Journal of Statistical Physics, Vol. 104, Nos. 1/2, 2001

Entropy and Random Vectors

Oliver Johnson[1] **and Yurii Suhov**[1]

Received June 27, 2000; revised February 21, 2001

Barron[(1)] produced a proof of the Central Limit Theorem for real-valued IID random variables, in the sense of convergence in relative entropy. Here, we establish a similar result for independent real-valued random vectors, not necessarily identically distributed. The main developments required are a generalisation of De Bruijn's identity, and various inequalities proposed in ref. 2.

KEY WORDS: Normal convergence; entropy; Fisher information.

1. INTRODUCTION AND NOTATION

The Central Limit Theorem shows convergence to the multivariate normal of normalised sums of random variables with fixed covariance. Since the multivariate normal maximises information-theoretic entropy for fixed covariance, we can therefore view the Central Limit Theorem as a "maximum entropy" result. This approach produced a proof of convergence in the sense of relative entropy for IID real random vectors.[(1)] It also unites our view of convergence on the real line with convergence to Haar measure on compact groups,[(7)] which is also an entropy maximisation result.

In this paper, the result of ref. 1 is extended to independent, not necessarily identical random vectors. Previously, Johnson[(6)] extended the results to independent, not necessarily identical random variables. In this case we require control over the covariance, as well as conditions which correspond directly to those imposed in ref. 6. The central limit theorem for independent random vectors was previously considered in ref. 3. This paper follows a similar structure of argument, but uses weaker assumptions. The role and significance of entropy in Statistical Mechanics is well established. Here, we only mention one possible application of our result:

[1] Statistical Laboratory, CMS, Wilberforce Road, Cambridge, CB3 DWB, United Kingdom; e-mail: otj1000@cam.ac.uk

absence of spontaneous magnetisation in 2D lattice models with continuous symmetries. See refs. 5 and 12 and the bibliography there.

Throughout this paper, given an $(n \times n)$ matrix $B = (B_{ij})$, $|B|$ denotes $\max_{i,j} |B_{ij}|$ and $\operatorname{tr}(B)$ stands for the trace $\sum_i B_{ii}$. We write $B > 0$ to mean that B is positive definite and non-singular.

We will consider a sequence of independent random vectors $\mathbf{X}^{(i)} = (X_1^{(i)},\ldots, X_n^{(i)})$ $(i = 1, 2,\ldots)$ taking values in the Euclidean space $\mathscr{R}^n$, with densities, mean zero and finite covariance matrices $C^{(i)} = \mathsf{E}\mathbf{X}^{(i)}\mathbf{X}^{(i)\mathsf{T}} > 0$ and set $v_i = \operatorname{tr}(C^{(i)})$. For a non-empty set of positive integers S, define $\mathbf{X}^{(S)} = \sum_{i \in S} \mathbf{X}^{(i)}$, $C^{(S)} = \sum_{i \in S} C^{(i)}$, $v_S = \operatorname{tr}(C^{(S)})$ and $\mathbf{U}^{(S)} = (\sum_{i \in S} \mathbf{X}^{(i)} / \sqrt{v_S}$ with density $g^{(S)}$. We require certain conditions below.

Condition 1. Convergent Covariance Condition. There exists a matrix $C > 0$ such that

$$\lim_{W \to \infty} \left(\sup_{S:\, v_S \geqslant W} \left| \frac{C^{(S)}}{v_S} - C \right| \right) = 0.$$

For the rest of this paper, C stands for this matrix. We will write $C^{-1} = (C_{ij}^{-1})$ for the inverse of C, and I for the identity matrix. This Condition implies that for all $\tau > 0$, $\operatorname{tr}((C^{(S)}/v_S + C\tau)^{-1} C - I/(1+\tau)) \to 0$.

Definition 1.1. For $t > 0$, define the n-dimensional ellipsoids: $\mathscr{E}(t) = \{\mathbf{z} = (z_1,\ldots, z_n) : \sum_{i,j} C_{ij}^{-1} z_i z_j \leqslant t\}$.

If random vector $\mathbf{X} = (X_1,\ldots, X_n)$ has mean zero and covariance B then the expectation $\mathsf{E} \sum_{i,j} C_{ij}^{-1} X_i X_j = \operatorname{tr}(C^{-1}B)$, so by Chebyshev:

$$\mathsf{P}(\mathbf{X} \notin \mathscr{E}(t)) \leqslant \frac{1}{t} \mathsf{E} \sum_{i,j} C_{ij}^{-1} X_i X_j = \frac{1}{t} \operatorname{tr}(C^{-1}B).$$

Take $K_0 = \max(\text{eigenvalue of } C^{-1})$ then if $\operatorname{tr}(B) = 1$ then $\operatorname{tr}(C^{-1}B) \leqslant K_0$.

Condition 2. Uniform Lindeberg Condition. There exists a decreasing function $\psi(R)$, such that $\psi(R) \to 0$ as $R \to \infty$ and for all $R > 0$, $i = 1, 2,\ldots,$ and $r = 1,\ldots, n$:

$$\mathsf{E} X_r^{(i)2} I(\mathbf{X}^{(i)} \notin \mathscr{E}(R v_i)) \leqslant v_i \psi(R).$$

By Cauchy–Schwarz, if this condition holds then $\mathsf{E} X_r^{(i)} X_s^{(i)} I(\mathbf{X}^{(i)} \notin \mathscr{E}(R v_i)) \leqslant v_i \psi(R)$.

Condition 3. Bounded Trace Condition. There exists V_0 such that $v_i \leqslant V_0 \forall i$.

Given a function p, write ∇p for the gradient vector $(\partial p/\partial x_1, \ldots, \partial p/\partial x_n)$ and $\nabla^2 p$ for the Hessian matrix $(\nabla^2 p)_{ij} = \partial^2 p/\partial x_i\, \partial x_j$.

Definition 1.2. For a random vector $\mathbf{U}$ with differentiable density f and covariance matrix $B > 0$, define the score vector function $\rho_{\mathbf{U}}(\mathbf{x}) = \nabla \log f(\mathbf{x}) = \nabla f(\mathbf{x})/f(\mathbf{x})\, I(f(\mathbf{x}) > 0)$. Define the Fisher Information matrix J and its standardised version J_{st} by:

$$J(\mathbf{U}) = \mathsf{E}_{\mathbf{U}}(\rho_{\mathbf{U}}(\mathbf{U})\, \rho_{\mathbf{U}}(\mathbf{U})^{\mathsf{T}}),$$
$$J_{\mathrm{st}}(\mathbf{U}) = J(\mathbf{U}) - B^{-1}.$$

Since $\mathsf{E}(\mathbf{U} \rho_{\mathbf{U}}(\mathbf{U})^{\mathsf{T}}) = -I$, we know $J_{\mathrm{st}}(\mathbf{U}) = \mathsf{E}(\rho_{\mathbf{U}}(\mathbf{U}) + B^{-1}\mathbf{U})(\rho_{\mathbf{U}}(\mathbf{U}) + B^{-1}\mathbf{U})^{\mathsf{T}}$ is positive definite.

We use $\mathbf{Z}_B$ and $\mathbf{Z}_B^{(\cdot)}$ to represent multivariate normal, or $N(\mathbf{0}, B)$ random vectors with mean zero and covariance B; the corresponding probability density is denoted by ϕ_B. Given $\tau > 0$, set $\mathbf{Y}_\tau^{(S)} = \mathbf{U}^{(S)} + \mathbf{Z}_{C\tau}$, where $\mathbf{Z}_{C\tau}$ is $N(\mathbf{0}, C\tau)$, independent of $\mathbf{U}^{(S)}$.

Definition 1.3. Define $\kappa(W, \tau) = \sup_{S:\, v_S \geqslant W} \operatorname{tr}(C J_{\mathrm{st}}(\mathbf{Y}_\tau^{(S)}))$.

Condition 4. Integrability Condition. There exists V_1 such that $\int_0^\infty \kappa(V_1, \tau)\, d\tau$ is finite.

Note. This condition will hold if $\int_0^\varepsilon \operatorname{tr}(C J(\mathbf{Y}_\tau^{(S)}))\, d\tau$ is finite for some ε.

Definition 1.4. Given probability densities f and g, define the Kullback–Leibler distance (or relative entropy) to be:

$$D(f \,\|\, g) = \int f(\mathbf{x}) \log\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) d\mathbf{x}.$$

Kullback–Leibler distance is shift and scale invariant. Furthermore $D(f \,\|\, g) \geqslant 0$ with equality iff $f(\mathbf{x}) = g(\mathbf{x})$ for almost all $\mathbf{x}$. However, $D(f \,\|\, g)$ is not a metric; it does not satisfy the triangle inequality and is asymmetric. The principal theorem of the paper is the following:

Theorem 1.5. If the convergent covariance, uniform Lindeberg, bounded trace and integrability conditions listed above hold then:

$$\lim_{W \to \infty} \left(\sup_{S:\, v_S \geqslant W} D(g^{(S)} \,\|\, \phi_C) \right) = 0$$

Historically, the first attempt to use entropy theoretic methods to prove weak convergence for random vectors was made by Linnik.[11] However, as in the one-dimensional situation,[10] he does not prove convergence in D.

Theorem 1.5 is reminiscent of Theorem 2.2 of Carlen and Soffer,[3] which holds under stricter conditions. Firstly, in place of our Condition 1, they require the stronger condition that $C^{(S)}/v_S \to I/n$. Secondly, in place of our Condition 3, they need that the variances be bounded above and below (Eq. (2.12)). Our Condition 2 corresponds directly to their Eq. (2.11).

Our method of proof is significantly different from that of Carlen and Soffer. They use topological arguments developed in their Lemma 1.4 to show that a certain set of distributions is compact, and that the convolution map is continuous, and hence in Theorem 1.2 provide an existence proof of a positive lower bound on the rate of increase of entropy. In contrast, our Proposition 3.3 constructs an explicit lower bound on the rate of decrease of Fisher information.

An example where these conditions hold in the "nearly identical" case is where each of the $\mathbf{X}^{(i)}$ are of the form $\mathbf{W}^{(i)} + \mathbf{Z}^{(i)}_{\varepsilon C}$ for an IID sequence $\mathbf{Z}^{(i)}_{\varepsilon C}$. In this case, the covariance conditions on $\mathbf{X}^{(i)}$ to corresponding conditions on the covariance of $\mathbf{W}^{(i)}$ which can be easily verified. The uniform Lindeberg condition holds by an argument similar to that in Lemma 5.1 below. The Integrability Condition holds because $J(\mathbf{X}^{(i)}) \leqslant J(\mathbf{Z}^{(i)}_{\varepsilon C})$, which is finite.

In the IID case, matters are simpler, and we recover a natural extension to the vector case of Theorem 1 of ref. 1.

