Steffen L. Lauritzen

Examples of Extreme Point Models in Statistics

Preprint
November
1980
7

Institute of Mathematical Statistics
University of Copenhagen

S.L. Lauritzen

EXAMPLES OF EXTREME POINT MODELS IN STATISTICS. *

Preprint 1980 No. 7

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

November 1980

Abstract

We give a survey of examples of so-called extreme point models in statistics, i.e. statistical models that are given as the extreme points of the convex set of probability measures satisfying (in a general sense) a symmetry condition.

Some of the examples are only partially solved, some are classical and some are recent. A sketch of the general theory is also provided.

Key words: exchangeability, de Finetti's theorem, Rasch models, sufficiency.

0. Introduction and summary

Many statistical models have the common structure that they, in a given 'repetetive structure', are given as the extreme points of the convex set of probability measures satisfying a certain symmetry condition or, which amounts to the same, have certain given systems of statistics as sufficient. In particular this includes models given by i.i.d. repetitions of exponential families.

The aim of the present paper is to give examples of this, to point out some unsolved problems and to indicate the possibility of constructing new models deserving interest in theoretical statistics.

The proofs of the various results are omitted and the interested reader is referred to the cited works for those.

An important aspect of the considerations here is a kind of 'duality' between model and analysis.

In textbooks in theoretical statistics it is common to consider only a one-way connection among a statistical model and a statistical analysis, namely the way of deducing the statistical analysis from the model and some general inference principles.

Through examples we want here to indicate a 'dual' relation. From a specified symmetry or specified systems of calculations to be performed (the sufficient reduction) it is possible by mathematical construction to generate a corresponding statistical model in a canonical fashion. Thus the role of the statistical model can be said to be the way in which we express what we do when making a given analysis.

It is the opinion of the author that this way of thinking is useful, conceptually as well as for statistical practice where it is quite important to realise that the model certainly is not given beforehand.

There are many other examples in the literature than those given here and the reader is in that respect advised to consult the references to the present paper.

1. De Finetti's theorem

The starting point of the present survey is the following form of de Finetti's theorem:

Let $X_1,\dots,X_n,\dots$ be a sequence of random variables taking values in $\{0,1\}$ and suppose that their joint distribution is <u>exchangeable</u>

$$(X_1,\dots,X_n) \stackrel{\mathcal{D}}{=} (X_{\pi(1)},\dots,X_{\pi(n)})$$

for all $n \in \mathbb{N}$ and all permutations $\pi \in S(n)$, the symmetric group of order n. Here $X \stackrel{\mathcal{D}}{=} Y$ means that X and Y have the same distribution.

Then there is a unique probability measure μ on $[0,1]$ such that for all $n \in \mathbb{N}$,

$$P\{X_1 = x_1,\dots,X_n = x_n\} = \int_0^1 \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} \mu(d\theta) \ . \quad (1.1)$$

Moreover, the limit

$$\bar{X}_\infty = \lim_{n\to\infty} \frac{X_1 + \dots + X_n}{n}$$

exists almost surely and μ <u>is the distribution of</u> $\bar{X}_\infty$.

We can twist this result slightly by realising that

$$(X_1,\dots,X_n) \stackrel{\mathcal{D}}{=} (X_{\pi(1)},\dots,X_{\pi(n)}) \ \forall \ \pi \in S(n)$$

if and only if for all $t \in \{0,\dots,n\}$

$$P\{X_1 = x_1, \dots, X_n = x_n \mid X_1 + \dots + X_n = t\}$$

$$= \frac{1}{\binom{n}{t}} 1_{\{t\}} (x_1 + \dots + x_n) \ , \qquad (1.2)$$

and thereby noticing that the class of exchangeable probability measures on $\{0,1\}^{\mathbb{N}}$ exactly is <u>the largest class of probabilities</u> for which

$$t_n : \{0,1\}^n \to \{0,1,\dots,n\}$$

$$t_n(x_1,\dots,x_n) = x_1 + x_2 + \dots + x_n$$

for each n is <u>sufficient</u> with (1.2) as conditional distributions given $t_n(X_1,\dots,X_n) = t$.

