SECTION 10

Functional Central Limit Theorems

When does the standardized partial-sum processes converge in distribution, in the sense of the previous section, to a Gaussian process with nice sample paths? This section will establish a workable sufficient condition.

Part of the condition will imply finiteness (almost everywhere) of the envelope functions, which will mean that $S_n(\omega, \cdot)$ is a bounded function on T, for almost all ω. Ignoring negligible sets of ω, we may therefore treat S_n as a random element of the space $B(T)$ of all bounded, real-valued functions on T. The natural metric for this space is given by the uniform distance,

$$d(x, y) = \sup_t |x(t) - y(t)|.$$

One should take care not to confuse d with any metric, or pseudometric, ρ defined on T. Usually such a ρ will have something to do with the covariance structure of the partial-sum processes. The interesting limit distributions will be Gaussian processes that concentrate on the set

$$U_\rho(T) = \{x \in B(T) : x \text{ is uniformly } \rho \text{ continuous}\}.$$

Under the uniform metric d, the space $U_\rho(T)$ is separable if and only if T is totally bounded under ρ. [Notice that total boundedness excludes examples such as the real line under its usual metric.] In the separable case, a Borel probability measure P on $U_\rho(T)$ is uniquely determined by its finite dimensional projections,

$$P(B \mid t_1, \ldots, t_k) = P\{x \in U_\rho(T) : (x(t_1), \ldots, x(t_k)) \in B\},$$

with $\{t_1, \ldots, t_k\}$ ranging over all finite subsets of T and B ranging over all Borel sets in $\mathbb{R}^k$, for $k = 1, 2, \ldots$.

Let us first consider a general sequence of stochastic processes indexed by T,

$$\{X_n(\omega, t) : t \in T\} \qquad \text{for } n = 1, 2, \ldots,$$

and then specialize to the case where X_n is a properly standardized partial-sum process. Let us assume that the finite dimensional projections of X_n converge in distribution. That is, for each finite subset $\{t_1, \dots, t_k\}$ of T there is a Borel probability measure $P(\cdot \mid t_1, \dots, t_k)$ on $\mathbb{R}^k$ such that

$$(X_n(\cdot, t_1), \dots, X_n(\cdot, t_k)) \rightsquigarrow P(\cdot \mid t_1, \dots, t_k). \tag{10.1}$$

Usually classical central limit theorems will suggest the standardizations needed to ensure such finite dimensional convergence.

(10.2) THEOREM. *Let $\{X_n(\cdot, t) : t \in T\}$ be stochastic processes indexed by a totally bounded pseudometric space (T, ρ). Suppose:*

(i) the finite dimensional distributions converge, as in (10.1);

(ii) for each $\epsilon > 0$ and $\eta > 0$ there is a $\delta > 0$ such that

$$\limsup \mathbb{P}^* \left\{ \sup_{\rho(s,t)<\delta} |X_n(\omega, s) - X_n(\omega, t)| > \eta \right\} < \epsilon.$$

Then there exists a Borel measure P concentrated on $U_\rho(T)$, with finite dimensional projections given by the distributions $P(\cdot \mid t_1, \dots, t_k)$ from (10.1), such that X_n converges in distribution to P.

Conversely, if S_n converges in distribution to a Borel measure P on $U_\rho(T)$ then conditions (i) and (ii) are satisfied.

SKETCH OF A PROOF. The converse part of the theorem is a simple exercise involving almost sure representations.

For the direct part, first establish existence of the measure P concentrating on $U_\rho(T)$. Let $T(\infty) = \{t_1, t_2, \dots\}$ be a countable dense subset of T. Define $T(k) = \{t_1, \dots, t_k\}$. The Kolmogorov extension theorem lets us build a measure P on the product σ-field of $\mathbb{R}^{T(\infty)}$ with the finite dimensional distributions from the right-hand side of (10.1). By passing to the limit in (ii) we get

$$P\left\{x \in \mathbb{R}^{T(\infty)} : \max_{\substack{\rho(s,t)<\delta \\ s,t \in T(k)}} |x(s) - x(t)| \geq \eta \right\} \leq \epsilon \qquad \text{for every } k.$$

Letting $k \to \infty$, then casting out various sequences of negligible sets, we find that P concentrates on the set $U_\rho(T(\infty))$ of all uniformly continuous functions on $T(\infty)$. Each function in $U_\rho(T(\infty))$ has a unique extension to a function in $U_\rho(T)$; the extension carries P up to the sought-after Borel measure on $U_\rho(T)$.