Theorem 1.6. In the case of IID random vectors with densities, finite covariance $C > 0$ and $v = \mathrm{tr}(C)$, denote the density of $(\sum_{i=1}^m \mathbf{X}^{(i)})/\sqrt{mv}$ by g_m. If the Kullback–Leibler distance $D(g_m \,||\, \phi_C)$ is ever finite then:

$$D(g_m \,||\, \phi_C) \to 0.$$

Kullback and Leibler[9] also introduced a symmetrized version of D, expressed as $\tilde{D}(f, \phi) = D(f \,||\, \phi) + D(\phi \,||\, f)$. Although this restores symmetry, it is unknown what conditions will guarantee its convergence, since for example unless $\mathrm{supp} f$ is the whole space, $\tilde{D}(f, \phi) = \infty$. In particular, we feel that the simplicity of Theorem 1.6 suggests that $D(f \,||\, \phi)$ is a natural measure of distance.

2. TECHNICAL BACKGROUND

We will use a generalisation of the well-known de Bruijn identity expressing Kullback–Leibler distance as an integral of Fisher informations, using the natural inner product space for random vectors and the Fisher information matrix.

Definition 2.1. For random vectors $\mathbf{U}$, $\mathbf{V}$, define the inner product $\langle \mathbf{U}, \mathbf{V} \rangle = \mathsf{E}(\mathbf{U}^{\mathsf{T}}\mathbf{V})$ and the norm $\langle \mathbf{U} \rangle^2 = \mathsf{E}(\mathbf{U}^{\mathsf{T}}\mathbf{U})$.

Note that for any matrix P, $\langle P\mathbf{U}, P\mathbf{V} \rangle = \mathsf{E}\,\mathrm{tr}(P^{\mathsf{T}}P\mathbf{V}\mathbf{U}^{\mathsf{T}})$.

Throughout this section, $\mathbf{X}$ stands for a random vector with density f and covariance matrix $B > 0$. Furthermore, given $\tau > 0$, $\mathbf{Y}_\tau = \mathbf{X} + \mathbf{Z}_{C\tau}$, where $\mathbf{Z}_{C\tau}$ is independent of $\mathbf{X}$ and $\mathbf{Y}_\tau$ has density f_τ.

Lemma 2.2. For $\mathbf{x} \in \mathcal{R}^n$:

$$2\frac{\partial f_\tau}{\partial \tau}(\mathbf{x}) = \sum_{i,j} C_{ij} \frac{\partial^2 f_\tau}{\partial x_i\, \partial x_j}(\mathbf{x}) = \mathrm{tr}(C(\nabla^2 f_\tau)(\mathbf{x})).$$

Proof. f_τ is twice continuously differentiable, so we know that $\nabla^2 f_\tau$ exists. Now:

$$\frac{\partial \phi_{C\tau}}{\partial z_i}(\mathbf{z}) = \left(- \frac{\sum_k C_{ik}^{-1} z_k}{\tau} \right) \phi_{C\tau}(\mathbf{z}), \qquad \mathbf{z} \in \mathcal{R}^n$$

and so:

$$\frac{\partial^2 \phi_{C\tau}}{\partial z_i\, \partial z_j}(\mathbf{z}) = \left(- \frac{C_{ij}^{-1}}{\tau} + \frac{(\sum_{k,l} C_{ik}^{-1} z_k C_{jl}^{-1} z_l)}{\tau^2} \right) \phi_{C\tau}(\mathbf{z}).$$

Hence, we deduce that

$$\mathrm{tr}(C\nabla^2 \phi_{C\tau}(\mathbf{z})) = \left(- \frac{n}{\tau} + \frac{\mathbf{z}^{\mathsf{T}} C^{-1} \mathbf{z}}{\tau^2} \right) \phi_{C\tau}(\mathbf{z}) = 2\frac{\partial \phi_{C\tau}}{\partial \tau}(\mathbf{z}),$$

and taking expectations provides the result.

Observe that:

$$D(f \,\|\, \phi_C) = \frac{1}{2}\log((2\pi e)^n \det C) - H(f) + \frac{\log e}{2}(\mathrm{tr}(C^{-1}B) - n)$$

$$= H(\phi_C) - H(f) + \frac{\log e}{2}(\mathrm{tr}(C^{-1}B) - n),$$

where $H(p) = -\int p(\mathbf{x}) \log p(\mathbf{x})\, d\mathbf{x}$ represents the differential entropy of density p.

Theorem 2.3 (*n*-dimensional de Bruijn's identity).

$$D(f \,\|\, \phi_C) = \frac{\log e}{2} \int_0^\infty \operatorname{tr}(C J_{\mathrm{st}}(\mathbf{Y}_\tau))\, d\tau + \frac{\log e}{2} (\operatorname{tr}(C^{-1}B) - n)$$
$$+ \frac{\log e}{2} \int_0^\infty \operatorname{tr}\left(C \left((B + C\tau)^{-1} - \frac{C^{-1}}{1+\tau} \right) \right) d\tau$$

Note that if $|B - C| \to 0$, the second and third terms tend to zero.

Proof. This is an integral form of Lemma 2.2. As $\tau \to \infty$, f_τ tends to a normal, so: $\lim_{\tau \to \infty} (H(f_\tau) - n \log \sqrt{1+\tau}) = 1/2 \log((2\pi e)^n \det C)$. As $\tau \to 0$, $f_\tau \to f$ in probability, so $H(f_\tau) \to H(f)$ by upper semi-continuity of H, where $H(f)$ may be $-\infty$. Hence if the integral is finite, by Lemma 2.2, for all $t > 0$, we can write $H(f)$ as:

$$H(f_\tau) - \int_0^t \frac{\partial H}{\partial \tau}(f_\tau)\, d\tau$$
$$= H(f_t) - \frac{1}{2} \int_0^t \sum_{i,j} C_{ij} \int (\nabla^2 f_\tau)_{ij} \log f_\tau(\mathbf{x})\, d\mathbf{x}\, d\tau$$
$$= H(f_t) - n \log \sqrt{1+t} - \frac{\log e}{2} \int_0^t \left(\sum_{i,j} C_{ij} \int \frac{\partial f_\tau}{\partial x_i} \frac{\partial f_\tau}{\partial x_j} \frac{1}{f_\tau}\, d\mathbf{x} - \frac{n}{1+\tau} \right) d\tau$$
$$= H(f_t) - n \log \sqrt{1+t} - \frac{\log e}{2} \int_0^t \operatorname{tr}\left(C \left(J(\mathbf{Y}_\tau) - \frac{C^{-1}}{1+\tau} \right) \right) d\tau.$$

We obtain the result by taking the limit as $t \to \infty$. If the integral is $-\infty$, then by Fatou $H(f) = -\infty$. Rearranging, we obtain the first form.

The second generalisation that we require is Hermite polynomials in n dimensions. Our exposition follows that of Dattoli *et al.*,[4] and extends their method from two dimensions in the obvious way. In what follows $\mathscr{Z}^n_+$ denotes the set of nonnegative integer vectors $\mathbf{m} = (m_1, m_2, \ldots, m_n)$ and $|\mathbf{m}| = \sum_{i=1}^n m_i$. We fix a matrix B throughout the rest of this section.

Definition 2.4. Define Hermite polynomials $G_{\mathbf{m}}$ and $H_{\mathbf{m}}$, $\mathbf{m} \in \mathscr{Z}^n_+$, with respect to multivariate normal weight ϕ_B, via the generating functions $G(\mathbf{x}, \mathbf{t})$ and $H(\mathbf{x}, \mathbf{t})$, where $\mathbf{x}, \mathbf{t} \in \mathscr{R}^n$:

1. $$G(\mathbf{x}, \mathbf{t}) = \sum_{\mathbf{m} \in \mathscr{Z}_+^n} \frac{t_1^{m_1} \cdots t_n^{m_n}}{m_1! \cdots m_n!} G_{\mathbf{m}}(\mathbf{x}) = \exp(\mathbf{t}^\top B^{-1}\mathbf{x} - \mathbf{t}^\top B^{-1}\mathbf{t}/2),$$

2. $$H(\mathbf{y}, \mathbf{s}) = \sum_{\mathbf{l} \in \mathscr{Z}_+^n} \frac{s_1^{l_1} \cdots s_n^{l_n}}{l_1! \cdots l_n!} H_{\mathbf{l}}(\mathbf{y}) = \exp(\mathbf{s}^\top \mathbf{y} - \mathbf{s}^\top B\mathbf{s}/2).$$

Write (a) $\mathbf{0} = (0,\ldots, 0)$ (b) $\mathbf{e}_i = (0,\ldots, 1,\ldots, 0)$ (1 at the ith position) (c) $\mathbf{e}_{ij} = (0,\ldots, 1,\ldots, 1,\ldots, 0)$ (1 at the ith and jth positions, $i \neq j$). Then $G_{\mathbf{0}}(\mathbf{x}) = H_{\mathbf{0}}(\mathbf{y}) = 1$, $G_{\mathbf{e}_i}(\mathbf{x}) = (B^{-1}\mathbf{x})_i$, $H_{\mathbf{e}_i}(\mathbf{y}) = y_i$, $G_{\mathbf{e}_{ij}}(\mathbf{x}) = \sum_{k,l} B_{ik}^{-1} B_{jl}^{-1} x_k x_l - B_{ij}^{-1}$ and $H_{\mathbf{e}_{ij}}(\mathbf{y}) = y_i y_j - B_{ij}$. Observe that in the one-dimensional case with $B = (\sigma^2)$, $H_m(x)$ have a standard form (see, for example, Szegő,[13] pp. 104–109), and $G_m(x) = H_m(x)/\sigma^{2m}$.

Each set $\{G_{\mathbf{m}}\}$ and $\{H_{\mathbf{m}}\}$ spans the space $L^2(\phi_B(\mathbf{x})\, d\mathbf{x})$, although neither set is orthogonal. We need a formula for $\langle G_{\mathbf{m}}, H_{\mathbf{l}} \rangle$:

Proposition 2.5.

$$\langle G_{\mathbf{m}}, H_{\mathbf{l}} \rangle = \int G_{\mathbf{m}}(\mathbf{x})\, H_{\mathbf{l}}(\mathbf{x})\, \phi_B(\mathbf{x})\, d\mathbf{x} = \delta_{\mathbf{ml}} m_1!\, m_2! \cdots m_n!$$

Proof. For all $\mathbf{s}, \mathbf{t} \in \mathscr{R}^n$:

$$\sum_{\mathbf{m},\mathbf{l}} \frac{t_1^{m_1} \cdots t_n^{m_n}}{m_1! \cdots m_n!} \frac{s_1^{l_1} \cdots s_n^{l_n}}{l_1! \cdots l_n!} G_{\mathbf{m}}(\mathbf{x})\, H_{\mathbf{l}}(\mathbf{y})$$
$$= \exp(\mathbf{t}^\top B^{-1}\mathbf{x} - \mathbf{t}^\top B^{-1}\mathbf{t}/2 + \mathbf{s}^\top \mathbf{y} - \mathbf{s}^\top B\mathbf{s}/2).$$

Taking $\mathbf{y} = \mathbf{x}$, multiplying by $\phi_B(\mathbf{x})$ and integrating, the left hand side becomes:

$$\sum_{\mathbf{m},\mathbf{l}} \left(\frac{t_1^{m_1} \cdots t_n^{m_n}}{m_1! \cdots m_n!}\right)\left(\frac{s_1^{l_1} \cdots s_n^{l_n}}{l_1! \cdots l_n!}\right) \int G_{\mathbf{m}}(\mathbf{x})\, H_{\mathbf{l}}(\mathbf{x})\, \phi_B(\mathbf{x})\, d\mathbf{x}$$
$$= \exp(-\mathbf{s}^\top B\mathbf{s}/2 - \mathbf{t}^\top B^{-1}\mathbf{t}/2)\, \mathsf{E}[\exp((\mathbf{s} + B^{-1}\mathbf{t})^\top\, \mathbf{Z}_B)]$$
$$= \exp(-\mathbf{s}^\top B\mathbf{s}/2 - \mathbf{t}^\top B^{-1}\mathbf{t}/2)\, \exp((\mathbf{s} + B^{-1}\mathbf{t})^\top\, B(\mathbf{s} + B^{-1}\mathbf{t})/2)$$
$$= \exp(\mathbf{s}^\top \mathbf{t}),$$

so comparing coefficients we deduce the result.

