sum_{\alpha \in T(k)} k!\, \{D^{\Sigma \alpha_j} g(x_2)\} \left\{ (Df(x_1)(v))^{\alpha_1}, \ldots, (D^k f(x_1)(v^k))^{\alpha_k} \right\} / \prod \{\alpha_j!\, j!^{\alpha_j}\} \end{aligned} \tag{2.22}$$

for any $v \in V_1$, where $T(k)$ is the set

$$T(k) = \left\{ (\alpha_1, \ldots, \alpha_k) \in \mathbb{N}_0^k : \sum_{j=1}^k j\alpha_j = k \right\}. \tag{2.23}$$

Notice that in (2.22) each component of the form $D^j f(x_1)(v^j)$ is a vector in V_2. Sometimes it is more convenient to apply the formula for higher order differentials for composite functions in terms of the sets

$$S_m(k) = \left\{ (a_1, \ldots, a_m) \in \mathbb{N}^m : \sum a_j = k \right\}, \tag{2.24}$$

defined for $m, k \in \mathbb{N}$ satisfying $m \le k$. Each sequence a in $S_m(k)$ corresponds to a sequence $\alpha(a) = (\alpha_1, \ldots, \alpha_k)$ in $T(k)$, namely by identifying α_j with the number of a's from the sequence in $S_m(k)$ that are equal to j. Now, it may be seen that for any function $\alpha \mapsto g(\alpha)$ on $T(k)$ this identification induces a function $\tilde{g}$, say, on

$S_m(k)$ for $m = 1, \ldots, k$, namely by $\tilde{g}(a) = g(\alpha(a))$, satisfying

$$\sum_{\alpha \in T(k)} \left(\prod \alpha_j!\right)^{-1} g(\alpha) = \sum_{m=1}^{k} \sum_{a \in S_m(k)} \frac{1}{m!} \tilde{g}(a). \tag{2.25}$$

By use of this relation we may re-express formula (2.22) as

$$D^k(g \circ f)(x_1)(v^k)$$
$$= \sum_{m=1}^{k} \sum_{a \in S_m(k)} \frac{k!}{m!} \{D^m g(x_2)\} \{(D^{a_1} f(x_1)(v^{a_1})), \ldots, (D^{a_k} f(x_1)(v^{a_k}))\} / \prod a_j! . \tag{2.26}$$

For computations related to expressions of the form (2.22) or (2.26) we shall be using the relation,

$$\sum_{\alpha \in T_m(k)} \left(\prod \alpha_j!\right)^{-1} = \sum_{a \in S_m(k)} \frac{1}{m!} = \frac{1}{m!} \binom{k-1}{m-1}, \tag{2.27}$$

where $T_m(k)$ is the subset of $T(k)$ on which $\sum \alpha_j = m$. From this it is furthermore seen that

$$\sum_{\alpha \in T(k)} \left(\prod \alpha_j!\right)^{-1} \left(\sum \alpha_j\right)! \gamma^{\Sigma \alpha_j} = \sum_{m=1}^{k} \binom{k-1}{m-1} \gamma^m$$
$$= \gamma(\gamma+1)^{k-1}, \tag{2.28}$$

for any $\gamma \in \mathbf{R}$, cf. Federer (1969, Section 3.1.24).

3 Analytic functions between vector spaces

In this section let V and W denote finite dimensional normed real (or complex) vector spaces and consider a mapping $f : B \to W$ where $B \subseteq V$. The definition of analytic functions in this setting, cf., e.g., Federer (1969, Section 3.1.24), does not rely on complex function analysis, but agrees with the traditional definition of analytic functions in that framework.

Definition 3.1. *A function $f : B \to W$, $B \subseteq V$, is said to be analytic at a point $x_0 \in \mathrm{int}(B)$ if it is infinitely often differentiable at x_0 and satisfies the following two conditions:*

(1) There is a constant $c < \infty$ such that

$$\|D^k f(x_0)(v^k)\| \leq k!\, c^k \|v\|^k \tag{3.1}$$

for all $k \in \mathbf{N}$.

(2) For all x *in some neighbourhood of* x_0,

$$f(x) = \sum_{k=1}^{\infty} \frac{1}{k!} D^k f(x_0)(x - x_0)^k. \tag{3.2}$$

Notice that the sum in (3.2) converges absolutely in some neighbourhood of x_0 because of (1). The condition (2) then states that the function actually agrees locally with the sum. That this is not a consequence of (1) is seen, e.g., from the standard counter example,

$$f(x) = \exp(-1/|x|)$$

of a function of $\mathbf{R}$ into $\mathbf{R}$. For this function all derivatives vanish at $x = 0$.

