APPENDIX I

Some Limit Theorems of Probability Theory

In this appendix a brief survey of the basic facts dealing with the convergence of probability distributions in spaces $R^k$ and some infinite dimensional spaces which are used throughout this book is presented. Those proofs which can be found in elementary texts are omitted. Basic references are M. Loève [94] and P. Billingsley [10]. All the random variables are considered over a general probability space $\{\Omega, \mathcal{F}, P\}$; the members of $\Omega$ are denoted by $\omega$.

1 Convergence of Random Variables and Distributions in $R^k$

We shall consider distributions in the measurable space $(R^k, \mathcal{B}^k)$, where $\mathcal{B}^k$ is a $\sigma$-algebra of Borel subsets of $R^k$. We say that a sequence of probability distributions $\mathcal{P}_n$ in $R^k$ is weakly convergent to the probability distribution $\mathcal{P}$ in $R^k$ if for any continuous bounded function $\phi: R^k \rightarrow R$,

$$\int_{R^k} \phi(x) d\mathcal{P}_n \rightarrow \int_{R^k} \phi(x) d\mathcal{P}.$$ 

It is the usual practice, which is also followed in this book, to call "weak convergence of distributions" simply "convergence of distributions." The convergence of $\mathcal{P}_n$ to $\mathcal{P}$ will be denoted symbolically by: $\mathcal{P}_n \rightarrow \mathcal{P}$.

A random vector $\xi$ with values in $R^k$ defines in $R^k$ a distribution $\mathcal{P}_\xi(A) = P(\xi \in A)$. If $\mathcal{P}_{\xi_n} \rightarrow \mathcal{P}_\xi$ we say that a sequence of vectors $\xi_n$ converges in distribution to the vector $\xi$ and we write $\mathcal{L}(\xi_n) \rightarrow \mathcal{L}(\xi)$. 

363
Note that if \( \xi_n \to \xi \) with probability 1 or in probability then also \( \mathcal{L}(\xi_n) \to \mathcal{L}(\xi) \).

A set \( A \subset \mathbb{R}^k \) is called the set of continuity of the distribution \( \mathcal{P} \) (or a \( \mathcal{P} \)-continuous set) if \( \mathcal{P}(\partial A) = 0 \), where \( \partial A \) is the boundary of the set \( A \).

**Theorem 1** (see [10]). Let \( \mathcal{P}_n \) and \( \mathcal{P} \) be distributions in \( \mathbb{R}^k \). The following conditions are equivalent:

1. \( \mathcal{P}_n \to \mathcal{P} \).
2. \( \int_{\mathbb{R}^k} \varphi \, d\mathcal{P}_n \to \int_{\mathbb{R}^k} \varphi \, d\mathcal{P} \) for all finite infinitely differentiable functions \( \varphi : \mathbb{R}^k \to \mathbb{R}^1 \).
3. \( \lim_{n \to \infty} \mathcal{P}_n(A) = \mathcal{P}(A) \) for all \( \mathcal{P} \)-continuous sets \( A \).
4. \( \lim_{n \to \infty} \mathcal{P}_n(A) = \mathcal{P}(A) \) for all continuous parallelepipeds \( A \).

Theorem 1 immediately implies that the following conditions are equivalent:

1. \( \mathcal{L}(\xi_n) \to \mathcal{L}(\xi) \).
2. \( \lim_{n \to \infty} \mathbb{E}\varphi(\xi_n) = \mathbb{E}\varphi(\xi) \) for any finite infinitely differentiable function \( \varphi : \mathbb{R}^k \to \mathbb{R}^1 \).
3. \( \lim_{n \to \infty} \mathbb{P}\{\xi_n \in A\} = \mathbb{P}\{\xi \in A\} \) for any set \( A \) with the property \( \mathbb{P}(\partial A) = 0 \).

**Theorem 2** (see [10], p. 50). If \( \mathcal{L}(\xi_n) \to \mathcal{L}(\xi) \) then

\[
\lim_{n \to \infty} \mathbb{E}|\xi_n| \geq \mathbb{E}|\xi|.
\]

The random variables \( \{\xi_n\} \) are called uniformly integrable if

\[
\lim_{N \to \infty} \sup_{n} \int_{\Omega} |\xi_n(\omega)| \mathbb{1}(|\xi_n(\omega)| > N) \, d\mathbb{P} = \lim_{N \to \infty} \sup_{n} \mathbb{E}\{|\xi_n| \mathbb{1}(|\xi_n| > N)\} = 0.
\]

**Theorem 3** (see [10], p. 51). Let the random variables \( \xi_n, \xi \) be nonnegative, \( \mathbb{E}\xi < \infty \) and \( \mathcal{L}(\xi_n) \to \mathcal{L}(\xi) \). Then

\[
\lim_{n \to \infty} \mathbb{E}\xi_n = \mathbb{E}\xi
\]

if and only if the variables \( \{\xi_n\} \) are uniformly integrable.

**Corollary.** If \( \mathcal{L}(\xi_n) \to \mathcal{L}(\xi) \) and for some \( r > 0 \) \( \sup_n \mathbb{E}|\xi_n|^r < \infty \), then \( \lim_{n \to \infty} \mathbb{E}|\xi_n|^r = \mathbb{E}|\xi|^r \) for all \( 0 < r' < r \). (Indeed the sequence \( |\xi_n|^r \) is uniformly integrable.

The following theorem is also a corollary of the two preceding ones.

**Theorem 4.** Let the random variables \( \xi_n \) converge in probability to the random variable \( \xi \).

1. If \( \lim_{n \to \infty} \mathbb{E}|\xi_n| \leq \mathbb{E}|\xi| \), then the sequence \( \xi_n \) is uniformly integrable and

\[
\lim_{n \to \infty} \mathbb{E}|\xi_n - \xi| = 0.
\]
(2) If $\lim_{n \to \infty} E|\xi_n|^2 \leq E|\xi|^2$, then $\lim_{n \to \infty} E|\xi_n - \xi|^2 = 0$.

Let $\xi$ be a random vector in $\mathbb{R}^k$ with distribution $\mathcal{P}_\xi$. Its characteristic function is defined by

$$\varphi_\xi(t) = \mathbb{E}e^{it, \xi} = \int_{\mathbb{R}^k} e^{it, x} d\mathcal{P}_\xi.$$ 

As it is known, $\varphi_\xi$ and $\mathcal{P}_\xi$ determine each other in a one-to-one manner. The following well known theorem asserts that this correspondence between characteristic functions in distributions in $\mathbb{R}^k$ is continuous.

**Theorem 5** (see [10] p. 72). Convergence $\mathcal{P}_n \to \mathcal{P}$ is valid if and only if $\lim_{n \to \infty} \varphi_n(t) = \varphi(t)$, where $\varphi_n$ and $\varphi$ are characteristic functions of $\mathcal{P}_n$ and $\mathcal{P}$ respectively. This convergence is necessarily uniform on any compact set.

The last theorem allows us to reduce the convergence of distributions in $\mathbb{R}^k$ to the convergence of distributions in $\mathbb{R}^1$ by means of a simple device due to Cramér and Wold.

**Theorem 6.** Let $\xi_n, \xi$ be random variables with values in $\mathbb{R}^k$. In order that $\mathcal{L}(\xi_n) \to \mathcal{L}(\xi)$ it is necessary and sufficient that $\mathcal{L}(\xi_n, t) \to \mathcal{L}(\xi, t)$ for any $t \in \mathbb{R}^k$.

Indeed, in view of Theorem 5, $\mathcal{L}(\xi_n, t) \to \mathcal{L}(\xi, t)$ if and only if for all $\lambda \in \mathbb{R}^1, t \in \mathbb{R}^k$, $\mathbb{E}e^{i\lambda(\xi_n, t)} \to \mathbb{E}e^{i\lambda(\xi, t)}$, i.e., if $\varphi_{\xi_n}(t) \to \varphi_{\xi}(t)$ for all $t \in \mathbb{R}^k$. □

Let the distributions $\mathcal{P}_{n\theta}, \mathcal{P}_\theta$ depend on the parameter $\theta \in \Theta$. We say that the distribution $\mathcal{P}_{n\theta}$ converges to $\mathcal{P}_\theta$ uniformly in $\Theta$ if for any continuous bounded function $g : \mathbb{R}^k \to \mathbb{R}^1$,

$$\int_{\mathbb{R}^k} g(x) d\mathcal{P}_{n\theta} \to \int_{\mathbb{R}^k} g(x) d\mathcal{P}_\theta$$

uniformly in $\theta \in \Theta$.

**Theorem 7.** Let the distributions $\mathcal{P}_{n\theta}, \mathcal{P}_\theta, \theta \in \Theta$, possess the following properties uniformly for all $\theta \in \Theta$.

(1) $\sup_{\theta \in \Theta} \mathcal{P}_{n\theta}(\{x : |x| > H\}) \to 0$, $H \to \infty$;

(2) $\varphi_{n\theta}(t) = \int_{\mathbb{R}^k} e^{it, x} d\mathcal{P}_{n\theta} \to \int_{\mathbb{R}^k} e^{it, x} d\mathcal{P}_\theta = \varphi_\theta(t)$.

Then $\mathcal{P}_{n\theta} \to \mathcal{P}_\theta$ uniformly in $\Theta$.

**Proof.** In view of (3), relation 1 is valid for all functions $e_i(x) = e^{it, x}$. Therefore it is fulfilled for functions $g$ of the form $\sum c_k e_i(x)$. By the Weierstrass theorem on approximating continuous functions by means of trigonometric
polynomials (and taking (2) into account), (1) is valid for all finite continuous $g$. But then in view of (2), it is valid for all bounded continuous $g$. \hfill \Box

If $w$ is bounded but not a continuous function then the convergence $P_n \to P$ in general does not imply the convergence of $\int w \, dP_n$ to $\int w \, dP$. Nevertheless the following result is valid.

