PREPRINT

SEPT 1971

3

Jørgen Larsen

Estimation in Exponential Families

UNIVERSITY OF COPENHAGEN
INSTITUTE OF MATHEMATICAL STATISTICS

Jørgen Larsen

ESTIMATION IN EXPONENTIAL FAMILIES

Preprint 1971 No. 3

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

September 1971

1. Introduction and summary.

In this paper the existence and uniqueness of maximum likelihood estimators in exponential families is discussed, and an example demonstrates a method of extending discrete models so that maximum likelihood estimation always is possible.

The maximization of the likelihood function $L(\cdot,t_0)$ is equivalent to the minimization of $-\log L(\cdot,t_0)$ which is convex. Therefore a result about minimization of l.s.c. quasi-convex functions is presented in section 2, using some elementary results from the theory of convex sets. In section 3 concepts as polar cone and the support of a measure are presented. Section 4 contains the main result: a necessary and sufficient condition for t_0 so that the maximum likelihood estimator $\hat{\theta}(t_0)$ exists. - Barndorff-Nielsen (1970) has given a comprehensive discussion of estimation in exponential families using convex duality theory.

In section 5 the logistic dose-response model is considered as an example, and we deduce how to extend the model so that estimation always is possible. - Barndorff-Nielsen (1970) discusses the same example (and the problem in general), and explains the extension in a different way.

M Davis (1970) has dealt with the estimation problems in the logistic model in a way that has given some of the inspiration to this paper.

2. Convex sets. Recession cone. Quasi-convex functions.

In this section E denotes a finite dimensional real Banach space.

2.1 Definition: For any subset $M \subseteq E$,

$$\mathrm{aff}M$$

denotes the smallest affine subspace in E containing M, i.e. the affine hull of M.

2.2 Definition: For any convex subset $A \subseteq E$ we define the relative interior of A,

$$\mathrm{ri}A,$$

as the interior of A considered as a subset of affA, which should be endowed with the subspace topology. (Since affA is a closed subset of E, the closure $\bar{A}$ of A in E coincides with the closure of A in the subspace topology in affA).

The convex subsets of E have the following important property.

2.3 Proposition: Let A be a convex subset of E. If $x \in \mathrm{ri}A$ and $y \in \bar{A}$, then

$$\{x + \lambda(y-x) \mid \lambda \in [0,1[\} \subseteq \mathrm{ri}A.$$

The proof will not be given here; see e.g. Bourbaki (1966), ch. II, Rockafellar (1970), Stoer and Witzgall (1970).

2.4 Definition: Let $x,y \in E$. The ray from the point x in direction y is the set

$$\mathrm{ray}(x,y) := \{x + \lambda y \mid \lambda \in [0, +\infty[\}.$$

2.5 Definition: Let A be a convex subset of E. The recession cone of A is the set

$$0^+A := \{y \in E \mid \forall x \in A : \mathrm{ray}(x,y) \subseteq A\}.$$

2.6 Proposition: Let A be a convex subset of E. Then the following properties hold:

(i). 0^+A is a convex cone.
If A is closed then 0^+A is closed.

(ii). For any $x \in A$ we have $0^+A \subseteq \{y \in E \mid ray(x,y) \subseteq A\}$. Furthermore, if $x \in riA$ then

$$\{y \in E \mid ray(x,y) \subseteq A\} = 0^+\bar{A} = 0^+riA.$$

(iii). Suppose also B is a convex subset of E.

If $\bar{A} \subseteq \bar{B}$ then $0^+\bar{A} \subseteq 0^+\bar{B}$.

If $riA \subseteq riB$ then $0^+riA \subseteq 0^+riB$.

Proof: It is easy to verify (i), and the first statement of (ii) follows from 2.5. Since (iii) follows from (ii) choosing an x from $ri\bar{A} \subseteq ri\bar{B}$ or from $riA \subseteq riB$, it therefore remains to show that

$$\{y \in E \mid ray(x,y) \subseteq A\} = 0^+\bar{A} = 0^+riA$$

for any convex subset $A \subseteq E$ and any $x \in riA$. First we shall show that

$$(*) \qquad \{y \in E \mid ray(x,y) \subseteq \bar{A}\} \subseteq 0^+riA.$$

Let $y \in \{y \in E \mid ray(x,y) \subseteq \bar{A}\}$, $z \in riA$; since $\bar{A}$ is convex and $ray(x,y) \subseteq \bar{A}$,

$$z + \frac{1}{n}(x-z) + \lambda y = (1 - \frac{1}{n})z + \frac{1}{n}(x + n\lambda y) \in \bar{A}$$

for every $n \in \mathbb{N}$ and $\lambda \in [0,+\infty[$. Letting $n \to \infty$ we see that $z + \lambda y \in \bar{A}$ for every $\lambda \in [0,+\infty[$, and hence $ray(z,y) \subseteq \bar{A}$. As $z \in riA$, 2.3 shows that $ray(z,y) \subseteq riA$, and (*) is thus established.

Next we show that

$$(**) \qquad 0^+riA \subseteq 0^+\bar{A}.$$

To this end, we consider $y \in 0^+riA$ and $w \in \bar{A}$. Choosing $x \in riA$ and putting $x_n = x + \frac{n-1}{n}(w-x)$, $(n \in \mathbb{N})$, $(x_n)_{n \in \mathbb{N}}$ is a sequence on riA (prop. 2.3) converging to w. Therefore $ray(x_n,y) \subseteq riA \subseteq A$ and hence $x_n + \lambda y \in A$ for every $n \in \mathbb{N}$ and $\lambda \in [0,+\infty[$. Letting $n \to \infty$ we see that $w + \lambda y \in \bar{A}$ for every $\lambda \in [0,+\infty[$, so that $ray(w,y) \subseteq \bar{A}$. Since this holds for every $w \in \bar{A}$, (**) is proved.

Using (**), the first statement of (ii), and (*), we obtain for any $x \in riA$

$$0^+riA \subseteq 0^+\bar{A}$$
$$\subseteq \{y \in E \mid ray(x,y) \subseteq \bar{A}\},$$

and

$$0^+\mathrm{ri}A \subseteq \{y \in E \mid \mathrm{ray}(x,y) \subseteq \mathrm{ri}A\}$$
$$\subseteq \{y \in E \mid \mathrm{ray}(x,y) \subseteq A\}$$
$$\subseteq \{y \in E \mid \mathrm{ray}(x,y) \subseteq \bar{A}\}$$
$$\subseteq 0^+\mathrm{ri}A,$$

which gives the desired results.

The following two propositions show how some topological properties of convex sets can be described by means of rays.

2.7 Proposition: Let A be a convex subset of E, and let $x \in \mathrm{ri}A$. Then A is closed, if (and only if) all the sets

$$A \cap \mathrm{ray}(x,y), \quad y \in \mathrm{aff}A - \mathrm{aff}A,$$

are closed.

Proof (of "if"): Applying 2.3 we have for $z \in \bar{A}$:

$$z \in \overline{\{x + \lambda(z-x) \mid \lambda \in [0,1[\}}$$
$$\subseteq \overline{\mathrm{ri}A \cap \mathrm{ray}(x,z-x)}$$
$$\subseteq \overline{A \cap \mathrm{ray}(x,z-x)}$$
$$= A \cap \mathrm{ray}(x,z-x) \subseteq A.$$

2.8 Proposition: Let A be a convex subset of E, and let $x \in \mathrm{ri}A$. Then the following three statements are equivalent:

(i). A is bounded (i.e. $A \subseteq \{w \in E \mid ||w|| \leq \lambda\}$ for some $\lambda \in \mathbb{R}_+$).

(ii). $0^+\bar{A} = \{\underline{0}\}$.

(iii). All the sets

$$A \cap \mathrm{ray}(x,y), \quad y \in \mathrm{aff}A - \mathrm{aff}A,$$

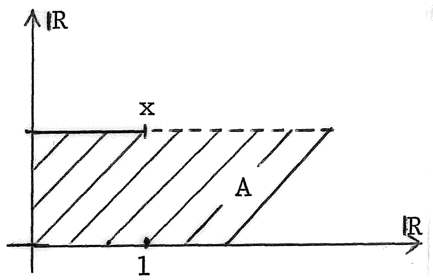
are bounded.

Proof: The equivalence (ii) $\Leftrightarrow$ (iii) is a consequence of 2.6 (ii). Since (i) $\Rightarrow$ (iii) is obvious, it remains to show that (iii) $\Rightarrow$ (i).

Suppose that A is unbounded. Then there exists a sequence $(y_n)_{n \in \mathbb{N}}$ of unit vectors with $x + n\, y_n \in A$, $\forall n \in \mathbb{N}$. As the unit ball (in affA - affA) is compact it contains a clusterpoint y for (y_n). It is easily seen that $\mathrm{ray}(x,y) \subseteqq \bar{A}$ and, by 2.3, that $\mathrm{ray}(x,y) \subseteqq \mathrm{ri}A \subseteqq A$. Thus the set $A \cap \mathrm{ray}(x,y)$ is unbounded.

<u>2.9 Example</u>, demonstrating the importance in 2.7 and 2.8 for x to be a point from $\mathrm{ri}A$.

Suppose that $E = \mathbb{R}^2$, $A = ([0,+\infty[\times [0,1[) \cup ([0,1] \times \{1\})$:



Choosing

$$x = (0,1) \in A \setminus \mathrm{ri}A$$

we have that all the sets

$$A \cap \mathrm{ray}(x,y) \quad , \quad y \in \mathbb{R}^2$$

are closed and bounded, although A is neither closed nor bounded. It is seen that

$$0^+A = \{\underline{0}\},$$

$$0^+\bar{A} = [0,+\infty[\times \{0\}.$$

If $B = [0,+\infty[\times [0,1[$, we have $B \subset A$,

$$0^+B = 0^+\bar{B} = [0,+\infty[\times \{0\},$$

that is $0^+A \subset 0^+B$.