To complete the proof let us construct a new probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ that supports perfect maps ϕ_n into Ω, such that $X_n \circ \phi_n$ converges to an $\widetilde{X}$ with distribution P, in the strengthened almost sure sense of the Representation Theorem from the previous section. This is not the circular argument that it might appear; we do not need to assume the convergence $X_n \rightsquigarrow P$ in order to adapt some of the ideas from the proof of that theorem. Indeed, we can break into the proof between its second and third steps by establishing directly that $\liminf \mathbb{P}_*\{X_n \in B\} \geq PB$ for every B that is a finite intersection of closed balls in $B(T)$ with centers in $U_\rho(T)$ and zero P measure on their boundaries.

Such a set B has a simple form; it is defined by a pair of functions g, h in $U_\rho(T)$:

$$B = \{x \in B(T) : g(t) \le x(t) \le h(t) \text{ for all } T\}.$$

It has zero P measure on its boundary. For $\eta > 0$ define

$$B_\eta = \{x \in B(T) : g(t) + \eta < x(t) < h(t) - \eta \text{ for all } t\}.$$

As $\eta \to 0$, the sets B_η expand up to the interior of B. The fact that P puts zero measure on the boundary of B lets us choose η so that $PB_\eta \ge PB - \epsilon$. This inequality gives us room to approximate the paths of the X_n processes from their values on a finite subset of T.

Fix an $\epsilon > 0$. The fact that P concentrates on $U_\rho(T)$ lets us choose a $\delta > 0$ so that the set

$$F = \left\{x \in B(T) : \sup_{\rho(s,t)<\delta} |x(s) - x(t)| \le \eta/2\right\}$$

has P measure at least $1 - \epsilon$. Condition (ii) of the theorem lets us assume that δ is small enough to ensure $\limsup \mathbb{P}^*\{X_n \in F^c\} < \epsilon$. We may also assume that both g and h belong to F, because both are uniformly continuous.

Now let $T(k) = \{t_1, \ldots, t_k\}$ be a finite set that approximates within a distance δ to every point of T. For a function x in F and a t with $\rho(t, t_i) < \delta$, if $x(t_i) < h(t_i) - \eta$ then $x(t) \le x(t_i) + \eta/2 < h(t_i) - \eta/2$. The upper bound is less than $h(t)$, because $h \in F$. A similar argument with g would give a lower bound. It follows that the set

$$\{X_n \in F : g(t_i) + \eta < X_n(\cdot, t_i) < h(t_i) - \eta \text{ for } t_i \in T(k)\}$$

is contained within $\{X_n \in B\}$, and hence

$$\mathbb{P}_*\{X_n \in B\} \ge \mathbb{P}\{g + \eta < X_n < h - \eta \text{ on } T(k)\} - \mathbb{P}^*\{X_n \in F^c\}.$$

The first term on the right-hand side may be reexpressed as

$$\mathbb{P}\{\big(X_n(\cdot, t_1), \ldots, X_n(\cdot, t_k)\big) \in G\},$$

where G is the open subset of $\mathbb{R}^k$ defined by the inequalities

$$g(t_i) + \eta < x_i < h(t_i) - \eta \qquad \text{for } i = 1, \ldots, k.$$

From assumption (i), the lim inf of the last probability is greater than

$$P\{x \in U_\rho(T) : (x(t_1), \ldots, x(t_k)) \in G\} \ge PB_\eta.$$

It follows that

$$\liminf \mathbb{P}_*\{X_n \in B\} \ge PB - 2\epsilon \qquad \text{for each } \epsilon > 0.$$

By copying Steps 3 through 8 in the proof of the Representation Theorem we could now complete the construction of versions of the X_n that converge in the strong sense to an $\widetilde{X}$ with distribution P. The assertion of the theorem would then follow easily. □