**Theorem 8.** Let the distributions $P_{n\theta} \to P_{\theta}$ uniformly in $\Theta$, where $\Theta$ is a closed bounded subset in $R^k$. Assume that the family $\{P_{\theta}\}$ is continuous in $\Theta$ with respect to weak convergence. Then for any bounded Borel measurable function $w$ such that $P_{\theta}$-measures of the set of points of discontinuity of the function $w$ equal zero,

$$\lim \int_{R^k} w(x) \, dP_{\theta} = \int_{R^k} w(x) \, dP_{\theta}.$$  

**Proof.** Fix a number $\varepsilon > 0$. Given $\theta \in \Theta$, construct an open set $O$ which contains all the discontinuities of $w$ and such that $P_{\theta}(O) \leq \varepsilon$. In view of the continuity of the family $\{P_{\theta}\}$, the set $O$ can be chosen to be the same for all $\theta \in \Theta$. Moreover, it may be considered that $\lim_{n \to \infty} P_{n\theta}(O) \leq 2\varepsilon$ uniformly in $\theta$. The function $w$ is continuous on the closed set $\overline{O}$; we extend it continuously over all of $R^k$ and denote the extended function by $\tilde{w}$. Then uniformly in $\Theta$ we have

$$\lim_{n \to \infty} \int_{R^k} \tilde{w} \, dP_{n\theta} = \int_{R^k} \tilde{w} \, dP_{\theta},$$

and hence

$$\limsup_{n \to \infty} \sup_{\theta} \left| \int_{R^k} w \, dP_{n\theta} - \int_{R^k} w \, dP_{\theta} \right| \leq \limsup_{n \to \infty} \sup_{\theta} \left| \int_{R^k} \tilde{w} \, dP_{n\theta} - \int_{R^k} \tilde{w} \, dP_{\theta} \right|$$

$$+ \limsup_{n \to \infty} \sup_{\theta} \int_{O} \tilde{w} \, dP_{n\theta} + \sup_{\theta} \int_{O} \tilde{w} \, dP_{\theta} + \limsup_{n \to \infty} \int_{O} w \, dP_{n\theta}$$

$$+ \sup_{\theta} \int_{O} w \, dP_{\theta} \leq 3\varepsilon \sup_{x} |w(x)|. \hfill \Box$$


2 Some Limit Theorems for Sums of Independent Random Variables

The proof of all the theorems presented in this subsection is given in Chapter V of the text [94]. First we note the following well known result (see [94], Section 16.3).
Theorem 9. Let $\xi_1, \xi_2, \ldots$ be independent random variables. If the (numerical) series $\Sigma E\xi_j, \Sigma D\xi_j$ converges then the series $\Sigma \xi_j$ converges with probability 1.

The following theorems are a different version of the law of large numbers. By the law of large numbers we mean assertions concerning the convergence of arithmetic means $n^{-1} \sum_1^n \xi_j$ of the random variables $\xi_j$ to the arithmetic mean of their expectations $n^{-1} \sum_1^n E\xi_j$, or more generally theorems concerning the convergence of normalized sums of random variables to a sequence of nonrandom quantities. First we shall present a uniform version of a well known theorem.

Theorem 10. Let $\xi_1, \xi_2, \ldots$ be a sequence of independent identically distributed random variables with a common distribution $P_\theta$ which depends on the parameter $\theta \in \Theta$. If

$$\sup_{\theta} E_\theta(\{\xi_j \geq H\}) \longrightarrow 0,$$

then for any $\varepsilon > 0$ we have uniformly in $\theta \in \Theta$,

$$\lim_{n \to \infty} P_\theta\left\{\left|n^{-1} \sum_1^n \left(\xi_j - E_\theta \xi_j\right)\right| > \varepsilon\right\} = 0.$$

Proof. It may be assumed that $E_\theta \xi_j = 0$. Denote by $\varphi(t)$ the characteristic function of the variables $\xi_j$. The characteristic function of the sum $n^{-1} \sum_1^n \xi_j$ equals $(\varphi(tn^{-1}))^n$ and in view of the condition of the theorem the difference

$$(\varphi(tn^{-1}))^n - 1| \leq n|\varphi(tn^{-1}) - 1|$$

$$= n \left| \int_{-\infty}^{\infty} \left( e^{itxn^{-1}} - 1 - itxn^{-1} \right) dP_\theta \right|$$

$$\leq \sup_{\theta} \left( \frac{t^2}{n} E_\theta|\xi_1| + |t| \int_{|x| > \sqrt{n}} |x| dP_\theta \right)$$

converges to zero uniformly in $\theta$. It remains to note that the interval $[-\varepsilon, \varepsilon]$ is the interval of continuity of a distribution concentrated at zero and then refer to Theorems 7 and 8.

Theorem 11 (Theorem on relative stability). Let $\xi_1, \xi_2, \ldots$ be a sequence of independent nonnegative random variables whose distributions $P_{\theta\xi_j}$ depend on parameter $\theta \in \Theta$, where $\Theta$ is a closed subset in $R^k$. If $E_\theta \xi_j < \infty, c_n = \sum_1^n E_\theta \xi_j$, and for any $\varepsilon > 0,

$$\sup_{\theta} C_n^{-1} \sum_1^n E_\theta\{\xi_j \geq c_n \varepsilon\} \longrightarrow 0,$$

then for any $\varepsilon > 0$ uniformly in $\theta \in \Theta$ we have

$$\lim_{n \to \infty} P_\theta\left\{|C_n^{-1} \sum_1^n \xi_j - 1| > \varepsilon\right\} = 0.$$
The proof is similar to the proof of the preceding theorem. In the same manner as in the preceding theorem it is proved that

$$\sup_{\theta} \left| \mathbb{E}_\theta \exp \left\{ itC_n^{-1} \sum_{j=1}^{n} \xi_j \right\} - 1 \right| \xrightarrow{n \to \infty} 0. \tag*{\square}$$

We present yet two other theorems on the strong law of large numbers.

**Theorem 12** (Kolmogorov's strong law of large numbers; see [94], Section 16.3). Let \( \xi_1, \xi_2, \ldots \) be a sequence of independent identically distributed random variables with \( \mathbb{E}|\xi_j| < \infty \). Then with probability 1,

$$\lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} \xi_j = \mathbb{E}\xi_1.$$  

The following generalization of the preceding theorem is due to Marcinkiewicz (see [94] Section 16.4).

**Theorem 13.** Let \( \xi_1, \xi_2, \ldots \) be a sequence of independent identically distributed random variables with \( \mathbb{E}|\xi_j|^r < \infty, 0 < r < 2 \). Then if \( r \geq 1 \) we have with probability 1,

$$\lim_{n \to \infty} n^{-1/r} \sum_{j=1}^{n} (\xi_j - \mathbb{E}\xi_j) = 0;$$

if, however, \( r < 1 \), then with probability 1,

$$\lim_{n \to \infty} n^{-1/r} \sum_{j=1}^{n} \xi_j = 0.$$  

We now proceed to theorems on convergence to the normal law.

**Theorem 14** (see [94], Section 20.2). Let \( \xi_1, \xi_2, \ldots \) be a sequence of independent random variables with the distributions \( \mathcal{F}_j \) and let \( \mathbb{E}\xi_j^2 < \infty \). Set \( B_n = \sum_{j=1}^{n} D\xi_j \). If for some \( \varepsilon > 0 \),

$$B_n^{-2} \sum_{j=1}^{n} \mathbb{E}(\xi_j^2 \chi(|\xi_j - \mathbb{E}\xi_j| > \varepsilon B_n))$$

$$= B_n^{-2} \sum_{j=1}^{n} \int_{|x - \mathbb{E}\xi_j| > \varepsilon B_n} (x - \mathbb{E}\xi_j)^2 d\mathcal{F}_j \xrightarrow{n \to \infty} 0, \tag{4}$$

then the normalized sum

$$B_n^{-1} \sum_{j=1}^{n} (\xi_j - \mathbb{E}\xi_j)$$

is asymptotically normal with parameters \((0, 1)\), i.e., the distribution of these sums converges to the normal distribution with parameters \((0, 1)\).
Condition (4) is called the *Lindeberg condition*; it is a fortiori fulfilled if there exist moments $E|\xi_j|^{2+\delta}, \delta > 0$, and Liapunov's ratio tends to zero:

$$ L_{n\delta} = B_n^{-1-\delta/2} \sum_{i=1}^{n} E|\xi_j|^{2+\delta} \longrightarrow 0. $$

The last condition is called Liapunov’s condition.

**Theorem 15.** Assume the conditions of Theorem 14 are fulfilled, the distribution $P_{\theta^n}$ of the random variables $\xi_j$ depends on the parameter $\theta \in \Theta$, and for some $\varepsilon > 0$ we have uniformly in $\theta \in \Theta$,

$$ \Lambda_n(\varepsilon) = B_n^{-2} \sum_{j=1}^{n} \int_{|x-E_\theta \xi_j| \geq \varepsilon B_n} (x-E_\theta \xi_j) dP_{\theta^n} \rightarrow 0. $$

Then

$$ \mathcal{L} \left( B_n^{-1} \sum_{j=1}^{n} (\xi_j - E_\theta \xi_j) \right) \rightarrow \mathcal{N}(0, 1) $$

uniformly in $\Theta$.

**Proof.** It may be assumed that $E_\theta \xi_j = 0$. Then $E_\theta \xi_j^2 = \sigma_j^2$. Then

$$ \max_{1 \leq j \leq n} \sigma_j^2 B_n^{-2} \leq \Lambda_n(\varepsilon) + \varepsilon. $$

If $\varphi_j(t)$ is the characteristic function of $\xi_j B_n^{-1}$, then

$$ \left| E_\theta \exp \left\{ it B_n^{-1} \sum_{j=1}^{n} \xi_j \right\} - e^{-t^2/2} \right| $$

$$ \leq \sum_{j=1}^{n} \left| \varphi_j(t) - \exp \left\{ - \frac{\sigma_j^2 t^2}{2B_n^2} \right\} \right| $$

$$ \leq \sum_{j=1}^{n} \left| \exp \left\{ - \frac{\sigma_j^2 t^2}{2B_n^2} \right\} - 1 + \frac{\sigma_j^2 t^2}{2B_n^2} \right| $$

$$ + \sum_{j=1}^{n} \left| \int_{-\infty}^{\infty} \left( e^{itx B_n^{-1}} - 1 - \frac{itx}{B_n} + \frac{t^2 x^2}{2B_n^2} \right) dP_{\theta^n} \right| $$

$$ \leq t^4 \max_{1 \leq j \leq n} \frac{\sigma_j^2}{B_n^2} + \varepsilon |t|^3 + t^2 \Lambda_n(\varepsilon) $$

and it is sufficient to refer to Theorem 7.

### 3 Weak Convergence on Function Spaces

Let $X$ be a metric space of real-valued functions $x(t), t \in T, T \subset R^k$ with the metric $\rho(x, y)$. The minimal $\sigma$-algebra of subsets of $X$ containing open sets is called the class of *Borel sets* $\mathfrak{B}$ in $X$. Let $P_n$ and $P$ be probability distributions
on the measurable space \((X, \mathcal{B})\). We say that \(\mathcal{P}_n\) converges weakly to \(\mathcal{P}\) if for any continuous function \(g: X \to \mathbb{R}'\), bounded on \(X\),

\[
\lim_{n \to \infty} \int_X g(x) \, d\mathcal{P}_n = \int_X g(x) \, d\mathcal{P}.
\]

As above we shall write “convergence” in place of “weak convergence.”

Let \(\xi(t), t \in T\), be a random process on \(T\) such that \(\xi(\cdot, \omega)\) are \(\mathcal{B}\)-measurable; if with probability 1 the realizations of this process \(\xi(\cdot) \in X\), then \(\xi(t)\) determines a probability distribution \(\mathcal{P}_\xi\) on \((X, \mathcal{B})\). Namely, \(\mathcal{P}_\xi(A) = \mathbb{P}\{\xi \in A\}\). Therefore analogously to Subsection 1 one may define convergence of \(\mathcal{P}_n(t)\) to \(\mathcal{P}_\xi(t)\) in distribution as the convergence of \(\mathcal{P}_\xi\) to \(\mathcal{P}_\xi\).

We note that Theorem 1 is carried over without any alterations to the case of convergence in metric spaces (see [10], p. 21). The notion of the uniform convergence of parametric families of distributions is also readily carried over to this case.

