Z. Wahrscheinlichkeitstheorie verw. Gebiete
58, 407-421 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Thermodynamics and Statistical Analysis of Gaussian Random Fields

H. Künsch

Mathematics Department, ETH Zentrum, CH-8092 Zürich

Summary. Gaussian fields are considered as Gibbsian fields. Thermodynamic functions are calculated for them and the variational principle is proved. As an application we get an approximation of log likelihood and an information theoretic interpretation of the asymptotic behaviour of the maximum likelihood estimator for Gaussian Markov fields.

0. Introduction

For the statistical analysis of spatial data observed on a regular lattice two approaches have been used. One of them generalizes techniques of time series analysis and is based on a rather artificial unilateral innovation type representation (Whittle [19], Tjøstheim [18]) or on periodic continuation of the data (Besag-Moran [5], Besag [4]). The other approach (Besag [3]) uses models of statistical mechanics, so called Gibbsian fields, where the conditional density given the field outside a finite set is considered. In this paper we want to push the second approach further since we believe that there are surprisingly many connections between statistical mechanics and statistical analysis: for instance both need an approximation of log likelihood, and Föllmer [9] discovered that the Gibbs variational principle can also be formulated in information theoretic terms which allow directly a statistical interpretation.

More precisley our paper contains the following results: In Sect. 1 we recall the results of Rosanov [16] and Dobrushin [7] on Gibbs representation and phase transition of Gaussian fields. In Sect. 2 we prove the convergence of the thermodynamic functions and the variational principle. There exist already results for more general real valued fields (Pirlot [14], Künsch [12]) but they do not cover all Gaussian cases and moreover in the Gaussian case the thermodynamic functions can be calculated explicitely and the proofs are easier. In Sect. 3 these results are applied to estimation problems: First we get a new and exact proof for Whittle's approximation of log likelihood. Later we specialise to the Markovian case where the maximum likelihood equations say that the

0044-3719/81/0058/0407/$03.00

covariances of the fitted model should coincide with the sample covariances in a neighborhood of lag zero. As a kind of converse it follows from the variational principle that Gaussian Markov fields are the natural choice if we take only the covariances near zero into account because they have maximal entropy. The variational principle implies also that the maximum likelihood estimator asymptotically minimizes the information gain of the true model with respect to the fitted model. The relation of this with Akaike's information criterion AIC is sketched. Finally we show in the last section that the covariances of an arbitrary field coincide with the covariances of a Gaussian Markov field for lags near zero if and only if the dimension of the lattice is smaller than three. This is closely connected with the absence of phase transition in dimension one and two.

Acknowledgement. I wish to express my gratitude to H. Föllmer and F. Hampel for their constant encouragement and advice while I was writing my Ph.D. thesis were these results are taken from. Thanks are also due to B. Ripley for useful comments, to X. Guyon who drew my attention to the importance of the difference between biased and unbiased sample covariances and to the referee who helped me to improve Theorem (2.11).

1. Gaussian Random Fields as Gibbsian Fields

By a *random field* we mean a stochastic process indexed by the d-dimensional lattice $\mathbb{Z}^d$, i.e. a family of random variables $(X_i)_{i\in\mathbb{Z}^d}$ or equivalently a probability measure v on the configuration space $(\mathbb{R})^{\mathbb{Z}^d}$. A random field will be called *stationary* if all finite dimensional distributions are invariant under translations of the index set $\mathbb{Z}^d$.

Let v be a stationary Gaussian field with mean value 0 and spectral density $f(x)$, $x\in[-\pi,\pi]^d$, satisfying

$$\sum_{k\in\mathbb{Z}^d}|a_k|<\infty \text{ where } a_k=(2\pi)^{-d}\int f(x)^{-1}\exp(-ikx)\,dx \quad (1.1)$$
$$\text{(in particular } \int f(x)^{-1}\,dx<\infty).$$

Since $f(x)=f(-x)$, we have $a_k=a_{-k}$.

Now we put $c=a_0^{-1}$ and $\alpha_k=-a_k/a_0$. Then it is straightforward to see that $R_k=E(X_iX_{i+k})=(2\pi)^{-d}\int f(x)\,e^{ikx}\,dx$ satisfies

$$R_k-\sum_{j\neq 0}\alpha_j R_{k+j}=c\,\delta_{0k} \quad (1.2)$$

which is equivalent to

$$E(X_i\,|\,X_j,\ j\neq i)=\sum_{j\neq 0}\alpha_j X_{i+j} \text{ and} \quad (1.3.a)$$

$$E\Big(X_i-\sum_{j\neq 0}\alpha_j X_{i+j}\Big)^2=c. \quad (1.3.b)$$

(The sum $\sum_j \alpha_j X_{i+j}$ converges a.s. and in L_1).

Since the conditional distribution of a Gaussian family is again Gaussian, it follows from (1.3) that the conditional distribution of X_i given $X_j = x_j$, $j \neq i$, has the density

(1.4) $$\pi^i(y \mid x) = Z_i(x)^{-1} \exp\Big(-\tfrac{1}{2}\big(a_0 y^2 + 2 \sum_{k \neq 0} a_k y x_{i+k}\big)\Big)$$

where the normalizing constant $Z_i(x)$ is equal to

$$(2\pi)^{1/2} \exp\left(\frac{1}{2}\frac{1}{a_0}\Big(\sum_{k \neq 0} a_k x_{i+k}\Big)^2\right) a_0^{-1/2}.$$

We can calculate the conditional expectation and distribution not only for single points but also for arbitrary finite sets:

(1.5) **Theorem.** *Let v be as above. Then we have for any finite set $V \subseteq \mathbb{Z}^d$: $E(X_i \mid X_j, j \notin V) = \sum_{j \notin V} h_{ij}^V X_j$ $(i \in V)$ and $E[(X_i - \sum_{j \notin V} h_{ij}^V X_j)\ (X_k - \sum_{j \notin V} h_{kj}^V X_j)] = g_{ik}^V$ where g^V is the matrix inverse of $(a_{i-j})_{i,j \in V}$ and $h_{ij}^V = -\sum_{k \in V} g_{ik}^V a_{j-k}$.*

