Math-Net.Ru

Общероссийский математический портал

О. Johnson, An information-theoretic central limit theorem for finitely susceptible FKG systems, *Теория вероятн. и ее примен.*, 2005, том 50, выпуск 2, 331–343

DOI: <https://doi.org/10.4213/tvp110>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 3.95.254.165

26 апреля 2023 г., 18:17:40

© 2005 г.

JOHNSON O.*

AN INFORMATION-THEORETIC CENTRAL LIMIT THEOREM
FOR FINITELY SUSCEPTIBLE FKG SYSTEMS

Рассуждения, касающиеся теоретико-энтропийной сходимости, переносятся с независимого случая на случай FKG-величин (случайных величин Фортюина–Кастелейна–Жиинбра). FKG-системы выбраны потому, что структура зависимости в них определяется только ковариацией, хотя затем многие из этих рассуждений мы используем для слабозависимых случайных величин. Как и в предыдущих работах Э. Р. Барона и автора, рассматриваются случайные величины с малыми нормальными возмущениями, поскольку FKG-свойство позволяет контролировать получающиеся в результате плотности. Мы приходим к необходимости накладывать условие «конечной чувствительности», заключающееся в том, что ковариация между одной случайной величиной и суммой всех случайных величин должна оставаться конечной.

Ключевые слова и фразы: сходимость к нормальному закону, энтропия, информация Фишера, FKG-величины (величины Фортюина–Кастелейна–Жиинбра).

1. Introduction and notation. Gnedenko and Korolev [4] discuss the relationship between probabilistic limit theorems and the increase of entropy, saying that

The formal coincidence of definitions of entropies in physics and in information theory gives rise to the question, whether analogs of the second law of thermodynamics exist in probability theory.

It is indeed striking that whilst the principle of increase of physical entropy is taken for granted, the increase of information-theoretic entropy is much less well understood. The Gaussian is both the distribution of maximum entropy (under a variance constraint) and the limit distribution of convolutions in the Central Limit regime (which preserves variance). There is clear physical interest in asking questions such as whether entropy always increases on convolution, and whether it tends to this maximum.

By showing that the entropy tends to its maximum, we prove the Central Limit Theorem in a stronger sense than classical methods achieve. This form

---

* Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK; e-mail: otj1000@cam.ac.uk

of convergence implies the classical weak convergence proved in Theorem 2 of Newman [13]. Lemma 5 of Takano [16] and Theorem 3.1 of Carlen and Soffer [3] also only prove weak convergence, though under different conditions. Furthermore by understanding how score functions become more linear on convolution, we gain an insight into the workings of the limit theorem, and why the Gaussian is the limiting distribution. Since relative entropy appears as a distance in large deviations and coding theory, it is interesting to know whether it tends to zero.

Gnedenko and Korolev propose a program to investigate the relationship between results like the Central Limit Theorem and maximum entropy distributions. Perhaps the earliest paper to investigate this link was by Linnik [11], though he only proves weak convergence. This program has been developed by Brown [2], Barron [1], Johnson [6] and Barron and Johnson [7], who have used information-theoretic techniques to prove the Central Limit Theorem. These papers only deal with the case of independent random variables, [2] and [1] in the case of identically distributed variables, and [6] and [7] for nonidentical variables satisfying a Lindeberg-like condition.

This paper extends these results and develops new techniques to consider the case of dependent random variables satisfying the FKG (Fortuin–Kasteleyn–Ginibre) inequalities. The fact that proofs of entropy-theoretic convergence have only previously existed for independent variables is unfortunate, particularly given the natural physical interest in dependent systems. In extending Barron's work, we have shown the link between physical and information-theoretic entropies holds in more generality than Gnedenko and Korolev discussed.

Definition 1.1. A set of random variables $\{X_1, \dots, X_m\}$ is said to be FKG if $\text{cov}(F(X_1, \dots, X_m), G(X_1, \dots, X_m)) \geq 0$ for all increasing functions F, G (here by «increasing», we mean increasing in each individual coordinate).

FKG inequalities hold for many physical models with «positive correlation», as discussed by Newman [13]. For example, in the Ising model with Hamiltonian $H = \sum_{j,k} J(j-k) X_j X_k - h \sum X_j$, the FKG inequalities hold if $J(r) \geq 0$ for all r . Further, FKG inequalities hold for percolation models, where $X_j = I$ (vertex j is in an infinite cluster), and Yukawa models of quantum field theory.

To obtain convergence in relative entropy we use de Bruijn's identity, which relates the relative entropy to Fisher information of perturbed random variables, which have densities we can control.

Definition 1.2. For a random variable U with variance σ^2 and smooth density f , we consider the score function $\rho(u) = f'(u)/f(u)$, the Fisher information $J(U) = \mathbf{E}\rho^2(U)$, and the standardised Fisher information $J_{\text{st}}(U) = \sigma^2 J(U) - 1 = \mathbf{E}(\sigma\rho(U) + U/\sigma)^2 \geq 0$.

Lemma 1.1 (de Bruijn). *For U with mean 0, variance 1, and density f , the relative entropy distance $D(f \parallel \varphi)$ from a standard normal can be expressed in terms of the Fisher information of U perturbed by normals $Z^{(\tau)} \sim N(0, \tau)$:*

$$D(f \parallel \varphi) = \frac{1}{2} \int_0^\infty \left(J(U + Z^{(\tau)}) - \frac{1}{1 + \tau} \right) d\tau = \frac{1}{2} \int_0^\infty \frac{J_{\text{st}}(U + Z^{(\tau)})}{1 + \tau} d\tau.$$

Lemma 3 in [13] shows that for (S, T) FKG, we can control $\text{cov}(f(S), g(T))$. In our case this is useful because $Z_S^{(\tau)}$, $Z_T^{(\tau)}$ are normal $N(0, \tau)$, independent of S , T and of each other, this means we can control the densities $p_{X,Y} - p_X p_Y$, where $(X, Y) = (S + Z_S^{(\tau)}, T + Z_T^{(\tau)})$. See Lemma 3.4 for a discussion of these methods.

