Chapter 1

ASYMPTOTIC EFFICIENCY OF STATISTICAL TESTS AND MATHEMATICAL MEANS FOR ITS COMPUTATION

• 1.1 General Approach to Computation of Asymptotic Efficiency

Let $(\mathfrak{X}, \mathfrak{A})$ be a sample space corresponding to the observation X. It is assumed that the distribution $\mathbf{P}_\theta$ of this observation is determined by parameter θ taking on values in a parametric set Θ. Let $s = \{X_1, X_2, \dots\}$ be a sequence of independent identically distributed random variables with values in $\mathfrak{X}$ and having the distribution $\mathbf{P}_\theta$, $\theta \in \Theta$. For any positive integer n put $\mathbf{X}^{(n)} := \big(X_1, X_2, \dots, X_n\big)$ and denote by $\big(\mathfrak{X}^{(n)}, \mathfrak{A}^{(n)}\big)$ the corresponding sample space and by $\mathbf{P}_\theta^{(n)}$ the distribution of $\mathbf{X}^{(n)}$ on $\mathfrak{A}^{(n)}$. In the sequel $\mathbf{P}_\theta^{(n)}$ will be usually abbreviated to $\mathbf{P}_\theta$.

Consider the problem of testing the hypothesis

$$H\colon\ \theta \in \Theta_0 \subset \Theta$$

against the alternative

$$A\colon\ \theta \in \Theta_1 = \Theta \setminus \Theta_0$$

on the basis of observations $X_1, X_2, \dots, X_n$. For this purpose we dispose of a sequence of statistics $\{T_n\}$, $T_n(s) := T_n(X_1, X_2, \dots, X_n)$, assuming (without essential loss of generality) large values of T_n to be significant. Thus the acceptance region of H is given by

$$\big\{\, s\colon\ T_n(s) \ge c \,\big\}$$

where c is some real number.

The *power function* of this test is the quantity $\mathbf{P}_\theta(T_n \geq c)$ considered as a function of θ and its size is equal to

$$\sup\left\{ \mathbf{P}_\theta\,(T_n \geq c):\ \theta \in \Theta_0 \right\}.$$

Now define for any $\beta \in (0,\,1)$ and $\theta \in \Theta_1$ a real sequence $c_n := c_n(\beta,\theta)$ with the aid of double inequality

$$\mathbf{P}_\theta\,(T_n > c_n) \leq \beta \leq \mathbf{P}_\theta\,(T_n \geq c_n)\,. \tag{1.1.1}$$

Then

$$\alpha_n(\beta,\theta) := \sup\left\{ \mathbf{P}_{\theta'}(T_n \geq c_n):\ \theta' \in \Theta_0 \right\}$$

is the minimal size of the test based on $\{T_n\}$ for which the power at the point θ is not less than β. Let us define for any given level of significance α, $0 < \alpha < \beta$, the positive integer

$$N_T(\alpha,\beta,\theta) := \min\left\{n:\ \alpha_m(\beta,\theta) \leq \alpha \quad \text{for all } m \geq n\right\}.$$

It is clear that $N_T(\alpha,\beta,\theta)$ is the minimal sample size necessary for the test at a level α, based on $\{T_n\}$, to have the power not less than β at the point θ.

Suppose that for testing H against A we dispose of two sequences of test statistics $\{T_n\}$ and $\{V_n\}$. Define by $\mathbf{e}_{V,T}(\alpha,\beta,\theta)$ the relative efficiency of the sequence $\{V_n\}$ with respect to $\{T_n\}$ in the following way:

$$\mathbf{e}_{V,T}(\alpha,\beta,\theta) := N_T(\alpha,\beta,\theta)\big/\,N_V(\alpha,\beta,\theta)\,. \tag{1.1.2}$$

The value $\mathbf{e}_{V,T}(\alpha,\beta,\theta)$ larger than 1 indicates that for given α, β and θ one should prefer the sequence $\{V_n\}$ to $\{T_n\}$ because the first sequence requires less observations for reaching the power β for the level α and the alternative value θ.

As has been already noted in the Introduction, relative efficiency (1.1.2) has indisputable merits. Unfortunately it has also two substantial drawbacks. One consists in that the value of $\mathbf{e}_{V,T}(\alpha,\beta,\theta)$ depends on three arguments (and two sequences of statistics), the other is connected with the fact that it is extremely difficult or simply impossible to calculate this value. It is possible at present to overcome these difficulties by calculating the limiting values $\mathbf{e}_{V,T}(\alpha,\beta,\theta)$ as $\alpha \to 0$, $\beta \to 1$ and as $\theta \to \theta_0 \in \partial\Theta_0$ (in a certain topology on Θ) keeping fixed the values of two remaining parameters. As a result one obtains three fundamental types of the asymptotic relative efficiency (ARE).

If for $\beta \in (0, 1)$ and $\theta \in \Theta_1$ there exists the limit

$$\mathbf{e}^{\,\mathrm{B}}_{V,T}(\beta,\theta) := \lim_{\alpha\downarrow 0} \mathbf{e}_{V,T}(\alpha,\beta,\theta)\,, \tag{1.1.3}$$

it is called the *Bahadur ARE of the sequence* $\{V_n\}$ *with respect to* $\{T_n\}$.