The conditional distribution of $X_i, i \in V$, given $X_j = x_j, j \notin V$, has the density

$$\pi^V(y \mid x) = Z_V(x)^{-1} \exp\Big(-\tfrac{1}{2}\big(\sum_{i,j \in V} a_{i-j} y_i y_j + 2 \sum_{i \in V, j \notin V} a_{i-j} y_i x_j\big)\Big)$$

where

$$Z_V(x) = (2\pi)^{|V|/2} \exp\Big(\tfrac{1}{2} \sum_{i,j \in V} g_{ij}^V \big(\sum_{k \notin V} a_{j-k} x_k\big)\big(\sum_{k \notin V} a_{i-k} x_k\big)\Big) \det(a_{i-j})^{-1/2}.$$

Proof. Either one uses the fact that the π^V are determined by the π^i and calculates then the conditional expectation and covariance from the π^V (see Dobrushin [7] for this approach), or one calculates first the conditional covariance and the conditional expectation from Eq. (1.3), then the formula for π^V follows immediately (this approach is used in Spitzer [17] and Künsch [11]). □

In statistical mechanics one studies usually the set of all random fields for which $\sum_j |a_{i-j}|\,|X_j|$ converges a.s. and for which $\pi^V(y \mid x)\, d^V y$ is a version of the conditional distribution given $X_j, j \notin V$. We denote this set by $\mathscr{G}(\pi)$. It has been described completely by Dobrushin [7]: Every field in $\mathscr{G}(\pi)$ is a mixture of Gaussian fields with covariance $\operatorname{Cov}(X_i X_{i+k}) = (2\pi)^{-d} \int f(x) e^{ikx} dx$ and a mean value $E(X_i) = m_i$ satisfying $\sum_{j \in \mathbb{Z}^d} a_j m_{i-j} \equiv 0$. This equation has in general many non constant solutions which means that there are non stationary fields with translation invariant conditional distributions. If $\sum_{k \in \mathbb{Z}^d} a_k \neq 0$, i.e. $f(0) < \infty$, then the only constant solution of the above equation is $m_i \equiv 0$, nevertheless it is possible by a mixture of non stationary fields to have several stationary fields in $\mathscr{G}(\pi)$: for instance Rosanov [16] has shown that a stationary Gaussian field with mean 0 is in $\mathscr{G}(\pi)$ iff the spectral measure is $(2\pi)^{-d}(\sum_k a_k e^{ikx})^{-1} dx + dF_s(x)$ where $dF_s(x)$ is an arbitrary measure concentrated on $\{x, \sum_k a_k e^{ikx} = 0\}$. However it can be shown that there is only one stationary field in $\mathscr{G}(\pi)$ with finite second moment iff $\sum_k a_k e^{ikx}$ has no zero's, i.e. if $f(x)$ is bounded. If $\mathscr{G}(\pi)$ contains more than one (stationary) field, we say that (stationary) *phase transition* occurs. In the statistical part we will mainly deal with the case where all but finitely many $a_k=0$. Because then the conditional density π^V depends only on finitely many $X_j, j \notin V$, the fields in $\mathscr{G}(\pi)$ are called *Markovian.*

We close this section with a result on the convergence of g_{ij}^V which we will need later.

(1.6) **Theorem.** *g_{jk}^V, V finite, converges for fixed j and k to $R_{j-k}=(2\pi)^{-d} \cdot \int \exp(i(j-k)x) f(x)\,dx$ in the sense that for any $\varepsilon>0$ $|g_{jk}^V - R_{j-k}|<\varepsilon$ for all $V \supseteq V_0(\varepsilon, j, k)$.*

Proof. Since $\mathbb{R}$ is complete, it is sufficient to prove convergence for all increasing sequences $V_n \uparrow \mathbb{Z}^d$. By the martingale convergence theorem $E(X_j \mid X_i, i \notin V_n)$ converges in L_2 to $E(X_j \mid \mathscr{B}_\infty)$ where $\mathscr{B}_\infty = \bigcap_n \sigma(X_i, i \notin V_n)$. That $\mathscr{B}_\infty$ is trivial mod ν follows from Theorem 1 of Rosanov [16] or also from Theorem 2.1 of Preston [15]. Therefore the assertion follows because g_{jk}^V is the conditional covariance given X_i, $i \notin V$. □

2. Thermodynamics of Gaussian Fields

We fix a spectral density $f(x)$ satisfying (1.1). Unless stated otherwise ν denotes the Gaussian field with mean 0 and spectral density $f(x)$. In analogy to statistical physics the function

$$U_V(x) = \tfrac{1}{2} \sum_{i,j \in V} a_{i-j} x_i x_j \tag{2.1}$$

is called the energy of the configuration x in the domain V. In order to get a function independent of V, we will divide U_V by the number of points $|V|$ and take the limit as V increases to $\mathbb{Z}^d$. Other thermodynamic functions will be defined in a similar way and relations between them established. For simplicity we consider only the sequence of boxes $V_n = [-n, n]^d$, but the results are also true for more general sequences like the ones considered by Nguyen-Zessin [13].

(2.2) **Theorem.** *Let μ be stationary with $E_\mu(X_i^2)<\infty$. Then $(2n+1)^{-d} U_{V_n}(x)$ converges μ-a.s. and in $L_1(d\mu)$ to $\frac{1}{2}\sum_k a_k E_\mu(X_0 X_k \mid \mathscr{I})$ where $\mathscr{I}$ denotes the σ-field of invariant events. If especially $\mu=\nu$, then the limit equals $\frac{1}{2}$.*

Proof. By the d-dimensional ergodic theorem (see e.g. Dunford-Schwartz [8], VIII, 6.9) $(2n+1)^{-d} \sum_{i \in V_n} X_i \sum_{j \in \mathbb{Z}^d} a_{i-j} X_j$ converges a.s. and in L_1 to $\sum_k a_k E(X_0 X_k \mid \mathscr{I})$. In order to show that $2(2n+1)^{-d} U_{V_n}(x)$ has the same limit, we observe that for $k<n$

$$
\begin{aligned}
&(2n+1)^{-d}\Big|\sum_{i\in V_n} X_i \sum_{j\notin V_n} a_{i-j}X_j\Big| \\
&\quad\leqq (2n+1)^{-d} \sum_{\substack{i\in V_n\\ i\notin V_{n-k}}} |X_i| \sum_{j\in\mathbb{Z}^d} |a_{i-j}|\,|X_j| \\
&\qquad + (2n+1)^{-d} \sum_{i\in V_{n-k}} |X_i| \sum_{|i-j|>k} |a_{i-j}|\,|X_j|.
\end{aligned}
$$

The first expression converges for every fixed k to zero since by the ergodic theorem

$$\lim_n (2n+1)^{-d} \sum_{i\in V_n} Y_i = E(Y_0\,|\,\mathscr{I}) = \lim_n (2n+1)^{-d} \sum_{i\in V_{n-k}} Y_i$$

for any stationary random field (Y_i) with $E|Y_i|<\infty$. Applying the ergodic theorem once more we see that for fixed k the second expression converges to $E\big(\sum_{|j|>k} |a_j|\,|X_0 X_j|\,|\mathscr{I}\big)$. But this last expression is arbitrarily small since $\sum_{|j|>k} |a_j|\,|X_0 X_j|$ converges to zero for k going to infinity and is bounded by the integrable function $\sum_j |a_j|\,|X_0 X_j|$.