Let \(\mathcal{X} = \mathcal{X}(T)\) be a metric space of functions on \(T\). Let \(\mathcal{P}\) be a distribution on \((\mathcal{X}(T), \mathcal{B})\); choose points \(t_1, \ldots, t_k \in T\). Assume that all the sets of the form \(\{x: (x(t_1), \ldots, x(t_k)) \in A\}, A \in \mathcal{B}_k\), belong to \(\mathcal{B}\). Then the distribution \(\mathcal{P}\) induces the distribution \(\mathcal{P}_{t_1, \ldots, t_k}\) in \(R^k\) according to the formula

\[
\mathcal{P}_{t_1, \ldots, t_k}(A) = \mathbb{P}\{x: (x(t_1), \ldots, x(t_k)) \in A\}.
\]

The distributions \(\mathcal{P}_{t_1, \ldots, t_k}\) are called marginal distributions (induced by the distributions \(\mathcal{P}\)). If \(\mathcal{P}\) is a distribution of the random process \(\xi(t)\) then

\[
\mathcal{P}_{t_1, \ldots, t_k}(A) = \mathbb{P}\{(\xi(t_1), \ldots, \xi(t_k)) \in A\}
\]

are marginal distributions of the random process \(\xi(t)\).

The following question is of interest to us: under what conditions does the convergence of all the marginal distributions \(\mathcal{P}_{t_1, \ldots, t_k}^{(n)}\) to the marginal distributions \(\mathcal{P}_{t_1, \ldots, t_k}\) imply the convergence of distributions \(\mathcal{P}_{t_1, \ldots, t_k}^{(n)}\) to \(\mathcal{P}\)? Evidently it is necessary that distributions \(\mathcal{P}\) on \((\mathcal{X}, \mathcal{B})\) be uniquely determined by their marginal distributions. But are there any other conditions?

We say that the family \(\{\mathcal{P}\}\) of distributions on \((\mathcal{X}, \mathcal{B})\) is relatively compact if for any \(\varepsilon > 0\) there exists a compact set \(K\) in \(\mathcal{X}\) such that for all \(\mathcal{P}\) belonging to \(\{\mathcal{P}\}\), \(\mathbb{P}(K) > 1 - \varepsilon\).

The following theorem due to Yu. V. Prokhorov presents a useful test for relative compactness of families of distributions.
Theorem 16 (see [10], p. 58). If a family \{\mathcal{P}\} of distributions in \((\mathcal{X}, \mathcal{B})\) is dense then it is relatively compact.

For the purposes of this book it is sufficient to study in more detail the convergence of distributions in spaces of continuous functions \(R^k \rightarrow R^1\). Two types of such spaces will be required. First, the space \(C(F)\) of functions continuous on a compact subset \(F\) of the Euclidean space \(R^k\) and, second, the space \(C_0(R^k)\) of functions \(\psi(t)\) continuous on \(R^k\) which tend to zero as \(|t| \rightarrow \infty\). Both these spaces are considered as Banach spaces with the norm

\[ \|\psi\| = \sup_t |\psi(t)|.\]

In the case of spaces \(C(F)\) and \(C_0(R^k)\) the sets of the form \(\{\psi: (\psi(t_1), \ldots, \psi(t_k)) \in A\}, A \in \mathcal{B}^k\), are Borel sets and also sets generating the whole Borel set in \(C(F)\) or in \(C_0(R^k)\). Thus the distributions in \(C(F)\) and \(C_0(R^k)\) are uniquely determined by their marginal distributions (see [10], p. 33).

4 Conditions for the Density of Families of Distributions in \(C(F)\) and \(C_0(R^k)\) and Criteria for Uniform Convergence

Compact subsets of \(C(F)\) are described by the following well-known theorem of Arzela.

Theorem 17. A subset \(\Phi \subset C(F)\) possesses a compact closure in \(C(F)\) if and only if

1. The functions \(\varphi \in \Phi\) are uniformly bounded, i.e., \(\sup_{\Phi} \|\varphi\| < \infty\);
2. The functions \(\varphi \in \Phi\) are equicontinuous, i.e.,

\[ \lim_{\varepsilon \to 0} \sup_{\Phi} \sup_{|x-y| \leq \varepsilon} |\varphi(x) - \varphi(y)| = 0.\]

From Arzela's theorem the following criterion for compactness in \(C_0(R^k)\) is deduced.

Theorem 18. A subset \(\Phi \subset C_0(R^k)\) possesses a compact closure in \(C_0(R^k)\) if and only if

1. The functions \(\varphi(t) \in \Phi\) are uniformly bounded;
2. The functions \(\varphi(t) \in \Phi\) are equicontinuous; and
3. The functions \(\varphi(t) \in \Phi\) tend to zero uniformly as \(t \rightarrow \infty\), i.e.,

\[ \lim_{R \to \infty} \sup_{\Phi} \sup_{|t| > R} |\varphi(t)| = 0.\]
The criteria for the density of families of distributions in $C(F)$ and $C_0(R^k)$ are deduced from these two theorems and the following theorem on the properties of realizations of random functions.

**Theorem 19.** Let $\xi(t)$ be a real-valued random function defined on a closed subset $F$ of the Euclidean space $R^k$. We shall assume the random process $\xi(t)$ is measurable and separable. Assume that the following condition is fulfilled: there exist numbers $m \geq r > k$ and a function $H(x) : R^k \rightarrow R^1$ bounded on compact sets such that for all $x, h \in F, x + h \in F,$

$$E|\xi(x)|^m \leq H(x),$$

$$E|\xi(x + h) - \xi(x)|^m \leq H(x)|h|^r.$$  

(7)

Then with probability 1 the realizations of $\xi(t)$ are continuous functions on $F$. Moreover, set

$$\omega(\delta; \xi, L) = \sup |\xi(x) - \xi(y)|,$$

where the upper bound is taken over $x, y \in F$ with $|x - y| \leq \delta, |x| \leq L, |y| \leq L$; then

$$E(\omega(h; \xi, L)) \leq B_0 \left( \sup_{|x| < L} H(x) \right)^{1/m} L^{k(r-k)/m},$$

(8)

where the constant $B_0$ depends on $m, r$ and $k$.

**PROOF.** We shall consider separately three cases.

1) First we shall assume that the function $\xi(t)$ is defined on $R^k$ and vanishes outside a sphere. In this case it may be assumed that in the inequalities (7) $H(x) \equiv H$ does not depend on $x$. We choose and fix an infinitely differentiable nonnegative function $c : R^k \rightarrow R^1$ vanishing outside the unit sphere and such that $\int_{R^k} c(x) \, dx = 1$. Since $\int_{R^k} E|\xi(t)| \, dt < \infty$, it follows from Fubini's theorem that with probability 1 that $\int_{R^k} |\xi(t)| \, dt < \infty$ and one can define random functions

$$\eta(t; \varepsilon) = \varepsilon^{-k} \int_{R^k} \xi(y)c((t - y)\varepsilon^{-1}) \, dy, \quad \varepsilon > 0.$$ 

We have

$$E|\xi(t) - \eta(t; \varepsilon)| = E\left| \varepsilon^{-k} \int_{R^k} (\xi(y) - \xi(t))c((t - y)\varepsilon^{-1}) \, dy \right|$$

$$\leq \int_{R^k} E|\xi(t - \varepsilon) - \xi(t)|c(v) \, dv$$

$$\leq H^{1/m} \varepsilon^{r/m} \int_{R^k} |v|^{r/m} c(v) \, dv.$$
This, together with the separability of the random functions under consideration, implies that

\[ \xi(t) = \eta_1(t) + \sum_{n=1}^{\infty} (\eta_{n+1}(t) - \eta_n(t)), \]  

(9)

where \( \eta_n(t) = \eta(t; 2^{-n}) \).

Evidently all the functions \( \eta_n(t) \) are infinitely differentiable. If \( \eta^{(l)}_n(t) \) is an arbitrary derivative of order \( l \), then

\[
E|\eta^{(l)}_n(t)|^m = e^{-lm} E \left| \int_{R^k} (\xi(t + \varepsilon v) - \xi(t)) c^{(l)}(v) \ dv \right|^m \\
\leq e^{-lm} \int_{R^k} E|\xi(t + \varepsilon v) - \xi(t)|^m |c^{(l)}(v)| \ dv \\
\times \left( \int_{R^k} |c^{(l)}(v)| \ dv \right)^{m(m-1)} \leq HB^2e^{-lm}, \quad \varepsilon = 2^{-n},
\]

where here and below \( B \) denotes a constant which depends only on the dimension and the numbers \( m \) and \( r \). In particular,

\[
E|\eta^{(l)}_n(t)|^m \leq HB^2 2^{-nm}. \quad (10)
\]

We now bound \( \sup_{|t-s| \leq h} |\eta_n(t) - \eta_n(s)| \). We take a cube of the minimal size with center at the origin and edges parallel to the coordinate axes, outside of which all the functions \( \eta_n(t) = 0 \) and inside of which all \( \eta_n(t) = 0 \) if the point \( t \) is less than one unit distant from the boundary. Let an edge of this cube be of length \( L \). We subdivide this cube into smaller cubes with edges of length \( h \) parallel to the coordinates (if necessary \( L \) will be slightly increased). There are \( N = (Lh^{-1})^k \) of the smaller cubes, which are denoted by \( \Delta_j \). If \( |t - s| \leq h, t, s \in Q \), then \( t \) and \( s \) either belong to one cube \( \Delta_j \) or to two neighboring cubes. Therefore for any \( \lambda > 0, \)

\[
\left( \sup_{|t-s| \leq h} |\eta_n(t) - \eta_n(s)| \right)^\lambda \leq B \sum_j \left( \sup_{t, s \in \Delta_j} |\eta_n(t) - \eta_n(s)| \right)^\lambda. \quad (11)
\]

Furthermore, let \( t = (t_1, \ldots, t_k), i = (i_1, \ldots, i_k) \). We shall assume that \( i \) is a vertex of a cube \( \Delta_j \) and is such that \( t_j > i_j \) for all \( t \in \Delta_j \). We then have

\[
\eta_n(t) - \eta_n(i) = \int_{i_k}^{t_k} \frac{\partial}{\partial x_k} \eta_k(i_1, \ldots, i_k-1, x_k) \ dx_k + \int_{i_k}^{t_k} \left[ \frac{\partial}{\partial x_k} \eta_n(t_1, \ldots, t_k-1, x_k) - \frac{\partial}{\partial x_k} \eta_n(i_1, \ldots, i_k-1, x_k) \right] \ dx_k + [\eta_n(t_1, \ldots, t_k-1, i_k) - \eta_n(i)].
\]
Continuing in a similar way we arrive at the inequality for \( t \in \Delta_j \),

\[
|\eta_n(t) - \eta_n(t)| \leq \sum_{l=1}^{k} \sum_{i_1, \ldots, i_l} \int_{i_1}^{i_1 + h} \cdots \int_{i_l}^{i_l + h} \frac{\partial^l}{\partial x_{i_1} \cdots \partial x_{i_l}} \times \eta_n(t_1, \ldots, x_{i_l}, \hat{t}_i, \ldots, \hat{t}_k) \, dx_{i_1} \cdots dx_{i_l}.
\]