First we establish conditions under which convergence $J_{\text{st}}(U) \rightarrow 0$ holds, which implies more conventional forms of convergence.

Lemma 1.2 (Shimizu [14]). *If U has variance σ^2 , density f , and distribution function F , then denoting the density and the distribution function of a $N(0, \sigma^2)$ by φ and Φ , respectively, we have*

$$\sup_u |F(u) - \Phi(u)| \leq \int |f(u) - \varphi(u)| du \leq 4\sqrt{3}\sqrt{J_{\text{st}}(U)},$$

$$\sup_u |f(u) - \varphi(u)| \leq \left(1 + \sqrt{\frac{6}{\pi}}\right) \sqrt{J_{\text{st}}(U)}.$$

Indeed weak convergence implies that $\lim_n \mathbf{E}h(S_n) = \mathbf{E}h(Z)$ for all bounded uniformly continuous functions h . Convergence in relative entropy extends this to the class of measurable functions bounded by some multiple of $x^2 + 1$ (see [1] for further details).

Definition 1.3. Consider a stationary d -dimensional system of random variables $X_{\mathbf{u}}$ (where the index $\mathbf{u} \in \mathbf{Z}^d$) with mean zero and a finite variance. For a particular vector $\mathbf{u} = (u_1, \dots, u_d)$, we define the box

$$B_{\mathbf{u}} = \{\mathbf{y}: 0 \leq y_i \leq u_i \text{ for all } i\}$$

with volume $|\mathbf{u}| = |B_{\mathbf{u}}| = \prod_i u_i$. We can define $v(\mathbf{x}) = \mathbf{D}(\sum_{\mathbf{u} \in B_{\mathbf{x}}} X_{\mathbf{u}})$ and $U_{\mathbf{x}} = (\sum_{\mathbf{u} \in B_{\mathbf{x}}} X_{\mathbf{u}})/\sqrt{|\mathbf{x}|}$. Define perturbed random variables $Y_{\mathbf{u}}^{(\tau)} = X_{\mathbf{u}} + Z_{\mathbf{u}}^{(\tau)}$, for $Z_{\mathbf{u}}^{(\tau)}$ a sequence of $N(0, \tau)$ independent of $X_{\mathbf{u}}$ and of each other. We introduce $V_{\mathbf{x}}^{(\tau)} = (\sum_{\mathbf{u} \in B_{\mathbf{x}}} Y_{\mathbf{u}}^{(\tau)})/\sqrt{|\mathbf{x}|} \sim U_{\mathbf{x}} + Z^{(\tau)}$.

Condition 1 (finite susceptibility).

$$v = \sum_{\mathbf{u}} \text{cov}(X_{\mathbf{0}}, X_{\mathbf{u}}) < \infty.$$

Definition 1.4. For a function ψ , define the class of random variables X with variance σ^2 such that

$$\mathcal{C}_{\psi} = \left\{ X: \mathbf{E}X^2 I(|X| \geq R\sigma) \leq \sigma^2 \psi(R) \text{ for all } R \right\}.$$

Condition 2 (uniform integrability). *There exists a function $\psi(R)$ which tends to zero as $R \rightarrow \infty$ such that $V_{\mathbf{u}}^{(\tau)} \in \mathcal{C}_\psi$ for all $\mathbf{u}$.*

Theorem 1.1. *Consider a stationary collection of mean zero, finite variance random variables $X_{\mathbf{u}}$ obeying the FKG inequalities and finite susceptibility (Condition 1). Then*

$$\lim_{n \rightarrow \infty} \sup_{\mathbf{x}: |\mathbf{x}|=n} J_{\text{st}}(V_{\mathbf{x}}^{(\tau)}) = 0,$$

if and only if Condition 2 (uniform integrability) holds.

Condition 2 (for stationary FKG variables with finite variance) is actually implied by Condition 1. This follows by Newman's proof [13] that these conditions imply the central limit theorem, since if some sequence of distribution functions $F_n(x) \rightarrow \Phi(x)$, then

$$\begin{aligned} \int z^2 I(|z| > N) dF_n(z) &= 1 - \int z^2 I(|z| \leq N) dF_n(z) \\ &\rightarrow 1 - \int z^2 I(|z| \leq N) d\Phi(z) = \int z^2 I(|z| > N) d\Phi(z). \end{aligned}$$

Carlen and Soffer claim in [3, p. 369] that if $\text{cov}(X_0, X_{\mathbf{i}})$ decays at a rate of $|\mathbf{i}|^{-t}$, where $t > 2d$, d the dimension of the lattice, then Condition 2 will hold. This roughly corresponds to requiring that $\sum_{\mathbf{i}} (\text{cov}(X_0, X_{\mathbf{i}}))^{1/2} < \infty$.

Note: We do not need to assume that the X_i themselves have densities — even if not, by Lemma 1.2 we obtain weak convergence of the normalized sums of the original variables.

Definition 1.5. Define $\kappa(n, \tau) = \sup_{|\mathbf{u}| \geq n} J_{\text{st}}(V_{\mathbf{u}}^{(\tau)})$.

Condition 3. *For some n , $\int \kappa(n, \tau)/(1 + \tau) d\tau$ is finite.*

Theorem 1.2. *Consider a stationary collection of mean zero, finite variance random variables $X_{\mathbf{u}}$ with densities, obeying the FKG inequalities and Conditions 1 and 3. Let $g_{\mathbf{u}}$ be the density of $V_{\mathbf{u}}^{(\tau)}$. Then*

$$\lim_{n \rightarrow \infty} \sup_{\mathbf{x}: |\mathbf{x}|=n} D(g_{\mathbf{u}} \| \varphi) = 0,$$

if and only if Condition 2 (uniform integrability) holds.

