Z. Wahrscheinlichkeitstheorie verw. Gebiete 29, 123–151 (1974)

Central Limit Theorems and Statistical Inference for Finite Markov Chains

Thomas Höglund

1. Introduction

In this paper we will deal with four topics in connexion with Markov chains: *saddle point approximation* (Section 3), *equivalence of ensembles* (Section 4), *central limit theorems* (Section 5), and *statistical inference* (Section 6). The saddle point approximation is a common key to the other topics, but Sections 4, 5, and 6 can be read independently of each other except that 4 has to precede 6. The key words above will be explained in some more detail when we have introduced the necessary notations.

We consider a finite set X, a sequence $\{q_{xy}\}_{(x,y)\in X\times X}$ of nonnegative numbers with support

$$Y=\{(x,y)\in X\times X \mid q_{xy}>0\} \tag{1.1}$$

and a statistic $Y\ni(x,y)\to t(x,y)\in Z^p$. Y is supposed to be an *irreducible* subset of $X\times X$: For each $(x,y)\in X\times X$ there is a sequence $z_0, z_1, \ldots, z_k$ $(k\geqq 1)$ such that $z_0=x$, $z_k=y$, and $(z_{i-1}, z_i)\in Y$ for $i=1,\ldots,k$.

Let C^X (where C is the set of complex numbers) denote the set of complex valued sequences $g=\{g(x)\}_{x\in X}$ and define the linear operators (or matrices)

$$C^X\ni g\to F(t)\,g\in C^X \qquad (t\in Z^p) \tag{1.2}$$

by $F(t)\,g(x)=\sum\limits_{y\in X} f_{xy}(t)\,g(y)$, where

$$f_{xy}(t)=\begin{cases} q_{xy} & \text{if } (x,y)\in Y \text{ and } t(x,y)=t\\ 0 & \text{otherwise.}\end{cases} \tag{1.3}$$

In analogy with the terminology of Hinčin [9] the function $Z^p\ni t\to F(t)$ will be called the *structure function.* Using matrix notation we will sometimes write $F(t)=(f_{xy}(t))$.

We further define

$$F^{n*}(t)=\sum_{t_1+\cdots+t_n=t} F(t_1)\ldots F(t_n) \tag{1.4}$$

where the product to the right is matrix product. Then we have $F^{n*}(t)=(f_{xy}^{(n)}(t))$,

$$f_{x_0x_n}^{(n)}(t)=\sum q_{x_0x_1}\,q_{x_1x_2}\cdots q_{x_{n-1}x_n} \tag{1.5}$$

where the summation is extended over all $x_1,\ldots,x_{n-1}$ such that $(x_{i-1},x_i)\in Y$ for $i=1,\ldots,n$ and $\sum\limits_1^n t(x_{i-1},x_i)=t$.

The saddle point approximation (which concerns $F^{n*}(t)$) will be obtained via the *transform* of F

$$\Phi(a+i\alpha)=\sum_{t\in Z^p} e^{(a+i\alpha)\cdot t} F(t) \qquad (a\in R^p,\ \alpha\in R^p); \tag{1.6}$$

then $\Phi(z)=(\varphi_{xy}(z))$ $(z\in C^p)$ where

$$\varphi_{xy}(z)=\begin{cases} e^{z\cdot t(x,y)}\, q_{xy} & \text{if } (x,y)\in Y \\ 0 & \text{otherwise.}\end{cases} \tag{1.7}$$

The sum to right in (1.6) converges for all $z\in C^p$ since $F(t)$ has finite support. It is easy to verify the following identities:

$$\Phi(a+i\alpha)^n=\sum_{t\in Z^p} e^{(a+i\alpha)\cdot t} F^{n*}(t) \tag{1.8}$$

and

$$F^{n*}(t)=\frac{1}{(2\pi)^p}\int_{(-\pi,\pi]^p} e^{-(a+i\alpha)\cdot t}\,\Phi(a+i\alpha)^n\, d\alpha \tag{1.9}$$

where the integration is to be interpreted as componentwise integration. The latter identity, which will be our main tool, gives us reason to examine Φ closer. This will be done in Section 2.

In order to give a summary of our results we impose certain simplifying conditions (not to be specified here). Under these conditions the following holds. $\Phi(a)$ has a unique positive eigenvalue $\lambda(a)$ which exceeds the modulus of any other eigenvalue of $\Phi(a)$. $\Phi(a)$ and $\Phi(a)^*$ (the adjoint of $\Phi(a)$) have strictly positive eigenfuntions e_a and e_a^* corresponding to the eigenvalue $\lambda(a)$. Furthermore, $\lambda(a)$ is analytic and the function $a\to \operatorname{grad}\log\lambda(a)$ is one to one. Write $\hat{a}$ for its inverse. ($\hat{a}$ has an interpretation as a statistical estimator of a certain parameter. See Section 6.)

The *saddle point approximation* states that

$$(2\pi n)^{p/2}(\det V_{\hat{a}(t/n)})^{\frac{1}{2}}\, e^{-nH_{\hat{a}(t/n)}}\, F^{n*}(t)=E(\hat{a}(t/n))+O(1/n), \tag{1.10}$$

uniformly in t when $\hat{a}(t/n)$ stays within any fixed compact subset of R^p. Here

$$E(a)=\left(\frac{e_a(x)\,e_a^*(y)}{e_a\cdot e_a^*}\right)$$

is the eigenprojection corresponding to $\lambda(a)$, V_a is the matrix $\left(\frac{\partial^2}{\partial a_i\,\partial a_j}\log\lambda(a)\right)$ and $H_a=\log\lambda(a)-a\cdot\operatorname{grad}\log\lambda(a)$ is the specific entropy. (The connexion between H_a and entropy is explained in Section 4.)

Equivalence of ensembles. In statistical mechanics one considers several kinds of "ensembles" two of which are the microcanonical and the canonical ensemble. It is believed that they are equivalent in a certain sense. We apply a variant of our saddle point approximation and show, more precisely, that if

$$p^{M,N}_{Tx_{-M}x_N}(x_0,\dots,x_n)$$

denotes the density of the distribution for $x_0, \ldots, x_n$ induced by the uniform (the microcanonical) distribution on the surface

$$\{x_{-M+1}, \ldots, x_0, \ldots, x_n, \ldots, x_{N-1} | \sum_{-M<i\leqq N} t(x_{i-1}, x_i) = T\},$$

then

$$p_{Txy}^{M,N}(x_0, \ldots, x_n) = (e_{\hat{a}}^* \cdot e_{\hat{a}})^{-1} e_{\hat{a}}^*(x_0) \frac{e^{\hat{a}\cdot t_n}}{\lambda(\hat{a})^n} e_{\hat{a}}(x_n) \left[1 + O(1/(M+N))\right], \quad (1.11)$$

uniformly in T when $\hat{a} = \hat{a}(T/(M+N))$ stays within any fixed compact subset of R^p, provided $M/(M+N)$ remains bounded away from zero and one. Here

$$t_n = \sum_1^n t(x_{i-1}, x_i),$$

and e_a, e_a^*, $\lambda(a)$ and $\hat{a}$ are determined by the structure function for which $q_{xy} = 1$ for all x and y. The densities to the right in (1.11) determine a stationary measure on X^Z: the canonical Markov chain.

Central Limit Theorems. We let $\{q_{xy}\}$ be transition probabilities $(\sum_{y\in X} q_{xy} = 1$ for all $x \in X)$, and consider the stationary Markov chain which has $\{q_{xy}\}$ as transition probabilities. The asymptotic expression for the structure function gives various approximations of the density and distribution function of $\sum_1^n t(x_{i-1}, x_i)$, some of which hold far out in the tails. For example, when $p=1$ we obtain an approximation for Prob $\left(\sum_1^n t(x_{i-1}, x_i) > t\right)$ which is non-trivial not only when t/n is close to the mean value but also when t/n belongs to compact subsets of the preimage of $\hat{a}$.

Statistical Inference. Martin-Löf [12, 13] gave a statistical theory which can be divided into two parts: exact and asymptotic (or microcanonical and canonical). The saddle point approximation provides us with a tool to derive the asymptotic part from the exact one in the Markov case.

Consider now the following more general situation (we will refer to this situation as "the m-dependence case"): Let X be finite, $\{q_{x_1 \ldots x_m}\}_{x_1, \ldots, x_m \in X^m}$ $(m \geqq 1)$ a non-negative sequence with support

$$Y = \{(x_1, \ldots, x_m) \in X^m \,|\, q_{x_1 \ldots x_m} > 0\} \quad (1.12)$$

and let $Y \ni (x_1, \ldots, x_m) \to t(x_1, \ldots, x_m) \in Z^p$ be a statistic (i.e. a function). Suppose that Y is an irreducible subset of X^m: For each $(x_1, \ldots, x_m) \in X^m$ there is a sequence $z_1, \ldots, z_k$ $(k \geqq m)$ such that

$$z_1 = x_1, \ldots, z_{m-1} = x_{m-1}, z_k = x_m \text{ and } (z_i, z_{i+1}, \ldots, z_{i+m-1}) \in Y$$

for $i = 1,, \ldots, k-m+1$.

We identify C^{X^0} with C and define the operators

$$C^{X^{m-1}} \ni g \to F(t)\, g \in C^{X^{m-1}} \qquad (t \in Z^p) \quad (1.13)$$

by $F(t)\,g(x_1,\ldots,x_{m-1})=\sum_{x_m\in X} f_{x_1,\ldots,x_m}(t)\,g(x_2,\ldots,x_m)$ where

$$f_{x_1\ldots x_m}(t)=\begin{cases} q_{x_1\ldots x_m} & \text{if } (x_1,\ldots,x_m)\in Y \text{ and } t(x_1,\ldots,x_m)=t \\ 0 & \text{otherwise.}\end{cases} \tag{1.14}$$

$\phi(a+i\alpha)$ is defined as in (1.6) and $F^{n*}(t)$ as in (1.4). Then the identities (1.8) and (1.9) holds and

$$F^{n*}(t)\,g(x_1,\ldots,x_{m-1})=\sum_{y_1\cdots y_{m-1}} f^n_{x_1\ldots x_{m-1}\,y_1\ldots y_{m-1}}(t)\,g(y_1,\ldots,y_{m-1})$$

where for $n\geqq m$

$$f^n_{x_1\ldots x_{m-1}\,y_1\ldots y_{m-1}}(t)=\sum q_{x_1\ldots x_m}\,q_{x_2\ldots x_{m+1}}\cdots q_{x_n\ldots x_{n+m-1}} \tag{1.15}$$

the summation being extended over all $x_m, x_{m+1},\ldots,x_n$ such that

$$(x_i,x_{i+1},\ldots,x_{i+m-1})\in Y \quad\text{for } i=1,\ldots,n \quad\text{and}\quad \sum_{i=1}^{n} t(x_i,x_{i+1},\ldots,x_{i+m-1})=t.$$

The case $m=1$ – the independence case – is trivial in the present context and the case $m>2$ may be reduced to the case $m=2$ in the following well-known way: Define $X^*=X^{m-1}$,

$$Y^*=\{(x^*,y^*)\in X^*\times X^*\,|\,x^*=(x_1,\ldots,x_{m-1}),\ y^*=(y_1,\ldots,y_{m-1}),$$
$$y_1=x_2,\ldots,y_{m-2}=x_{m-1} \quad\text{and } (x_1,\ldots,x_{m-1},y_{m-1})\in Y\}.$$

Also

$$t^*(x^*,y^*)=t(x_1,\ldots,x_{m-1},y_{m-1}) \quad\text{and}\quad q^*_{x^*y^*}=q_{x_1\ldots x_{m-1}\,y_{m-1}} \quad\text{for}\quad (x^*,y^*)\in Y^*,$$

$x^*=(x_1,\ldots,x_{m-1})$, $y^*=(y_1,\ldots,y_{m-1})$. It is a straightforward matter to verify that Y^* is irreducible, and hence the 2-dependence case contains the m-dependence case.

It will be convenient to consider only the case $m=2$. It follows, however, from the argument above that our results are valid for all $m\geqq 1$.

2. The Transform

The *period* of Y is defined to be the largest integer r for which there is a partition $\{X_i\}_{i=1}^r$ of X such that

$$Y\subset\bigcup_{i=1}^{r} X_{i-1}\times X_i. \tag{2.1}$$

Here we used the convention $X_0=X_r$. Throughout this paper we will denote the period of Y by r and let $\{X_i\}_{i=1}^r$ stand for the unique partition which satisfies (2.1). We will give meaning to X_i for all $i\in Z$ via the formula $X_{k+nr}=X_k$, $n\in Z$, $1\leqq k\leqq r$, and let $r(x)$ $(x\in X)$ stand for the unique integer which satisfies

$$x\in X_{r(x)},\qquad 1\leqq r(x)\leqq r. \tag{2.2}$$

$\Phi(a)$ is (considered as a matrix) a matrix with non-negative elements. Perron and Frobenius examined such matrices and their results are collected in Wielandt [19]. Wielandt says that a matrix $A=(a_{ij})$ is irreducible (unzerlegbar) if one cannot bring it into a form

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

by a consistent relabeling of the rows and columns, where A_{11} and A_{22} are quadratic submatrices. For irreducible matrices we have [19], p. 642 and p. 645:

I. Die charakteristische Gleichung

$$\det(A - xI) = 0 \tag{2.3}$$

besitzt eine einfache positive Wurzel λ, die dem Betrage nach von keiner anderen Wurzel übertroffen wird. Die zu dieser „Maximalwurzel" λ gehörige Eigenlösung kann positiv gewählt werden; λ ist der einzige Eigenwert, zu dem eine nicht negative Eigenlösung existiert.