Further, that this class is a <u>convex set</u>, with the independent Bernoulli measures as <u>extreme points</u>, i.e. those where the <u>representing measure</u> μ is degenerate and equal to the one-point measure ε_θ at θ and we can write

$$P\{\cdot\} = \int_{[0,1]} P_\theta\{\cdot\}\mu(d\theta)$$

where

$$P_\theta\{X_1 = x_1,\dots,X_n = x_n\} = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i} \ ,$$

such that, according to $P_\theta, X_1, X_2, \dots$ are just independent Bernoulli trials with probability of "success" equal to θ.

Moreover, since the representing measure exactly is "the limiting distribution of the sufficient statistic", <u>the probability</u> P <u>can be identified from complete observation of the entire sequence</u> $(X_1,X_2,\dots)$ <u>if and only if</u> P <u>is an extreme point</u>, since for only a single outcome $(X_1, X_2, \ldots)$ only a single value of $\bar{X}_\infty$ is even approximately observable.

The fact that the independent Bernoulli measures are the extreme points of the convex set of measures, for which $x_1 + \ldots + x_n$ for all n is sufficient and (1.2) are the corresponding conditionals, shall in the sequel be expressed as 'the model of independent identical Bernoulli trials is <u>an extreme point model</u>'.

We shall see that such extreme point models in fact occur quite commonly in statistics and that results like the above integral representation and interpretation of the representing measure as the limiting distribution of the sufficient statistic, are of a quite general nature.

Generalisations of de Finetti's theorem can be made in several directions. One is to exchange the spaces $\{0,1\}$ with more general measure spaces, as done by Hewitt and Savage (1955).

Another is to consider different groups like e.g. the group of rotations of $\mathbb{R}^n$ where the random variables $X_1, \ldots, X_n$ all take values in $\mathbb{R}$, cf. e.g. Kingman (1972).

Again, to specify other sufficient statistics than $x_1 + \ldots + x_n$ is one way of generalising de Finetti's theorem.

All these lines are special cases of the same general theory that we shall return to after having considered some examples related to statistics, where we shall focus on the last version of de Finetti's theorem.

2. Exponential families (discrete case)

A large class of examples well known to statisticians are the <u>exponential families of distributions</u>.

Let X be a discrete, almost countable sample space, $t : X \to \mathbb{R}^d$ a statistic and $a : X \to]0,\infty[$ a fixed function. Consider the family of distributions on the infinite product space $X^{\mathbb{N}}$ with marginal point probabilities

$$P_\theta(x_1,\ldots,x_n) = \prod_{i=1}^{n} \left(\frac{a(x_i)}{\phi(\theta)} e^{<\theta, t(x_i)>}\right), \theta \in D \qquad (2.1)$$

where $< , >$ is usual inner product and

$$\theta \in D = \{\theta \in \mathbb{R}^d : \phi(\theta) = \sum_x a(x) e^{<\theta, t(x)>} < +\infty\} .$$

According to P_θ, $X_1,\ldots,X_n$ are i.i.d. with distribution

$$P_\theta(X_1 = x) = \frac{a(x)}{\phi(\theta)} e^{<\theta, t(x)>}$$

for all n

$$t_n(x_1,\ldots,x_n) = t(x_1) + \ldots + t(x_n) \qquad (2.2)$$

is sufficient and the conditional distribution of $X_1,\ldots,X_n$ given $t_n(X_1,\ldots,X_n) = t$ is given as

$$\frac{\prod_{i=1}^{n} a(x_i)}{b^{*n}(t)} 1_{\{t\}}(t(x_1) + \ldots + t(x_n)) \qquad (2.3)$$

where $b^{*n}(t)$ is the n'th convolution of the measure $b = a \circ t^{-1}$, i.e.

$$b(t_0) = \sum_{x \in t^{-1}(t_0)} a(x) \; .$$

The extreme points of the convex set of probabilities having t_n as a sufficient statistic and (2.3) as conditionals, consists exactly of the exponential families (2.1) plus weak limits of these (Martin-Löf (1974)). That is, the extreme point model corresponding to the statistics (2.2) and the conditionals (2.3) is the so-called extended exponential families, introduced by Barndorff-Nielsen (1973,1978).

