Joint Asymptotic Normality

By Lucien Le Cam,

University of California, Berkeley

1. Introduction. The purpose of the present paper is to answer a question about joint asymptotic normality of random variables (X_n, Y_n) when the marginals are asymptotically normal. The result given here is easy, but we do not know of adequate references. It answers a question that continues to be asked by students year after year.

Simple examples will show that a pair of variables (X,Y) may have Gaussian marginals, be uncorrelated and yet be far from jointly Gaussian. So, to obtain joint asymptotic normality one needs some restriction. The one used here is approximate infinite divisibility. This occurs often for sums of independent or nearly independent summands and is obviously necessary.

2. Notations and Definitions.

We shall work in the plane metrized by the maximum coordinate norm. Extension to higher dimensions is easy. If P is a measure on the plane, its first coordinate marginal will be P' and the second marginal will be P''. For two measures P and Q their (Kolmogorov-style) distance will be $K(P,Q) = \sup_A |P(A) - Q(A)|$ where A ranges of the class $\mathcal{A}$ of all rectangles with sides parallel to the coordinate axes. For $A \in \mathcal{A}$ and $\varepsilon > o$, let A^ε be the set of points whose distance to A does not exceed ε.

The (Lévy-style) distance of P,Q is a number $L(P,Q)$, infimum of the $\varepsilon > o$ such that $P(A) \leq Q(A^{\varepsilon}) + \varepsilon$ and $Q(A) \leq P(A^{\varepsilon}) + \varepsilon$ for all $A \in \mathcal{A}$.

We shall also need a cross between the metrics K and L. To define it if $A \in \mathcal{A}$, let $A(o,\varepsilon)$ be the set of pairs (x,y) such that there is a pair $(x,v) \in A$ with $|y-v| \leq \varepsilon$. Take for $KL(P,Q)$ the infimum of the numbers $\varepsilon > o$ such that $P(A) \leq Q[A(o,\varepsilon)] + \varepsilon$ and $Q(A) \leq P[A(o,\varepsilon)] + \varepsilon$ for all $A \in \mathcal{A}$.

3. The theorem for the Kolmogorov distance.

THEOREM 1. *Given an $\varepsilon > o$ there is a $\delta > o$ with the following property.*

Let P be a probability measure on the plane such that

a) *There is a Gaussian measure G' such that $K(P',G') < \delta$.*

b) *There is a Gaussian measure G'' such that $K(P'',G'') < \delta$.*

c) *There is an infinitely divisible Q such that $K(P,Q) < \delta$.*

Then there is a Gaussian G such that $K(P,G) < \varepsilon$.

Proof. Assume that the statement is incorrect. Then there is some $\varepsilon_o > o$ for which it is violated. If so, for each integer n one can find a measure P_n such that $\max\{K(P_n',G_n'), K(P_n'',G_n'')\} < 1/n$ for some Gaussian measures G_n' and G_n'' but $K(P_n,G) \geq \varepsilon_o$ for all Gaussian G.

The distance K is invariant by shifts and by scale transformations on the separate coordinate axes. Thus one may assume that G_n' and G_n'' are both $\mathcal{N}(0,1)$. If so, the sequence $\{P_n\}$ is relatively compact for the Lévy distance L. Taking a subsequence if necessary, one can assume that $L(P_n,P) \to o$ for some P. By construction P' and P'' are $\mathcal{N}(0,1)$. Given any $\varepsilon > o$ one can find an $\alpha > o$ such that $P(A^{\alpha}) \leq P(A) + \varepsilon$ for all rectangles $A \in \mathcal{A}$. Thus $L(P_n,P) \to o$ implies that $K(P_n,P) \to o$. Since $K(P_n,G) \geq \varepsilon_o$ for any Gaussian G one concludes that $K(P,G) \geq \varepsilon_o$ for all G. However P is infinitely divisible with Gaussian marginals. Thus it must be a Gaussian measure. This contradition proves the desired result.

4. The theorem for the Lévy distance.

THEOREM 2. The statement in Theorem 1 remains valid if the Kolmogorov distance K is everywhere replaced by the Lévy distance L.

Proof. The Lévy distance L is shift but not scale invariant. Proceed as in the proof of Theorem 1, obtaining a sequence P_n with $L(P_n,Q_n) < 1/n$ for some infinitely divisible Q, $L(P_n',G_n') < 1/n$, $L(P_n'',G_n'') < 1/n$ but $L(P_n,G) \geq \varepsilon_o$ for all Gaussian G. Here we can assume that G_n' is $N(o,\sigma_n^2)$ and G_n'' is $N(o,\tau_n^2)$. Now suppose that there exists a number $b > o$ such that $\min(\sigma_n,\tau_n) > b$ for all n. Then, just as in the proof of Theorem 1, convergence in Lévy distance implies convergence in Kolmogorov distance. Thus this case is covered by Theorem 1.

A second possibility is that one of the two variances, say σ_n^2, stays larger than a fixed $b > o$ but $\tau_n^2 \to o$. (The σ_n may tend to infinity.)

In such a case convergence of $L(P_n,Q_n)$ to zero implies that the cross metric $KL(P_n,Q_n)$ also tends to zero. Since KL is invariant by scale changes on the first coordinates, one can change scale and assume that $\sigma_n = 1$. Thus one is again reduced to the relatively compact case and the argument of Theorem 1 applies. If both $\sigma_n \to o$ and $\tau_n \to o$ the situation is trivial. This completes the proof of the theorem.

5. Concluding remarks.

Note that Theorems 1 and 2 have been stated as approximation theorems, without assuming existence of limits. Of course convergence to normality of marginal distributions does not imply convergence of the joint distribution, because the covariance of the approximating Gaussian may vary. However convergence of the marginals implies relative compactness of the joint distributions. Then the Lévy distances can be replaced by any topologically equivalent metrics, such as the Prokorov or the dual Lipschitz metrics. We do not know however whether a result such as Theorem 1 holds for these metrics. A result such as Theorem 1 or 2 can be formulated for classes other than the class A of rectangles with sides parallel to the coordinate axes. We do not know for what classes the theorems are true.

In some cases one wants to know the covariance structure of the approximating Gaussian measures. A simple result is as follows.

Suppose that the marginal P' is approximated by a Gaussian G' that has the same mean and variance as P'. Suppose that the analogous result holds for P''. Then the joint Gaussian G approximating P can be taken with the same means and covariance structure as P.

To show this one can argue as in Theorems 1 and 2 that it is enough to look at relatively compact sequences. Then the variances of the approximating Gaussian marginals can also be replaced by truncated variances of P'_n and P''_n respectively. The approximability of the covariances then follows from the inequality $|xy| \leq \frac{1}{2}[x^2+y^2]$.

Theorems 1 and 2 are expressions of the fact that an infinitely divisible distribution cannot be close to Gaussian unless its Lévy measure is concentrated near the origin. One could attempt to make this more precise and give a relation between the ε and δ of Theorem 1.

We conjecture that it may be possible to prove a result of the following nature:

Conjecture. There is a constant C such that if (a)(b)(c) of Theorem 1 hold and if $0 < \delta < C\varepsilon^5$ then $K(P,G) < \varepsilon$.

(This is based on a rough argument to the effect that the Lévy measure M' of Q' satisfies

$$M' \{x; |x| \geq \tau\} \leq C, \delta\tau^{-4}$$

for $\tau \in (0,1]$. We have not actually carried out the proof because we do not know what are the best achievable results.)