Definition 3.2. *A function* $f : B \to W$, $B \subseteq V$, *is said to be analytic in* $B_0 \subseteq B$, *if* B_0 *is an open subset of* V *and* f *is analytic at every point in* B_0.

Polynomials are analytic at all points of their domain, linear combinations of analytic functions are analytic and compositions of analytic functions are analytic. Also the inverse mapping of an analytic mapping is analytic at any point at which the differential is non-singular, cf. Federer (1969, Section 3.1.24).

It is evident from the definition that a Taylor series expansions of an analytic function in a sufficiently small neighbourhood is bounded by a *geometric series*, i.e., a series of the form

$$c \sum_{k=0}^{\infty} a^k \tag{3.3}$$

for some $c \geq 0$ and $0 \leq a < 1$.

The main virtue of analytic functions in the present context is that the remainder from the Taylor series expansion,

$$R_k(x) = f(x) - f(x_0) - \sum_{m=1}^{k} \frac{1}{m!} D^m f(x_0)(x - x_0)^m$$

around a point x_0 at which the function is analytic, may be bounded in terms of the derivatives at x_0 only. More precisely, if $f : B \to W$ is analytic at $x_0 \in \mathrm{int}(B)$ and c is the constant from (3.1), then

$$\begin{aligned} |R_k(x)| &= \Big| \sum_{m=k+1}^{\infty} \frac{1}{m!} D^m f(x_0)(x - x_0)^m \Big| \\ &\leq \sum_{m=k+1}^{\infty} c^m \|x - x_0\|^m \\ &= c^{k+1} \|x - x_0\|^{k+1} / (1 - c\|x - x_0\|), \end{aligned} \tag{3.4}$$

for all x in some neighbourhood of x_0. In fact, (3.4) holds for all x satisfying $\|x - x_0\| < c^{-1} \leq \infty$ if f is analytic in this area.

Inequalities similar to (3.4) hold for the derivatives of f at any point where f is analytic. In particular, if f is analytic at $x_0 \in B$, so is any of the functions $x \mapsto D^k f(x)$.

4 Moments and cumulants of random vectors

Throughout this section let V be a finite dimensional real vector space, V^* its dual, and let X be a random variable with values in V and with a distribution denoted P. Some definitions of moments and cumulants are given below, but only for cases in which all moments up to a certain order exist. Thus, we do not extend the definitions to cases where, e.g., a certain moment exists for one coordinate, but not for another.

Notice that for each $t \in V^*$ the variable $t(X)$ is a one-dimensional real random variable. Our definitions of moments and cumulants are based on such random variables.

Definition 4.1. *The random variable $X \in V$ is said to have finite kth moment if*

$$\mathrm{E}\left\{|t(X)^k|\right\} < \infty$$

for any $t \in V^*$. In that case the kth moment of X is defined as the k-linear symmetric form

$$\begin{aligned} &\mu_k \in \mathrm{Sym}_k(V^*;\mathbf{R}) \\ &\mu_k(t_1,\ldots,t_k) = \mathrm{E}\{t_1(X)\cdots t_k(X)\}. \end{aligned} \tag{4.1}$$

Notice that because μ_k is a k-linear symmetric form it is determined by its values on the diagonal, i.e., by the values $\mu_k(t^k)$, $t \in V^*$. The coordinate representation of μ_k is

$$\begin{aligned} [\mu_k]_{i_1\cdots i_k} &= \mathrm{E}\left\{[X]_{i_1}\cdots[X]_{i_k}\right\}, \\ \mu_k(t^k) &= [\mu_k]_{i_1\cdots i_k}[t]^{i_1}\cdots[t]^{i_k}, \qquad t \in V^*, \end{aligned} \tag{4.2}$$

where $[X]_i$ and $[t]^i$ denote the ith coordinate of X and t, respectively, with respect to dual bases.

Definition 4.2. *The characteristic function of the random variable X is the function*

$$\begin{aligned} &\xi : V^* \to \mathbf{C}, \\ &\xi(t) = \mathrm{E}\exp\{it(X)\}. \end{aligned} \tag{4.3}$$

It is well known that if the kth moment of X exists then $D^k\xi(t)$ exists and is uniformly continuous in $t \in V^*$, and that

$$D^k\xi(0) = i^k\mu_k. \tag{4.4}$$

Thus, in that case, we have the Taylor series expansion

$$\xi(t) = 1 + \sum_{j=1}^{k} \frac{i^j}{j!}\mu_j(t^j) + o\left(\|t\|^k\right) \tag{4.5}$$

as $t \to 0$. Since $\xi(0) = 1$ the equation (4.4) implies the existence of $\operatorname{Log}\xi(t)$ as a k times differentiable function in a neighbourhood of $t = 0$.