Condition (ii) of Theorem 10.2 is sometimes called *stochastic equicontinuity* or, less precisely, uniform tightness. It is equivalent to the requirement: for every sequence $\{r_n\}$ of real numbers converging to zero,

$$\text{(10.3)} \qquad \sup\{|X_n(s) - X_n(t)| : \rho(s, t) < r_n\} \to 0 \qquad \text{in probability.}$$

It also implies (and actually is implied by) that for every sequence of random elements $\{\sigma_n\}, \{\tau_n\}$ of T with $\rho(\sigma_n, \tau_n) \to 0$ in probability,

$$X_n(\sigma_n) - X_n(\tau_n) \to 0 \quad \text{in probability.} \tag{10.4}$$

One has only to choose r_n converging to zero so slowly that $\mathbb{P}^*\{\rho(\sigma_n, \tau_n) \geq r_n\} \to 0$ to establish the implication. Notice that (10.4) is much stronger than the corresponding assertion for deterministic sequences $\{\sigma_n\}, \{\tau_n\}$ with $\rho(s_n, t_n) \to 0$. Verification of the weaker assertion would typically involve little more than an application of Tchebychev's inequality, whereas (10.4) corresponds to a much more powerful maximal inequality.

Let us now specialize Theorem 10.2 to random processes constructed from a triangular array $\{f_{ni}(\omega, t) : t \in T, 1 \leq i \leq k_n, n = 1, 2, \ldots\}$, with the $\{f_{ni}\}$ independent within each row. Define

$$X_n(\omega, t) = \sum_{i \leq k_n} \Big(f_{ni}(\omega, t) - \mathbb{P} f_{ni}(\cdot, t) \Big),$$

$$\rho_n(s, t) = \Big(\sum_{i \leq k_n} \mathbb{P} |f_{ni}(\cdot, s) - f_{ni}(\cdot, t)|^2 \Big)^{1/2}.$$

The double subscripting allows us to absorb into the notation the various standardizing constants needed to ensure convergence of finite dimensional distributions. If we also arrange to have

$$\rho(s, t) = \lim_{n \to \infty} \rho_n(s, t) \tag{10.5}$$

well defined for each pair s, t in T, then such a ρ will be an appropriate choice for the pseudometric on T. In the frequently occurring case where $f_{ni}(\omega, t) = f_i(\omega, t)/\sqrt{n}$, with the $\{f_i\}$ independent and identically distributed, we have $\rho(s, t) = \rho_n(s, t)$, and condition (v) of the next theorem is trivially satisfied.

(10.6) FUNCTIONAL CENTRAL LIMIT THEOREM. *Suppose the processes from the triangular array $\{f_{ni}(\omega, t)\}$ are independent within rows and satisfy:*

- *(i) the $\{f_{ni}\}$ are manageable, in the sense of Definition 7.9;*
- *(ii) $H(s, t) = \lim_{n \to \infty} \mathbb{P} X_n(s) X_n(t)$ exists for every s, t in T;*
- *(iii) $\limsup \sum_i \mathbb{P} F_{ni}^2 < \infty$;*
- *(iv) $\sum_i \mathbb{P} F_{ni}^2 \{F_{ni} > \epsilon\} \to 0$ for each $\epsilon > 0$;*
- *(v) the limit $\rho(\cdot, \cdot)$ is well defined by (10.5) and, for all deterministic sequences $\{s_n\}$ and $\{t_n\}$, if $\rho(s_n, t_n) \to 0$ then $\rho_n(s_n, t_n) \to 0$.*

Then

- *(a) T is totally bounded under the ρ pseudometric;*
- *(b) the finite dimensional distributions of X_n have Gaussian limits, with zero means and covariances given by H, which uniquely determine a Gaussian distribution P concentrated on $U_\rho(T)$;*
- *(c) X_n converges in distribution to P.*

PROOF. Conditions (ii) and (iv) imply (Lindeberg central limit theorem) that the finite dimensional distributions have the stated Gaussian limits.