The last assertion of the theorem follows from the fact that $\mathscr{I}$ is trivial modulo ν and from Eq. (1.2). □

(2.3) *Remark.* The arguments used here are very similar to the ones used by Nguyen-Zessin [13], and the theorem follows also from their results.

(2.4) *Definition.* For a stationary random field μ with $E_\mu(X_i^2)<\infty$ we define the *energy* $e(\mu)$ as $\lim_n (2n+1)^{-d} E_\mu(U_{V_n}(x)) = \frac{1}{2}\sum_k a_k E_\mu(X_0 X_k)$.

We consider now the conditional density π^V of ν given the values of the field outside V (see Theorem (1.5)). First we prove

(2.5) **Theorem.** *The* $\lim_n (2n+1)^{-d} \log Z_{V_n}(0)$ *exists and equals* $\frac{1}{2}(\log(2\pi) + (2\pi)^{-d}\int \log(f(x))\,dx)$ (0 *means the boundary condition identically zero for all* $j\notin V$).

Proof. We consider the following family of spectral densities $f(x,\theta) = (a_0 + \theta\sum_{k\neq 0} a_k e^{ikx})^{-1}$ $(0\leqq\theta\leqq 1)$. Because of

(2.6) $$f(x,\theta) \leqq \max(a_0^{-1}, f(x))$$

$f(x,\theta)$ is really a spectral density, and we denote the corresponding Gaussian field by ν_θ. g_{ij}^V and $Z_V(x)$ are now also functions of θ.

From the rules for the derivation of a determinant and the definition of g_{ij}^V and $Z_V(0)$ it follows

(2.7) $$\frac{d}{d\theta}\log Z_V(0,\theta) = -\frac{1}{2}\sum_{i\neq j} g_{ji}^V a_{i-j} = -\frac{1}{2}\theta^{-1}\Big(|V| - a_0\sum_i g_{ii}^V\Big).$$

The convergence of g_{ii}^V (Theorem (1.6)) and translation invariance imply therefore that $(2n+1)^{-d}\frac{d}{d\theta}\log Z_{V_n}(0,\theta)$ converges to

$$-\tfrac{1}{2}(2\pi)^{-d}\int\theta^{-1}(1-a_0 f(x,\theta))\,dx=-\tfrac{1}{2}(2\pi)^{-d}\int\sum_{k\neq 0}a_k e^{ikx}f(x,\theta)\,dx.$$

By Schwartz inequality $|g_{ij}^V(\theta)|\leqq(g_{ii}^V(\theta)\,g_{jj}^V(\theta))^{1/2}\leqq E_{\nu_\theta}(X_i^2)$, so it follows from (2.6) and (2.7) that

$$\left|\frac{d}{d\theta}\log Z_V(0,\theta)\right|\leqq\tfrac{1}{2}E_{\nu_\theta}(X_i^2)\sum_{k\neq 0}|a_k|\,|V|\leqq\text{const.}\,|V|.$$

Integrating with respect to θ and using dominated convergence we see that $(2n+1)^{-d}\log Z_{V_n}(0,1)$ converges to

$$\tfrac{1}{2}(\log(2\pi)-\log a_0-(2\pi)^{-d}\int_0^1\int\sum_{k\neq 0}a_k e^{ikx}f(x,\theta)\,dx\,d\theta).$$

Therefore with Fubinis theorem

$$\lim_n(2n+1)^{-d}\log Z_{V_n}(0)=\tfrac{1}{2}(\log(2\pi)+(2\pi)^{-d}\int\log(f(x))\,dx).\quad\square$$

(2.8) *Definition.* We denote $\lim_n\,(2n+1)^{-d}\log Z_{V_n}(0)$ by p. In statistical physics it is called the *pressure.*

If μ is a random field and if the distribution of X_i, $i\in V$, has a density we denote it by $\mu_V(x)$.

(2.9) **Theorem.** *If* $f(x)$ *is in addition to* (1.1) *also bounded, then* $(2n+1)^{-d}\cdot(\log\nu_{V_n}(x)-\log\pi^{V_n}(x|0))$ *converges* μ-a.s. *and in* $L_1(d\mu)$ *to zero for any stationary field* μ *with* $E_\mu(X_i^2)<\infty$.

Proof. By the definition of the conditional density $\nu_V(x)=\int\pi^V(x\,|\,y)\,\nu(dy)$. Therefore we have

$$\begin{aligned}&\nu_V(x)/\pi^V(x\,|\,0)\\ &=\int\exp\Big(-\sum_{\substack{i\in V\\ j\notin V}}a_{i-j}x_i y_j-\tfrac{1}{2}\sum_{i,j\in V}g_{ij}^V\big(\sum_{k\notin V}a_{i-k}y_k\big)\cdot\big(\sum_{p\notin V}a_{j-p}y_p\big)\Big)\,\nu(dy).\end{aligned}\tag{2.10}$$

So by Jensen's inequality

$$\log\nu_V(x)-\log\pi^V(x\,|\,0)\geqq-\tfrac{1}{2}\sum_{i,j\in V}g_{ij}^V\sum_{k,p\notin V}a_{i-k}a_{j-p}R_{k-p}.$$

Now Theorem (1.5) shows that

$$\sum_{j\in V,\,p\notin V}g_{ij}^V a_{j-p}R_{k-p}=E_\nu\big(\sum_{p\notin V}h_{ip}^V X_p\cdot X_k\big)=E_\nu(X_i X_k),$$

and therefore $\log \nu_V(x) - \log \pi^V(x \mid 0) \geqq -\text{const.} \sum_{i \in V, k \notin V} |a_{i-k}|$. With the same arguments as in the proof of Theorem (2.2) we see that this lower bound is of the order $o(n^d)$.