In general, Lauritzen (1975), one can show that the so-called generalised exponential families are examples of extreme point models, these being given by a countable set X, a statistic $t : X \to S$ where S is a commutative semigroup, an $a : X \to [0,\infty[$ as the family of probabilities on $X^{\mathbb{N}}$ with marginal point probabilities

$$P_\theta(x_1,\ldots,x_n) = \sum_{i=1}^{n} \left[\frac{a(x_i)}{\phi(\theta)} \; \theta(t(x_i))\right] \quad \theta \in D \qquad (2.4)$$

where D is a subset of EXP(S), the exponential functions on the semigroup S, i.e. those satisfying for $s,t \in S$

$$\theta(s)\theta(t) = \theta(s+t)$$

and

$$D = \{\theta \in \mathrm{EXP}(S) : \phi(\theta) = \sum_x a(x)\theta(t(x)) < \infty\} \; .$$

For these families

$$t_n(x_1,\ldots,x_n) = t(x_1) + \ldots + t(x_n)$$

is sufficient (+ denotes the semigroup composition), and the conditional distributions are also given by (2.3), where convolution is the convolution on the semigroup.

These examples include families like the uniform distributions on $\{1,\dots,\theta\}$, where the corresponding semigroup operation is maximum.

It is important that the 'repetitive structure' in these exponential family cases are that of identically distributed repetitions. Consider the following family of exponential models. Let $X_1,\dots,X_n,\dots$ be a sequence of independent random variables with distribution

$$P\{X_n = 1\} = 1 - P\{X_n = 0\} = \frac{\pi_n e^{\theta}}{1+\pi_n e^{\theta}} \qquad \theta \in \mathbb{R}$$

where $(\pi_n)_{n\in\mathbb{N}}$ is a fixed and known sequence of positive real numbers.

The joint distribution of $X_1,\dots,X_n$ is thus given by the point probabilities

$$P_\theta(x_1,\dots,x_n) = \prod_{i=1}^{n} \frac{\pi_i^{x_i}}{(1+\pi_i e^{\theta})} \cdot e^{\theta \sum_{i=1}^{n} x_i}$$

such that we for each n have an exponential family of distributions with

$$t_n(x_1,\dots,x_n) = x_1 + \dots + x_n$$

being sufficient and the conditional distributions

$$P(x_1,\dots,x_n | X_1 + \dots + X_n = t) = \left\{\frac{\prod_{i=1}^{n} \pi_i^{x_i}}{\gamma_t(\pi_1,\dots,\pi_n)}\right\}$$

if $x_1 + \dots + x_n = t$, 0 otherwise where $\gamma_t(\pi_1,\ldots,\pi_n)$ are the elementary symmetric functions

$$\gamma_t(\pi_1,\ldots,\pi_n) = \sum_{\substack{x_1+\ldots+x_n=t \\ x_i\in\{0,1\}}} \left(\prod_{i=1}^{n} \pi_i^{x_i}\right) . \qquad (2.5)$$

It follows from the results of Pitman (1978) that the above model is an extreme point model _if and only if_

$$\sum_{n=1}^{\infty} \frac{\pi_n}{(1+\pi_n)^2} = \infty . \qquad (2.6)$$

Again, this fact implies that θ is consistently estimable from one realisation of $X_1,\ldots,X_n,\ldots$ if and only if (2.6) is fulfilled.

Condition (2.6) is easily seen to be equivalent to

$$V_\theta\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \frac{\pi_n e^\theta}{(1+\pi_n e^\theta)^2} = \infty .$$

A related example is the model given by $X_1,\ldots,X_n,\ldots$ independent and Poisson distributed with

$$P_\theta\{X_n = x\} = \frac{(\theta^n)^x}{x!} e^{-\theta^n} .$$

Again the marginal point probabilities become

$$P_\theta(x_1,\ldots,x_n) = \frac{\theta^{\sum_{i=1}^{n} i x_i}}{\prod_{i=1}^{n} x_i!} e^{-\sum_{i=1}^{n} \theta^n}$$

and we have for _each_ n an exponential family with the statistic

$$t_n(x_1,\ldots,x_n) = \sum_{i=1}^{n} i\, x_i$$

as sufficient and the conditional distributions given as

$$P_\theta\{x_1,\ldots,x_n \mid t_n(X_1,\ldots,X_n) = t\} = \frac{1}{c_n(t)}\, 1_{\{t\}}\left(\sum_1^n i\, x_i\right)$$

where

$$c_n(t) = \sum_{\nu=0}^{\infty} \frac{1}{\nu!} \sum_{x_1+2x_2+\ldots+nx_n = t} \binom{\nu}{x_1,\ldots,x_n}$$

The situation is here slightly more complicated, but one can show (Lauritzen (1980)) that

i) P_θ is an extreme point if and only if $\theta \geqq 1$

ii) The distributions P^y, $y \in \mathbb{N}$ obtained by conditioning as

$$P^y\{\cdot\} = P^y_{\frac{1}{2}}\{\cdot \mid Y_\infty = y\}$$

where

$$Y_\infty = \sum_{i=1}^{\infty} i\ X_i$$

are extreme points as well.