The stochastic equicontinuity requirement of Theorem 10.2 will be established largely by means of maximal inequalities implied by manageability. Recall from Section 7 that manageability of the $\{f_{ni}\}$ means that there exists a deterministic function λ with $\sqrt{\log \lambda}$ integrable and

$$(10.7) \qquad D_2(x|\boldsymbol{\alpha} \odot \mathbf{F}_n|_2, \boldsymbol{\alpha} \odot \mathcal{F}_{n\omega}) \le \lambda(x) \qquad \text{for } 0 < x \le 1, \text{ all } \omega, \text{ all } \boldsymbol{\alpha}, \text{ all } n.$$

For manageable arrays of processes we have the moment bounds, for $1 \le p < \infty$,

$$(10.8) \qquad \mathbb{P}\sup_t |\sum_i f_{ni}(\omega, t) - \mathbb{P}f_{ni}(\cdot, t)|^p \le \mathbb{P}|\mathbf{F}_n|_2^p \Lambda_p(\delta_n/|\mathbf{F}_n|_2),$$

where $\delta_n^2 = \sup_t \sum_i f_{ni}(\omega, t)^2$ and Λ_p is a continuous, increasing function that depends only on λ and p, with $\Lambda_p(0) = 0$ and $\Lambda_p(1) < \infty$.

The presence of the rescaling vector $\boldsymbol{\alpha}$ in (10.7) will allow us to take advantage of the Lindeberg condition (iv) without destroying the bound. Because (iv) holds for each fixed $\epsilon > 0$, it also holds when ϵ is replaced by a sequence $\{\epsilon_n\}$ converging to zero slowly enough:

$$\sum_i \mathbb{P}F_{ni}^2\{F_{ni} > \epsilon_n\} \to 0.$$

We can replace f_{ni} by $f_{ni}\{F_{ni} \le \epsilon_n\}$ and F_{ni} by $F_{ni}\{F_{ni} \le \epsilon_n\}$ without disturbing inequality (10.7); the indicator function $\{F_{ni} \le \epsilon_n\}$ is absorbed into the weight α_i. The same truncation has no bad effect on the other four assumptions of the theorem. We therefore lose no generality by strengthening (iv) to:

$$(iv)' \qquad F_{ni}(\omega) \le \epsilon_n \qquad \text{for all } n, \text{ all } i, \text{ all } \omega.$$

Henceforth assume this inequality holds.

The idea will be to apply a maximal inequality analogous to (10.8) to the processes

$$h_{ni}(\omega, s, t) = f_{ni}(\omega, s) - f_{ni}(\omega, t),$$

at least for those pairs s, t with $\rho(s,t) < r_n$, with the aim of establishing stochastic equicontinuity in the form (10.3). The maximal inequality will involve the random variable

$$\theta_n(\omega) = \sup\{|\mathbf{h}_n(\omega, s, t)|_2 : \rho(s,t) < r_n\}.$$

We will use manageability to translate the convergence $r_n \to 0$ into the conclusion that $\theta_n \to 0$ in probability.

From the stability results for packing numbers in Section 5, the doubly indexed processes $\{h_{ni}(\omega, s, t)\}$ are also manageable, for the envelopes $H_{ni} = 2F_{ni}$, with capacity bound $\lambda(x/2)^2$. And the processes $\{h_{ni}(\omega, s, t)^2\}$ are manageable for the envelopes $\{H_{ni}^2\}$, by virtue of inequality (5.2) for packing numbers of pointwise products. The analogue of (10.8) therefore holds for the $\{h_{ni}^2\}$ processes, with envelopes $\{H_{ni}^2\}$ and the Λ_p function increased by a constant multiple. In particular, there is a constant C such that

$$\begin{aligned}\mathbb{P}\sup_{s,t}\left|\sum_i h_{ni}(\omega,s,t)^2 - \mathbb{P}h_{ni}(\cdot,s,t)^2\right|^2 &\le C\,\mathbb{P}\sum_i F_{ni}^4 \\ &\le C\sum_i \epsilon_n^2 \mathbb{P}F_{ni}^2 \\ &\to 0.\end{aligned}$$

Consequently,

$$\sup_{s,t}\left|\,|\mathbf{h}_n(\omega,s,t)|_2^2 - \rho_n(s,t)^2\right| \to 0 \qquad \text{in probability.} \tag{10.9}$$

The second part of assumption (v) implies that

$$\sup_{s,t}\{\rho_n(s,t) : \rho(s,t) < r_n\} \to 0.$$

Together these two uniformity results give $\theta_n \to 0$ in probability.