Proof. Via monotone convergence: $\kappa(n, \tau)$ converges monotonically to 0 in n , and hence $\int \kappa(n, \tau) d\tau$ converges to zero.

Newman [13] claims that if instead of scaling by $|\mathbf{x}|$, we scale by $v(\mathbf{x})$, Condition 1 can be relaxed to Condition 2 and Condition 4.

Condition 4. *If $K(R) = \sum_{|j| \leq R} \text{cov}(X_0, X_j)$, then $K(R)$ is slowly varying (that is, for any λ , $\lim_{R \rightarrow \infty} K(\lambda R)/K(R) = 1$).*

He remarks that Condition 2 can be checked if for example $\mathbf{E}(V_n^{(\tau)})^4 \leq 3(\mathbf{E}(V_n^{(\tau)})^2)^2$, which itself holds in many cases as a consequence of results such as the Lebowitz inequality [9] or the GHS (Griffiths–Hurst–Sherman) inequality [12].

Takano [15], [16] deals with the behavior of entropy, under a δ_4 -mixing condition which seems hard to check in most useful cases, since it is defined by a ratio of densities. Further, Takano only proves convergence in relative entropy of the «rooms» (in Bernstein’s terminology), equivalent to weak convergence of the original variables. Our conclusion holds in the stronger sense of relative-entropy convergence of the full sequence. Another paper to use entropy-theoretic methods in the dependent case is by Carlen and Soffer [3]. They use a variety of conditions which are different to ours, but again only prove weak convergence for dependent variables.

2. Fisher information and convolution. In the independent case, Fisher information is a subadditive quantity on convolution. In the dependent case, we prove that Fisher information is «almost subadditive» — the interest comes in trying to bound the error term. Takano [15], [16] produces bounds which depend on his δ_4 -mixing coefficient, which is hard to understand, and hard to check since it depends on ratios of densities. Our calculations provide weaker, and more standard conditions under which the CLT will hold in the sense of convergence of Fisher information.

Definition 2.1. Given random variables X, Y with score functions ρ_X, ρ_Y , for any β , we define $\tilde{\rho}$ to be the score function of $\sqrt{\beta}X + \sqrt{1-\beta}Y$ and then

$$\Delta(X, Y, \beta) = \mathbf{E}\left(\sqrt{\beta}\rho_X(X) + \sqrt{1-\beta}\rho_Y(Y) - \tilde{\rho}(\sqrt{\beta}X + \sqrt{1-\beta}Y)\right)^2 \geq 0.$$

The principal theorem of this section is the following.

Theorem 2.1. *Let S and T be FKG random variables with mean zero and variance $\leq K$. Consider $Z_S^{(\tau)}$ and $Z_T^{(\tau)}$ distributed as $N(0, \tau)$, independent of S, T and of each other. Define $X = S + Z_S^{(\tau)}$ and $Y = T + Z_T^{(\tau)}$ with score functions ρ_X and ρ_Y . There exists a constant $C = C(K, \tau, \varepsilon)$ such that for any β*

$$\beta J(X) + (1-\beta)J(Y) - J(\sqrt{\beta}X + \sqrt{1-\beta}Y) + C \operatorname{cov}(S, T)^{1/3-\varepsilon} \geq \Delta(X, Y, \beta).$$

If S, T have bounded $(2+\delta)$ -th moment, we can replace $\frac{1}{3}$ by $(2+\delta)/(6+\delta)$.

Proof. The proof requires some involved analysis and is deferred to Section 3.

Next, we need lower bounds on the term $\Delta(X, Y, \beta)$. As discussed in [7], in the case of independent variables, such terms are equal to zero exactly when all the functions concerned are linear. In general, if such an expression is small, then the derivatives of ρ_X and ρ_Y are close to constant, so long as we have uniform control over the tails of X and Y .

Proposition 2.1. *For any ψ , there exists a function $\nu = \nu_\psi$ with $\nu(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$ such that if X, Y lie in $\mathcal{C}_\psi$, then*

$$\beta(1 - \beta) J_{\text{st}}(X) \leq \nu(\Delta(X, Y, \beta)).$$

P r o o f. For each τ , define a semi-norm $\|\cdot\|_{\Theta(\tau)}$ on functions via

$$\|f\|_{\Theta(\tau)}^2 = \inf_{a,b} \mathbf{E}(f(Z_{\tau/2}) - aZ_{\tau/2} - b)^2,$$

where $Z_{\tau/2}$ is $N(0, \tau/2)$.

Using [8, Lemma 3.1] for $K > 0$, there exists a constant $\xi_K > 0$ such that for any dependent random variables (S, T) with variances $\leq K$ the sum $(X, Y) = (S + Z_S^{(\tau)}, T + Z_T^{(\tau)})$ has joint density $p^{(\tau)}(x, y)$ bounded from below by $\xi_K \varphi_{\tau/2}(x) \varphi_{\tau/2}(y)$.

Hence writing $h(x, y) = \sqrt{\beta} \rho_X(x) + \sqrt{1 - \beta} \rho_Y(y) - \tilde{\rho}(\sqrt{\beta} x + \sqrt{1 - \beta} y)$, we have

$$\begin{aligned} \Delta(X, Y, \beta) &= \int p^{(\tau)}(x, y) h(x, y)^2 dx dy \geq \xi_K \int \varphi_{\tau/2}(x) \varphi_{\tau/2}(y) h(x, y)^2 dx dy \\ &\geq \frac{\beta(1 - \beta)\xi_K}{2} (\|\rho_X\|_{\Theta(\tau)}^2 + \|\rho_Y\|_{\Theta(\tau)}^2), \end{aligned}$$

by [6, Proposition 3.2]. The crucial result of [6] implies for a fixed ψ , if the sequence $X_n \in \mathcal{C}_\psi$ have score functions f_n , then $\|f_n\|_{\Theta(\tau)} \rightarrow 0$ implies that $J_{\text{st}}(X_n) \rightarrow 0$.

