ON REGULAR VARIATION AND LOCAL LIMIT THEOREMS

WILLIAM FELLER
PRINCETON UNIVERSITY

1. Introduction

Recent work on limit theorems in probability is marked by two tendencies. The old limit theorems are being supplemented and sharpened by a variety of so-called local limit theorems (which sometimes take the form of asymptotic expansions). Even more striking is the increasing role played by functions of regular variation. They made their debut in W. Doblin's pioneer work of 1940 where he gave a complete description of the domains of attraction of the nonnormal stable distributions. A long series of investigations started by E. B. Dynkin and continued by J. Lamperti, S. Port, and others have shown that essential limit theorems connected with renewal theory depend on regular variation. The same is true of the asymptotic behavior of the maximal term of a sequence of independent random variables and of D. A. Darling's theorems concerning the ratio of this term to the corresponding partial sum.

It seems that each of these problems still stands under the influence of its own history and that, therefore, a great variety of methods is used. Actually a considerable unification and simplification of the whole theory could be achieved by a systematic exploitation of two powerful tools: J. Karamata's beautiful theory of regular variation and the method of estimation introduced by A. C. Berry in his well-known investigation of the error term in the central limit theorem.

[It seems that proofs of Karamata's theorems can be found only in his paper of 1930 in the Rumanian journal *Mathematica* (Vol. 4), which is not easily accessible. For purposes of probability theory, one requires a generalization from Lebesgue to Stieltjes integrals. A streamlined version is contained in the forthcoming second volume of my *Introduction to Probability Theory*, but this book does not contain the inequalities derived in the sequel.]

Berry's method is of wide applicability and not limited to the normal distribution. It leads to an estimate for the discrepancy between distributions in terms of the discrepancy between the corresponding characteristic functions. In the case of the normal distribution, the latter discrepancy can be estimated in terms of the moments, and the theory of regular variation leads readily to similar

Research supported by Army Research Office (Durham) Project in Probability at Princeton University.

estimates, in the case of convergence to other stable distributions. For example, the asymptotic expansions connected with the central limit theorem have their natural analogues in this general setting.

[See H. Cramér, "On the approximation to a stable probability distribution," pp. 70–76 in *Studies in Mathematical Analysis and Related Topics*, Stanford University Press, 1962, and "On asymptotic expansions for sums of independent random variables with a limiting stable distribution," *Sankhyā*, Vol. 25 (1963), pp. 13–24. The basic relations for regularly varying functions make it possible to avoid Cramér's severe restrictions simplifying at the same time the calculations.] Indeed, we shall see that expansions and error estimates of this sort are more general than the basic limit theorems in the sense that they may apply even when the leading terms do not converge. A typical example is treated in section 9.

We proceed to a brief sketch of the basic properties of regular variation and of its connection with limit theorems. Part of the material of sections 2–7 will be contained in the second volume of my book on probability, but encumbered by details and spread over many places. For a better understanding of the whole theory we shall in this address generalize the notion of regular variation by considering inequalities instead of equalities (section 7). In probabilistic terms, we shall replace the condition that a sequence of distributions F_n converges to a limit by the requirement that it be compact in the sense of the following.

DEFINITION 1. *A family $\{F_n\}$ of probability distributions is stochastically compact if every sequence $\{F_{n_k}\}$ contains a further subsequence converging to a probability distribution not concentrated at one point.*

It is essential that the limit be nondegenerate. We shall see that compactness is related to our one-sided regular variation much in the same way as convergence is to regular variation. Furthermore, the typical local limit theorems and error estimates depend on compactness rather than actual convergence, and they can be formulated in this more general setting.

2. The basic property of regular variation

A positive finite valued function U defined on $(0, \infty)$ is said to *vary regularly at* ∞ if for each x

$$\frac{U(tx)}{U(t)} \to x^p, \qquad t \to \infty \tag{2.1}$$

where p is a constant. If this exponent is 0, one says that U *varies slowly at* ∞. In other words, for a slowly varying function one has

$$\frac{\mathfrak{L}(tx)}{\mathfrak{L}(t)} \to 1, \qquad t \to \infty, \tag{2.2}$$

and U varies regularly iff it is of the form

$$U(x) = x^p \mathfrak{L}(x). \tag{2.3}$$

Regular variation at 0 is defined in like manner, that is, U varies regularly at 0 iff $U(1/x)$ varies regularly at ∞. [Since regular variation at ∞ is not affected by the behavior of U in a finite interval, it suffices to assume that U is defined and positive in some interval (a, ∞). By the same token, regular variation at 0 is a local property.]

At first sight the condition (2.1) appears rather artificial, but it may be replaced by the more natural condition that

$$\frac{U(tx)}{U(t)} \to \psi(x). \tag{2.4}$$

Indeed, if such a limit exists and does not vanish identically, then it is either nonmeasurable or of the form $\psi(x) = x^p$. We prove this assertion together with a variant streamlined for applications to probabilistic limit theorems. It refers to convergence of monotone functions and, as usual, it is understood that convergence need take place only at point of continuity of the limit.

LEMMA 1. (a) *Let U be positive and monotone, and suppose that there exists a sequence of numbers $a_n \to \infty$ such that*

$$nU(a_n x) \to \psi(x) > 0. \tag{2.5}$$

Then $\psi(x) = Cx^p$ and U varies regularly at ∞.