Definition 2.6. Let $H_{\mathbf{r}}$ and $G_{\mathbf{r}}$ be Hermite polynomials with respect to ϕ_B. Given a function $u = \sum_{\mathbf{r}} a_{\mathbf{r}} G_{\mathbf{r}} = \sum_{\mathbf{s}} b_{\mathbf{s}} H_{\mathbf{s}} \in L^2(\phi_B(\mathbf{x})\, d\mathbf{x})$, define the projection map $u \mapsto \Theta u$ by:

$$(\Theta u)(\mathbf{x}) = \sum_{\mathbf{r}:\, |\mathbf{r}| \geqslant 2} a_{\mathbf{r}} G_{\mathbf{r}}(\mathbf{x}) = \sum_{\mathbf{r}:\, |\mathbf{r}| \geqslant 2} b_{\mathbf{r}} H_{\mathbf{r}}(\mathbf{x})$$

$$= u(\mathbf{x}) - \int u(\mathbf{y})(1 + \mathbf{y} B^{-1} \mathbf{x})\, \phi_B(\mathbf{y})\, d\mathbf{y},$$

so that $\int (\Theta u)^2\, (\mathbf{x})\, \phi_B(\mathbf{x}) = \sum_{\mathbf{r}:\, |\mathbf{r}| \geqslant 2} a_{\mathbf{r}} b_{\mathbf{r}} r_1! \cdots r_n!$.

Define a seminorm $\|u\|_\Theta$ using:

$$\|u\|_\Theta^2 = \sum_i \int (\Theta u_i)^2\, (\mathbf{x})\, \phi_B(\mathbf{x})\, d\mathbf{x}.$$

Note that if $\mathbf{l} = \int \mathbf{u}(\mathbf{y})\, \mathbf{y} \phi_{\mathbf{B}}(\mathbf{y})\, \mathbf{dy}$, then $a_{\mathbf{e}_i} = l_i$, $b_{\mathbf{e}_i} = (B^{-1}\mathbf{l})_i$.

3. SANDWICH INEQUALITY

First we need a technical lemma:

Lemma 3.1. For $\tau, K > 0$, there exists a constant $\xi_{\tau, K} > 0$ such that for any random vector $\mathbf{X}$ with covariance matrix B where $\operatorname{tr}(C^{-1}B) \leqslant K$, the sum $\mathbf{X} + \mathbf{Z}_{C\tau}$, where $\mathbf{Z}_{C\tau}$ is independent of $\mathbf{X}$, has density f bounded below by $\xi_{\tau, K} \phi_{C\tau/2}$.

Proof. By Definition 1.1 and Chebyshev: $\int I(\mathbf{x} \in \mathscr{E}(2K))\, dF_{\mathbf{U}}(\mathbf{x}) \geqslant 1 - K/2K = 1/2$. Hence if $F_{\mathbf{U}}$ is the distribution function of $\mathbf{U}$, for any $\mathbf{y} \in \mathscr{R}^n$:

$$f_\tau(\mathbf{y}) = \int \phi_{C\tau}(\mathbf{y} - \mathbf{x})\, dF_{\mathbf{U}}(\mathbf{x})$$

$$\geqslant \min\{\phi_{C\tau}(\mathbf{y} - \mathbf{x}) : \mathbf{x} \in \mathscr{E}(2K)\}/2$$

$$= \frac{\phi_{C\tau/2}(\mathbf{y})}{2^{(n/2+1)}} \exp\left(\min_{\mathbf{x} \in \mathscr{E}(2K)} \left\{ \frac{(\mathbf{x}+\mathbf{y})^{\mathsf{T}}\, C^{-1}\, (\mathbf{x}+\mathbf{y}) - 2\mathbf{x}^{\mathsf{T}} C^{-1} \mathbf{x}}{2\tau} \right\} \right)$$

$$\geqslant \frac{\phi_{C\tau/2}(\mathbf{y})}{2^{(n/2+1)}} \exp\left(\min_{\mathbf{x} \in \mathscr{E}(2K)} \left\{ -\frac{\mathbf{x}^{\mathsf{T}} C^{-1} \mathbf{x}}{\tau} \right\} \right)$$

$$\geqslant 2^{-(n/2+1)} \exp(-2K/\tau)\, \phi_{C\tau/2}(\mathbf{y}) = \xi_{\tau, K} \phi_{C\tau/2}(\mathbf{y}).$$

Note that $\xi_{\tau, K} = 2^{-(n/2+1)} \exp(-2K/\tau)$ depends only on K and τ.

Lemma 3.2. Let $\mathbf{V}^{(1)}$, $\mathbf{V}^{(2)}$ be independent random vectors with densities p and q and let $\mathbf{V}^{(3)} = \mathbf{V}^{(1)} + \mathbf{V}^{(2)}$, with score functions $\rho^{(j)} = \rho_{\mathbf{V}^{(j)}}$, $i = 1, 2, 3$. With probability one, for any $\beta \in [0, 1]$:

$$\rho^{(3)} = \mathsf{E}[\beta \rho^{(1)}(\mathbf{V}^{(1)}) + (1-\beta)\, \rho^{(2)}(\mathbf{V}^{(2)}) \mid \mathbf{V}^{(3)}].$$

Here and below, $\mathsf{E}[\,\cdot \mid \mathbf{V}^{(3)}]$ stands for the conditional expectation with respect to the σ-algebra generated by $\mathbf{V}^{(3)}$.

Proof. Since $\mathbf{V}^{(3)}$ has the density $r(\mathbf{w}) = \int p(\mathbf{x})\, q(\mathbf{w}-\mathbf{x})\, d\mathbf{x}$, for all $i = 1,\ldots, n$:

$$(\rho^{(3)}(\mathbf{w}))_i = \frac{1}{r(\mathbf{w})} \int \frac{\partial p}{\partial x_i}(\mathbf{x})\, q(\mathbf{w}-\mathbf{x})\, d\mathbf{x} = \int (\rho_{\mathrm{U}}(\mathbf{x}))_i \frac{p(\mathbf{x})\, q(\mathbf{w}-\mathbf{x})}{r(\mathbf{w})}\, d\mathbf{x},$$

which equals, almost surely, the expected value $\mathsf{E}[\rho^{(1)}(\mathbf{V}^{(1)})_i \mid \mathbf{V}^{(3)}](\mathbf{w})$. Similarly, we can produce an expression in terms of the score function $\rho^{(2)}$. Now add β times the first expression to $1-\beta$ times the second one.

Until the end of this section $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ denote independent random vectors with mean zero and covariances $B^{(1)}$, $B^{(2)} > 0$. Given $\tau > 0$, we set $\mathbf{Y}_\tau^{(j)} = \mathbf{U}^{(j)} + \mathbf{Z}_{C\tau}^{(j)}$, $j = 1, 2$, where $\mathbf{Z}_{C\tau}^{(j)}$ are independent of the $\mathbf{U}^{(j)}$ and of each other. Furthermore, for $0 < \alpha < 1$, define $\mathbf{Y}_\tau^{(3)} = \sqrt{\alpha}\, \mathbf{Y}_\tau^{(1)} + \sqrt{1-\alpha}\, \mathbf{Y}_\tau^{(2)}$, and denote by $f^{(j)}$ the density and $\rho^{(j)}$ the score function of $\mathbf{Y}_\tau^{(j)}$, $j = 1, 2, 3$. Finally let $\tilde{\mathbf{Z}}^{(1)}$ and $\tilde{\mathbf{Z}}^{(2)}$ be independent $N(\mathbf{0}, C\tau/2)$ and $\tilde{\mathbf{Z}}^{(3)} = \sqrt{\alpha}\, \tilde{\mathbf{Z}}^{(1)} + \sqrt{1-\alpha}\, \tilde{\mathbf{Z}}^{(2)}$.

The main proposition of this section is the following:

Proposition 3.3 (Sandwich Inequality). If $\mathrm{tr}(C^{-1}B^{(i)}) \leqslant K$ then there exists a constant $\xi_{\tau, K}$ such that for any matrix P:

$$\begin{aligned}
&\alpha\, \mathrm{tr}(P^\top P J(\mathbf{Y}_\tau^{(1)})) + (1-\alpha)\, \mathrm{tr}(P^\top P J(\mathbf{Y}_\tau^{(2)})) - \mathrm{tr}(P^\top P J(\mathbf{Y}_\tau^{(3)})) \\
&\quad = \alpha \langle P\rho^{(1)}(\mathbf{Y}_\tau^{(1)})\rangle^2 + (1-\alpha)\langle P\rho^{(2)}(\mathbf{Y}_\tau^{(2)})\rangle^2 - \langle P\rho^{(3)}(\mathbf{Y}_\tau^{(3)})\rangle^2 \\
&\quad \geqslant (\|P\rho^{(1)}\|_\Theta^2 + \|P\rho^{(2)}\|_\Theta^2) \left(\frac{\xi_{\tau, K}^2\, \alpha(1-\alpha)}{2} \right).
\end{aligned}$$

The proof uses a series of lemmas.