Define

$$J(n) = \sup\{J_{\text{st}}(V_{\mathbf{x}}^{(\tau)}): |\mathbf{x}| = n\}.$$

Note that in the one-dimensional case, there is only one set of this form, $\{1, \dots, n\}$.

Corollary 2.1. *If Condition 2 holds for some ψ , then there exists $d(m) \rightarrow 0$ as $m \rightarrow \infty$ such that for $m \geq n$*

$$J(n + m) \leq \frac{m}{m + n} J(m) + \frac{n}{m + n} J(n) + d(m) - \nu_\psi^{-1} \left(\frac{J(m)mn}{(m + n)^2} \right).$$

P r o o f. For any $\mathbf{x}$, we can decompose the box into smaller distinct ones: $B_{\mathbf{x}} = B_{\mathbf{y}} \cup \tilde{B}_{\mathbf{z}}$, where $B_{\mathbf{y}} \cap \tilde{B}_{\mathbf{z}} = \emptyset$, and $\mathbf{x} = \mathbf{y} = \mathbf{z}$ for all but the j -th coordinate, so that

$$\tilde{B}_{\mathbf{z}} = \{\mathbf{u}: 0 \leq u_i \leq x_i \text{ for all } i \neq j \text{ and } y_j + 1 \leq u_j \leq y_j + z_j = x_j\}.$$

This corresponds to splitting the box into two smaller ones by making a cut parallel to the j -th face. We write $\tilde{U}_{\mathbf{z}} = (\sum_{\mathbf{u} \in \tilde{B}_{\mathbf{z}}} X_{\mathbf{u}})/\sqrt{|\mathbf{z}|}$, and $\tilde{V}_{\mathbf{z}}^{(\tau)} = (\sum_{\mathbf{u} \in \tilde{B}_{\mathbf{z}}} Y_{\mathbf{u}}^{(\tau)})/\sqrt{|\mathbf{z}|}$.

Taking $\beta = m/(m+n) = |\mathbf{y}|/|\mathbf{x}|$, and by substituting in Theorem 2.1, since $J(m) \leq 1/\tau$, we obtain

$$J_{\text{st}}(V_{\mathbf{x}}^{(\tau)}) \leq \frac{m}{m+n} J_{\text{st}}(V_{\mathbf{y}}^{(\tau)}) + \frac{n}{n+m} J_{\text{st}}(V_{\mathbf{z}}^{(\tau)}) \\ + C' \text{cov}(V_{\mathbf{y}}^{(\tau)}, \tilde{V}_{\mathbf{z}}^{(\tau)})^{1/3-\varepsilon} - \Delta\left(V_{\mathbf{y}}^{(\tau)}, \tilde{V}_{\mathbf{z}}^{(\tau)}, \frac{m}{m+n}\right).$$

Define

$$c(m, n) = \sup \{ \text{cov}(U_{\mathbf{y}}, \tilde{U}_{\mathbf{z}}) : |\mathbf{y}| = m, |\mathbf{z}| = n \}.$$

Under the finite susceptibility condition (Condition 1), Lemma 4 in [13] shows that this quantity is bounded above in a suitable way, since $\mathbf{D}(U_{\mathbf{u}} - U_{\mathbf{v}}) \rightarrow 0$ if $|\mathbf{u}|/|\mathbf{v}| \rightarrow 0$.

We are able to complete the proof of the CLT, under FKG conditions.

P r o o f o f T h e o r e m 1.1. We first establish convergence along the «powers of 2 subsequence». Condition 2 implies that $V_n^{(\tau)} \in \mathcal{C}_{\psi}$ for some ψ and hence that $J(V_n^{(\tau)}) \leq \nu_{\psi}(\Delta(V_n^{(\tau)}, \vartheta_n, \frac{1}{2}))$. We can write $D(k) = \nu^{-1}(J_{\text{st}}(V_{2^k}^{(\tau)}/4))$. By Corollary 2.1, we know that

$$J(2^{k+1}) \leq J(2^k) + d(k) - D(k),$$

where $d(k) \rightarrow 0$.

We use an argument structured like Linnik's proof [11]. Given ε , we can find K such that $d(k) \leq \varepsilon/2$, for all $k \geq K$. Now either

1) For all $k \geq K$, $2d(k) \leq D(k)$, so $J(2^k) - J(2^{k+1}) \geq D(k)/2$, and summing the telescoping sum, we deduce that $\sum_k D(k)$ is finite, and hence there exists L such that $D(L) \leq \varepsilon$;

2) otherwise for some $L \geq K$, $2d(L) \geq D(L)$, then $D(L) \leq \varepsilon$.

Thus, in either case, there exists L such that $D(L) \leq \varepsilon$, and hence by Proposition 2.1, $J(2^L) \leq 4\nu(\varepsilon)$.

Now, for any $k \geq L$, either $J(2^{k+1}) \leq J(2^k)$, or $D(k) \leq d(k) \leq \varepsilon$. In the second case, $J(2^k) \leq 4\nu(\varepsilon)$, so that $J(2^{k+1}) \leq 4\nu(\varepsilon) + \varepsilon$. In either case, we prove by induction for all $k \geq L$, that $J(2^{k+1}) \leq 4\nu(\varepsilon) + \varepsilon$.

Now, we can «fill in the gaps» to gain control of the whole sequence, adapting the proof of the standard subadditive inequality, using the methods described in [5, Appendix 2].

3. Proof of subadditive relation. This is the key part of the argument, proving the bounds at the heart of the limit theorems. However, although the analysis is somewhat involved, it is not too technically difficult.