(b) *The same conclusion holds if U and ψ are assumed continuous. (It suffices that ψ is finite valued and positive in some interval. The coefficients n may be replaced by arbitrary $\lambda_n > 0$ such that $\lambda_{n+1}/\lambda_n \to 1$.)*

PROOF. If ψ is monotone there is no loss of generality in assuming that 1 is a point of continuity and $\psi(1) = 1$. For fixed t determine n as the *last* index such that $a_n \leq t$. Then $U(t)$ lies between $U(a_n)$ and $U(a_{n+1})$. Since $nU(a_n) \to 1$, it follows easily that (2.4) holds. But then the relation

$$\frac{U(txy)}{U(t)} = \frac{U(txy)}{U(ty)} \frac{U(ty)}{U(t)} \tag{2.6}$$

implies that $\psi(xy) = \psi(x)\psi(y)$. This equation differs only notationally from the famous Hamel equation, and its unique measurable solution is given by $\psi(x) = x^\alpha$. Part (b) is proved in like manner.

The most usual applications in probability theory refer to the truncated second moment

$$U(x) = \int_{-x}^{x} y^2 F\{dy\}, \qquad x > 0 \tag{2.7}$$

of a probability distribution F, or to the tail sum

$$T(x) = 1 - F(x) + F(-x-). \tag{2.8}$$

Generally speaking, a distribution with a regularly varying tail sum T is well-behaved, except if T is slowly varying. This exceptional role is illustrated by the following example which shows that distributions with slowly varying tails can exhibit severe pathologies.

EXAMPLE. Let F be a distribution concentrated on $(0, \infty)$ (that is, let

$F(0) = 0$) and denote its characteristic function by $\varphi = u + iv$. If F has a finite expectation μ then $v'(0) = \mu$, and hence $v(\zeta) > 0$ for all sufficiently small positive values of ζ. This is not so if $\mu = \infty$. Indeed, we shall exhibit *an arithmetic distribution F with slowly varying tail $1 - F$ such that φ has infinitely many zeros accumulating to* 0. *Furthermore,*

$$\limsup_{\zeta \to 0+} \frac{v(\zeta)}{\zeta} = \infty, \qquad \liminf_{\zeta \to 0-} \frac{v(\zeta)}{\zeta} = -\infty. \tag{2.9}$$

In other words, the values of φ oscillate wildly, and the curious nature of these oscillations becomes clear if one reflects that the integral of v over any positive interval $(0, a)$ is strictly positive. The set at which $v(\zeta)/\zeta$ is strongly negative is therefore rather sparse.

To obtain the desired example choose an integer of the form $a = 4\nu + 1$, and let F attribute weight $1/(n(n+1))$ to the point a^n, $(n = 1, 2, \cdots)$. Then

$$\varphi(\zeta) = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \exp(ia^n\zeta). \tag{2.10}$$

If $\zeta = 3\pi/2$, then $a^n\zeta$ is congruent $(3/2)\pi$ modulo 2π, and hence for every positive integer r,

$$\varphi\left(a^{-r}\tfrac{3}{2}\pi\right) = \sum_{n=1}^{r-1} \frac{1}{n(n+1)} \exp\left(ia^{n-r}\tfrac{3}{2}\pi\right) - \frac{i}{r}. \tag{2.11}$$

Since $|\sin t| \le |t|$, it is obvious that as $r \to \infty$,

$$v\left(a^{-r}\tfrac{3}{2}\pi\right) = -\frac{1}{r} + O\left(\frac{1}{r^2}\right), \tag{2.12}$$

which proves the second relation in (2.9). The first one is even easier to verify.

3. Applications

(a) *Distribution of maxima.* Let $X_1, X_2, \cdots$ be independent random variables with a common distribution function F such that $F(x) < 1$ for all x. Let $M_n = \max[X_1, \cdots, X_n]$. We inquire whether there exist constants $a_n \to \infty$ such that the random variables M_n/a_n have a nondegenerate limit distribution $\mathcal{G}$, that is,

$$F^n(a_n x) \to \mathcal{G}(x) \tag{3.1}$$

at points of continuity. By assumption there exist values $x > 0$ for which $0 < \mathcal{G}(x) < 1$, and at such points (3.1) is equivalent to

$$n[1 - F(a_n x)] \to -\log \mathcal{G}(x). \tag{3.2}$$

By lemma 1, therefore, *the possible limits are of the form*

$$\begin{aligned} \mathcal{G}(x) &= e^{-Cx^{-\alpha}}, \qquad & x > 0, \\ &= 0, & x < 0. \end{aligned} \tag{3.3}$$

A limit distribution exists iff $1 - F$ *varies regularly.* Indeed, the lemma asserts the necessity of regular variation, and the sufficiency is obvious on choosing a_n such that $n[1 - F(a_n)] \to 1$. What first appeared to be a relatively deep theorem thus becomes a simple corollary to a simple lemma.