The convergence (10.9) also establishes total boundedness of T under the ρ pseudometric, with plenty to spare. First note that assumption (iii) and the fact that $\sum_i \mathbb{P}F_{ni}^4 \to 0$ together imply that $|\mathbf{F}_n|_2$ is stochastically bounded: for some constant K there is probability close to one for all n that $|\mathbf{F}_n|_2 \le K$. Now suppose $\{t_1, \ldots, t_m\}$ is a set of points with $\rho(t_i, t_j) > \epsilon K$ for $i \ne j$. By definition of ρ and by virtue of (10.9), with probability tending to one,

$$|\mathbf{f}_n(\omega,t_i) - \mathbf{f}_n(\omega,t_j)|_2 > \epsilon K \qquad \text{for } i \ne j.$$

Eventually there will be an ω (in fact, a whole set of them, with probability close to one) for which $m \le D_2(\epsilon|\mathbf{F}_n|_2, \mathcal{F}_{n\omega})$. It follows from (10.7) that $m \le \lambda(\epsilon)$. That is, λ is also a bound on the packing numbers of T under the ρ pseudometric.

To complete the proof of stochastic equicontinuity, invoke the analogue of (10.8) with $p = 1$ for the processes of differences $h_{ni}(\omega,s,t)$ with $\rho(s,t) < r_n$. By manageability, there is a continuous, increasing function $\Gamma(\cdot)$ with $\Gamma(0) = 0$ such that

$$\mathbb{P}\sup\{|X_n(s) - X_n(t)| : \rho(s,t) < r_n\} \le \mathbb{P}|\mathbf{F}_n|_2\Gamma(\theta_n/|2\mathbf{F}_n|_2).$$

For a fixed $\epsilon > 0$, split the right-hand side according to whether $|\mathbf{F}_n|_2 > \epsilon$ or not, to get the upper bound

$$\epsilon\,\Gamma(1) + \mathbb{P}|\mathbf{F}_n|_2\Gamma\left(1 \wedge \frac{\theta_n}{2\epsilon}\right).$$

The Cauchy-Schwarz inequality bounds the second contribution by

$$\left[\mathbb{P}|\mathbf{F}_n|_2^2\,\mathbb{P}\Gamma^2\left(1 \wedge \frac{\theta_n}{2\epsilon}\right)\right]^{1/2}.$$

Assumption (iii) keeps $\mathbb{P}|\mathbf{F}_n|_2^2$ bounded; the convergence in probability of θ_n to zero sends the second factor to zero. Stochastic equicontinuity of $\{X_n\}$ follows. □

REMARKS. The original functional central limit theorem for empirical distribution functions is due to Donsker (1952). Dudley (1978) extended the result to empirical processes indexed by general classes of sets. He used the term *Donsker class* to describe those classes of sets (and later, also those classes of functions—see Dudley (1987), for example) for which a functional central limit theorem holds.

The literature contains many examples of such limit theorems for empirical processes and partial-sum processes, mostly for identically distributed summands. Some of the best recent examples may be found in the papers of Dudley (1984), Giné and Zinn (1984), Alexander and Pyke (1986), Ossiander (1987), Alexander (1987a, 1987b), Talagrand (1987), and Andersen and Dobrić (1987, 1988). My Theorem 10.6 extends a central limit theorem of Kim and Pollard (1990), refining the earlier result from Pollard (1982). It could also be deduced from the theorems of Alexander (1987b). The assumption of manageability could be relaxed.

Theorem 10.2 is based on Theorem 5.2 of Dudley (1985), which extends a line of results going back at least to Dudley (1966). See also Dudley (1984). My method of proof is different, although similar in spirit to the methods of Skorohod (1956).