Steffen L. Lauritzen

Discrete Exponential Models

Preprint
August
1973
7

Institute of Mathematical Statistics
University of Copenhagen

Steffen L. Lauritzen

Complete Canonical Exponential Models for Discrete Observations.

Preprint 1973 No. 7

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

August 1973

1. Introduction and Summary

The purpose of the present paper is to illustrate the concept of an extreme family as defined in [3], to give an alternative formulation of the theory of exponential models for discrete observations using simple algebraic properties of these instead of more complicated analytic results, and to show how some problems in connection with non-existence of maximum likelihood estimators in discrete exponential families can be solved.

In the classical formulation a discrete exponential family is a family of probability measures $(p_\theta, \ \theta \in \Theta)$, where the parameter set Θ is a subset of a k-dimensional Euclidean vector space, the probability function being given by

$$p_\theta(x) = a(\theta)b(x)e^{\sum_{i=1}^{k} \theta_i t_i(x)} \quad ,$$

where $x \in X$, a discrete set, t_i are real-valued functions and $\theta = (\theta_1,\ldots,\theta_k)$. One of the essential properties of the exponential family comes from the fact that if we repeat an experiment with the above probability function, the combined experiment will have the probability function given by

$$p_\theta^{(2)}(x,y) = a^2(\theta)b(x)b(y)e^{\sum_{i=1}^{k} \theta_i(t_i(x)+t_i(y))}$$

and will again be a family of the first kind with the same parameter space. Now we shall try to focus the elementary properties of the different elements in the exponential family. The function b(x) plays the role as a "reference measure" and a measure defining the support of the other measures in the family and $a(\theta)$ is a normalizing constant. The functions $(t_1,\ldots,t_k)$ are the sufficient statistics, defining an equivalence relation on X by saying that x_1 and x_2 are statistically equivalent if they give rize to the same value of the sufficient statistics. The equivalence relation as well as the statistics extend immediately to outcomes of n repetitions of the experiment by letting

$$t_i^{(n)}(x_1,\ldots,x_n) = t_i(x_1)+\ldots+t_i(x_n).$$

The essential reason for this is that the function

$$(t_1,\ldots,t_k) \to e^{\sum_{i=1}^{k} \theta_i t_i}$$

is a homomorphism of the range space of $(t_1,\ldots,t_k)$ with addition as composition into the non-negative real numbers with multiplication as composition.

The results in the present paper will be derived concentrating on the "support-defining measure", the statistical equivalence relation, the homomorphisms and the connection between an experiment and its repetitions.

In section 7, part I of [1] , there is a detailed discussion of problems connected to maximum likelihood estimation in exponential families. The maximum likelihood estimator in regular canonical exponential families is shown to exist iff the observation happens to be so, that the value of the sufficient statistic falls within the interior of the convex hull of the support of the measures in the family, transformed by the sufficient statistics. This means that if the boundary of this convex hull has positive probability, one might very well get an observation from which it is impossible to estimate. To solve this problem it is proposed in [1] to make a suitable extension of the model, the extension being defined for families where the set of possible values of the set of sufficient statistics is assumed to be finite. The extension is in [1] called the completion of an exponential family.

In the present paper the families defined are shown under weak assumptions to be "complete" in the sense that the maximum likelihood estimator of the parameters always exist. In the case with finite support treated in [1], it is shown that "the complete canonical exponential family" defined in the present paper is identical to the "completion of the canonical exponential family" as defined in [1].

The approach and the mathematical framework might be unusual to many statisticians, but an example is used throughout the paper to illustrate the various concepts introduced.

In section 2 some notation and simple algebraic concepts used in the paper is introduced as well as the notion of an extreme family of Markov chains as defined in [3] .

In section 3 the basic set-up is formulated, the algebraic structure, the equivalence relations,and the complete canonical families are defined in a general framework. Section 4 is pure algebra, the notion of a face of a monoid is introduced, and the results here seem to be very useful to describe the models in detail when dealing with concrete problems.