(b) *Stable distributions and their domains of attraction.* Let $X_1, X_2, \cdots$ be independent random variables, with the common distribution F and characteristic function $\varphi = u + iv$. Put $S_n = X_1 + \cdots + X_n$. The variables X_j are said to belong to the domain of attraction of a nondegenerate distribution $\mathcal{G}$ iff there exist real constants $a_n > 0$ and b_n such that the distributions of $a_n^{-1}S_n - b_n$ tend to $\mathcal{G}$. If F is symmetric one can put $b_n = 0$. Then φ is real, and convergence takes place iff

$$\varphi^n\left(\frac{\zeta}{a_n}\right) \to \gamma(\zeta), \tag{3.4}$$

where γ is a continuous function. It is obvious that $a_n \to \infty$, and so (3.4) refers to the behavior of φ near the origin. Otherwise there is no essential difference between (3.4) and (3.1), and we see again that for $\zeta > 0$, the limit γ is necessarily of the form $\gamma(\zeta) = \exp(-C\zeta^\alpha)$ where C and α are positive constants. This solves the problem as far as symmetric distributions are concerned: *only stable characteristic functions have a domain of attraction, and for real φ a relation* (3.4) *holds iff* $1 - \varphi$ *varies regularly at the origin.*

In the case of asymmetric distributions, the same argument still applies to the real parts of φ and $\log \gamma$. Now it is well known that for $\zeta > 0$ a stable characteristic function is not necessarily of the form $\log \gamma(\zeta) = -(a + ib)\zeta^\alpha$, but may be also of the form

$$\log \gamma(\zeta) = -a\zeta + i(b\zeta + c\zeta \log \zeta). \tag{3.5}$$

The appearance of the logarithm on the right would seem to preclude the application of lemma 1, but this is not so. We proceed to show that a minor modification of lemma 1 explains the general form of stable characteristic functions. If the random variables $a_n^{-1}S_n - b_n$ have a limit distribution $\mathcal{G}$ with characteristic function γ, then

$$nv\left(\frac{\zeta}{a_n}\right) - ib_n\zeta \to \psi(\zeta) \tag{3.6}$$

where ψ is the imaginary part of $\log \gamma$. (It is easily seen that γ can have no zeros). We now prove the following lemma.

LEMMA 2. *Suppose that* (3.6) *holds for two real continuous functions v and ψ, and that $a_n \to \infty$ and $a_{n+1}/a_n \to 1$. Then for $\zeta > 0$,*

$$\psi(\zeta) = c\zeta^\alpha \quad or \quad \psi(\zeta) = c_1\zeta + c_2\zeta \log \zeta. \tag{3.7}$$

PROOF. Choose $\lambda > 0$ arbitrarily and put

$$w(\zeta) = \frac{v(\lambda\zeta)}{\lambda\zeta} - \frac{v(\zeta)}{\zeta}, \qquad \zeta > 0. \tag{3.8}$$

Then

$$\frac{n}{a_n} w\left(\frac{\zeta}{a_n}\right) \to \frac{\psi(\lambda\zeta)}{\lambda\zeta} - \frac{\psi(\zeta)}{\zeta}. \tag{3.9}$$

It follows easily that w *varies regularly at the origin*, and hence

$$\frac{\psi(\lambda\zeta)}{\lambda\zeta} - \frac{\psi(\zeta)}{\zeta} = C\zeta^{p}, \qquad \zeta > 0 \tag{3.10}$$

where C and p are constants. In principle, p could depend on λ, but from the fact that the norming constants a_n are independent of λ, one concludes easily that p is an absolute constant. Put $f(\zeta) = (\psi(\zeta)/\zeta) - 1$. Without loss of generality, one may suppose that $\psi(1) = 1$. Then $C = f(\lambda)$ and (3.10) takes on the form $f(\lambda\zeta) - f(\zeta) = f(\lambda)\zeta^{p}$. Interchanging the roles of λ and ζ, we conclude that

$$f(\zeta)[1 - \lambda^{p}] = f(\lambda)[1 - \zeta^{p}]. \tag{3.11}$$

Thus $f(\zeta) = C[1 - \zeta^{p}]$ unless $p = 1$, in which case (3.10) reduces to $f(\lambda\zeta) = f(\lambda) + f(\zeta)$. Since $f(1) = 0$, the only continuous solution of the last equation is given by $f(\zeta) = \log \zeta$, and this concludes the proof of the lemma.

The proof is admittedly not as simple as the proof of lemma 1, but it is nevertheless remarkable that so elementary an argument leads to the general form of the stable characteristic functions and gives at the same time *the precise conditions under which a given characteristic function φ belongs to a domain of attraction.*

4. Karamata's relations

In this section we denote by U an arbitrary nondecreasing function with $U(0) = 0$. We are interested only in the asymptotic behavior of U at infinity, and so there is no loss of generality in assuming that U vanishes identically in some neighborhood of 0. Together with U we consider the one-parametric family of truncated moments

$$U_p(x) = \int_0^{x+} y^{p} U(dy), \qquad x > 0, \tag{4.1}$$

for all values of p for which the integral diverges (at infinity). When considering U_p it is always understood that $U_p(\infty) = \infty$. For other values we change the notation and consider the tails

$$V_q(x) = \int_{x-}^{\infty} y^{-q} U(dy), \qquad x > 0. \tag{4.2}$$

In probabilistic applications U will be identified with the truncated second moment (2.7) of a probability distribution F. Then V_2 coincides with the tail sum T defined in (2.8), and U_p is the truncated moment of order $2 + p$, provided it diverges. The slightly greater generality will contribute to the understanding of the various phenomena of attraction. The following propositions generalize Karamata's basic relations from Lebesgue to Stieltjes integrals. (Only the limiting cases $\rho = 0$, $\rho = -p$, and $q = \rho$ present new features.) We shall later replace the asymptotic relations by asymptotic inequalities, and the new proof applies also to the following propositions.

The first proposition states that if U is of regular variation, namely

$$U(x) \sim x^{\zeta}\mathfrak{L}(x), \qquad x \to \infty, \tag{4.3}$$

then the integrals U_p and V_q are related to U as would be the case if $\mathfrak{L}$ were a constant. (The sign $\sim$ indicates that the ratio of the two sides tends to 1.) Note that $\zeta \geq 0$, for otherwise U could not increase.