Taking the bound (10) into account we obtain from the last inequality that

\[
E \left( \sup_{t, s \in \Delta_j} |\eta_n(t) - \eta_n(s)| \right)^m \leq BH \sum_{j=1}^{k} h^{mj} 2^{-n(r-mj)}.
\]

If \( 2^n h \leq 1 \) this inequality implies in particular the bound

\[
E \left( \sup_{t, s \in \Delta_j} |\eta_n(t) - \eta_n(s)| \right)^m \leq BH 2^{-nr} (2^n h)^m
\]  
(12)

and hence for such \( n \) and \( h \) we have

\[
E \left( \sup_{|l-s| \leq h} |\eta_n(t) - \eta_n(s)| \right)^m \leq L^k HB 2^{-n(r-m)h^{m-k}}.
\]  
(13)

We now bound \( E \sup_t |\eta_n(t) - \eta_{n+1}(t)| \). First,

\[
E|\eta_n(t) - \eta_{n+1}(t)|^m \leq 2^m E|\eta_n(t) - \zeta(t)|^m + E|\eta_{n+1}(t) - \zeta(t)|^m.
\]

Next

\[
E|\eta_n(t) - \zeta(t)|^m \leq \int_{R^k} E|\zeta(t + y 2^{-n}) - \zeta(t)|^m c(y) \, dy \leq HB 2^{-nr}.
\]  
(14)

We subdivide again the original cube into small cubes \( \Delta_j \) with edges of size \( \delta = 2^{-n} \). If \( \Delta_j \) is an arbitrary subdividing cube and \( t_0 \) is its center then, in view of (12) and (14),

\[
E \left( \sup_{\Delta_j} |\eta_n(t) - \eta_{n+1}(t)| \right)^m \leq 3^m (E|\eta_n(t_0) - \eta_{n+1}(t_0)|^m)
\]

\[
+ E \left( \sup_{t, s \in \Delta_j} |\eta_n(t) - \eta_n(s)| \right)^m
\]

\[
+ E \left( \sup_{t, s \in \Delta_j} |\eta_{n+1}(t) - \eta_{n+1}(s)| \right)^m
\]

\[
\leq 3^m HB(2^{-nr} + 2^{-nr}(2^n \delta)^m) \leq HB 2^{-nr}.
\]

Consequently,

\[
E \left( \sup_t |\eta_n(t) - \eta_{n+1}(t)| \right)^m \leq L^k HB 2^{-nr}.
\]
4 Conditions for the Density of Families of Distributions in $C(F)$ and $C_0(R')$ 375

and hence

$$
\mathbf{E} \sup_t |\eta_n(t) - \eta_{n+1}(t)| \leq L^{k/m} H^{1/m} B \cdot 2^{-(r-k)/m}.
$$

(15)

Fixing a number $h$, $0 < h < 1$, we choose $N$ in such a manner that

$$
2^N h \leq 1; 2^{N+1} h > 1.
$$

Relations (13) and (15) yield the following inequality:

$$
\mathbf{E} \sup_{|t-s| \leq h} |\xi(t) - \xi(s)| \leq 2 \sum_{n=1}^N \mathbf{E} \left( \sup_{|t-s| \leq h} |\eta_n(t) - \eta_n(s)| \right)
+ 2 \sum_{N+1}^\infty \mathbf{E} \left( \sup_t |\eta_{n+1}(t) - \eta_n(t)| \right)
\leq BH^{1/m} L^{k/m} h^{(r-k)/m}.
$$

(16)

Thus this theorem is proved under the conditions of Subsection 1.

2) The set $F$ coincides with $R^k$. Introduce the following function

$$
\varphi: R^k \to R^1:
$$

$$
\varphi(x) = \begin{cases} 
1, & \text{if } |x| \leq 1, \\
2 - |x|, & \text{if } 1 \leq |x| \leq 2, \\
0, & \text{if } |x| \geq 2.
\end{cases}
$$

Along with the function $\xi(t)$ consider the function

$$
\xi_L(t) = \xi(t) \varphi \left( \frac{t}{L} \right).
$$

For this random function the condition of the preceding subsection is fulfilled; consequently if $|t|, |s| \leq L$, then

$$
\mathbf{E} \sup_{|t-s| \leq h} |\xi(t) - \xi(s)| \leq B \left( \sup_{|t| \leq L} K(t) \right)^{1/m} L^{k/m} h^{(r-k)/m},
$$

(17)

which is equivalent to the assertion of the lemma.

3) The general case: $F$ is an arbitrary closed subset of $R^k$. Utilizing Whitney’s construction (see Chapter VI in the text [113]) we construct an extension $\tilde{\xi}(t)$ of the function $\xi(t)$ from the set $F$ over the whole $R^k$ retaining the condition (7). After that it will be sufficient to refer to Subsection 2.

Let $F$ be the complement for $F$. It was shown in Sections 1 and 2 of Chapter VI of [113] that one can define on $\bar{F}$ a family of functions $\{\varphi^*_F(x)\}$ possessing the following properties:

(1) The set $F$ is subdivided into pairwise disjoint cubes $Q_k$ and, moreover, $\text{diam } Q_k \leq \text{dist } (Q_k, F) \leq 4 \text{ diam } Q_k$, where $\text{diam } A$ is the diameter of the set $A$ and $\text{dist}(A, B)$ is the distance between the sets $A$ and $B$. 

(2) All the functions $\varphi_r^*$ are such that $0 \leq \varphi_r^* \leq 1$, $\varphi_r^* > 0$ on $Q_r$, are infinitely differentiable, and possess bounded support $Q_r^*$ such that $\text{diam } Q_r \leq \text{diam } Q_r^* \leq 2 \text{ diam } Q_r$.

(3) Each point $t \in \tilde{F}$ belongs to at most $N$ of the sets $Q_r^*$, where $N$ depends only on the dimension $k$ of the space ($N \leq 12^k$).

(4) $\sum_r \varphi_r^*(x) = 1, \quad x \in \tilde{F}$.

(5) $\left| \frac{\partial}{\partial x} \varphi_r^*(x) \right| \leq B (\text{diam } Q_r)^{-1}$.

Now let a point $p_r \in F$ be defined by the relationship $\text{dist } (Q_r, F) = \text{dist } (Q_r, p_r)$. Set

$$\tilde{\xi}(t) = \xi(t) \quad \text{for } t \in F, \quad \tilde{\xi}(t) = \sum_r \xi(p_r) \varphi_r^*(t) \quad \text{for } t \in \tilde{F}.$$ 

We will prove that $\tilde{\xi}(t)$ is the required extension of the function $\xi(t)$. (For each $t \in \tilde{F}$ the sum in (20) possesses at most $N$ summands.)

Denote by $\delta(x)$ the distance from $x$ to $F$. We shall prove the inequality

$$\mathbb{E} \left| \frac{\partial}{\partial t_i} \tilde{\xi}(t) \right|^m \leq B \max_{j: t \in Q_r^*} H(p_j)(\delta(t))^{-m}.$$ 

Indeed, in view of (18),

$$\sum_r \frac{\partial \varphi_r^*}{\partial t_i} (t) \equiv 0,$$

so that

$$\frac{\partial}{\partial t_i} \tilde{\xi}(t) = \sum_j \left( \xi(p_j) - \xi(y) \right) \frac{\partial \varphi_r^*}{\partial t_i} (t),$$

where the point $y \in F$ is chosen in such a manner that $|t - y| = \delta(t)$.

The sum in (22) contains only $N$ summands: the summation is taken over $j$ such that $t \in Q_r^*$. For all such $j$,

$$|y - p_j| \leq B |t - y| = B\delta(t).$$

Hence utilizing (19) we have

$$\mathbb{E} \left| \frac{\partial}{\partial t_i} \tilde{\xi}(t) \right|^m \leq B \max_{j: t \in Q_r^*} \mathbb{E} |\xi(p_j) - \xi(y)|^m(\delta(t))^{-m} \leq B \max_{j: t \in Q_r^*} [H(p_j)|p_j - y|^r](\delta(t))^{-m} \leq B \max_{j: t \in Q_r^*} H(p_j)(\delta(t))^{-m},$$

which proves (21).
Similarly from (19) follows the inequality
\[ E \left| \frac{\partial}{\partial t_i} \xi(t) \right|^m \leq B \max_{t \in Q_j} H(p_j)(\delta(t))^{-m}, \] (23)
which is preferable to (21) provided \( \delta(t) \geq 1 \).

We shall now prove that \( \xi(t) \) satisfies the relations of (7). The first one is obvious: if \( t \in F \) then
\[ E |\xi(t)|^m \leq E \left( \sum_j |\xi(p_j)|^m \varphi_j^*(t) \right)^{m/(m-1)} \leq \sup_{j: t \in Q_j^*} H(p_j). \]
For the proof of the second inequality in (7) we shall consider separately several cases.

(a) Both points \( t \) and \( s \) belong to \( F \); this is a trivial case.
(b) The point \( s \in F \) and the point \( t \in \overline{F} \). We then have
\[ |\xi(s) - \xi(t)| = \left| \sum_{t \in Q_j^*} (\xi(s) - \xi(p_j))\varphi_j^*(t) \right|. \]
Observe that \( |s - p_j| \leq |s - t| + |p_j - t| \). Clearly \( |s - t| \geq \text{dist} (Q_j^*, F) \), and \( |t - p_j| \sim \text{diam} Q_1 \sim \text{dist} (Q_j^*, F) \sim \delta(t) \). (Here and below the notation \( a \sim b \) means that the ratio \( a/b \) is contained between two positive constants which may depend only on the dimension.) Therefore \( |s - p_j| \leq B|s - t| \) and, in view of Minkowski's inequality,
\[ E |\xi(s) - \xi(t)|^m \leq \sum_{t \in Q_j^*} E |\xi(s) - \xi(p_j)|^m \varphi_j^*(t) \leq B \max_{j: t \in Q_j^*} H(p_j)|s - t|^m. \] (24)
It remains to consider the cases when both \( t \) and \( s \in \overline{F} \). Denote by \( L \) the segment joining the points \( t \) and \( s \).

(c) Both \( t \) and \( s \in \overline{F} \) and \( \text{dist}(L, F) \geq 1 \). The required bound immediately follows from (23).

(d) Both \( t \) and \( s \in \overline{F} \) and \( 1 > \text{dist}(L, F) > |s - t| \). We have, taking inequality (21) into account, that
\[ E |\xi(t) - \xi(s)|^m \leq B \sup_{t \in Q_j^*} H(p_j)|t - s|^m \sup_{y \in L} \text{dist} (y, F)^{-m}) \leq B \sup_{t \in Q_j^*} H(p_j)|t - s|^m, \] (25)
where the upper bound on the right-hand side is taken over all \( j \) such that the support of \( Q_j^* \) intersects \( L \).