Definition 4.3. *If X has finite kth moment, then the kth cumulant of X is defined as the k-linear symmetric form*

$$\begin{aligned}&\kappa_k \in \operatorname{Sym}_k(V^*;\mathbf{R})\\ &\kappa_k(t_1,\ldots,t_k) = i^{-k}\{D^k\operatorname{Log}\xi(0)\}(t_1,\ldots,t_k),\end{aligned} \tag{4.6}$$

where Log *denotes the (principal branch of) the complex logarithmic function.*

The kth cumulant is, as the moments, determined by the values on the diagonal, i.e., by the values of $\kappa_k(t^k)$ for $t \in V^*$, which for any fixed t equal the kth cumulants of the real random variables $t(X)$. It follows from above that if the kth moment of X is finite then

$$\operatorname{Log}\xi(t) = \sum_{j=1}^{k} \frac{i^j}{j!}\kappa_j(t^j) + o\left(\|t\|^k\right) \tag{4.7}$$

as $t \to 0$. For real random variables $X_1,\ldots,X_k \in \mathbf{R}$ we use the notation

$$\operatorname{cum}(X_1,\ldots,X_k) \in \mathbf{R}$$

to denote the joint cumulant of the variables, i.e.,

$$\operatorname{cum}(X_1,\ldots,X_k) = i^{-k}\frac{\partial}{\partial h_1}\cdots\frac{\partial}{\partial h_k}\operatorname{Log}\{\operatorname{E}\exp\left(ih_1X_1 + \cdots + ih_kX_k\right)\}, \tag{4.8}$$

evaluated at $h_1 = \cdots = h_k = 0$, whenever this derivative exists. When $X_1 = \cdots = X_k$ in (4.8) we abbreviate to

$$\operatorname{cum}_k(X) = \operatorname{cum}(X,\ldots,X). \tag{4.9}$$

It follows that (4.6) can be expressed as

$$\kappa_k(t_1,\ldots,t_k) = \operatorname{cum}\left(t_1(X),\ldots,t_k(X)\right) \tag{4.10}$$

and that, in particular,

$$\kappa_k(t^k) = \operatorname{cum}_k(t(X)), \tag{4.11}$$

which gives the definition of cumulants entirely in terms of cumulants of real random variables.

If X has finite second moment its variance is defined as the bilinear symmetric form

$$\{\operatorname{var}(X)\}(t^2) = \kappa_2(t^2) = \operatorname{var}\{t(X)\} \geq 0. \tag{4.12}$$

Thus, κ_2 is (semi)-definite and defines a (pseudo) inner product on V^*, denoted $\langle \cdot, \cdot \rangle_{\kappa_2}$, cf. (1.14). This (pseudo) inner product then induces a (semi)-norm on V^*. If κ_2 is positive definite it has an inverse $\kappa_2^{-1} \in \operatorname{Sym}_2(V; \mathbf{R})$, cf. (1.16), which is also positive definite and therefore induces an inner product and a norm on V. We shall frequently use this construction of an inner product from the variance of the random variable considered.

If $X_1 \in V_1$ and $X_2 \in V_2$ are both finite dimensional random variables and if $(X_1, X_2) \in V_1 \times V_2$ has finite second moment, then we define the covariance between X_1 and X_2 as the bilinear form $\operatorname{cov}(X_1, X_2)$ in $\operatorname{Lin}(V_1^*, V_2^*; \mathbf{R})$ defined as

$$\{\operatorname{cov}(X_1, X_2)\}(t_1, t_2) = \operatorname{cov}(t_1(X_1), t_2(X_2)), \tag{4.13}$$

where the right side denotes the familiar covariance of real random variables.

By repeated differentiation of $\operatorname{Log} \xi(t)$ and of $\exp \operatorname{Log} \xi(t)$ according to the formula (2.22) we obtain the relations between moments and cumulants,

$$\mu_k(t^k) = \sum_{\alpha \in T(k)} k! \left(\prod j!^{\alpha_j} \alpha_j!\right)^{-1} \prod \{\kappa_j(t^j)\}^{\alpha_j} \tag{4.14}$$

and

$$\kappa_k(t^k) = \sum_{\alpha \in T(k)} k! \left(\prod j!^{\alpha_j} \alpha_j!\right)^{-1} (-1)^{\Sigma \alpha_j - 1} \left(\sum \alpha_j - 1\right)! \prod \{\mu_j(t^j)\}^{\alpha_j}, \tag{4.15}$$

whenever μ_k exists, where $T(k)$ is the set defined in (2.23).