If for $\alpha \in (0, 1)$ and $\theta \in \Theta_1$ there exists the limit

$$\mathbf{e}^{\,\mathrm{HL}}_{V,T}(\alpha,\theta) := \lim_{\beta\uparrow 1} \mathbf{e}_{V,T}(\alpha,\beta,\theta)\,, \tag{1.1.4}$$

it is called the *Hodges–Lehmann ARE of the sequence* $\{V_n\}$ *with respect to* $\{T_n\}$.

If for $0 < \alpha < \beta < 1$ and $\theta \to \theta_0 \in \partial\Theta_0$ (in a certain topology on Θ) there exists the limit

$$\mathbf{e}^{\,\mathrm{P}}_{V,T}(\alpha,\beta,\theta_0) := \lim_{\theta\to\theta_0} \mathbf{e}_{V,T}(\alpha,\beta,\theta)\,, \tag{1.1.5}$$

it is called the *Pitman efficiency of the sequence* $\{V_n\}$ *with respect to* $\{T_n\}$.

It is also difficult to calculate these three types of the ARE, but still much easier than relative efficiency (1.1.2). Moreover, the Bahadur ARE usually does not depend on β, the Hodges–Lehmann ARE does not depend on α and the Pitman ARE in most cases does not depend neither on α nor on β and turns out to be a constant. We emphasize once again that from the practical point of view the very cases of small levels, high powers and close alternatives are the most important. That is why one may suppose that the knowledge of three types of the ARE such as (1.1.3)–(1.1.5) will in a sense help to put in order principal tests used in a concrete problem and will permit to give the well-founded recommendations for their applications in practice.

Note that there exist the intermediate approaches to measuring the ARE not coinciding with the approaches of Bahadur, Hodges and Lehmann, and Pitman. The typical examples are the Chernoff ARE (see Chernoff (1952), Kallenberg (1982) and Ronzhin (1985)) when for a fixed θ the other parameters α and β tend to 0, and the case of the intermediate or Kallenberg ARE (see Kallenberg (1983a)), when β is fixed, but θ and α tend to θ_0 or 0 at a controlled rate.

The different definitions of the ARE belong also to Rubin and Sethuraman (1965b), and to Borovkov and Mogulskii (1992). The first exploits the notion of the Bayes risk whereas the second deals with a certain modification of the number $N_T(\alpha, \beta, \theta)$ being more symmetric with respect to the errors of the first and second kinds. Their values for more or less complicated nonparametric statistics are unknown.

It would be most interesting to learn if the values of ARE's calculated under different approaches are close to the values of the relative efficiency for finite samples and reasonable values of α, β and θ arising in practical problems. The first experience of such comparison has been realized by Groeneboom and Oosterhoff (1981) with the aid of statistical modelling. The results of Groeneboom and Oosterhoff (1981) are connected with some simplest examples only and for the present do not give any reasons for definite conclusions.

It may happen that the value of the ARE is equal to 1 and this circumstance prevents the asymptotic comparison of tests. In that case one may recommend to follow Hodges and Lehmann (1970) and to use more sensitive means for comparing different tests, namely the *deficiency*

$$\operatorname{def}(V, T; \alpha, \beta, \theta) := N_V(\alpha, \beta, \theta) - N_T(\alpha, \beta, \theta), \qquad \theta \in \Theta_1 .$$

Asymptotic approximations to the deficiency as $\alpha \to 0$ and $\theta \to \theta_0 \in \partial\Theta_0$ have been considered, mainly for parametric statistics, by Albers (1974), Chandra and Ghosh (1978), Groeneboom and Oosterhoff (1981), Kallenberg (1981, 1982) as well as by Borovkov and Mogulskii (1992). Not much is known in the nonparametric case and we will not use the notion of the deficiency in the sequel.

• 1.2 Bahadur Asymptotic Relative Efficiency

The Bahadur approach to measuring the ARE is opposite to the classical approach of Neyman and Pearson and prescribes one to fix the power of tests and to compare the rate of decreasing their sizes for the increasing number of observations. This point of view, expressed for the first time by Cochran (1952), has been deeply and systematically developed by Bahadur (1960, 1967, 1971). The other expositions of the Bahadur theory with the reviews of publications in this area may be found in Savage (1969), Groeneboom and Oosterhoff (1977), and Serfling (1980).

Denote for any θ, t and any sequence of statistics $\{T_n\}$

$$F_n(t;\theta) := \mathbf{P}_\theta\left(s\colon\ T_n(s) < t\right), \qquad G_n(t) := \inf\left\{F_n(t;\theta)\colon\ \theta \in \Theta_0\right\}.$$

The quantity

$$L_n(s) := 1 - G_n(T_n(s))$$

is called the *attained level* or the $\mathbf{P}$*-value.* This is a random variable representing the degree to which the test statistic T_n rejects H.

For $\theta \in \Theta_0$ the $\mathbf{P}$-value is distributed approximately uniformly on $[0,\ 1]$. Anyway the following inequality is valid:

$$\mathbf{P}_\theta\left(L_n \le u\right) \le u \qquad \text{for any} \qquad u \in [0,\ 1]\,. \tag{1.2.1}$$

In the case of continuous distribution function $F_n(t;\theta_0)$ this inequality follows directly from the definition of L_n; the general case is based on a suitable approximation (see Bahadur (1971), Theorem 7.4). Therefore working with finite samples one compares the $\mathbf{P}$-value with the preassigned level α and rejects the basic hypothesis if $L_n < \alpha$. The substantial discussion and the interpretation of $\mathbf{P}$-values were described by Gibbons and Pratt (1975), Serfling (1980), Lambert and Hall (1982).

