and

$$\left|\int f\,dP_n - S_n\right| \leqq \varepsilon + MF_n(a) + M(1 - F_n(b)) \to \varepsilon + MF(a) + M(1 - F(b)) < (2M + 1)\varepsilon,$$

the limits superior and inferior of $\int f\,dP_n$ are within $(4M + 2)\varepsilon$ of $\int f\,dP$, which completes the proof since ε was arbitrary.

A probability measure on the class $\mathscr{R}^k$ of Borel sets in k-dimensional Euclidean space R^k is determined by its distribution function

$$F(x_1, \cdots, x_k) = P\{y = (y_1, \cdots, y_k) : y_1 \leqq x_1, \cdots, y_k \leqq x_k\}.$$

The analogue to Theorem 2.3 holds and can be proved by a consideration of k-dimensional Riemann sums.

3. Random elements and convergence in distribution. Let S be a metric space as before, and let $(\Omega, \mathscr{B}, \mathsf{P})$ be a probability measure space. A mapping $X : \Omega \to S$ is called a random element of S if it is measurable in the sense that $\{\omega : X(\omega) \in A\} = X^{-1}A \in \mathscr{B}$ for each $A \in \mathscr{S}$. Special cases are random variables ($S = R^1$) and random vectors ($S = R^k$). The distribution of X is the measure $P = \mathsf{P}X^{-1}$ on $\mathscr{S}$:

$$P(A) = \mathsf{P}X^{-1}(A) = \mathsf{P}\{\omega : X(\omega) \in A\} = \mathsf{P}\{X \in A\}.$$

Suppose in addition to X we have a sequence of random elements X_n of S, defined on spaces $(\Omega_n, \mathscr{B}_n, \mathsf{P}_n)$, with distributions $P_n = \mathsf{P}_n X_n^{-1}$. If $P_n \Rightarrow P$ we say X_n converges in distribution to X and write $X_n \Rightarrow X$. Every result about weak convergence has an anologue about convergence in distribution, and vice versa, and to pass from the one to the other requires not a proof, but merely a translation. For example, by Theorem 2.1, $X_n \Rightarrow X$ if and only if $\limsup_n \mathsf{P}_n\{X_n \in F\} \leqq \mathsf{P}\{X \in F\}$ for all closed sets F.

Suppose X_n and Y_n are both random elements of S defined on Ω_n. We shall drop the subscript n from P_n.

THEOREM 3.1. *If $X_n \Rightarrow X$ and the distance $\rho(X_n, Y_n)$ converges to 0 in probability, then $Y_n \Rightarrow X$.*

Proof. Clearly $\mathsf{P}\{Y_n \in F\} \leqq \mathsf{P}\{X_n \in (F^\delta)^-\} + \mathsf{P}\{\rho(X_n, Y_n) \geqq \delta\}$. The second term here goes to 0 since $\rho(X_n, Y_n)$ converges in probability to 0, and since $X_n \Rightarrow X$, we have

$$\limsup_n \mathsf{P}\{Y_n \in F\} \leqq \mathsf{P}\{X \in (F^\delta)^-\}.$$

Since $(F^\delta)^- \downarrow F$ as $\delta \downarrow 0$ if F is closed, the result follows.

Let us use the term *Lebesgue interval* to refer to the probability measure space $(\Omega, \mathscr{B}, \mathsf{P})$, where Ω is the unit interval, $\mathscr{B}$ is the σ-field of Borel sets in Ω and P is Lebesgue measure on $\mathscr{B}$.

THEOREM 3.2. *For each probability measure P on $(S, \mathscr{S})$, there is a random element of S, defined on the Lebesgue interval, with distribution P.*

Proof. For each k, construct a decomposition $\mathcal{A}_k = \{A_{k1}, A_{k2}, \cdots\}$ of S into disjoint sets of diameter less than $1/k$, and arrange that $\mathcal{A}_{k+1}$ refines $\mathcal{A}_k$. And construct a decomposition $\mathcal{I}_k = \{I_{k1}, I_{k2}, \cdots\}$ of the unit interval into subintervals whose lengths satisfy $|I_{ku}| = P(A_{ku})$, and arrange that $\mathcal{I}_{k+1}$ refines $\mathcal{I}_k$. Finally, arrange the indexing so that $A_{ku} \supset A_{k+1,v}$ if and only if $I_{ku} \supset I_{k+1,v}$. The construction can be carried out inductively because, if $l_1, l_2, \cdots$ are nonnegative numbers adding to the length of an interval, that interval can be split into subintervals of lengths $l_1, l_2, \cdots$.