A more advanced relation gives the joint cumulant in (4.8) in terms of moments as

$$\operatorname{cum}(X_1, \ldots, X_k) = \sum_S (-1)^{A-1}(A-1)!\, \mu(S_1) \cdots \mu(S_A), \tag{4.16}$$

where the sum is over all partitions $S = (S_1, \ldots, S_A)$; $A = 1, \ldots, k$, of $\{1, \ldots, k\}$ into A non-empty subsets, i.e.,

$$S_1 \cup \cdots \cup S_A = \{1, \ldots, k\}$$

and the subsets are pairwise disjoint, and

$$\mu(S_a) = \mathrm{E}\left\{ \prod_{j \in S_a} X_j \right\}. \tag{4.17}$$

The proof of this relation may be found in McCullagh (1987, Section 3.6.2).

Some inequalities for cumulants and moments in terms of each other follow from (4.14) and (4.15). These are stated in the following two lemmas.

Lemma 4.4. *If $X \in V$ has finite kth moment and there exist two constants $c \geq 0$ and $\lambda \geq 0$ such that*

$$|\mu_j(t^j)| \leq cj!\,\lambda^j \|t\|^j \tag{4.18}$$

for all $j = 1, \ldots, k$, where $\|\cdot\|$ is any norm on V, then

$$\begin{aligned} |\kappa_k(t^k)| &\leq (k-1)!\,\lambda^k\{(1+c)^k - 1\}\,\|t\|^k \\ &\leq (k-1)!\,\{(1+c)\lambda\}^k\,\|t\|^k. \end{aligned} \tag{4.19}$$

Proof. Insertion of (4.18) into (4.15) together with an application of (2.25) and (2.27) yields

$$\begin{aligned} |\kappa_k(t^k)| &\leq \sum_{\alpha \in T(k)} k!\,\left(\sum \alpha_j - 1\right)!\,\prod_{j=1}^{k} \left\{ \left(c\lambda^j \|t\|^j\right)^{\alpha_j} \alpha_j!^{-1} \right\} \\ &= k!\,\lambda^k \|t\|^k \sum_{m=1}^{k} \frac{1}{m} c^m \binom{k-1}{m-1} \\ &= (k-1)!\,\lambda^k\{(1+c)^k - 1\}\|t\|^k \end{aligned}$$

as claimed in the first inequality in (4.19). The second is trivial. ∎

Lemma 4.5. *If $X \in V$ has finite kth moment and there exist two constants $c \geq 0$ and $\lambda \geq 0$ such that*

$$|\kappa_j(t^j)| \leq c(j-1)!\,\lambda^j\|t\|^j \tag{4.20}$$

for all $j = 1, \ldots, k$, then

$$|\mu_k(t^k)| \leq \frac{(c+k-1)!}{(c-1)!}\lambda^k\|t\|^k, \tag{4.21}$$

where the definition $x! = \Gamma(x+1)$ covers cases of non-integer valued c.

Proof. From (4.14) and (4.20) we see that

$$|\mu_k(t^k)| \leq \sum_{\alpha \in T(k)} k! \prod_{j=1}^{k} \left\{ \left(c\lambda^j\|t\|^j/j\right)^{\alpha_j} \alpha_j!^{-1} \right\}.$$

But all terms in this sum are positive and the sum equals the expression for the kth moment of a Γ-distribution in terms of its cumulants. The Γ-distribution with shape parameter θ and expectation $\theta\beta$ has jth cumulant $(j-1)!\,\theta\beta^j$ and jth moment $(\Gamma(\theta+j)/\Gamma(\theta))\,\beta^j$, so if we let $\theta = c$ and $\beta = \lambda\|t\|$ the claim is justified and we see that the sum equals the right hand side of (4.21). ∎

In place of the characteristic function it is often convenient to work with the moment generating function, $\mu : V^* \to \mathbf{R}$, given by

$$\mu(t) = \mathrm{E}\exp\{t(X)\} \leq \infty.$$

This definition is extended to admit complex arguments for which we need the complex dual $V^*_{\mathbf{C}} = \mathrm{Lin}(V;\mathbf{C})$. A strict definition requires the extension of V to a complex vector space, but for our purpose the representation $V^*_{\mathbf{C}} = V^* + iV^*$ will suffice, where $i^2 = -1$.