Lemma 3.4. For any matrix P:

$$\alpha\langle P\rho^{(1)}(\mathbf{Y}_\tau^{(1)})\rangle^2+(1-\alpha)\langle P\rho^{(2)}(\mathbf{Y}_\tau^{(2)})\rangle^2-\langle P\rho^{(3)}(\mathbf{Y}_\tau^{(3)})\rangle^2$$
$$\geqslant \xi_{\tau,K}^2\langle\sqrt{\alpha}\,P\rho^{(1)}(\tilde{\mathbf{Z}}^{(1)})+\sqrt{1-\alpha}\,P\rho^{(2)}(\tilde{\mathbf{Z}}^{(2)})\rangle^2-\xi_{\tau,K}^2\langle\mathbf{w}(\tilde{\mathbf{Z}}^{(3)})\rangle^2$$

where $\xi_{\tau,K}^2$ is the constant from Lemma 3.1 and:

$$\mathbf{w}(\mathbf{x})=\mathsf{E}[\sqrt{\alpha}\,P\rho^{(1)}(\tilde{\mathbf{Z}}^{(1)})+\sqrt{1-\alpha}\,P\rho^{(2)}(\tilde{\mathbf{Z}}^{(2)})\mid\tilde{\mathbf{Z}}^{(3)}=\mathbf{x}]$$
$$=\sqrt{\alpha}\int P\rho^{(1)}(\sqrt{\alpha}\,\mathbf{x}+\sqrt{1-\alpha}\,\mathbf{v})\,\phi_{C\tau/2}(\mathbf{v})\,d\mathbf{v}$$
$$+\sqrt{1-\alpha}\int P\rho^{(2)}(\sqrt{1-\alpha}\,\mathbf{x}-\sqrt{\alpha}\,\mathbf{v})\,\phi_{C\tau/2}(\mathbf{v})\,d\mathbf{v}.$$

Proof. Taking $\mathbf{V}^{(1)}=\sqrt{\alpha}\,\mathbf{Y}_\tau^{(1)}$, $\mathbf{V}^{(2)}=\sqrt{1-\alpha}\,\mathbf{Y}_\tau^{(2)}$ in Lemma 3.2 gives $\mathbf{Y}_\tau^{(3)}=\mathbf{V}^{(3)}$. Hence:

$$\alpha\langle P\rho^{(1)}(\mathbf{Y}_\tau^{(1)})\rangle^2+(1-\alpha)\langle P\rho^{(2)}(\mathbf{Y}_\tau^{(2)})\rangle^2-\langle P\rho^{(3)}(\mathbf{Y}_\tau^{(3)})\rangle^2$$
$$=\langle\sqrt{\alpha}\,P\rho^{(1)}(\mathbf{Y}_\tau^{(1)})+\sqrt{1-\alpha}\,P\rho^{(2)}(\mathbf{Y}_\tau^{(2)})-P\rho^{(3)}(\mathbf{Y}_\tau^{(3)})\rangle^2$$
$$\geqslant\xi_{\tau,K}^2\langle\sqrt{\alpha}\,P\rho^{(1)}(\tilde{\mathbf{Z}}^{(1)})+\sqrt{1-\alpha}\,P\rho^{(2)}(\tilde{\mathbf{Z}}^{(2)})-P\rho^{(3)}(\tilde{\mathbf{Z}}^{(3)})\rangle^2,$$

which is minimised when $P\rho^{(3)}(\tilde{\mathbf{Z}}^{(3)})$ is replaced by the orthogonal projection $w(\tilde{\mathbf{Z}}^{(3)})$.

Now rearranging, the conditional density $p^{(1)}(\mathbf{s}\mid\mathbf{x})$ of $\tilde{\mathbf{Z}}^{(1)}$ given $(\tilde{\mathbf{Z}}^{(3)}=x)$ is $((\pi\tau)^n\det C)^{-1/2}\exp(-(\mathbf{s}^\top C^{-1}\mathbf{s}+\mathbf{u}^\top C^{-1}\mathbf{u}-\mathbf{x}^\top C^{-1}\mathbf{x})/\tau)$, where $\sqrt{\alpha}\,\mathbf{s}+\sqrt{1-\alpha}\,\mathbf{u}=\mathbf{x}$. This equals $\phi_{C\tau/2}(\mathbf{v})$, where $\mathbf{v}=(\mathbf{s}-\sqrt{\alpha}\,\mathbf{x})/\sqrt{1-\alpha}$. Similarly the conditional density $p^{(2)}(\mathbf{s}\mid\mathbf{x})$ of $\tilde{\mathbf{Z}}^{(2)}$ given that $\tilde{\mathbf{Z}}^{(3)}=x$ is $\phi_{C\tau/2}(\mathbf{v})$, where $\mathbf{v}=(\mathbf{r}-\sqrt{1-\alpha}\,\mathbf{x})/\sqrt{\alpha}$. Substituting for $\mathbf{r}$, $\mathbf{s}$, the result follows.

Lemma 3.5. If $G_\mathbf{r}$ and $H_\mathbf{r}$ are the Hermite polynomials with respect to ϕ_B, then the linear map Λ_β, $0<\beta<1$, defined by:

$$\Lambda_\beta u(\mathbf{x})=\int u(\sqrt{\beta}\,\mathbf{x}+\sqrt{1-\beta}\,\mathbf{v})\,\phi_B(\mathbf{v})\,d\mathbf{v},\quad u\in L^2(\phi_B(\mathbf{x})\,d\mathbf{x})$$

takes $G_\mathbf{r}$ into $(\sqrt{\beta})^{|\mathbf{r}|}\,G_\mathbf{r}$ and $H_\mathbf{r}$ into $(\sqrt{\beta})^{|\mathbf{r}|}\,H_\mathbf{r}$.

Proof. Consider the action of Λ_β on the generating functions $G(\mathbf{x}, \mathbf{a})$ and $H(\mathbf{x}, \mathbf{a})$. For given $\mathbf{a}, \mathbf{x} \in \mathscr{R}^n$, $\Lambda_\beta G(\mathbf{x}, \mathbf{a})$ equals:

$$\int \exp\left(-\frac{\mathbf{a}^\top B^{-1}\mathbf{a}}{2} + \mathbf{a}^\top B^{-1}(\sqrt{\beta}\,\mathbf{x} + \sqrt{1-\beta}\,\mathbf{v})\right)\phi_B(\mathbf{v})\,d\mathbf{v}$$

$$= \int \phi_B(\mathbf{v} - \sqrt{1-\beta}\,\mathbf{a})\exp(\sqrt{\beta}\,\mathbf{a}^\top B^{-1}\mathbf{x} - \beta\mathbf{a}^\top B^{-1}\mathbf{a}/2)\,d\mathbf{v}$$

$$= G(\mathbf{x}, \sqrt{\beta}\,\mathbf{a}).$$

Similarly Λ_β: $H(\mathbf{x}, \mathbf{a}) \to H(\mathbf{x}, \sqrt{\beta}\,\mathbf{a})$.

An immediate consequence of Lemma 3.5 is that given a matrix P, the vector-function $\mathbf{w}$ defined in Lemma 3.4 has the form $\Lambda_\alpha(P\rho^{(1)}) + \lambda_{1-\alpha}(P\rho^{(2)})$:

$$\mathbf{w}(\mathbf{x}) = \sum_{\mathbf{r}} [(\sqrt{\alpha})^{|\mathbf{r}|+1} a_{\mathbf{r}}^{(1)} + (\sqrt{1-\alpha})^{|\mathbf{r}|+1} a_{\mathbf{r}}^{(2)}]\, G_{\mathbf{r}}(\mathbf{x})$$

$$= \sum_{\mathbf{r}} [(\sqrt{\alpha})^{|\mathbf{r}|+1} b_{\mathbf{r}}^{(1)} + (\sqrt{1-\alpha})^{|\mathbf{r}|+1} b_{\mathbf{r}}^{(2)}]\, H_{\mathbf{r}}(\mathbf{x}), \qquad \mathbf{x} \in \mathscr{R}^n,$$

where $P\rho^{(j)} = \sum_{\mathbf{r}} a_{\mathbf{r}}^{(j)} G_{\mathbf{r}}(\mathbf{x}) = \sum_{\mathbf{r}} b_{\mathbf{r}}^{(j)} H_{\mathbf{r}}(\mathbf{x})$, $j = 1, 2$.

Lemma 3.6. If for $j = 1, 2$, $(P\rho^{(j)})(\mathbf{x}) = \sum a_{\mathbf{r}}^{(j)} G_{\mathbf{r}}(\mathbf{x}) = \sum b_{\mathbf{r}}^{(j)} H_{\mathbf{r}}(\mathbf{x})$ then:

$$\langle \sqrt{\alpha}\, P\rho^{(1)}(\tilde{\mathbf{Z}}^{(1)}) + \sqrt{1-\alpha}\, P\rho^{(2)}(\tilde{\mathbf{Z}}^{(2)})\rangle^2 - \langle \mathbf{w}(\tilde{\mathbf{Z}}^{(3)})\rangle^2$$

$$\geqslant (\|P\rho^{(1)}\|_\Theta^2 + \|P\rho^{(2)}\|_\Theta^2)\,\frac{\alpha(1-\alpha)}{2}.$$

Proof. Defining $\mathbf{w}' = (\alpha(1-\alpha))^{1/4}\,\Lambda_{\sqrt{\alpha(1-\alpha)}}\,(P\rho^{(1)} - P\rho^{(2)})$, we have:

$$\mathbf{w}'(\mathbf{x}) = \sum_{\mathbf{r}} (\alpha(1-\alpha))^{(|\mathbf{r}|+1)/4}\,(a_{\mathbf{r}}^{(1)} - a_{\mathbf{r}}^{(2)})\, G_{\mathbf{r}}(\mathbf{x})$$

$$= \sum_{\mathbf{r}} (\alpha(1-\alpha))^{|\mathbf{r}|+1)/4}\,(b_{\mathbf{r}}^{(1)} - b_{\mathbf{r}}^{(2)})\, H_{\mathbf{r}}(\mathbf{x})$$

and

$$\langle \mathbf{w}(\tilde{\mathbf{Z}}^{(3)})\rangle^2 \leqslant \langle \mathbf{w}(\tilde{\mathbf{Z}}^{(3)})\rangle^2 + \langle \mathbf{w}'(\tilde{\mathbf{Z}}^{(3)})\rangle^2$$

$$= \sum_{\mathbf{r}} r_1! \cdots r_n!\, [a_{\mathbf{r}}^{(1)} b_{\mathbf{r}}^{(1)}(\alpha^{r+1} + \alpha^{(r+1)/2}(1-\alpha)^{(r+1)/2})$$

$$+ a_{\mathbf{r}}^{(2)} b_{\mathbf{r}}^{(2)}((1-\alpha)^{r+1} + \alpha^{(r+1)/2}(1-\alpha)^{(r+1)/2})].$$

Expanding, and using the normalisation of the $H_{\mathbf{r}}$ and $G_{\mathbf{r}}$, we deduce that:

$$\begin{aligned}
&\langle \sqrt{\alpha}\, P\rho^{(1)}(\tilde{\mathbf{Z}}^{(1)}) + \sqrt{1-\alpha}\, P\rho^{(2)}(\tilde{\mathbf{Z}}^{(2)})\rangle^2 - \langle \mathbf{w}(\tilde{\mathbf{Z}}^{(3)})\rangle^2 \\
&\quad = \sum_{\mathbf{r}} [\alpha a_{\mathbf{r}}^{(1)} b_{\mathbf{r}}^{(1)} + (1-\alpha)\, a_{\mathbf{r}}^{(2)} b_{\mathbf{r}}^{(2)}]\, r_1! \cdots r_n! - \langle \mathbf{w}(\tilde{\mathbf{Z}}^{(3)})\rangle^2 \\
&\quad \geqslant \sum_{\mathbf{r}\neq \mathbf{0}} [A_{|\mathbf{r}|}(\alpha)\, a_{\mathbf{r}}^{(1)} b_{\mathbf{r}}^{(1)} + A_{|\mathbf{r}|}(1-\alpha)\, a_{\mathbf{r}}^{(2)} b_{\mathbf{r}}^{(2)}]\, r_1! \cdots r_n!,
\end{aligned}$$

where $A_r(x) = x - x^{r+1} - x^{(r+1)/2}(1-x)^{(r+1)/2}$. As $A_1(x) \equiv 0$, terms involving A_1 may be removed. For fixed $x \in [0, 1]$, $A_r(x)$ is increasing in r, so we may replace $A_r(\alpha)$, $A_r(1-\alpha)$ by the (positive) values $A_2(\alpha)$, $A_2(1-\alpha)$. Finally, note that for $x \in [0, 1]$, $A_2(x) = x - x^3 - x^{3/2}(1-x)^{3/2} = x(1-x)(1+x-x^{1/2}(1-x)^{1/2}) \geqslant x(1-x)/2$, which completes the proof.