PROPOSITION 1. *If U is of the form (4.3) with $\mathfrak{L}$ slowly varying, then*

$$U_p(x) \sim \frac{\rho}{p+\rho} x^{p+\rho}\mathfrak{L}(x), \qquad p > -\rho, \tag{4.4}$$

$$V_q(x) \sim \frac{\rho}{q-\rho} x^{\rho-q}\mathfrak{L}(x), \qquad q > \rho. \tag{4.5}$$

The trouble with these relations is that they break down in the interesting limit cases $p = -\rho$ and $q = \rho$. However, they may be rewritten in the form

$$\frac{U_p(x)}{x^pU(x)} \to \frac{\rho}{p+\rho}, \tag{4.6}$$

$$\frac{x^qV_q(x)}{U(x)} \to \frac{\rho}{q-\rho}, \tag{4.7}$$

and in this form they remain valid for all admissible combinations of the parameters p, q, ρ (with the obvious interpretation when a denominator vanishes. Under any circumstances $p \geq -\rho$, $q \geq \rho$, and $\rho \geq 0$).

PROPOSITION 2. *The relation* (4.3) *implies* (4.6) *and* (4.7).

The most interesting point is that the conditions (4.6) and (4.7) are not only necessary, but also sufficient for the regular variation of U, except in the limiting cases $p = -\rho$ and $q = \rho$ if they arise.

PROPOSITION 3. *If either* (4.6) *or* (4.7) *hold with a nonzero denominator, then U varies regularly.*

If (4.6) holds with $p = -\rho \neq 0$, we may interchange the role of U and U_p to conclude that U_p varies regularly. In other words, if *either U or U_p varies regularly, then*

$$\frac{U_p(x)}{x^pU(x)} \to \lambda \leq \infty \tag{4.8}$$

where $\lambda = \rho/(p+\rho)$. Conversely, if (4.8) *holds with $0 < \lambda < \infty$ then both U and U_p vary regularly. Finally,* (4.8) *with $\lambda = 0$ or ∞ implies regular variation of U and U_p, respectively.* A similar remark applies to the V_q.

5. Compactness and convergence criteria for triangular arrays

In order to explain the application of the preceding propositions and to motivate the proposed generalization of the notion of regular variation, we recall a basic fact concerning triangular arrays of random variables. For each n we consider n mutually independent random variables $X_{1,n}, \cdots, X_{n,n}$ with a common distribution F_n. As usual, we put $S_n = X_{1,n} + \cdots + X_{n,n}$. The familiar convergence theorems for triangular arrays imply the following.

CRITERION. (i) *In order that every sequence $\{S_{n_k}\}$ contains a subsequence whose* *distributions converge to a probability distribution, it is necessary and sufficient that for each* $t > 0$,

$$\limsup_{n\to\infty} n \int_{-t}^{t} y^2 F_n(dy) < \infty, \tag{5.1}$$

$$\limsup_{n\to\infty} n \left| \int_{-t}^{t} y F_n(dy) \right| < \infty, \tag{5.2}$$

and that for given $\epsilon > 0$ *there exists a* $\tau > 0$ *such that*

$$n[1 - F_n(\tau) + F_n(-\tau)] < \epsilon \tag{5.3}$$

for all n.

(ii) *If none of the limit distributions is concentrated at a single point, then*

$$\liminf_{n\to\infty} n \int_{-\tau}^{\tau} y^2 F_n(dy) > 0 \tag{5.4}$$

for some $\tau > 0$.

(iii) *The distributions of* S_n *converge to a probability distribution iff*

$$\lim_{n\to\infty} n \int_{-s}^{t} y^2 F_n\{dy\} = \psi(s, t) < \infty \tag{5.5}$$

exists for almost all positive s *and* t, *and*

$$\lim_{n\to\infty} n \int_{-t}^{t} y F_n(dy) \tag{5.6}$$

exists for some (and therefore almost all) $t > 0$.

6. Domains of attraction

We shall now show that Karamata's relations enable us to derive from the preceding criterion not only Doblin's original characterization of the nonnormal stable domains of attraction, but also the analogue for the normal distribution as well as certain variants which were derived by various authors at the expense of cumbersome calculations.

We return to a sequence $\{X_n\}$ of independent random variables with a common distribution F and partial sums S_n. We seek conditions for the existence of a limit distribution of S_n/a_n with appropriately chosen $a_n > 0$. For that purpose we apply the criterion to the triangular array defined by $X_{k,n} = X_k/a_n$ with distribution $F_n(x) = F(a_n x)$. If U denotes the truncated second moment (2.7), then condition (5.5) specialized to $s = t$ requires the existence of a limit of $n a_n^{-2} U(a_n t)$ for almost all t. This implies regular variation of U, and hence we can write

$$U(t) = t^{2-\alpha} \mathcal{L}(t) \tag{6.1}$$

with $\mathcal{L}$ slowly varying at ∞.

[The same consideration applies to the variables $a_n^{-1} S_n - b_n$. The variables of the triangular array are then $a_n^{-1}(X_k - \beta_n)$ where $b_n = n\beta_n/a_n$, and since obviously $\beta_n = o(a_n)$ it is easily seen that $n \int_{-t}^{t} y^2 F(a_n\, dy + \beta_n)$ behaves essentially as the integral in (6.2).]