II. Besitzt (2.3) insgesamt r Wurzeln vom Betrage λ, so sind diese einfach und haben die Werte $\lambda \exp(i2\pi j/r)$, $(j=1,2,\ldots,r)$. Die Gesamtheit der n Wurzeln von (2.3) gestattet die Drehung um den Nullpunkt mit dem Winkel $2\pi/r$, aber nicht mit einem kleineren Winkel. A hat die Gestalt

$$A = \begin{pmatrix} 0 & A_{12} & 0 & \ldots 0 \\ 0 & 0 & A_{23} & \ldots 0 \\ 0 & 0 & 0 & \ldots 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots A_{r-1\,r} \\ A_{r1} & 0 & 0 & \ldots 0 \end{pmatrix}$$

(mit quadratischen Teilmatrizen in der Diagonale) oder kann durch Umstellung der Zeilen und gleichlautende Umstellung der Spalten auf diese Form gebracht werden.

III. Es sei $A=(a_{jk})$ eine unzerlegbare Matrix mit nicht negativen Elementen, $B=(b_{jk})$ eine Matrix mit komplexen Elementen $(j,k=1,2,\ldots,n)$. Es sei $|b_{jk}| \leqq a_{jk}$ für alle j, k. Ist λ die Maximalwurzel von A und β ein beliebiger Eigenwert von B, so ist $|\beta| \leqq \lambda$. Gilt das Gleichheitszeichen, ist also $\beta = \lambda \exp(i\varphi)$, so hat die Gestalt $B = \exp(i\varphi)\, DAD^{-1}$ mit einer Diagonalmatrix D, deren Diagonalelemente sämtlich den absoluten Betrag 1 haben; insbesondere ist dann stets $|b_{jk}| = a_{jk}$.

For any function $u \in (R^p)^X$: $X \ni x \to u(x) \in R^p$ and any $\alpha \in R^p$ we define the set $T_u \subset R^p$, the diagonal operator $C^X \ni g \to \Delta_u(\alpha)\, g \in C^X$, and the additive group $G_\alpha \subset R^p$ by

$$\begin{aligned} T_u &= \{t(x,y) - (u(x) - u(y)) \mid (x,y) \in Y\} \\ \Delta_u(\alpha)\, g(x) &= e^{i\alpha \cdot u(x)}\, g(x) \\ G_\alpha &= \{t \in R^p \mid e^{i\alpha \cdot t} = 1\}. \end{aligned} \tag{2.4}$$

With this notation we have the following consequences of I, II and III (Lemmas 2.1 and 2.2).

Lemma 2.1. *For each $a \in R^p$ there is a positive number $\lambda(a)$, and a strictly positive sequence $\{e_a(x)\}_{x \in X}$ such that the numbers $\lambda(a)\, e^{i2\pi j/r}$, $j=0,\ldots,r-1$ are simple eigenvalues of $\phi(a)$ corresponding to the eigenfunctions $e_a(x)\, e^{i2\pi j r(x)/r}$, $j=0,\ldots,r-1$*

respectively. The absolute value of any other eigenvalue of $\phi(a)$ is strictly less than $\lambda(a)$.

Lemma 2.2. *Let $\lambda(a)$ denote the maximal positive eigenvalue of $\phi(a)$. The spectrum of $\Phi(a+i\alpha)$ is contained in the disc $\{z\in C \mid |z|\leqq\lambda(a)\}$, and the following three statements are equivalent:*

(a) *$e^{i\alpha\cdot s}\lambda(a)$ $(s\in R^p)$ is an eigenvalue of $\Phi(a+i\alpha)$.*

(b) *$\Phi(z+i\alpha)=e^{i\alpha\cdot s}\Delta_u(\alpha)\,\Phi(z)\,\Delta_u(\alpha)^{-1}$ for some $u\in(R^p)^X$ and all $z\in C^p$.*

(c) *$T_u\subset s+G_\alpha$ for some $u\in(R^p)^X$.*

The proper definition of T_u in the m-dependence case is

$$T_u=\{t(x_1,\dots,x_m)-[u(x_1,\dots,x_{m-1})-u(x_2,\dots,x_m)] \mid (x_1,\dots,x_m)\in Y\},\qquad u\in(R^p)^{X^{m-1}}.$$

In particular when $m=1$, $T_u=T=\{t(x)\mid q_x>0\}$ for all $u\in R^p=(R^p)^{X^0}$. Compare Lemma 3, p. 475 of [5].

Proof of Lemma 2.1. It is clear that if Y is irreducible in our sense, then $\Phi(a)$ is irreducible in the sense of Wielandt. A glance at I and II shows that it remains to verify that

$$\Phi(a)\,e_a^j(x)=\lambda_j(a)\,e_a^j(x) \tag{2.5}$$

where $e_a^j(x)=e_a(x)\,e^{i2\pi j r(x)/r}$ and $\lambda_j(a)=e^{i2\pi j/r}$. Since $\varphi_{xy}(a)=0$ unless $y\in X_{r(x)+1}$ we have

$$\begin{aligned}\Phi(a)\,e_j^a(x)&=e^{i2\pi j(r(x)+1)/r}\Big(\sum_{y\in X_{r(x)+1}}\varphi_{xy}(a)\,e_a(y)\Big)\\&=e^{i2\pi j(r(x)+1)/r}\,\lambda(a)\,e_a(x)=\lambda_j(a)\,e_a^j(x).\end{aligned}$$

Proof of Lemma 2.2. If $T_u\subset s+G_\alpha$, then $t(x,y)=s+u(x)-u(y)+t^*(x,y)$, where t^* satisfies $e^{i\alpha\cdot t^*(x,y)}=1$. Hence $e^{(z+i\alpha)\cdot t(x,y)}=e^{z\cdot t(x,y)}\,e^{i\alpha\cdot s}\,e^{i\alpha\cdot u(x)}\,e^{-i\alpha\cdot u(y)}$, $(x,y)\in Y$. That is $\Phi(z+i\alpha)=e^{i\alpha\cdot s}\Delta_u(\alpha)\,\Phi(z)\,\Delta_u(\alpha)^{-1}$. Hence (c) implies (b). If (b) holds, then $\Phi(a+i\alpha)(\Delta_u(\alpha)\,e_a(x))=e^{i\alpha\cdot s}\Delta_u(\alpha)(\Phi(a)\,e_a(x))=e^{i\alpha\cdot s}\lambda(a)\,\Delta_u(\alpha)\,e_a(x)$. So that $e^{i\alpha\cdot s}\lambda(a)$ is an eigenvalue of $\Phi(a+i\alpha)$. Hence (b) implies (a).

It remains to show that (a) implies (c). Suppose $\alpha\neq 0$, for otherwise there is nothing to prove. It follows from III that there is a diagonal matrix Δ and a $s\in R^p$ such that $\Phi(a+i\alpha)=e^{i\alpha\cdot s}\Delta\Phi(a)\,\Delta^{-1}$ which is equivalent to

$$e^{i\alpha\cdot t(x,y)}=e^{i\alpha\cdot s}\,e^{i\delta(x)}\,e^{-i\delta(y)}$$

for some function $\delta\in R^X$. Choose $u\in(R^p)^X$ such that $\alpha\cdot u(x)=\delta(x)$. Lemma 2.2 is proved.

We count the $N=|X|=$(cardinality of X) eigenvalues of $\Phi(z)$ repeatedly according to their (algebraic) multiplicities and let $\Lambda(\Phi(z))$ denote the unordered N-tuple consisting of the N repeated eigenvalues of $\Phi(z)$. If $\Lambda=(\lambda_1,\dots,\lambda_N)$ and $\Lambda'=(\lambda_1',\dots,\lambda_N')$ we define

$$d(\Lambda,\Lambda')=\min_{\sigma\in\pi}\ \max_{j=1,\dots,N}|\lambda_j-\lambda_{\sigma(j)}|$$

where π is the set of all permutations of the numbers $1,\dots,N$. Then d is *a* distance function. Lemma 2.3 contains the perturbation theory we will need.

Lemma 2.3. (i) *For each matrix $M(z)$ which is analytic in $D\subset C^p$, and for each*

$$z_0\in D\ d(\Lambda(M(z)),\Lambda(M(z_0)))\to 0 \quad \text{as } z\to z_0.$$

(ii) *There is an open neighbourhood U of R^p (considered as a subset of C^p), and analytic functions $U \ni z \to \lambda(z) \in C$ and $U \ni z \to E^j(z)$, where E^j is operator valued, $j=0, \ldots, r-1$ such that $\lambda(z)\exp(i2\pi j/r)$ is a simple eigenvalue of $\Phi(z)$ and $E^j(z)$ is the corresponding eigenprojection for each $z \in U$ and $j=0, \ldots, r-1$. Furthermore, $\lambda(a)$ coincides with the maximal positive eigenvalue of $\Phi(a)$ for each $a \in R^p$.*

Proof. (i) follows directly from Theorem 5.14, p. 118 of Kato [8].

To show (ii) we note that the analytic function $C \times C^p \ni (w, z) \to p(w; z) = \det(\Phi(z) - wI)$ is a polynomial in w and that the eigenvalues of $\Phi(z)$ are the roots of this polynomial. Since the maximal positive eigenvalue $\lambda(a)$ is a simple eigenvalue of $\Phi(a)$ i.e. a simple root of $p(w; z)$ we have $\left.\frac{d}{dw} p(w, a)\right|_{w=\lambda(a)} \neq 0$, and it follows from the implicit function theorem (see [2], p. 138) that there is an eigenvalue $\lambda_a(z)$ of $\Phi(z)$ which is analytic in a neighbourhood of a and which satisfies $\lambda_a(a) = \lambda(a)$. If we recall (i) we therefore see that to each $a \in R^p$ there is an open neighbourhood U_a of a and a function $\lambda_a(z)$ with the following properties:

(α) $\lambda_a(z)$ is analytic in U_a.

(β) $\lambda_a(z)$ is a simple eigenvalue of $\Phi(z)$ for each $z \in U_a$, and $\lambda_a(a) = \lambda(a)$.

(γ) $\operatorname{Re}\lambda_a(z)$ is larger than the real part of any other eigenvalue of $\Phi(z)$ ($z \in U_a$). (Here we made use of (i).)

If $z \in U_a \cap U_b$, then $\lambda_a(z) = \lambda_b(z)$ (for otherwise $\operatorname{Re}\lambda_a(z) > \operatorname{Re}\lambda_b(z)$ and $\operatorname{Re}\lambda_b(z) > \operatorname{Re}\lambda_a(z)$), and we may therefore unambiguously define the function $\bigcup_{a \in R^p} U_a \ni z \to l(z)$ by $l(z) = \lambda_a(z)$ if $z \in U_a$. Then $l(z)$ is an eigenvalue of $\phi(z)$ which is analytic in the open set $U = \bigcup_{a \in R^p} U_a$. If we note that $l(a) = \lambda_a(a) = \lambda(a)$, and remember the verification of (2.5) we therefore see that $l(z)$ may serve as the function $\lambda(z)$ of the lemma.

Concerning the analyticity of the eigenprojections we refer to Kato [8], p. 67. Kato treats just the case $p=1$ but the proof applies to a general $p \geqq 1$. This completes the proof of Lemma 2.3.

From now on we let $\lambda(z)$ stand for the function λ of Lemma 2.3. This lemma permits us to introduce the notations

$$\begin{aligned} m_a &= \operatorname{grad} \log \lambda(a) \\ V_a &= \left(\frac{\partial^2}{\partial a_i \, \partial a_j} \log \lambda(a)\right). \end{aligned} \tag{2.6}$$

The reader may think of m_a and V_a as asymptotic expectation and covariance matrix respectively.

Lemma 2.4. *V_a is a positive semidefinite matrix for each $a \in R^p$, and the following two statements are equivalent:*

(a) *There is an $u \in (R^p)^X$ such that T_u is contained in some cofactor of the additive group $G_\alpha^0 = \{t \in R^p \mid \alpha \cdot t = 0\}$.*

(b) *$\alpha \cdot V_a \alpha = 0$ for all $a \in R^p$.*

It would be preferable to add a third equivalence to the lemma:

(c) *$\alpha \cdot V_a \alpha = 0$ for some $a \in R^p$.*

However, I am not able to prove it. Instead we have to impose an additional assumption in Section 3.

Proof of Lemma 2.4. For fixed a and α we choose a disc $W \subset C$ with center at the origin so small that $W \ni w \to \lambda(a+w\alpha)$ is analytic in W, $\operatorname{Re}\lambda(a+w\alpha)>0$, and $\lambda(a+w\alpha)$ is a simple eigenvalue for $w \in W$. Put $f(w)=\log\lambda(a+w\alpha)$, then f is analytic in W and $f''(w)=\alpha\cdot V_{a+w\alpha}\alpha$.

It follows from Lemma 2.2 that $\operatorname{Re}(f(i\eta)-f(0))\leqq 0\,(\eta\in R\cap W)$, and the Taylor expansion shows that $\operatorname{Re}(f(i\eta)-f(0))=-\eta^2\alpha\cdot V_a\alpha+O(|\eta|^3)$. Hence $\alpha\cdot V_a\alpha\geqq 0$. a and α being arbitrary we conclude V_a is positive semidefinite for each $a\in R^p$.

$G_\alpha^0=\bigcap_{\eta\in R\cap W} G_{\eta\alpha}$, and hence $T_u\subset s+G_\alpha^0$ if and only if $T_u\subset s+G_{\eta\alpha}$ for all $\eta\in R\cap W$. According to Lemma 2.2 the latter is the case if and only if $\lambda(a)\exp(i\eta\alpha\cdot s)$ is an eigenvalue of $\Phi(a+i\eta\alpha)$ for all $\eta\in R\cap W$. This eigenvalue coincides with $\lambda(a+i\eta\alpha)$ at $\eta=0$. Since $\lambda(a+i\eta\alpha)$ is a simple eigenvalue as $\eta\in W$ we therefore must have $\lambda(a+i\eta\alpha)=\lambda(a)\exp(i\eta\alpha\cdot s)$ when η belongs to some real neighbourhood of the origin, and hence for all $i\eta\in W$ (because both sides are analytic in η).

