The Annals of Probability
1974, Vol. 2, No. 1, 131–135

ON CONSTRUCTIVE CONVERGENCE OF MEASURES ON THE REAL LINE[1]

BY Y. K. CHAN

University of Washington

This is a discussion how Lévy's Continuity Theorem can be proved without idealistic compactness arguments, and how it can then be used to give constructive proofs to some limit theorems.

Modern probability theory, as opposed to the classical works including those of Lévy and Wiener, contain theorems which are devoid of numerical content, despite their elegance and power. This perhaps reflects the prevalent situation in mathematics in general. But one certainly expects a discipline so application oriented to be more computational.

As elucidated by Brouwer and more recently Bishop, the source of non-constructivity lies in the unrestricted use of the principle of the excluded middle in order to avoid honest (and sometimes hard) constructions.

It has often been argued that platonic theorems, non-constructive as they are, are first steps to obtain constructive versions. The fact is that they are steps in the wrong direction. Never has a purely idealistic proof contributed anything towards the construction.

When a constructivist looks at theorems already in existence, the most interesting ones are perhaps those whose computational intents are so vague that finding the proper constructive interpretation alone poses challenging questions. Nonetheless, it is perphaps illustrative to look at some examples where the numerical interpretation is clear cut, and compare a classical (i.e., idealistic) proof to a constructive one. One such example concerns the convergence of measures on the real line.

Suppose one wants to prove that a sequence of probability measures μ_n on R converges vaguely to a limit μ, i.e., $\mu_n(g) \to \mu(g)$ for all bounded continuous function g on R. One way is to prove that (A) every subsequence of μ_n contains a convergent subsequence, and that (B) the limits for all the convergent subsequences must be the same. Suppose we have such a proof. Take a specific bounded continuous function g and a positive real number ε. Armed with the proof, one sets out to find an integer n_0 such that $|\mu_n(g) - \mu_m(g)| < \varepsilon$ for all $m, n \geqq n_0$. But no matter how closely the proof is examined it does not provide this integer. So this proof, that the limit exists, does not mean a prescription for the construction of the desired integer $n_0(\varepsilon, g)$. It means something else. (It is unclear what that something is.) It is not hard to see why this proof is

Received June 26, 1972; revised June 25, 1973.

[1] Research partially supported by NSF Grant GP-29233.

AMS 1970 *subject classifications.* Primary 60B10; Secondary 02E99.

Key words and phrases. Constructive mathematics; convergence of measures on the real line.

non-constructive. Indeed, statement (A) is proved by showing that it is equivalent to the statement (C) that every bounded sequence of real numbers has a convergent subsequence. But a constructive proof of (C) would yield a finite routine to prove or disprove all statements of a certain type, Fermat's last theorem included (see [1]).

A second way to prove vague convergence $\mu_n \to \mu$ is to prove that the characteristic function (Fourier transform) f_n of μ_n converges to some f and then use Lévy's Continuity Theorem to conclude that μ_n must converge to some μ which has characteristic function f. The question, then, is whether there is a constructive proof for Lévy's theorem. If one examines the proofs given for Lévy's theorem in the standard textbooks, one finds that they are of the first kind described above! ([2] and the references there.) Fortunately, on the other hand, Lévy's original proof in terms of a metric (now known as Lévy's metric) on distribution functions ([5] page 199) is constructive. A reformulation of Lévy's theorem is given below.

LEMMA. *Let f_1 and f_2, the characteristic functions of μ_1 and μ_2 respectively, have the common modulus of continuity γ. Let g be a continuous function on R satisfying $|g| \leqq 1$, and having modulus of continuity β. Then for every $\varepsilon > 0$, there exist positive numbers δ and θ, depending on ε, β, γ and otherwise independent of μ_1, μ_2, f_1, f_2 and g, such that*

$$|\mu_1(g) - \mu_2(g)| < \varepsilon$$

provided

$$|f_1 - f_2| < \delta \qquad \text{on } [-\theta, \theta] .$$

(For a continuous function g on R, a modulus of continuity β means an operation which assigns to every $\varepsilon > 0$ and $a > 0$ a real $\beta(\varepsilon, a) > 0$ such that $|g(x) - g(y)| < \varepsilon$ if $|y| \leqq a$, $|x| \leqq a$ and $|x - y| < \beta$. If the continuity is uniform, β is independent of a. Thus, since characteristic functions are uniformly continuous, γ is understood to be dependent only on ε.) The lemma contains the essence of Lévy's theorem. To be precise, let μ_n be a sequence of probability measures on R whose characteristic functions f_n converges uniformly on compact intervals to the function f. Let r be the rate of convergence, (i.e., an operation assigning to every given $\varepsilon > 0$ and $a > 0$ a positive integer $r(\varepsilon, a)$ such that $|f_m - f_n| < \varepsilon$ on $[-a, a]$ whenever $m, n \geqq r$). Let γ denote a common modulus of continuity for the functions f_n. (The functions f_n must be equicontinuous. They certainly are equicontinuous on $[-1, 1]$. Thus if h is small enough, one can make $\int (1 - \cos hx)\, d\mu_n(x) = |\mathrm{Re}\,(f_n(0) - f_n(h))|$ arbitrarily small uniformly over n. Equicontinuity then follows from