Section 5 is devoted to the estimation problem, existence and uniqueness of maximum likelihood estimators of the parameters in complete canonical models is established. In section 6 the results in the present paper are referred to those of [1] in a rather explicit manner.

2. Various Notation

The set of non-negative integers is denoted by $\mathbb{N}$, and $\mathbb{R}_+$ is used to denote the non-negative real numbers. Certain commutative monoids will play an essential role throughout the paper. A commutative monoid is a set M equipped with a composition rule + which is associative, commutative and has an identity. A detailed discussion of monoids is given in e.g. [2].

A commutative monoid $(M,+)$ satisfying $a + x = a + y \Rightarrow x = y$ for all $x,y,a \in M$ is in the present paper called a commutative semigroup.

The concept of an extreme family was introduced in [3] . Let $(E_n,\ n = 1,2,\dots)$ be a family of discrete, at most denumerable spaces and $Q = (Q_{mn})_{m\leq n}$ a family of matrices with elements $q_{mn}(x,y)$, $x \in E_m, y \in E_n$, satisfying

$$q_{mn}(x,y) \geqq 0, \qquad \sum_{x \in E_m} q_{mn}(x,y) = 1$$

and

$$Q_{mn}Q_{np} = Q_{mp} \text{ for } m \leqq n \leqq p.$$

$M(Q)$ denotes the set of sequences of probability measures $\mu = (\mu_n, n = 1,2,\ldots)$ so that μ_n is a probability measure on E_n and $\mu_m = Q_{mn}\mu_n$ for $m \leqq n$. $M(Q)$ is convex, $E(Q)$ denotes the extreme points of $M(Q)$ and is called the extreme family generated by Q.

3. Complete Canonical Exponential Families

Let X be a set, at most denumerable. Let ${}^{X}\mathbb{N}$ be the free, commutative monoid over X. ${}^{X}\mathbb{N}$ consists of all maps from X to $\mathbb{N}$ with finite support, equipped with addition as composition. One can interpret an element $f \in {}^{X}\mathbb{N}$ as a result of a sample taken from X with replacement, the sample size being $N(f) = \sum_{x \in X} f(x)$ and $f(x)$ the number of times x was observed. ${}^{X}\mathbb{N}$ is also a commutative semigroup.

Let R be an equivalence relation on ${}^{X}\mathbb{N}$, compatible with the composition rule, i.e. $fRg \wedge uRv \Rightarrow f + uRg + v$. By ${}^{X}\mathbb{N}|R$ we denote the quotient monoid of ${}^{X}\mathbb{N}$ over R, i.e. the set of R-equivalence classes with the composition rule given by

$$A_f + A_g = A_{f+g},$$

where

$$A_f = \{g \in {}^{X}\mathbb{N} : fRg\} .$$

The canonical mapping $t_R : {}^{X}\mathbb{N} \to {}^{X}\mathbb{N}|R$ given by $t_R(f) = A_f$ is a monoid-homomorphism.

Let R_0 be the equivalence relation given by $fR_0 g \Leftrightarrow \sum_{x \in X} f(x) = \sum_{x \in X} g(x)$, i.e.. that f and g correspond to samples of the same size. By X_n, $n \in \mathbb{N}$, we denote the R_0-equivalence classes so that $f \in X_n \Leftrightarrow \sum_{x \in X} f(x) = n$.

In the following we shall only consider equivalence relations R on $X_N^{\prime}$ satisfying $fRg \Rightarrow fR_0g$. We shall also assume that R satisfies $fRg \wedge f+cRg+d \Rightarrow cRd$, which ensures that $X_N^{\prime}|R$ is a commutative semigroup.

Example 1

The following example will be used throughout the paper to illustrate the various concepts.