We have thus shown that part (*a*) of the lemma is equivalent to: $f''(w)=0$ for all $w\in W$, i.e. $f''(\xi)=0$ for all $\xi\in R\cap W$. The lemma is proved since a was arbitrary.

Let $\Phi^*(a)$ stand for the transpose of $\Phi(a)$

$$\Phi^*(a)\,g(x)=\sum_{y\in X}\varphi_{yx}(a)\,g(y). \tag{2.7}$$

Then $\Phi^*(a)$ has properties similar to those of $\Phi(a)$: $\Phi^*(a)$ is non-negative and irreducible with period r, and the eigenvalues of $\Phi^*(a)$ coincide with those of $\Phi(a)$. It follows from the preceding theory that there is a strictly positive sequence $e_a^*=\{e_a^*(x)\}_{x\in X}$ such that

$$\Phi^*(a)\,e_{a,j}^*=\lambda(a)\,e^{-i2\pi j/r}\,e_{a,j}^*,\qquad j=0,\dots,r-1 \tag{2.8}$$

where

$$e_{a,j}^*(x)=e_a^*(x)\exp(-i2\pi j\,r(x)/r),\qquad j=0,\dots,r-1. \tag{2.9}$$

For any two sequences $f=\{f(x)\}_{x\in X}$ and $g=\{g(x)\}_{x\in X}$ we define

$$f\cdot g=\sum_{x\in X}f(x)\,g(x). \tag{2.10}$$

It is well known that the eigenprojections $E^j(a)$ are given by

$$E^j(a)\,g(x)=\frac{e_{a,j}^*\cdot g}{e_{a,j}^*\cdot e_{a,j}}\,e_{a,j}(x) \tag{2.11}$$

where $e_{a,j}(x)=e_a(x)\exp(i2\pi j\,r(x)/r)$, and also that

$$E^j(a)\,E^k(a)=\delta_{jk}\,E^k(a). \tag{2.12}$$

If we use (2.11) and (2.12) and note that $e_{a,j}^*\cdot e_{a,j}=e_a^*\cdot e_a$ we conclude

$$\sum_{k=0}^{r-1}\exp(i2\pi h k/r)\sum_{x\in X_k}\frac{e_a^*(x)\,e_a(x)}{e_a^*\cdot e_a}=\delta(h),\qquad h=0,\dots,r-1, \tag{2.13}$$

where $\delta(h)=1$ if $h=0$, $=0$ otherwise. But (2.13) cannot hold unless

$$\sum_{x\in X_k} \frac{e_a^*(x)\, e_a(x)}{e_a^* \cdot e_a} = \frac{1}{r}, \qquad k=1, \ldots, r. \tag{2.14}$$

3. The Saddle Point Approximation

Except when it is apparent from the context, we will in this and the following sections suppose that the regularity condition below is satisfied.

Condition. We require that it is not the case that there is an $\alpha \neq 0$ in $(-\pi, \pi]^p$ and an $u \in (R^p)^X$ such that T_u is contained in some coset of G_α.

Since $G_\alpha^0 \subset G_\alpha$ we conclude (Lemma 2.4):

$$V_a = \left(\frac{\partial^2}{\partial a_i\, \partial a_j} \log \lambda(a)\right)$$

is a positive semidefinite matrix for each $a \in R^p$, and for each $0 \neq \alpha \in R^p$ there is an $a \in R^p$ such that $\alpha \cdot V_a \alpha > 0$. It is believed but not proved that the condition above also implies:

Additional Assumption. V_a is a positive definite matrix for each $a \in R^p$.

This assumption implies that the mapping $a \to m_a = \operatorname{grad} \log \lambda(a)$ is one to one. Thus we may write

$$\hat{a} = m^{-1}. \tag{3.1}$$

That is $m_{\hat{a}(x)} = x$ for all $x \in m(R^p) = \{m_a \in R^p \mid a \in R^p\}$.

The tensor product $F_1 \otimes \cdots \otimes F_k = \bigotimes_{j=1}^{k} F_j$ of mappings $F_j = (f_{xy}^j)\, j=1, \ldots, k$ from C^X into C^X is defined to be the mapping from C^{X^k} into C^{X^k} which satisfies

$$\bigotimes_{j=1}^{k} F_j\, g(x_1, \ldots, x_k) = \sum_{(y_1, \ldots, y_k)\in X^k} f_{x_1 y_1}^1 \cdot \cdots \cdot f_{x_k y_k}^k\, g(y_1, \ldots, y_k). \tag{3.2}$$

This multiplication induces another convolution operation (denoted $\circledast$) on structure functions $F_1(t), \ldots, F_k(t)$

$$F_1 \circledast F_2 \circledast \cdots \circledast F_k(t) = \sum_{t_1+\cdots+t_k=t} F_1(t_1) \otimes \cdots \otimes F_k(t_k). \tag{3.3}$$

The convolution (3.3) is thus a mapping from C^{X^k} into C^{X^k}. If $F_j(t) = \left(f_{xy}^j(t)\right)$ $j=1, \ldots, k$ and $g \in C^{X^k}$, then

$$F_1 \circledast \cdots \circledast F_k(t)\, g(x_1, \ldots, x_k) = \sum_{(y_1, \ldots, y_k)\in X^k} f_{x_1 y_1}^1 * \cdots * f_{x_k y_k}^k(t)\, g(y_1, \ldots, y_k) \tag{3.4}$$

where we this time have ordinary convolution to the right in (3.4).

We further define $E_n(a) = \sum_{j=0}^{r-1} \exp(i 2\pi j n/r)\, E^j(a)$, then

$$E_n(a)\, g(x) = r(e_a^* \cdot e_a)^{-1} \Big(\sum_{y \in X_{r(x)+n}} e_a^*(y)\, g(y) \Big)\, e_a(x).$$

Also

$$E_{n+kr}(a)=E_n(a), \qquad k\in Z. \tag{3.6}$$

$|z|$ $(z\in C^p)$ stands for the euclidian norm $(|z_1|^2+\cdots+|z_p|^2)^{\frac{1}{2}}$, and $\|\cdot\|$ for the norms

$$\|(f_{\mathbb{X}\mathbb{Y}})\|=\sum_{\substack{\mathbb{X}\in X^k\\ \mathbb{Y}\in X^k}}|f_{\mathbb{X}\mathbb{Y}}|, \qquad k\geqq 1.$$

Then

$$\begin{aligned}\|F_1\otimes\cdots\otimes F_k\|&=\|F_1\|\cdot\cdots\cdot\|F_k\|\\ \|F_1\cdot\cdots\cdot F_k\|&\leqq\|F_1\|\cdot\cdots\cdot\|F_k\|.\end{aligned} \tag{3.8}$$

Theorem 3.1. *For each integer $k\geqq 1$ and each compact $K\subset R^p$*

$$\begin{aligned}\Big\|&(2\pi n)^{p/2}(\det V_a)^{\frac{1}{2}}\frac{e^{a\cdot t}}{\lambda(a)^n}F^{n_1}*\circledast\cdots\circledast F^{n_k}*(t)\\ &-e^{-\frac{1}{2}|t^*|^2}\Big[(1+n^{-\frac{1}{2}}P_a(t^*))\bigotimes_{j=1}^{k}E_{n_j}(a)+t^*\cdot(nV_a)^{-\frac{1}{2}}\operatorname{grad}\bigotimes_{j=1}^{k}E_{n_j}(a)\Big]\Big\|=O(1/n),\end{aligned} \tag{3.9}$$

uniformly in t and in a when $a\in K$ and n_j/n, $j=1,\ldots,k$ remain bounded away from zero and one. Here $n=n_1+\cdots+n_k$, $t^*=(nV_a)^{-\frac{1}{2}}(t-nm_a)$, and P_a is a polynomial of third degree with coefficients which are analytic functions of a. Furthermore $P_a(0)=0$.

Remark 1. The rightmost operator of the second line in (3.9) is defined via the formula

$$\begin{aligned}&\Big[t^*\cdot(nV_a)^{-\frac{1}{2}}\operatorname{grad}\bigotimes_{1}^{k}E_{n_j}(a)\Big]g(x_1,\ldots,x_k)\\ &\quad=t^*\cdot(nV_a)^{-\frac{1}{2}}\operatorname{grad}\Big[\bigotimes_{1}^{k}E_{n_j}(a)\,g(x_1,\ldots,x_k)\Big].\end{aligned} \tag{3.10}$$

Remark 2. P_a is given by

$$e^{-\frac{1}{2}|x|^2}P_a(x)=\tfrac{1}{6}(2\pi)^{-p/2}\int_{R^p}e^{-i\alpha\cdot x}e^{-\frac{1}{2}|\alpha|^2}Q_a(V_a^{-\frac{1}{2}}\alpha)\,d\alpha \qquad (x\in R^p) \tag{3.11}$$

where $Q_a(x)=\left(\sum_1^p ix_j\frac{\partial}{\partial a_j}\right)^3\log\lambda(a)$. In particular

$$P_a(x)=\frac{1}{6}(x^3-3x)\,V_a^{-\frac{3}{2}}\frac{d}{da}V_a$$

when $p=1$.

Theorem 3.1 is the Markov version of a result of [12] (which is the independence version of the present result, but without the extra assumption that X is finite). The starting point of the considerations presented here was a question of Per Martin-Löf concerning an extension of his result to the Markov case.

Corollary 3.1. *For each integer $k\geqq 1$ and each compact $K\in R^p$*

$$\begin{aligned}\Big\|(2\pi n)^{p/2}(\det V_a)^{\frac{1}{2}}\frac{e^{a\cdot t}}{\lambda(a)^n}F^{n_1}*\circledast\cdots\circledast F^{n_k}*(t)-e^{-\frac{1}{2}|t^*|^2}\bigotimes_{j=1}^{k}E_{n_j}(a)\Big\|\\ =O(n^{-\frac{1}{2}}),\end{aligned} \tag{3.12}$$

uniformly in t and in a when $a \in K$ and n_j/n, $j=1, \ldots, k$ remain bounded away from zero and one.

Proof. The properties of P_a mentioned in the formulation of Theorem 3.1 give $\max_{a\in K} |P_a(x)| \leqq \text{Const}\,|x|\,(1+|x|^2)$. Here and below Const stands for a number that may depend on k and K but not on a, t or $n_1, \ldots, n_k$. Since $\left\|\bigotimes_1^k E_{n_j}(a)\right\|$ is dominated by the continuous function $\left(\sum_{j=1}^r \|E_j(a)\|\right)^k$ for all n_j we also have

$$\max_{a\in K} \left\|\bigotimes_1^k E_{n_j}(a)\right\| \leqq \text{Const}.$$

Therefore

$$\max_{t\in Z^p,\, a\in K} \left\| e^{-\frac{1}{2}|t^*|^2} P_a(t^*) \bigotimes_1^k E_{n_j}(a)\right\| \leqq \text{Const} \max_{x\in R^p} e^{-\frac{1}{2}|x|^2} |x|\,(1+|x|^2) \leqq \text{Const}.$$

We treat the term containing the gradient in a similar way, and the corollary follows.

For reasons explained in Section 4

$$H_a = \log \lambda(a) - a \cdot m_a \tag{3.13}$$

will be called the specific entropy. If we let $a=\hat{a}(t/n)$ in Theorem 3.1 and note that $t - n m_{\hat{a}(t/n)} = 0$ we obtain:

Corollary 3.2. *For each integer $k \geqq 1$ and each compact $K \subset R^p$*

$$\left\|(2\pi n)^{p/2} (\det V_{\hat{a}(t/n)})^{\frac{1}{2}} e^{-nH_{\hat{a}(t/n)}} F^{n_1} * \circledast \cdots \circledast F^{n_k} * (t) - \bigotimes_{j=1}^k E_{n_j}(\hat{a}(t/n))\right\| = O(1/n), \tag{3.14}$$

uniformly in t when $\hat{a}(t/n) \in K$ and n_j/n, $j=1, \ldots, k$ remain bounded away from zero and one.

In the independence case it is possible to replace the approximation above by an approximation which holds uniformly for *all* $t \in Z$. See [7].

Proof of Theorem 3.1. Given the compact K, choose $\delta > 0$ and $\eta > 0$ so small that

(i) $K_\delta = \{a + i\alpha \mid a \in K, |\alpha| \leqq \delta\}$ is contained in the set U of Lemma 2.3.

(ii) $\log \lambda(a+i\alpha)$ has no branching point for $a + i\alpha \in K_\delta$.

(iii)
$$|\log \lambda(a+i\alpha) - \log \lambda(a) - i\alpha \cdot m_a + \tfrac{1}{2}\alpha \cdot V_a \alpha| \leqq \tfrac{1}{4}\alpha \cdot V_a \alpha \tag{3.15}$$

when $a \in K$, $|\alpha| \leqq \delta$.

(iv)
$$\sigma(\Phi(a+i\alpha)) - \bigcup_{j=1}^r \{\lambda(a+i\alpha)\exp(i2\pi j/r)\} \subset \{w \in C \mid |w| < (I-\eta)|\lambda(a+i\alpha)|\} \tag{3.16}$$

and $|\lambda(a+i\alpha)| \geqq \eta$ as $a \in K$, $|\alpha| \leqq \delta$. Here and below $\sigma(\Phi(z))$ denotes the spectrum of $\Phi(z)$.

(v) $\sigma(\Phi(a+i\alpha)) \subset \{w \in C \mid |w| < (1-\eta)\lambda(a)\}$ when $a \in K$, $|\alpha| \geqq \delta$ and $\alpha \in (-\pi, \pi]^p$.

It is seen in the following way that such a choise is possible.

(i) Every open set which K contains K_δ for all δ sufficiently small.