The asymptotic behaviour of L_n under the alternative (then $\theta \in \Theta_1$) is of considerable interest for comparing sequences of test statistics. It occurs usually that for $\theta \in \Theta_1$ the convergence in $\mathbf{P}_\theta$-probability takes place:

$$\mathbf{P}_\theta\text{-}\lim_{n\to\infty} L_n = -\tfrac{1}{2}\,\mathrm{c}_T(\theta) \tag{1.2.2}$$

where $\mathrm{c}_T(\theta)$ is a nonrandom positive function of parameter θ on Θ_1, that is called the *Bahadur exact slope of the sequence* $\{T_n\}$. The factor $\frac{1}{2}$ is present in (1.2.2) due to historical reasons. Some authors, e.g., Groeneboom and Oosterhoff (1977) as well as Chandra and Ghosh (1978) called $\mathrm{c}_T(\theta)$ the *"weak" slope* unlike the *"strong" slope* when the convergence in (1.2.2) takes place with $\mathbf{P}_\theta$-probability 1. In the sequel we will use the term slope mainly for weak slopes as it is sufficient to consider the convergence in probability in most of statistical problems.

One may rewrite (1.2.2) in terms of sample sizes $N_T(\alpha, \beta, \theta)$.

Theorem 1.2.1 (Bahadur (1967), Groeneboom and Oosterhoff (1977)).
If (1.2.2) *is valid for a sequence of statistics* $\{T_n\}$ *with* $c_T(\theta) > 0$, *then*

$$N_T(\alpha, \beta, \theta) \sim \frac{2 \ln 1/\alpha}{c_T(\theta)} \quad \text{as} \quad \alpha \to 0 . \tag{1.2.3}$$

Proof. Let us prove at first that for any $\beta \in (0, 1)$ we have

$$\lim_{n\to\infty} n^{-1} \ln \alpha_n(\beta, \theta) = -\tfrac{1}{2}\, c_T(\theta) . \tag{1.2.4}$$

Suppose that (1.2.4) fails for some β. Let us define for this β the sequence $\{c_n\}$ by formula (1.1.1). There exist an increasing subsequence $\{n_k\}$ and $\varepsilon > 0$ such that one of the following two inequalities

$$-n_k^{-1} \ln \alpha_{n_k}(\beta, \theta) < \tfrac{1}{2}\, c_T(\theta) - \varepsilon , \tag{a}$$

$$-n_k^{-1} \ln \alpha_{n_k}(\beta, \theta) > \tfrac{1}{2}\, c_T(\theta) + \varepsilon \tag{b}$$

is valid for all k.

In the case (a) we obtain, due to (1.2.2), that

$$\begin{aligned}
\mathbf{P}_\theta\Big(-n_k^{-1} \ln L_{n_k} > \tfrac{1}{2}\, c_T(\theta) - \varepsilon\Big) &\le \mathbf{P}_\theta\Big(-n_k^{-1} \ln L_{n_k} > -n_k^{-1} \ln \alpha_{n_k}(\beta, \theta)\Big) \\
&= \mathbf{P}_\theta\Big(L_{n_k} < \alpha_{n_k}(\beta, \theta)\Big) \\
&= \mathbf{P}_\theta\Big(1 - G_{n_k}(T_{n_k}) < 1 - G_{n_k}(c_{n_k})\Big) \\
&\le \mathbf{P}_\theta\Big(T_{n_k} > c_{n_k}\Big) \le \beta \qquad \text{for all } k .
\end{aligned}$$

The left-hand side of the considered inequality tends to 1 as $k \to \infty$ which contradicts the assumption $0 < \beta < 1$. The other case (b) may be examined analogously.

Now let us derive from (1.2.4) the conclusion of the theorem. Put for brevity $N_\alpha := N_T(\alpha, \beta, \theta)$. It follows from (1.2.4) that

$$\alpha_n(\beta, \theta) > \exp\{-n\, c_T(\theta)\}$$

for sufficiently large n. Therefore for any such n the inequality

$$\alpha < \exp\{-n\,\mathrm{c}_T(\theta)\}$$

entails $N_\alpha > n$, ensuring that $N_\alpha \to \infty$ as $\alpha \to 0$.

The definition of N_α implies that

$$\alpha_{N_\alpha}(\beta,\theta) \le \alpha \le \alpha_{N_\alpha - 1}(\beta,\theta)$$

or, equivalently,

$$-N_\alpha^{-1}\ln \alpha_{N_\alpha-1}(\beta,\theta) \le -N_\alpha^{-1}\ln\alpha \le -N_\alpha^{-1}\ln\alpha_{N_\alpha}(\beta,\theta)\,.$$

The transition to the limit as $\alpha \to 0$ together with (1.2.4) completes the proof of (1.2.3). $\diamondsuit$

Thus, if two sequences of statistics $\{V_n\}$ and $\{T_n\}$ are such that (1.2.2) holds, their Bahadur ARE $\mathbf{e}^{\,\mathrm{B}}_{V,T}(\beta,\theta)$ can be calculated, according to (1.1.2), by means of the formula

$$\mathbf{e}^{\,\mathrm{B}}_{V,T}(\beta,\theta) = \mathrm{c}_V(\theta)\big/\mathrm{c}_T(\theta)\,. \tag{1.2.5}$$

If $\mathbf{e}^{\,\mathrm{B}}_{V,T} > 1$ for some θ then we should prefer the sequence $\{V_n\}$ to $\{T_n\}$.