Let A and B play chess. Let $X = \{+,\sim,-\}$, interpreted the following way:

+: A wins

~: draw

-: A loses.

$X_N^{\prime}$ is here all triples $f = (f_+,f_\sim,f_-)$ with $f_+,f_\sim,f_-$ all being non-negative integers, f being interpreted as the outcome of a match, A winning f_+ games, losing f_-, and $f_\sim$ games ending up as a draw. The composition rule is pointwise addition.

The equivalence relation ot be considered is the following:

$$fRg \Leftrightarrow$$
$$f_+ - f_- = g_+ - g_- \wedge f_+ + f_\sim + f_- = g_+ + g_\sim + g_- ,$$

i.e. the outcome of two matches are considered to be equivalent if they contain the same number of games and if A in both matches won the same number of games more than B. The equivalence classes are

$$A_f = \{(f_+ + s, f_\sim - 2s, f_- + s): -\min(f_+,f_-) \leq s \leq \tfrac{1}{2}f_\sim\}.$$

It is immediate to verify that R satisfies the properties demanded in the text above.

Let Ξ_R denote all monoid-homomorphisms of $X_N^{\prime}|R$ into $(R_+,\cdot)$.

For α being a probability measure on X'_N with support X_1 we define α^{*n}, the n-fold convolution of α, in the obvious way, $\alpha^{*1} = \alpha$ and

$$\alpha^{*n}(h) = \sum_{f+g=h} \alpha(f)\alpha^{*(n-1)}(g).$$

It is immediate that supp $\alpha^{*n} = X_n$.

Define Ξ_R^α by

$$\Xi_R^\alpha = \{\xi \in \Xi_R : \sum_{x \in X_n} \alpha^{*n}(x)\xi \circ t_R(x) = 1\} .$$

For $\xi \in \Xi_R^\alpha$ define the sequence $p^\xi = (p_n^\xi, n = 1,2,\ldots)$ of probability measures on X'_N by

$$p_n^\xi(x) = \alpha^{*n}(x)\xi \circ t_R(x) \qquad \text{for } x \in X'_N.$$

<u>Definition 1</u>: With the above notation the family $(p^\xi, \xi \in \Xi_R^\alpha)$ is said to be <u>the complete canonical exponential family</u> generated by α, X and R.

<u>Remark</u>: As supp $p_n^\xi \subseteq X_n$ for all $\xi \in \Xi_R^\alpha$ and $p_n^\xi * p_m^\xi = p_{n+m}$, we can interpret p_n^ξ as the distribution of a sample of size n of elements of X, taken with replacement. We shall use the families above to define the following statistical models. We let X_n be the sample space corresponding to an experiment of size n, n being arbitrary but fixed. Let the parameter space be Ξ_R^α and the distribution of the observation $f \in X_n$ be given by $p_n^\xi(f) = \alpha^{*n}(f)\xi \circ t_R(f)$. The models so described shall be referred to as the canonical models associated with α, X and R.

<u>Example 1</u> (continued)

Let
$$\alpha(f) = \begin{cases} \frac{1}{2} & \text{for } f = (0,1,0) \\ \frac{1}{4} & \text{for } f = (1,0,0) \\ \frac{1}{4} & \text{for } f = (0,0,1). \end{cases}$$

Then we have

$$\alpha^{*n}(f) = \binom{n}{f_+, f_\sim, f_-}\left(\frac{1}{2}\right)^{f_\sim}\left(\frac{1}{4}\right)^{f_+ + f_-} .$$

It is technically complicated to derive Ξ_R and Ξ_R^α without using the results about faces of monoids in section 5, so this will be done later in the paper.