Let x_{ku} be some point in A_{ku}, and define a random element X_k by

$$X_k(\omega) = x_{ku} \quad \text{if} \quad \omega \in I_{ku}. \tag{3.1}$$

Since $\{X_k(\omega), X_{k+1}(\omega), \cdots\}$ is contained in some one element of $\mathcal{A}_k$, its diameter is at most $1/k$; thus $\{X_k(\omega)\}$ is a Cauchy sequence for each ω, the limit $X(\omega) = \lim_k X_k(\omega)$ exists, and this limit satisfies

$$\rho(X(\omega), X_k(\omega)) \leqq \frac{1}{k}. \tag{3.2}$$

If a prime denotes a sum or union extended over those u for which A_{ku} meets a given set F, then

$$\begin{aligned} \mathsf{P}\{X_k \in F\} &\leqq \mathsf{P}\{X_k \in \textstyle\bigcup' A_{ku}\} = \textstyle\sum' \mathsf{P}\{X_k \in A_{ku}\} \\ &= \textstyle\sum' |I_{ku}| = \textstyle\sum' P(A_{ku}) \leqq P((F^{1/k})^-). \end{aligned}$$

If F is closed, it follows that

$$\limsup_k \mathsf{P}\{X_k \in F\} \leqq P(F).$$

Thus the distribution of X_k converges to P, and hence, by (3.2) and Theorem 3.1 (with $Y_k = X$), X has distribution P.

We turn now to the Skorokhod representation theorem, an extension of Theorem 3.2. In addition to P, consider a sequence $\{P_n\}$ of probability measures on S.

THEOREM 3.3. *If $P_n \Rightarrow P$, then there exist on the Lebesgue interval random elements X_n and X which have respective distributions P_n and P and satisfy* $\lim_n X_n(\omega) = X(\omega)$ *for each ω.*

Proof. Construct the decompositions $\mathcal{A}_k$ of the preceding proof, but this time require that each A_{ku} be a P-continuity set. (Since $\partial\{y:\rho(x, y) < \delta\} \subset \{y:\rho(x, y) = \delta\}$, the spheres about x are P-continuity sets except for countably many radii, so S can be covered by countably many P-continuity sets of diameter less than $1/k$. The usual procedure for rendering the sets disjoint preserves P-continuity because $\partial(A \cap B) \subset (\partial A) \cup (\partial B)$.)

Consider the decompositions $\mathcal{I}_k$ as before, and, for each n, construct successively finer partitions $\mathcal{I}_k^{(n)} = \{I_{k1}^{(n)}, I_{k2}^{(n)}, \cdots\}$ with $|I_{ku}^{(n)}| = P_n(A_{ku})$. Inductively arrange the indexing so that (here $I < J$ for intervals means the right endpoint of I does not exceed the left endpoint of J) $I_{ku}^{(n)} < I_{kv}^{(n)}$ if and only if $I_{ku} < I_{kv}$. In other words, ensure that for each k the families $\mathcal{I}_k, \mathcal{I}_k^{(1)}, \mathcal{I}_k^{(2)}, \cdots$ are ordered similarly.

Define X_k by (3.1), as before, where $x_{ku} \in A_{ku}$, and define

$$X_k^{(n)}(\omega) = x_{ku} \quad \text{if} \quad \omega \in I_{ku}^{(n)}.$$

Again $X_k(\omega)$ converges to an $X(\omega)$ satisfying (3.2), and $X_k^{(n)}(\omega)$ converges ($k \to \infty$) to an $X^{(n)}(\omega)$ satisfying

$$\rho(X^{(n)}(\omega), X_k^{(n)}(\omega)) \leqq \frac{1}{k}. \tag{3.3}$$

And again X has distribution P and $X^{(n)}$ has distribution P_n.