It is an open problem whether these are all the extreme points.

3. Models for 0 - 1 matrices

An interesting class of examples different from the usual class of exponential families are extreme point models for 0 - 1 matrices.

We consider a doubly infinite array $(X_{ij})_{i,j \in \mathbb{N}}$ of random variables taking values in $\{0,1\}$.

Aldous (1979) has investigated the class of row-column exchangeable (RCE) arrays, i.e. arrays where

$$(X_{ij}, i = 1,\ldots,m, j = 1,\ldots,n)$$

$$\stackrel{\mathcal{D}}{=} (X_{\pi(i)\sigma(j)}, i = 1,\ldots,m, j = 1,\ldots n)$$

for all $m,n,\pi \in S(m)$ $\sigma \in S(n)$.

Or, equivalently where the maximal invariant under this group is sufficient and the conditional distribution given this statistic is uniform on the corresponding orbit.

The extreme points of this class of distributions are given as those RCE-arrays that are <u>dissociated</u>, i.e.

$$(X_{ij}, i \leq m, j \leq m) \text{ and } (X_{ij}, i > m, j > n)$$

<u>are independent</u> for all m and n.

Further, any such array can be matched in distribution by choosing

i) a measurable function $h :]0,1[\times]0,1[\times]0,1[\to \{0,1\}$

ii) independent sequences $(\xi_i)_{i \in \mathbb{N}}$ $(\eta_j)_{j \in \mathbb{N}}$ $(\lambda_{ij})_{ij \in \mathbb{N} \times \mathbb{N}}$ of i.i.d. uniform $]0,1[$ random variables.

and letting

$$X^h_{ij} = h(\xi_i, \eta_j, \lambda_{ij})$$

i.e. X^h_{ij} is composed by h from a random '<u>row-effect</u>' ξ_i, '<u>column-effect</u>' η_j and 'interaction' or 'error' λ_{ij}.

An open problem is in which sense the model here is <u>overparametrised</u> by h, since obviously different h's can give the same joint distribution of $(X^h_{ij})_{i,j \in \mathbb{N} \times \mathbb{N}}$.

In fact Aldous' result is not restricted to the state spaces {0,1} but holds for rather general spaces (Polish).

A different, but clearly related model is <u>Rasch's model for item analysis</u>. In this model, the random variables X_{ij} should be interpreted as the response of a person j to a question i and the model is that the X_{ij} are all independent with

$$P_{\underset{\sim}{\alpha},\underset{\sim}{\beta}}\{X_{ij} = 1\} = 1 - P\{X_{ij} = 0\} = \frac{\alpha_i \beta_j}{1 + \alpha_i \beta_j}$$

where $\underset{\sim}{\alpha} = (\alpha_i)_{i \in \mathbb{N}}$ and $\underset{\sim}{\beta} = (\beta_j)_{j \in \mathbb{N}}$ are sequences of unknown non-negative parameters. The marginal point probabilities for $(X_{ij})_{i \leq m, j \leq n}$ are given as

$$P_{\underset{\sim}{\alpha}\underset{\sim}{\beta}}\{(x_{ij})\} = \prod_{i=1}^{m} \prod_{i=1}^{n} \frac{(\alpha_i \beta_j)^{x_{ij}}}{(1 + \alpha_i \beta_j)} = \frac{\prod_{i=1}^{m} \alpha_i^{r_i} \prod_{j=1}^{n} \beta_j^{s_i}}{\prod_{i=1}^{m} \prod_{i=1}^{n} (1 + \alpha_i \beta_j)} \qquad (3.1)$$

where $r_i = \sum_{j=1}^{n} x_{ij}$ $s_j = \sum_{i=1}^{m} x_{ij}$ are the row- and column sums of the matrix $\{x_{ij}\}_{i=1,\dots,m;\ j=1,\dots,n}$.

The conditional distribution of the matrix given the sufficient statistics, the row- and column sums is uniform on the set of 0 - 1 matrices with the given row- and column sums.