(e) Both \( t \) and \( s \in \overline{F} \) and
\[ \text{dist} (L, F) < |s - t|, \quad \text{dist} (L, F) \leq 1. \]
Then there exists a point \( t' \in L \) and a point \( y' \in F \) such that \( |t' - y'| \leq |t - s| \). Consequently, \( |y' - t| \leq 2|t - s| \) or \( |y' - s| \leq 2|t - s| \).
Since \( y' \in F \) and \( t \) and \( s \in F \) we have, in view of inequality (24)
\[
\mathbb{E} |\tilde{\xi}(t) - \tilde{\xi}(s)|^m \leq 2^m(\mathbb{E} |\tilde{\xi}(t) - \tilde{\xi}(y')|^m + \mathbb{E} |\tilde{\xi}(s) - \tilde{\xi}(y')|^m)
\leq B|s - t|^r \left( \max_{j: t \in Q_j^s} H(p_j) + \max_{j: s \in Q_j^t} H(p_j) \right).
\]

Using the theorem just proved it is easy to construct convenient criteria for convergence of distributions in \( C(F) \) or \( C_0(R^k) \). We now present several such criteria.

**Theorem 20.** Let the random processes \( \xi_{n\theta}(t), n = 1, 2, \ldots \) and \( \xi_\theta(t) \) which depend on the parameter \( \theta \in \Theta \) be defined and continuous with probability 1 on a bounded closed set \( F \subset R^k \). Assume that the following conditions are fulfilled:

1. There exist numbers \( m \geq r > k \) and a number \( H > 0 \) such that
\[
\mathbb{E} |\xi_{n\theta}(t + h) - \xi_{n\theta}(t)|^m \leq H|h|^r,
\]
\[
\mathbb{E} |\xi_{n\theta}(t)|^m \leq H.
\]

2. The marginal distributions of the random functions \( \xi_{n\theta}(t) \) converge as \( n \to \infty \) to the marginal distributions of \( \xi_\theta(t) \) uniformly in \( \Theta \).

Then the distributions \( \mathcal{P}_{n\theta} \) in \( C(F) \) generated by the random processes \( \xi_{n\theta} \) converge to the distribution \( \mathcal{P}_\theta \) in \( C(F) \) generated by \( \xi_\theta \) uniformly in \( \Theta \).

**Proof.** Theorem 3 implies that condition 1 is fulfilled with the same constants also for the function \( \xi_\theta(t) \). Let \( \psi \) be a bounded continuous functional in \( C(F) \). It is required to prove that
\[
\lim_{n \to \infty} \mathbb{E} \psi(\xi_{n\theta}) = \mathbb{E} \psi(\xi_\theta)
\]
uniformly in \( \theta \in \Theta \). Denote by \( K = K(M_1, M_2) \) the set of functions \( g \) belonging to \( C(F) \) possessing the following properties

1. \( \|g\| \leq M_1 \).
2. \( |g(t + h) - g(t)| \leq M_2|h|^{(r-k)/2m} \).

In view of Theorem 19, for any \( \varepsilon > 0 \) there exist \( M_1 = M_1(\varepsilon) \) and \( M_2 = M_2(\varepsilon) \) such that for all \( n \) and
\[
P\{\xi_{n\theta} \in K_\varepsilon\} > 1 - \varepsilon, \quad P\{\xi_\theta \in K_\varepsilon\} > 1 - \varepsilon, \quad (26)
\]
where \( K_\varepsilon = K(M_1(\varepsilon), M_2(\varepsilon)) \).

Subdivide the whole space \( R^k \) into cubes \( \Delta_j \) of diameter \( \delta \) with edges parallel to the coordinate axes. Let \( Q \) be the cube of unit length with center at the origin. Choose and fix an infinitely differentiable function \( \varphi, 0 \leq \varphi \leq 1, \)
which equals one on the cube \( Q \) and zero outside the cube \( 2Q \). If \( t_{0j} \) is the center of the cube \( \Delta_j \) and \( \delta \) is the length of its edge, we then set

\[
\varphi_j(t) = \varphi \left( \frac{t - t_{0j}}{\delta} \right).
\]

Finally, let

\[
\phi^*_j(t) = \frac{\varphi_j(t)}{\sum \varphi_j(t)};
\]

clearly the functions \( \phi^*_j \) form a partition of unity: \( \sum \phi^*_j(t) \equiv 1 \).

From each cube \( \Delta_j \) which intersects \( F \) we choose a point \( t_j \in F \). Define on \( C(F) \) an operator \( A_\delta \) acting on functions \( g \in C(F) \) in the following manner:

\[
A_\delta(g) = \sum_j g(t_j)\phi^*_j(t).
\]

Clearly, if \( K \) is a compact set in \( C(F) \) it follows that

\[
\|A_\delta(g) - g\| \leq \sup_{|t - s| \leq 2\delta \sqrt{K}} |g(t) - g(s)| \leq \omega(\delta; K), \tag{27}
\]

where

\[
\omega(\delta; K) = \sup_{g \in K} \sup_{|t - s| \leq 2\delta \sqrt{K}} |g(t) - g(s)|.
\]

Now construct the random functions

\[
\eta_{n\theta}(t) = A_\delta(\xi_{n\theta})(t), \quad \eta_{\theta}(t) = A_\delta(\xi_{\theta})(t).
\]

Let \( \psi \) be an arbitrary continuous bounded functional on \( C(F) \). It is uniformly continuous on any compact set \( K \) belonging to \( C(F) \) and we set

\[
\Omega(\psi; K; \delta) = \sup \{ |\psi(g_1) - \psi(g_2)| : g_1, g_2 \in K, \|g_1 - g_2\| < \delta \}.
\]

Let

\[
M_\psi = \sup_{g \in C(F)} |\psi(g)|.
\]

By virtue of the convergence of the marginal distributions uniformly in \( \Theta \) we have \( \lim_n E\psi(\eta_{n\theta}) = E\psi(\eta_{\theta}) \). Furthermore, in view of (26) and (27) we have

\[
E\psi(\xi_{n\theta}) - E\psi(\eta_{n\theta}) \leq 2M_\psi \mathbb{P}\{\xi_{n\theta} \notin K_\varepsilon\}
\]

\[
+ E\{ |\psi(\xi_{n\theta}) - \psi(\eta_{n\theta})| \chi(\xi_{n\theta} \in K_\varepsilon) \}
\]

\[
\leq 2\varepsilon M_\psi + \Omega(\psi; K_\varepsilon; \omega(\delta, K_\varepsilon)).
\]

Analogously, also

\[
|E\psi(\xi_{\theta}) - E\psi(\eta_{\theta})| \leq 2M_\psi \varepsilon + \Omega(\psi; K_\varepsilon; \omega(\delta, K_\varepsilon)).
\]

Hence uniformly in \( \Theta \), \( \lim_n E\psi(\xi_{n\theta}) = E\psi(\xi_{\theta}) \).
Theorem 21. Let random functions \( \xi_{n\theta} \) and \( \xi_\theta(t) \) which depend on a parameter \( \theta \) be defined on \( \mathbb{R}^k \) and let \( \xi_{n\theta} \in C_0(\mathbb{R}^k) \) and \( \xi_\theta \in C_0(\mathbb{R}^k) \) with probability 1. Assume that the following conditions are fulfilled.

1. There exist numbers \( m \geq r > k \) such that for any cube \( Q \) in \( \mathbb{R}^k \) and \( t \), \( t + h \in Q \),
   \[
   E|\xi_{n\theta}(t + h) - \xi_{n\theta}(t)|^m \leq H(Q)|h|^r,
   
   E|\xi_{n\theta}(t)|^m \leq H(Q),
   
   E|\xi_\theta(t + h) - \xi_\theta(t)|^m \leq H(Q)|h|^r,
   
   E|\xi_\theta(t)|^m \leq H(Q),
   
   \text{where the constant } H(Q) \text{ depends only on the cube } Q.
   
2. There exists a function \( \gamma(y) \downarrow 0 \) such that
   \[
   \lim_{y \to \infty} \sup_{n, \theta} P \left\{ \sup_{|t| > y} \left| \xi_{n\theta}(t) \right| > \gamma(y) \right\} = 0,
   
   \lim_{y \to \infty} \sup_{\theta} P \left\{ \sup_{|t| > y} \left| \xi_\theta(t) \right| > \gamma(y) \right\} = 0.
   
3. The marginal distributions of functions \( \xi_{n\theta}(t) \) converge to the corresponding marginal distributions of \( \xi_\theta(t) \) uniformly in \( \Theta \).

Then the distributions of \( \xi_{n\theta} \) in \( C_0(\mathbb{R}^k) \) converge to the distribution of \( \xi_\theta \) in \( C_0(\mathbb{R}^k) \) uniformly in \( \Theta \).

The proof of this theorem is identical to the proof of the preceding one.

5 A Limit Theorem for Integrals of Random Functions

Theorem 22. Let \( Z_{n\theta}(t), n = 1, 2, \ldots, \) and \( Z_\theta(t) \) be measurable random functions which depend on parameter \( \theta \in \Theta \) and are defined on a closed bounded set \( F \subset \mathbb{R}^k \). Assume that the following conditions are satisfied:

1. \[
   \sup_{n, \theta} E \int_F |w(t)||Z_{n\theta}(t)| \, dt < \infty.
   
2. There exists a number \( \alpha > 0 \) such that
   \[
   \sup_{n, \theta} E|Z_{n\theta}(t) - Z_{n\theta}(s)| \leq H|t - s|^{\alpha},
   
   \text{where } H > 0 \text{ is a constant.}
   
3. The marginal distributions of \( Z_{n\theta} \) converge to the marginal distributions of \( Z_\theta \) uniformly in \( \Theta \).
Then the distributions of the integrals \( \int_F w(t)Z_{n\theta}(t) \, dt \) converge to the distribution of the integral \( \int_F w(t)Z_{\theta}(t) \, dt \) uniformly in \( \Theta \).