$$\begin{aligned} |f_n(t) - f_n(t+h)| &= \left|\int (e^{itx} - e^{i(t+h)x})\, d\mu_n(x)\right| \\ &\leqq \int |1 - e^{ihx}|\, d\mu_n(x) \\ &\leqq \int (1 - \cos hx)\, d\mu_n(x) + \int |\sin hx|\, d\mu_n(x) \\ &\leqq \int (1 - \cos hx)\, d\mu_n(x) + \varepsilon + \varepsilon^{-1} \int (1 - \cos hx)\, d\mu_n(x) . \end{aligned}$$

The last inequality can be proved for all $\varepsilon > 0$ by elementary means.) Let g be a continuous function R satisfying $|g| \leqq 1$ and having modulus of continuity β. Then the lemma can be used: for every $\varepsilon > 0$

$$|\mu_m(g) - \mu_n(g)| < \varepsilon$$

whenever $m, n \geqq r(\delta(\varepsilon, \beta, \gamma), \theta(\varepsilon, \beta, \gamma))$. In particular μ_n converges vaguely to some probability measure μ on R. This is Lévy's theorem.

A sketch of an elementary proof of the lemma is as follows. First, the probability measure μ_i puts almost all its mass in some finite interval $[-a, a]$. On this finite interval, the integral of g with respect to μ_i is close to that with respect to the convolution $\mu_i * \varphi_r$ (where φ_r is the normal density $(2\pi r)^{-\frac{1}{2}}e^{-x^2/2r^2}$), provided that r is small enough. So it suffices to prove that the integrals of g on this interval with respect to $\mu_1 * \varphi_r$ and $\mu_2 * \varphi_r$ differ by very little. This is easy because $\mu_i * \varphi_r$ has a density conveniently expressed in terms of f_i ([4] page 507). So all that is necessary is to make sure that f_1 and f_2 differ very little where they should. Following these steps, one can find $\delta(\varepsilon, \beta, \gamma)$ and $\theta(\varepsilon, \beta, \gamma)$.

Lévy's theorem is easily generalized to the case when μ_n are nonnegative measures on R (not necessarily with equal total mass). Suppose $f_n \to f$ on compact intervals. Classically one would first determine whether $f(0) = 0$ or $f(0) > 0$. In the first case, the total mass $f_n(0) = \mu_n(R)$ converges to 0 and so μ_n trivially converges to the 0 measure. In the second case, one would consider $f_n/f_n(0)$ which are the characteristic functions for probability measures. Now since the dichotomy $f(0) = 0$ or $f(0) > 0$ does not hold constructively, one has to use a little care. But the argument is essentially the same. Let $\varepsilon > 0$ and let g be a continuous function with $|g| \leqq 1$. One does have the "dichotomy": either $f(0) < \varepsilon/2$ or $f(0) > 0$ ([1] page 24). In the first case there exists n_1 such that the total masses $f_n(0) < \varepsilon/2$ if $n \geqq n_1$, which implies that $|\mu_n(g) - \mu_m(g)| < \varepsilon$ if $n, m \geqq n_1$. In the second case Lévy's theorem is applicable to $f_n/f_n(0)$ and yields an integer n_2 such that $|\mu_n(g) - \mu_m(g)| < \varepsilon$ if $n, m \geqq n_2$. Combining, we see that $\mu_n(g)$ converges; i.e., μ_n converges vaguely.

Now consider two examples where Lévy's theorem is used. The first is the simplest version of the Central Limit Theorem. Let X_n be a sequence of independent identically distributed random variables which have first moment equal to 0 and a finite second moment $\sigma^2 > 0$. Then the Central Limit Theorem asserts that the properly normalized partial sum $(X_1 + \cdots + X_n)/n^{\frac{1}{2}}\sigma$ has probability measure converging vaguely to the normal distribution. (The probability measure μ_X associated with a random variable X is defined by $\mu_X(g) = E(g(X))$ for all bounded continuous function g.) One should of course be careful to interprete the hypothesis constructively. "X has second moment" means "X^2 is integrable" and not "$E(X^2) = \infty$ implies a contradiction." One needs not go into details of constructive measure theory here. It suffices to say that the integrability of X^2 implies the existence of an operation $r(\varepsilon)$ such that

$$E(X^2; X^2 > r(\varepsilon)) < \varepsilon ;$$

similarly for the first moment. Under this circumstance, one can prove that

$$f(t) = 1 - \frac{\sigma^2}{2} t^2 + o(t^2) .$$

It should be emphasized that as $t \to 0$ the expression $o(t^2)/t^2$ converges to zero in the constructive sense—the rate of convergence for $o(t^2)/t^2$ comes from the operation r.