Since $\sum_u [P(A_{ku}) - P_n(A_{ku})] = 0$, we have

$$\begin{aligned}\sum_u \big| |I_{ku}| - |I_{ku}^{(n)}| \big| &= \sum_u |P(A_{ku}) - P_n(A_{ku})| \\ &= 2 \sum_u{}'' [P(A_{ku}) - P_n(A_{ku})] = 2 \sum_u [P(A_{ku}) - P_n(A_{ku})]^+,\end{aligned}$$

where the next-to-last sum extends over those u for which the summand is positive. Each summand goes to 0 as $n \to \infty$ because the A_{ku} are P-continuity sets, and it follows by dominated convergence that

$$\lim_{n \to \infty} \sum_u \big| |I_{ku}| - |I_{ku}^{(n)}| \big| = 0. \tag{3.4}$$

Fix k and u_0, let α and α_n be the left endpoints of I_{ku_0} and $I_{ku_0}^{(n)}$ respectively, and let $\sum'$ indicate summation over the set of u for which $I_{ku} < I_{ku_0}$ (which is the same as the set for which $I_{ku}^{(n)} < I_{ku_0}^{(u)}$). Then (3.4) implies

$$\alpha = \sum_u{}' |I_{ku}| = \lim_{n \to \infty} \sum_u{}' |I_{ku}^{(n)}| = \lim_{n \to \infty} \alpha_n.$$

Similarly the right endpoint of $I_{ku}^{(n)}$ converges as $n \to \infty$ to the right endpoint of I_{ku}.

Hence, if ω is interior to I_{ku}, then ω lies in $I_{ku}^{(n)}$ for all sufficiently large n, so that, by (3.2) and (3.3),

$$\rho(X(\omega), X^{(n)}(\omega)) \leqq \frac{2}{k}. \tag{3.5}$$

Thus, if ω is not an endpoint of any I_{ku}, then, for each k, (3.5) holds for all sufficiently large n. In other words, $\lim_n X^{(n)}(\omega) = X(\omega)$ if ω is not in the set of endpoints of the I_{ku}. This last set, being countable, has Lebesgue measure 0, so that, if $X^{(n)}(\omega)$ is redefined as $X(\omega)$ on this set, $X^{(n)}$ still has distribution P_n, and there is now convergence for all ω. This proves the theorem (with $X^{(n)}$ for X_n).

The theorem can be restated: Consider random elements X and X_n of S; they may all be defined on different probability spaces.

COROLLARY 1. *If $X_n \Rightarrow X$, then there exist on the Lebesgue interval random elements Y_n and Y which have the distributions of X_n and X respectively and which satisfy $\lim_n Y_n(\omega) = Y(\omega)$ for all ω.*

If P is a measure on S, and if f is a mapping from S to another metric space S', measurable in the sense that $f^{-1}A \in \mathscr{S}$ if A is in the σ-field $\mathscr{S}'$ of Borel sets in S', then Pf^{-1} is the probability measure on $\mathscr{S}'$ defined by $Pf^{-1}(A) = P(f^{-1}A)$. Suppose, in addition, we have a sequence of measures P_n. Let D_f be the set of discontinuities of f and assume it lies in $\mathscr{S}$.

Corollary 2. *If* $P_n \Rightarrow P$ *and* $P(D_f) = 0$, *then* $P_n f^{-1} \Rightarrow Pf^{-1}$.

Proof. Consider the random elements of Theorem 3.3. Now $\lim_n X_n(\omega) = X(\omega)$ for each ω, and if f is continuous at $X(\omega)$, which by hypothesis holds except on an ω-set of Lebesgue measure 0, then

$$\lim_n f(X_n(\omega)) = f(X(\omega)). \tag{3.6}$$

Thus (3.6) holds for almost all ω, and, since $f(X_n)$ and $f(X)$ have respective distributions $P_n f^{-1}$ and Pf^{-1}, it follows that $P_n f^{-1} \Rightarrow Pf^{-1}$.

Corollary 3. *If* $X_n \Rightarrow X$ *and* $\mathsf{P}\{X \in D_f\} = 0$, *then* $f(X_n) \Rightarrow f(X)$.

This corollary is a direct translation of Corollary 2. It can also be deduced from Corollary 1.