The following theorem contains the most simple method of the calculation of exact slopes.

Theorem 1.2.2 (Bahadur (1967, 1971)). *Let for a sequence $\{T_n\}$ the following two conditions be fulfilled:*

$$T_n \xrightarrow{\mathbf{P}_\theta} b(\theta)\,, \qquad \theta \in \Theta_1\,, \tag{1.2.6}$$

where $-\infty < b(\theta) < \infty$;

$$\lim_{n\to\infty} n^{-1}\ln\bigl[1 - G_n(t)\bigr] = -f(t) \tag{1.2.7}$$

for each t from an open interval I on which f is continuous and $\{b(\theta),\, \theta\in\Theta_1\} \subset I$. Then (1.2.2) is valid and, moreover, for any $\theta \in \Theta_1$

$$\mathrm{c}_T(\theta) = 2\,f\bigl(b(\theta)\bigr)\,. \tag{1.2.8}$$

Formula (1.2.8) plays the exceptionally important role in the Bahadur theory. It shows that for the calculation of exact slopes it is necessary to solve the problem of determining large deviation asymptotics of a sequence $\{T_n\}$ under the null-hypothesis. This problem is always nontrivial as Bahadur himself notes (1971). On the contrary the verification of (1.2.6) does not usually present any difficulties.

In the cases when a sequence $\{T_n\}$ does not satisfy the conditions of Theorem 1.2.2, usually one succeeds in selecting strictly monotone functions ψ_n such that a new sequence $\{T_n^*\}$, $T_n^* := \psi_n(T_n)$, satisfies already these conditions. Since the **P**-value L_n stays invariable under this transform, the exact slope of $\{T_n\}$ coincides with the exact slope of $\{T_n^*\}$, that might be calculated by (1.2.8).

Proof of Theorem 1.2.2. Fix an arbitrary $\theta \in \Theta_1$ and $\varepsilon > 0$ such that $(b-\varepsilon,\, b+\varepsilon) \subset I$. Put

$$\mathbf{\Omega} := \left\{ s \in \mathfrak{X}^{(\infty)} \colon\ b-\varepsilon < T_n(s) < b+\varepsilon \right\}.$$

It follows from (1.2.6) that for sufficiently large n the estimate $\mathbf{P}_\theta(\mathbf{\Omega}) > 1-\delta$ is valid for any $\delta > 0$. As F_n is monotone, the following inequalities

$$1 - F_n(b+\varepsilon) \le L_n(s) \le 1 - F_n(b-\varepsilon)$$

are valid for the same $s \in \mathbf{\Omega}$. Taking logarithms and passing to the limit as $n \to \infty$, we obtain under condition (1.2.7) that

$$-f(b+\varepsilon) \le \varliminf_{n\to\infty} n^{-1} \ln L_n \le \varlimsup_{n\to\infty} n^{-1} \ln L_n \le -f(b-\varepsilon)$$

for each $s \in \mathbf{\Omega}$. By virtue of the continuity of f and of ε being arbitrary it follows now that in $\mathbf{P}_\theta$-probability

$$\lim_{n\to\infty} n^{-1} \ln L_n = -f(b)\,. \qquad \diamondsuit$$

The right-hand side of (1.2.2) may be, generally speaking, a random variable. Note that such situations were discussed by Bahadur and Raghavachari (1972). It is natural in that case to call the exact slope *stochastic* (see Berk and Brown (1978), Kallenberg (1981)) as distinct from nonstochastic exact slopes defined by (1.2.4). Fortunately it follows from Theorem 1.2.1 that, if the limit in (1.2.2) is nonrandom, both notions coincide. In the sequel we shall meet only nonrandom limits in (1.2.2) the values of which may be calculated via Theorem 1.2.2.

Another fundamental result in the Bahadur theory is the existence of an upper bound for exact slopes, that is sometimes compared in the literature with the Cramér–Rao inequality in the estimation theory.

Define for any two elements $\mathbf{P}_\theta$ and $\mathbf{P}_{\theta'}$ from the basic family of distributions the *Kullback–Leibler information number* (or simply *information*) by means of the formula

$$K(\mathbf{P}_\theta, \mathbf{P}_{\theta'}) := \begin{cases} \displaystyle\int_{\mathfrak{X}} \ln \frac{d\,\mathbf{P}_\theta}{d\,\mathbf{P}_{\theta'}}\, d\,\mathbf{P}_\theta & \text{if } \mathbf{P}_\theta \ll \mathbf{P}_{\theta'}\,, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{1.2.9}$$

We shall henceforth often write $K(\theta, \theta')$ instead of $K(\mathbf{P}_\theta, \mathbf{P}_{\theta'})$. The properties of the Kullback–Leibler information have been examined by Kullback (1959), Bahadur (1971) and Borovkov (1984) among others. It is well-known that $K(\theta, \theta') \geq 0$ and $K(\theta, \theta') = 0$ only if $\mathbf{P}_\theta = \mathbf{P}_{\theta'}$.