In view of the independence of X_n, the normalized partial sum $(X_1 + \cdots + X_n)/n$ has characteristic function

$$f\left(\frac{t}{n^{\frac{1}{2}}\sigma}\right)^n = \left\{1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n\sigma^2}\right)\right\}^n$$

which converges on compact intervals to $e^{-t^2/2}$, the characteristic function for the normal probability measure. Lévy's theorem now applies to yield the Central Limit Theorem. Thus, given a bounded continuous function g and a positive $\varepsilon > 0$, one can tell how large n must be in order for $E(g((X_1 + \cdots + X_n)/n^{\frac{1}{2}}\sigma))$ to differ from $(2\pi)^{-\frac{1}{2}} \int g(x)e^{-x^2/2}\, dx$ by less than ε—one starts with the operation r and eventually produces an integer n_0 so that every $n \geqq n_0$ is large enough. This proof is of course well known. The point is that n_0 can be computed now that Lévy's theorem is constructive.

The second example concerns an infinitely divisible distribution, i.e., a probability measure μ on R with the property that for every $\varepsilon > 0$ the characteristic function f can be written in the form

$$f = \prod_1^n f_j$$

when $f_1, \cdots, f_n$ are characteristic functions with $|1 - f_j| < \varepsilon$ on $[-\varepsilon^{-1}, \varepsilon^{-1}]$. The characteristic functions for infinitely divisible distribution are characterized by the Lévy–Khintchine formula

$$\text{(1)} \qquad \log f(t) = it\gamma + \int_{-\infty}^{\infty} \left(e^{it\lambda} - 1 - \frac{it\lambda}{1 + \lambda^2}\right) \frac{1 + \lambda^2}{\lambda^2}\, dG(\lambda)$$

when γ is some real number and G is a finite, nonnegative measure on R. Given $f(t)$, consider the analytic proof for the existence of G and γ in ([3] page 134). First, for every $\varepsilon > 0$, a real number γ and a measure G_ε on R whose total mass is uniformly bounded in ε are constructed satisfying

$$\text{(2)} \qquad \log f(t) = it\gamma_\varepsilon + \int_{-\infty}^{\infty} \left(e^{it\lambda} - 1 - \frac{it\lambda}{1 + \lambda^2}\right) \frac{1 + \lambda^2}{\lambda^2}\, dG_\varepsilon(\lambda) + o_{\varepsilon,t}(1)$$

where $o_{\varepsilon,t}(1) \to 0$ uniformly on compact t-intervals as $\varepsilon \to 0$. Up to this point the proof is constructive. Then the compactness argument (A) described at the beginning of this paper is invoked to select a subsequence of G_ε which converges to some G. It follows from (2) that the corresponding subsequence of γ_ε converges to some γ and the Lévy–Khintchine formula is satisfied by G and γ. This measure G is later shown to be unique. As observed earlier, such a proof is not constructive. For example, the proof does not indicate a routine for finding the integral of a given bounded continuous function g with respect to the measure G. But the use of Helly's Theorem (A) is really not essential. Modifying the uniqueness proof given in ([3] page 134) and using Lévy's theorem, one can obtain a constructive proof, as follows. Subtract from (2) its average on $[t-1, t+1]$:

$$\log f(t) - \tfrac{1}{2}\int_{t-1}^{t+1} \log f(s)\,ds = \int_{-\infty}^{\infty} e^{it\lambda}\left(1 + \frac{\sin\lambda}{\lambda}\right)\frac{1+\lambda^2}{\lambda^2}\,dG_\varepsilon(\lambda) + o'_{\varepsilon,t}(1) \tag{3}$$

$$\equiv \int_{-\infty}^{\infty} e^{it\lambda}\,dH_\varepsilon(\lambda) + o'_{\varepsilon,t}(1)\,.$$

From this we see that the characteristic function of H_ε converges (uniformly on compact t-intervals) as $\varepsilon \to 0$. So Lévy's continuity theorem can be applied to yield the desired limit H for H_ε and the corresponding limit G for G_ε. In particular, if one desires the integral $G(g)$, one can compute a small enough ε, and $G_\varepsilon(g)$ will be a good approximate to $G(g)$.

REFERENCES

[1] BISHOP, E. (1967). *Foundations of Constructive Analysis*. McGraw-Hill, New York.
[2] CHUNG, K. L. (1968). *A Course in Probability Theory*. Harcourt, Brace & World, New York.
[3] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.
[4] FELLER, W. (1967). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.
[5] LÉVY, P. (1925). *Calcul des probabilités*. Gauthier-Villars, Paris.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195