It is not completely clear at present what the corresponding extreme point model is. Some positive results can be stated (Lauritzen (1980)).

The following condition is <u>necessary</u> for $P_{\underset{\sim}{\alpha},\underset{\sim}{\beta}}$ to be an extreme point:

$$A: \sum_{n=1}^{\infty} \frac{\alpha_n}{(1+\alpha_n)^2} = \sum_{n=1}^{\infty} \frac{\beta_n}{(1+\beta_n)} = \infty$$

and the condition below is <u>sufficient</u>

$$B: \sum_{n=1}^{\infty} \frac{\alpha_n \beta_n}{(1+\alpha_n)(1+\beta_n)(1+\alpha_n\beta_n)} = \infty \ .$$

Note that $B \Rightarrow A$, since

$$\frac{\alpha}{(1+\alpha)^2} - \frac{\alpha\beta}{(1+\alpha)(1+\beta)(1+\alpha\beta)} = \frac{\alpha(1+\beta^2)}{(1+\alpha)^2(1+\beta)(1+\alpha\beta)} \geq 0 \ .$$

Thus, the ratios $(\frac{\alpha_i}{\alpha_{i'}})$ and $(\frac{\beta_j}{\beta_{j'}})$ are consistently estimable as $m \to \infty$, $n \to \infty$ if B is satisfied and they are <u>not</u> if A is violated.

If we consider the modified version of this Rasch-model where m is <u>fixed</u> and only $n \to \infty$, the $P_{\underset{\sim}{\alpha},\underset{\sim}{\beta}}$ are <u>not</u> extreme points. The corresponding extreme point model can be shown to be the <u>conditional model</u>, i.e. the model obtained from the above by conditioning on the column sums $s_1,\ldots,s_n,\ldots$ and considering these as fixed, to obtain the probabilities

$$P_{\underset{\sim}{\alpha},s_1,\ldots,s_n,\ldots}\{X_{ij}=x_{ij},\ i=1,\ldots,m;\ j=1,\ldots,n\}$$

$$= \prod_{j=1}^{n} \frac{\prod_{i=1}^{m} \alpha_i^{x_{ij}}}{\gamma_{s_j}(\alpha_1,\ldots,\alpha_m)}$$

where $\gamma_s(\cdot,\ldots,\cdot)$ are the elementary symmetric functions.

To be exact, the above probability measures have to be supplemented by certain degenerate ones to make up <u>all</u> the extreme points, see Lauritzen (1980).

Note how in this example, the idea of looking for the extreme point model leads to <u>conditioning</u> on the statistics $(s_1,\ldots,s_n)$, a procedure also supported by other inference principles, since those conditional distributions now are free from the <u>nuisance parameters</u> $(\beta_1,\ldots,\beta_n,\ldots)$.

4. "Exponential" models for Markov chains

Consider an array (X_i), $i = 0,1,2,\ldots$ of random variables taking values in the countable set $\mathcal{X}$.

Consider the statistics for $i < j$

$$t_n(x_0,x_1,\ldots,x_n) = (x_0,\{n_{xy}\}_{(x,y) \in \mathcal{X} \times \mathcal{X}},x_n)$$

where n_{xy} are the transition counts

$$n_{xy} = \#\{k : (x_k,x_{k+1}) = (x,y)\} .$$

Diaconis and Freedman (1980) have shown that the extreme points of the class of probability measures, for which t_n is sufficient for all n, and the conditional distribution of $(X_k)_{k=i,\ldots,j}$ given t_n is uniform on the set of strings $(x_0,\ldots,x_n)$ with the given first and last value, and the given transition counts, fall in three classes

1) Recurrent Markov chains

2) Processes starting with a fixed string of transient states and continuing as recurrent Markov chains

3) Totally transient processes.

Results of Höglund (1974) indicate that a similar result can be obtained by considering the statistic

$$t_n(x_0,\ldots,x_n) = (x_0, \sum_{i=0}^{n-1} t(x_i,x_{i+1}),x_n)$$

where

$$t : \mathcal{X} \times \mathcal{X} \to Z^d$$

is a fixed statistic and the corresponding extreme point model consists then of the 'exponential families' of homogeneous Markov chains with initial distribution degenerate at x_0 and transition probabilities given as

$$P_\theta(x,y) = \frac{e^{<\theta, t(x,y)>}}{\phi(\theta)} \frac{e_\theta(y)}{e_\theta(x)}$$

where $\phi(\theta)$ is the maximal eigenvalue of the matrix

$$\left\{e^{<\theta, t(x,y)>}\right\}_{(x,y) \in X \times X}$$

and e_θ is the corresponding eigenvector. To get all the extreme points, certain probabilities of the type 2) and 3) in the case considered by Diaconis and Freedman should no doubt be added.