**Proof.** In view of Theorem 3, conditions (1) and (2) are fulfilled also for \( Z_{\theta} \). Subdivide the space \( R^k \) into cubes with edges of length \( \delta \) parallel to the coordinate axes. Let \( \Delta_j \) be the intersection of the \( j \)-th cube with the set \( F \). Choose in each set \( \Delta_j \) a point \( t_j \). We have

\[
\left| \exp \left\{ i\lambda \int_F w(t)Z_{n\theta}(t) \, dt \right\} - \exp \left\{ i\lambda \int_F w(t)Z_{\theta}(t) \, dt \right\} \right| \\
\leq \left| \exp \left\{ i\lambda \sum_j Z_{n\theta}(t_j) \int_{\Delta_j} w(t) \, dt \right\} - \exp \left\{ i\lambda \sum_j Z_{\theta}(t_j) \int_{\Delta_j} w(t) \, dt \right\} \right| \\
+ \left| \exp \left\{ i\lambda \sum_j Z_{n\theta}(t_j) \int_{\Delta_j} w(t) \, dt \right\} - \exp \left\{ i\lambda \int_F w(t)Z_{n\theta}(t) \, dt \right\} \right| \\
+ \left| \exp \left\{ i\lambda \sum_j Z_{\theta}(t_j) \int_{\Delta_j} w(t) \, dt \right\} - \exp \left\{ i\lambda \int_F w(t)Z_{\theta}(t) \, dt \right\} \right|.
\]

The first summand on the right-hand side, in view of condition (3), approaches zero uniformly in \( \theta \). The second and third summands are estimated in the same manner. For example, the second summand does not exceed

\[
|\lambda| \sum_j \int_{\Delta_j} |w(t)| \mathbb{E} |Z_{n\theta}(t) - Z_{n\theta}(t_j)| \, dt \leq H \sup_t |w(t)||\lambda| \text{mes } F \delta^2 \longrightarrow 0. \quad \square
\]

**Remark.** The conditions of the theorem will be fulfilled if \( Z_{n\theta}(t) \) is a non-negative function with

\[
\sup_{n,t,\theta} \mathbb{E} Z_{n\theta}(t) < \infty
\]

and

\[
\mathbb{E} |Z_{n\theta}^{1/2}(t) - Z_{n\theta}^{1/2}(s)|^2 \leq H|t - s|^z.
\]

Indeed, in this case

\[
\mathbb{E} |Z_{n\theta}(t) - Z_{n\theta}(s)| \leq \sqrt{2} \mathbb{E}^{1/2} |Z_{n\theta}^{1/2}(t) - Z_{n\theta}^{1/2}(s)|^2 \sup_{t,n,\theta} \mathbb{E}^{1/2} Z_{n\theta}(t).
\]

More details on the convergence of the functional of the integral type are given in [34], Chapter IX, Section 7.
APPENDIX II

Stochastic Integrals and Absolute Continuity of Measures

1 Stochastic Integrals over $b(t)$

In Chapters I, II, III, and VII an observational process of an unknown signal $S(t)$ (or $S(t, \theta)$) with Gaussian white noise was considered. By this was meant a process of observations $X(t)$, $0 \leq t \leq T$, which is defined by the formula

$$X(t) = \int_0^t S(u) \, du + b(t), \quad 0 \leq t \leq T,$$

(1)

where $b(t)$ is a Wiener process, i.e., a continuous Gaussian process with independent increments such that $E b(t) = 0$, $E b^2(t) = t$. Let $\mathcal{F}_t$ be a monotone ($\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2}$ for $t_1 < t_2$) family of $\sigma$-algebras such that the random variable $b(t)$ is $\mathcal{F}_t$-measurable and the process $b(t + h) - b(t)$, $h > 0$, does not depend on $\mathcal{F}_t$.

In order to study likelihood ratios connected with the process of observations (1) it is necessary to consider stochastic integrals over the process $b(t)$.

Let $f(t)$ be an arbitrary (nonrandom) function such that

$$\int_0^T f^2(t) \, dt < \infty.$$

In this case we write $f \in L_2(0, T)$. For step functions $f \in L_2(0, T)$, continuous from the right, i.e., functions taking on value $f(t_k)$ on the interval $t_k \leq t < t_{k+1}$, $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$, the stochastic integral

$$I(f) = \int_0^T f(t) \, db(t)$$
is defined by the expression
\[ I(f) = \sum_{k=0}^{n-1} f(t_k)(b(t_{k+1}) - b(t_k)). \] (2)

Clearly for the definite integral defined in this manner on step functions the properties
\[ I(\alpha f_1 + \beta f_2) = \alpha I(f_1) + \beta I(f_2), \] (3)
\[ I(\chi_{(r_1, r_2)}(t)) = b(r_1) - b(r_2), \] (4)
\[ EI(f) = 0, \] (5)
\[ EI^2(f) = \int_0^T f^2(t) \, dt \] (6)

are valid. Equation (2) establishes a correspondence between step functions \( f(t) \in L_2(0, T) \) and random variables \( \xi \). It is evident from (5) that this relation is an isometry if \( \| f \| \) is the usual \( L_2(0, T) \)-norm and \( \| \xi \| = (E\xi^2)^{1/2} \).

Therefore the set of \( f \) for which the stochastic integral \( I(f) \) is defined can be completed to form \( L_2(0, T) \) while relations (3), (4), and (5) will be preserved. Moreover, the stochastic integral is defined up to its values on sets of probability 0. By definition, it is assumed that
\[ \int_0^T f(s) \, db(s) = \int_0^T f(s)\chi(s) \, db(s). \]

Furthermore, one can choose a variant of definition such that the latter stochastic integral be a.s. continuously dependent on the upper limit \( t \). (see [34], p. 497)

From the construction of \( I(f) \) the equality
\[ E(I(f_1)I(f_2)) = \int_0^T f_1(t)f_2(t) \, dt = (f_1, f_2) \] (6)
and the relation
\[ I(f_n) \to I(f) \text{ a.s. as } \| f_n - f \| \to 0 \] (7)
follow. It is also clear that for any \( f \in L_2(0, T) \) the random variable \( I(f) \) is Gaussian, \( \mathcal{L}(I(f)) = \mathcal{N}(0, \| f \|^2) \), so that, for example,
\[ E \exp I(f) = \exp \left( \frac{\| f \|^2}{2} \right). \] (8)

Let \( \varphi_1, \varphi_2, \ldots, \varphi_n, \ldots \) be an orthonormal system of functions in \( L_2(0, T) \) and let
\[ f(t) = \sum_{i=1}^\infty c_i \varphi_i(t). \]
Then it follows from (3) and (6) that

\[ I(f) = \sum_{i=1}^{\infty} c_i I(\varphi_i). \]  

(9)

In view of (7), \( E(I(\varphi_i)I(\varphi_j)) = \delta_{ij} \) and therefore \( I(\varphi_i) \) are independent Gaussian random variables with parameters \((0, 1)\).

The stochastic integral

\[ \int_{0}^{T} f(t) \, dX(t), \]

where \( X(t) \) is given by the equality

\[ \int_{0}^{T} f(t) \, dX(t) = \int_{0}^{T} f(t) S(t) \, dt + \int_{0}^{T} f(t) \, db(t). \]

2 Some Definitions and Theorems of Measure Theory

We shall briefly recall several definitions and theorems of measure theory (see [126]). Let \( (\mathcal{X}, \mathcal{U}) \) be a measurable space and \( \mu_1 \) and \( \mu_2 \) be two \( \delta \)-finite measures on \( \mathcal{U} \). We say that \( \mu_2 \) is absolutely continuous with respect to \( \mu_1 \) if \( \mu_2(A) = 0 \) for any set \( A \in \mathcal{U} \) such that \( \mu_1(A) = 0 \). Mutually absolutely continuous measures are called equivalent. The measures \( \mu_1 \) and \( \mu_2 \) are singular if for some set \( A \), \( \mu_1(A) = 0 \) and \( \mu_2(\mathcal{X} \setminus A) = 0 \). For any measures \( \mu_1 \) and \( \mu_2 \) the representation \( \mu_2 = \nu_1 + \nu_2 \) is valid where \( \nu_1 \) is absolutely continuous and \( \nu_2 \) is singular with respect to \( \mu_1 \). The measures \( \nu_1 \) and \( \nu_2 \) are called respectively the absolutely continuous and the singular components of measure \( \mu_2 \) with respect to \( \mu_1 \). The Radon–Nikodym theorem asserts that \( \mu_2 \) is absolutely continuous with respect to \( \mu_1 \) if and only if for any \( A \in \mathcal{U} \),

\[ \mu_2(A) = \int_{A} \rho(x) \, d\mu_1(x) \]

for some \( \mathcal{U} \)-measurable nonnegative function \( \rho(x) \). The function \( \rho(x) \) is called the density of the measure \( \mu_2 \) with respect to \( \mu_1 \) and is denoted by \( d\mu_2/d\mu_1 \).

If \( \mu_2 \) is absolutely continuous with respect to \( \mu_1 \) and the function \( f \) is integrable relative to the measure \( \mu_2 \) then the formula

\[ \int f \, d\mu_2 = \int f \frac{d\mu_2}{d\mu_1} \, d\mu_1 \]

(10)

is valid.
Let $P_i$ be a probability measure on $(C(0, T), \mathcal{F})$ associated with the process
\[ X_i(t) = \int_0^t S_i(s) \, ds + b(t), \quad i = 1, 2, \quad 0 \leq t \leq T, \tag{11} \]
and let $E_i G(X) = E G(X_i)$. We shall derive a formula for the density of measure $P_2$ with respect to $P_1$ (see [93]).

**Theorem 1.** If $S_i \in L^2(0, T)$ then the measures $P_1$ and $P_2$ are equivalent and, moreover,
\[ \frac{dP_2}{dP_1}(X_1) = \left\{ \int_0^T (S_2(t) - S_1(t)) \, dX_1(t) - \frac{1}{2}(\|S_2\|^2 - \|S_1\|^2) \right\}. \]

**Proof.** Let
\[ J_j(X_i) = \int_0^T \phi_j(t) \, dX_i(t), \quad j = 1, \ldots, m, \]
where as above $\phi_j(t)$ is a complete orthonormal system in $L^2(0, T)$. Since
\[ S_i(t) = \sum_{j=1}^{\infty} c_{ij} \phi_j(t), \quad c_{ij} = (S_i, \phi_j), \]
it follows that
\[ J_j(X_i) = \int_0^T \phi_j(t) \, dX_i(t) = c_{ji} + J_i(\phi_j). \tag{12} \]

Consequently, for any arbitrary bounded measurable function $G(x_1, \ldots, x_m)$,
\[
E_2 G(J_1(X), \ldots, J_m(X)) = \int_{R^m} \cdots \int G(x_1, \ldots, x_m) \exp \left\{ -\frac{1}{2} \sum_{j=1}^{m} (x_j - c_{j2})^2 \right\} \, dx_1 \cdots dx_m \\
= \int_{R^m} \cdots \int G(x_1, \ldots, x_m) \exp \left\{ \sum_{j=1}^{m} (c_{j2} - c_{j1})x_j + \frac{1}{2} \sum_{j=1}^{m} (c_{j1}^2 - c_{j2}^2) \right\} \exp \left\{ -\frac{1}{2} \sum_{j=1}^{m} (x_j - c_{j1})^2 \right\} \, dx_1 \cdots dx_m \\
= E_1 \left\{ G(J_1(X), \ldots, J_m(X)) \exp \left\{ \sum_{j=1}^{m} (c_{j2} - c_{j1})J_j(X) + \frac{1}{2} \sum_{j=1}^{m} (c_{j1}^2 - c_{j2}^2) \right\} \right\}. \tag{13} \]
Furthermore, the random variables $J_x(X_2)$ are jointly independent and, in view of (12) and (8),

\[
E_1 \exp \{ (c_{j_2} - c_{j_1}) J_x(X) + \frac{1}{2} (c_{j_1}^2 - c_{j_2}^2) \} = E \exp \{ (c_{j_2} - c_{j_1}) (c_{j_1} + I(\phi_j)) + \frac{1}{2} (c_{j_1}^2 - c_{j_2}^2) \} = E \exp \{ (c_{j_2} - c_{j_1}) I(\phi_j) - \frac{1}{2} (c_{j_2} - c_{j_1})^2 \} = 1.
\]