For an upper bound we notice that g_{ij}^V is positive definite, so from (2.10)

$$\log \nu_V(x) - \log \pi^V(x \mid 0) \leqq \log \int \exp\left(-\sum_{i \in V, j \notin V} a_{i-j} x_i y_j\right) \nu(dy).$$

For a Gaussian random variable Z we find by direct computation $E(\exp(Z)) = \exp(E(Z) + \frac{1}{2}\operatorname{Var}(Z))$, therefore

$$\begin{aligned}
\log \nu_V(x) - \log \pi^V(x \mid 0) &\leqq \tfrac{1}{2} \sum_{i, k \in V} x_i x_k \sum_{j, p \notin V} a_{i-j} a_{k-p} R_{j-p} \\
&= \tfrac{1}{2} \sum_{j, p \notin V} \left(\sum_{i \in V} x_i a_{i-j}\right)\left(\sum_{i \in V} x_i a_{i-p}\right) R_{j-p} \\
&\leqq \tfrac{1}{2} \sup_x f(x) \sum_{j \notin V} \left(\sum_{i \in V} x_i a_{i-j}\right)^2.
\end{aligned}$$

This upper bound is also of the order $o(n^d)$ since with $Y_i = \sum_{k \in \mathbb{Z}^d} |a_{i-k}| \, |X_k|$ we have:

$$\begin{aligned}
\sum_{j \notin V} \left(\sum_{i \in V} X_i a_{i-j}\right)^2 &\leqq \sum_{i \in V, j \notin V} |X_i| \, |a_{i-j}| \, Y_j \\
&\leqq \sum_{i \in V_n, \notin V_{n-k}} |X_i| \sum_{j \in \mathbb{Z}^d} |a_{i-j}| \, Y_j + \sum_{i \in V_{n-k}} |X_i| \sum_{|i-j| > k} |a_{i-j}| \, Y_j,
\end{aligned}$$

and we can repeat the arguments of the proof of Theorem (2.2). □

Without the condition of a bounded spectral density, it is much more difficult to prove the above theorem. This has to do with the fact that then the so-called superstability condition is violated, see Künsch [12], Example (1.12) i). We will have to impose a stronger condition on the decay of the a_k than just summability.

(2.11) **Theorem.** *Suppose that the spectral density $f(x)$ is such that $|a_k| \leqq \text{const.} |k|^{-\gamma}$ for $\gamma > 3d/2$, and let ν be any stationary Gaussian field in $\mathcal{G}(\pi)$. Then for any stationary field μ with $E_\mu(X_i^2) < \infty$ $(2n+1)^{-d}(\log \nu_{V_n}(x) - \log \pi^{V_n}(x \mid 0))$ converges to zero μ-a.s. and in $L_1(d\mu)$.*

Proof. The lower bound is proved as in the previous theorem, but for the upper bound we need a different argument: It is obvious that all eigenvalues of $(a_{i-j})_{i, j \in V}$ are $\leqq \sum_k |a_k|$, and therefore all eigenvalues of (g_{ij}^V) are bounded below. Therefore it follows from (2.10) that for every $k < n$

$$\begin{aligned}
\nu_{V_n}(x)/\pi^{V_n}(x \mid 0) \leqq \int &\exp\left(-\sum_{i \in V_{n-k}} x_i \sum_{j \notin V_n} a_{i-j} y_j\right) \\
&\cdot \exp\left(\sum_{i \in V_n, \notin V_{n-k}} \left(|x_i| \left|\sum_{j \in V_n} a_{i-j} y_j\right| - C\left(\sum_{j \notin V_n} a_{i-j} y_j\right)^2\right)\right) \nu(dy).
\end{aligned}$$

Estimating the second factor of the integrand by its maximum and then applying the same argument as in the proof of (2.9) we find

$$|\log \nu_{V_n}(x) - \log \pi^{V_n}(x \mid 0)| \leqq \text{const.} \sum_{i \in V_n, \notin V_{n-k}} x_i^2$$
$$- E_\nu(X_i) \sum_{i \in V_{n-k}} x_i \sum_{j \notin V_n} a_{i-j} + \tfrac{1}{2} E_\nu(X_i^2) \Big(\sum_{i \in V_{n-k}} |x_i| \sum_{j \notin V_n} |a_{i-j}| \Big)^2.$$

We choose now $k = k(n) = n^\alpha$ with $d(2\gamma - 2d)^{-1} < \alpha < 1$. Then the first and the second term above are of the order $o(n^d)$ by similar arguments as in the proof of (2.2). The third term is bounded by

$$\Big(\sum_{|j| > k(n)} |a_j| \Big)^2 \Big(\sum_{i \in V_n} |x_i| \Big)^2 \leqq \text{const.}\, k(n)^{2(d-\gamma)} \Big(\sum_{i \in V_n} |x_i| \Big)^2$$
$$\leqq \text{const.}\, n^d \cdot n^{2\alpha(d-\gamma)+d} \Big((2n+1)^{-d} \sum_{i \in V_n} |x_i| \Big)^2$$

which is of the order $o(n^d)$. □

Before we can state now the main results of this section, we need two more definitions:

(2.12) *Definition.* Let μ be a stationary field with $E_\mu(x_i^2) < \infty$ and such that the density $\mu_V(x)$ exists. Then we define the *entropy* $s(\mu)$ of μ to be the $\lim (2n+1)^{-d} \cdot E_\mu(-\log \mu_{V_n}(x))$. A general result of Ruelle (see e.g. Preston [15], Theorem 8.1) says that this limit always exists (it can be equal to $-\infty$).

(2.13) *Definition.* Let μ and ν be stationary fields for which the densities $\mu_V(x)$ and $\nu_V(x)$ exist. Then we define the *information gain* $h(\mu, \nu)$ of μ with respect to ν to be the $\lim (2n+1)^{-d} E_\mu(\log(\mu_{V_n}(x)/\nu_{V_n}(x)))$ if the integral and the limit exist. Otherwise we put $h(\mu, \nu) = +\infty$.