(ii) Put $\eta=\frac{1}{2}\min_{a\in K}\lambda(a)$, then $\eta>0$. $\log\lambda(a+i\alpha)$ has no branching point in the open set $\{a+i\alpha|\operatorname{Re}\lambda(a+i\alpha)>\eta\}$ which contains K and hence also K_δ for all δ sufficiently small.

(iii) The expression to the left in (3.15) equals (according to Taylors formula) $O(|\alpha|^3)$ as $\alpha\to 0$, uniformly when $a\in K$. V_a is a strictly positive definite matrix for each $a\in R^p$, and it follows from part (*a*) of Lemma 2.3 (applied to V_a) that the smallest eigenvalue of V_a is a continuous function of a. The smallest eigenvalue of V_a is therefore bounded away from zero when $a\in K$. Hence $\alpha\cdot V_a\alpha\geqq c|\alpha|^2$ for all $a\in K$, $\alpha\in R^p$ and some $c>0$. We conclude that the inequality (3.15) is valid for all $a\in K$ and $|\alpha|\leqq\delta$, provided δ is sufficiently small.

(iv) Write $\sum(a+i\alpha)$ for the set to the left in (3.16), and $D(c)$ for the open disc $\{w\in C^p|\,|w|<c\}$. Our first aim is to show

$$\sum(a+i\alpha)\subset D\big(\lambda(a)\big)\qquad (a\in R^p,\ \alpha\in R^p). \tag{3.17}$$

When $\alpha\neq 0$ (3.17) follows from Lemma 2.2 and the regularity condition (p. 14). When $\alpha=0$, (3.17) is a consequence of Lemma 2.1.

It now follows from the continuity of the spectrum (part (*a*) of Lemma 2.1) that to each $b\in R^p$ there is an open neighbourhood $U_b\subset C^p$ of b and a number $\eta_b>0$ such that $\sum(a+i\alpha)\subset D\big((1-\eta_b)|\lambda(a+i\alpha)|\big)$ for all $a+i\alpha\in U_b$. Choose δ so small that $K_\delta\subset\bigcup_{b\in R^p}U_b$, and let $U_{b_1},\dots,U_{b_N}$ be a finite covering of K_δ, then

$$\sum(a+i\alpha)\subset D\big((1-\eta)|\lambda(a+i\alpha)|\big)$$

for all $a\in K$, $|\alpha|\leqq\delta$ and $\eta\leqq\min(\eta_{b_1},\dots,\eta_{b_N})$

(v) It follows from Lemma 2.2 and the regularity condition that

$$\sigma\big(\Phi(a+i\alpha)\big)\subset D\big(\lambda(a)\big)$$

when $a\in K$, $|\alpha|\geqq\delta$ and $\alpha\in[-\pi,\pi]^p$. The set $\{a+i\alpha|a\in K,|\alpha|\geqq\delta,\alpha\in[-\pi,\pi]^p\}$ is however compact and it follows in a similar way as in the proof of (iv) that

$$\sigma\big(\Phi(a+i\alpha)\big)\subset D\big((1-\eta)\,\lambda(a)\big)\quad\text{for all }\ a\in K,\ |\alpha|\geqq\delta,\qquad \alpha\in[-\pi,\pi]^p$$

and all η sufficiently small.

Let $n=n_1+\dots+n_k$ and

$$\begin{aligned}
\psi_1(\alpha)&=(2\pi)^{-p}\,e^{-i\alpha\cdot t}\,\lambda(a)^{-n}\bigotimes_{j=1}^{k}\Phi(a+i\alpha)^{n_j}\\
\psi_2(\alpha)&=(2\pi)^{-p}\,e^{-i\alpha\cdot t}\left(\frac{\lambda(a+i\alpha)}{\lambda(a)}\right)^{n}\bigotimes_{j=1}^{k}E_{n_j}(a+i\alpha)\\
\psi_3(\alpha)&=(2\pi)^{-p}\,e^{-i\alpha\cdot(t-nm_a)}\,e^{-\frac{1}{2}n\alpha\cdot V_a\alpha}\left[\big(1+\tfrac{1}{2}nQ_a(\alpha)\big)\bigotimes_{j=1}^{k}E_{n_j}(a)+i\alpha\cdot\operatorname{grad}\bigotimes_{j=1}^{k}E_{n_j}(a)\right]
\end{aligned}\tag{3.18}$$

where $Q_a(\alpha)$ is defined just below (3.11). Then

$$\frac{e^{a\cdot t}}{\lambda(a)^n} F^{n_1 *} \circledast \cdots \circledast F^{n_k *}(t) = \int_{(-\pi,\pi]^p} \psi_1(\alpha)\, d\alpha \tag{3.19}$$

and

$$\begin{aligned}(2\pi n)^{-p/2}(\det V_a)^{-\frac{1}{2}} e^{-\frac{1}{2}|t^*|^2} \Big[\big(1+n^{-\frac{1}{2}} P_a(t^*)\big) \bigotimes_{j=1}^{k} E_{n_j}(a) \\ + t^* \cdot (n V_a)^{-\frac{1}{2}} \operatorname{grad} \bigotimes_{j=1}^{k} E_{n_j}(a)\Big] = \int_{R^p} \psi_3(\alpha)\, d\alpha.\end{aligned} \tag{3.20}$$

Hence the norm in (3.9) equals

$$(2\pi n)^{p/2} (\det V_a)^{\frac{1}{2}} \Big\| \int_{(-\pi,\pi]^p} \psi_1(\alpha)\, d\alpha - \int_{R^p} \psi_3(\alpha)\, d\alpha \Big\|, \tag{3.21}$$

which is dominated by

$$\begin{aligned}(2\pi n)^{p/2} (\det V_a)^{\frac{1}{2}} \Big[&\int_{\substack{\alpha\in(-\pi,\pi]^p \\ |\alpha|>\delta}} \|\psi_1(\alpha)\|\, d\alpha + \int_{|\alpha|>\delta} \|\psi_3(\alpha)\|\, d\alpha \\ &+ \int_{|\alpha|<\delta} \|\psi_1(\alpha)-\psi_2(\alpha)\|\, d\alpha + \int_{|\alpha|<\delta} \|\psi_2(\alpha)-\psi_3(\alpha)\|\, d\alpha \Big] \\ &= I_1 + \cdots + I_4.\end{aligned} \tag{3.22}$$

We choose a number $0<\varepsilon<\frac{1}{2}$ and have to show that $I_j = O(1/n)$ for $j=1,2,3,4$, uniformly in t and in a when $a\in K$ and $\varepsilon \leqq n_j/n \leqq 1-\varepsilon$ for $j=1,\dots,k$. The symbol "Const" will be used for numbers that may depend on k, K, ε, δ and η but not on t, α, a or $n_1,\dots,n_k$ when $a\in K$ and $\varepsilon \leqq n_j/n \leqq 1-\varepsilon, j=1,\dots,k$. We will begin with I_4.

I_4. Note that I_4 may be written

$$(2\pi n)^{p/2} (\det V_a)^{\frac{1}{2}} \int_{|\alpha|<\delta} \big\| e^{i\alpha\cdot(t-nm_a)} \big(\psi_2(\alpha)-\psi_3(\alpha)\big)\big\|\, d\alpha. \tag{3.23}$$

Write $r(a,\alpha)$ for the expression within the absolute value signs in (3.15). Then we have

$$e^{\frac{1}{2}n\alpha\cdot V_a \alpha}\, e^{-i\alpha\cdot nm_a} \big(\lambda(a+i\alpha)/\lambda(a)\big)^n = e^{nr(a,\alpha)}. \tag{3.24}$$

But $|e^z-1-w| \leqq |e^z-1-z| + |z-w| \leqq \frac{1}{2}|z|^2 e^{|z|} + |z-w| \leqq e^{|z|}(|z|+|z-w|)$. If we apply this inequality with $z=n\,r(a,\alpha)$, $w=\frac{n}{6} Q_a(\alpha)$, remember (iii) and note that $Q_a(\alpha) = O(|\alpha|^3)$, $r(a,\alpha) - \frac{1}{6} Q_a(\alpha) = O(|\alpha|^4)$ uniformly when $a\in K$, then we conclude

$$\Big| e^{nr(a,\alpha)} - 1 - \frac{n}{6} Q_a(\alpha) \Big| \leqq \text{Const}\, e^{\frac{n}{6}\alpha\cdot V_a\alpha} (n|\alpha|^4 + n^2|\alpha|^6) \tag{3.25}$$

when $a\in K$. Also

$$\Big\| \bigotimes_{j=1}^{k} E_{n_j}(a+i\alpha) - \bigotimes_{j=1}^{k} E_{n_j}(a) + i\alpha\cdot \operatorname{grad} \bigotimes_{j=1}^{k} E_{n_j}(a) \Big\| \leqq \text{Const}\, |\alpha|^2. \tag{3.26}$$

Hence

$$\begin{aligned}\Big\| \Big(1+\frac{n}{6} Q_a(\alpha)\Big) \bigotimes_{j=1}^{k} E_{n_j}(a) + i\alpha\cdot \operatorname{grad} \bigotimes_{j=1}^{k} E_{n_j}(a) - e^{nr(a,\alpha)} \bigotimes_{j=1}^{k} E_{n_j}(a+i\alpha) \Big\| \\ \leqq \text{Const}\, e^{\frac{n}{4}\alpha\cdot V_a\alpha} |\alpha|^2 (1+n^2|\alpha|^4)\end{aligned} \tag{3.27}$$

when $a \in K$ and $|\alpha| \leqq \delta$. Therefore

$$I_4 \leqq \text{Const}\, n^{p/2} (\det V_a)^{\frac{1}{2}} \int_{|\alpha|<\delta} e^{-\frac{n}{4}\alpha \cdot V_a \alpha} |\alpha|^2 (1+n^2 |\alpha|^4)\, d\alpha. \tag{3.28}$$

The substitution $\alpha \to (n V_a)^{-\frac{1}{2}} \alpha$ together with the inequality $|V_a^{-\frac{1}{2}} \alpha| \leqq \text{Const}\, |\alpha|$ finally yields $I_4 \leqq \text{Const}/n$.

I_3. We have $I_3 \leqq \text{Const}\, n^{p/2} I'_3$, where

$$I'_3 = \int_{|\alpha|<\delta} \lambda(a)^{-n} \left\| \bigotimes_{j=1}^{k} \Phi(a+i\alpha)^{n_j} - \bigotimes_{j=1}^{k} \lambda(a+i\alpha)^{n_j} E_{n_j}(a+i\alpha) \right\| d\alpha. \tag{3.29}$$

Let $A_j = \Phi(a+i\alpha)^{n_j}$ and $B_j = \lambda(a+i\alpha)^{n_j} E_{n_j}(a+i\alpha)$ then

$$\bigotimes_1^k A_j - \bigotimes_1^k B_j = \sum_{j=1}^{k} A_1 \otimes \cdots \otimes A_{j-1} \otimes (A_j - B_j) \otimes B_{j+1} \otimes \cdots \otimes B_k$$

and hence

$$\left\| \bigotimes_1^k A_j - \bigotimes_1^k B_j \right\| \leqq \sum_{j=1}^{k} \|A_1\| \ldots \|A_{j-1}\|\, \|A_j - B_j\|\, \|B_{j+1}\| \ldots \|B_k\|.$$

We are going to show that $\|A_j - B_j\|$ $j=1, \ldots, k$ are small.

Let $\Gamma(a+i\alpha)$ denote the circle with radius $(1-\eta/2)|\lambda(a+i\alpha)|$ and center 0. Choose circles $\gamma_j(a+i\alpha)$ with centers at $\lambda(a+i\alpha)\exp(i 2\pi j/r)$ $j=1, \ldots, r$ and with so small radii that the sets $\Gamma(a+i\alpha)$, $\gamma_1(a+i\alpha), \ldots, \gamma_r(a+i\alpha)$ are disjoint.

The resolvent $R(w, a+i\alpha) = (\Phi(a+i\alpha) - wI)^{-1}$ $(w \in C)$ is a meromorphic function of w the poles of which coincides with $\sigma(\Phi(a+i\alpha))$. $\lambda(a+i\alpha)\exp(i2\pi j/r)$ $j=1, \ldots, r$ are simple poles since these eigenvalues are simple (remember (i) and Lemma 2.3).

If we remember (2.11) and (2.9) we see that if $z \in R^p$ then

$$E_m(z) = \sum_{j=1}^{r} \exp(i 2\pi j m/r) E^j(z). \tag{3.30}$$

The identity (3.30) defines $E_m(z)$ in a neighbourhood of R^p and we have (see [6] 1966, p. 39 and p. 44)

$$\begin{aligned} E^j(z) &= -(2\pi i)^{-1} \int_{\gamma_j(z)} R(w,z)\, dw \\ \Phi(z)^n &= -(2\pi i)^{-1} \int_{\Gamma(z)+\gamma_1(z)+\cdots+\gamma_r(z)} w^n R(w,z)\, dw. \end{aligned} \tag{3.31}$$

Hence

$$\begin{aligned} \Phi(a+i\alpha)^m - \lambda(a+i\alpha)^m E_m(a+i\alpha) &= -(2\pi i)^{-1} \int_{\Gamma(a+i\alpha)} w^m R(w, a+i\alpha)\, dw \\ &+ \sum_{j=1}^{r} -(2\pi i)^{-1} \int_{\gamma_j(a+i\alpha)} (w^m - [\lambda(a+i\alpha)\exp(i2\pi j/r)]^m) R(w, a+i\alpha)\, dw. \end{aligned} \tag{3.32}$$

The integrand of the j-th integral in the sum to the right in (3.32) is in view of (iv) analytic on the disc whose boundary is $\gamma_j(a+i\alpha)$ $j=1, \ldots, r$. The r rightmost integrals in (3.32) therefore equal zero. It follows from (iv) and the definition of Γ

that the distance between $\sigma(\Phi(a+i\alpha))$ (i.e. the poles of $w \to R(w, a+i\alpha)$) and $\Gamma(a+i\alpha)$ is at least $\frac{1}{2}\eta|\lambda(a+i\alpha)|>\eta^2/2$. The norm of the remaining integral in (3.32) is therefore dominated by

$$\mathrm{Const}(1-\eta/2)^m|\lambda(a+i\alpha)|^m \leqq \mathrm{Const}(1-\eta/2)^m \lambda(a)^m. \tag{3.33}$$

Hence $\|A_j - B_j\| \leqq \mathrm{Const}(1-\eta/2)^{n_j}\lambda(a)^{n_j}$. Also

$$\|B_j\| = |\lambda a+i\alpha)|^{n_j}\|E_{n_j}(a+i\alpha)\| \leqq \lambda(a)^{n_j}\sum_1^r \|E_j(a+i\alpha)\| \leqq \mathrm{Const}\,\lambda(a)^{n_j}$$

when $a\in K$ and $|\alpha|\leqq\delta$. But $\|A_j\| \leqq \|A_j - B_j\| + \|B_j\| \leqq \mathrm{Const}\,\lambda(a)^{n_j}$, and hence

$$\left\|\bigotimes_1^k A_j - \bigotimes_1^k B_j\right\| \leqq \mathrm{Const}\sum_{j=1}^k \lambda(a)^n(1-\eta/2)^{n_j}$$
$$\leqq \mathrm{Const}\,\lambda(a)^n(1-\eta/2)^{\varepsilon n} \quad \text{when } a\in K, \quad |\alpha|\leqq\delta.$$

A glance at (3.29) finally yields $I_3 \leqq \mathrm{Const}\, n^{p/2}(1-\eta/2)^{\varepsilon n} \leqq \mathrm{Const}/n$.