We introduce notation, where it will be clear whether densities or score functions are associated with joint or marginal distributions, by their number of arguments: $\rho_X(x)$ will be the score function of X , and $p'_X(x)$ the derivative of its density. For joint densities $p_{X,Y}(x, y)$, we denote by $p_{X,Y}^{(1)}(x, y)$ the

derivative of the density with respect to the first argument and $\rho_{X,Y}^{(1)}(x, y) = p_{X,Y}^{(1)}(x, y)/p_{X,Y}(x, y)$, and so on.

Note that a similar equation to the independent case tells us about the behavior of Fisher information of sums.

Lemma 3.1. *If X, Y are random variables with joint density $p(x, y)$ and score functions $\rho_{X,Y}^{(1)}$ and $\rho_{X,Y}^{(2)}$, then $X + Y$ has score function $\tilde{\rho}$ given by*

$$\tilde{\rho}(z) = \mathbf{E}[\rho_{X,Y}^{(1)}(X, Y) \mid X + Y = z] = \mathbf{E}[\rho_{X,Y}^{(2)}(X, Y) \mid X + Y = z].$$

P r o o f. Since $X + Y$ has density p_{X+Y} given by $p_{X+Y}(z) = \int p(z - y, y) dy$, then

$$p'_{X+Y}(z) = \int \frac{\partial p}{\partial z}(z - y, y) dy.$$

Hence dividing, we obtain that

$$\tilde{\rho}(z) = \frac{p'_{X+Y}(z)}{p_{X+Y}(z)} = \int \rho_{X,Y}^{(1)}(z - y, y) \frac{p(z - y, y)}{p_{X+Y}(z)} dy,$$

as claimed.

For given a, b , define the function $M(x, y) = M_{a,b}(x, y)$ by

$$M(x, y) = a(\rho_{X,Y}^{(1)}(x, y) - \rho_X(x)) + b(\rho_{X,Y}^{(2)}(x, y) - \rho_Y(y)),$$

which is zero if X and Y are independent. We will show that if $\text{cov}(X, Y)$ is small, then M is close to zero.

Proposition 3.1. *If X, Y are random variables with score functions ρ_X, ρ_Y and if the sum $\sqrt{\beta}X + \sqrt{1-\beta}Y$ has a score function $\tilde{\rho}$, then*

$$\begin{aligned} & \beta J(X) + (1 - \beta) J(Y) - J(\sqrt{\beta}X + \sqrt{1-\beta}Y) \\ & + 2\sqrt{\beta(1-\beta)} \mathbf{E}\rho_X(X) \rho_Y(Y) + \mathbf{E}M_{\sqrt{\beta}, \sqrt{1-\beta}}(X, Y) \tilde{\rho}(X + Y) \\ & = \mathbf{E}\left(\sqrt{\beta} \rho_X(X) + \sqrt{1-\beta} \rho_Y(Y) - \tilde{\rho}(\sqrt{\beta}X + \sqrt{1-\beta}Y)\right)^2. \end{aligned}$$

P r o o f. By the two-dimensional version of Stein's equation, for any function $f(x, y)$:

$$\mathbf{E}\rho_{X,Y}^{(1)}(X, Y) f(X, Y) = -\mathbf{E} \frac{\partial f}{\partial x}(X, Y).$$

In particular, if $f(x, y) = \tilde{\rho}(x + y)$:

$$\mathbf{E}\rho_X(X) \tilde{\rho}(X + Y) = -\mathbf{E} \tilde{\rho}'(X + Y) - \mathbf{E}(\rho_{X,Y}^{(1)}(X, Y) - \rho_X(X)) \tilde{\rho}(X + Y).$$

Hence, we know that for any a, b :

$$\mathbf{E}(a\rho_X(X) + b\rho_Y(Y)) \tilde{\rho}(X + Y) = (a + b) J(X + Y) - \mathbf{E}M_{a,b}(X, Y) \tilde{\rho}(X + Y).$$

By considering $\int p(x, y)(a\rho_X(x) + b\rho_Y(y) - (a + b)\tilde{\rho}(x + y))^2 dx dy$, dealing with the cross term with the expression above, we deduce that

$$\begin{aligned} & a^2 J(X) + b^2 J(Y) - (a + b)^2 J(X + Y) + 2ab \mathbf{E} \rho_X(X) \rho_Y(Y) \\ & + 2(a + b) \mathbf{E} M_{a,b}(X, Y) \tilde{\rho}(X + Y) \\ & = \mathbf{E} \left(a\rho_X(X) + b\rho_Y(Y) - (a + b)\tilde{\rho}(X + Y) \right)^2 \geq 0. \end{aligned}$$

As in the independent case, we can rescale and consider $X' = \sqrt{\beta} X$, $Y' = \sqrt{1 - \beta} Y$, and take $a = \beta$, $b = 1 - \beta$. Note that $\sqrt{\beta} \rho_{X'}(u) = \rho_X(u/\sqrt{\beta})$, $\sqrt{1 - \beta} \rho_{Y'}(v) = \rho_Y(v/\sqrt{1 - \beta})$. Proposition 3.1 is proved.

We will show that the two terms on the second line of Proposition 3.1 can be controlled when $(X, Y) = (S + Z_S^{(\tau)}, T + Z_T^{(\tau)})$, by controlling $\text{cov}(S, T)$ alone. We need control of the score functions of perturbed variables. We obtain this in two regions, first in Lemma 3.2 over the tail, and then in Lemma 3.3 over the rest of the real line.

We require an extension of Lemma 3 in [1] applied to single and bivariate random variables.