Definition 4.6. *Consider the set*

$$T_{\mathbf{C}}(X) = \{\, t \in V^*_{\mathbf{C}} : \mathrm{E}|\exp\{t(X)\}| < \infty \,\}. \tag{4.22}$$

The function $\mu : T_{\mathbf{C}}(X) \to \mathbf{C}$,

$$\mu(t) = \mathrm{E}\exp\{t(X)\} \tag{4.23}$$

is called the moment generating function of X*. The same name is used for the mapping (4.23) considered as a function on the subset*

$$T_{\mathbf{R}}(X) = T_{\mathbf{C}}(X) \cap V^*$$

of real linear forms on V*.*

Notice that we always have

$$T_{\mathbf{R}}(X) + iV^* \subseteq T_{\mathbf{C}}(X), \tag{4.24}$$

i.e., if $z = t_1 + it_2 \in V^*_{\mathbf{C}}$, where $t_1, t_2 \in V^*$, is a complex linear form on V, then $z \in T_{\mathbf{C}}(X)$ if $t_1 \in T_{\mathbf{R}}(X)$, simply because $|\exp z(X)|$ equals $\exp t_1(X)$.

The name, moment generating function, refers to the fact that the moments are the derivatives of μ at zero, if these exist. Obviously,

$$\xi(t) = \mu(it) \tag{4.25}$$

so that μ is always defined on iV^*, and the moments may always, if they exist, be found as derivatives of $\mu(z)$ at $z = 0$ for purely imaginary arguments z.

Definition 4.7. *The function*

$$\kappa(z) = \operatorname{Log}\mu(z) = \operatorname{Log}\{\mathrm{E}\exp z(X)\}, \tag{4.26}$$

$z \in V_{\mathbf{C}}^*$, defined on the subset of $T_{\mathbf{C}}(X)$ on which $\mu(z)$ does not meet the non-positive real axis, is called the cumulant generating function. The same name applies to the function (4.26) restricted to the subset obtained by considering $z \in V^*$ only.

Of special interest are the random variables for which the moment generating function is finite in a neighbourhood of the origin in V^*.

Definition 4.8. *A random variable $X \in V$ is said to have finite exponential moments if*

$$\mu(t) = \mathrm{E}\exp\{t(X)\} < \infty \tag{4.27}$$

for all $t \in V^*$ in some neighbourhood of zero.

It follows easily from Hölder's inequality that a finite sum of random variables with finite exponential moments, itself has finite exponential moments.

Lemma 4.9. *A random variable $X \in V$ has finite exponential moments if and only if*

$$\mathrm{E}\{\exp(s\|X\|)\} < \infty \tag{4.28}$$

for some $s > 0$, where $\|\cdot\|$ is any norm on V.

Proof. If (4.28) holds then (4.27) holds for all $t \in V^*$ with $\|t\| \leq s$, because $|t(X)| \leq \|t\|\,\|X\|$. Assume now that (4.27) holds and that $X = X_1 v_1 + \cdots + X_p v_p$, where $(v_1, \ldots, v_p)$ is a basis on V and $X_1, \ldots, X_p$ are real random variables. Then

$$\|X\| \leq |X_1|\,\|v_1\| + \cdots + |X_p|\,\|v_p\|.$$

But each of the variables X_j are linear functions of X and therefore

$$\mathrm{E}\,\exp\{s_j\,|X_j|\,\|v_j\|\} < \mathrm{E}\exp\{s_j\,\|v_j\|\,X_j\} + \mathrm{E}\exp\{-s_j\,\|v_j\|\,X_j\}$$

which is finite for sufficiently small $s_j > 0$. The result now follows from Hölder's inequality. ∎

In some cases we consider a family $\{P_\beta; \beta \in B\}$ of distributions for a random variable X, where β ranges over some set B. Then we say that X has *uniformly bounded exponential moments* with respect to $\beta \in B$, if there exists an $s > 0$ such that

$$\sup\{\,\mathrm{E}_\beta \exp(s\|X\|) : \beta \in B\,\} < \infty. \tag{4.29}$$

If a random variable X has finite exponential moments, then the moment generating function, μ, is analytic in the set

$$\{\, z \in V_{\mathbf{C}}^* : z = t_1 + it_2;\; t_1, t_2 \in V^*;\; t_1 \in \mathrm{int}\,(T_{\mathbf{R}}(X))\,\}, \tag{4.30}$$

where $T_{\mathbf{R}}(X)$ is defined in Definition 4.6. Hence, it follows by analytic continuation that μ is determined throughout this set from its derivatives at zero. We conclude that for a random variable with finite exponential moments, the moments determine the distribution uniquely. It is useful to be able to see whether this is the case from the moments themselves, without the assumption that X has finite exponential moments.