I_2. This integral tends to zero faster than any power of n. The verification is left to the reader.

I_1. We have

$$I_1 \leqq \int_{\substack{\alpha\in(-\pi,\pi]^p \\ |\alpha|\geqq\delta}} \lambda(a)^{-n}\prod_{j=1}^k \|\Phi(a+i\alpha)^{n_j}\|\,d\alpha. \tag{3.34}$$

It follows from (v) that the poles of $R(w, a+i\alpha)$ lies inside the circle $\Gamma(a)$ and that the distance between the poles and $\Gamma(a)$ is at least $\frac{1}{2}\eta\lambda(a)>\eta^2/2$ if $a\in K$, $|\alpha|\geqq\delta$ and $\alpha\in(-\pi,\pi]^p$. The representation

$$\Phi(a+i\alpha)^m = -(2\pi i)^{-1}\int_{\Gamma(a)} w^m R(w, a+i\alpha)\,dw \tag{3.35}$$

therefore yields $\|\Phi(a+i\alpha)^m\| \leqq \mathrm{Const}(1-\eta/2)^m\lambda(a)^m$ if $a\in K$, $|\alpha|>\delta$ and $\alpha\in(-\pi,\pi]^p$. Hence also $I_1 \leqq \mathrm{Const}(1-\eta/2)^n$.

4. Equivalence of Ensembles

In this section $q_{xy}=1$ on Y, and we will assume that Y is aperiodic ($r=1$). The canonical Markov chain is defined by the transition probabilities

$$p_{xy}(a) = \frac{e^{a\cdot t(x,y)}}{\lambda(a)}\frac{e_a(y)}{e_a(x)}, \quad (x,y)\in Y \tag{4.1}$$

and the stationary initial distribution

$$p_a(x) = (e_a^*\cdot e_a)^{-1} e_a^*(x)\, e_a(x), \quad x\in X. \tag{4.2}$$

We define

$$p_a(x_0, \dots, x_n) = p_a(x_0)\prod_{i=1}^n p_{x_{i-1}x_i}(a) \tag{4.3}$$

then

$$p_a(x_0, \dots, x_n) = (e_a^*\cdot e_a)^{-1} e_a^*(x_0)\frac{e^{a\cdot t_n}}{\lambda(a)^n} e_a(x_n), \quad (x_{i-1}, x_i)\in Y,\ i=1, \dots, n, \tag{4.4}$$

where $t_n = \sum_{i=1}^{n} t(x_{i-1}, x_i)$. We will write Π_a for the probability measure on X^Z determined by the densities (4.4).

The reason why we consider the canonical Markov chain will now be explained. Consider a long chain $x_{-M}, \ldots, x_0, \ldots, x_n, \ldots, x_N$. The uniform, or the microcanonical, distribution on the surface

$$\{x_{-M+1}, \ldots, x_{N-1} \mid \sum_{-M<i\leq N} t(x_{i-1}, x_i) = T,\ x_{-M} = x,\ x_N = y\} \tag{4.5}$$

is given by the density which equals $(\delta_x \cdot F^{(M+N)*}(T)\,\delta_y)^{-1}$ if $\sum_{-M<i\leq N} t(x_{i-1}, x_i) = T$, and which is zero otherwise. Here $\delta_x(y) = 1$ if $y = x$, $= 0$ otherwise. The distribution for $x_0, \ldots, x_n$ induced by the microcanonical distribution on the surface (4.6) is then given by the density

$$p_{Txy}^{M,N}(x_0, \ldots, x_n) = \frac{\delta_{xx_n} \cdot F^{M*} \circledast F^{(N-n)*}(T - t_n)\,\delta_{x_0 y}}{\delta_x \cdot F^{(M+N)*}(T)\,\delta_y}, \tag{4.6}$$

where $t_n = \sum_{i=1}^{n} t(x_{i-1}, x_i)$, and $\delta_{x_1 x_2}(y_1, y_2) = 1$ if $(y_1, y_2) = (x_1, x_2)$, $= 0$ otherwise.

Theorem 4.1 (Boltzmann's law). *For each compact $K \subset R^p$*

$$p_{Txy}^{M,N}(x_0, \ldots, x_n) = p_{\hat{a}}(x_0, \ldots, x_n)\left[1 + O(1/(M+N))\right], \tag{4.7}$$

uniformly when $\hat{a} = \hat{a}(T/(M+N)) \in K$, and $M/(M+N)$ remains bounded away from zero and one.

There are many results that are related to Theorem 4.1. One approach to the problem can be found in [11]. Thompson at Cornell university has obtained a related result which is applicable also when the interaction is more complicated than the ordinary Markov interaction considered here.

Proof. In this proof we will write $\hat{a}$ instead of $\hat{a}(T/(M+N))$, and $E(a)$ instead of $E_1(a)$. Since $r = 1$ we have $E_n(a) = E(a)$ for all n.

We apply Theorem 3.1 (with $k = 2$ and $a = \hat{a}$) to the operator in the numerator of (4.6). The result is

$$\begin{aligned} F^{M*} \circledast F^{(N-n)*}(T-t) = {} & (2\pi(M+N-n))^{-p/2} (\det V_{\hat{a}})^{-\frac{1}{2}} \\ & \cdot \frac{e^{-\hat{a}\cdot(T-t)}}{\lambda(\hat{a})^{M+N-n}} \{[e^{-|u|^2/2}(1 + (M+N-n)^{-\frac{1}{2}} P_{\hat{a}}(u))\, E(\hat{a}) \otimes E(\hat{a}) \\ & + u \cdot ((M+N-n)\, V_{\hat{a}})^{-\frac{1}{2}} \operatorname{grad} E(a) \otimes E(a)|_{a=\hat{a}}] \\ & + O(1/(M+N))\}, \end{aligned}$$

uniformly in T when $\hat{a} \in K$ and $M/(M+N)$ remains bounded away from zero and one. Here

$$\begin{aligned} u &= ((M+N-n)\, V_{\hat{a}})^{-\frac{1}{2}} (T - t - (M+N-n)\, m_{\hat{a}}) \\ &= -((M+N-n)\, V_{\hat{a}})^{-\frac{1}{2}} (t - n m_{\hat{a}}) = O((M+N)^{-\frac{1}{2}}). \end{aligned} \tag{4.9}$$

Therefore

$$\begin{aligned} e^{-|u|^2} &= 1+O((M+N)^{-1}) \\ (M+N-n)^{-\frac{1}{2}} P_{\hat{a}}(u) &= O((M+N)^{-1}) \\ u \cdot ((M+N-n) V_{\hat{a}})^{-\frac{1}{2}} \text{ grad } E(a) \otimes E(a)|_{a=\hat{a}} &= O((M+N)^{-1}). \end{aligned} \tag{4.10}$$

Hence the numerator of (4.6) equals

$$(2\pi(M+N))^{-p/2} (\det V_{\hat{a}})^{-\frac{1}{2}} \frac{e^{-\hat{a}\cdot(T-t)}}{\lambda(\hat{a})^{M+N-n}} [\delta_{xx_n} \cdot E(\hat{a}) \otimes E(\hat{a}) \delta_{x_0 y} + O((M+N)^{-1}], \tag{4.11}$$

uniformly when $\hat{a} \in K$ and $M/(M+N)$ remains bounded away from zero and one.

Corollary 3.2 applied (with $k=1$) ro the denominator of (4.6) yields

$$\begin{aligned} \delta_x \cdot F^{(M+N)*}(T) \delta_y &= (2\pi(M+N))^{-p/2} (\det V_{\hat{a}})^{-\frac{1}{2}} \frac{e^{-\hat{a}\cdot T}}{\lambda(\hat{a})^{M+N}} \\ &\cdot [\delta_x \cdot E(\hat{a}) \delta_y + O((M+N)^{-1})], \end{aligned} \tag{4.12}$$

uniformly when $\hat{a} \in K$ and $M/(M+N)$ remains bounded away from zero and one.

From (4.11) and (4.12) we finally obtain

$$p_{Txy}^{M,N}(x_0, \ldots, x_n) = \frac{e^{\hat{a}\cdot t}}{\lambda(\hat{a})^n} \frac{\delta_{xx_n} \cdot E(\hat{a}) \otimes E(\hat{a}) \delta_{x_0 y}}{\delta_x \cdot E(\hat{a}) \delta_y} + O(1/(M+N)), \tag{4.13}$$

which is the desired result.

Example. The one dimensional Ising Model.

Here $X = \{-1, 1\}$, $Y = X \times X$ and $t(x, y) = \begin{pmatrix} xy \\ \frac{x+y}{2} \end{pmatrix}$. Thus Y is irreducible and has period 1.

Calculations show that $\lambda(a) = e^b(\cosh c + \sigma(a))$ and

$$e_a(x) = e_a^*(x) = (1 + x(\sinh c)/\sigma(a))^{\frac{1}{2}},$$

where $a = \begin{pmatrix} b \\ c \end{pmatrix}$ and $\sigma(a) = (e^{-4b} + \sinh^2 c)^{\frac{1}{2}}$.

However, we have to do a modification to be able to apply Theorem 4.1. If we let $s = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\alpha = 2\pi \begin{pmatrix} \frac{1}{4} \\ \frac{1}{2} \end{pmatrix}$ and choose u such that $u(-1) - u(1) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, then $T_u \subset s + G_\alpha$ and hence our regularity condition is not satisfied. But if we define $\tilde{t}$ by

$$t(x, y) = s + u(x) - u(y) + A\tilde{t}(x, y) \tag{4.14}$$

where $A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$, then $\tilde{t}$ is a statistic which takes its values in Z^2 and it is straightforward to verify that $\tilde{t}$ satisfies the regularity condition.

Using the relations between the structure functions, eigenfunctions and eigenvalues corresponding to t and $\tilde{t}$ induced by (4.14) we conclude that Theorem 4.1 holds not only for $\tilde{t}$ but also for the original t.

m_a and V_a may be expressed in terms of the moments of the canonical Markov chain

$$\begin{aligned} m_a &= E_a t(x_0, x_1) \\ V_a &= \sum_{k=-\infty}^{\infty} E_a\big(t(x_0, x_1) - m_a\big)\big(t(x_k, x_{k+1}) - m_a\big)'. \end{aligned} \tag{4.15}$$

Here the prime denotes the transpose of a column vector and the multiplication is matrix multiplication. The series converges.

The first of these identities follows from

$$0 = \operatorname{grad} \sum_{\mathbb{X}_n} p_a(\mathbb{X}_n) = E_a \operatorname{grad} \log p_a(\mathbb{X}_n) \text{ where } \mathbb{X}_n = (x_0, \ldots, x_n).$$

To see that the second holds we note that

$$0 = \frac{\partial^2}{\partial a_i \partial a_j} \sum_{\mathbb{X}_n} p_a(\mathbb{X}_n) = E_a \frac{\partial^2}{\partial a_i \partial a_j} \log p_a(\mathbb{X}_n) + E_a \frac{\partial}{\partial a_i} \log p_a(\mathbb{X}_n) \frac{\partial}{\partial a_j} \log p_a(\mathbb{X}_n)$$

and that as $h - k \to \infty$.

$$\sum_{x_{k+1} \ldots x_{h-1}} p_a(x_0, \ldots, x_n) = p_a(x_0 \ldots x_k)\, p_a(x_h, \ldots, x_n) + O(\theta^{h-k})$$

for some $0 < \theta < 1$. The latter because $\lambda(a)^{-n} \Phi(a)^n = E(a) + O(\theta^n)$ for some $0 < \theta < 1$. Hence

$$V_a = \lim_{n \to \infty} n^{-1} E_a \sum_0^{n-1} s_j \left(\sum_0^{n-1} s_j\right)',$$

where $s_j = t(x_j, x_{j+1}) - m_a$. But $E_a s_i s_j = E_a s_0 s_{j-i} = O(\theta^{|j-i|})$ and hence

$$V_a = \lim \left(E_a s_0^2 + 2 \sum_{k=1}^{n} (1 - k/n) E_a s_0 s_k\right) = E_a s_0^2 + 2 \sum_{k=1}^{\infty} E_a s_0 s_k = \sum_{-\infty}^{+\infty} E_a s_0 s_k,$$

which was to be shown.