Now, define $E_n = t_r(X_n)$ and for $a \in E_n$, $b \in E_{n+k}$

$$q_{n,n+k}(a,b) = \frac{\alpha_R^{*n}(a)\alpha_R^{*k}(c)}{\alpha_R^{*(n+k)}(b)}$$

where c is the unique element in E_k so that $b = a + c$ (the semigroup-property of $X_N^-|R$) and

$$\alpha_R^{*n}(a) = \sum_{t_R(x)=a} \alpha^{*n}(x).$$

Let $Q = (Q_{mn})_{m \leq n}$ as defined above. We can then obtain a connection between the complete canonical exponential family generated by α,X and R, and the extreme family defining the distributions of $t_R(f)$, whereas the complete canonical family defined the distribution of f. This is stated in the following result.

<u>Proposition 1:</u>

$$\mu \in E(Q) \Leftrightarrow \exists\ \xi \in \Xi_R^\alpha : \mu_n(a) = \alpha_R^{*n}(a)\,\xi(a)$$

<u>Proof:</u> Suppose $\mu \in E(Q)$ for $x \in E_k$, define

$$T_{x,k}\mu_n(a) = \begin{cases} \dfrac{\mu_{n+k}(a+x)}{\mu_k(x)} \cdot \dfrac{\alpha_R^{*n}(a)\alpha_R^{*k}(x)}{\alpha_R^{*(n+k)}(a+x)} & \text{if } \mu_k(x) > 0 \\ \mu_n(a) & \text{otherwise} \end{cases} .$$

It is straightforward to show that $\mu \in M(Q) \Rightarrow T_{x,k}\mu \in M(Q)$.

As we have

$$\mu_n(a) = \sum_{b \in E_k} \frac{\alpha_R^{*k}(b)\alpha_R^{*n}(a)}{\alpha_R^{*n+k}(a+b)} \mu_{n+k}(a+b)$$

and $\mu_k(b) = 0 \Rightarrow \mu_{n+k}(a+b) = 0$, some rearrangement of the above equation gives

$$\mu_n(a) = \sum_{b \in E_k} \mu_k(b) T_{b,k}\mu_n(a),$$

which expresses μ as a convex combination of $(T_{b,k}\mu, b \in E_k)$. Hence if μ is extreme, we must have

$$\mu_n(a) = T_{b,k}\mu_n(a) \text{ whenever } \mu_k(b) > 0.$$

If we let

$$h_n(a) = \frac{\mu_n(a)}{\alpha_k^{*n}(a)}$$

we then have $h_n(a)h_k(b) = h_{n+k}(a+b)$ whenever $h_k(b) > 0$. Because of symmetry the equation must hold also when $h_k(b) = 0$. As we supposed that $fRg \Rightarrow fR_0g$ we have $E_m \cap E_n = \emptyset$ whenever $m \neq n$, and we can therefore define $\xi(a) = h_n(a)$, for $a \in E_n$, and the equation becomes $\xi(a)\xi(b) = \xi(a+b)$, i.e. that $\xi \in \Xi_R$. As μ_n should be a probability measure we have

$$\sum_{a \in E_n} \alpha_R^{*n}(a)\xi(a) = 1 \Leftrightarrow \sum_{x \in X_n} \alpha^{*n}(x)\xi \circ t_R(x) = 1,$$

i.e. that $\xi \in \Xi_R^\alpha$.

The reverse implication follows from the fact that the Markov chains $Z_1, Z_2, \ldots$ as defined above can be thought of as $Z_n = t_R(Y_1+\ldots+Y_n)$, where $Y_1,\ldots,Y_n$, are independent, and the random variable $\lim_{n\to\infty}$ a.s. $q_{mn}(x,Z_n)$ is measurable with regard to the tail σ-algebra of $Y_1,\ldots,Y_n,\ldots$, which by the classical 0-1 law is degenerate. Proposition 2 of [3] together with the theorem of [3] gives that $\mu \in E(Q)$.

4. Faces of Monoids

We should like to get a more detailed description of the measures $(p_n^\xi, \xi \in \Xi_R^\alpha)$. In this section we shall investigate the structure of positivity regions of homomorphisms of a monoid into $(R_+,\cdot)$.