Lemma 4.10. *Assume that $X \in V$ has finite moments of all orders, satisfying*

$$\left|\mu_k(t^k)\right| \leq ck!\,\lambda^k\|t\|^k \tag{4.31}$$

for some $c \geq 0$, $\lambda \geq 0$ and all $k \in \mathbb{N}$ and $t \in V^*$. Then the distribution is uniquely determined by the moments, X has finite exponential moments, and for any $t \in V^*$ with $\|t\| < \lambda^{-1}$,

$$\mu(t) = \sum_{k=0}^{\infty} \frac{1}{k!}\mu_k(t^k), \tag{4.32}$$

where the sum is absolutely convergent.

For the case of one-dimensional random variables the proof of Lemma 4.10 may be found in Feller (1971, Chapter XV, Section 5). The extension to several dimensions is trivial.

From Lemma 4.10 it is easy to extend the result to the cumulants for which the inequalities will often be given in the form (4.20).

Lemma 4.11. *Assume that all cumulants of $X \in V$ exist and satisfy*

$$\left|\kappa_k(t^k)\right| \leq c(k-1)!\,\lambda^k\|t\|^k \tag{4.33}$$

for some $c \geq 0$, $\lambda \geq 0$, and all $k \in \mathbb{N}$ and $t \in V^*$. Then the distribution of X is uniquely determined by its cumulants, X has finite exponential moments, and for any $t \in V^*$ with $\|t\| < \lambda^{-1}$,

$$\kappa(t) = \sum_{k=1}^{\infty} \frac{1}{k!}\kappa_k(t^k) \tag{4.34}$$

where the sum is absolutely convergent.

Proof. . From Lemma 4.5 and the inequality

$$\binom{c+k-1}{k} \leq (c+k-1)^c$$

it follows that the assumption and hence the conclusion of Lemma 4.10 holds with λ replaced by any $\tilde{\lambda}$, say, with $\tilde{\lambda} > \lambda$ and a suitably chosen constant c. Then we know that $\mu(z)$ and hence $\kappa(z)$ is analytic in a set containing the subset of $V_{\mathbb{C}}^*$ on which $\|t\| < \lambda^{-1}$, and the result follows from the theory of analytic functions. ∎

It is not implied by Lemma 4.10 or Lemma 4.11 that a distribution with the given series of moments or cumulants exists, only that there is at most one.

For later reference we conclude this section with a few well-known lemmas.

Lemma 4.12. *If $X \in \mathbf{R}$ is non-negative with probability one, then*

$$\mathrm{E}\,X = \int_0^\infty \mathrm{P}(X > x)\,dx \leq \infty. \tag{4.35}$$

Proof. From Tonelli's theorem it follows that the two integrals can be interchanged in the computation

$$\begin{aligned}\mathrm{E}\,X &= \int_0^\infty x\,dF(x) = \int_0^\infty \int_0^\infty I_{\{y<x\}}\,dy\,dF(x)\\ &= \int_0^\infty \mathrm{P}(X > y)\,dy.\end{aligned}$$

∎

Lemma 4.13. *For any positive random variable X with $\mu(t) < \infty$ for some fixed $t \in \mathbf{R}$, we have*

$$\mathrm{P}(X \geq x) \leq \mu(t)\exp(-tx). \tag{4.36}$$

Proof. This is Chebychev's inequality applied to $\exp(tX)$. ∎

5 Some inequalities for symmetric multilinear mappings

In this section let V and W denote finite dimensional real vector spaces and let $p = \dim V$. We assume that W is equipped with a norm denoted $|\cdot|$, and that V is equipped with an inner product $\langle\cdot,\cdot\rangle$ and a corresponding norm denoted $\|\cdot\|$.

Recall from (1.19) and (1.20) that for any $A \in \mathrm{Sym}_k(V;W)$,

$$\begin{aligned}\|A\| &= \sup\{\,|A(v^k)| : \|v\| \leq 1\,\}\\ &= \sup\{\,|A(v_1,\dots,v_k)| : \|v_j\| \leq 1; j = 1,\dots,k\,\}.\end{aligned} \tag{5.1}$$

To estimate factorials we shall need Stirling's formula in the form

$$k! = (2\pi)^{1/2}(k+1)^{k+\frac{1}{2}}e^{-(k+1)}(1+\delta_k), \tag{5.2}$$

where $\delta_k = O(k^{-1})$ as $k \to \infty$ and $0 < \delta_k < 1/23$ when $k \geq 1$. In particular, we shall be using the following consequence,

$$k!^{\,-1}k^k < (\pi k)^{-1/2}e^k \tag{5.3}$$

for all $k \geq 1$. This follows by simple manipulations from (5.2).