<u>2.10 Lemma</u>: Let $\grave{F}$ be a filterbase of closed path-connected subsets of E, and let M be the set of clusterpoints of $\grave{F}$:

$$M = \bigcap_{F \in \grave{F}} F.$$

Then M is bounded and $\neq \emptyset$, if and only if $\grave{F}$ contains a bounded set F_0.

<u>Proof</u>: If $F_0 \in \grave{F}$ is bounded, $\{F_0 \cap F \mid F \in \grave{F}\}$ is a filterbase on the compact set F_0 and hence

$$\emptyset \neq \bigcap_{F \in \grave{F}} (F_0 \cap F) \subseteqq M \subseteqq F_0,$$

that is, M is bounded and $\neq \emptyset$.

Suppose next that M is bounded and $\neq \emptyset$:

$$\emptyset \neq M \subseteq \{x \in E \mid ||x|| < \lambda\}.$$

The set $K := \{x \in E \mid ||x|| = \lambda\}$ is compact, and to each $x \in K$ we can find $F_x \in \grave{F}$ so that $x \in F_x^c$. Since $(F_x^c, x \in K)$ constitutes an open covering of K, there exists a finite subset K_0 of K with

$$\bigcap_{x \in K_0} F_x \subseteq K^c,$$

and the filterbase axioms give the existence of $F_0 \in \grave{F}$ so that

$$F_0 \subseteq \bigcap_{x \in K_0} F_x \subseteq K^c;$$

the set F_0 being path-connected, this implies that

$$F_0 \subseteq \{x \in E \mid ||x|| < \lambda\},$$

so F_0 is bounded.

We shall now introduce the quasi-convex functions, and using the preceeding results it is possible to prove a result concerning minimization of quasi-convex lower semicontinuos functions.

<u>2.11 Definition</u>: A function $f : E \to \mathbb{R} \cup \{+\infty\}$ is called <u>quasi-convex</u> if all the sets

$$f^{-1}(]-\infty, a]) = \{x \in E \mid f(x) \leq a\}, \; a \in \mathbb{R},$$

are convex.

A function $f : D \to \mathbb{R}$, $D \subseteq E$, is called quasi-convex if the function $\tilde{f} : E \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if} \quad x \in D \\ +\infty & \text{if} \quad x \notin D \end{cases}$$

is quasi-convex.

<u>2.12 Examples</u>: Every convex function $f : D \to \mathbb{R}$, where $D \subseteq E$ is convex, is quasi-convex on E.

For any quasi-convex function $f : \mathbb{R} \to \mathbb{R}$ there exists $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}$, $a \leq b$, so that f is decreasing on $\{x \in \mathbb{R} \mid x \leq b\}$ and increasing on

$\{x \in \mathbb{R} \mid a \leq x\}$. Furthermore, if $f : \mathbb{R} \to \mathbb{R}$ satisfies the latter conditions, f is quasi-convex.

Recall that a function $f : E \to \mathbb{R} \cup \{-\infty,+\infty\}$ is called <u>lower semi-continuous</u> (l.s.c.) if and only if all the sets

$$f^{-1}(]-\infty,a]) = \{x \in E \mid f(x) \leq a\},\quad a \in \mathbb{R},$$

are closed.

<u>2.13 Proposition</u>: Sufficient for a quasi-convex function $f : E \to \mathbb{R} \cup \{+\infty\}$ to be l.s.c. is that for every $a \in \mathbb{R}$ and every pair $(x,y) \in E^2$ the following holds:

$$\text{if } f(x + \lambda y) \leq a \text{ for every } \lambda \in [0,1[,$$

$$\text{then } f(x+y) \leq a.$$

<u>Proof</u>: This is an immediate consequence of proposition 2.7 with $A := f^{-1}(]-\infty,a])$.

<u>2.14 Proposition</u>: For a quasi-convex l.s.c. function $f : E \to \mathbb{R} \cup \{+\infty\}$ the following three statements are equivalent:

(i): The minimum set

$$M := \{x \in E \mid f(x) = \inf_{y\in E} f(y)\}$$

is compact and $\neq \emptyset$.

(ii): There exists an $a \in \mathbb{R}$ so that to every $y \in E\setminus\{\underline{0}\}$ there exists an $x \in f^{-1}(]-\infty,a])$ with $\lim_{\rho\to+\infty} f(x + \rho y) > a$,

$\Leftrightarrow$:

$$\exists a \in \mathbb{R}\ \forall y \in E\setminus\{\underline{0}\}\ \exists x \in f^{-1}(]-\infty,a]): \lim_{\rho\to+\infty} f(x + \rho y) > a.$$

(iii): There exists an $a \in \mathbb{R}$ so that for some (and then for any) $x \in \operatorname{ri} f^{-1}(]-\infty,a])$ $\lim_{\rho\to+\infty} f(x + \rho y) > a$ for every $y \in E\setminus\{\underline{0}\}$, $\Leftrightarrow$:

$$\exists a \in \mathbb{R}\ \exists(\forall)x \in \operatorname{ri} f^{-1}(]-\infty,a])\ \forall y \in E\setminus\{\underline{0}\}: \lim_{\rho\to+\infty} f(x + \rho y) > a.$$

<u>Proof</u>: We shall apply lemma 2.10 to the filterbase

$$\grave{F} := \{f^{-1}(]-\infty,a]) \mid a \in f(E)\setminus\{+\infty\}\},$$

which consists of closed convex sets, and it is seen that (i) is equivalent to the existence of an $a \in \mathbb{R}$ so that $f^{-1}(]-\infty,a])$ is bounded and $\neq \emptyset$. According to 2.8 ((i) $\Leftrightarrow$ (ii)), this can be expressed by saying that $f^{-1}(]-\infty,a])$ $\neq \emptyset$ and $0^+f^{-1}(]-\infty,a]) = \{\underline{0}\}$, i.e. (cfr. 2.5)

$$\exists a \in \mathbb{R}\ \forall y \in E\setminus\{\underline{0}\}\ \exists x \in f^{-1}(]-\infty,a]) : \mathrm{ray}(x,y) \nsubseteq f^{-1}(]-\infty,a])$$

The equivalence (i) $\Leftrightarrow$ (ii) is thus established by remarking, that if $f(x) \leqq a$ and $f(x + ry) > a$ $(r > 0)$ then $f(x + \rho y) > a$ for $\rho \geqq r$ since $f^{-1}(]-\infty,a])$ is convex.

In a completely analogous way (i) $\Leftrightarrow$ (iii) is proved, using 2.8 ((i) $\Leftrightarrow$ (iii)).

3. Dual space. Polar cone. The support of a measure.

We are still considering a finite dimensional real Banach space E. The totality of all (continuous) linear forms on E is E^*, the dual space of E, and constitutes a Banach space itself with the norm

$$x^* \mapsto ||x^*|| := \sup\{||x^*(x)|| \mid ||x|| \leq 1\},\ x^* \in E^*.$$

Furthermore E and E^* have the same dimension. Now, consider the natural embedding $\psi : E \to E^{**} = (E^*)^*$, where

$$\psi x : x^* \mapsto (\psi x)(x^*) = x^*(x),\ (x^* \in E^*,\ x \in E).$$

In our case ψ is an isometric isomorphism of E onto E^{**}.

In the sequel we shall use some well known results*):

3.1 The Hahn-Banach Theorem in its geometric formulation: Let A be an open convex subset of E and let M be an affine subspace of E, $M \cap A = \emptyset$. Then there exists a closed hyperplane H in E so that $M \subseteq H$ and $H \cap A = \emptyset$.

3.2 Separation Theorem: Let A be a closed convex subset of E and let B be a compact convex subset of E, $A \cap B = \emptyset$. Then there exists a closed hyperplane $H = (x^*)^{-1}(\gamma)$ in E ($x^* \in E^*$, $\gamma \in \mathbb{R}$) which separates A and B strictly, i.e. $x^*(A) \subseteq]-\infty,\gamma[$ and $x^*(B) \subseteq]\gamma,+\infty[$.

3.3 Corollary: Let C be a convex cone with vertex $\underline{0}$, $C \subset E$. Then $\bar{C}$ is the intersection of all closed halfspaces containing C whose boundary hyperplane contains $\underline{0}$.

3.4 Definition: Let A be a convex subset of F, F = E or E^*. The polar cone of A is the set

$$A^o := \{y^* \in F^* \mid \forall y \in A : y^*(y) \leq 0\}.$$

In the case $F = E^*$ we shall use the notation

$$A^p := \psi^{-1}(A^o) = \{x \in E \mid \forall x^* \in A : x^*(x) \leq 0\}.$$

If M is a subset of E then cone M denotes the smallest cone with vertex $\underline{0}$ containing M: cone $M = \{\lambda x \mid x \in M,\ \lambda \in \mathbb{R}_+\}$. If M is convex, then cone M is convex.

---

*) see e.g. Bourbaki (1966).

<u>3.5 Proposition</u>: Let A be a convex subset of $F = E$ or E^*. If $F = E$, then $A^o = (\text{cone } A)^o$ is a closed convex cone, and $\overline{\text{cone } A} = \psi^{-1}(A^{oo}) = A^{op}$. If $F = E^{**}$, then $A^o = (\text{cone } A)^o$ and $A^p = (\text{cone } A)^p$ are closed convex cones, and $\overline{\text{cone } A} = [\psi^{-1}(A^o)]^o = A^{po}$.