Corollary 4. *If random variables* X_n *and* X *satisfy* $X_n \Rightarrow X$, *then*

$$\mathsf{E}\{|X|\} \leqq \limsup_n \mathsf{E}\{|X_n|\}. \tag{3.7}$$

Corollary 5. *If random variables* X_n *and* X *satisfy* $X_n \Rightarrow X$, *and if the* X_n *are uniformly integrable in the sense that*

$$\lim_{\alpha\to\infty} \sup_n \int_{\{|X_n| > \alpha\}} |X_n| \, d\mathsf{P} = 0, \tag{3.8}$$

then X *is integrable and*

$$\lim_n \mathsf{E}\{X_n\} = \mathsf{E}\{X\}.$$

To prove Corollary 4, consider the random variables Y_n and Y guaranteed by Corollary 1. Fatou's lemma implies $\mathsf{E}\{|Y|\} \leqq \limsup_n \mathsf{E}\{|Y_n|\}$, and (3.7) follows because Y_n and Y have the distributions of X_n and X. (The E in (3.7) denotes expected value with respect to whatever probability measure governs the random variable in question.) Corollary 5 similarly reduces to a standard fact of integration theory.

Theorem 3.3 can be used to give simple proofs of many results in statistics, for example, those connected with the δ-method.

4. Prokhorov's theorem. A family Π of probability measures on S is said to be relatively compact if each sequence $\{P_n\}$ of elements of Π contains some subsequence $\{P_{n_i}\}$ converging weakly to some probability measure P. The limit P is not required to lie in Π, but of course it must be a probability measure on S.

It is possible to metrize the space of probability measures on S (see the remarks preceding Theorem 2.2), and Π is relatively compact if and only if it has compact closure in this metric. It is not necessary to go into this matter, however, because the definition above makes good sense as it stands.

The following theorem, due to Prokhorov, is basic to the application of weak convergence in probability theory. The family Π is said to be *tight* if, for each positive ε, there is a compact set K_ε for which $P(K_\varepsilon) > 1 - \varepsilon$ for every P in Π.

THEOREM 4.1. *The family* Π *is relatively compact if and only if it is tight.*

Proof. Suppose that Π is tight. There is a sequence $\{K_u\}$ of compact sets such that $K_1 \subset K_2 \subset \cdots$ and $P(K_u) > 1 - 1/u$ for all u. Let $\mathscr{A}$ be a countable collection of open spheres forming a base for the topology of S, and let $\mathscr{H}$ consist of the finite unions of sets of the form $A^- \cap K_u$ with $u \geqq 1$ and A an element of $\mathscr{A}$. Then $\mathscr{H}$ is countable and is closed under the formation of finite unions, and each set in $\mathscr{H}$ is compact.

Given a sequence $\{P_n\}$ in Π, select by the diagonal procedure a subsequence $\{P_{n_i}\}$ along which limits

$$\alpha(H) = \lim_{i\to\infty} P_{n_i}(H) \tag{4.1}$$

exist for all H in $\mathscr{H}$. Suppose there exists a probability measure P such that

$$P(G) = \sup_{H\subset G} \alpha(H) \tag{4.2}$$

for all open sets G. Then $P_{n_i} \Rightarrow P$ as $i \to \infty$ because, if $H \subset G$, $\alpha(H) = \lim_i P_{n_i}(H) \leqq \liminf_i P_{n_i}(G)$, whence $P(G) \leqq \liminf_i P_{n_i}(G)$ follows via (4.2), proving weak convergence. Thus it suffices to produce a P satisfying (4.2).

Clearly $\alpha(H)$, defined by (4.1) for all H in $\mathscr{H}$, has these properties:

$$\alpha(H_1) \leqq \alpha(H_2) \quad \text{if} \quad H_1 \subset H_2; \tag{4.3}$$

$$\alpha(H_1 \cup H_2) = \alpha(H_1) + \alpha(H_2) \quad \text{if} \quad H_1 \cap H_2 = 0; \tag{4.4}$$

$$\alpha(H_1 \cup H_2) \leqq \alpha(H_1) + \alpha(H_2). \tag{4.5}$$

Define

$$\beta(G) = \sup_{H\subset G} \alpha(H) \tag{4.6}$$

for open sets G, and then define

$$\gamma(M) = \inf_{M\subset G} \beta(G)$$

for arbitrary subsets M of S. Clearly $\gamma(G) = \beta(G)$ for open G.