(2.14) **Theorem.** *Let $f(x)$ be a spectral density which is either bounded and satisfies* (1.1) *or $|a_k| \leqq \text{const.}\, |k|^{-\gamma}$ for $\gamma > 3d/2$, let ν be a stationary Gaussian field in $\mathscr{G}(\pi)$ and let μ be any stationary field such that $E_\mu(X_i^2) < \infty$, the density $\mu_V(x)$ exists and $s(\mu) > -\infty$. Then*

i) $e(\nu) - s(\nu) = -p = \inf_\mu (e(\mu) - s(\mu))$,

ii) $s(\nu)$ *is equal to* $\frac{1}{2}(1 + \log(2\pi) + (2\pi)^{-d} \int \log(f(x))\, dx)$,

iii) $(2n+1)^{-d} \log \nu_V(x)$ *converges* ν-a.s. *and in* $L_1(d\nu)$ *to* $\frac{1}{2} \sum_k a_k E_\nu(X_0 X_k | \mathscr{I}) + p$,

iv) $h(\mu, \nu) = e(\mu) - s(\mu) + p$. *If μ is Gaussian with mean* 0 *and spectral density $g(x)$ satisfying the same conditions as $f(x)$, then* $h(\mu, \nu) = \frac{1}{2}(2\pi)^{-d} \int (g(x)/f(x) - 1 - \log(g(x)/f(x)))\, dx$.

Proof. The inequality $e(\mu) - s(\mu) \geqq -p$ follows from Jensens inequality, see Pirlot [14]. All other assertions follow from the previous results, observing that

$$\log \nu_V(x) = -\log Z_V(0) - U_V(x) + (\log \nu_V(x) - \log \pi^V(x \mid 0)). \quad \square$$

(2.15) *Remarks.* i) is one direction of *Gibbs variational principle.* For the converse (i.e. every μ for which $e(\mu) - s(\mu) = -p$ is in $\mathscr{G}(\pi)$) see Preston [15], Sect. 7, and Künsch [12]. iii) is a *d-dimensional version of the theorem of McMillan and Breiman.* i) and iv) imply that in particular $h(\mu, \nu)=0$ if also μ is in $\mathcal{G}(\pi)$. i) and ii) can also be proved if only condition (1.1) holds: Theorems (2.2) and (2.5) were already proved under this condition, and then one can proceed generalizing Pirlots arguments in [14] slightly. However we do not know if iii) and iv) for which Theorem (2.9) or (2.11) are used remain true under the weaker assumption (1.1). It would be very interesting to find a spectral density $f(x)$ for which the statement of the Theorems (2.9) and (2.11) does not hold, but we have no idea how to construct such a counterexample.

3. Estimation Problems

Let $(X_i)_{i\in\mathbb{Z}^d}$ be a stationary Gaussian field with mean 0 and spectral density $f(x, \theta)$ depending on unknown parameters $\theta\in\Theta\subseteq\mathbb{R}^K$ and suppose we have observed the field in the box $V_n=[-n, n]^d$. We treat here the problem how to estimate θ from that sample. For this we want to use some approximate maximum likelihood estimator. If $f(x, \theta)$ is bounded and satisfies (1.1) or if $f(x, \theta)$ is such that $|a_k|\leqq\text{const.}|k|^{-\gamma}$ with $\gamma>3d/2$ for all θ, we have from Theorems (2.5), (2.9) and (2.11)

$$(3.1)\quad -2(2n+1)^{-d}\log \nu_{V_n}(x, \theta)=\log(2\pi)+(2\pi)^{-d}\int\log(f(x, \theta))\,dx$$
$$+\sum_k a_k(\theta)\,C_k+o(1)=\log(2\pi)+(2\pi)^{-d}\int(\log(f(x, \theta))+I(x)/f(x, \theta))\,dx+o(1),$$

where

$$(3.2)\quad C_k=(2n+1)^{-d}\sum_{i\in V_n(k)} X_i X_{i+k} \text{ with } V_n(k)=\{i\in V_n, \text{ such that } i+k\in V_n\}, \text{ and}$$

$$(3.3)\quad I(x)=\sum_k C_k\,e^{ikx}$$

are the sample covariances and the sample spectral density. This approximation was first given by Whittle [19], but it seems us difficult to make his justification exact, see also Guyon [10], Sect. 5. Our arguments here are exact, but completely different.

To the same degree of accuracy we may also use instead of C_k the unbiased sample covariances

$$(3.4)\qquad C_k^*=\left(\prod_1^d(2n+1-|k_i|)\right)^{-1}\sum_{i\in V_n(k)} X_i X_{i+k}\qquad (k=(k_1, \ldots, k_d)).$$

The differences will later become important.

In the following we specialise to the Markovian case where all a_k are unknown. Because we must have $a_k=a_{-k}$, we can write

$$(3.5)\qquad f(x, \theta)=\Big(\sum_{k\in M}\theta_k\cos(kx)\Big)^{-1},$$

where M is some finite subset, $M \ni 0$ and $M \subseteq \{i \leqq 0\}$ ($\leqq$ denotes the lexicographic order on $\mathbb{Z}^d$). The admitted parameter set is then $\Theta = \{\theta \in \mathbb{R}^{|M|}, \sum_{k \in M} \theta_k \cos(kx) \geqq 0, \int f(x, \theta)\, dx < \infty\}$. We denote the interior of Θ by $\mathring{\Theta}$.

Taking derivatives of (3.1) gives the following equations

(3.6) $(2\pi)^{-d} \int \cos(kx) f(x, \hat{\theta})\, dx = C_k$, $k \in M$, respectively

(3.7) $(2\pi)^{-d} \int \cos(kx) f(x, \hat{\theta})\, dx = C_k^*$, $k \in M$.

These equations appear for the first time in Besag [3], he used Whittle's result. They say that the covariances of the fitted model should be equal to the sample covariances for lags $k \in M$.