Lemma 3.2. *For any random variables S, T as before we define $(X, Y) = (S + Z_S^{(\tau)}, T + Z_T^{(\tau)})$ and define $p_{U,V}^{(2\tau)}$ for the density of $(U, V) = (S + Z_S^{(2\tau)}, T + Z_T^{(2\tau)})$. Now there exists a constant $c_{\tau,k} = \sqrt{2}(2k/(\tau e))^{k/2}$ such that for all x*

$$\begin{aligned} p_X^{(\tau)}(x) |\rho_X(x)|^k &\leq c_{\tau,k} p_U^{(2\tau)}(x), \\ p^{(\tau)}(x, y) |\rho_{X,Y}^{(1)}(x, y)|^k &\leq c_{\tau,k} p_{U,V}^{(2\tau)}(x, y), \\ p^{(\tau)}(x, y) |\rho_{X,Y}^{(2)}(x, y)|^k &\leq c_{\tau,k} p_{U,V}^{(2\tau)}(x, y), \end{aligned}$$

and hence for any $k > 1$

$$(\mathbf{E} |\rho_X(X)|^k)^{1/k} \leq \sqrt{\frac{2^{1/k} 2k}{\tau e}}.$$

P r o o f. We adapt the proof given in [1] using Hölder's inequality and the bound $(u/\tau)^k \varphi_\tau(u) \leq c_{\tau,k} \varphi_{2\tau}(u)$ for all u :

$$\begin{aligned} p'_X(x)^k &= \left(\mathbf{E} \left(\frac{x - S}{\tau} \right) \varphi_\tau(x - S) \right)^k \\ &\leq \left(\mathbf{E} \left(\frac{x - S}{\tau} \right)^k \varphi_\tau(x - S) \right) (\mathbf{E} \varphi_\tau(x - S))^{k-1} \\ &\leq c_{\tau,k} (\mathbf{E} \varphi_{2\tau}(x - S)) p_X(x)^{k-1}. \end{aligned}$$

A similar argument gives the other bounds.

Now, the normal perturbation ensures that the density does not decrease too fast, and so the modulus of the score function can not grow too fast. By considering S normal, so that ρ grows linearly with u , we know that the B^3 rate of growth established in Lemma 3.3 is a sharp bound.

Lemma 3.3. *If S is a random variable with variance $\leq K$, for $X = S + Z_S^{(\tau)}$ with score function ρ , for $B > 1$, there exists a function $f_1(\tau, K)$ such that*

$$\int_{-B\sqrt{\tau}}^{B\sqrt{\tau}} \rho(u)^2 du \leq f_1(\tau, K) B^3.$$

P r o o f. Now: $p(u) \geq (2 \exp(2K/\tau))^{-1} \varphi_{\tau/2}(u)$, so that for $u \in (-B\sqrt{\tau}, B\sqrt{\tau})$, $(B\sqrt{\tau} p(u))^{-1} \leq 2\sqrt{\pi} \exp(B^2 + 4\tau^{-1}) B^{-1} \leq 2\sqrt{\pi} \exp(B^2 + 4\tau^{-1})$. Hence for any $k \geq 1$, by Hölder's inequality:

$$\begin{aligned} \int_{-B\sqrt{\tau}}^{B\sqrt{\tau}} \rho(u)^2 du &\leq \left(\int_{-B\sqrt{\tau}}^{B\sqrt{\tau}} |\rho(u)|^{2k} du \right)^{1/k} (2B\sqrt{\tau})^{1-1/k} \\ &\leq \left(\int_{-B\sqrt{\tau}}^{B\sqrt{\tau}} \frac{p(u) |\rho(u)|^{2k}}{2B\sqrt{\tau} \inf_u p(u)} du \right)^{1/k} (2B\sqrt{\tau}) \\ &\leq \left(\frac{8B}{\sqrt{\tau} e} \right) k \left(2\sqrt{2\pi} \exp \left(B^2 + \frac{2K}{\tau} \right) \right)^{1/k}. \end{aligned}$$

We have a free choice of $k \geq 1$ to maximise $k \exp(v/k)$, and since $v \geq 1$, taking $k = v$ means that $k \exp(v/k) \exp(-1) = v$. Hence we obtain a bound of

$$\int_{-B\sqrt{\tau}}^{B\sqrt{\tau}} \rho(u)^2 du \leq \frac{8B}{\sqrt{\tau}} \left(B^2 + \frac{2K}{\tau} + \log(2\sqrt{2\pi}) \right) \leq \frac{8B^3}{\sqrt{\tau}} \left(3 + \frac{2K}{\tau} \right).$$

Lemma 3.3 is proved.

We continue by considering $L_B = \{|x| \leq B\sqrt{\tau}, |y| \leq B\sqrt{\tau}\}$.

Lemma 3.4. *For random variables S, T , let $X = S + Z_S^{(\tau)}$ and $Y = T + Z_T^{(\tau)}$. If S, T satisfy the FKG inequalities, then there exists a function $f_2(\tau, K)$ such that for $B \geq 1$*

$$\mathbf{E} M_{a,b}(X, Y) \tilde{\rho}(X + Y) I((X, Y) \in L_B) \leq f_2(\tau, K) (a + b) B^4 \text{cov}(S, T).$$

P r o o f. Lemma 3 in [13] uses the fact that FKG inequalities imply «positive quadrant dependence» originally due to Lehmann [10]. That is, defining $H(s, t) = \mathbf{P}\{S \geq s, T \geq t\} - \mathbf{P}\{S \geq s\} \mathbf{P}\{T \geq t\}$, we have that S, T are positive quadrant dependent if and only if $H(s, t) \geq 0$ for all s, t , which is a consequence of S, T being FKG. Since $\text{cov}(S, T) = \int H(s, t) ds dt \geq 0$, it follows that

$$\text{cov}(f(S), g(T)) = \int f'(s) g'(t) H(s, t) ds dt \leq \|f'\|_{\infty} \|g'\|_{\infty} \text{cov}(S, T).$$