Put for any $\theta \in \Theta_1$

$$K(\theta, \Theta_0) := \inf \left\{ K(\theta, \theta_0)\colon\ \theta_0 \in \Theta_0 \right\}. \tag{1.2.10}$$

Theorem 1.2.3 (Raghavachari (1970) and Bahadur (1971)). *For any* $\theta \in \Theta_1$ *with* $\mathbf{P}_\theta$*-probability* 1 *we have*

$$\varliminf_{n\to\infty} n^{-1} \ln L_n(s) \geq -K(\theta, \Theta_0)\,. \tag{1.2.11}$$

Proof. Only the case when $K(\theta,\, \Theta_0) < \infty$ is of interest. Fix a $\theta \in \Theta_1$ for which it is valid. For any $\varepsilon > 0$ there exists $\theta_0 \in \Theta_0$ such that

$$0 \leq K(\theta, \theta_0) < K(\theta, \Theta_0) + \varepsilon < +\infty\,. \tag{1.2.12}$$

For any fixed θ and θ_0 denote for brevity $K(\theta, \theta_0)$ by K. If $K < \infty$ we have $\mathbf{P}_\theta \ll \mathbf{P}_{\theta_0}$ and consequently

$$d\,\mathbf{P}_\theta = r(x)\; d\,\mathbf{P}_{\theta_0} \qquad \text{on} \quad \big(\mathfrak{X}, \mathfrak{A}\big)\,,$$

$$d\,\mathbf{P}_\theta^{(n)} = r_n(s)\; d\,\mathbf{P}_{\theta_0}^{(n)} \qquad \text{on} \quad \big(\mathfrak{X}^{(n)}, \mathfrak{A}^{(n)}\big)$$

where $r_n(s) := \Pi_{i=1}^n r(X_i)$. Note that by the strong law of large numbers one can state that with $\mathbf{P}_\theta$-probability 1

$$\lim_{n\to\infty} n^{-1} \ln r_n(s) = K\,. \tag{1.2.13}$$

For any positive integer n let us introduce the events

$$A_n := \big\{L_n < \exp\left[-n\,(K + 2\,\varepsilon)\right]\big\}\,, \qquad B_n := \big\{r_n < \exp\left[n\,(K + \varepsilon)\right]\big\}\,.$$

Then

$$\mathbf{P}_\theta(A_n B_n) = \int\limits_{A_n B_n} d\,\mathbf{P}_\theta^{(n)} = \int\limits_{A_n B_n} r_n\, d\,\mathbf{P}_{\theta_0}^{(n)} \tag{1.2.14}$$

$$\le \exp\left\{n\,(K+\varepsilon)\right\} \int\limits_{A_n} d\,\mathbf{P}_{\theta_0}^{(n)}$$

$$= \exp\left\{n\,(K+\varepsilon)\right\} \;\cdot\; \mathbf{P}_{\theta_0}(A_n) \le \exp\left\{-n\,\varepsilon\right\},$$

where the last inequality follows from (1.2.1).

It follows from (1.2.14) that $\sum_n \mathbf{P}_\theta(A_n B_n) < \infty$. By the Borel–Cantelli lemma only a finite number of events $A_n B_n$ occurs with $\mathbf{P}_\theta$-probability 1. Taking into account (1.2.13) we obtain that the inequality

$$L_n(s) \ge \exp\big\{-n\,(K + 2\,\varepsilon)\big\}$$

holds almost surely for sufficiently large n. Under condition (1.2.12) we establish the conclusion of Theorem 1.2.3 due to the arbitrary choice of ε. Some generalizations are contained in Bahadur and Raghavachari (1972), Bahadur, Chandra and Lambert (1982). $\diamondsuit$

Theorem 1.2.3 implies that the exact slope $\mathrm{c}_T(\theta)$ of any sequence of statistics $\{T_n\}$ satisfies the inequality

$$\mathrm{c}_T(\theta) \le 2\,K(\theta, \Theta_0)\,. \tag{1.2.15}$$

If the equality takes place in (1.2.15) for all $\theta \in \Theta_1$, the sequence $\{T_n\}$ is said to be *asymptotically optimal (AO) in the Bahadur sense.* The class of such statistics is apparently rather narrow, though it contains under certain conditions the likelihood ratio statistics (see Bahadur (1965, 1967) and Rublik (1989)). But if for each $\theta_0 \in \partial\Theta_0$ the weaker condition holds, namely

$$c_T(\theta) \sim 2K(\theta, \Theta_0), \qquad \theta \to \theta_0, \tag{1.2.16}$$

then the sequence $\{T_n\}$ is said to be *locally asymptotically optimal (LAO) in the Bahadur sense.*

In initial papers on the Bahadur ARE it had been impossible to find the function f in (1.2.7) because of insufficient development of large deviations theory. In this connection it had been proposed by Bahadur (1960) to replace the exact distribution of $\{T_n\}$ in the definition of the **P**-value by its limiting distribution. Suppose that for all $\theta_0 \in \Theta_0$ and $t \in \mathbf{R}^1$

$$F_n(t, \theta_0) \longrightarrow F(t) \qquad \text{as} \qquad n \to \infty.$$

Then the substitute of $L_n(s)$ is

$$L_n^*(s) := 1 - F\big(T_n(s)\big),$$

and what's more in typical cases there exists the limit in $\mathbf{P}_\theta$-probability

$$\lim_{n\to\infty} \big(-n^{-1} \ln L_n^*\big) = \tfrac{1}{2}\, c_T^*(\theta) > 0. \tag{1.2.17}$$

If (1.2.17) is actually valid, the function $c_T^*(\theta)$ is called the *approximate* (unlike exact) *slope of the sequence* $\{T_n\}$. The ratio of approximate slopes of two sequences of statistics $\{V_n\}$ and $\{T_n\}$ is called their *approximate Bahadur* ARE and is denoted by $\mathbf{e}^{*\,\mathrm{B}}_{V,T}(\theta)$.