The asymptotic behaviour of $\hat{\theta}$ – also in cases when the true distribution is not a Gaussian Markov field – are obtained easily. Let Φ denote the mapping $\Theta \to \mathbb{R}^{|M|}$ given by $\Phi_k(\theta) = (2\pi)^{-d} \int \cos(kx) f(x, \theta)\, dx$, $k \in M$. Then we have

(3.8) **Theorem.** *Let μ be stationary such that $E_\mu(X_i^2) < \infty$ and $\mathscr{I}$ is trivial mod μ. If there is a $\theta^0 \in \mathring{\Theta}$ with $\Phi_k(\theta^0) = \mathrm{Cov}_\mu(X_0 X_k)$, then this θ^0 is unique and $\hat{\theta}$ defined by* (3.6) *or* (3.7) *converges μ-a.s. to θ^0.*

Proof. Φ is differentiable in $\mathring{\Theta}$ with Jacobian

$$\frac{d\Phi_k}{d\theta_j} = -(2\pi)^{-d} \int \cos(kx) \cos(jx) f(x, \theta)^2\, dx. \tag{3.9}$$

Therefore

$$\sum_{j, k \in M} c_k \frac{d\Phi_k}{d\theta_j} c_j = -(2\pi)^{-d} \int \Big(\sum_{k \in M} c_k \cos(kx)\Big)^2 f(x, \theta)^2\, dx,$$

and because the zeros of a trigonometric polynomial $\not\equiv 0$ have Lebesgue measure zero, the Jacobian is strictly negative definite. The uniqueness of θ^0 follows then by standard arguments and the convergence of $\hat{\theta}$ holds because C_k and C_k^* converge to $\mathrm{Cov}_\mu(X_0 X_k)$ by the ergodic theorem. □

We postpone the question when there is such a θ^0 to the next section. With the variational principle of Theorem (2.14) we can interprete the above result. Fix a $\theta^0 \in \Theta$ and put $R_k = \Phi_k(\theta^0)$. We then consider the following sets

$\mathscr{M}_1 = \{\mu$ stationary, $E_\mu(X_i X_{i+k}) = R_k$ for $k \in M$, $s(\mu) > -\infty\}$,
$\mathscr{M}_2 = \{\nu$ stationary, Gaussian, $E_\nu(X_i) = 0$, with spectral density $(\sum_{k \in M} \theta_k \cos kx)^{-1}$, $\theta \in \Theta\}$, and let $\nu_0 \in \mathscr{M}_1 \cap \mathrm{M}_2$.

(3.10) **Theorem.** i) $s(\nu_0) = \sup\{s(\mu), \mu \in \mathscr{M}_1\}$.
ii) *For any $\mu \in \mathscr{M}_1$ $h(\mu, \nu_0) = \inf\{h(\mu, \nu), \nu \in \mathscr{M}_2\}$.*
iii) *For any $\nu \in \mathscr{M}_2$ $h(\nu_0, \nu) = \inf\{h(\mu, \nu), \mu \in \mathscr{M}_1\}$.*
(*s and h are the entropy and the information gain as in* (2.12) *and* (2.13)).

Proof. For $\theta \in \Theta$ we denote by $e_\theta(\cdot)$ the corresponding energy. First we observe that $e_\theta(\mu)$ is the same for all $\mu \in \mathscr{M}_1$. Therefore i) follows from (2.14) i). iii) follows from (2.14) iv) and from the first assertion of this theorem. Furthermore using again (2.14) i) and iv) we have:

$$h(\mu, \nu)=e_\theta(\mu)-s(\mu)+p_\theta=e_\theta(\nu_0)-s(\nu_0)+p_\theta+e_0(\nu_0)-s(\mu)-e_0(\nu_0)$$
$$+s(\nu_0)=h(\nu_0, \nu)+e_0(\mu)-s(\mu)+p_0=h(\nu_0, \nu)+h(\mu, \nu_0).$$

So ii) is also proved because h is always nonnegative. □

(3.11) *Remark.* The same results in the case of independent observations of a k-dimensional Gaussian sample were given by Dempster [6].

The entropy is a measure for the simplicity of a distribution whereas the information gain measures the goodness of fit. The intuitive meaning of Theorem (3.10) is therefore as follows:

i) If we take only the covariances of lags $k \in M$ into consideration, we should choose the Gaussian Markov field ν_0.

ii) Maximum-Likelihood gives asymptotically the best fit to the true model μ.

iii) If we use wrong information on the covariances, the Gaussian Markov field ν_0 gives the best fit provided the true model is also a Gaussian Markov field.

In practice the set M is not known, and so one will fit the parameters for different M's and then choose the final M by a criterion which depends also on the sample $X_i, i \in V_n$. We remark that this results also in smoothing the sample covariances and estimating the spectral density. Akaike [1] proposed a general rule for this type of problems, namely to minimize the so-called Akaike information criterion (AIC). In our situation it says to select that set M for which

$$\sum_{k \in M} \hat{\theta}_k^M \cdot C_k^* - (2\pi)^{-d} \int \log\Big(\sum_{k \in M} \hat{\theta}_k^M \cos(kx)\Big)\, dx + 2|M|(2n+1)^{-d} \tag{3.12}$$

is minimal. We briefly sketch its justification because the basic idea is to choose the information gain $h(\mu, \hat{\nu}_M)$ as loss function for fitting the model $\hat{\nu}_M$ to the true model μ. The risk function is then $E(h(\mu, \hat{\nu}_M))$ (expectation with respect to the distribution of the estimated parameters). We are going to show that the above rule means to choose that set M which minimizes an unbiased estimate of this risk function.

In order to make computational progress, we assume that the true model μ is stationary and satisfies

(3.13) $E_\mu(X_i^2) < \infty$, and for all M considered there is a $\theta^{M,0} \in \mathring{\Theta}^M$ with $R_k = \mathrm{Cov}_\mu(X_0 X_k) = \Phi_k(\theta^{M,0})$, $k \in M$.

(3.14) $(2n+1)^{d/2}(C_k^* - R_k)$, $k \in \mathbb{Z}^d$, is asymptotically normal with mean 0 and a certain covariance Γ_{jk}.