<u>Proof</u>: It is rather trivial that $A^o = (\text{cone } A)^o$ and $A^p = (\text{cone } A)^p$ are closed convex cones. The assertions about $\overline{\text{cone } A}$ are obvious if $C := \text{cone } A = F$, and if $C \subset F$ they are simply reformulations of corollary 3.3; note however the importance in the case $F = E^*$ of ψ being surjective.

Let $\grave{X}$ be a locally compact space (e.g. $\grave{X} = E$). The <u>Borel-σ-algebra</u> on $\grave{X}$, $\grave{B}(\grave{X})$, is the σ-algebra generated by the open sets in $\grave{X}$.

<u>3.6 Definition</u>: Given a positive σ-finite measure m on $(\grave{X}, \grave{B}(\grave{X}))$. The <u>support</u> of m is the set

$$\text{supp}(m) := \{x \in \grave{X} \mid m(U) > 0 \text{ for every open neighbourhood } U \text{ of } x\}.$$

If $\grave{X} = E$ we define the <u>affine support</u> of m as the set

$$S(m) := \text{aff}(\text{supp}(m)).$$

We note that $\text{supp}(m)^c$ is open. If $\grave{X} = E$ it follows from 3.7 that $\text{supp}(m)$ is the largest set $M \subseteq \grave{X}$ so that $m(U) > 0$ for every non empty relative open subset U of M, and $\text{supp}(m)^c$ is the largest open set $N \subseteq \grave{X}$ so that $m(N) = 0$.

<u>3.7 Proposition</u>: Let m be as in 3.6 with $\grave{X} = E$. For every $B \in \grave{B}(E)$ we have

$$B \cap \text{supp}(m) = \emptyset \quad \Rightarrow \quad m(B) = 0.$$

<u>Proof</u>: Choose open sets U_x, $x \in B$, so that $x \in U_x$ and $m(U_x) = 0$. Since E is a Lindelöf space one can find a countable subset B_0 of B with

$$\bigcup_{x \in B} U_x = \bigcup_{x \in B_0} U_x \supseteq B;$$

hence
$$m(B) \leq \sum_{x \in B_0} m(U_x) = 0.$$

<u>3.8 Theorem</u>: Given a positive σ-finite measure m on $(E, \grave{B}(E))$, and a closed convex cone $K \subseteq (S(m) - S(m))^*$ with vertex $\underline{0}$. Then we have the identity

$$\text{ri}(\text{conv}(\text{supp}(m))) + K^p$$
$$= \{x \in S(m) \mid \forall x^* \in K \setminus \{\underline{0}\}: m\{y \in S(m) \mid x^*(y-x) > 0\} > 0\},$$

where K is considered as a subset of $F = (S(m) - S(m))^*$ and thus

$$K^p = \{z \in S(m) - S(m) \mid \forall x^* \in K : x^*(z) \leqq 0\}.$$

<u>Proof</u>: For shortness we put

$$A := ri(conv(supp(m))) + K^p$$

$$B := \{x \in S(m) \mid \forall x^* \in K \setminus \{\underline{0}\} : m\{y \in S(m) \mid x^*(y-x) > 0\} > 0\}.$$

First we show that $A^c \subseteqq B^c$: The set A is convex and open (relative to S(m)), so according to the Hahn-Banach theorem we can to $x \in A^c$ find $x^* \in (S(m) - S(m))^*$ with

(1) $$A \subseteqq \{y \in S(m) \mid x^*(y-x) < 0\}.$$

Using proposition 2.6, (iii), we obtain

$$\begin{aligned} K^p = 0^+(K^p) &\subseteqq 0^+A \\ &\subseteqq 0^+\{y \in S(m) \mid x^*(y-x) < 0\} \\ &= 0^+\{z \in S(m) - S(m) \mid x^*(z) < 0\} \\ &\subseteqq \{z \in S(m) - S(m) \mid x^*(z) \leqq 0\}, \end{aligned}$$

which shows that

$$x^* \in K^{po} = \bar{K} = K$$

(proposition 3.5). Inclusion (1) implies that

$$conv(supp(m)) \subseteqq \{y \in S(m) \mid x^*(y-x) \leqq 0\},$$

so

$$supp(m) \cap \{y \in S(m) \mid x^*(y-x) > 0\} = \emptyset,$$

that is,

$$m\{y \in S(m) \mid x^*(y-x) > 0\} = 0$$

(proposition 3.7), and hence $x \notin B$.

To see that $B^c \subseteqq A^c$ it sufficies to show that

$$x - z \notin ri(conv(supp(m)))$$

for all $x \in B^c$, $z \in K^p$. If $x \in B^c$ then

$$m\{y \in S(m) \mid x^*(y-x) > 0\} = 0$$

for some $x^* \in K$, and hence for $z \in K^p$

$$0 = m\{y \in S(m) \mid x^*(y-x) > 0\}$$
$$= m\{y \in S(m) \mid x^*(y-(x-z)) > x^*(z)\}$$
$$\geq m\{y \in S(m) \mid x^*(y-(x-z)) > 0\}$$

since $x^*(z) \leq 0$. It follows from 3.6 that

$$\text{supp}(m) \cap \{y \in S(m) \mid x^*(y-(x-z)) > 0\}$$

is empty, since it is a relative open subset of supp(m) with m-measure 0. This implies that

$$\text{ri}(\text{conv}(\text{supp}(m)))$$
$$\subseteq \text{ri}(\text{conv}\{y \in S(m) \mid x^*(y-(x-z)) \leq 0\})$$
$$= \{y \in S(m) \mid x^*(y-(x-z)) < 0\}$$

so

$$x-z \notin \text{ri}(\text{conv}(\text{supp}(m))).$$

Sometimes the following concept may be usefull:

<u>3.9 Definition</u>: Let A be convex subset of $F = E$ or E^*. The <u>normal cone</u> of A at $x \in A$ is

$$(A-x)^{\circ} = \{y^* \in F^* \mid \forall y \in A : \; y^*(y-x) \leq 0\}.$$

We shall use normal cones in section 5.

4. Maximum likelihood estimation in exponential families.

In this section we consider

V, a finite dimensional real Banach space with the Borel-σ-algebra $\grave{B}(V)$,

$(\grave{X},\grave{A})$, an arbitrary measurable space,

T, a measurable mapping $\grave{X} \to V$,

μ, a σ-finite positive measure on $(\grave{X},\grave{A})$,

S is the affine support of the measure $B \mapsto \mu(T^{-1}B)$, $B \in \grave{B}(V)$, and $\grave{B}(S)$ is the Borel-σ-algebra on S,

μ_T is the measure on $(S, \grave{B}(S))$ given by $\mu_T(B) = \mu(T^{-1}B)$.

$S_0 = S-S$

T_0 is an arbitrary fixed point i S.

We suppose, that the set

$$\Theta := \{\theta \in S_0^* \mid \int \exp(\theta(Tx - T_0))\ \mu(dx) < +\infty\}$$

is non empty. (Θ does not depend on the choice of T_0). In this case the measure μ_T is σ-finite.

Consider the exponential family

$$\grave{P} = \{P_\theta \mid \theta \in \Theta\}$$

of probability measures on $(\grave{X},\grave{A})$ defined by

$$\frac{dP_\theta}{d\mu}(x) = \frac{\exp(\theta(Tx - T_0))}{\int \exp(\theta(Ty - T_0))\ \mu(dy)}$$

(not depending on T_0). Corresponding to $\grave{P}$ we have the family

$$\grave{P}_T = \{P_{\theta,T} \mid \theta \in \Theta\}$$

of probability measures on $(S,\grave{B}(S))$ defined by

$$\frac{dP_{\theta,T}}{d\mu_T}(t) = \frac{\exp(\theta(t - T_0))}{\int \exp(\theta(s - T_0))\ \mu_T(ds)}.$$

(The measure $P_{\theta,T}$ is obtained from P_θ in the same way as μ_T from μ).

The likelihood function is

$$L : \Theta \times S \to [\,0,+\infty[$$

$$(\theta,t) \mapsto \frac{\exp(\theta(t - T_0))}{\int \exp(\theta(s - T_0))\, \mu_T(ds)} .$$

Now, suppose we have an observation $t_0 = Tx_0 \in S$ and that we want to estimate the parameter θ under the hypothesis $\theta \in H$, where H should be a closed convex subset of S_0^* and $H \cap \Theta \neq \emptyset$. The principle of maximum likelihood estimation then tells us to maximize $L(\theta,t_0)$ with respect to $\theta \in \Theta$, subject to the constraint $\theta \in H$, and if there exists a unique value $\hat\theta_H(t_0) \in \Theta \cap H$ with

$$L(\hat\theta_H(t_0),t_0) = \sup_{\theta \in \Theta \cap H} L(\theta,t_0)$$

then to use $\hat\theta_H(t_0)$ as an estimator for θ.

4.1 Theorem: Under the above conditions, it is necessary and sufficient for the existence of $\hat\theta_H(t_0)$ that

$$t_0 \in \operatorname{ri}(\operatorname{conv}(\operatorname{supp}(\mu_T))) + (0^+\, \overline{\Theta \cap H})^p,$$

(where the polar cone operation goes from S_0^* to S_0, cf. 3.4). In any case $L(\cdot,t_0)$ attains its supremum at at most one point in $\Theta \cap H$.

Moreover, the equation

$$E_\theta T = t_0$$

has at most one solution $\tilde\theta \in \Theta$, and if $\hat\theta_H(t_0) \in \operatorname{ri}(\Theta \cap H)$ then $\tilde\theta = \hat\theta_H(t_0)$.