The method of calculating approximate slopes analogous with Theorem 1.2.2 takes place (see Bahadur (1960)): *suppose a sequence* $\{T_n\}$ *satisfies* (1.2.6) *for some function* b, *and for some constant* a, $0 < a < \infty$, *the limiting distribution function* F *submits to the condition that*

$$\ln\big[1 - F(t)\big] \sim -\tfrac{1}{2}\, a\, t^2, \qquad t \to \infty. \tag{1.2.18}$$

Then (1.2.17) *holds and, besides,*

$$c_T^*(\theta) = a\, b^2(\theta).$$

Approximate slopes are not very reliable as means of the comparison of tests because monotone transforms of test statistics may lead to entirely different approximate slopes (see, e.g., Groeneboom and Oosterhoff (1977)). But nevertheless they are still used in the statistical literature, mainly for the following reasons:

the approximate Bahadur ARE may be easier calculated than any other known type of ARE's;

the approximate and exact slopes are often locally (as $\theta \to \theta_0$) equivalent, so that the approximate ARE gives a notion of the local exact ARE;

the approximate slopes give a simple method of the calculation of the Pitman ARE (see more detail in Section 1.4).

A definite merit of the Bahadur ARE in the opinion of a number of authors (see Savage (1969), Groeneboom and Oosterhoff (1977) and Singh (1984)) lies in the fact that it permits one to distinguish tests in the cases when other types of ARE's are useless. We quote some typical examples.

Let $X_1, X_2, \ldots, X_n$ be a normally distributed sample with parameters $(\theta, 1)$. The null-hypothesis H: $\theta = 0$ is tested against the alternative A: $\theta > 0$. With that end two sequences of statistics are proposed: the sample means $\{\bar{X}_n\}$ and the Student ratios $\{t_n\}$. As the Student test does not use the information that the true variance is equal to 1, it should lose to the test based on $\{\bar{X}_n\}$. The Bahadur exact slopes prove it correct (see Bahadur (1971)) because for any $\theta > 0$

$$\mathrm{c}_t(\theta) \equiv \ln(1+\theta^2) < \mathrm{c}_{\bar{X}}(\theta) \equiv \theta^2 .$$

Meanwhile these tests are indistinguishable from the point of view of the Hodges–Lehmann and Pitman ARE's as

$$\mathbf{e}^{\mathrm{HL}}_{t,\bar{X}}(\alpha,\theta) = \mathbf{e}^{\mathrm{P}}_{t,\bar{X}}(\alpha,\beta) \equiv 1 .$$

The other example is based on the paper by Mason (1984) where it has been shown that the statistics based on sequential ranks had lower Bahadur efficiency than usual rank statistics whereas they are equivalent from the standpoint of the Pitman efficiency.

Now we return to the definition of exact slopes. Formula (1.2.2) may be interpreted as the law of large numbers for logarithms of **P**-values. Lambert and

Hall (1982) initiated a new stage in studying the asymptotics of **P**-values by proving that for large n and under appropriate conditions the logarithms of **P**-values have approximately the normal distribution. The following theorem is the precise formulation of their result.

Theorem 1.2.4 (Lambert and Hall (1982)). *Suppose for $\theta \in \Theta_1$ there exist constants $b(\theta)$ and $\sigma(\theta)$, $-\infty < b(\theta) < \infty$, $0 < \sigma(\theta) < \infty$, such that for any $z \in \mathbf{R}^1$*

I. $\mathbf{P}_\theta\big(\sqrt{n}\,(T_n - b(\theta)) < z\big) \to \Phi\big(z/\sigma(\theta)\big)$, $n \to \infty$, *where Φ is the standard normal distribution function. In addition, assume that I is an open interval containing $b(\theta)$ and there exists on I a real function g, possessing such property that if $b \in I$ and $\{b_n\}$ is a sequence of numbers, $b_n = b + O\big(n^{-1/2}\big)$, then under the hypothesis H we have*

II. $\ln\big[1 - G_n(b_n)\big] = -n\,g(b_n) + o(\sqrt{n})\,, \qquad n \to \infty.$

Moreover, assume that

III. *the function g is continuously differentiable on I.*

Then for all $\theta \in \Theta_1$ and $z \in \mathbf{R}^1$

$$\mathbf{P}_\theta\Bigg(\frac{\ln L_n + \frac{1}{2}\,n\,\mathrm{c}(\theta)}{\sqrt{n}} < z \Bigg) \longrightarrow \Phi\Bigg(\frac{z}{\tau(\theta)} \Bigg) \qquad \text{as} \qquad n \to \infty\,,$$

where

$$\mathrm{c}(\theta) := 2\,g\big(b(\theta)\big)\,, \qquad \tau(\theta) := \sigma(\theta)\,g'\big(b(\theta)\big)\,.$$