Now suppose we succeed in proving that γ is an outer measure and that each closed set is γ-measurable (measurable with respect to γ). Then all sets in $\mathscr{S}$ will be γ-measurable (recall the γ-measurable sets form a σ-field) and the restriction P of γ to $\mathscr{S}$ will be a measure satisfying $P(G) = \gamma(G) = \beta(G)$, so (4.2) will hold for open G as required, and P will be a probability measure because

$$1 \geqq P(S) = \beta(S) \geqq \sup_u \alpha(K_u) \geqq \sup_u \left(1 - \frac{1}{u}\right).$$

We first prove that β *is finitely subadditive* (on open sets): If $H \subset G_1 \cup G_2$ and $H \in \mathscr{H}$, define $F_1 = \{x \in H : \rho(x, G_1^c) \geqq \rho(x, G_2^c)\}$ and $F_2 = \{x \in H : \rho(x, G_2^c)$

$\geqq \rho(x, G_1^c)\}$ (see Fig. 2). If $x \in F_1$ and $x \notin G_1$, then $x \in G_2$, so that, since G_2^c is closed, $\rho(x, G_1^c) = 0 < \rho(x, G_2^c)$, a contradiction. Thus $F_1 \subset G_1$; similarly $F_2 \subset G_2$. Since F_1 is compact, being a closed subset of the compact set H, and F_1 is inside the open set G_1, it follows by the definition of $\mathscr{H}$ that $F_1 \subset H_1 \subset G_1$ for some H_1 in $\mathscr{H}$; similarly $F_2 \subset H_2 \subset G_2$ for some H_2 in $\mathscr{H}$. But then $\alpha(H) \leqq \alpha(H_1 \cup H_2) \leqq \alpha(H_1) + \alpha(H_2) \leqq \beta(G_1) + \beta(G_2)$ by (4.3), (4.5), and (4.6). Since we can vary H inside $G_1 \cup G_2$, $\beta(G_1 \cup G_2) \leqq \beta(G_1) + \beta(G_2)$ follows.

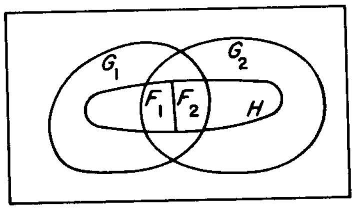

FIG. 2

Next, β *is countably subadditive* (on open sets): For, if $H \subset \bigcup_n G_n$, then, since H is compact, $H \subset \bigcup_{n \leqq n_0} G_n$ for some n_0, and therefore, by finite subadditivity, $\alpha(H) \leqq \beta(\bigcup_{n \leqq n_0} G_n) \leqq \sum_{n \leqq n_0} \beta(G_n) \leqq \sum_n \beta(G_n)$. Taking the supremum over H inside $\bigcup_n G_n$ gives $\beta(\bigcup_n G_n) \leqq \sum_n \beta(G_n)$.

And γ *is an outer measure*: Since γ is clearly monotone, we need only prove it countably subadditive. Given a positive ε and arbitrary subsets M_n of S, choose open sets G_n such that $M_n \subset G_n$ and $\beta(G_n) < \gamma(M_n) + \varepsilon/2^n$. Then, by the countable subadditivity of β, $\gamma(\bigcup_n M_n) \leqq \beta(\bigcup_n G_n) \leqq \sum_n \beta(G_n) < \sum_n \gamma(M_n) + \varepsilon$, whence, ε being arbitrary, we conclude $\gamma(\bigcup_n M_n) \leqq \sum_n \gamma(M_n)$.

It remains only to prove that *each closed set is γ-measurable*. We must show that, if F is closed and M arbitrary,

$$\gamma(M) \geqq \gamma(M \cap F) + \gamma(M \cap F^c) \tag{4.7}$$

(the reverse inequality follows by the subadditivity of γ). To prove (4.7) it suffices to prove

$$\beta(G) \geqq \gamma(G \cap F) + \gamma(G \cap F^c) \tag{4.8}$$

for open G, because then $G \supset M$ implies $\beta(G) \geqq \gamma(M \cap F) + \gamma(M \cap F^c)$ and taking the infimum over G gives (4.7).

To prove (4.8), choose, for given positive ε, an H_0 in $\mathscr{H}$ for which $H_0 \subset G \cap F^c$ and $\alpha(H_0) > \beta(G \cap F^c) - \varepsilon$. Now choose an H_1 in $\mathscr{H}$ for which $H_1 \subset G \cap H_0^c$ and $\alpha(H_1) > \beta(G \cap H_0^c) - \varepsilon$. Since H_0 and H_1 are disjoint and are contained in G (see Fig. 3), it follows by (4.4) that $\beta(G) \geqq \alpha(H_0 \cup H_1) = \alpha(H_0) + \alpha(H_1) > \beta(G \cap F^c) + \beta(G \cap H_0^c) - 2\varepsilon \geqq \gamma(G \cap F^c) + \gamma(G \cap F) - 2\varepsilon$. Since ε was arbitrary, this proves (4.8).