Proof: For the sake of simplification we choose $T_0 = t_0$, so that

$$L(\theta,t_0) = \Phi(\theta)^{-1},$$

$$\Phi(\theta) := \Phi(\theta,t_0) := \int \exp(\theta(t - t_0))\, \mu_T(dt) \quad , \quad \theta \in \Theta.$$

It is a convenient trick to introduce the function

$$f_H : S_0^* \to \mathbb{R} \cup \{+\infty\}$$

$$f_H(\theta) := \begin{cases} \log \Phi(\theta) & \text{for} \quad \theta \in \Theta \cap H \\ +\infty & \text{for} \quad \theta \in S_0^* \setminus (\Theta \cap H). \end{cases}$$

Obviously, every minimum point for f_H will be a maximum point for $L(\cdot, t_0)$ restricted to $\Theta \cap H$, and vice versa. It is well known (see e.g. Barndorff-Nielsen (1970) or Johansen (1970)) that Θ is convex and $\log \Phi$ is strictly convex on Θ. Hence we conclude that f_H is quasi-convex and that f_H attains its infimum at at most one point (which will be $\hat{\theta}_H(t_0)$). The idea is now to show that f_H is l.s.c. using 2.13, and then to apply 2.14. First we shall find $\lim_{\rho \uparrow r} f_H(\theta + \rho\xi)$, $\theta, \xi \in S_0^*$, $r \in \,]0, +\infty]$.

Given $\theta \in \Theta \cap H$, $\xi \in S_0^* \setminus \{\underline{0}\}$, $r \in \,]0, +\infty]$, each of the following statements is either true or false:

$$\text{(st 1)}: r = +\infty$$

$$\text{(st 2)}: \{\theta + \rho\xi \mid \rho \in [0, r[\,\} \subseteq \Theta \cap H$$

$$\text{(st 3)}: \mu_T\{\xi(t - t_0) > 0\} > 0$$

$$\text{(st 4)}: \mu_T\{\xi(t - t_0) = 0\} > 0.$$

If the logical values of the statements are known we can find the desired limit:

case	(st 1)	(st 2)	(st 3)	(st 4)	$\lim_{\rho \uparrow r} f_H(\theta + \rho\xi)$
(1)		f			$+\infty$
(2)	f	t			$f_H(\theta + r\xi)$
(3)	t	t	t		$+\infty$
(4)	t	t	f	t	$-\log L(\theta, t_0 \mid \xi(T - t_0) = 0) = \log \int_{\{\xi(t - t_0) = 0\}} \exp(\theta(t - t_0))\mu_T(dt)$
(5)	t	t	f	f	$-\infty$

(t = true, f = false)

These results are, of course, obtained by rewriting $\log \Phi(\theta)$ as

$$\log\left(\int_{\{\xi(t - t_0) > 0\}} + \int_{\{\xi(t - t_0) \leq 0\}} \exp(\theta(t - t_0))\mu_T(dt) \right)$$

and using the monotone convergence theorem. Case (2) needs a little more attention: The limit is

$$\log \int \exp((\theta + r\xi)(t - t_0))\mu_T(dt) = \log \Phi(\theta + r\xi)$$

if the expression under the integral sign is integrable, and $+\infty$ otherwise. In the non-integrable case $\theta + r\xi \notin \Theta$ and hence $f_H(\theta + r\xi) = +\infty$. If the expression in fact is integrable, $\theta + r\xi \in \Theta$, and since $\{\theta + \rho\xi | \rho \in [0,r[\} \subseteq H$ and H is closed, $\theta + r\xi \in H$, and so $f_H(\theta + r\xi) = \log \Phi(\theta + r\xi)$.

In case (4) $L(\cdot,\cdot|\xi(T - t_0) = 0)$ is the likelihood function in the distribution of T conditional on $\xi(T - t_0) = 0$,

$$L(\omega,t_1|\xi(T - t_0) = 0) := \frac{\exp(\omega(t_1 - T_0))}{\int_{\{\xi(t - t_0) = 0\}} \exp(\omega(t - T_0))\ \mu_T(dt)}$$

for $t_1 \in \{t|\xi(t - t_0) = 0\}$, $\omega \in \Theta$.

For every $\theta \in \Theta \cap H$ and $\xi \in S_0^*$ we have

$$f_H(\theta + \xi) = \lim_{\lambda \uparrow 1} f_H(\theta + \lambda\xi)$$

(case (1) or case (2)), so proposition 2.13 gives that f_H is l.s.c.

We can now examine for which t_0's the condition 2.14 (iii) is fulfilled. Choose $a \in \mathbb{R}$, $\theta \in \text{ri} f_H^{-1}(]-\infty,a])$, $\xi \in S_0^* \setminus \{\underline{0}\}$. Then

$$\lim_{\rho \uparrow +\infty} f(\theta + \rho\xi) > a,$$

if and only if we are in case (1) or case (3) (in case (4)

$$\log \int_{\{\xi(t - t_0) = 0\}} \exp(\theta(t - t_0))\mu_T(dt) \leq \log \Phi(\theta) \leq a).$$

The following bi-implications hold (since $r = +\infty$):

[we are in case (1) or case (3)]

$\Leftrightarrow$ [((st 1) $\wedge$ (st 2) $\wedge$ (st 3)) $\vee$ non (st 2)]

$\Leftrightarrow$ [(st 2) $\Rightarrow$ (st 3)]

$\Leftrightarrow$ [$\text{ray}(\theta,\xi) \subseteq \Theta \cap H \Rightarrow \mu_T\{\xi(t - t_0) > 0\} > 0$]

$\Leftrightarrow$ [$\xi \in 0^+ \overline{\Theta \cap H} \Rightarrow \mu_T\{\xi(t - t_0) > 0\} > 0$]

according to 2.6 (ii). (Note the independence on a!)

We have thus found that condition 2.14 (iii) is fullfilled if and only if

$$\forall \xi \in S_0^* \setminus \{\underline{0}\} : [\xi \in 0^+ \overline{\Theta \cap H} \Rightarrow \mu_T\{\xi(t - t_0) > 0\} > 0].$$

Applying 3.8 it is seen to be equivalent to

$$t_0 \in \mathrm{ri}(\mathrm{conv}(\mathrm{supp}(\mu_T))) + (0^+ \overline{\Theta \cap H})^p.$$

This finishes the deduction of the criterion for existence of maximum likelihood estimators.

It is known (see e.g. Barndorff-Nielsen (1970) or Johansen (1970)) that the mapping $\theta \mapsto E_\theta T$ is injective and that

$$E_\theta T = t_0 + D \log \Phi(\theta)$$

for $\theta \in \mathrm{ri}\Theta$. If $\hat{\theta}_H(t_0) \in \mathrm{ri}(\Theta \cap H)$, it is a stationary point, i.e. $D \log \Phi(\hat{\theta}_H(t_0)) = \underline{0}$, and therefore solution to

$$E_\theta T = t_0.$$

This completes the proof of 4.1.

4.2 <u>Corollary</u>: The functions Φ and $\log\Phi$ $S_0^* \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\Phi(\theta) = \begin{cases} \int \exp(\theta(t - T_0))\mu_T(dt) & , \quad \theta \in \Theta \\ + \infty & , \quad \theta \in S_0^* \setminus \Theta, \end{cases}$$

are l.s.c.

<u>Proof</u>: $\Phi = f_H$, $H = S_0^*$, with f_H as in the proof of 4.1.

4.3 <u>Corollary</u>: The level sets

$$\{\theta \in S_0^* \mid -\log L(\theta, t_0) \leqq a\} \quad , \quad a > \inf(-\log L(\cdot, t_0)),$$

are all bounded or all unbounded.

<u>Proof</u>: In the proof of 4.1 we found a criterion for $\lim_{\rho \uparrow +\infty} f(\theta + \rho\xi)$ to be $>a$ for every $\xi \in S_0^* \setminus \{\underline{0}\}$, and this criterion did not depend on a. The result then follows from 2.8 (i, iii).

4.4 Corollary: Suppose $\grave{X}$ is a locally compact vector space and $\grave{A}$ the Borel-σ-algebra on $\grave{X}$, and suppose μ is discrete (i.e. supp(μ) is discrete in the subspace-topologi). If $t_0 \in \mathrm{conv}(\mathrm{supp}(\mu_T))$ and $\xi \in 0^+\mathrm{ri}(\Theta \cap H)$ so that

$$\mu_T\{t \mid \xi(t - t_0) > 0\} = 0$$

$$\mu_T\{t \mid \xi(t - t_0) = 0\} > 0,$$

then for $\theta_0 \in \mathrm{ri}(\Theta \cap H)$

$$P_{\theta_0 + \rho\xi} \xrightarrow{w} P_{\theta_0}(\cdot \mid \xi(T - t_0) = 0)$$

for $\rho \to +\infty$. Since $P_{\theta_0,T}\{t_0\} \leqq P_{\theta_0,T}(\{t_0\} \mid \xi(T - t_0) = 0)$, this implies that

$$\sup_{\theta \in \Theta \cap H} P_{\theta,T}\{t_0\} = \sup_{\theta \in \Theta \cap H} P_{\theta,T}(\{t_0\} \mid \xi(T - t_0) = 0).$$

Proof: This is an immediate consequence of the examinations of $\lim f(\theta + \rho\xi)$ in the proof of 4.1, since

$$P_\theta\{x\} = \exp(-f(\theta))\,\mu\{x\}, \quad f(\theta) = \log\Phi(\theta, Tx),$$

so that it is seen that

$$\lim_{\rho\to+\infty} P_{\theta_0 + \rho\xi}\{x\} = P_{\theta_0}(\{x\} \mid \xi(T - t_0) = 0)$$

for all $x \in \mathrm{supp}(\mu)$.