Let $(A,+)$ be a commutative monoid and $\hat{A}$ all homomorphisms of A into $(R_+,\cdot)$ not identically zero. Let $h \in \hat{A}$ be given. Consider $A^+(h) = \{a \in A \mid h(a) < 0\}$.

From the equation $h(a)h(b) = h(a+b)$ it easily follows that $A^+(h)$ is a submonoid of A. But $A^+(h)$ also satisfies $c \in A^+(h) \wedge c = a+b \Rightarrow a \in A^+(h) \wedge b \in A^+(h)$.

Conversely, let $A_0 \subseteqq A$ be a submonoid so that $c \in A_0 \wedge c = a+b \Rightarrow a,b \in A_0$. Let

$$h(a) = \begin{cases} 1 & \text{if } a \in A_0 \\ 0 & \text{otherwise} \end{cases} .$$

Clearly $h \in \hat{A}$ and $A^+(h) = A$ so we define

<u>Definition 2</u>: $F \subseteqq A$ is said to be a <u>face</u> of A if

i) F is submonoid of A
ii) $c \in F \wedge c = a+b \Rightarrow a \in F \wedge b \in F$.

The faces of A are exactly the possible positivity regions for elements in $\hat{A}$ which shall be formulated as a proposition.

Proposition 2: Let F be a subset of A. F is a face of A iff there is an $h \in \hat{A}$ so that $F = \{a \in A \mid h(a) \neq 0\}$.

The proof is given in the arguments above. We have furthermore

Proposition 3: A is a face of A.

Proposition 4: If $(F_i)_{i \in I}$ is a family of faces of A, then $\bigcap_{i \in I} F_i$ is a face of A.

The proofs of propositions 3 and 4 are immediate and left to the reader. It follows that to any $a \in A$, there is a unique smallest face of A, F(a), so that $a \in F(a)$.

Now, let A and B be monoids and $u: A \to B$ a homomorphism then

Proposition 5: If F is a face of B, then $u^{-1}(F)$ is a face of A.

The proof is trivial.

It will be of some use to have a characterization of all faces of X_N^- contained in the following result:

Proposition 6: For $f \in X_N^-$

$$F(f) = \{g \in X_N^- \mid \operatorname{supp} g \subseteq \operatorname{supp} f\} .$$

Proof: Let $A_0 = \{g \in X_N^- \mid \operatorname{supp} g \subseteq \operatorname{supp} f\}$. It is clear that A_0 is a face of X_N^- containing f, so $A_0 \supseteq F(f)$. Now let ε_{x_0} be the element of X_N^- given by

$$\varepsilon_{x_0}(x) = \begin{cases} 1 & \text{if } x = x_0 \\ 0 & \text{otherwise} \end{cases} .$$

We have $f = \sum_{x \in \text{supp } f} f(x)\varepsilon_x$ and therefore $\varepsilon_x \in F(f)$ for all $x \in \text{supp } f$. As A_0 is the submonoid of ${}^{X'}N$ generated by $(\varepsilon_x,\ x \in \text{supp } f)$ we must have $A_0 \subseteqq F(f)$ and the proof is complete.

Example 1 (continued)

We shall find the faces of the monoid ${}^{X'}N|R$ in the chess-example. The equivalence class corresponding to an element of the form $(f_+,0,0)$ contains only one element, and it is clear that $(f_+,0,0) = h+g$ implies that both h and g are of the form $(h_+,0,0)$ respectively $(g_+,0,0)$. Hence, the equivalence classes corresponding to elements of the form $(f_+,0,0)$ form a face of ${}^{X'}N|R$, which we shall denote as A_+. Analogously, A_- consisting of all equivalence classes corresponding to elements of the form $(0,0,f_-)$ is a face of ${}^{X'}N|R$. All other equivalence classes contain elements with $f_\sim$ being positive. Any submonoid of ${}^{X'}N|R$ containing an equivalence class with an $f_\sim$ positive, contains an equivalence class with an element $f_\sim > 2$, and hence also an element $(f_+ + 1, f_\sim - 2, f_- + 1)$ with all coordinates positive. Propositions 5 and 6 together implies then that the smallest face containing an equivalence class neither being of the form $(f_+,0,0)$ nor $(0,0,f_-)$ is ${}^{X'}N|R$ itself. Hence, the