From now on we always use the $\hat{\theta}_k^M$ defined by (3.7). From (3.14) and the δ-technique we see immediately that $(2n+1)^{d/2}(\hat{\theta}_k^M - \theta_k^{M,0})$, $k \in M$, is asymptotically normal with mean 0 and covariance matrix $\left(\frac{\partial \Phi}{\partial \theta}\right)^{-1} \Gamma \left(\frac{\partial \Phi}{\partial \theta}\right)^{-1}$. So applying the δ-technique again and using that $R_k = \phi_k(\theta^{M,0})$ we find

$$h(\mu, \hat{v}_M) = \tfrac{1}{2} \sum_{k \in M} \hat{\theta}_k^M R_k - s(\mu) - \tfrac{1}{2}(2\pi)^{-d} \int \log\big(\sum_{k \in M} \hat{\theta}_k^M \cos(kx)\big)\, dx$$
$$= h(\mu, v_M^0) - \tfrac{1}{4} \sum_{k, j \in M} (\hat{\theta}_k^M - \theta_k^{M,0}) \frac{\partial \Phi_k}{\partial \theta_j} (\hat{\theta}_j^M - \theta_j^{M,0}) + o((2n+1)^{-d}).$$

So the risk function is asymptotically equal to

$$h(\mu, v_M^0) + \tfrac{1}{4} \operatorname{trace}\left(\left(-\frac{\partial \Phi}{\partial \theta}\right)^{-1} \Gamma\right)(2n+1)^{-d}.$$

Finally in order to estimate $h(\mu, v_M^0)$ unbiasedly, we observe

$$\sum_{k \in M} \hat{\theta}_k^M C_k^* - (2\pi)^{-d} \int \log\big(\sum_{k \in M} \hat{\theta}_k^M \cos(kx)\big)\, dx = \sum_{k \in M} \theta_k^{M,0} C_k^*$$
$$- (2\pi)^{-d} \int \log\big(\sum_{k \in M} \theta_k^{M,0} \cos(kx)\big)\, dx + \sum_{k \in M} (\hat{\theta}_k^M - \theta_k^{M,0})(C_k^* - R_k)$$
$$- \tfrac{1}{2} \sum_{j, k \in M} (\hat{\theta}_k^M - \theta_k^{M,0}) \frac{\partial \Phi_k}{\partial \theta_j} (\hat{\theta}_j^M - \theta_j^{M,0}) + o((2n+1)^{-d}).$$

Therefore because

$$C_k^* - R_k = \sum_{j \in M} \frac{\partial \Phi_k}{\partial \theta_j} (\hat{\theta}_j^M - \theta_j^{M,0}) + o((2n+1)^{-d/2}),$$
$$\sum_{k \in M} \hat{\theta}_k^M C_k^* - (2\pi)^{-d} \int \log\big(\sum_{k \in M} \hat{\theta}_k^M \cos(kx)\big)\, dx$$
$$+ \tfrac{1}{2} \operatorname{trace}\left(\left(\frac{\partial \Phi}{\partial \theta}\right)^{-1} \Gamma\right)(2n+1)^{-d}$$

will be an asymptotically unbiased estimate of $\sum_{k \in M} \theta_k^{M,0} R_k - (2\pi)^{-d} \int \log\big(\sum_{k \in M} \theta_k^{M,0} \cos(kx)\big)\, dx$. $s(\mu)$ is independent of M and need not be estimated. So taking all these results together we see that we have to choose that M for which

$$\sum_{k \in M} \hat{\theta}_k^M C_k^* - (2\pi)^{-d} \int \log\big(\sum_{k \in M} \hat{\theta}_k^M \cos(kx)\big)\, dx$$
$$+ \operatorname{trace}\left(\left(-\frac{\partial \Phi}{\partial \theta}\right)^{-1} \Gamma\right)(2n+1)^{-d}$$

is minimal. Off course Γ is still unknown, but Theorem (3.13) below shows that $\Gamma_{kj} = 2(2\pi)^{-d} \int \cos(kx) \cos(jx) g(x)^2\, dx$ if the true model μ is Gaussian with spectral density $g(x)$. Therefore (3.12) is justified if the true model is Gaussian and the spectral density can be well approximated by a density of the form $\big(\sum_{k \in M} \theta_k \cos(kx)\big)^{-1}$.

(3.13) **Theorem.** *Let μ be stationary Gaussian with mean* 0 *and spectral density* $g(x)$. *If* $\sum_{k \in \mathbb{Z}^d} \left|\int \cos(kx) g(x)^{1/2}\, dx\right| < \infty$, *then* $(2n+1)^{d/2}(C_k^* - R_k)$, $k \in \mathbb{Z}^d$, *is asymp-*

totically normal with mean 0 *and covariance matrix* $2(2\pi)^{-d}\int\cos(kx)\cos(jx)g(x)^2\,dx$. *If in particular* $g(x)^{-1}=\sum_{k\in M}\theta_k\cos(kx)>0$ *the above condition is satisfied.*

Proof. $g(x)^{-1/2}e^{ikx}$, $k\in\mathbb{Z}^d$, is a complete orthonormal system in $L_2((2\pi)^{-1}g(x)\,dx)$. This implies that we can represent X_i as $\sum_{k\in\mathbb{Z}^d}g_k U_{i+k}$ with U_i i.i.d. $\mathcal{N}(0,1)$ and $g_k=(2\pi)^{-1}\int\cos(kx)g(x)^{1/2}\,dx$. The proof can then be carried over from the case $d=1$ (Anderson [2], 8.4.2) without any difficulties. Finally let $g(x)^{-1}=\sum_{k\in M}\theta_k\cos(kx)$ and put $c=\inf g(x)^{-1}$, $C=\sup g(x)^{-1}$. Using the power series for $z^{-1/2}$ at $(C+c)/2$ it is easily seen that in this case the g_k decay even exponentially. □

(3.14) *Remark.* For Theorem (3.13) and the justification of the AIC it is important to use C_k^* instead of C_k because $E(C_k)=R_k+O(n^{-1})$, so this bias can be much bigger than the random fluctuations and the correction terms in the AIC. See also Guyon [10].

4. Existence of Solutions of the Maximum Likelihood Equations

As we have seen in Sect. 1 there exist more than one stationary measure in $\mathcal{G}(\pi)$ with $E(X_i^2)<\infty$ iff $f(x)^{-1}$ has zeros. In the Markovian case, i.e. when $f(x)^{-1}$ is a trigonometric polynomial, $f(x)^{-1}$ can have zero's iff the dimension of the lattice is bigger than two. This is easily seen as follows: If $f(x_0)^{-1}=0$, then x_0 is a minimum of f^{-1} and therefore by the Taylor formula $f(x)^{-1}\leqq\text{const.}\,\|x-x_0\|^2$ near x_0. But $\|x-x_0\|^{-2}$ is integrable only in dimensions $\geqq 3$. The example $f(x)^{-1}=1-\frac{1}{d}\sum_{k=1}^{d}\cos x_k$ shows that zero's can really occur for $d\geqq 3$.