5. The dose-response model.

In this section we shall - as an example - discuss the estimation problems in the dose-response model.

5.1

Consider mutually independent random variables $X_1,\ldots,X_k$, so that X_i is binominally distributed with known number parameter $n_i \in \mathbb{N}$ and unknown probability parameter $p^{(i)} \in [0,1]$, $i = 1,\ldots,k$. Furthermore, $z_1 < \ldots < z_k$ are given real numbers. Now the statistical problem is obtained assuming that for some $\theta = (\alpha,\beta) \in \mathbb{R}^2$

$$p^{(i)} = p_\theta(z_i) := \frac{1}{1 + \exp(-\alpha-\beta z_i)}, \quad i = 1,\ldots,k.$$

Note, that the logistic function

$$p_\theta = p_{\alpha,\beta} : z \mapsto \frac{1}{1 + \exp(-\alpha-\beta z)} = \frac{\exp(\alpha+\beta z)}{1 + \exp(\alpha+\beta z)}, \quad z \in \mathbb{R},$$

for $\beta > 0$ is increasing, for $\beta = 0$ constant, and for $\beta < 0$ decreasing. For $\beta \neq 0$ $p_{\alpha,\beta}$ is a bijection from $\mathbb{R}$ to $]0,1[$; the inverse mapping of $p_{0,1}$ is

$$\lambda : u \mapsto \log \frac{u}{1-u}, \quad u \in]0,1[.$$

($\lambda(u)$ is sometimes called the logistic transform of u). Finally, $\{p_\theta(z) | \theta \in \mathbb{R}^2\} =]0,1[$ for every $z \in \mathbb{R}$.

5.2

The distribution of $X = (X_1,\ldots,X_k)$ when $p_\theta(z_1),\ldots,p_\theta(z_k)$ are the parameters is P_θ given by

$$P_\theta\{(x_1,\ldots,x_k)\} := P\{X_1 = x_1,\ldots,X_k = x_k\}$$

$$= \begin{cases} \dfrac{\exp(\alpha \sum x_i + \beta \sum z_i x_i)}{\prod_{i=1}^{k} [1 + \exp(\alpha+\beta z_i)]^{n_i}} \prod_{i=1}^{k} \binom{n_i}{x_i}, & \text{for } (x_1,\ldots,x_k) \in \prod_{i=1}^{k} \{0,1,\ldots,n_i\} \\ 0, & \text{for } (x_1,\ldots,x_k) \notin \prod_{i=1}^{k} \{0,1,\ldots,n_i\} \end{cases}$$

so we are concerned with an exponential family of order 2. Introducing the measure μ on $(\mathbb{R}^k, \mathcal{B}^k)$ with

$$\mu(\{x\}) = \begin{cases} \prod_{i=1}^{k} \binom{n_i}{x_i} & \text{for } x = (x_1,\ldots,x_k) \in \prod_{i=1}^{k} \{0,1,\ldots,n_i\} \\ 0 & \text{else} \end{cases},$$

and the functions $T : \mathbb{R}^k \to \mathbb{R}^2$ and $\Phi : \mathbb{R}^2 \to \mathbb{R}$:

$$T : x \mapsto (T_1x, T_2x) := (\sum_i x_i, \sum_i z_i x_i) = \sum_i x_i(1,z_i),$$

$$\Phi : \theta = (\alpha,\beta) \mapsto \int \exp(\alpha T_1 x + \beta T_2 x)\, \mu(dx)$$

$$= \prod_{i=1}^{k} [1 + \exp(\alpha + \beta z_i)]^{n_i},$$

we have

$$\frac{dP_\theta}{d\mu}(x) = \frac{\exp(\theta \cdot Tx)}{\Phi(\theta)}, \quad \theta \in \mathbb{R}^2,$$

(where $\cdot$ denotes the ordinary inner product in $\mathbb{R}^2$), to be compared with the family of section 4.

The support of the transformed measure μ_T is

$$\text{supp}(\mu_T) = T(\text{supp}(\mu))$$

$$= \{\sum x_i(1,z_i) \mid x_i \in \{0,1,\ldots,n_i\},\ i=1,\ldots,k\},$$

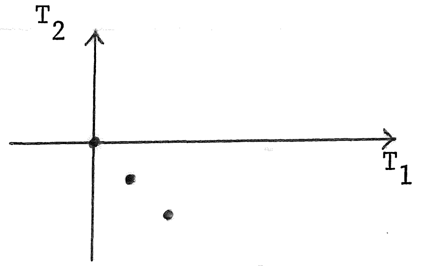
and the convex hull of $\text{supp}(\mu_T)$ is a (convex) polygon with 2k sides and the corners

$$(0,0), \quad \sum_{i=j}^{k} n_i(1,z_i), \quad \sum_{i=1}^{j} n_i(1,z_i), \quad j=1,2,\ldots,k.$$

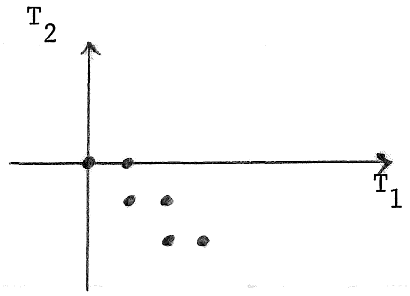
Since $(0^+ \mathbb{R}^2)^p = (\mathbb{R}^2)^p = \{(0,0)\}$, it follows from theorem 4.1 that the maksimum likelihood estimator $\hat{\theta}(t_0)$, $t_0 = Tx_0$, exists if and only if t_0 belongs to the interior of the polygon $\text{conv}(\text{supp}(\mu_T))$.

5.3

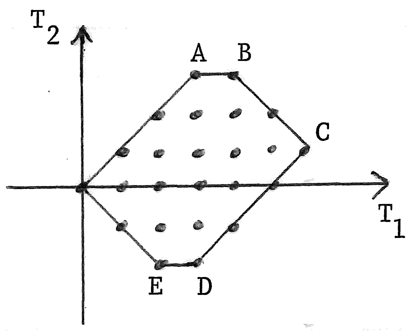
Example: The case $k = 3$; $n_1 = 2$, $n_2 = 1$, $n_3 = 3$; $z_1 = -1$, $z_2 = 0$, $z_3 = 1$.



$$\{x_1(1,-1) \mid x_1 = 0,1,2\}$$



$$\{x_1(1,-1) \mid x_1 = 0,1,2\}$$
$$+ \{x_2(1,0) \mid x_2 = 0,1\}$$



$$\{x_1(1,-1) \mid x_1 = 0,1,2\}$$
$$+ \{x_2(1,0) \mid x_2 = 0,1\}$$
$$+ \{x_3(1,1) \mid x_3 = 0,1,2,3\}$$
$$= \{ \sum_{i=1}^{3} x_i(1,z_i) \mid x_i=1,\ldots,n_i\}$$

A: $n_3(1,z_3)$

B: $n_3(1,z_3) + n_2(1,z_2)$

C: $n_3(1,z_3) + n_2(1,z_2) + n_1(1,z_1)$

D: $n_1(1,z_1) + n_2(1,z_2)$

E: $n_1(1,z_1)$.

5.4

It is possible in an explicit manner to describe the observations x_0 leading to a $t_0 = Tx_0$ on the boundary of $\mathrm{conv}(\mathrm{supp}(\mu_T))$, since the corners are known; $t_0 = Tx_0$ is a boundary point if and only if x_0 is of the form

$$(0,\ldots,0, x_{0j}, n_{j+1},\ldots,n_k)$$

or

$$(n_1,\ldots,n_{j-1}, x_{0j}, 0,\ldots,0),$$

where $x_{0j} \in \{0,1,\dots,n_j\}$, $j \in \{1,\dots,k\}$, and t_0 is a corner point if and only if also $x_{0j} \in \{0,n_j\}$.

Indeed, it is quiet an inconsistent behavior to want to estimate the parameter θ by the maximum likelihood method when it is possible to get observations x_0 with $P_\theta(\{x_0\}) > 0$, $\forall\theta \in \mathbb{R}^2$, so that no maximum likelihood estimator exists. Therefore an extension of the model is needed.

In the case of $t_0 = Tx_0 \in \mathrm{supp}(\mu_T)$ on the boundary, it turns out that for some sequences $(\theta_n)_{n\in\mathbb{N}}$ so that the likelihood function converges to its supremum (and therefore (θ_n) converges to infinity), the corresponding sequence of logistic functions (p_{θ_n}) converges to a kind of a degenerate logistic function which fits the observed x_{0i}-values perfectly; this has been discussed by Silverstone (1957). One might then use the family of degenerate and ordinary logistic functions as a parametrization of an extended model.

Passing to polar coordinates for θ and allowing the module to be $+\infty$ is another convenient method, which has been used by Davis (1970), who also notices that the t_0's giving rise to nonsolvable maximum likelihood equations are those on the boundary of $\mathrm{conv}(\mathrm{supp}(\mu_T))$.

A different approach has been made by Barndorff-Nielsen (1970), who proves a general result about extending certain types of exponential families; Barndorff uses the mean value parametrization, which makes many things very nice.