From the definition of U it is clear that $0 \leq \alpha \leq 2$. Now the function V_2 introduced in (4.2) coincides with the tail sum T (defined in (2.8)), and for it the relation (4.7) holds with $\rho = 2 - \alpha$ and $q = 2$. When $\alpha = 0$ it follows that either $nT(a_n\tau) \to \infty$ or else $na_n^{-2}U(a_n\tau) \to 0$, which excludes convergence.

We have thus found that condition (6.1) with $0 < \alpha \leq 2$ is *necessary*. Assume now that it is satisfied. Since U is right continuous we can choose a_n such that $na_n^{-2}U(a_n) = 1$, in which case

$$n \int_{-t}^{t} y^2 F_n(dy) = \frac{n}{a_n^2} U(a_n t) \to t^{2-\alpha}. \tag{6.2}$$

The condition (5.3) is automatically satisfied, since in consequence of (4.7),

$$T(x) \sim \frac{2-\alpha}{\alpha} x^{-\alpha} \mathfrak{L}(x) \qquad \text{if} \quad \alpha < 2 \tag{6.3a}$$

$$T(x) = o(\mathfrak{L}(x)) \qquad \text{if} \quad \alpha = 2. \tag{6.3b}$$

When $\alpha > 1$ the distribution F has an expectation μ, and we may suppose $\mu = 0$. Using (4.7) (with $\rho = 2 - \alpha$ and $q = 1$) when $\alpha > 1$ and (4.6) (with $\rho = 2 - \alpha$ and $p = 1$) when $\alpha < 1$, one sees that the condition (5.6) is an immediate consequence of (6.2). We skip over the case $\alpha = 1$ in which (5.6) is not necessarily satisfied and centering constants may be essential to achieve convergence.

To assure the proper convergence of our distributions it remains to establish the convergence in (5.5) when $s \neq t$. Now when $\alpha = 2$ the left side in (6.2) tends to the constant 1, and this trivially implies that $\psi(s, t) = 1$ for all s, t. Thus (6.1) with $\alpha = 2$ represents the necessary and sufficient condition for convergence to the normal distribution.

When $\alpha < 2$ the existence of the limit $\psi(s, t)$ is equivalent to the existence of the limits

$$\lim_{x \to \infty} \frac{x^2[1 - F(x)]}{U(x)}, \qquad \lim_{x \to -\infty} \frac{x^2 F(-x)}{U(x)}, \tag{6.4}$$

and this requires a certain balance between the right and left tails. Under any circumstances *F belongs to a domain of attraction iff* (6.1) *holds with* $0 < \alpha \leq 2$ *and the limits in* (6.4) *exist.* They are always finite, and they vanish when $\alpha = 2$. In this formulation the only difference between the normal and other stable distributions is that the subsidiary condition relating to (6.4) is automatically satisfied when $\alpha = 2$.

When $\alpha < 2$ the tail sum T varies regularly at ∞, but this is not necessarily true when $\alpha = 2$. However, regularly varying tails play a noticeable role even if F possesses a variance or finite higher order moments. This is not visible in the usual formulation of the central limit theorem because (as we have seen) the norming constants a_n are such as to emphasize the central part of F and to obliterate the extreme tails. An entirely different picture is presented if one introduces norming constants which emphasize the tails. In fact, suppose that the tail sum T varies regularly, say

$$T(x) = x^{-p}\Lambda(x), \qquad x > 0 \tag{6.5}$$

where Λ varies slowly at infinity. Then F has finite absolute moments of order $<p$, but none of order $>p$. Whatever $p > 0$, *the tail sum for S_n varies regularly and is* $\sim nx^{-p}\Lambda(x)$. With the standard norming constants a_n, this is noticeable only when $p < 2$, that is, when F does not belong to the normal domain of attraction, but the assertion remains true when the central limit theorem applies. To see the probabilistic consequences, note that our assertion implies that as $t \to \infty$,

$$P\{|X_1| > t \mid |S_2| > t\} \to \tfrac{1}{2}, \tag{6.6}$$

whenever (6.5) *holds.* (This observation is due to B. Mandelbrot.) Because of the symmetry between X_1 and X_2 this may be expressed thus: if the tail sum varies regularly then *a large observed value of $|X_1 + X_2|$ is likely to be due entirely to one of the two components.* By contrast, if F is the exponential distribution with density e^{-x}, the left side in (6.6) equals $(1 + t)^{-1}$ and tends to 0. It is thus apparent that as far as the extreme tails are concerned, regular variation plays the same role within the domain of attraction of the normal distribution as it does for other stable distributions.

In conclusion let us remark that the Karamata relations enable us to reformulate the conditions for domains of attraction in terms of truncated moments of arbitrary orders.

7. Local limit theorems

We now show that the regular variation of U enables us to use uniformly for all domains of attraction the methods originally developed for distributions with a variance (or higher order moments). We shall be satisfied to give a typical example which, however, is of special interest.

Denote by $I_{x,h}$ the open interval of length $2h$ centered at the point x. We assume that

$$P\{S_n - b_n \leq a_n t\} \to \mathcal{G}(t) \tag{7.1}$$

where $\mathcal{G}$ is a (necessarily stable) distribution with density g and characteristic function γ. To avoid trivialities we assume F to be nonarithmetic. (A systematic reduction of arithmetic distributions to nonarithmetic ones will be described in chapter XVI of my second volume.) We are interested in the probability

$$p_n(I_{x,h}) = P\{S_n - b_n \in I_{x,h}\} = F^{n*}(x + h + b_n) - F^{n*}(x - h + b_n) \tag{7.2}$$

(at points of continuity). The following theorem states that in the limit, p_n becomes independent of x; it could be sharpened by various estimates.