Since $|\varphi_c(u)'| \leq \exp(-\frac{1}{2})/\sqrt{2\pi c^2}$, and $|(u\varphi_c(u)/c)'| \leq (2\exp(-\frac{3}{2}))/\sqrt{2\pi c^3}$, we deduce that

$$\begin{aligned} |p_{X,Y}(x, y) - p_X(x) p_Y(y)| &\leq \frac{\text{cov}(S, T)}{2\pi\tau^2 e}, \\ |p_{X,Y}^{(1)}(x, y) - p'_X(x) p_Y(y)| &\leq \frac{\text{cov}(S, T)}{\pi\tau^{5/2} e^2}, \\ |p_{X,Y}^{(2)}(x, y) - p_X(x) p'_Y(y)| &\leq \frac{\text{cov}(S, T)}{\pi\tau^{5/2} e^2}. \end{aligned}$$

We can rearrange $M_{a,b}$ to get

$$\begin{aligned} M_{a,b}(x, y) &= a \left(\frac{p_{X,Y}^{(1)}(x, y) - p'_X(x) p_Y(y)}{p_{X,Y}(x, y)} \right) + b \left(\frac{p_{X,Y}^{(2)}(x, y) - p_X(x) p'_Y(y)}{p_{X,Y}(x, y)} \right) \\ &\quad + (a\rho_X(x) + b\rho_Y(y)) \left(\frac{p_X(x) p_Y(y) - p_{X,Y}(x, y)}{p_{X,Y}(x, y)} \right), \end{aligned}$$

and hence, writing c for $\text{cov}(S, T)/(2\pi\tau^{5/2}e^2)$, for $(x, y) \in L_B$:

$$p_{X,Y}(x, y) |M_{a,b}(x, y)| \leq c \left(\sqrt{\tau} e(a\rho_X(x) + b\rho_Y(y)) + 2(a+b) \right).$$

By the Cauchy-Schwarz inequality

$$\begin{aligned} &\int p_{X,Y}(x, y) M_{a,b}(x, y) \tilde{\rho}(x+y) I((x, y) \in L_B) dx dy \\ &\leq c \int \left(\sqrt{\tau} e(a\rho_X(x) + b\rho_Y(y)) + 2(a+b) \right) \tilde{\rho}(x+y) I((x, y) \in L_B) dx dy \\ &\leq c(a+b) \left(\sqrt{2B^4\sqrt{\tau} f_1} \sqrt{16B^4\sqrt{\tau} f_1} + \sqrt{4B^2\tau} \sqrt{16B^4\sqrt{\tau} f_1} \right). \end{aligned}$$

This follows firstly since

$$\begin{aligned} \int \rho_X(x)^2 I((x, y) \in L_B) dx dy &\leq (2B\sqrt{\tau}) \int_{-B\sqrt{\tau}}^{B\sqrt{\tau}} \rho_X(x)^2 dx \\ &\leq (2B\sqrt{\tau}) B^3 f_1(\tau, K), \end{aligned}$$

and secondly

$$\begin{aligned} &\int \tilde{\rho}(x+y)^2 I((x, y) \in L_B) dx dy \\ &\leq \int \tilde{\rho}(x+y)^2 I(|x+y| \leq 2B\sqrt{\tau}) I(|y| \leq B\sqrt{\tau}) dx dy \\ &\leq 2B\sqrt{\tau} \int_{-2B\sqrt{\tau}}^{2B\sqrt{\tau}} \tilde{\rho}(z)^2 dz \leq 16 B^4 \sqrt{\tau} f_1(\tau, K). \end{aligned}$$

Lemma 3.4 is proved.

Lemma 3.5. *For any random variables S, T with mean zero and variance $\leq K$, let $X = S + Z_S^{(\tau)}$ and $Y = T + Z_T^{(\tau)}$. There exists a function $f_3(\tau, K, \varepsilon)$ such that*

$$\mathbf{E} M_{a,b}(X, Y) \tilde{\rho}(X+Y) I((X, Y) \notin L_B) \leq (a+b) \frac{f_3(\tau, K, \varepsilon)}{B^{2-\varepsilon}};$$

for S, T with k -th moment ($k \geq 2$) bounded from above, we can achieve a rate of decay of $1/B^{k-\varepsilon}$.

P r o o f. By Chebyshev

$$\mathbf{P}\{(S + Z_S^{(2\tau)}, T + Z_T^{(2\tau)}) \notin L_B\} \leq \int \frac{p^{(2\tau)}(x, y)(x^2 + y^2)}{2B^2\tau} dx dy \leq \frac{K + 2\tau}{B^2\tau}$$

so by the Hölder–Minkowski inequality for $1/p + 1/q = 1$:

$$\begin{aligned} \mathbf{E}\rho_{X,Y}^{(1)}(X, Y) \tilde{\rho}(X + Y) I((X, Y) \notin L_B) \\ \leq \left(\mathbf{E}|\rho_{X,Y}^{(1)}(X, Y)|^p I((X, Y) \notin L_B) \right)^{1/p} (\mathbf{E}|\tilde{\rho}(X + Y)|^q)^{1/q} \\ \leq c_{\tau,p}^{1/p} c_{\tau,q}^{1/q} \mathbf{P}\{(S + Z_S^{(2\tau)}, T + Z_T^{(2\tau)}) \notin L_B\}^{1/p} \\ \leq \frac{2\sqrt{2} \exp(-1)}{\tau} \left(2 + \frac{K}{\tau} \right) \sqrt{pq} \frac{1}{B^{2/p}}. \end{aligned}$$

By choosing p arbitrarily close to 1, we can obtain a constant term, as required. The other terms work in a similar way. Lemma 3.5 is proved.

Similarly we bound the remaining product term.