Lemma 5.1 below gives, like (5.1), a bound for off-diagonal values of a symmetric multilinear mapping in terms of diagonal values, but for the case when the diagonal values are bounded by a geometric average of factors of the form $\|v\|_i$, for some semi-norms $\|\cdot\|_i$ that need not be generated by (pseudo) inner products.

Lemma 5.1. *Let $A \in \mathrm{Sym}_k(V;W)$ and assume that there exist m semi-norms on V, denoted $\|\cdot\|_1, \ldots, \|\cdot\|_m$, $m \in \mathbb{N}$, such that for some constants $c \geq 0$ and $a_1, \ldots, a_m \geq 0$ with $a_1 + \cdots + a_m = k$, the inequality*

$$|A(v^k)| \leq c \prod_{i=1}^{m} \|v\|_i^{a_i} \tag{5.4}$$

holds for all $v \in V$. Then, for all $v_1, \ldots, v_k \in V$,

$$|A(v_1, \ldots, v_k)| \leq c\gamma(k) \prod_{i=1}^{m} \left\{ \frac{1}{k} \sum_{j=1}^{k} \|v_j\|_i \right\}^{a_i} \tag{5.5}$$

where

$$\gamma(k) = k!^{-1} k^k < (\pi k)^{-1/2} e^k. \tag{5.6}$$

Proof. The polarization identity (1.1.6) shows that

$$\begin{aligned}
|A(v_1, \ldots, v_k)| &= \left(k!\, 2^k\right)^{-1} \left| \sum_{\alpha} \left(\prod \alpha_j \right) A \left\{ \left(\sum \alpha_j v_j \right)^k \right\} \right| \\
&\leq \left(k!\, 2^k\right)^{-1} 2^k c \prod_{i=1}^{m} \left(\sup_{\alpha} \left\{ \left\| \sum \alpha_j v_j \right\|_i \right\} \right)^{a_i} \\
&\leq k!^{-1} c \prod_{i} \left(\sum_{j} \|v_j\|_i \right)^{a_i},
\end{aligned}$$

where the subscript α to the sum and the supremum refers to all sequences $(\alpha_1, \ldots, \alpha_k)$ in $\{-1, 1\}^k$. Hence (5.5) follows with $\gamma(k) = k!^{-1} k^k$. The second inequality in (5.6) is just (5.3). ∎

The next result is of importance for probabilistic calculations with random symmetric multilinear mappings, because it establishes a bound for the norm of such a mapping, cf. (5.1), in terms of its values on a finite subset of arguments on the diagonal. For bilinear symmetric forms such an inequality is well-known. In probabilistic terms it says that the variance of any linear combination of p random variables is at most p times the maximal variance of the variables, provided that the sum of squares of the coefficients of the linear combination does not exceed one. For multilinear symmetric mappings a similar inequality can be derived, although a bit weaker in terms of the constant factor. The result is given in Lemma 5.3 below; in Lemma 5.2 we give a slightly more general version. First, however, we introduce the finite set of vectors to be used in the lemmas.

Let $(e_1, \ldots, e_p)$ be an orthonormal basis with respect to the inner product on V

and define the subset

$$V_k = \left\{ \sum_{j=1}^{k} \alpha_j e_{i_j} : \alpha_j \in \{-1,1\}; i_j \in \{1,\ldots,p\}; j = 1,\ldots,k \right\} \tag{5.7}$$

of vectors in V. For the number, $|V_k|$, of elements in V_k we obtain

$$|V_k| \le 2^p \binom{p+k}{k} < 2^p p!^{-1}(p+k)^p \tag{5.8}$$

from the combinatorial argument to the number of allocations of at most k units on p base vectors combined with a sign on each base vector. The second inequality in (5.8) is trivial.