THEOREM. *As $n \to \infty$,*

$$a_n p_n(I_{x,h}) \to g(0) \cdot 2h. \tag{7.3}$$

PROOF. Let δ_τ denote the density defined by

$$\delta_\tau(x) = \frac{1}{\pi} \frac{1 - \cos \tau x}{\tau x^2} \tag{7.4}$$

whose characteristic function vanishes for $|\zeta| \geq \tau$ and is given by $1 - |\zeta|/\tau$ when $|\zeta| \leq \tau$.

The starting observation is that $(1/2h)p_n(I_{x,h})$ is the value at the point $x + b_n$ of the convolution of F^{n*} and the uniform distribution with characteristic function $\sin h\zeta/h\zeta$. If φ^n were in $\mathfrak{L}_2$, we could apply the Fourier inversion formula directly, but to cover the most general case we take a further convolution with δ_τ. By the Fourier inversion formula,

$$(7.5) \qquad \frac{a_n}{2h}\int_{-\infty}^{+\infty} p_n(I_{x-y,h})\delta_\tau(y)\,dy = \frac{a_n}{2\pi}\int_{-\tau}^{\tau} e^{-i\zeta(x+b_n)}\varphi^n(\zeta)\left(1 - \frac{|\zeta|}{\tau}\right)d\zeta.$$

A relation of this form was the starting point of Berry's investigation, and the same technique is of much wider applicability than is generally realized. We proceed to estimate the two sides in (7.5).

(a) *Proof that the right side tends to* $g(0)$. With the obvious change of variables the right side becomes

$$(7.6) \qquad \frac{1}{2\pi}\int_{-a_n\tau}^{a_n\tau} e^{-i\zeta(x+b_n)/a_n}\varphi^n\left(\frac{\zeta}{a_n}\right)\frac{\sin h\zeta/a_n}{h\zeta/a_n}\left(1 - \frac{|\zeta|}{a_n\tau}\right)d\zeta.$$

From the assumption (7.1) it follows that for each fixed ζ the integrand tends to $\gamma(\zeta)$. By the Fourier inversion formula the formal limit of (7.6) equals $g(0)$, and to prove the assertion it suffices to show that the contribution of the intervals $|\zeta| > A$ is negligible when A is sufficiently large. More precisely, we show that given ϵ there exists an A such that

$$(7.7) \qquad \int_{A<\zeta<\tau a_n}\left|\varphi^n\left(\frac{\zeta}{a_n}\right)\right|d\zeta < \epsilon.$$

Since F is nonarithmetic, $\varphi(\zeta)$ is bounded away from 0 in every closed interval excluding the origin. There exists, therefore, a number $q < 1$ such that the contribution of $\eta a_n < |\zeta| < \tau a_n$ to (7.7) is bounded by $\tau a_n q^n$, which tends to 0. The only difficulty consists in proving that *there exist numbers A and η such that*

$$(7.8) \qquad \int_{A<\zeta<\eta a_n}\left|\varphi^n\left(\frac{\zeta}{a_n}\right)\right|d\zeta < \epsilon.$$

Since we can pass from φ to the characteristic function $|\varphi|^2$, there is no loss in generality in assuming φ to be real and positive. Using the inequality $1 - t < e^{-t}$ it is seen that it suffices to prove that for n sufficiently large and η sufficiently small,

$$(7.9) \qquad n\left[1 - \varphi\left(\frac{\zeta}{a_n}\right)\right] > c\zeta^\alpha \qquad \text{for} \quad 0 < \zeta < \eta a_n$$

where α is the characteristic exponent of the stable distribution $\mathcal{G}$ and c a positive constant independent of η. Obviously

$$(7.10) \qquad n\left[1 - \varphi\left(\frac{\zeta}{a_n}\right)\right] \geq n\int_{-a_n/\zeta}^{a_n/\zeta}\left(1 - \cos\frac{\zeta x}{a_n}\right)F(dx) \geq \tfrac{1}{3}\,n\,\frac{a_n^2}{\zeta^2}\,U\left(\frac{a_n}{\zeta}\right).$$

We know that

$$\lim \frac{n}{a_n} U(a_n) = c_0 \tag{7.11}$$

exists, and so the first inequality in (7.9) will hold with $c = c_0/6$ if

$$\frac{U(a_n)}{U\left(\frac{a_n}{\zeta}\right)} < 2\zeta^{2-\alpha}. \tag{7.12}$$

But U varies regularly and so (7.12) will hold for all n sufficiently large provided only that $\zeta > 1$ and a_n/ζ is sufficiently large.

(b) *The left side in* (7.5). We now describe Berry's method of estimation, which is by no means restricted to our special problem. Put $\tau = 2\epsilon^{-2}$. The density δ_τ attributes mass $<\epsilon$ to $|y| \geq \epsilon$. For $|y| < \epsilon$ the interval $I_{x-y,h}$ contains the interval $I_{x,h-\epsilon}$, and so the integral on the left side is $\geq (1 - \epsilon)p_n(I_{x,h-\epsilon})$. Replacing h by $h + \epsilon$ we have thus obtained an upper estimate of the form

$$\frac{a_n}{2h} p_n(I_{x,h}) \leq (1 + \epsilon)\gamma(0) + \epsilon \tag{7.13}$$

for all n sufficiently large. Using this we get an upper bound for the contribution of $|y| \geq \epsilon$ to the integral on the left in (7.5). For $|y| < \epsilon$ we have $p_n(I_{x-y,h}) \leq p_n(I_{x,h+\epsilon})$, and we get thus a lower bound for $p_n(I_{x,h})$ similar to (7.13).