Lemma 3.6. *For FKG random variables S, T with mean zero and variance $\leq K$, let $X = S + Z_S^{(\tau)}$ and $Y = T + Z_T^{(\tau)}$. There exist functions $f_4(\tau, K)$ and $f_5(\tau, K)$ such that*

$$\mathbf{E}\rho_X(X) \rho_Y(Y) \leq f_4(\tau, K) B^4 \text{cov}(S, T) + f_5(\tau, K) B^{-2}.$$

P r o o f. We know (see the proof of Lemma 3.4) that $p_{X,Y}(x, y) - p_X(x)p_Y(y) \leq \text{cov}(S, T)/(2\pi e\tau^2)$. Hence by an argument similar to those of Lemmas 3.5 and 3.6, we obtain that

$$\begin{aligned} \mathbf{E}\rho_X(X) \rho_Y(Y) &= \int (p_{X,Y}(x, y) - p_X(x)p_Y(y)) \rho_X(x) \rho_Y(y) dx dy \\ &\leq \frac{\text{cov}(S, T)}{2\pi e\tau^2} \int |\rho_X(x)| |\rho_Y(y)| I((x, y) \in L_B) dx dy \\ &\quad + \int p(x, y) |\rho_X(x)| |\rho_Y(y)| I((x, y) \notin L_B) dx dy \\ &\quad + \int p(x)p(y) |\rho_X(x)| |\rho_Y(y)| I((x, y) \notin L_B) dx dy \\ &\leq \frac{\text{cov}(S, T)}{2\pi e\tau^2} \left(\int_{-B\sqrt{\tau}}^{B\sqrt{\tau}} |\rho_X(x)|^2 dx \right)^2 \\ &\quad + 2 \left(\int p_{X,Y}(x, y) |\rho_X(x)|^2 I((x, y) \notin L_B) dx dy \right) \end{aligned}$$

as required.

P r o o f o f T h e o r e m 2.1. Combining Lemmas 3.4, 3.5, and 3.6 we obtain for given K, τ, ε that there exist constants C_1, C_2 such that

$$\mathbf{E}M_{\sqrt{\beta}, \sqrt{1-\beta}} \tilde{\rho} + \sqrt{\beta(1-\beta)} \mathbf{E}\rho_X \rho_Y \leq C_1 \text{cov}(S, T) B^4 + \frac{C_2}{B^{2-\varepsilon}},$$

so taking $B = (K/\text{cov}(S, T))^{1/6} > 1$, we get a bound of $C(\text{cov}(S, T))^{1/3-\varepsilon}$.

By Lemma 3.5, note that if X, Y have bounded k -th moment, then we obtain decay at the rate $C_1 \text{cov}(S, T) B^4 + C_2/B^{k'}$, for any $k' < k$. Choosing $B = (\text{cov}(S, T))^{-1/(k'+4)}$, we obtain a rate of $(\text{cov}(S, T))^{k'/(k'+4)}$. Hence for $k \rightarrow \infty$ we can find a rate arbitrarily close to $\text{cov}(S, T)$.

REFERENCES

1. Barron A. R. Entropy and the central limit theorem. — Ann. Probab., 1986, v. 14, № 1, p. 336–342.
2. Brown L. D. A proof of the central limit theorem motivated by the Cramér–Rao inequality. — Statistics and Probability: Essays in Honour of C. R. Rao. Ed. by G. Kallianpur, P. R. Krishnaiah, J. K. Ghosh. Amsterdam–New York: North-Holland, 1982, p. 141–148.
3. Carlen E. A., Soffer A. Entropy production by block variable summation and central limit theorems. — Commun. Math. Phys., 1991, v. 140, № 2, p. 339–371.
4. Gnedenko B. V., Korolev V. Y. *Random Summation: Limit Theorems and Applications*. Boca Raton: CRC Press, 1996, 267 p.
5. Grimmett G. R. Percolation. Berlin: Springer-Verlag, 1999, 444 p.
6. Johnson O. T. Entropy inequalities and the central limit theorem. — Stochastic Process. Appl., 2000, v. 88, № 2, p. 291–304.
7. Johnson O. T., Barron A. R. Fisher information inequalities and the central limit theorem. — Probab. Theory Related Fields, 2004, v. 129, № 3, p. 391–409.
8. Johnson O. T., Suhov Y. M. Entropy and random vectors. — J. Statist. Phys., 2001, v. 104, № 1–2, p. 145–165.
9. Lebowitz J. L. GHS and other inequalities. — Comm. Math. Phys., 1974, v. 35, p. 87–92.
10. Lehmann E. L. Some concepts of dependence. — Ann. Math. Statist., 1966, v. 37, p. 1137–1153.
11. Линник Ю. В. Теоретико-информационное доказательство центральной предельной теоремы в условиях Линдберга. — Теория вероятн. и ее примен., 1959, т. 4, в. 3, с. 311–321.
12. Newman C. M. Moment inequalities for ferromagnetic Gibbs distributions. — J. Math. Phys., 1975, v. 16, № 9, p. 1956–1959.
13. Newman C. M. Normal fluctuations and the FKG inequalities. — Comm. Math. Phys., 1980, v. 74, № 2, p. 119–128.
14. Shimizu R. On Fisher's amount of information for location family. — A Modern Course on Statistical Distributions in Scientific Work. V. 3: Characterizations and Applications. Ed. by G. P. Patil, S. Kotz, and J. K. Ord. Dordrecht: Reidel, 1975, p. 305–312.
15. Takano S. The inequalities of Fisher information and entropy power for dependent variables. — Proceedings of the 7th Japan–Russia Symposium on Probability Theory and Mathematical Statistics (Tokyo, 1995). Ed. by S. Watanabe et al. River Edge: World Scientific, 1996, p. 460–470.
16. Takano S. Entropy and a limit theorem for some dependent variables. — Prague Stochastics '98. V. 2. Prague: Union of Czech Mathematicians and Physicists, 1998, p. 549–552.

Поступила в редакцию
12.X.2001