Here $\mathrm{c}(\theta)$ is the exact slope of the sequence $\{T_n\}$ and $\tau^2(\theta)$ is the asymptotic variance of $\ln L_n$. The arguments of Lambert and Hall (1982) show that the pair $\big(\mathrm{c}(\theta),\, \tau^2(\theta)\big)$ may serve as a more sensitive measure of the test comparison than the single exact slope $\mathrm{c}(\theta)$. This corresponds to the well-known situation in the estimation theory when the quality and the choice of the unbiased estimator depend on its variance. If two different tests have the same or very close exact slopes, but the asymptotic variances of logarithms of their **P**-values are distinct, then just the comparison of the latter enables one to realize the well-founded choice of a suitable test. Moreover it is possible to obtain the following refinement of Theorem 1.2.1. Put for brevity

$$\mathrm{c} := \mathrm{c}(\theta)\,, \qquad A := -2\,\ln\alpha\,, \qquad \tau^2 := \tau^2(\theta).$$

Lambert and Hall (1982) proved that

$$N_T(\alpha,\beta,\theta) = \frac{A}{\mathrm{C}}\left\{1 + \frac{\tau\ \Phi^{-1}(\beta)}{(A\,\mathrm{C})^{1/2}} + o\big(A^{-1/2}\big)\right\}, \qquad A \to \infty\,.$$

So precisely the asymptotic variance τ^2 determines the correction term with respect to the main term A/C from Theorem 1.2.1. Generalizations of these results may be found in Bahadur, Chandra and Lambert (1982), Kallenberg (1983b), Berk (1984), Bahadur and Gupta (1985), and Chandra (1989).

The verification of conditions of Theorem 1.2.4, especially of condition II, is a rather complicated task. It has been done by Lambert and Hall (1982) for some important parametric statistics. The Ph.D. Thesis by Leont'yev (1990) was devoted to the proof of assertions close to Theorem 1.2.4 for nonparametric and asymptotically nonparametric goodness-of-fit statistics; specifically the Kolmogorov statistic was considered by Leont'yev (1987) and the ω^2-statistic also by Leont'yev (1988).

In conclusion of this Section we note that the attempt of constructing the sequential variant of the Bahadur theory has been made by Berk and Brown (1978), and Kourouklis (1984). It turns out, however, that the Bahadur efficiency is not well-adapted to the discrimination of sequential tests.

• 1.3 Hodges–Lehmann Asymptotic Relative Efficiency

This type of the ARE proposed by Hodges and Lehmann (1956) is at most in the conformity with the classical approach of Neyman and Pearson, but at the same time is studied inadequately. This is apparent because of the purely technical difficulties connected with the calculation of large deviation asymptotics of a given sequence of statistics not under the null-hypothesis (as in the case of the Bahadur ARE) but under the alternative. These difficulties are pointed out by Bahadur (1967), Singh (1984) and also by Rao (1962) by the discussion of different measures of efficiency. After the paper of Hodges and Lehmann (1956) there was a standstill in studying the Hodges–Lehmann ARE until the appearance of the paper of Brown (1971) on the conditions of asymptotic optimality of likelihood ratio tests. Then later appeared papers of Raghavachari (1983), Brown, Ruymgaart and Truax (1984), Baringhaus (1987) and Kourouklis (1988) which were also devoted essentially to parametric tests. As a matter of fact the Hodges–Lehmann efficiency has not been used for comparing nonparametric tests.

Let in the framework of Section 1.1 $\beta_n(\alpha,\theta)$ be the power of the significance test based on the sequence $\{T_n\}$ having the size not exceeding $\alpha \in (0,\,1)$ and considered under the alternative $\theta \in \Theta_1$. If there exists a function $d_T(\theta)$, $0 < d_T(\theta) < \infty$, such that

$$\lim_{n\to\infty} n^{-1} \ln\bigl(1-\beta_n(\alpha,\theta)\bigr) = -\tfrac{1}{2}\, d_T(\theta)\,, \tag{1.3.1}$$

then $d_T(\theta)$ is called the *Hodges–Lehmann index* (or simply *index*) of the sequence $\{T_n\}$. For two such sequences the Hodges–Lehmann ARE (1.1.4) is equal to the ratio of corresponding indices.

Indeed, it follows from the definition of $N := N_T(\alpha,\beta,\theta)$ and condition (1.3.1) that $N \to \infty$ as $\beta \to 1$ and

$$\beta_N(\alpha,\theta) \geq \beta \geq \beta_{N-1}(\alpha,\theta)\,.$$

This implies

$$N^{-1}\ln(1-\beta) \sim -\tfrac{1}{2}\, d_T(\theta) \qquad \text{as} \qquad \beta \to 1\,,$$

or, equivalently,

$$N_T(\alpha,\beta,\theta) \sim -2\,\ln(1-\beta)\,/\,d_T(\theta) \qquad \text{as} \qquad \beta \to 1\,. \tag{1.3.2}$$

It remains to use (1.3.2) and the definition of the Hodges–Lehmann ARE.

Thus the indices serve as a measure of the Hodges–Lehmann ARE. Formula (1.3.1) shows that their computation is based on the large deviation asymptotics of $\{T_n\}$ under the alternative.

There exists an upper bound for the Hodges–Lehmann indices analogous to the upper bound for exact slopes stated by inequality (1.2.15). Introduce for any $\theta \in \Theta_1$

$$K(\Theta_0,\theta) := \inf\bigl\{K(\theta_0,\theta)\colon\ \theta_0 \in \Theta_0\bigr\}\,. \tag{1.3.3}$$

The quantity $K(\Theta_0,\theta)$ is "dual" to the quantity specified in (1.2.10), but does not coincide with it since the Kullback–Leibler information is nonsymmetric.