[For distributions with variance (and therefore belonging to the domain of attraction of the normal distribution), the theorem was proved by L. A. Shepp using different methods: "A local limit theorem," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 419–423. After presenting this address, I noticed that our version of the theorem is contained in more general results recently obtained by Charles Stone in "A local limit theorem for non-lattice multi-dimensional distributed functions," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 546–551. (See also section 9.)]

8. Dominated variation

We proceed to investigate how much of the theory of regular variation remains if the requirement that a unique limit exists is replaced by a compactness condition. For definiteness we focus our attention on measures.

DEFINITION. *A positive monotone function on* $(0, \infty)$ *varies dominatedly at* ∞ *if every sequence* $\{t'_k\}$ *converging to* ∞ *contains a subsequence* $\{t_k\}$ *such that*

$$\frac{U(t_k x)}{U(t_n)} \to \tau(x) < \infty \tag{8.1}$$

for almost all x.

An equivalent requirement is the existence of numbers $a_n \to \infty$ such that every subsequence of $\{nU(a_n x)\}$ contains a convergent subsequence.

We now show the close relationship between dominated and regular variation.

THEOREM 1. *A nondecreasing function* U *varies dominatedly at* ∞ *iff there exist positive constants such that*

$$\frac{U(tx)}{U(t)} < Cx^{\rho} \qquad \text{for} \quad t > \tau, \quad x > 1. \tag{8.2}$$

PROOF. The condition is sufficient by virtue of Helly's selection theorem. If U varies dominatedly, it is possible to choose $\rho > 0$ such that

$$\frac{U(2t)}{U(t)} < 2^{\rho} \qquad \text{for} \quad t > \tau. \tag{8.3}$$

Then

$$\frac{U(2^n t)}{U(t)} < 2^{n\rho}, \tag{8.4}$$

and if $2^{n-1} < x \leq 2^n$, this implies (8.2) with $C = 2^{\rho}$. This condition is therefore necessary.

We now preserve the notations and conventions introduced in (4.1) and (4.2), and proceed to prove the counterpart to the basic relation (4.7). Also (4.6) has a similar counterpart, but the proof is slightly more delicate.

THEOREM 2. *If* $q > \rho$, *then* (8.2) *implies*

$$\limsup_{n \to \infty} \frac{x^q V_q(x)}{U(x)} \leq \gamma \tag{8.5}$$

with

$$\gamma = -1 + C \frac{q}{q - \rho}. \tag{8.6}$$

Conversely, (8.5) *implies* (8.2) *with*

$$C = 1 + \gamma, \qquad \rho = \frac{\gamma}{\gamma + 1} q. \tag{8.7}$$

PROOF. (i) Assume (8.2) and choose $\lambda > 1$. Then

$$\begin{aligned} V_q(t) &= \sum_{n=1}^{\infty} [V_q(\lambda^{n-1}t) - V_q(\lambda^n t)] \leq \sum_{n=1}^{\infty} (\lambda^{n-1}t)^{-q}[U(\lambda^n t) - U(\lambda^{n-1}t)] \\ &= -t^{-q}U(t) + (\lambda^q - 1)t^{-q} \sum_{n=1}^{\infty} \lambda^{-nq} U(\lambda^n t) \end{aligned} \tag{8.8}$$

therefore, for $t > \tau$ and arbitrary $\lambda > 1$

$$\frac{t^q V_q(t)}{U(t)} \leq -1 + C \frac{\lambda^q - 1}{\lambda^{q-\rho} - 1}. \tag{8.9}$$

Letting $\lambda \to 1$, one gets (8.6).

(ii) Assume that for $y > \tau$,

$$\frac{y^q V_q(y)}{U(y)} \leq \gamma. \tag{8.10}$$

An integration by parts shows that

$$U(y) = -y^q V_q(y) + q \int_0^y s^{q-1} V_q(s)\, ds. \tag{8.11}$$

Putting for abbreviation

$$\int_0^y s^{q-1} V_q(s)\, ds = W_q(y), \tag{8.12}$$

we see, therefore, from (8.10) that for $y > \tau$,

$$\frac{y^{q-1}V_q(y)}{W_q(y)} \leq \frac{\gamma}{\gamma + 1} q \cdot \frac{1}{y} = \rho \frac{1}{y}. \tag{8.13}$$

Integrating between t and tx one gets for $x > 1$ and $t > \tau$

$$\frac{W_q(tx)}{W_q(t)} \leq x^\rho. \tag{8.14}$$

Referring again to (8.11) we have, therefore,

$$U(tx) \leq qW_q(tx) \leq qx^\rho W_q(x) = x^\rho[U(x) + x^q V_q(x)] \leq x^\rho U(x)[1 + \gamma], \tag{8.15}$$

and so (8.2) holds with the constants, given in (8.7).