Proposition 3.3 follows on substituting Lemma 3.6 into Lemma 3.4.

4. STABILITY RESULT FOR CRAMÉR–RAO LOWER BOUND

Definition 4.1. For a function $\psi: \mathscr{R}_+ \mapsto \mathscr{R}_+$, with $\psi(R) \to 0$ as $R \to \infty$, define the set of random vectors $\mathscr{C}_\psi$:

$$\mathscr{C}_\psi = \left\{ \begin{aligned} &\mathbf{X}: \mathsf{E}\mathbf{X} = \mathbf{0},\ v_{\mathbf{X}} = \operatorname{tr}(\mathsf{E}\mathbf{X}^{\mathsf{T}}\mathbf{X}) < \infty, \\ &\mathsf{E}X_j^2 I(\mathbf{X} \notin \mathscr{E}(Rv_{\mathbf{X}})) \leqslant v_{\mathbf{X}}\psi(R) \ \text{ for all } j, R \end{aligned} \right\}.$$

$\mathscr{C}_\psi$ is a scale-invariant class, and the uniform Lindeberg condition guarantees that each $\mathbf{X}^{(i)} \in \mathscr{C}_\psi$. Lemma 5.1 below ensures that there exists a fixed positive function Ψ with $\Psi(R) \to 0$ as $R \to \infty$ such that for all S, $\mathbf{X}^{(S)} \in \mathscr{C}_\Psi$.

Throughout this section, $\mathbf{X}$ will represent a random vector with mean $\mathbf{0}$ and covariance $B^{(\mathbf{X})}$, with $v_{\mathbf{X}} = \operatorname{tr}(B^{(\mathbf{X})})$. We fix $\tau > 0$ and define random vector $\mathbf{Y}_\tau = \mathbf{X}/\sqrt{v_{\mathbf{X}}} + \mathbf{Z}_{C\tau}$, where $\mathbf{Z}_{C\tau}$ is independent of $\mathbf{X}$. We write $B^{(\mathbf{Y}_\tau)}$, $f_{\mathbf{Y}_\tau}$ and $\rho_{\mathbf{Y}_\tau}$ for the covariance, density and score function of $\mathbf{Y}_\tau$; sometimes we will use a simplified notation B_τ, f_τ and ρ_τ. Observe that $B^{(\mathbf{Y}_\tau)} = B^{(\mathbf{X})}/v_{\mathbf{x}} + C\tau$. As before, $G_{\mathbf{m}}$, $H_{\mathbf{m}}$ are the Hermite polynomials with respect to $\phi_{C\tau/2}$.

Proposition 4.2. Fix a positive function ψ with $\lim_{R\to\infty} \psi(R) = 0$, and as above fix $\tau > 0$. Given a random vector $\mathbf{X} \in \mathscr{C}_\psi$, for any invertible matrix P there exists a function $\nu_P(\varepsilon)$ (depending only on P), with $\nu_P(\varepsilon) \to 0$ monotonically as $\varepsilon \to 0$, such that

$$\operatorname{tr}(P^{\mathsf{T}} P J_{\mathrm{st}}(\mathbf{Y}_\tau)) \leqslant \nu_P(\|P\rho_{\mathbf{Y}_\tau}\|_\Theta).$$

The proof is based on Lemmas 4.3 to 4.5.

Lemma 4.3. Consider density $\tilde{f}$ with score function $\tilde{\rho} \in L^2(\phi_{C\tau/2}(\mathbf{x})\, d\mathbf{x})$. There exists $\tilde{\kappa}(\mathbf{x}) = \tilde{f}(\mathbf{0}) \exp(\sum_i u_i x_i + 1/2 \sum_{i,j} V_{ij} x_i x_j)$, such that for any $t > 0$:

$$\lim_{\zeta \to 0+} [\sup_{\mathbf{x} \in \mathscr{E}(t)} |\tilde{f}(\mathbf{x})/\tilde{\kappa}(\mathbf{x}) - 1| : \|\tilde{\rho}\|_{\Theta} \leqslant \zeta] = 0.$$

Coefficients u_i and V_{ij} relate to the Hermite expansion $\tilde{\rho} = \sum_{\mathbf{m} \in \mathscr{Z}_+^n} \mathbf{a}_{\mathbf{m}} G_{\mathbf{m}}$ as follows: vector $(u_1 \cdots u_n) = \mathbf{a}_{\mathbf{0}}$, and matrix $V = (V_{ij}) = AC^{-1}$, where $A = (\mathbf{a}_{\mathbf{e}_1} | \cdots | \mathbf{a}_{\mathbf{e}_n})$ is the matrix with ith column equal to the vector $\mathbf{a}_{\mathbf{e}_i}$.

Proof. We write $E(t)$ for the volume of $\mathscr{E}(t)$. If $\mathbf{x} \in \mathscr{E}(t)$ then $\phi_{C\tau/2}(\mathbf{x}) \geqslant \lambda(t, \tau/2)$, where $\lambda(t, \tau/2) = ((\pi\tau)^n \det C)^{-1/2} \exp(-t/\tau) > 0$. Hence, for any $t > 0$ and vector-function h: $\mathscr{R}^n \to \mathscr{R}^n$:

$$\begin{aligned}
&\sum_k \int |h_k(\mathbf{x})|\, I(\mathbf{x} \in \mathscr{E}(t))\, d\mathbf{x} \\
&\leqslant \left(n \int I(\mathbf{x} \in \mathscr{E}(t))\, d\mathbf{x} \right)^{1/2} \left(\int \sum_k h_k^2(\mathbf{x})\, I(\mathbf{x} \in \mathscr{E}(t))\, d\mathbf{x} \right)^{1/2} \\
&\leqslant \left(\frac{nE(t)}{\lambda(t, \tau/2)} \right)^{1/2} \left(\sum_k \int h_k^2(\mathbf{x})\, I(\mathbf{x} \in \mathscr{E}(t))\, \phi_{C\tau/2}(\mathbf{x})\, d\mathbf{x} \right)^{1/2} \\
&\leqslant \left(\frac{nE(t)}{\lambda(t, \tau/2)} \right)^{1/2} \left(\sum_k \int h_k^2(\mathbf{x})\, \phi_{C\tau/2}(\mathbf{x})\, d\mathbf{x} \right)^{1/2}.
\end{aligned}$$

Using the above expansion $\tilde{\rho} = \sum_{\mathbf{m} \in \mathscr{Z}_+^n} \mathbf{a}_{\mathbf{m}} G_{\mathbf{m}}$, since $G_{\mathbf{e}_i}(\mathbf{x}) = (C^{-1}\mathbf{x})_i$, the best linear approximation to $\tilde{\rho}$ becomes $\mathbf{a}_{\mathbf{0}} + (AC^{-1})\, \mathbf{x}$, with $A = (\mathbf{a}_{\mathbf{e}_1} | \cdots | \mathbf{a}_{\mathbf{e}_n})$. Thus taking $h(\mathbf{x}) = (\Theta\tilde{\rho})(\mathbf{x}) = \tilde{\rho}(\mathbf{x}) - \mathbf{a}_{\mathbf{0}} - (AC^{-1})\, \mathbf{x}$, we know that:

$$\sum_k \int |(\Theta\tilde{\rho})(\mathbf{x})|_k\, I(\mathbf{x} \in \mathscr{E}(t))\, d\mathbf{x} \leqslant \|\rho\|_{\Theta} \left(\frac{nE(t)}{\lambda(t, \tau/2)} \right)^{1/2}.$$

Now, since $\tilde{f}$ is continuously differentiable, $(\Theta\tilde{\rho})$ is continuous, so for any $\mathbf{y} \in \mathscr{E}(t)$, if $|(\Theta\tilde{\rho})(\mathbf{y}) - (\Theta\tilde{\rho})(\mathbf{0})| > \varepsilon$, then considering a thin tube L around the line segment connecting $\mathbf{0}$ and $\mathbf{y}$, for some ε':

$$\varepsilon' \leqslant \sum_k \int |(\Theta\tilde{\rho})_k\, (\mathbf{x})|\, I(\mathbf{x} \in L)\, d\mathbf{x} \leqslant \sum_k \int |(\Theta\tilde{\rho})_k\, (\mathbf{x})|\, I(\mathbf{x} \in \mathscr{E}(t))\, d\mathbf{x}.$$

Hence as $\|\tilde{\rho}\|_{\Theta} \to 0$, we have control uniformly in $\mathbf{y} \in \mathscr{E}(t)$ over:

$$[\log \tilde{f}(\mathbf{x}) - \mathbf{x}^{\mathsf{T}} AC^{-1}\mathbf{x}/2 - \mathbf{x}^{\mathsf{T}} \mathbf{a}_{\mathbf{0}}]_{\mathbf{0}}^{\mathbf{y}}.$$

To complete the proof, observe that $|g(\mathbf{x})-g(\mathbf{0})|<c$ implies $|\exp(g(\mathbf{x})-g(\mathbf{0}))-1| \leqslant \exp c-1$, so we take κ as suggested, with $(u_1,\ldots,u_n)=\mathbf{a}_0$ and $V=AC^{-1}$.