Hence, the faces of ${}^{X'}N|R$ are A_+, A_-, ${}^{X'}N|R$ and the neutral element $(0,0,0)$, the last being uninteresting in connection with the determination of Ξ_R^α.

The homomorphisms positive on A_+ are

$$\xi_+^\theta(A_f) = \begin{cases} 0 & \text{if f is not of the form } (f_+,0,0) \\ \theta^{f_+} & \text{otherwise} \end{cases},$$

those positive on A_-:

$$\xi_-^\theta(A_f) = \begin{cases} 0 & \text{if f is not of the form } (0,0,f_-) \\ \theta^{f_-} & \text{otherwise} \end{cases},$$

those positive on $X'_N|R$:

$$\xi^{\theta_1,\theta_2}(A_f) = \theta_1^{2f_+ + f_\sim} \cdot \theta_2^{2f_- + f_\sim},$$

θ, θ_1, and θ_2 being arbitrary positive real numbers.

A straightforward calculation shows that

$$\xi_+^\theta, \xi_-^\theta \in \Xi_R^\alpha \Leftrightarrow \theta = 4,$$

and

$$\xi^{\theta_1,\theta_2} \in \Xi_R^\alpha \Leftrightarrow \theta_1 + \theta_2 = 2.$$

This is put together in the following statement:

$$\xi \in \Xi_R^\alpha \Leftrightarrow \exists\ \theta \in [0,1]: \quad \xi(f) = 4^{f_+ + f_\sim + f_-}\, \theta^{2f_+ + f_\sim} (1-\theta)^{2f_- + f_\sim}.$$

The corresponding probability measures are

$$p_n^\xi(f) = \binom{n}{f_+, f_\sim, f_-} \theta^{2f_+} (2\theta(1-\theta))^{f_\sim} (1-\theta)^{2f_-}$$

So, the model consists of probability measures where A loses or wins with probability one, but if there is positive probability of a draw then there is positive probability both of A winning and losing.

5. Maximum Likelihood Estimation

We shall now estimate in the previously defined canonical models associated with α, X and R.

Let n be arbitrary but fixed. The likelihood function becomes

$$L(f,\xi) = \alpha^{*n}(f)\xi \circ t_R(f),$$

and a maximum likelihood estimator is a mapping $\hat{\xi}_n : X_n \to \Xi_R^{\alpha}$, satisfying for all $f \in X_n$

$$L(f,\hat{\xi}_n^f) = \sup_{\xi \in \Xi_R^{\alpha}} L(f,\xi). \qquad (*)$$

<u>Proposition 7</u>: Under the assumption that for all n,k = 1,2,... and $a \in E_n$

$$\sum_{c \in E_k} \frac{\alpha_R^{*k}(c)}{\alpha_R^{*(n+k)}(a+b)} < +\infty ,$$

there exists a mapping $\hat{\xi}_n$ satisfying (*), and if $\tilde{\xi}_n$ satisfies (*) then $\tilde{\xi}_n^f = \hat{\xi}_n^f$ for all $f \in X_n$.

<u>Proof</u>: The existence follows from the regularity assumption and proposition 3 of [3].