Here we shall proceed in the following way. Since a parametrization of our family $\grave{P} := \{P_\theta | \theta \in \mathbb{R}^2\}$ - from a mathematical point of view - just serves to define the subset $\grave{P}$ of the set of all probability measures on $\mathbb{R}^k$, let us for a while reformulate our problem of estimating θ to a problem of estimating a probability measure from $\grave{P}$. If we want to extend $\grave{P}$ in order to make maximum likelihood estimation, it is good to have a non parametric definition of "likelihood function". Obviously we can use the function

$$Q \mapsto Q\{T = t_0\} = Q_T\{t_0\}.$$

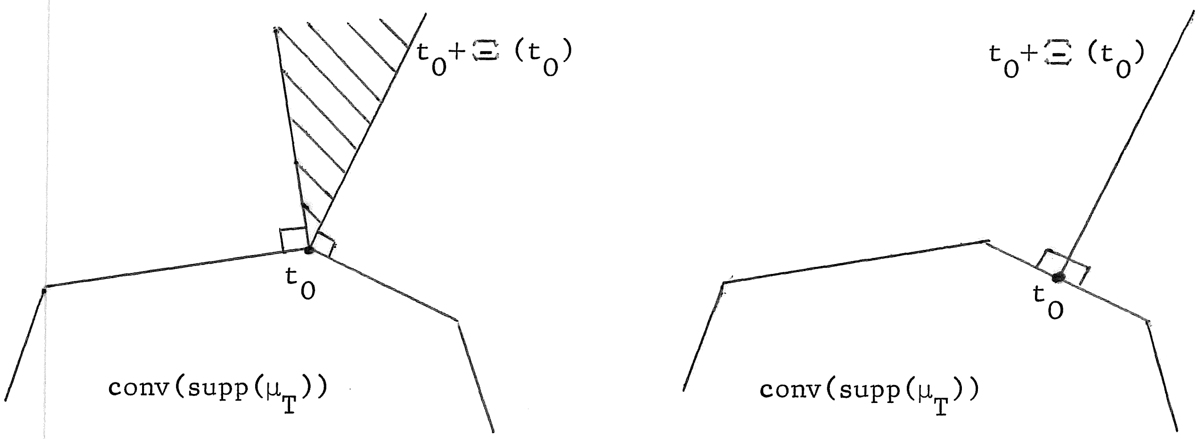
We are now able to find the smallest family of probability measures on $\mathbb{R}^k$ containing $\grave{P}$, so that maximum likelihood estimation always is possible.

5.6

If t_0 is a boundary point of $\mathrm{conv}(\mathrm{supp}(\mu_T))$ let $\Xi(t_0)$ denote the normal cone of $\mathrm{conv}(\mathrm{supp}(\mu_T))$ at t_0, i.e.

$$\Xi(t_0) := \{\xi = (\xi_1,\xi_2) \in \mathbb{R}^2 \mid \forall t \in \mathrm{supp}(\mu_T) : \xi\cdot(t-t_0) \leqq 0\}.$$

$\Xi(t_0)$ is a closed convex cone with vertex $\underline{0}$.



Let x_0 be an observed value, $t_0 = Tx_0$. If t_0 belongs to the relative interior of $\mathrm{conv}(\mathrm{supp}(\mu_T))$ we can (as already mentioned) solve the estimation problem. If t_0 is a boundary point then choose $\xi \in \Xi(t_0)\setminus\{\underline{0}\}$; from 4.4 we know that for $\theta \in \mathbb{R}^2$ the conditional distribution $P_\theta(\cdot\,|\,\xi\cdot(T - t_0) = 0)$ belongs to the closure $\bar{\grave{P}}$ of $\grave{P}$ and that

$$\sup_{Q \in \grave{P}} Q_T\{t_0\} = \sup_{Q \in \grave{P}} Q_T(\{t_0\}\,|\,\xi\cdot(T - t_0) = 0)$$

Since the family

$$\grave{P}^\xi := \{Q(\cdot\,|\,\xi\cdot(T - t_0) = 0) \mid Q \in \grave{P}\}$$

of conditional distributions can be written in the same way as in section 4, - with the same statistic T and another measure μ^ξ, $\mu^\xi\{x\} = \mu\{x\}\cdot 1_{\{\xi\cdot(Tx - t_0) = 0\}}$, - we are in a similar situation as before: given $t_0 \in \mathrm{supp}(\mu_T^\xi)$ we seek the $Q \in \grave{P}^\xi$ that maximizes the likelihood function $Q_T\{t_0\}$; and again this problem is solvable if and only if t_0 is a point of the relative interior of $\mathrm{conv}(\mathrm{supp}(\mu_T^\xi))$. If

$t_0 \notin \mathrm{ri}(\mathrm{conv}(\mathrm{supp}(\mu_T^{\xi})))$ we choose an η from the normal cone of $\mathrm{conv}(\mathrm{supp}(\mu_T^{\xi}))$ at t_0 so that ξ and η are linearly independent, and form the conditional distributions

$$\grave{P}^{\xi,\eta} := \{Q(\cdot \mid \eta\cdot(T - t_0) = 0) \mid Q \in \grave{P}^{\xi}\}.$$

From 4.4 we know that $\grave{P}^{\xi,\eta} \subseteq \overline{P}$ and that

$$\sup_{Q \in \grave{P}} Q_T\{t_0\} = \sup_{Q \in \grave{P}^{\xi}} Q_T\{t_0\} = \sup_{Q \in \grave{P}^{\xi,\eta}} Q_T\{t_0\}.$$

The sequence $(\grave{P}, \grave{P}^{\xi}, \grave{P}^{\xi,\eta})$ is a sequence of families of decreasing order, and because $\grave{P}$ is of order 2 $\grave{P}^{\xi,\eta}$ will contain only one element so that estimation is trivial. This means that it is always possible to find $\hat{Q} \in \overline{P}$ so that

$$\hat{Q}_T\{t_0\} = \sup_{Q \in \grave{P}} Q_T\{t_0\}.$$

It is a reasonable demand to $\hat{Q}$ that it does not depend on our choise of ξ's and η's, and it is seen from the following investigation that this demand indeed is fulfilled.

<u>5.7</u>

Consider an observation x_0 with $t_0 = Tx_0$ on the boundary of $\mathrm{conv}(\mathrm{supp}(\mu_T))$. We will confine ourselves to the case

$$x_0 = (0,\dots,0, x_{0,k-1}, n_k),\ x_{0,k-1} \in \{1,2,\dots,n_{k-1}\},$$

$$t_0 = n_k\cdot(1,z_k) + x_{0,k-1}\cdot(1,z_{k-1}),$$

but the results are generalized in an obvious way.

If $x_{0,k-1} \in \{1,\dots,n_{k-1} - 1\}$, t_0 is not corner point (cf. 5.4), and the normal cone is

$$\Xi(t_0) = \{\sigma\cdot(-z_{k-1},1) \mid \sigma \in [\,0,+\infty[\,\}.$$

If $x_{0,k-1} = n_{k-1}$, t_0 is a corner point, and the normal cone is

$$\Xi(t_0) = \{\sigma\cdot(-z_{k-1},1) + \tau\cdot(-z_k,1) \mid \sigma,\tau \in [\,0,+\infty[\,\}.$$

The distribution of $X = (X_1,\dots,X_k)$ conditionally on $\xi\cdot(T-t_0) = 0$, $\xi \in \Xi(t_0)\setminus\{\underline{0}\}$, is easily found; there are two cases:

A: $\text{supp}(\mu_T) \cap \{t \mid \xi\cdot(t - t_0) = 0\} = \text{supp}(\mu_T^\xi) = \{t_0\}$.

This happens if and only if t_0 is a corner point ($x_{0,k-1} = n_{k-1}$) and $\xi = \sigma\cdot(-z_{k-1},1) + \tau\cdot(-z_k,1)$ for $\sigma,\tau \in \mathbb{R}_+$. For every $\theta = (\alpha,\beta) \in \mathbb{R}^2$

$$\hat{Q} = P_\theta(\cdot \mid \xi\cdot(T - t_0) = 0) = \text{the one-point distribution on } \mathbb{R}^2 \text{ at } x_0 = (0,\dots,0, n_{k-1}, n_k).$$

B: $\text{supp}(\mu_T) \cap \{t \mid \xi\cdot(t - t_0) = 0\} = \text{supp}(\mu_T^\xi) \supset \{t_0\}$.

Here we must distinguish between two cases:

B1: $\xi = \sigma\cdot(-z_{k-1},1) \quad , \quad \sigma \in \mathbb{R}_+$.

B2: $\xi = \tau\cdot(-z_k,1) \quad , \quad \tau \in \mathbb{R}_+$.

We shall only discuss B1; B2 can be treated in a completely analogous way.

B1: For every $\theta = (\alpha,\beta) \in \mathbb{R}^2$ and $x_{k-1} \in \{0,1,\dots,n_{k-1}\}$ we find

$$P_\theta(\{(0,\dots,0, x_{k-1}, n_k)\} \mid \xi\cdot(T - t_0) = 0)$$

$$= \binom{n_{k-1}}{x_{k-1}} p_\theta(z_{k-1})^{x_{k-1}} (1 - p_\theta(z_{k-1}))^{n_{k-1} - x_{k-1}}$$

and

$$\frac{dP_\theta(\cdot \mid \xi\cdot(T - t_0) = 0)}{d\mu^\xi}(x) = \left.\frac{\exp(\omega\cdot(Tx - T_0))}{\int \exp(\omega\cdot(Ty - T_0))\mu^\xi(dy)}\right|_{\omega = \omega_\theta} , x \in \mathbb{R}^k,$$

where

$$\mu^\xi\{x\} = \mu\{x\}\cdot 1_{\{\xi\cdot(Tx - t_0) = 0\}} \quad , x \in \mathbb{R}^k,$$

$$T_0 = n_k(1,z_k) \quad (\text{so that } Tx - T_0 = x_{k-1}\cdot(1,z_{k-1}),\ x \in \text{supp}(\mu^\xi))$$

$$\omega_\theta = \lambda(p_\theta(z_{k-1}))\cdot(1,0) = (\alpha + \beta z_{k-1})\cdot(1,0),$$

$$\omega = \gamma\cdot(1,0), \quad \gamma \in \mathbb{R} = \{\lambda(p_\theta(z_{k-1})) \mid \theta \in \mathbb{R}^2\}$$

(λ is the logistic transformation, see 5.1); thus

$$\frac{\exp(\omega\cdot(Tx - T_0))}{\int \exp(\omega\cdot(Ty - T_0))\mu^\xi(dy)}\mu^\xi\{x\} = \binom{n_{k-1}}{x_{k-1}} \frac{e^{\gamma x_{k-1}}}{(1 + e^\gamma)^{n_{k-1}}}$$

for $x \in \mathrm{supp}(\mu^{\xi})$, i.e. X_{k-1} is binomially distributed with parameters n_{k-1} and $e^{\gamma}/1+e^{\gamma}$.