It seems also plausible that these models could be generalised to semigroup valued statistics.

5. General theory

We shall here give a sketch of the general framework unifying the examples given previously in the paper.

The way of expressing the repetitive structure is to start with a projective system of sample spaces and maps, i.e. and indexed family

$$(\mathcal{X}_i)_{i \in I}$$

of spaces that we assume to be Polish (complete, separable and metrisable), where I is partially ordered (<) and directed to the right, i.e.

$$\forall i,j \in I \;\exists\, k \in I : i < k,\; j < k \;.$$

The spaces $\mathcal{X}_i$ are connected with continuous maps $(p_{ij})_{i<j}$

$$p_{ij} : \mathcal{X}_j \to \mathcal{X}_i$$

that are coherent in the sense that

$$p_{ij} p_{jk} = p_{ik} \quad \text{if } i < j < k \quad .$$

The maps p_{ij} are called projections.

We further consider another system of Polish spaces $(\mathcal{Y}_i)_{i \in I}$ and a system of statistics, i.e. continuous surjective maps

$$t_i : \mathcal{X}_i \to \mathcal{Y}_i \;, \quad i \in I \quad .$$

We then consider probability measures on the projective limit

$$\varprojlim_{i \in I} \mathcal{X}_i = \{\underset{\sim}{x} \simeq (x_i)\, i \in I \mid x_i \in \mathcal{X}_i : p_{ij}(x_j) = x_i \;\; \forall i < j\}$$

for which the conditional distribution of X_i, where

$$X_i(\underset{\sim}{x}) = x_i$$

given

$$Y_i = t_i(X_i) = y$$

are given by a specified system of continuous Markov kernels

$$Q_i(y,B) \quad y \in \mathcal{Y}_i, \; B \in \mathcal{B}_i \; .$$

$\mathcal{B}_i$ is the Borel σ-algebra of $\mathcal{X}_i$ and the Q_i's satisfy the following consistency conditions:

i) $\quad Q_i(y, t_i^{-1}\{y\}) = 1 \quad$ for all $y \in \mathcal{Y}_i$, $i \in I$

ii) $\quad Q_j(z, p_{ij}^{-1}(B)) = \int_{\mathcal{Y}_i} Q_i(y,B) \; \nu_{ij}(z,dy)$

for all $i < j$, $z \in \mathcal{Y}_j$, $B \in \mathcal{B}_i$,

where for C being a Borel-subset of $\mathcal{Y}_i$

$$\nu_{ij}(z,C) = Q_j(z,(t_i \circ p_{ij})^{-1}(C)) \; .$$

In words, conditions i) and ii) ensure that Q_i is the conditional distributions of X_i given $t_i(X_i)$, calculated in any of the measures on $\mathcal{X}_i$

$$\mu_{ij}(z,B) = Q_j(z,p_{ij}^{-1}(B)) \quad .$$

If we now assume I to contain a <u>cofinal sequence</u> $(i_n)_{n \in \mathbb{N}}$, i.e. a sequence satisfying

$$\forall \; i \in I \; \exists \; n \in \mathbb{N} : i_n > i \; ,$$

one can show the following:

The set of probability measures $\mathcal{P}$ on $\mathcal{X} = \varprojlim_{i \in I} \mathcal{X}_i$ for which Q_i is the conditional distribution of X_i given $t_i(X_i)$ is convex and each member of $\mathcal{P}$ has a unique representation by a probability measure on the Polish space $\mathcal{C}$ of extreme points of $\mathcal{P}$, considered in the weak topology, i.e.

$$P(B) = \int_{\theta \in \mathcal{C}} \theta(B)\nu(d\theta)$$

for any Borel subset B of $\mathcal{X}$.