The uniqueness is proved in the following way: suppose that there is an $f \in X_n$ so that $\tilde{\xi}_n^f \neq \hat{\xi}_n^f$. As $L(f,\tilde{\xi}_n^f) = L(f,\hat{\xi}_n^f)$ we have $\tilde{\xi}_n^f \circ t_R(f) = \hat{\xi}_n^f \circ t_R(f)$. Define

$$\xi(a) = \frac{\sqrt{(\hat{\xi}_n^f(a)\tilde{\xi}_n^f(a))}}{\sum_{b \in E_n} \alpha_R^{*n}(b)\sqrt{(\hat{\xi}_n^f(b)\tilde{\xi}_n^f(b))}} .$$

Clearly, $\xi \in \Xi_R^{\alpha}$. As when $\tilde{\xi}_n^f \neq \hat{\xi}_n^f$

$$\sum_{b \in E_n} \alpha_R^{*n}(b)\sqrt{(\tilde{\xi}_n^f(b)\hat{\xi}_n^f(b))} < 1$$

by the Cauchy-Schwarz inequality, $\xi \circ t_R(f) > \hat{\xi}_n^f \circ t_R(f)$, contradicting the fact that $\hat{\xi}_n^f$ should satisfy (*). Hence $\hat{\xi}_n^f = \tilde{\xi}_n^f$, which was to be proved.

Remark: I do not know of any example where the regularity assumption is not fulfilled.

The following result giving some more detailed information about the maximum likelihood estimator should be compared to some of the results in sec. 7, part I of [1] .

Proposition 8: The positivity region of $\hat{\xi}_n^f$ is exactly the face $F(t_R(f))$.

Proof: Suppose $a_0 \notin F(t_R(f))$ and $\hat{\xi}_n^f(a_0) > 0$. Define

$$\xi'(a) = \begin{cases} \hat{\xi}_n^f(a) & \text{for } a \in F(t_R(f)) \\ 0 & \text{otherwise} \end{cases},$$

and

$$\xi(a) = \frac{\xi'(a)}{\sum_{b \in E_n} \alpha_R^{*n}(b)\xi'(b)} .$$

Clearly $\xi \in \Xi_R^\alpha$ and $\xi \circ t_R(f) > \hat{\xi}_n^f \circ t_R(f)$, contradicting that $\hat{\xi}_n^f$ was a maximum likelihood estimator.

Example 1 (continued)

From proposition 8 we get that if f is of the form $(f_+,0,0)$, the maximum likelihood estimator will tell that A wins all games with probability 1, and if $f = (0,0,f_-)$ that A loses with probability one. Otherwise the maximum likelihood estimator is

$$\hat{\xi}^f_n(A_g) = 4^{g_+ + g_\sim + g_-}\left(\frac{f_+ + \frac{1}{2}f_\sim}{n}\right)^{2g_+ + g_\sim}\left(\frac{f_- + \frac{1}{2}f_\sim}{n}\right)^{2g_- + g_\sim} .$$

6. Additional Comments

Let X now be finite, $t: X \to \mathbb{R}^k$ a mapping into k-dimensional Euclidean space with t taking values in the integer lattice points of $\mathbb{R}^k$. Let P_0 be a probability measure on X and consider π , the canonical exponential family generated by P_0 and t as defined in [1] , i.e. $P \in \pi \Leftrightarrow \exists \theta \in \mathbb{R}^k$:

$$P(x) = P_0(x)a(\theta)e^{\theta \cdot t(x)} .$$

It is not difficult for the reader to identify the "completion" of π (as defined in section 7, part I of [1]) with the complete canonical model generated by P_0, X, and the equivalence relation R given by

$$fRg \Leftrightarrow \sum_{x \in X} f(x) = \sum_{x \in X} g(x) \wedge \sum_{x \in X} f(x)t(x) = \sum_{x \in X} g(x)t(x).$$

The only problem is to establish a connection between faces of the convex hull of t(X) in the convexity sense and faces of the quotient monoid $X_{\mathbb{N}}|R$.