Lemma 5.2. *Let $\|\cdot\|_B$ denote a (semi)-norm on V induced by a positive (semi)-definite bilinear form $B \in \mathrm{Sym}_2(V;\mathbf{R})$. Let V_k be the set defined in (5.7) in terms of a basis $(e_1,\ldots,e_p)$ that is orthonormal with respect to the inner product $\langle\cdot,\cdot\rangle$ and orthogonal with respect to B. If $A \in \mathrm{Sym}_k(V;W)$ satisfies*

$$|A(v^k)| \le c\,\|v\|^{k-1}\|v\|_B \tag{5.9}$$

for some constant $c \ge 0$ and all $v \in V_k$, then

$$|A(v^k)| \le c\gamma(p,k)\|v\|^{k-1}\|v\|_B \tag{5.10}$$

for all $v \in V$, where

$$\gamma(p,k) = k!^{-1}k^k p^k < (\pi k)^{-1/2}(pe)^k. \tag{5.11}$$

Proof. Let $b_j = \|e_j\|_B^2$ denote the eigenvalues of B with respect to $\langle\cdot,\cdot\rangle$, and let $v = \beta_1 e_1 + \cdots + \beta_p e_p$ be any vector in V. Then

$$A(v^k) = \sum_i \left(\prod_{j=1}^{k} \beta_{i_j}\right) A(e_{i_1},\ldots,e_{i_k}) \tag{5.12}$$

where the sum is over all sequences $i = (i_1,\ldots,i_k)$ in $\{1,\ldots,p\}^k$. For any such sequence the assumption (5.9) and the polarization identity (1.6) show that

$$\begin{aligned} &|A(e_{i_1},\ldots,e_{i_k})| \\ &\quad = \left(k!\,2^k\right)^{-1} \left|\sum_\alpha \left(\prod \alpha_j\right) A\left\{\left(\sum \alpha_j e_{i_j}\right)^k\right\}\right| \\ &\quad \le \left(k!\,2^k\right)^{-1} 2^k k^k c\left(\sup\{\,b_j : j \in \{i_1,\ldots,i_k\}\,\}\right)^{1/2}, \end{aligned} \tag{5.13}$$

where $\sum_\alpha$ denotes the sum over all sequences in $\{-1,1\}^k$ and the factor 2^k in the last expression is the number of terms in this sum, while the factor k^k stems from the estimate

$$\left\|\sum \alpha_j e_{i_j}\right\| \leq k$$

and the inequality

$$\left\|\sum \alpha_j e_{i_j}\right\|_B \leq k\,(\sup\{\,b_j : j \in \{i_1,\ldots,i_k\}\,\})^{1/2}.$$

Now, notice that for any $j = 1,\ldots,p$,

$$\|v\|_B^2 = \sum_{i=1}^{p} b_i\beta_i^2 \geq b_j\beta_j^2.$$

Therefore, by use of the trivial inequality $|\beta_j| \leq \|v\|$ for all j, combination of (5.12) and (5.13) gives

$$\begin{aligned}|A(v^k)| &\leq ck!^{\,-1}k^kp^k\,\|v\|^{k-1}\sup\{\,|\beta_j|\,\sqrt{b_j} : j = 1,\ldots,p\,\}\\ &\leq ck!^{\,-1}k^kp^k\|v\|^{k-1}\|v\|_B,\end{aligned}$$

which proves (5.10) with the first expression in (5.11) for $\gamma(p,k)$. The inequality in (5.11) follows directly from (5.3). ∎

Of particular interest is the case when only one norm is involved, i.e., when $\|v\|_B = \|v\|$. Then Lemma 5.2 implies that

$$\|A\| \leq k!^{\,-1}k^kp^kM_k(A),$$

where

$$M_k(A) = \inf\{\,M \in \mathbf{R} : |A(v^k)| \leq M\|v\|^k : v \in V_k\,\}. \tag{5.14}$$

For this case we can, however, do slightly better as shown in the following lemma.

Lemma 5.3. *For any symmetric k-linear mapping $A \in \mathrm{Sym}_k(V;W)$ we have*

$$\|A\| \leq k!^{\,-1}k^kp^{k/2}M_k(A) < (\pi k)^{-1/2}(e\sqrt{p})^kM_k(A) \tag{5.15}$$

where $M_k(A)$ is given in (5.14) in terms of the set V_k defined in (5.7).

Proof. As in the proof above we see that for any sequence $(i_1,\ldots,i_k)$ in $\{1,\ldots,p\}^k$, the inequality

$$|A(e_{i_1},\ldots,e_{i_k})| \leq \left(2^kk!\right)^{-1}2^kk^kM_k(A) \tag{5.16}$$

holds. From Federer (1969, Section 1.10.6), applied to the case of an inner product norm on V, for which (5.1) holds, it follows that

$$||A||^2 \leq \sum_i |A(e_{i_1}, \ldots, e_{i_k})|^2, \tag{5.17}$$

where $\sum_i$ is the sum over all sequences $(i_1, \ldots, i_k)$ in $\{1, \ldots, p\}^k$. Since the number of terms in this sum is p^k, the result is obtained by insertion of (5.16) into (5.17). The inequality in (5.15) is an application of (5.3). ∎