Now $t_0 \in \mathrm{ri}(\mathrm{conv}(\mathrm{supp}(\mu_T^{\xi})))$ if and only if $x_{0,k-1} \in \{1,\dots,n_{k-1}-1\}$, and in this case the maximum likelihood estimator $\hat{\omega} = \hat{\gamma}\cdot(1,0)$ is of course given by

$$\hat{\gamma} = \lambda(\hat{p}), \quad \hat{p} = \frac{x_{0,k-1}}{n_{k-1}},$$

obtained from the relation

$$E_{\hat{\omega}}(T \mid \xi(T - t_0) = 0) = t_0 ,$$

that is,

$$n_k\cdot(1,z_k) + n_{k-1}\frac{e^{\hat{\gamma}}}{1+e^{\hat{\gamma}}}\cdot(1,z_{k-1}) = n_k\cdot(1,z_k) + x_{0,k-1}\cdot(1,z_{k-1});$$

consequently

$$\hat{Q}\{x\} = \frac{d\hat{Q}}{d\mu^{\xi}}(x)\mu^{\xi}\{x\} = \frac{\exp(\hat{\omega}\cdot(Tx - T_0))}{\int \exp(\hat{\omega}\cdot Ty - T_0))\mu^{\xi}(dy)}\mu^{\xi}\{x\}$$

$$= \begin{cases} \binom{n_{k-1}}{x_{k-1}}\hat{p}^{x_{k-1}}(1-\hat{p})^{n_{k-1}-x_{k-1}} & \text{if } \begin{cases} x_1 = \dots = x_{k-2} = 0, \\ x_{k-1} \in \{0,1,\dots,n_{k-1}\} \\ x_k = n_k \end{cases} \\ 0 & \text{else.} \end{cases}$$

If $x_{0,k-1} = n_{k-1}$ no $\hat{\omega}$ exists. In this case we shall choose an η from the normal cone of $\mathrm{conv}(\mathrm{supp}(\mu_T^{\xi}))$ so that ξ and η are linearly independent:

$$\eta \in \{\eta = (\eta_1,\eta_2) \in \mathbb{R}^2 \mid \eta_1 + \eta_2 z_{k-1} < 0\}$$

For any such η

$$\mathrm{supp}(\mu_T^{\xi}) \cap \{t \mid \eta\cdot(T - t_0) = 0\} = \{t_0\},$$

so we are in a situation similar to A; we find

$$\hat{Q} = P_{\theta}(\cdot \mid \xi\cdot(T - t_0) = 0,\ \eta\cdot(T - t_0) = 0)$$

$$= \text{the one-point distribution on } \mathbb{R}^k \text{ at } x_0 = (0,\dots,n_{k-1}, n_k)$$

for all $\theta \in \mathbb{R}^2$.

5.8

The distribution of $X = (X_1,\dots,X_k)$ is estimated as follows:

$X_1,\dots,X_k$ are independent binomially distributed with parameters $n_1,\dots,n_k \in \mathbb{N}^k$ (known) and $p^{(1)},\dots,p^{(k)} \in [0,1]$. The observation is x_0.
If $t_0 = Tx_0 \in ri(conv(supp(\mu_T)))$, then $p^{(i)} = p_{\hat\theta}(z_i)$, $i=1,\dots,k$; if t_0 is on the boundary of $conv(supp(\mu_T))$, then $p^{(i)} = \frac{x_{0i}}{n_i}$, $i=1,\dots,k$, (see 5.4). Thus if we put

$$P_p\{(x_1,\dots,x_k)\} := \prod_{i=1}^{k} p_i^{x_i}(1-p_i)^{n_i-x_i} \cdot \mu\{x\},\ x \in \mathbb{R}^k,$$
$$p \in [0,1]^k,$$

the smallest extension $\grave{P}_1$ of $\grave{P} = \{P_p \mid p_i = p_\theta(z_i),\ i=1,\dots,k\}$ ($\theta \in \mathbb{R}^2$) so that maximum likelihood estimation always is possible (and unique) is

$$\grave{P}_1 = \grave{P} \cup \{P_p \mid p = (0,\dots,0, \tfrac{x_i}{n_i}, 1,\dots,1),\ x_i=1,\dots,n_i;\ i=1,\dots,k\}$$

$$\cup \{P_p \mid p = (1,\dots,1, \tfrac{x_i}{n_i}, 0,\dots,0),\ x_i=1,\dots,n_i;\ i=1,\dots,k\}.$$

It seems, however, more natural to consider the extension

$$\grave{P}_2 = \grave{P} \cup \{P_p \mid p = (0,\dots,0, p_i,1,\dots,1),\ p_i \in [0,1];\ i=1,\dots,k\}$$

$$\cup \{P_p \mid p = (1,\dots,1, p_i,0,\dots,0),\ p_i \in [0,1];\ i=1,\dots,k\},$$

since

$$\{p = (p_1,\dots,p_k) \in [0,1]^k \mid P_p \in \grave{P}_2\}$$

is independent of $n_1,\dots,n_k$.
With each element P_θ, $\theta = (\alpha,\beta) \in \mathbb{R}^2$, of P we can associate the logistic function

$$\mathbb{R} \to [0,1]$$
$$z \mapsto p_\theta(z)$$

If $P_p \in \grave{P}_2 \setminus \grave{P}$ we may associate with P_p the degenerate logistic function which is the pointwise limit of $p_{\theta+\rho\xi+\rho'\eta}$ for any θ,ξ,η so that $P_{\theta+\rho\xi+\rho'\eta} \to P_p$ for $\rho,\rho' \to +\infty$. This leads to the following functions:

if $p = (0,\dots,0, p_i, 1,\dots,1)$, $p_i \in]0,1[$, $i \in \{1,\dots,k\}$:

$$\begin{aligned} \mathbb{R} &\to [0,1] \\ z &\mapsto \begin{cases} 0 & \text{for } z \in]-\infty, z_i[\\ p_i & \text{for } z = z_i \\ 1 & \text{for } z \in]z_i,+\infty[; \end{cases} \end{aligned}$$

if $p = (0,\ldots,0,\underset{\uparrow \text{i'th place}}{1},\ldots,1)$, $i \in \{2,3,\ldots,k\}$:

$$\begin{aligned} \mathbb{R}\backslash]z_{i-1},z_i[&\to [0,1] \\ z &\mapsto \begin{cases} 0 & \text{if } z \in]-\infty,z_{i-1}] \\ 1 & \text{if } z \in [z_i,+\infty[; \end{cases} \end{aligned}$$

if $p = (0,\ldots,0)$:

$$\begin{aligned} [z_1,z_k] &\to [0,1] \\ z &\mapsto 0 ; \end{aligned}$$

if $p = (1,\ldots,1)$:

$$\begin{aligned} [z_1,z_k] &\to [0,1] \\ z &\mapsto 1 \end{aligned}$$

(plus some analogous functions for the p_i-sequence decreasing).

The reason why some of the functions are undefined for some $z \in \mathbb{R}$ is that for these z $\lim_{\rho,\rho'} p_{\theta+\rho\xi+\rho'\eta}(z)$ is a non-constant function of (θ,ξ,η) on the set of all applicable (θ,ξ,η)'s. Thinking of the information contained in the observations $(x_{01},\ldots,x_{0k})$ about the graph of the logistic function, it is indeed very reasonable that the function is indetermined in some intervals.