Therefore equality (13) can also be written in the form

\[
E_2 G(J_1(X), \ldots, J_m(X)) = E_1 \left\{ G(J_1(X), \ldots, J_m(X)) \times \exp \left[ \sum_{j=1}^{\infty} (c_{j_2} - c_{j_1}) J_x(X) + \frac{1}{2} \sum_{j=1}^{\infty} (c_{j_1}^2 - c_{j_2}^2) \right] \right\}.
\]

Taking into account (9) and Parseval's equality $\sum_j c_{ji}^2 = \|S_i\|^2$, we finally obtain

\[
E_2 G(J_1(X), \ldots, J_m(X)) = E_1 \left\{ \exp \left[ \int_0^T (S_2(t) - S_1(t)) dX(t) \right] - \frac{1}{2} (\|S_2\|^2 - \|S_1\|^2) \right\} G(J_1(X), \ldots, J_m(X)) \right\}. \quad (14)
\]

Observe now that relation (14) is valid for any orthonormal system of functions $\phi_1(t), \ldots, \phi_n(t), \ldots$. In particular we can consider a sequence of Haar functions

\[
\psi_{2^n, k}(t) = \begin{cases} 2^{(n-1)/2} T^{-1/2}, & \text{if } T(k - 1)2^{-n} \leq t \leq k2^{-n}T, \\
-2^{(n-1)/2} T^{-1/2}, & \text{if } Tk2^{-n} \leq t \leq T(k + 1)2^{-n}, \\
0, & \text{for other } t \in [0, T], \\
1 \leq k \leq 2^{n-1}, k \text{ is odd. }
\end{cases}
\]

Clearly, for such a choice

\[
G(J_1(X), \ldots, J_{2^n}(X)) = \tilde{G}(X(kT2^{-n})), \quad k = 1, \ldots, 2^n. \quad (15)
\]

Substituting (15) into (14) and setting $\tilde{G}$ equal to the indicator of a set $A$ determined by the values of $X(t)$ at the points $kT2^{-n}, k = 1, \ldots, 2^n$, we shall obtain the formula

\[
P_2(A) = \int_A \exp \left\{ \int_0^T (S_2(t) - S_1(t)) dX(t) - \frac{1}{2} (\|S_2\|^2 - \|S_1\|^2) \right\} dP_1(X).
\]

Since such sets $A$ generate a $\sigma$-algebra of Borel sets in $C(0, T)$ (see, for example, [10], p. 33) the theorem is proved.
From this theorem and (10) the following formula
\[ E_2 f(X) = E_1 \left[ \exp \left\{ \int_0^T (S_2(t) - S_1(t)) \, dX(t) - \frac{1}{2}(\|S_2\|^2 - \|S_1\|^2) \right\} \right] f(X) \] (16)
which is often used in this book results.

From this formula we also obtain for the density \( dP_2/dP_1 \) and a form for the process \( X_1 \) the relationship
\[ \ln \frac{dP_2}{dP_1} (X_1) = \int_0^T (S_2 - S_1) \, dX_1(t) - \frac{1}{2}\|S_2\|^2 + \frac{1}{2}\|S_1\|^2 \]
\[ = \int_0^T (S_2 - S_1) \, db(t) - \frac{1}{2}\|S_2 - S_1\|^2. \] (17)

From (17) and the formula
\[ \frac{dP_2}{dP_1} (\cdot) = \left( \frac{dP_1}{dP_2} (\cdot) \right)^{-1} \]
valid for equivalent measures \( P_1 \) and \( P_2 \) we have
\[ \ln \frac{dP_1}{dP_2} (X_1) = \frac{1}{2}\|S_2 - S_1\|^2 + \int_0^T (S_1 - S_2) \, db(t). \] (18)

3 Stochastic Integrals over Orthogonal Random Measure

By means of the process \( b(t) \) one can define a random measure on the half-line \( t \geq 0 \) by setting
\[ \mu([t_1, t_2)) = b(t_2) - b(t_1) \]
and extending it over the \( \sigma \)-algebra \( \mathfrak{B} \) of Borel sets on this half-line. The measure thus defined is orthogonal, i.e.,
\[ E[\mu(\Delta_1)\mu(\Delta_2)] = 0, \text{ if } \Delta_1 \cap \Delta_2 = \emptyset. \]
Moreover, \( E[\mu(\Delta_1)]^2 = \text{mes} \Delta_1 \), where \( \text{mes} \Delta_1 \) is the Lebesgue measure.

The stochastic integral constructed in Section 1 can be viewed as a stochastic integral over an orthogonal random measure \( \mu \).

In Chapter VI integrals over more general orthogonal random measures are used. Let \( \nu(\Delta), \Delta \in \mathfrak{B} \), be a family of random measures with the properties
\[ \nu(\Delta_1 \cup \Delta_2) = \nu(\Delta_1) + \nu(\Delta_2) \quad \Delta_1 \cap \Delta_2 = \emptyset, \]
\[ E(\nu(\Delta_1)\nu(\Delta_2)) = m(\Delta_1 \cap \Delta_2), \]
where \( m(\cdot) \) is a countably-additive set function in \( \mathcal{B} \). The family \( \nu(\Delta) \) is called an orthogonal random measure and \( m \) is its structure function. A structure function defines a (nonrandom) measure on \( \mathcal{B} \).

An integral over an orthogonal random measure is constructed exactly in the same manner as an integral over the Wiener measure in Section 1. First we define the integral for step functions—which are constant for \( t \in [t_i, t_{i+1}) = \Delta_i \)—by means of the formula

\[
I(f) = \int f(t) \, d\nu(t) = \sum f(t_i) \nu(\Delta_i).
\]

For this integral we then establish properties (3), (4), and the property

\[
\mathbb{E}I^2(f) = \int f^2(t) m(dt).
\]  \hfill (19)

By means of the last equality one can extend the set of functions for the stochastic integral is defined for all \( f \in L^2(m) \). Moreover, properties (3), (4), and (19) remain valid. Furthermore, the equality

\[
\mathbb{E}(I(f)I(g)) = \int f(t)g(t)m(dt)
\]

is also satisfied.
Remarks

The bibliographical comments presented below intend in no way to sketch the complete history of the problems discussed in this book. The authors' aim was rather to note the sources from which the results were taken and to aid the reader's orientation as far as the literature is concerned. In many instances, references to the original papers are missing and are replaced by references to monographs, textbooks, and survey articles.

The absence of comments on a certain theorem does not indicate that the theorem is due to the authors of the book but simply means that we have no knowledge as to the identity of the original author.

Chapter 1

Sections 1 and 2. The statement of the problem of statistical estimation presented here is the one commonly accepted. Its modern statement is due to A. Wald (see, for example [18], [21]). The history of the problem is well treated in [83], [90], and [91]. Basic notions of statistical decision theory are presented in [91]. The reader is also advised to consult A. N. Kolmogorov's works [70], [71]. An interesting geometric treatment of the theory of statistical decisions is contained in N. N. Čencov's monograph [136].

See [91], [92] in connection with the theory of sufficient statistics. A study of problems connected with a description of functions admitting unbiased estimators is contained in paper [71], [9] and others. Estimation of scale parameters is discussed in [64].
Section 3. Results presented in Subsection 1 are adapted from Hodges and Lehmann's paper [128]; Example 2 is also taken from this paper. The inadmissibility of $\overline{X}$ in the case of estimation of the mean of a multivariate normal population was established by Ch. Stein [114]; we have followed his paper [116] here in presenting Example 1. The conditions for admissibility of various estimators of the mean of a normal population were studied in detail by Brown [15]. See also lectures given by Ch. Stein at the Leningrad Branch of the Mathematical Institute of the Academy of Sciences of the U.S.S.R [117].

Section 4. The notion of consistency of statistical estimators can be traced to R. A. Fisher's works, see [124]. The method of moments was originated with K. Pearson, see [74] in this connection. Maximum likelihood estimators have appeared in D. Bernoulli's works (see [68]) but received wide popularity as a result of Fisher's contributions (see, for example [124]). A very general theorem dealing with the consistency of maximum likelihood estimators was proved by A. Wald [20], the theorem of consistency of Bayesian estimators is due to L. LeCam [83]; see also [140] concerning Bayesian estimators.

Section 5. The basic results presented in this section are due to the authors. The theorems of subsection 2 (in a somewhat different formulation) were published in [47]; Theorems 5.1 and 5.2 are new. For further development of results of subsection 2, see [5].

Section 6. A part of the authors' paper [52] is presented; see also [89].

Section 7. The essence of the notions and results of subsection 1 is due to R. A. Fisher and is presented in some form in many books. The role of differentiability in the mean square when distinguishing the class of regular experiments was first noticed by L. LeCam [87] and J. Hájek [29].

First versions of the Cramér–Rao inequality (also called the information inequality) were published by H. Cramér [73] C. R. Rao [105] and M. Fréchet [125]. Later this inequality was studied by many authors, see, for example [12], [13], [139] and [135]. Our version of the proof follows C. Blyth and D. Roberts's paper [11]; Bhattacharya's inequality appeared in paper [17]; the Chapman–Robbins inequality in [137]; Barankin–Kiefer's in [6] and [69]; see also A. M. Kagan's paper [63]. The integral variant of Cramér–Rao's inequality was proved by N. N. Čencov [135], see also [48]. Hodges and Lehmann [129] were first to notice the feasibility of proving the admissibility of estimators using the Cramér–Rao inequality; see also [11] and [12]; Theorem 1.7.5. is due to the authors [48].

Section 8. In this Section the authors' results [49] are given.

Section 9. We follow Wolfowitz's paper [26] in presenting Fisher's program; this paper presents a very clear exposition of the problem of definition and study of efficient estimators; see also Weiss and Wolfowitz's book
Chapter II

The idea of approximating the logarithm of likelihood ratios by a Gaussian family was expressed by A. Wald [19] and developed in papers of LeCam [83], [86], [88], [89] who, in particular, introduced the term "local asymptotic normality." In monographic literature the term was actively used for the first time in Roussas' book [112].

Section 1. Theorems 1.1 and 1.2 are due to LeCam [88] see also [112].

Section 2. Theorem 2.1 and the proof presented here is due to Hájek [31].

Section 3. The condition of local asymptotic normality for independent nonhomogeneous observations was considered in LeCam's [86], Kushnir's [78], Phillipou and Roussas' [112] and other works. Theorem 3.1 is due to the authors [53]. Lemma 3.2 which assures absolute continuity of a measure $P_{\theta}$ with respect to $P_{\theta}$ on the infinite product $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots$ can be deduced from the well-known theorem of Kakutani [65] if one notes that its condition is sufficient for the convergence of the infinite product

$$\prod_{j=1}^{\infty} \int_{\mathcal{X}_j} [f_j(x, \theta)f_j(x, \theta)]^{1/2} dv_j.$$ 

Section 5. Theorem 5.1 seems to be new, although closely related results are given by the authors' paper [46].

Section 6. Theorem 6.1 was proved in [53].

Section 7. The estimation problem of the observational process (7.1) was first studied by Kotel'nikov in [72]. Theorem 7.1 is due to the authors [53].

Section 8. The results presented here are basically due to LeCam [86]. We follow Hájek's paper [30] in the proof of Theorem 8.1.

Section 9. Theorem 9.1 is due to Hájek [30]. According to Roussas [112] the proof presented here is due to Bickel.