We will consider three kinds of entropy. The *microcanonical entropy*

$$H_{xy}^n(t) = \log f_{xy}^n(t), \tag{4.16}$$

the *canonical entropy*

$$H_a^n = -E_a \log p_a(x_0, \ldots, x_n), \tag{4.17}$$

and the *specific entropy*

$$H_a = \log \lambda(a) - a \cdot m_a. \tag{4.18}$$

They are related in the following way

$$\begin{aligned} H_a^n &= n H_a + H_a^0 \\ H_{xy}^n(t) &= n H_{\hat{a}(t/n)} + O(\log n), \end{aligned} \tag{4.19}$$

uniformly in t when $\hat{a}(t/n) \in K$.

The first of these identities is a consequence of (4.15) and the stationarity of the canonical Markov chain. The second follows from Corollary 3.2.

The canonical Markov chain $\Pi_{\hat{a}(m)}$ has the largest entropy among all strictly stationary processes $\{x_n\}_{n \in Z}$ satisfying $E t(x_0, x_1) = m$. See [17].

The definitions of the canonical Markov chain and the entropies make sense also when Y is periodic, and (4.19) still holds if we add "provided $r(y)=r(x)+n$ (mod r)" to the sentence just below (4.19). (2.14) can now be written

$$\Pi_a(x_0 \in X_k)=\frac{1}{r}\, k=1,\dots,r.$$

The same remark applies to the more general stationary Markov chain considered in the next section. In this case we lose, however, the interpretation of the micro-canonical entropy as the logaritm of the number of microstates wich realize a given macrostate t.

5. Central Limit Theorems

In this section we consider a general non-negative and irreducible sequence $\{q_{xy}\}$, and let Π_a stand for the stationary probability measure on X^Z determined by the densities

$$\begin{aligned} p_a(x_0,\dots,x_n)=(e_a^*\cdot e_a)^{-1}\, e_a^*(x_0)\left(\prod_1^n \frac{e^{a\cdot t(x_{i-1},x_i)}}{\lambda(a)}\, q_{x_{i-1}x_i}\right) e_a(x_n),\\ n\geqq 0, \quad (x_{i-1},x_i)\in Y \quad \text{for } i=1,\dots,n. \end{aligned} \tag{5.1}$$

We will investigate the asymptotic behavior of the distribution of

$$t_n=\sum_1^n t(x_{i-1},x_i)$$

induced by Π_a. When we do not want to stress upon uniformity in a, we will formulate our results for Π_0. In this case we suppose that $\{q_{xy}\}$ is normed in such a way that $\lambda(0)=1$ (this is always the case in case (*b*) below), and we will omit the zero, thus $\Pi=\Pi_0$, $e=e_0$, $m=m_0$ and so on.

We point out two cases

(a) The canonical Markov chain: $q_{xy}=1$ on Y.

(b) $\{q_{xy}\}$ is transition probabilities, $\sum_y q_{xy}=1$ for all $x\in X$. Then Π_0 is the stationary measure on X^Z induced by the Markov chain with transition probabilities $\{q_{xy}\}$. The modifications which are necessary when the initial distribution is not the stationary one will be left to the reader.

The results in this section are closely related to the results of Saulis and Statuljavičus [16] and Statuljavičus [18].

Let $p_a^n(t)$ stand for the density of t_n

$$p_a^n(t)=\Pi_a(t_n=t). \tag{5.2}$$

Then

$$p_a^n(t)=\frac{e^{a\cdot t}}{\lambda(a)^n}\,\frac{e_a^*\cdot F^{n*}(t)\, e_a}{e_a^*\cdot e_a}. \tag{5.3}$$

Theorem 3.1 and its two corollaries therefore induce three results for $p_a^n(t)$. If we note that $e_a^*\cdot E_n(a)\, e_n=e_a^*\cdot e_a$ and (recall (2.14)) $e_a^*\cdot(\operatorname{grad} E_n(a))\, e_a=0$, we see that these results may be stated in the following way.

Theorem 5.1. *For each compact $K \subset R^p$*

$$(2\pi n)^{p/2} (\det V_a)^{\frac{1}{2}} p_a^n(t) = e^{-\frac{1}{2}(t-nm_a)\cdot V_a^{-1}(t-nm_a)/n}\left[1+n^{-\frac{1}{2}} P_a\left((nV_a)^{-\frac{1}{2}}(t-nm_a)\right)\right]+O(1/n), \tag{5.4}$$

uniformly in t and in a when $a \in K$.

Theorem 5.2. *(Local central limit theorem.) For each compact $K \subset R^p$*

$$(2\pi n)^{p/2} (\det V_a)^{\frac{1}{2}} p_a^n(t) = e^{-\frac{1}{2}(t-nm_a)\cdot V_a^{-1}(t-nm_a)/n} + O(n^{-\frac{1}{2}}), \tag{5.5}$$

uniformly in t and in a when $a \in K$.

Theorem 5.3. *(Local central limit theorem for large deviations.) For each compact $K \subset R^p$*

$$(2\pi n)^{p/2} (\det V_{\hat{a}(t/n)})^{\frac{1}{2}} p_a^n(t) = e^{-(\hat{a}(t/n)-a)\cdot t}\left[\frac{\lambda(\hat{a}(t/n))}{\lambda(a)}\right]^n \frac{e_a^* \cdot E_n(\hat{a}(t/n))\, e_a}{e_a^* \cdot e_a}(1+O(1/n)) \tag{5.6}$$

uniformly in a and t when $a \in K$ and $\hat{a}(t/n) \in K$.

Theorem 5.2 implies convergence in distribution. For results of that kind we refer to the already cited paper of Statuljavičus and to part I §16 of [3] and the references given in the notes at the end of that paragraph. Another local limit theorem was given by Kolmogorov [10].

If we specialize (5.6) to the independence case and put $a=0$ we obtain

$$p^{n*}(t) = (2\pi n)^{-p/2} (\det V_{\hat{a}(t/n)})^{-\frac{1}{2}} e^{nH_{\hat{a}(t/n)}}(1+O(1/n)), \tag{5.7}$$

uniformly when $\hat{a}(t/n) \in K$. Compare [15].

Theorem 5.4 and 5.5 below supplements Theorem 5.3 by providing information concerning the tails. We use the word tail for halfplanes of form $c \cdot t_n > c \cdot t$, where $c \in Z^p$ and $c \cdot t > c \cdot m_a$. Since $c \cdot t_n$ is a one-dimensional statistic it suffices to consider the case $p=1$. The proofs of our statements concerning the tails will be given at the end of this section.

Theorem 5.4. *(Central limit theorem for large deviations.) For each compact $K \subset R$*

$$\Pi_a(t_n \geqq t) = \left[\frac{\lambda(\hat{a}(t/n))}{\lambda(a)}\right]^n \frac{e^{-(\hat{a}(t/n)-a)t}}{1-e^{-(\hat{a}(t/n)-a)}} (2\pi n V_{\hat{a}(t/n)})^{-\frac{1}{2}} \cdot \frac{e_a^* \cdot E_n(\hat{a}(t/n))\, e_a}{e_a^* \cdot e_a}\left[1+O\left(n^{-1}\left(\frac{t}{n}-m_a\right)^{-2}\right)\right] \tag{5.8}$$

uniformly in a and t when $K \ni a < \hat{a}(t/n) \in K$.

If we put $a=0$ in the theorem above and restrict ourselves to values of t for which $0 < t/n - m = o(1)$, we obtain a corollary which should be compared to Theorem 2 p. 520 of [5].

Corollary 5.1. *Let $x_t = (nV)^{-\frac{1}{2}}(t-nm)$. If $o(n^{\frac{1}{2}}) = x_t \to \infty$, then*

$$\Pi(t_n \geqq t) = (1-\mathfrak{N}(x_t))\, e^{x_t^2 l(x_t/n^{-\frac{1}{2}})}(1+O(x_t n^{-\frac{1}{2}})+O(x_t^{-2})). \tag{5.9}$$

Here $\mathfrak{N}(x)$ stands for the normal distribution with zero expectation and unit variance, and

$$l(x) = -\sum_{j=1}^{\infty} \frac{d^j}{dx^j} V_{\hat{a}(x)}^{-1}\bigg|_{x=m} \frac{V^{(j+2)/2}}{(j+2)!} x^j. \tag{5.10}$$

Theorem 5.5 presents an upper bound for the tails.

Theorem 5.5. *If $t \geqq n\, m_a$, then*

$$\Pi_a(t_n \geqq t) \leqq e^{-(\hat{a}(t/n)-a)t} \frac{e_a^* \cdot \Phi(\hat{a}(t/n))^n e_a}{\lambda(a)^n e_a^* \cdot e_a}. \tag{5.11}$$

In the independence case Theorem 5.5 takes the form

Theorem 5.5 a. *If $t \geqq n\, m_a$, then*

$$\Pi_a\left(\sum_1^n t(x_i) \geqq t\right) \leqq e^{-(\hat{a}(t/n)-a)t} \frac{\lambda(\hat{a}(t/n))^n}{\lambda(a)^n}. \tag{5.12}$$

Note that by Taylor's formula

$$\log \lambda(a) = \log \lambda(\hat{a}) + (a-\hat{a})\, m_{\hat{a}} + (a-\hat{a})^2 \int_0^1 (1-\xi)\, V_{\hat{a}+\xi(a-\hat{a})}\, d\xi.$$

The expression to the right in (5.12) may therefore equally well be written

$$e^{-n\delta(a, \hat{a}(t/n))}, \tag{5.13}$$

where

$$\delta(a, \hat{a}) = (a-\hat{a})^2 \int_0^1 (1-\xi)\, V_{\hat{a}+\xi(a-\hat{a})}\, d\xi. \tag{5.14}$$

Since V_a is strictly positive for each a and since $\hat{a}(t/n) = a$ only if $t/n = m_a$ we conclude that $\delta(a, \hat{a}(t/n))$ is positive and bounded away from zero when t/n is bounded away from m_a and $a \in K$, $\hat{a}(t/n) \in K$. Thus, when applicable, Theorem 5.5 a gives a sharpening of Čebyšev's inequality.

We also have the following corollary to Theorem 5.5 a

Corollary 5.2. *If $t \geqq n\, m_a$, then*

$$\Pi_a\left(\sum_1^n t(x_i) \geqq t\right) \leqq \exp\left(-\frac{n\left(\frac{t}{n} - m_a\right)^2}{2M}\right), \tag{5.15}$$

where $M = \max_{a \leqq b \leqq \hat{a}(t/n)} V_b$.

Example. Bernoulli Trials.

$x_1, \ldots, x_n$ are independent and takes the values 0 and 1 with probabilities $1-\theta$ and θ respectively. Here $t_n = x_1 + \cdots + x_n$ and $p_a(x_0, \ldots, x_n) = \theta^{t_n}(1-\theta)^{n-t_n} = e^{a t_n}(1+e^a)^{-n}$, where $\theta = e^a/(1+e^a)$. We also have $m_a = \theta$ and $V_a = \theta(1-\theta)$. It is easily verified that $V_b \leqq 1/4$ for all b, and that $V_b \leqq V_a$ for all $b \geqq a \geqq 0$. Corollary 5.2 therefore yields the inequalities

$$\sum_{s \geqq t} \binom{n}{s} \theta^s (1-\theta)^{n-s} \leqq \exp\left(-\frac{n\left(\frac{t}{n} - \theta\right)^2}{2\theta(1-\theta)}\right) \quad \text{if } \frac{t}{n} \geqq \theta \geqq \tfrac{1}{2} \tag{5.16}$$

and
$$\sum_{s\geq t}\binom{n}{s}\theta^s(1-\theta)^{n-s}\leqq\exp\left(-2n\left(\frac{t}{n}-\theta\right)^2\right)\quad\text{if }\frac{t}{n}\geqq\theta\geqq 0. \tag{5.17}$$

An inequality similar to those above was given by Bernstein. See [14], p. 387.

Proof of Theorem 5.4. In this proof we will use the following abbreviations: $\hat{a}=\hat{a}\left(\frac{t}{n}\right)$, $\theta=\exp(a-\hat{a})$ and
$$R_a^{n*}(t)=e^{at}\lambda(a)^{-n}F^{n*}(t). \tag{5.18}$$
We have
$$\Pi_a(t_n\geqq t)=\sum_{s=t}^{\infty}p_a^n(s)=(e_a^*\cdot e_a)^{-1}e_a^*\cdot\left(\sum_{s=t}^{\infty}R_a^{n*}(s)\right)e_a. \tag{5.19}$$
However
$$R_a^{n*}(s)=\theta^s\lambda(\hat{a})^n\lambda(a)^{-n}R_{\hat{a}}^{n*}(s) \tag{5.20}$$
and hence
$$\Pi_a(t_n\geqq t)=\lambda(\hat{a})^n\lambda(a)^{-n}\theta^t(1-\theta)^{-1}(e_a^*\cdot e_a)^{-1}e_a^*\cdot\mathscr{S}e_a, \tag{5.21}$$
where
$$\mathscr{S}=(1-\theta)\sum_{n=0}^{\infty}\theta^u R_{\hat{a}}^{n*}(t+u)=R_{\hat{a}}^{n*}(t)+\sum_{u=1}^{\infty}\theta^u\big(R_{\hat{a}}^{n*}(t+u)-R_{\hat{a}}^{n*}(t+u-1)\big). \tag{5.22}$$
Theorem 3.1 yields
$$\begin{aligned}R_{\hat{a}}^{n*}(t+u)=(2\pi nV_{\hat{a}})^{-\frac{1}{2}}\big[&g_n\big(u(nV_{\hat{a}})^{-\frac{1}{2}}\big)E_n(\hat{a})\\&+h_n\big(u(nV_{\hat{a}})^{-\frac{1}{2}}\big)\frac{d}{d\hat{a}}E_n(\hat{a})+O(1/n)\big],\end{aligned} \tag{5.23}$$
uniformly in u and t when $\hat{a}(t/n)\in K$.