Lemma 4.4. Random vector $\mathbf{Y}_\tau$ has $\rho_{\mathbf{Y}_\tau} \in L^2(\phi_{C\tau/2}(\mathbf{x})\,d\mathbf{x})$ and hence $\rho_{\mathbf{Y}_\tau}=\sum_{\mathbf{m}} \mathbf{a}_{\mathbf{m}} G_{\mathbf{m}}$. Let $\kappa(\mathbf{x})=k\exp(\sum_i u_i x_i+1/2\sum_{i,j} V_{ij}x_i x_j)$ where u_i and V_{ij} are as in Lemma 4.3. There exists a constant z_τ (depending only on τ) such that if $\sup_{x\in\mathscr{E}(z_\tau)}|f_\tau(\mathbf{x})/\kappa(\mathbf{x})-1| \leqslant 1/2$, then we can write $\kappa(\mathbf{x})=c\phi(\mathbf{x})$, where $c>0$ is a constant and ϕ is a multivariate $N(\mu,\Sigma)$ density, for $\mu\in\mathscr{R}^n$ and matrix $\Sigma>0$. Finally, there exist constants c_1, $c_2>0$ (depending only on τ) such that for $\mathbf{x}^\mathsf{T}C^{-1}\mathbf{x}\geqslant 4z_\tau$:

$$c\phi(\mathbf{x}) \leqslant c_1 \exp(-c_2\mathbf{x}^\mathsf{T}C^{-1}\mathbf{x}).$$

Proof. ForanyP: $\xi_{\tau,K}\int (P\rho)^2(\mathbf{x})\,\phi_{C\tau/2}(\mathbf{x})\,d\mathbf{x} \leqslant \int (P\rho)^2(\mathbf{x})\,f_\tau(\mathbf{x})\,d\mathbf{x} = \operatorname{tr}(P^\mathsf{T}PJ(\mathbf{X}+Z_{C\tau})) \leqslant \operatorname{tr}(P^\mathsf{T}PC^{-1})/\tau$, where the first inequality follows from Lemma 3.1. Notice that for $\mathbf{y}\in\mathscr{R}^n$:

$$f_\tau(\mathbf{y})=\frac{1}{(2\pi\tau)^{n/2}(\det C)^{1/2}}\,\mathsf{E}_{\mathbf{X}}\exp\left(-\frac{(\mathbf{y}-\mathbf{X}/\sqrt{v_\mathbf{X}})^\mathsf{T}C^{-1}(\mathbf{y}-\mathbf{X}/\sqrt{v_\mathbf{X}})}{2\tau}\right),$$

so that $f_\tau \leqslant ((2\pi\tau)^n(\det C))^{-1/2}$. In fact, better bounds are possible. Write $W=\mathbf{y}^\mathsf{T}C^{-1}\mathbf{y}$. Then $(\mathbf{y}-\mathbf{x}/\sqrt{v_\mathbf{X}})^\mathsf{T}C^{-1}(\mathbf{y}-\mathbf{x}/\sqrt{v_\mathbf{X}})$ is minimised as a function of $\mathbf{x}$ on $\mathscr{E}(Wv_\mathbf{X}/4)$ when $\mathbf{x}=\sqrt{v_\mathbf{X}}\,\mathbf{y}/2$, which gives the value $W/4$. For all other $\mathbf{y}$, $(\mathbf{x}-\mathbf{y})^\mathsf{T}C^{-1}(\mathbf{x}-\mathbf{y})$ is non-negative. Breaking up the region of integration into $\mathbf{y}\in\mathscr{E}(Wv_\mathbf{X}/4)$ and $\mathbf{y}\notin\mathscr{E}(Wv_\mathbf{X}/4)$, the result follows by Chebyshev since $\mathsf{E}(\mathbf{X}/\sqrt{v_\mathbf{X}})^\mathsf{T}C^{-1}(\mathbf{X}/\sqrt{v_\mathbf{X}})=\operatorname{tr}(C^{-1}B^{(\mathbf{X})}/v_\mathbf{X})\leqslant K_0=$ max(eigenvalue of C^{-1}). We deduce that

$$\begin{aligned} f_\tau(\mathbf{y}) &\leqslant \frac{1}{(2\pi\tau)^{n/2}(\det C)^{1/2}}\left(\exp\left(-\frac{\mathbf{y}^\mathsf{T}C^{-1}\mathbf{y}}{8\tau}\right)+\frac{4K_0}{\mathbf{y}^\mathsf{T}C^{-1}\mathbf{y}}\right)\\ &\leqslant \frac{1}{(2\pi\tau)^{n/2}(\det C)^{1/2}}\left(\frac{8\tau+4K_0}{\mathbf{y}^\mathsf{T}C^{-1}\mathbf{y}}\right). \end{aligned}$$

By Lemma 3.1,

$$f_\tau(\mathbf{0})\geqslant \xi_{\tau,K}\phi_{C\tau/2}(\mathbf{0})=\frac{\exp(-2K_0/\tau)}{2(2\pi\tau)^{n/2}(\det C)^{1/2}}.$$

Now for $\mathbf{x} \in \mathscr{E}(z_\tau)$:

$$\frac{\kappa(\mathbf{x})}{\kappa(\mathbf{0})} \leqslant \frac{f_\tau(\mathbf{x})}{f_\tau(\mathbf{0})/3} \leqslant \frac{3(8\tau + 4K_0)}{2\exp(-2K_0/\tau)\,\mathbf{x}^\mathsf{T} C^{-1}\mathbf{x}}.$$

Thus taking $z_\tau = 2(8\tau + 4K_0)\exp(2K_0/\tau)$, for any $\mathbf{y}$ such that $\mathbf{y}^\mathsf{T} C^{-1}\mathbf{y} = z_\tau$, we know that both $\kappa(\mathbf{y})/K(\mathbf{0}) \leqslant 3/4$ and $\kappa(-\mathbf{y})/\kappa(\mathbf{0}) \leqslant 3/4$. This means that both $\mathbf{u}^\mathsf{T}\mathbf{y} + 1/2\mathbf{y}^\mathsf{T} V\mathbf{y} \leqslant \log 3/4$ and $-\mathbf{u}^\mathsf{T}\mathbf{y} + 1/2\mathbf{y}^\mathsf{T} V\mathbf{y} \leqslant \log 3/4$. Choosing whichever of $\{\mathbf{y}, -\mathbf{y}\}$ makes $\mathbf{u}^\mathsf{T}\mathbf{y}$ positive, we deduce that $1/2\mathbf{y}^\mathsf{T} V\mathbf{y} \leqslant \log(3/4)$, and hence by scaling, for any $\mathbf{x}$, $1/2\mathbf{x}^\mathsf{T} V\mathbf{x} \leqslant \log(3/4)\,\mathbf{x}^\mathsf{T} C^{-1}\mathbf{x}/z_\tau$.

This implies that the matrix (V_{ij}) is negative definite, and we can write $\kappa = c\phi$, where ϕ is a non-degenerate $N(\mu, \Sigma)$ density, with $\Sigma > 0$, $\mu = \Sigma\mathbf{u}$ and $c = f_\tau(\mathbf{0})(2\pi)^{n/2}(\det\Sigma)^{1/2}\exp(\mathbf{u}^\mathsf{T}\Sigma^{-1}\mathbf{u}/2)$.

Now for any $\mathbf{y}$ such that $\mathbf{y}^\mathsf{T} C^{-1}\mathbf{y} = z_\tau$, $\phi(\mathbf{y}) < \phi(\mathbf{0})$ and $\phi(-\mathbf{y}) < \phi(\mathbf{0})$, so it must be the case that $\mu \in \mathscr{E}(z_\tau)$, since otherwise the triple $\phi(\mathbf{y})$, $\phi(\mathbf{0})$, $\phi(-\mathbf{y})$ would be monotonic. Since $\mu \in \mathscr{E}(z_\tau)$, $|f_\tau(\mu)/c\phi(\mu) - 1| \leqslant 1/2$, and so $c\phi(\mu) \leqslant 2f_\tau(\mu) \leqslant 2/((2\pi\tau)^n(\det C))^{1/2}$.

For $\mathbf{x}^\mathsf{T} C^{-1}\mathbf{x} \geqslant 4z_\tau \geqslant 4\mu^\mathsf{T} C^{-1}\mu$, and so by Cauchy–Schwarz $(\mathbf{x} - \mu)^\mathsf{T} C^{-1}(\mathbf{x} - \mu) \geqslant (\sqrt{\mathbf{x}^\mathsf{T} C^{-1}\mathbf{x}} - \sqrt{\mu^\mathsf{T} C^{-1}\mu})^2 \geqslant \mathbf{x}^\mathsf{T} C^{-1}\mathbf{x}/4$, and hence:

$$\begin{aligned} c\phi(\mathbf{x}) &= c\phi(\mu)\exp((\mathbf{x} - \mu)^\mathsf{T} V(\mathbf{x} - \mu)/2) \\ &\leqslant \frac{2}{((2\pi\tau)^n(\det C))^{1/2}}\exp\left(-\left(\frac{\log(4/3)}{4z_\tau}\right)\mathbf{y}^\mathsf{T} C^{-1}\mathbf{y}\right). \end{aligned}$$

Lemma 4.5. Fix $\tau > 0$ and a function ψ as in Definition 4.1. Given a random vector $\mathbf{X} \in \mathscr{C}_\psi$, there exist functions $r_j(\delta, z)$, $j = 0, 1, 2$, for $0 < \delta \leqslant 1/2$, $z \geqslant z_\tau$, such that:

1. If $\sup_{\mathbf{x} \in \mathscr{E}(z)} |f_\tau(\mathbf{x})/c\phi(\mathbf{x}) - 1| \leqslant \delta$, where ϕ is a $N(\mu, \Sigma)$ density then $|c - 1| < r_0(\delta, z)$, $|\mu| < r_1(\delta, z)$, and the covariance matrix B_τ satisfies $|B_\tau - \Sigma| < r_2(\delta, z)$.
2. $\lim_{\delta \to 0,\, z \to \infty} r_i(\delta, z) = 0$ for each i.

Proof. Since for $\mathbf{y} \in \mathscr{E}(z)$, $|f_\tau(\mathbf{y}) - c\phi(\mathbf{y})| < f_\tau(\mathbf{y})\,\delta/(1 - \delta)$, for any function h we can write:

$$\begin{aligned} &\left|\int h(\mathbf{x})(f_\tau(\mathbf{x}) - c\phi(\mathbf{x}))\,d\mathbf{x}\right| \\ &\quad \leqslant \frac{\delta}{1 - \delta}\int |h(\mathbf{x})|\,f_\tau(\mathbf{x})\,d\mathbf{x} + \frac{1 - 2\delta}{1 - \delta}\left|\int h(\mathbf{x})\,f_\tau(\mathbf{x})\,I(\mathbf{x} \notin \mathscr{E}(z))\,d\mathbf{x}\right| \\ &\qquad + \left|c\int h(\mathbf{x})\,\phi(\mathbf{x})\,I(\mathbf{x} \notin \mathscr{E}(z))\,d\mathbf{x}\right|. \end{aligned}$$

Taking $h(\mathbf{x}) = 1, x_i, x_i x_j$, we deduce the result. Now, $\mathsf{E}\mathbf{Y}_\tau^\top C^{-1}\mathbf{Y}_\tau = \mathrm{tr}(C^{-1}B_\tau) \leqslant K_0 + n\tau$, so Chebyshev will give us control over $\mathsf{E}I(\mathbf{Y}_\tau \notin \mathscr{E}(z))$ and $\mathsf{E}Y_i I(\mathbf{Y}_\tau \notin \mathscr{E}(z))$. Lemma 4.4 gives $c\phi(\mathbf{x}) \leqslant c_1 \exp(-c_2 \mathbf{x}^\top C^{-1}\mathbf{x})$, where c, c_1, c_2 depend only on τ.