Section 10. Basically we state here the results of Anderson [1].
Section 11. The interesting book of Wolfowitz and Weiss [23] is devoted to efficiency in Wolfowitz's sense. In this book some further results are given.

Section 12. Theorem 12.1 is due to Hájek [31] for the one dimensional case. A multidimensional variant was suggested in [53].

Section 13. See [83], [135] in connection with superefficient estimators. Theorem 13.3 seems to be new.

Chapter III

Sections 1 and 2. The results in these Sections are due to the authors. These are generalizations of the method of studying properties of estimators suggested by us in earlier papers [42] and [44]. Theorems 1.2 and 2.2 allow us to conclude that \( \varphi^{-1}(c, t)(\hat{T}_e - \hat{\theta}_e) \rightarrow 0 \) in \( P_e \)-probability as \( e \rightarrow 0 \). For a particular case the last result was established earlier in paper [113] where additional inferences are given. Theorems 1.1 and 2.1 may be used also for a study of limiting properties of estimators for dependent observations, see, for example, the papers by Bakirov [4], Efroimovits [40], Kutoyants [75], [76], [77] and others.

Section 3. Many papers are devoted to the study of properties of estimators for independent identically distributed observations. An excellent survey of these investigations up to 1953 is given in LeCam's paper [83]. See also his papers [84]–[89]. Huber [132], Inagaki and Ogata [62] studied the behavior of estimators in a multidimensional parameter set. The properties of Bayesian estimators were studied by Chao in [133] where additional references are given. But it was LeCam who, in paper [88] pointed out that in the case of \( k > 1 \), the existence of information amount is insufficient for the proof of asymptotic normality of the maximum likelihood estimators. Theorems 3.1 and 3.2 are due to the authors. For a one-dimensional parameter set these theorems were proved by us previously in [42], [44] under additional restrictions. See [101], p. 79, in connection with Lemma 3.1. Theorem 3.3 and a somewhat more general result was obtained in our papers [42] and [46]. Another proof for a particular case was given by Woodroofe [27].

Section 4. For nonhomogeneous observations properties of maximum likelihood estimators were studied quite intensively, cf. for example, Hoadley [130] where additional bibliography is presented. Theorem 4.1 is new; in a one-dimensional case it was previously proved in the authors' paper [55]. An example of estimation of a frequency of a periodic function was studied previously in Walker [25], Pisarenko [102] and others in connection with the important problem of detection of latent periodicities.
Section 5. The asymptotic normality of maximum likelihood estimators based on observations of a signal in a Gaussian white noise was shown by Kotel'nikov [72] with "physical" (rather than mathematical) rigor. Lemma 5.1 is a simple corollary to the well-known Neyman–Pearson lemma. Lemma 5.2 is new. Theorem 5.1 is due to the authors; for $k = 1$ an analogous result was published in the authors' paper [50]. Examples 2 and 3 are also borrowed from this paper. For development of the results presented in this Section see Burnashev [16], Kutoyants [75], [76] and [77].

Chapter IV

Sections 1–3. Definitions and results of these Sections are basically due to Levit, see [80], [81]. Example 5 in Section 2 was considered by Nevel'son [98]. Example 6 in Section 2 from a similar angle was analyzed in Stein's paper [115]. Asymptotically efficient (in a weaker than in our sense) estimators in this example were constructed in Beran's paper [8] where additional references may be found.

Section 4. Estimators of the form (4.1) were introduced by Parzen [100] and Rosenblatt [109]; an excellent survey of papers up to 1970 is given in Rosenblatt [110]. Theorem 4.1 is analogous to the result presented in Part 2 of Rosenblatt's survey, see also [96a].

Section 5. The statement of the problem in Section 5 is due to Čencov who solved it for the integral measure of deviation of $f_\alpha$ from $f$ given by

$$\int [f_\alpha(x) - f(x)]^2 \, dx.$$

Theorem 5.1 is due to Farrel [121]. See also Meyer's paper [95] and others.

Chapter V

It has been known for a long time (cf. for example [74], p. 527) that for samples from a distribution whose density with respect to Lebesgue's measure viewed as a function of the parameter has jumps one can in certain cases (such as the case of estimation of the location parameter for samples from a uniform or exponential distribution) construct unbiased estimators with variance of order $n^{-2}$. The first general result was probably considered in Chernoff and Rubin's paper [138]. This problem was later discussed in Wald's lecture delivered by W. Hoeffding in 1967 (the lecture was not published). The basic result of Chapter V was published by the authors in 1970 in the note [42]; this chapter is actually a modified version of the authors' paper [45].
H. Rubin in paper [111] considered the problem of estimating a multi-dimensional parameter in the case of a density with jumps; the same problem in a more general setting was studied by M. S. Ermakov [39].

**Chapter VI**

Prakasa Rao's work [104] is one of the first papers devoted to the study of the behavior of maximum likelihood estimators for the case of densities with singularities (of the second type); in particular, in this paper the expression (2.2) for the logarithm of the limiting process is presented. The late L. N. Bol'shev was very much interested in non-regular estimation problems; in particular the examples considered in Section 6 are due to him (in his lectures at the Moscow State University). The basic results of Chapter VI were published by the authors in 1970 in note [42]; this chapter is an improved version of the authors' papers [51], [57]. Here the representation of the limiting process as a stochastic integral (2.1) plays a substantial role; this representation was discovered by M. S. Ermakov [38]. A number of results on estimation theory in non-regular problems are given in [28] and [103].

**Chapter VII**

*Section 1.* Theorems 1.1 and 1.2 were derived by the authors in [50]. Lemma 1.1 is due to Fano, see, for example, [32], while lemma 1.2 is due to Shannon [141].

*Section 2.* The basic results for the scheme of observations (2.1) were obtained in the authors' paper [54]; the feasibility of extending these results to the case of a scheme of observations (2.1') was proved by Golub'ev in [37].

*Section 3.* Theorem 3.1 is due to Terent'yev [118] while Theorem 3.2 is due to the authors.

*Section 4.* Theorem 4.1 generalizes somewhat Theorem 4 presented in the authors' paper [56] while Theorem 4.2 is in Example 1 of paper [58]. Theorem 4.3 is presented also in [58]. More general results expressing the upper bounds on the precision of nonparametric estimation in Σ in terms of Kolmogorov's diameters (see [120]) of the set Σ are also given in [58].

*Section 5.* Theorem 5.1 was presented in the authors' paper [56], Theorems 5.2 and 5.3 in paper [58], while Theorem 5.4 and the example following the theorem in [59]. The last paper also contains more general results.
Bibliography*


*Translator's remark: The bibliography presented below is out of alphabetical order due to differences between the English and Russian alphabets. However, all the references are consecutively numbered and referred to by those numbers in the text. Books and journal articles by Russian authors that have been translated into English are cited as they appear in the English translation.*
Bibliography


[99] —, On asymptotic optimality of recurrent estimators, *ibid.* 14, 1 (1978), 50–76.
Index

Absolute continuous component of a measure 384
Absolutely continuous measure 384
Anderson's lemma 155
Anomaly, absence of 205
Arzela's theorem 371
$A$, condition 118
Asymptotic efficiency
in Bahadur's sense 95
in Fisher's sense 90, 159
in Rao's sense 102, 161
in Wolfowitz's sense 159, 160
with respect to a family of loss functions 6, 92
Asymptotically efficient estimator 90, 92, 162
Fisher's 90, 159
in Bahadur's sense 95
in Rao's sense 102, 161
in Wolfowitz's sense 159, 160
nonparametric 219
Average probability of error 323
Barankin-Kiefer inequality 75
Bayes(ian) risk 22
Bhattacharya's inequality 75
Blackwell's theorem 19

Capacity of a channel 328, 329
Chapman-Robbins' inequality 76
Class of Borel sets 369
Convergence of distributions 363
uniform 365
weak 363, 369, 370
Cramér-Rao inequality 73
Curve of a signal 144

Density of a measure 384
Distribution 10
Dirac function 233
Dominating measure 11

Efficient standard deviation of an estimator 97
Entropy 323
average conditional 323
Equivalent measures 384
Estimation problem in Bayesian formulation 22
parametric 18
two sample 227
Estimator
admissible 17
Bayesian 23, 24
consistent 30, 31
Estimator [cont.]
  efficient 78, 90
  equivariant 20
  \((F, R, w)\)–asymptotically efficient 219
  inadmissible 17
  maximum likelihood 35, 273
  minimax 17
  Pitman’s 21
  projection 349
  regular 151, 160
  Stein 27, 172
  Stein–James 27
  superefficient 91, 170
  unbiased 17
  \(w_e\)–asymptotically efficient 92

Experiment generated by observation 11

Factorization theorem 12
Fano’s lemma 323
Family of distributions relatively compact 370
dense 370
Family of estimators consistent 31
  locally asymptotically normal 120
  uniformly consistent 31
  \(w_e\)–asymptotically efficient 92
Family of estimators asymptotically efficient
  in Fisher’s sense 90, 159
  in Rao’s sense 102, 161
  in Wolfowitz’ sense 159
Fisher’s amount of information 63, 64
Fisher’s information matrix 63, 64
Functional differentiable in von Mises’ sense 220
Frequency modulation 206, 209

Hájek’s theorem 162
Hellinger’s distance 31, 81, 219, 243

Information amount in estimating a functional 217

Jackson’s inequality 353
Jensen’s inequality 19, 324

Law of large numbers 367
  strong (Kolmogorov’s) 368
Lindeberg’s condition 124, 128, 143, 369
Likelihood equation 35
Likelihood function 35
Likelihood ratio 120
Local asymptotic normality 7, 120
Loss function 16, 18
Lyapunov’s condition 128, 369

Marginal distributions 370
Maximum likelihood method 35
Method of moments 33
Modulator 206

Observation 10
Orthogonal random measure 388

Parametric estimation problems 18
Part of an experiment 12
Parzen–Rosenblatt estimator 233
\(\mathcal{P}\)–continuous set 364
Phase modulation 206, 208
Pitman’s estimator 21
Point of superefficiency 91
Posterior mean 23
Product of experiments 12
Projection of an experiment 11

Radon–Nikodym theorem 11, 384
Rate of transmission 328
Regular experiment 65, 114
Risk function 16

Sample space 11
Sequence of estimators, consistent 30
Sequence of statistical experiments, locally asymptotically exponential 276
Shannon’s information amount 324
Singular component of a measure 384
Singularity for a density of the first type 281
Singularity for a density of the second type 282
Index

Singularity for a density of the third type 283
Singularity of order $\alpha = 0$ 283
Statistic 12
asymptotically efficient in Rao's sense 102
sufficient 12, 14
Statistical experiment 11
generated by an observation 11
regular 65, 114
with a finite Fisher information 63, 199
Stein–James' estimator 27
Stein's estimator 27, 172
Stochastic integral over a Wiener process 382, 383
Stochastic integral over an orthogonal random measure 387
Structure function 388
Theorem on relative stability 367, 368
Time–pulse modulation 206, 207
Topology coordinated with estimation problem 218
Truncation of a vector 148
Uniform asymptotic normality 123, 143
Uniformly integrable random variables 364