If we let

$$T_n = \underbrace{t(X) + \ldots + t(X)}_{n \text{ terms}}, \quad T_0 = \{\underline{0}\}$$

and

$$T = \{(n,t): n \in \mathbb{N} \wedge t \in T_n\}$$

equipped with the composition

$$(p,s) + (q,t) = (p + q, s + t),$$

then $(T,+)$ is isomorphic to $X'_N|R$, the isomorphism being given by $\psi(A_f) = (p,t)$ whenever

$$\sum_{x \in X} f(x) = p \wedge \sum_{x \in X} f(x)t(x) = t.$$

Now, let F^* be a proper face of C, the convex hull of $t(X)$. F^* is given by a supporting hyperplane, i.e. a linear form ϕ, satisfying

$$\phi(t) > a \text{ for all } t \in C \setminus F^*$$

$$\phi(t) = a \text{ for all } t \in F^*.$$

Define

$$F_n^* = \underbrace{F^*+\ldots+F^*}_{n \text{ terms}} , \quad F_0^* = \{\underline{0}\}$$

and

$$F(F^*) = \{(n,t) \in T | t \in F_n^*\} .$$

Then $F(F^*)$ is a face of T as $(n,t) = (p,s)+(q,u)$ implies

$$\phi(t) = na \wedge \phi(s)+\phi(u) > pa + qa = na$$

if s and u are not in F_p^*, respectively F_q^*.

Now, if $t_0 \in t(X)$ and in the relative interior of F^*, where F^* is a face of C, there is a probability measure on $F^* \cap t(X)$ so that

$$t_0 = \sum_{i=1}^{p} \lambda_i t_i , \quad \lambda_i > o \ \forall i ,$$

where $F^* \cap t(X) = \{t_1,\ldots,t_p\}$, which follows from the results in [1] about existence of the maximum likelihood estimator in the relative interior of the convex support of an exponential family. As $\{t_1,\ldots,t_p\}$ are integer lattice points, the λ_i's can be chosen to be rational. Hence there is an N so that

$$Nt_0 = \sum_{i=1}^{p} (\lambda_i N) \cdot t_i \ ,$$

and so that $(\lambda_i N)$ are integers.

It follows that any face of T containing $(1,t_0)$ must contain (N,Nt_0) and also $(1,t_i)$ for all $t_i \in F^* \cap t(X)$. Hence, the smallest face of T containing t_0 is exactly $F(F^*)$ as defined above, where F^* is the uniquely determined face of C, so that $t_0 \in F^*$.

At last we shall give an example of the "completion" when X is not finite.

Example 2

Let $X = \{0,1,2,\dots\}$, $\alpha(x) = \frac{1}{x!} e^{-1}$, R given by

$$fRg \Leftrightarrow \sum_{x \in X} f(x) = \sum_{x \in X} g(x) \wedge \sum_{x \in X} xf(x) = \sum_{x \in X} xg(x).$$

It is easy to establish that

$$p \in (p_1^\xi,\ \xi \in \Xi_R^\alpha) \Leftrightarrow \exists\, 0 \leqq \lambda < +\infty:\ p(x) = \frac{(\lambda)^x}{x!} e^{-\lambda}.$$

The only interesting thing to notice is that $\lambda = 0$ is included.

7. Acknowledgements

I owe thanks to all my colleagues at the Institute of Mathematical Statistics, University of Copenhagen, for several inspiring discussions on the topics in the present paper, to Ann Mitchell, Imperial College, London, and Søren Johansen, Copenhagen, for reading the manuscript and giving very valuable suggestions.

Litterature

[1]: (1973).Barndorff-Nielsen, O.: Exponential Families and Conditioning. Copenhagen, Wiley.

[2]: (1970). Bourbaki, N.: Algèbre, Chap. I. Hermann, Paris.

[3]: (1973). Lauritzen, S.L.: Sufficiency, Prediction and Extreme Models. Preprint 1973 No. 6 , University of Copenhagen, Institute of Mathematical Statistics.