5.9

The dose-response model is often applicated when describing experiments where a number of animals are treated with different doses of a certain drug - n_i animals are treated with the i-th dose; z_i is most commonly the logarithme of the dose - and one observes the number X_i of animals that die in group i, i=1,...,k. It is often assumed that the probability of dying is an increasing function of the dose, leading to the consideration of the family

$$\grave{P}_0 := \{P_\theta \mid \theta = (\alpha,\beta) \in \mathbb{R} \times [0,+\infty[\}.$$

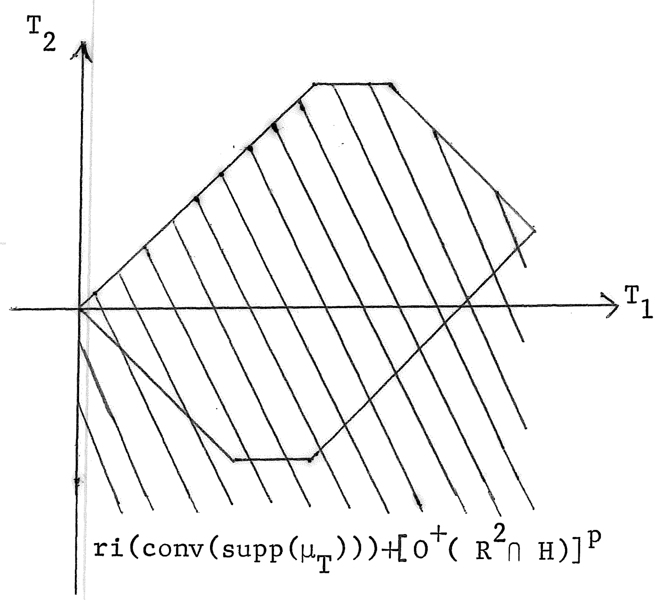
Here the parameter set thus is $\mathbb{R}^2 \cap H$, where

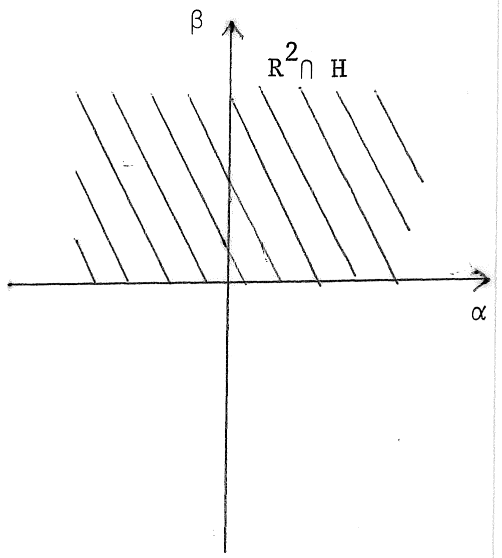
$$H = \{\theta = (\alpha,\beta) \in \mathbb{R}^2 \mid \beta \geqq 0\}$$

is closed and convex. Moreover

$$[0^+(\mathbb{R}^2 \cap H)]^p = \{(0,\delta) \mid \delta \in]-\infty,0]\}.$$

On applying Theorem 4.1 it is seen that the existence of a maximum likelihood estimator $\hat{\theta}_H(t_0) \in \mathbb{R}^2 \cap H$ is equivalent to the existence of $s_0 \in \mathrm{ri}(\mathrm{conv}(\mathrm{supp}(\mu_T)))$, $\delta \in]-\infty,0]$, so that $t_0 = s_0 + (0,\delta)$.





We shall now discuss the t_0's giving rise to a $\hat{\theta}_H(t_0) = (\hat{\alpha},\hat{\beta})$ in the interior of $\mathbb{R}^2 \cap H$, i.e. $\hat{\beta} > 0$; in this case $\hat{\theta}_H(t_0)$ is the solution $\tilde{\theta}$ to

$$E_{\tilde{\theta}}\, T = t_0.$$

According to 4.1 the mapping

$$\tau : \mathbb{R}^2 \to \mathrm{ri}(\mathrm{conv}(\mathrm{supp}(\mu_T)))$$
$$\theta \mapsto E_\theta T$$

is a bijection (as a matter of fact a homeomorphism) with the inverse mapping

$$\hat{\theta} : \text{ri}(\text{conv}(\text{supp}(\mu_T))) \to \mathbb{R}^2$$
$$t_0 \mapsto \hat{\theta}(t_0)$$

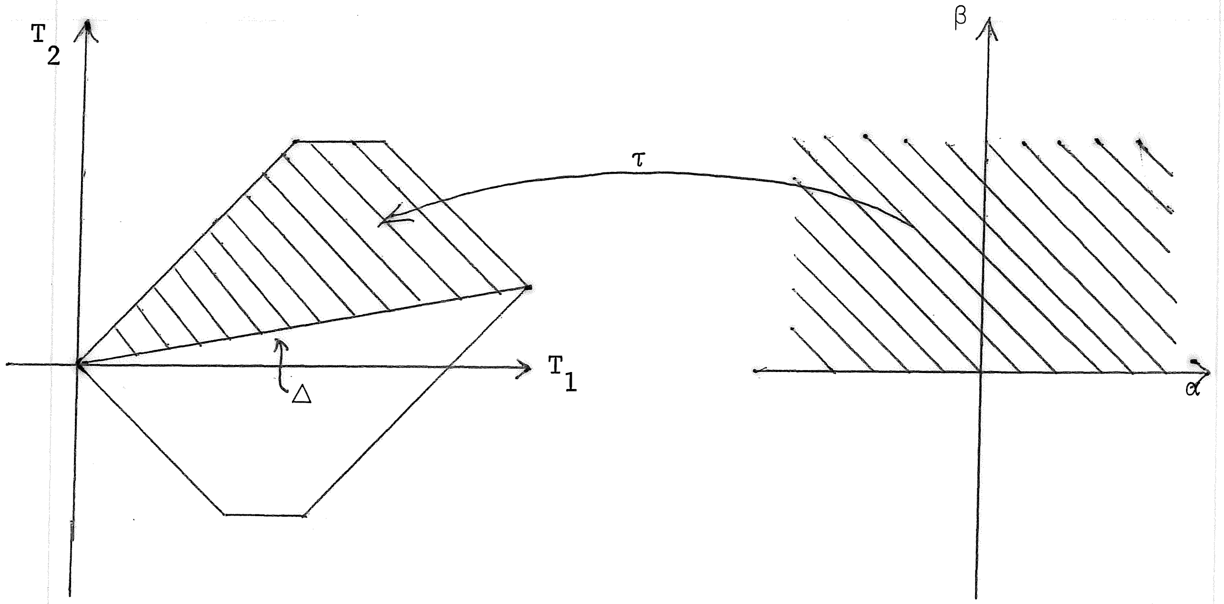
Since

$$\tau : \theta = (\alpha,\beta) \mapsto \sum_{i=1}^{k} \frac{n_i}{1 + \exp(-\alpha-\beta z_i)} (1,z_i) ,$$

the image of the α-axis is

$$\Delta : = \{\sigma \cdot \sum_{i=1}^{k} n_i(1,z_i) \mid \sigma \in]0,1[\}.$$

The set $\text{ri}(\text{conv}(\text{supp}(\mu_T)))\setminus\Delta$ consists of two path-connected components, as does $\mathbb{R}^2\setminus\{(\alpha,\beta)|\beta=0\}$, and as τ is continuous and bijective, the image by τ of $\text{ri}(\mathbb{R}^2 \cap H) = \{(\alpha,\beta) \in \mathbb{R}^2|\beta>0\}$ is one of the two components of $\text{ri}(\text{conv}(\text{supp}(\mu_T)))\setminus\Delta$; it is seen that it is the upper one:



If t_0 belongs to the interior of the upper sub-polygon we can find $\hat{\theta}_H(t_0)$ as the solution $\tilde{\theta}$ to

$$E_{\tilde{\theta}}T = t_0.$$

For any other t_0 for which $\hat{\theta}_H(t_0)$ exists, $\hat{\theta}_H(t_0)$ must be a point on the α-axis; because if $\hat{\theta}_H(t_0) \in \text{ri}(\mathbb{R}^2 \cap H)$ then $t_0 = E_{\hat{\theta}_H(t_0)}T$ was an interior point of the upper sub-polygon!

For $\theta = (\alpha,0)$, $\alpha \in \mathbb{R}$,

$$\frac{dP_\theta}{d\mu}(x) = \frac{\exp(\alpha \cdot T_1 x)}{(1+e^\alpha)^n} \quad , \quad x \in \mathbb{R}^k ,$$

where $n = \sum_{i=1}^{k} n_i$, $T_1 x = \sum_{i=1}^{k} x_i$, that is, $X_1,\ldots,X_k$ are independent, binominally distributed with the same probability parameter $e^\alpha/1+e^\alpha$. The maximum likelihood estimator $\hat{\alpha}$ thus exists if and only if

$$t_{01} = T_1 x_0 \in \mathrm{ri}(\mathrm{conv}(\mathrm{supp}(\mu_{T_1}))) =]\, 0,n[,$$

but this is implied by the assumption that $\hat{\theta}_H(t_0)$ exists, i.e. that $t_0 \in \mathrm{ri}(\mathrm{conv}(\mathrm{supp}(\mu_T))) + [0^+(\mathbb{R}^2 \cap H)]^p$.

We have, of course, that

$$\hat{\alpha} = \lambda(\frac{t_{01}}{n}).$$

In cases where $\hat{\theta}_H(t_0)$ does not exists, one should proceed in a similar way to 5.6, although for example the ξ's now should be chosen from $0^+(\mathbb{R}^2 \cap H) = H$. The results are not surprising. One should however be aware of the cases $t_0 = \underline{0}$ and $t_0 = \sum_{i=1}^{k} n_i(1,z_i)$; in the former case the degenerate logistic function is

$$\begin{aligned}]-\infty,z_k] &\to [0,1] \\ z &\mapsto 0 \end{aligned}$$

and in the latter case

$$\begin{aligned} [z_1,+\infty[&\to [0,1] \\ z &\mapsto 1. \end{aligned}$$

6. References

O. Barndorff-Nielsen (1970): Exponential Families.
Exact Theory.
Various Publication Series No. 191. Aarhus Universitet.

N. Bourbaki (1966): Espaces Vectorielles Topologiques, Ch. I et II.
2^e ed. Hermann.

M. Davis (1970): Geometric Representation of Designs for Biological Assay.
International Biometric Conference 1970.
Hannover, W. Germany.

S. Johansen (1970): Exponential Models.
Institute of Mathematical Statistics.
University of Copenhagen.

R. Tyrrell Rockafellar (1970): Convex Analysis.
Princeton University Press.

H. Silverstone (1957): Estimating the Logistic Curve.
J. Amer. Stat. Assoc. 52, 567-577.

J. Stoer, Chr. Witzgall (1970): Convexity and Optimization in Finite Dimensions I.
Springer Verlag.