Note. The occurrence of the factor C in (8.2) and (8.5) introduces a lack of reciprocity between the constants occurring in (8.2) and (8.5). In fact, starting from the relations (8.5)–(8.6), one does not get (8.2) with the original exponent ρ, but the new exponent is

$$\rho' = \frac{C-1}{C} q + \frac{1}{C} \rho. \tag{8.16}$$

Thus $\rho' = \rho$ only if $C = 1$. In the theory of regularly varying functions one could choose C arbitrarily close to 1, and this establishes a complete symmetry between (8.2) and (8.5). It is therefore natural to ask whether our inequalities can be improved to obtain more symmetric relations. The following examples show that our inequalities are, in a sense, the best. In both examples F is a probability distribution and U its second truncated moment (2.7). We take $q = 2$ so that V_q coincides with the tail sum T.

EXAMPLES. (a) Let $a > 1$ be fixed, and let F attribute mass $(a - 1)a^{-n}$ to the point a^n (here $n = 1, 2, \cdots$). For $a^n < t < a^{n+1}$ one has $U(t) = a^{n+1} - a$ and $V_2(t) = T(t) = a^{-n}$. The left side in (8.5) therefore equals a. Now for every $x > 1$,

$$\limsup \frac{U(tx)}{U(t)} \geq a. \tag{8.17}$$

Whatever the exponent ρ, the constant C in (8.2) is therefore at least a, whereas (8.7) leads to the estimate $(a + 1)$. Since a can be chosen arbitrarily large, the estimate (8.7) is essentially the best.

(b) Let F attribute mass $e^{-1}/n!$ to the point $\lambda_n = (2^n n!)^{1/2}$. For $\lambda_n < t < \lambda_{n+1}$, clearly $U(t) = e^{-1}(2^n - 1)$ and $V_2(t) = T(t) \backsim e/(n+1)!$. For every ϵ and t sufficiently large the inequality (8.2) holds with $C = 2 + \epsilon$ and $\rho = \epsilon$, while (8.5) is true with $\gamma = 1$.

9. Stochastic compactness

If a sequence of probability distributions is stochastically compact in the sense of the definition in section 1, the same is true of the sequence of distributions obtained by symmetrization. For our purposes it suffices, therefore, to consider symmetric distributions. This is not essential, but it simplifies the exposition.

THEOREM. *Let F be symmetric. In order that there exist constants $a_n > 0$ such that the family of distributions $F^{n*}(a_n x)$ is stochastically compact, each of the following conditions is necessary and sufficient.*

(a) *The following condition holds:*

$$\limsup_{x \to \infty} \frac{x^2 T(x)}{U(x)} < \infty. \tag{9.1}$$

(b) *There exist constants $\alpha > 0$, C, τ such that*

$$\frac{U(tx)}{U(t)} < Cx^{2-\alpha} \qquad \text{for} \quad x > 1, \quad t > \tau. \tag{9.2}$$

One admissible choice of a_n is such that

$$\frac{n}{a_n^2} U(a_n) = 1. \tag{9.3}$$

PROOF. Assume that $\{F^{n*}(a_n x)\}$ is stochastically compact. As stated in section 5, the sequence of numbers $na_n^{-2}U(a_n x)$ is bounded for each $x > 0$, and there exist some x for which it is bounded away from 0. Since a scale factor is inessential, we may suppose that this is the case for $x = 1$. Then

$$A^{-1} < \frac{n}{a_n^2} U(a_n) < A \tag{9.4}$$

for some constant A. Because of the monotonicity of U this implies

$$A^{-1} < \frac{a_{n+1}}{a_n} < A. \tag{9.5}$$

Again, because of the right continuity of U there exist numbers α_n such that $n\alpha_n^{-2}U(\alpha_n) = 1$, and obviously the ratios α_n/a_n remain between A^{-1} and A. It follows that we can replace a_n by α_n without affecting the stochastic compactness. This justifies (9.3).

Assume now that (9.1) is false. In consequence of (9.5) there exists then a sequence $n_1, n_2, \cdots$ such that as n runs through it

$$\frac{a_n^2 F(a_n)}{U(a_n)} \to \infty, \tag{9.6}$$

and hence

$$nF(a_n) \to \infty. \tag{9.7}$$

But by the theory of triangular arrays, $nF(a_n x)$ remains bounded for every fixed $x > 0$, and so the condition (9.1) *is necessary*. Using theorem 2 of section 8 with $q = 2$, it is seen that the conditions (9.1) and (9.2) imply each other. It remains to show that (9.2) is sufficient.

Choose a_n so as to satisfy (9.3). Then $na_n^{-2}U(a_n x) < Cx^{2-\alpha}$ for all $x > 1$. Furthermore, we see from (8.5) that

$$\limsup nT(a_n x) \leq C\gamma \cdot x^{-\alpha}, \tag{9.8}$$

and by the criteria of section 5 these relations suffice to guarantee the stochastic boundedness.

By way of application note that the error estimate in section 7 depended only on (9.2) but not on the regular variation of U. To be sure, if F does not belong to a domain of attraction, then the integral on the right in (7.5) need not converge, but stochastic compactness of $\{F^{n*}(a_n x + b_n)\}$ guarantees that it remains bounded away from 0 and ∞. The argument of section 7 then applies to each convergent subsequence, and the theorem may be replaced by the following more general theorem.

THEOREM 2. *Assume that F is nonarithmetic and that there exist constants a_n, h_n such that the sequence of distributions $F^{n*}(a_n x + b_n)$ is stochastically compact.*

There exist norming factors α_n such that

$$\alpha_n p_n(I_{x,h}) \rightarrow 2h, \tag{9.9}$$

and $A^{-1} < \alpha_n/a_n < A$.