This phenomenon is also the reason that the equation $\Phi_k(\theta)=R_k=(2\pi)^{-d}\int\cos(kx)g(x)\,dx$ ($k\in M$) does not have a solution for an arbitrary spectral density $g(x)$ if $d\geqq 3$. Namely take $M=\left\{i=(i_1,\dots,i_d),\ i\leqq 0,\ \sum_{k=1}^{d}|i_k|\leqq 1\right\}$ and $R_k=\rho R_0$ for $0\neq k\in M$ and $|\rho|<1$. By symmetry reasons we must have $\theta_k\equiv\theta_1$, $0\neq k\in M$, and $\theta\in\Theta$ iff $\theta_0>0$ and $|\theta_1/\theta_0|\leqq\frac{1}{d}$. Moreover $\Phi_k(\theta)/\Phi_0(\theta)$ is a continuous function of θ_1/θ_0 by (2.6), so the range of possible nearest neighbor correlation is a closed subset of $(-1,+1)$.

We are now going to show that solutions of $\Phi_k(\theta)=R_k$ always exist for $d\leqq 2$ because then $f(x)^{-1}$ cannot have zero's.

(4.1) **Theorem.** *If* $d\leqq 2$, *then for any finite* M *and any spectral density* $g(x)\not\equiv 0$ $\Phi_k(\theta)=R_k=(2\pi)^{-d}\int\cos(kx)g(x)\,dx$ ($k\in M$) *has exactly one solution.*

Proof. Uniqueness was already proved in Theorem (3.8). Let W be the set of possible covariances R_k, $k\in M$. We are going to show here that $\partial\Phi(\Theta)\cap W=\emptyset$ which implies the existence of a solution because W is connected and $\Phi(\Theta)$ is open. Assume that there is a $(R_k)_{k\in M}\in\partial\Phi(\Theta)\cap W$. Then there is a sequence $(\theta^n)\subseteq\Theta$ such that $\lim\Phi_k(\theta^n)=R_k$ and $\rho_k=\lim\theta^n_k/\theta^n_0$ exists. If we put $P(x)=\sum_{k\in M}\rho_k\cos(kx)$, then we have for any $\varepsilon>0$

$$\frac{\int P(x)g(x)\,dx}{\int g(x)\,dx}=\frac{\sum_{k\in M}\rho_k R_k}{R_0}=\lim\frac{\int P(x)(\sum_{k\in M}\theta^n_k/\theta^n_0\cos(kx))^{-1}\,dx}{\int(\sum_{k\in M}\theta^n_k/\theta^n_0\cos(kx))^{-1}\,dx}$$

$$\leqq\varepsilon+(\liminf\int(\sum_{k\in M}\theta^n_k/\theta^n_0\cos(kx))^{-1}\,dx)^{-1},$$

because $P(x)\leqq\sum_{k\in M}\theta^n_k/\theta^n_0\cos(kx)+\varepsilon$ for n big enough. Because $(R_k)\notin\Phi(\Theta)$ and $R_0\neq 0$, $P(x)$ must have at least one zero. Therefore with Fatou's lemma and because ε was arbitrary we have $\int P(x)g(x)\,dx=0$. But this leads to a contradiction because $P(x)$ is zero only on a set of Lebesgue measure zero. □

(4.2) *Remark.* The biased covariances C_k, $k\in\mathbb{Z}^d$, are strictly positive definite iff $X_i\not\equiv 0$, so (3.6) has always a solution for $d\leqq 2$. This need not to be true for the unbiased covariances.

Finally we remark that all the results in this paper can be generalized also to Gaussian random fields with values in $\mathbb{R}^k$ without any major difficulties. Details are given in the authors Ph.D. thesis.

References

1. Akaike, H.: Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory, eds. Petrov B.N. and Csaki F., Akademiai Kiado Budapest, 267–281 (1973)
2. Anderson, T.W.: The statistical analysis of time series. New York: Wiley 1971
3. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. J. Roy. Statist. Soc. Ser. B **36**, 192–236 (1974)
4. Besag, J.: Errors-in-variables estimation for Gaussian lattice schemes. J. Roy. Statist. Soc. Ser. B **39**, 73–78 (1977)
5. Besag, J., Moran, P.A.P.: On the estimation and testing of spatial interaction for Gaussian lattice processes. Biometrika **62**, 555–562 (1975)
6. Dempster, A.P.: Covariance selection. Biometrics **28**, 157–175, (1972)
7. Dobrushin, R.L.: Gaussian random fields – Gibbsian point of view. Multicomponent random systems, eds. Dobrushin R.L., Sinai Ya.G. New York: M. Dekker 119–151, 1980
8. Dunford, N., Schwartz, J.T.: Linear Operators I. New York: Interscience 1958
9. Föllmer, H.: On entropy and information gain in random fields, Z. Wahrscheinlichkeitstheorie verw. Geb. **26**, 207–217 (1973)
10. Guyon, X.: Parameter estimation for stationary processes on $\mathbb{Z}^d$. [To appear]
11. Künsch, H.: Gaussian Markov random fields. J. Fac. Sci. Univ. Tokyo Sec. 1A Math. **26**, 53–73 (1979)
12. Künsch, H.: Almost sure entropy and the variational principle for random fields with unbounded state space. [To appear in Z. Wahrscheinlichkeitstheorie verw. Geb.]
13. Nguyen, X.X., Zessin, H.: Ergodic theorems for spatial processes. Z. Wahrscheinlichkeitstheorie verw. Geb. **48**, 133–158 (1979)

14. Pirlot, M.: A strong variational principle for continuous spin systems. J. Appl. Probability **17**, 47–58 (1980)
15. Preston, C.: Random fields. Lecture Notes in Math. **534**, Berlin - Heidelberg - New York: Springer 1976
16. Rosanov, Yu.A.: On Gaussian fields with given conditional distributions. Theory Probability Appl. **12**, 381–391 (1967)
17. Spitzer, F.: Champs de Markov Gaussiens, Ecole d'été de probabilité Saint Flour III. Lecture Notes in Math. **390**, 179–185 Berlin - Heidelberg - New York: Springer 1974
18. Tjøstheim, D.: Statistical spatial series modeling. Adv. Appl. Probability **10**, 130–154 (1978)
19. Whittle, P.: On stationary processes in the plane. Biometrika **41**, 450–462 (1954)

Received February 19, 1981; in revised form July 31, 1981