We turn to the converse problem of showing that a relatively compact Π must be tight. Consider a sequence $A_1, A_2, \cdots$ of open spheres of radius δ that cover S. For each ε, there exists an n such that, if $B_n = \bigcup_{i \leqq n} A_i$, $P(B_n) > 1 - \varepsilon$ for all P in Π, because otherwise for each n we have $P_n(B_n) \leqq 1 - \varepsilon$ for some P_n in Π,

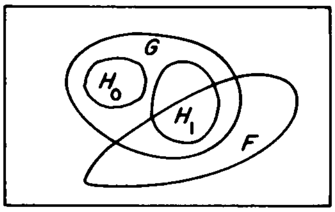

FIG. 3

and by relative compactness $P_{n_i} \Rightarrow P_0$ for some subsequence $\{P_{n_i}\}$ and probability measure P_0, which is impossible because $P_0(B_n) \leqq \liminf_i P_{n_i}(B_n) \leqq \liminf_i P_{n_i}(B_{n_i}) \leqq 1 - \varepsilon$ while $B_n \uparrow S$.

Thus for each positive ε and δ, there are finitely many spheres $A_1, \cdots, A_n$ of radius δ such that $P(\bigcup_{i \leqq n} A_i) > 1 - \varepsilon$ for all P in Π. Choose spheres $A_{k1}, \cdots, A_{kn_k}$ of radius $1/k$ such that $P(\bigcup_{i \leqq n_k} A_{ki}) > 1 - \varepsilon/2^k$. If K is the closure of the totally bounded set $\bigcap_{k \geqq 1} \bigcup_{i \leqq n_k} A_{ki}$, then K is compact and $P(K) > 1 - \varepsilon$ for all P in Π.

5. The space C. From here on we shall be concerned with the space C of continuous functions $x = x(t)$ on the closed unit interval, metrized by

$$\rho(x, y) = \sup_{0 \leqq t \leqq 1} |x(t) - y(t)|.$$

We denote by $\mathscr{C}$ the σ-field of Borel sets in C, and we shall be concerned with probability measures on $(C, \mathscr{C})$.

If $0 \leqq t_1 < t_2 < \cdots < t_k \leqq 1$, the mapping $\pi_{t_1 \cdots t_k}(x) = (x(t_1), \cdots, x(t_k))$ carries C continuously into R^k. Sets of the form $\pi_{t_1 \cdots t_k}^{-1} H$ with H an element of $\mathscr{R}^k$, a Borel set in R^k (k and $t_1, \cdots, t_k$ arbitrary), are called finite-dimensional sets, and the finite-dimensional sets form a finitely additive field. The closed sphere of radius r about x is the intersection of the finite-dimensional sets $\{y : |y(t) - x(t)| \leqq r\}$ with t ranging over the rationals; each open sphere is a countable union of closed spheres and each open set is a countable union of open spheres and hence lies in the σ-field generated by the finite-dimensional sets. Thus the finite-dimensional sets form a finitely additive field generating $\mathscr{C}$.

For a probability measure P on C, the various measures $P\pi_{t_1 \cdots t_k}^{-1}$ on the spaces R^k are called the finite-dimensional distributions of P. If two measures have the same finite-dimensional distributions, they agree for finite-dimensional sets, and hence, since these sets constitute a field generating $\mathscr{C}$, they are the same measure. Thus *the finite-dimensional distributions* $P\pi_{t_1 \cdots t_k}^{-1}$ *of* P *uniquely determine* P *itself.*

Suppose now that $x(t) \equiv 0$ and that x_n is the function given by

$$x_n(t) = \begin{cases} nt & \text{if } 0 \leqq t \leqq \dfrac{1}{n}, \\ 2 - nt & \text{if } \dfrac{1}{n} \leqq t \leqq \dfrac{2}{n}, \\ 0 & \text{if } \dfrac{2}{n} \leqq t \leqq 1; \end{cases}$$