By Cauchy–Schwarz, since $\mathbf{Y}_\tau \in \mathscr{C}_\Psi$ for some Ψ, for any i, j, $\mathsf{E}Y_i Y_j I(\mathbf{Y} \notin \mathscr{E}(z)) \leqslant (\mathsf{E}Y_i^2 I(\mathbf{Y}_\tau \notin \mathscr{E}(z)))^{1/2} (\mathsf{E}Y_j^2 I(\mathbf{Y}_\tau \notin \mathscr{E}(z)))^{1/2} \leqslant \mathrm{tr}(B_\tau)\, \Psi(z)$.

Proof of Proposition 4.2. We follow an argument from ref. 1. Write Φ_{B_τ} for the $N(\mathbf{0}, B_\tau)$ distribution function, ϕ_{B_τ} for the density and $\rho_{B_\tau} = B_\tau^{-1}\mathbf{x}$ for the score function.

By Lemma 4.3, $|f_\tau(\mathbf{y})/c\phi(\mathbf{y}) - 1| \leqslant \delta$ for $\mathbf{y} \in \mathscr{E}(z)$, where z, δ depend only on τ and the seminorm $\|\rho_{\mathbf{Y}_\tau}\|_\Theta$ and τ, and $z \to \infty$, $\delta \to 0$ as $\|\rho_{\mathbf{Y}_\tau}\|_\Theta \to 0$. Hence:

$$\int |f_\tau(\mathbf{y}) - \phi_{B_\tau}(\mathbf{y})|\, d\mathbf{y}$$

$$\leqslant \frac{\delta}{1-\delta} \mathsf{P}(\mathbf{Y}_\tau \in \mathscr{E}(z)) + 2\, |\mathscr{E}(z)| \sup_{\mathbf{y} \in \mathscr{E}(z)} |c\phi(\mathbf{y}) - \phi_{B_\tau}(\mathbf{y})| + \frac{2\mathrm{tr}(C^{-1}B_\tau)}{z},$$

By Proposition 4.5, we deduce that f_τ tends in $L^1(d\mathbf{x})$ to ϕ_{B_τ} as $\|\rho_{\mathbf{Y}_\tau}\|_\Theta \to 0$, uniformly in $\mathscr{C}_\psi$:

$$\lim_{\zeta \to 0+} \sup[\|f_\tau - \phi_{B_\tau}\|_{L^1(d\mathbf{x})} : \|\rho_{\mathbf{Y}}\|_\Theta \leqslant \zeta, \mathbf{X} \in \mathscr{C}_\psi] = 0.$$

Furthermore this implies that $f/\phi_{B_\tau} \to 1$ in $N(\mathbf{0}, B_\tau)$ probability. By Lemmas 4.3 and 4.5, we know that $\int (P\rho(\mathbf{x}) - P\rho_{B_\tau}(\mathbf{x}))^2\, \phi_{C\tau/2}(\mathbf{x})\, d\mathbf{x} \to 0$. But convergence in $L^2(\phi_{C\tau/2}(\mathbf{x})\, d\mathbf{x})$ implies convergence in $N(\mathbf{0}, C\tau/2)$ probability. So, since $N(\mathbf{0}, C\tau/2)$ and $N(\mathbf{0}, B_\tau)$ are equivalent measures, we deduce that $\rho \to \rho_{B_\tau}$ in $N(\mathbf{0}, B_\tau)$ probability.

The product of convergent sequences is also convergent. Hence we deduce that as $\zeta \to 0+$, uniformly over $\mathbf{X} \in \mathscr{C}_\psi$ with $\|\rho_\tau\|_\Theta \leqslant \zeta$:

$$\int I\left(\left|(P\rho_{\mathbf{Y}}(\mathbf{x}))^\top P\rho_{\mathbf{Y}}(\mathbf{x}) \frac{f_\tau(\mathbf{x})}{\phi_{B_\tau}(\mathbf{x})} - (PB_\tau^{-1}\mathbf{x})^\top PB_\tau^{-1}\mathbf{x}\right| \geqslant \varepsilon\right) \phi_{B_\tau}(\mathbf{x})\, d\mathbf{x} \to 0.$$

We want to show that $\{\rho_\tau^\top C \rho_\tau f_\tau / \phi_{B_\tau}\}$ form a uniformly Φ_{B_τ}-integrable family. By using an analogue of Lemma 3 from ref. 1 there exists a constant c_τ such that:

$$\rho_\tau^\top C \rho_\tau f_\tau(\mathbf{x}) = \sum_i (P\nabla f_\tau)_i^2 \frac{1}{f_\tau(\mathbf{x})} \leqslant c_\tau f_{2\tau}(\mathbf{x}),$$

which follows as $(P\partial f_\tau/\partial y)_i = \mathsf{E}_X(PC^{-1}(\mathbf{y}-\mathbf{X}))_i\ \phi_{C\tau}(\mathbf{y}-\mathbf{X})$, and since we know that $u\exp(-u/2\tau) \leqslant 4\tau\exp(-1)\exp(-u/4\tau)$.

Thus we need only that $h_\tau = f_{2\tau}/\phi_{B\tau}$ form a uniformly Φ_{B_τ}-integrable family. Since entropy increases on convolution: $\int h_\tau \log h_\tau\, d\Phi_{B_\tau} = -H(f_{2\tau}) + \int f_{2\tau}\log\phi_{B_\tau} \leqslant -H(\phi_{2C\tau}) + \int f_{2\tau}\log\phi_{B_\tau} := L(\tau)$. Since $\int |h_\tau|\, I(|h_\tau| > K)\, d\Phi_{B_\tau} \leqslant \int h_\tau(\log h_\tau/\log K)\, I(|h_\tau| > K)\, d\Phi_{B_\tau} \leqslant L(\tau)/\log K$. The uniform Φ_{B_τ}-integrability of the density ratios follows.

We deduce that, defining $\nu_P(\zeta)$ to be

$$\sup_{\mathbf{X}\in\mathscr{C}_\psi:\ \|\rho_\tau\|_\Theta \leqslant \zeta} \int \left((P\rho(\mathbf{x}))^\top (P\rho(\mathbf{x}))\frac{f_\tau}{\phi_{B_\tau}} - (PB_\tau^{-1}\mathbf{x})^\top PB_\tau^{-1}\mathbf{x} \right) d\Phi_{B_\tau}(\mathbf{x}),$$

then $\lim_{\zeta\to 0+}\nu_P(\zeta) = 0$, exactly as required.

Now the final step is to note that if P is invertible, then $\|P\rho_{\mathbf{Y}_\tau}\|_\Theta$ small implies that $\|\rho_{\mathbf{Y}_\tau}\|_\Theta$ is small. This follows since if h is the best linear approximation to ρ then Ph is the best linear approximation to $P\rho_{\mathbf{Y}_\tau}$. Hence $\Theta(P\rho_{\mathbf{Y}_\tau}) = P\Theta(\rho_{\mathbf{Y}_\tau})$.

Now for $Q = (Q_{ij}) = P^{-1}$, and for any function g, by Cauchy–Schwarz:

$$\begin{aligned}\sum_i \int (Qg(\mathbf{x}))_i^2\, \phi_C(\mathbf{x})\, d\mathbf{x} &= \sum_{i,j,k}\int Q_{ij}Q_{ik}g_j(\mathbf{x})\, g_k(\mathbf{x})\, \phi_C(\mathbf{x})\, d\mathbf{x}\\ &= \sum_{j,k}\int (Q^\top Q)_{jk}\, g_j(\mathbf{x})\, g_k(\mathbf{x})\, \phi_C(\mathbf{x})\, d\mathbf{x}\\ &\leqslant n(\max_{j,k}(Q^\top Q)_{jk})\left(\sum_j \int g_j(\mathbf{x})^2\, \phi_C(\mathbf{x})\, d\mathbf{x}\right),\end{aligned}$$

taking $g = P(\Theta\rho_{\mathbf{Y}_\tau})$, we know $\|\rho_{\mathbf{Y}_\tau}\|_\Theta \leqslant n(\max_{j,k}(Q^\top Q)_{jk})\, \|P\rho_{\mathbf{Y}_\tau}\|_\Theta$.

5. CONVERGENCE OF THE DENSITY

Throughout this section, $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots$ is a sequence of independent random vectors. We also set $\mathbf{Y}_\tau^{(S)} = \mathbf{U}^{(S)} + \mathbf{Z}_{C\tau}$, where $\mathbf{U}^{(S)} = \sum_{i\in S}\mathbf{X}^{(i)}/\sqrt{v_S}$.

Lemma 5.1. Let ψ be as in Definition 4.1. If for all i, $\mathbf{X}^{(i)}\in\mathscr{C}_\psi$ then for any $B>0$, there exists a positive function Ψ such that for all finite S, $\mathbf{X}^{(S)}\in\mathscr{C}_\Psi$ and $\mathbf{X}^{(S)}/\sqrt{v_S} + \mathbf{Z}_{B_\tau}\in\mathscr{C}_\Psi$, where $\mathbf{Z}_{B_\tau}$ is independent of $\mathbf{X}^{(S)}$.

Proof. As in the one-dimensional situation, we need only consider the case of $\mathbf{X}^{(i)}$ even, that is, having a probability distribution invariant under reflection about coordinate hyperplanes.

If $t > 0$ and k_0 is the smallest eigenvalue of B^{-1} then $\mathbf{u} \in \mathscr{E}(t)$ means that $t \geqslant \mathbf{u}^\mathsf{T} B^{-1}\mathbf{u} \geqslant k_0 \mathbf{u}^\mathsf{T}\mathbf{u} \geqslant k_0 \mathbf{u}_r^2$. If random vectors $\mathbf{V}^{(i)} = (V_1^{(i)}, \ldots, V_n^{(i)})$ are even, with $\mathbf{V}^{(i)} \in \mathscr{E}(tv_i)$ and $\mathsf{E}(V_r^{(i)})^2 \leqslant v_i$ then

$$\begin{aligned}\mathsf{E}\left(\sum_{i\in S} V_r^{(i)}\right)^4 &\leqslant \sum_{i\in S} \mathsf{E}(V_r^{(i)})^4 + 3\left(\sum_{i\in S}\mathsf{E}(V_r^{(i)})^2\right)^2 \\ &\leqslant \sum_{i\in S} tv_i/k_0 \mathsf{E}(V_r^{(i)})^2 + 3\left(\sum_{i\in S}\mathsf{E}(V_r^{(i)})^2\right)^2 \\ &\leqslant (t/k_0+3)\left(\sum_{i\in S} v_i\right)^2.\end{aligned}$$

Apply this to the sequence $\mathbf{V}^{(i)} = \mathbf{X}^{(i)} I(\mathbf{X}^{(i)} \in \mathscr{E}(tv_i))$, which is even, since for any r, $\mathscr{E}(r)$ is a symmetric set.