Here
$$\begin{aligned}g_n(x)&=e^{-x^2/2}\big(1+n^{-\frac{1}{2}}P_{\hat{a}}(x)\big),\\h_n(x)&=e^{-x^2/2}x(nV_{\hat{a}})^{-\frac{1}{2}}.\end{aligned} \tag{5.24}$$
Therefore
$$\begin{aligned}\mathscr{S}=(2\pi nV_{\hat{a}})^{-\frac{1}{2}}&\left[E_n(\hat{a})+O\left(\sum_{u=1}^{\infty}\theta^u\delta_1(u)\,\|E_n(\hat{a})\|\right)\right.\\&\left.+O\left(\sum_{u=1}^{\infty}\theta^u\delta_2(u)\left\|\frac{d}{d\hat{a}}E_n(\hat{a})\right\|\right)+O\big(n^{-1}(1-\theta)^{-1}\big)\right],\end{aligned} \tag{5.25}$$
uniformly in t when $\hat{a}(t/n)\in K$. Here
$$\begin{aligned}\delta_1(u)&=\big|g_n\big(u(nV_{\hat{a}})^{-\frac{1}{2}}\big)-g_n\big((u-1)(nV_{\hat{a}})^{-\frac{1}{2}}\big)\big|\\\delta_2(u)&=\big|h_n\big(u(nV_{\hat{a}})^{-\frac{1}{2}}\big)-h_n\big((u-1)(nV_{\hat{a}})^{-\frac{1}{2}}\big)\big|.\end{aligned} \tag{5.26}$$

Let $I_n(u)=[(u-1)(nV_{\hat{a}})^{-\frac{1}{2}},u(nV_{\hat{a}})^{-\frac{1}{2}}]$. Then
$$\delta_1(u)\leqq(nV_{\hat{a}})^{-\frac{1}{2}}\max_{x\in I_n(u)}|g_n'(x)|. \tag{5.27}$$
But
$$\begin{aligned}|g_n'(x)|=|-x\,g_n(x)+e^{-x^2/2}P_{\hat{a}}'(x)\,n^{-\frac{1}{2}}|&\leqq u(nV_{\hat{a}})^{-\frac{1}{2}}|g_n(x)|\\+n^{-\frac{1}{2}}e^{-x^2/2}|P_{\hat{a}}'(x)|&\leqq\text{Const}\,(u+1)\,n^{-\frac{1}{2}}\end{aligned} \tag{5.28}$$
when $x\in I_n(u)$. Hence $\delta_1(u)=O\big((u+1)/n\big)$, uniformly in u and t when $\hat{a}(t/n)\in K$.

In a similar way we obtain $\delta_2(u)=O(1/n)$, uniformly in u and t when $\hat{a}(t/n)\in K$. The facts $\sum_1^\infty \theta^u=\theta(1-\theta)^{-1}$, $\sum_1^\infty u\,\theta^u=\theta(1-\theta)^{-2}$, and $n^{-1}(1-\theta)^{-1}\leqq n^{-1}(1-\theta)^{-2}$ applied to (5.25) now yields

$$\mathcal{S}=(2\pi n V_{\hat{a}})^{-\frac{1}{2}}\left[E_n(\hat{a})+O\left(n^{-1}(1-\theta)^{-2}\right)\right], \tag{5.29}$$

uniformly in t when $\hat{a}(t/n)\in K$. The desired result follows if we substitute the expression to the right in (5.29) for $\mathcal{S}$ in (5.21).

Proof of Corollary 5.1. Let $\tilde{a}(x)=\hat{a}(m+x V^{\frac{1}{2}})$ and define $l(x)$ and $h_n(x)$ by

$$\begin{aligned} x^2\, l(x)-x^2/2 &= H_{\tilde{a}(x)} \\ h_n(x) &= x(1-e^{-\tilde{a}(x)})^{-1}\, V_{\tilde{a}(x)}\frac{e^*\cdot E_n(\tilde{a}(x))\,e}{e^*\cdot e}. \end{aligned} \tag{5.30}$$

It follows from Theorem 5.4 that

$$\Pi(t_n\geqq t)=(2\pi)^{-\frac{1}{2}}x_t^{-1}\,e^{-x_t^2/2}\,e^{x_t^2 l(x_t n^{-\frac{1}{2}})}\,h_n(x_t\, n^{-\frac{1}{2}})\left(1+O(x_t^{-2})\right), \tag{5.31}$$

uniformly in t when $0<\hat{a}(t/n)\in K$. Calculations show that $h_n(x)=1+O(x)$ uniformly in n. Also $\frac{d}{dx}H_{\hat{a}(x)}=-\hat{a}(x)$, $\frac{d^2}{dx^2}H_{\hat{a}(x)}=-V_{\hat{a}(x)}^{-1}$ and hence $l(x)$ is of the form (5.10). The estimate $(2\pi)^{-\frac{1}{2}}x^{-1}e^{-x^2/2}=\left(1-\mathfrak{N}(x)\right)\left(1+O(x^{-2})\right)$ (see [4], p. 166) finally yields the desired result. (Note that $o(n^{\frac{1}{2}})=x_t\to\infty$ implies $0<\hat{a}(t/n)\in K$.)

Proof of Theorem 5.4. For any $c\geqq 0$.

$$\begin{aligned} \Pi_a(t_n\geqq t) &\leqq \sum_{t_n\geqq t} e^{c(t_n-t)}\,p_a(x_0,\ldots,x_n) \\ &\leqq \sum e^{c(t_n-t)}\,p_a(x_0,\ldots,x_n)=e^{-ct}\,\lambda(a)^{-n}(e_a^*\cdot e_a)^{-1}\,e_a^*\cdot\Phi(a+c)^n\,e_a. \end{aligned} \tag{5.32}$$

The theorem follows if we let $c=\hat{a}(t/n)-a$. (Note that $t\geqq n m_a$ implies $\hat{a}(t/n)\geqq a$.) The reason for this choise is that for large n the expression to the right in (5.32) approximately equals

$$\left(e^{-ct/n}\frac{\lambda(a+c)}{\lambda(a)}\right)^n\frac{e_a^*\cdot E_n(a+c)\,e_a}{e_a^*\cdot e_a} \tag{5.33}$$

and that $c=\hat{a}(t/n)-a$ minimizes $e^{-ct/n}\,\lambda(a+c)/\lambda(a)$.

Proof of Corollary 5.3. The left hand side of (5.15) is by Theorem 5.5a dominated by (5.13). We make the substitution $\eta=m_{\hat{a}+\xi(a-\hat{a})}$ in the expression to the right in (5.14). The result is

$$\delta(a,\hat{a})=\int_{m_a}^{t/n}(\hat{a}(\eta)-a)\,d\eta. \tag{5.34}$$

But $\frac{d}{d\eta}\hat{a}(\eta)=1/V_{\hat{a}(\eta)}$ and hence

$$\hat{a}(\eta)-a\geqq\frac{\eta-m_a}{M}\qquad\text{if } m_a\leqq\eta\leqq t/n. \tag{5.35}$$

So that

$$\delta(a,\hat{a}) \geqq \int_{m_a}^{t/n} \frac{\eta - m_a}{M}\, d\eta = \frac{\left(\frac{t}{n} - m_a\right)^2}{2M}, \tag{5.36}$$

which proves the corollary.

6. Statistical Inference

We will in this section apply our results to the statistical theory presented in [12] and [13]. Another treatment of the subject treated here can be found in Billingsley [1].

Imagine an experiment the outcome of which may be listed $x_0, \ldots, x_n$. Suppose we feel that, given the endpoints x_0 and x_n, a statistic $t_n = \sum_1^n t(x_{i-1}, x_i)$ is sufficient for the experiment. Given x_0, x_n and $t_n = t$ we may consider $x_1, \ldots, x_{n-1}$ as being uniformly distributed on the surface

$$X^n_{x_0 x_n}(t) = \left\{ x_1, \ldots, x_{n-1} \,\middle|\, \sum_1^n t(x_{i-1}, x_i) = t,\ (x_{i-1}, x_i) \in Y,\ i = 1, \ldots, n \right\} \tag{6.1}$$

(the microcanonical point of view). Alternatively, using Boltzmann's law, we may think of $x_0, \ldots, x_n$ as being distributed according to the canonical Markov chain which the function $t(x, y)$ determines (the macrocanonical point of view).

We will here deal with the problems of testing statistical hypotheses and estimating parameters. We will begin with the former.

Let the statistic u be the image of t under a linear transformation, and suppose that u is of lower dimension than t. After a suitable coordinate transformation we may then write

$$t = \begin{pmatrix} u \\ v \end{pmatrix},$$

where $\dim u = q < p = \dim t$. *We want to test the hypothesis: The statistic may be replaced by the simpler statistic u.*

Let Π^n_{uxy} stand for the uniform (the microcanonical) distribution on the surface $X^n_{xy}(u)$, and let $p^n_{ux_0x_n}(v)$ denote the density of $v_n = \sum_1^n v(x_{i-1}, x_i)$ induced by $\Pi^n_{ux_0x_n}$. Then

$$p^n_{uxy}(v) = \frac{f^n_{xy}(u, v)}{f^n_{xy}(u)}. \tag{6.2}$$

Here and in the sequel the structure function is defined relative to the sequence $\{q_{xy}\}$ which equals one on Y.

The exact test of the hypothesis rejects at risk level ε if

$$p^n_{ux_0x_n}(v) < \lambda^n_\varepsilon(u; x_0, x_n), \tag{6.3}$$

where $\lambda^n_\varepsilon(u; x, y)$ is the largest number for which the probability to reject is at most ε,

$$\lambda^n_\varepsilon(u; x, y) = \max\{\lambda \mid \Pi^n_{uxy}(p^n_{uxy}(v) < \lambda) \leqq \varepsilon\}. \tag{6.4}$$

Introduce the projection matrices P and Q defined by

$$\underset{p\times p}{I}=\begin{pmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{pmatrix}=\begin{pmatrix}P\\ \cdots\\ Q\end{pmatrix}\begin{matrix}\}q\\ \\ \}p-q\end{matrix} \tag{6.5}$$

and let $\hat{b}_0(x)\in R^q$ be the solution of the equation

$$P\,m_{\hat{b}_0,O}=x \qquad (x\in R^q). \tag{6.6}$$

In order to get an asymptotic test we prove

Theorem 6.1. *For each compact $K\subset R^q$*

$$\begin{aligned}p^n_{uxy}(v)&=(2\pi n)^{-(p-q)/2}(\det Q V^{-1}_{\hat{b}_0,O}Q)^{\frac{1}{2}}\\ &\cdot\left(e^{-\frac{1}{2}(v-nQm_{\hat{b}_0,O})\cdot QV^{-1}_{\hat{b}_0,O}Q(v-nQm_{\hat{b}_0,O})/n}+O(n^{-\frac{1}{2}})\right)\end{aligned} \tag{6.7}$$

uniformly in v, and in u when $\hat{b}_0=\hat{b}_0(u/u)\in K$, provided $r(y)=r(x)+n$ (mod r).

Proof. We apply Corollary 3.2 (with $k=1$) to the structure function of u. The result is

$$\left(\text{note that } \left(\frac{\partial^2\log\lambda(b,O)}{\partial b_i\partial b_j}\right)=PV_{b,O}P'\right).$$

$$\begin{aligned}f^n_{xy}(u)&=\delta_x\cdot F^{n*}(u)\,\delta_y=(2\pi n)^{-q/2}(\det PV_{\hat{b}_0,O}P')^{-\frac{1}{2}}\\ &\cdot\lambda(\hat{b}_0,O)^n e^{-\hat{b}_0\cdot u}[\delta_x\cdot E_n(\hat{b}_0,O)\,\delta_y+O(1/n)],\end{aligned} \tag{6.8}$$

uniformly when $\hat{b}_0=\hat{b}_0(u/n)\in K$. If we put $a=(\hat{b}_0,O)$ in Corollary 3.1 we obtain

$$\begin{aligned}f^n_{xy}(u,v)&=(2\pi n)^{-p/2}(\det V_{\hat{b}_0,O})^{-\frac{1}{2}}\lambda(\hat{b}_0,O)^n e^{-\hat{b}_0\cdot u}\\ &\cdot\left[e^{-\frac{1}{2}(t-nm_{\hat{b}_0,O})\cdot V^{-1}_{\hat{b}_0,O}(t-nm_{\hat{b}_0,O})/n}\,\delta_x\cdot E_n(\hat{b}_0,O)\,\delta_y+O(n^{-\frac{1}{2}})\right],\end{aligned} \tag{6.9}$$

uniformly in $t=\begin{pmatrix}u\\ v\end{pmatrix}$ when $(\hat{b}_0(u/n),O)\in K\times\{O\}^{p-q}$ i.e. when $\hat{b}_0(u/n)\in K$. If we note that

$$\begin{aligned}&n^{-1}(t-nm_{\hat{b}_0,O})\cdot V^{-1}_{\hat{b}_0,O}(t-nm_{\hat{b}_0,O})\\ &\qquad=n^{-1}(v-nQm_{\hat{b}_0,O})\cdot QV^{-1}_{\hat{b}_0,O}Q'(v-nQm_{\hat{b}_0,O}),\end{aligned} \tag{6.10}$$

then we see that it remains only to verify the identity $\det PV_{b,O}P'=\det QV^{-1}_{b,O}Q'$ $\det V_{b,O}$. The latter is done in e.g. [10], p. 55. This completes the proof of Theorem 6.1.