The representing measure ν is determined as the limiting distribution of the random variables

$$Z(B) = \lim_{n \to \infty} Z_n(B)$$

where $B \in \mathcal{B}_i$

$$Z_n(B) = Q_{i_n}(t_{i_n}(X_{i_n}), p_{ii_n}^{-1}(B))$$

as

$$P\{Z(B) \leq x\} = \int_{\{\theta : \theta(B) \leq x\}} \theta(B)\nu(d\theta)$$

such that, in particular, if $P \in \mathcal{E}$,

$$Z_n(B) \to Z(B) = P(B) \quad \text{a.s.}$$

implying that $Z_n(\cdot)$ is a strongly consistent estimate of $P(\cdot)$ if and only if P is an extreme point.

For further details of the theory the reader is referred to Lauritzen (1980).

Literature

Aldous, D.J. (1979). Representations for partially exchangeable arrays of random variables. Manuscript. University of Cambridge.

Barndorff-Nielsen, O. (1973). Exponential families and conditioning. Sc.D.Thesis, Copenhagen.

Barndorff-Nielsen, O. (1978). Information and exponential families. Wiley.

Dawid, A.P. (1978). Extendibility of spherical matrix distributions. J. Mult. Analysis, 8, 567-572.

Diaconis, P. and Freedman, D. (1980). De Finetti's theorem for Markov chains. Annals of Probability, 8, 115-130.

de Finetti, B. (1931). Funzione carratteristica de un fenomeno aleatorio. Atti della R. Academia Nazionale dei Lincei, Ser.6.Memorie. Classe de Scienze Fisiche Mathematiche e Naturali. 4, 251-299.

Hewitt, E. and Savage, L.J. (1955). Symmetric measures on Cartesian products. Trans. Amer. Math. Soc., 80, 470-501.

Höglund, T.H. (1974). Central limit theorems and statistical inference for Markov chains. Z.f.Wahrscheinlichkeitstheorie u.verw.Gebiete, 29, 123-151.

Kingman, J.F.C. (1972). On random sequences with spherical symmetry. Biometrika, 59, 492-493.

Lauritzen, S.L. (1974). Sufficiency, prediction and extreme models. Scand.J.Statist., 1, 128-134.

Lauritzen, S.L. (1975). General exponential models for discrete observations. Scand.J.Statist., 2, 23-33.

Lauritzen, S.L. (1980). Statistical Models as extremal families. Manuscript, 238 pp. To appear.

Martin-Löf, P. (1970). Statistiska Modeller. Notes by Rolf Sundberg (Swedish).

Martin-Löf, P. (1974). Repetitive structures. In *Proceedings of conference on foundational questions in statistical inference*. Memoires No. 1. Ed. O. Barndorff-Nielsen, P.B. Blæsild, G. Schou. Dept. of Theoretical Statistics. Aarhus University.

Pitman, J.W. (1978). An extension of de Finetti's theorem. *Adv.Appl.Prob.*, **10**, 268-269.

PREPRINTS 1979

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Edwards, David: Large Sample Tests for Stationarity and Reversibility in Finite Markov Chains.

No. 2 Andersson, Steen A. : Distribution of Maximal Invariants Using Proper Action and Quotient Measures.

No. 3 Johansen, Søren: A Note on the Welch-James Approximation to the Distribution of the Residual Sum of Squares in a Weighted Linear Regression.

No. 4 Björnsson, Ottó: Four Simple Characterizations of Standard Borel Spaces.

No. 5 Johansen, Søren: Some Comments on Robustness

No. 6 Hald, Anders: T.N. Thiele's Contributions to Statistics.

No. 7 Jacobsen, Martin: Markov Chains: Birth and Death Times with Conditional Independence.

PREPRINTS 1980

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Olkin, Ingram and Væth, Michael: Maximum Likelihood Estimation in a Two-way Analysis of Variance with Correlated Errors in One Classification.

No. 2 Skovgaard, Ib M.: Transformation of an Edgeworth Expansion by a Sequence of Smooth Functions.

No. 3 Asmussen, Søren: Equilibrium Properties of the M/G/1 Queue.

No. 4 Johansen, Søren and Keiding, Susanne: A Family of Models for the Elimination of Substrate in the Liver.

No. 5 Skovgaard, Ib M.: Edgeworth Expansions of the Distributions of Maximum Likelihood Estimators.

No. 6 Tjur, Tue: A Connection Between Rasch's Item Analysis Model and a Multiplicative Poisson Model.

No. 7 Lauritzen, Steffen L. : Examples of Extreme Point Models in Statistice.