Hence we deduce the lemma, by Cauchy–Schwarz, since:

$$\begin{aligned}&\mathsf{E}X_r^{(S)2} I(\mathbf{X}^{(S)} \notin \mathscr{E}(Rv_S)) \\ &\quad = \mathsf{E}\left(\sum_{i\in S} X_r^{(i)} I(\mathbf{X}^{(i)} \notin \mathscr{E}(tv_i)) + I(\mathbf{X}^{(i)} \in \mathscr{E}(tv_i))\right)^2 I(\mathbf{X}^{(S)} \notin \mathscr{E}(Rv_S)) \\ &\quad \leqslant 2\sum_{i\in S} \mathsf{E}X_r^{(i)2} I(\mathbf{X}^{(i)} \notin \mathscr{E}(tv_i)) \\ &\qquad + 2\left(\mathsf{E}\left(\sum_{i\in S} X_r^{(i)} I(\mathbf{X}^{(i)} \in \mathscr{E}(tv_i))\right)^4\right)^{1/2} \mathsf{P}(\mathbf{X}^{(S)} \notin \mathscr{E}(Rv_S))^{1/2} \\ &\quad \leqslant 2v_S\psi(t) + (t/k_0+3)^{1/2}\, v_S (K_0/R)^{1/2},\end{aligned}$$

where as before K_0 is the largest eigenvalue of B^{-1}. Hence taking $t = R^{1/3}$, we obtain the result that $\mathbf{X}^{(S)} \in \mathscr{C}_\Psi$, with $\psi(R) = 2\Psi(R^{1/3}) + ((R^{1/3}/k_0+3)(K_0/R))^{1/2}$. By the same argument, we can add an independent vector $\mathbf{Z}_{B_\tau}$, which also has well behaved tails.

Proof of Theorem 1.5. By the Covariance Convergence Condition, given ε, there exists $V' = V'(\varepsilon)$ such that $v_S \geqslant V'$ implies that for all $\tau > 0$, $-\varepsilon/4 \leqslant \mathrm{tr}(C(C^{(S)}+C_\tau)^{-1} - C^{-1}/(1+\tau)) \leqslant \varepsilon/4$, and hence

$$-\varepsilon/4 \leqslant CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S)}) - \left(CJ(\mathbf{Y}_\tau^{(S)}) - \frac{n}{1+\tau}\right) \leqslant \varepsilon/4.$$

The next observation is that since the v_i are bounded by V_0, given a set S with $v_S \geqslant 3V_0$, we can write it as $S = S_1 \cup S_2$, where $S_1 \cap S_2 = \varnothing$, and $1/3 \leqslant \alpha \leqslant 2/3$ where $\alpha = v_{S_1}/v_S$. For example, defining $r = \min\{l: v_{S \cap (-\infty, l]} \geqslant v_S/3\}$, let $S_1 = S \cap (-\infty, r)$. Then $v_{S_1} \geqslant v_S/3$, and $v_{S_1} \leqslant v_S/3 + V_0 \leqslant 2v_S/3$.

Now given $\varepsilon > 0$, take S such that $v_S \geqslant 3\max(V', V_0)$. Partitioning $S = S_1 \cup S_2$ as above, we have that $v_{S_1}, v_{S_2} \geqslant V'$ and hence:

$$-\varepsilon/4 \leqslant \operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S)}) - (\operatorname{tr} CJ(\mathbf{Y}_\tau^{(S)}) - n/(1+\tau)) \leqslant \varepsilon/4$$

and so $-\varepsilon/4 \leqslant \operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S_i)}) - (\operatorname{tr} CJ(\mathbf{Y}_\tau^{(S_i)}) - n/(1+\tau)) \leqslant \varepsilon/4$, $i = 1, 2$.

Using Propositions 3.3 and 4.2 with $P^{\mathsf{T}}P = C$, there are two possibilities:

1. $\alpha \operatorname{tr} CJ(\mathbf{Y}_\tau^{(S_1)}) + (1-\alpha) \operatorname{tr} CJ(\mathbf{Y}_\tau^{(S_2)}) - \operatorname{tr} CJ(\mathbf{Y}_\tau^{(S)}) \geqslant \varepsilon$, and therefore: $\alpha \operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S_1)}) + (1-\alpha) \operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S_2)}) - \operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S)}) \geqslant \varepsilon/2$.

2. $\alpha \operatorname{tr} CJ(\mathbf{Y}_\tau^{(S_1)}) + (1-\alpha) \operatorname{tr} CJ(\mathbf{Y}_\tau^{(S_2)}) - \operatorname{tr} CJ(\mathbf{Y}_\tau^{(S)}) < \varepsilon$, then since $\alpha(1-\alpha) \geqslant 2/9$, by Proposition 3.3 both $\|P\rho^{(S_1)}\|_\Theta$ and $\|P\rho^{(S_2)}\|_\Theta$ are less than $3\varepsilon^{1/2}/\xi_{\tau, K_0}$. By Proposition 4.2 $\operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S_1)})$ and $\operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S_2)})$ are both less than $\nu_P(3\varepsilon^{1/2}/\xi_{\tau, K_0})$.

For any $V \geqslant 3\max(V', V_0)$, if $v_S \geqslant V$, then for $i = 1, 2$, $v_{S_i} \geqslant V/3$, so $\operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}^{(S_i)}) \geqslant \kappa(V/3, \tau)$. Hence in the first case, $\operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S)}) \leqslant \kappa(V/3, \tau) - \varepsilon/2$, and in the second, $\operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S)}) \leqslant \nu_P(3\varepsilon^{1/2}/\xi_\tau)$. We conclude that:

$$\kappa(V, \tau) = \sup_{S:\, v_S \geqslant V} \operatorname{tr} CJ_{\mathrm{ST}}(\mathbf{Y}_\tau^{(S)}) \leqslant \max\left(\kappa(V/3, \tau) - \varepsilon,\ \nu_P\left(\frac{3\varepsilon^{1/2}}{\xi_\tau}\right)\right).$$

Hence assuming that $\kappa(V, \tau) \geqslant \delta > 0$ for all V gives $\kappa(V, \tau) \leqslant \kappa(V/3, \tau) - (\xi_\tau \nu^{-1}(\delta)/3)^2$ for all V, providing a contradiction, since κ is bounded above and below. Thus $\liminf_{V \to \infty} \kappa(V, \tau) = 0$.

Now by Definition 1.3, $\kappa(V, \tau)$ is monotone decreasing in V, hence it converges to zero. By the Integrability Condition, $\int_0^\infty \kappa(V_1, \tau)\, d\tau$ is finite. Hence by the Monotone Convergence Theorem, $\lim_{V \to \infty} \int_0^\infty \kappa(V, \tau)\, d\tau \to 0$.

Now if set S has $v_S \geqslant V$ then by Definition 1.3, $\operatorname{tr} CJ_{\mathrm{st}}(\mathbf{Y}_\tau^{(S)}) \leqslant \kappa(V, \tau)$. By convergence of covariance, $\lim_{V \to \infty} \sup_{S:\, v_S \geqslant V} |C^{(S)}/v_S - C| = 0$, and hence the second and third terms in the n-dimensional de Bruijn identity (Theorem 2.3) converge to zero. Hence we deduce that by Theorem 2.3:

$$\lim_{V \to \infty} \Big(\sup_{S:\, v_S \geqslant V} D(g^{(S)} \,\|\, \phi)\Big) \leqslant \frac{\log e}{2} \Big(\lim_{V \to \infty} \kappa(V, \tau)\Big) = 0.$$

Proof of Theorem 1.6. The IID case is much easier, since Proposition 3.3 simplifies. If we define ρ_k for the score function of $\mathbf{W}^{(k)} \sum_{i=1}^{2^k} \mathbf{X}^{(i)}/\sqrt{2^k v} + \mathbf{Z}_{C\tau}$, it becomes:

$$\operatorname{tr}(P^{\top} P J(\mathbf{W}^{(k)})) - P^{\top} P J(\mathbf{W}^{(k+1)}) \geqslant \text{const. } \|P\rho_k\|_{\Theta}^2.$$

As in ref. 1, we deduce that $\operatorname{tr}(P^{\top} P J(\mathbf{W}^{(k)}))$ converges monotonically to its minimum, and this monotone convergence means that we only require the condition that $D(g_m \,\|\, \phi)$ is ever finite to ensure its convergence to zero.

As suggested in the introduction, such methods will also work in the case of 2 (or indeed higher) dimensional lattices of independent random variables, corresponding to 2D lattice models in the absence of spontaneous magnetisation. The key observation is that once again, we can decompose any large enough set of random variables S into two smaller sets S_1, S_2, each with comparable variance. Formally, the proof of Theorem 1.5 continues in the same way. In a forthcoming paper (ref. 8), we extend these techniques into the case of FKG systems of random variables, and prove convergence in relative entropy in a weakly dependent case.

ACKNOWLEDGMENTS

Andrew Barron of Yale University provided advice and insights into the results of this paper.

REFERENCES

1. A. R. Barron, Entropy and the Central Limit theorem, *Ann. Prob.* **14**:336–342 (1986).
2. L. D. Brown, A proof of the Central Limit Theorem motivated by the Cramer–Rao inequality, in *Statistics and Probability: Essays in Honour of C. R. Rao*, G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh, eds., pp. 141–148 (North-Holland, New York, 1982).
3. E. A. Carlen and A. Soffer, Entropy production by block variable summation and Central Limit Theorems, *Comm. Math. Phys.* **140**:339–371 (1991).
4. G. Dattoli, S. Lorenzutta, G. Maino, and A. Torre, Phase-space dynamics and Hermite polynomials of two variables and two indices, *J. Math. Phys.* **35**:4451–4462 (1994).
5. R. L. Dobrushin and S. B. Shlosman, Absence of breakdown of continuous symmetry in two-dimensional models of statistical physics, *Comm. Math. Phys.* **42**:31–40 (1975).
6. O. T. Johnson, Entropy inequalities and the Central Limit Theorem, *Stochastic Process. Appl.* **88**:291–304 (2000).
7. O. T. Johnson and Yu. M. Suhov, Entropy and convergence on compact groups, *J. Theor. Prob.* **13**:843–857 (2000).
8. O. T. Johnson, An information-theoretic Central Limit Theorem for finitely susceptible FKG systems. To be submitted (2001).
9. S. Kullback and R. A. Leibler, On information and sufficiency, *Ann. Math. Stat.* **22**:79–86 (1951).

10. Yu. V. Linnik, An information-theoretic proof of the Central Limit Theorem with the Lindeberg Condition, *Theor. Prob. Appl.* **4**:288–299 (1959).
11. Yu. V. Linnik, On certain connections between the information measures of Shannon and Fisher and the summation of random vectors, in *Transactions of the 2nd Prague Conference* (Czechoslovak Academy of Sciences, 1960). In Russian.
12. M. Romerio and P.-A. Vuillermot, Absence of long-range order for a class of two-dimensional systems, in *Group Theoretical Methods in Physics* (Proc. Third Internat. Colloq., Centre Phys. Théor., Marseille, 1974), Vol. 1, pp. 153–159. Centre Nat. Recherche Sci., Centre Phys. Théor., Marseille (1974).
13. G. Szegő, *Orthogonal Polynomials* (American Mathematical Society, New York, 1958), revised edition .