We will reject our hypothesis when $p^n_{ux_0x_n}(v)$ is too small. If we accept to approximate $p^n_{ux_0x_n}(v)$ with the normal density to the right in (6.7), this will occur when the quadratic form (6.10) is too large. It follows from Theorem 6.1 (by approximating the resulting Riemann sum with an integral) that the distribution of v induced by Π^n_{uxy} is asymptotically normal with expectation $nQm_{\hat{b}_0(u/n),O}$ and covariance matrix $n(QV^{-1}_{\hat{b}_0(u/n),O}Q')^{-1}$ and hence, in particular, that the quadratic form (6.10) is asymptotically chi-square distributed with $p-q$ degrees of freedom. Therefore:

The asymptotic test of the hypothesis rejects at risk level ε if

$$n^{-1}(v-nQm_{\hat{b}_0(u/n),O})\cdot QV^{-1}_{\hat{b}_0(u/n),O}\,Q'(v-nQm_{\hat{b}_0(u/n),O})>K^{-1}_{p-q}(1-\varepsilon), \tag{6.11}$$

where $K^{-1}_{p-q}(1-\varepsilon)$ is the $(1-\varepsilon)$-percentile of a chi-square distribution with $p-q$ degrees of freedom.

It is frequently convenient to calculate the quadratic form to the left in (6.11) via the modified likelihood quotient

$$\Lambda=\frac{\max\limits_{b\in R^q}\prod\limits_{i=1}^{n} p_{x_{i-1}x_i}(b,O)}{\max\limits_{a\in R^p}\prod\limits_{i=1}^{n} p_{x_{i-1}x_i}(a)} \tag{6.12}$$

where $p_{x_{i-1}x_i}(a)$ are the transition probabilities defined in (4.1). If $\prod_1^n p_{x_{i-1}x_i}(a)$ and $\prod_1^n p_{x_{i-1}x_i}(b,O)$ have maximum for $\tilde{a}$ and $\tilde{b}$ respectively, then $\tilde{a}-\hat{a}(t/n)=O(1/n)$ and $\tilde{b}-\hat{b}_0(u/n)=O(1/n)$. Hence

$$\log\Lambda=n(H_{\hat{a}}-H_{\hat{b}_0,O})+\log\frac{e_{\hat{b}_0,O}(x_n)\,e_{\hat{a}}(x_0)}{e_{\hat{b}_0,O}(x_0)\,e_{\hat{a}}(x_n)}+O(1/n). \tag{6.13}$$

Taylor expansion (in powers of $t/n-m_{\hat{b}_0,O}$) of the function to the right in (6.13) shows that the difference between $-2\log\Lambda$ and the quadratic form to the left in (6.11) equals $O(|v-nQm_{b_0,O}|/n)$, and hence they are asymptotically equivalent.

We now turn to the problem of parameter estimation. $x_0,\dots,x_n$ are considered as being distributed according to the canonical Markov chain and we want to estimate the unknown value of the parameter a.

The statistic $t_n=\sum_1^n t(x_{i-1},x_i)$ is, given the endpoints x_0 and x_n, sufficient for a, and the conditional distribution of t_n has the density

$$p^n_a(t\,|\,x_0,x_n)=\frac{e^{a\cdot t}}{\varphi^n_{x_0x_n}(a)}\,f^n_{x_0x_n}(t), \tag{6.14}$$

where $\varphi^n_{xy}(a)=\delta_x\cdot\Phi(a)^n\,\delta_y$.

The exact confidence region for a at risk level ε, $A^n_\varepsilon(t\,|\,x_0,x_n)$, is given by

$$A^n_\varepsilon(t\,|\,x_0,x_n)=\{a\in R^p\,|\,p^n_a(t\,|\,x_0,x_n)>\lambda^n_\varepsilon(a\,|\,x_0,x_n)\}, \tag{6.15}$$

where

$$\lambda^n_\varepsilon(a\,|\,x_0,x_n)=\max\{\lambda\,|\,\Pi^n_a(p^n_a(t\,|\,x_0,x_n)<\lambda\,|\,x_0,x_n)\leqq\varepsilon\}. \tag{6.16}$$

In order to obtain an asymptotic confidence region we prove

Theorem 6.2. *For each compact $K\subset R^p$*

$$p^n_a(t\,|\,x,y)=(2\pi n)^{-p/2}(\det V_a)^{-\frac{1}{2}}[e^{-\frac{1}{2}(t-nm_a)\cdot V_a^{-1}(t-nm_a)/n}+O(n^{-\frac{1}{2}})], \tag{6.17}$$

uniformly in t, and in a as $a\in K$, provided $r(y)=r(x)+n$ (mod r).

Proof. It follows from the argument centering around (3.33) that for each compact $K\subset R^p$ there is a $0<\theta<1$ such that

$$\varphi^n_{xy}(a)=\lambda(a)^n(\delta_x\cdot E_n(a)\,\delta_y+O(\theta^n)), \tag{6.18}$$

uniformly when $a \in K$, provided $r(y) = r(x) + n \pmod r$. The remainder of the proof is just another application of Theorem 3.1, and is left to the reader.

It follows from Theorem 6.2 that the inequality $p_a^n(t|x,y) > \lambda_\varepsilon^n(a|x,y)$ is approximately equivalent to that the quadratic form

$$n^{-1}(t - n\,m_a) \cdot V_a^{-1}(t - n\,m_a) \tag{6.19}$$

is smaller than a certain constant. Since this quadratic form is asymptotically chi-square distributed with p degrees of freedom, this constant approximately equals $K_p^{-1}(1-\varepsilon)$. Therefore

$$A_\varepsilon^n(t|x,y) \approx \{a \in R^p \,|\, n^{-1}(t - n\,m_a) \cdot V_a^{-1}(t - n\,m_a) < K_p^{-1}(1-\varepsilon)\}. \tag{6.20}$$

A further simplification is however possible. Calculations show that (6.19) equals

$$n\big(a - \hat a(t/n)\big) \cdot V_{\hat a(t/n)}\big(a - \hat a(t/n)\big) + O(|a - \hat a(t/n)|^3). \tag{6.21}$$

The asymptotic confidence region for a at risk level ε is given by the ellipsoid

$$\big\{a \in R^p \,|\, n\big(a - \hat a(t/n)\big) \cdot V_{\hat a(t/n)}\big(a - \hat a(t|n)\big) < K_p^{-1}(1-\varepsilon)\big\}. \tag{6.22}$$

The most precise statement we are in any right to make about the unknown value of a is to say that it belongs to *the set of exact estimates*

$$\check A_n(t|x_0, x_n) = \bigcap_{\varepsilon<1} A_\varepsilon^n(t|x_0, x_n). \tag{6.23}$$

It is straightforward to verify that

$$\check A_n(t|x,y) = \big\{a \in R^p \,\big|\, p_a^n(t|x,y) = \max_s p_a^n(s|x,y)\big\}, \tag{6.24}$$

and hence (recall that $H_{xy}^n(t) = \log f_{xy}^n(t)$)

$$\check A_n(t|x,y) = \{a \in R^p \,|\, H_{xy}^n(s) - H_{xy}^n(t) \leqq -a \cdot (s-t) \text{ for all } s \in Z^p\}, \tag{6.25}$$

provided $f_{xy}^n(t) > 0$. It is an immediate consequence of the latter representation that $\check A_n(t|x,y)$ is always a closed, convex set.

Theorem 6.3. *For each compact $K \subset R^p$ there is an integer $n(K)$ such that for any t, x and y: if $n \geqq n(K)$ and $\hat a(t/n) \in K$, then $\check A_n(t|x,y)$ is a non-empty, compact and convex set satisfying*

$$\max_{a \in \check A_n(t|x,y)} |a - \hat a(t/n)| = O(1/n) \tag{6.26}$$

uniformly in t when $\hat a(t/n) \in K$.

The counterpart to this result in the independence case was given in [6]. The proof of the present result is omited because the detailed proof is long and almost identical to the one in the independence case, provided we are acquainted with Sections 2 and 3 of the present paper.

In view of Theorem 6.3 it is natural to call $\hat a(t/n)$ *the asymptotic estimate.* Comparing (6.25) with the identity

$$\operatorname*{grad}_x H_{\hat a(x)} = -\hat a(x) \tag{6.27}$$

we see that the exact and the asymptotic estimate have the common interpretation: minus the local change of entropy.

Theorem 6.4. *The distribution of $\hat{a}(t_n/n)$ induced by Π_a is asymptotically normal with expectation a and covariance matrix $(n V_a)^{-1}$.*

Proof. A Taylor expansion yields $\hat{a}(t/n)=a+V_a^{-1}(t/n-m_a)+O(|t/n-m_a|^2)$ uniformly in t when $\hat{a}(t/n)\in K$. Hence $(n V_a)^{\frac{1}{2}}(\hat{a}(t/n)-a)=(n V_a)^{-\frac{1}{2}}(t-n m_a)+O(n^{-\frac{3}{2}}|(t-n m_a)|^2)$, uniformly in t when $\hat{a}(t/n)\in K$. It follows from Theorem 5.2 that the second term to the right tends to zero in probability and that the first term to the right is asymptotically normal with zero expectation and covariance matrix I. This completes the proof of Theorem 6.4.

We conclude this section with an example which I received from Per Martin-Löf. For further examples we refer to [1].

Example. Bird Navigation. A bird is caught in a circular cage in which there are s perches in s directions, the distances between adjacent perches being equal. We observe the successive positions $x_0, \dots, x_n$ of the bird, and we want to find out whether the bird navigates or not.

The sample space X may be identified with the group $\{0, \dots, s-1\}$ with addition modulo s. The frequencies n_{ij}, $i\in X, j\in X$ are sufficient for the experiment, and the hypothesis "no navigation" corresponds to a reduction in sufficiency from the frequencies to n^j, $j\in X$, where $n^j=\sum_{k-h=j} n_{hk}$. However, $\delta_i(x_0)+\sum_{j\in X} n_{ji}=\delta_i(x_n)+\sum_{j\in X} n_{ij}$ $i=0, \dots, s-1$, and $\sum_{j\in X} n^j=n$, and hence the proper dimension of t is $p=s^2-s$, and that of u is $q=s-1$. To calculate the likelihoodquotient is equivalent to calculate the quotient between

$$\max\left\{\prod_{j=1}^{s}\theta_j^{n^j}\,\middle|\,\sum_{j=1}^{s}\theta_j=1\right\} \quad \text{and} \quad \max\left\{\prod_{i,j}\theta_{ij}^{n_{ij}}\,\middle|\,\sum_{j=1}^{s}\theta_{ij}=1,\ i=1,\dots,s\right\}.$$

Hence

$$\Lambda=\frac{\prod_j (n^j/n)^{n^j}}{\prod_{i,j}(n_{ij}/n_{i\cdot})^{n_{ij}}}$$

where $n_{i\cdot}=\sum_{j=1}^{s} n_{ij}$. A Taylor expansion of $-2\log\Lambda$ finally shows that the criterion (6.11) takes the form

$$\sum_{i=1}^{s}\sum_{j=1}^{s}\frac{(n_{ij}-(n_{i\cdot}\, n^{j-i}/n))^2}{(n_{i\cdot}\, n^{j-i}/n)} > K^{-1}_{(s-1)^2}(1-\varepsilon).$$

References

1. Billingsley, P.: Statistical inference for Markov processes. Statistical research monographs, vol. **2**. Chicago: The university of Chicago press 1961
2. Cartan, H.: Théorie élémentaire des fonctions analytiques d'une ou plusieurs variables complexes. Paris: Hermann 1961
3. Chung, K. L.: Markov chains with stationary transition probabilities. Berlin-Heidelberg-New York: Springer 1967
4. Feller, W.: An introduction to probability theory and its applications, vol. **1**. New York: Wiley 1950
5. Ibid., vol. **2**

6. Höglund, T.: The exact estimate – a method of statistical estimation. To appear in Z. Wahrscheinlichkeitstheorie verw. Gebiete
7. Höglund, T.: A refined saddle point approximation. To appear in Arkiv för matematik (1974b)
8. Kato, T.: Perturbation theory for linear operators. Berlin-Heidelberg-New York: Springer 1966
9. Hinčin, A.I.: Mathematical foundations of statistical mechanics. New York: Dover 1949
10. Kolmogorov, A.N.: A local limit theorem for classical Markov chains. Izv. Akad. Nauk. SSSR, Ser. Mat. **13**, 281–300 (1949)
11. Lanford III, O.E.: Entropy and equilibrium states in classical statistical mechanics. In: Statistical mechanics and mathematical problems, pp. 1–113. Battelle Seattle 1971 Rencontres. Lecture notes in physics **20**. Berlin-Heidelberg-New York: Springer 1971
12. Martin-Löf, P.: Statistiska modeller, Anteckningar från seminarier läsåret 1969–1970 utarbetade av R. Sundberg, Institutionen for matematisk statistik, Stockholms universitet. 1970
13. Martin-Löf, P.: The notion of redundancy and its use as a quantitative measure of the deviation between a statistical hypothesis and a set of observational data. Scandinavian journal of statistics **1**, 3–18 (1974)
14. Rényi, A.: Probability theory. Amsterdam-London: North-Holland 1970
15. Richter, W.: Local limit theorems for large deviations. Theory of probability and its applications (transl.). Vol. 2, 206–220 (1957)
16. Saulis, L.I., Statuljavičus, V.A.: An asymptotic expansion for the probabilities of large deviations of sums of random variables that are connected in a Markov chain. Litovsk. Mat. Sb. **10**, 359–366 (1970)
17. Spitzer, F.: A variational characterization of finite Markov chains. Ann. Math. Statist. **43**, 303–307 (1972)
18. Statuljavičus, V.A.: Limit theorems fur sums of random variables that are connected in a Markov chain. I, II, III. Litovsk. Mat. Sb. **9**, 345–362 (1969); ibid. **9**, 635–672 (1969); ibid. **10**, 161–169 (1970)
19. Wielandt, H.: Unzerlegbare nicht-negative Matrizen. Math. Z. **52**, 642–648 (1950)

Thomas Höglund
Institut for matematisk statistik
Universitetsparken 5
DK-2100 Copenhagen
Denmark

(Received April 30, 1973)