Theorem 1.3.1. *For any sequence of statistics $\{T_n\}$ and any $\theta \in \Theta_1$ we have*

$$\varliminf_{n\to\infty} n^{-1}\ln\bigl(1-\beta_n(\alpha,\theta)\bigr) \geq -K(\Theta_0,\theta)\,. \tag{1.3.4}$$

The proof of Theorem 1.3.1 is very close to the proof of Theorem 1.2.3 and may be found in Nikitin (1986c, 1987a) or Kourouklis (1988). Generalization for several samples will be given below in Theorem 2.7.1.

It follows from (1.3.4) that the index of $\{T_n\}$ should satisfy the inequality

$$d_T(\theta) \le 2\,K(\Theta_0, \theta)\,. \tag{1.3.5}$$

As in the Bahadur theory the sequence of statistics $\{T_n\}$ is said to be *asymptotically optimal in the Hodges–Lehmann sense* if the equality takes place in (1.3.5). If

$$d_T(\theta) \sim 2\,K(\Theta_0, \theta) \qquad \text{as} \qquad \theta \to \theta_0 \in \partial\,\Theta_0\,, \tag{1.3.6}$$

this sequence is said to be *locally asymptotically optimal (LAO) in the Hodges–Lehmann sense.*

Hettmansperger (1973) has made the attempt to define the notion of the approximate Hodges–Lehmann ARE by analogy with the approximate Bahadur ARE. This concept has, however, only the limited value because in a number of important cases the exact and approximate Hodges–Lehmann ARE's do not coincide even locally. Some considerations in this connection may be found in Mikulski (1976). The Hodges–Lehmann ARE of sequential tests was touched upon by Berk (1976).

• 1.4 Pitman Asymptotic Relative Efficiency

The classical Pitman efficiency is very often used for the purpose of the asymptotic comparison of various tests. It has been introduced at the end of the fourties by Pitman (1949) in his unpublished lectures. Therefore the calculations of this type of the ARE has been for a long time based on the Pitman results stated in the publications by Noether (1950, 1955). The main Pitman's result looks as follows.

Theorem 1.4.1. *Let* $\Theta = \mathbf{R}^1$ *and* $\Theta_0 = (-\infty, \theta_0]$. *Suppose a sequence of statistics* $\{T_n\}$ *possesses the following three properties.*

(1) *There exist functions* μ *and* σ *such that*

$$\lim_{n\to\infty} \mathbf{P}_{\theta_n}\left(\frac{T_n - \mu(\theta_n)}{\sigma(\theta_n)} < z \right) = \Phi(z) \tag{1.4.1}$$

for all $z \in \mathbf{R}^1$ *and* $\theta_n = \theta_0 + k\,n^{-1/2}$, $k \ge 0$.

(2) *There exists the right-sided derivative* $\mu'(\theta_0) > 0$.

(3) $$\lim_{n\to\infty} \frac{\mu'(\theta_n)}{\mu'(\theta_0)} = 1 , \qquad \lim_{n\to\infty} \frac{\sigma(\theta_n)}{\sigma(\theta_0)} = 1 .$$

Let $\{\widetilde{T}_n\}$ *be another sequence of statistics satisfying the same conditions with functions* $\tilde{\mu}$ *and* $\tilde{\sigma}$. *Then the Pitman ARE exists for all* $0 < \alpha < \beta < 1$ *and may be calculated by the formula*

(1.4.2) $$\mathbf{e}^{\,\mathrm{P}}_{T,\tilde{T}}(\alpha,\beta,\theta_0) \equiv \mathbf{e}^{\,\mathrm{P}}_{T,\tilde{T}}(\theta_0) = \left\{ \frac{\mu'(\theta_0)}{\sigma(\theta_0)} \Big/ \frac{\tilde{\mu}'(\theta_0)}{\tilde{\sigma}(\theta_0)} \right\}^2 .$$

More general results may be found in Kendall and Stuart (1960) and Serfling (1980).

We quote now as an example one of the first Pitman's results that stimulated the development of nonparametric statistics. Consider one of classical problems of testing the hypothesis about mean of the Gaussian law under the location alternative. Let $\mathbf{e}^{\,\mathrm{P}}_{W,\,t}$ be the Pitman ARE of the one-sample Wilcoxon rank test with respect to the Student test. It turns out that

$$\mathbf{e}^{\,\mathrm{P}}_{W,\,t} = 3/\pi \approx 0.955 ,$$

and it shows that the ARE of the Wilcoxon test in the comparison with the Student test being optimal in this problem is unexpectedly high. Then Hodges and Lehmann (1956) proved that

$$0.864 \le \mathbf{e}^{\,\mathrm{P}}_{W,\,t} \le +\infty ,$$

if one rejects the assumption of normality and, moreover, the lower bound is attained at the density

$$f(x) = \begin{cases} 3\,(5 - x^2)/(20\sqrt{5}) & \text{if } |x| \le \sqrt{5} , \\ 0 & \text{otherwise} . \end{cases}$$

The question of the ARE bounds for various pairs of nonparametric statistics is of interest and is closely connected with the tail-ordering of distributions. Having no possibility to touch upon this topic, we address readers to the references of Hodges and Lehmann (1961), Sinha and Wieand (1977), and Weissfeld and Wieand (1984) as well as Loh Wei-Yin (1984) and Caperaà (1988), containing most promising advancements.

Returning to Theorem 1.4.1 we note that under its conditions the Pitman ARE is a constant independent on α and β. It is ensured by the asymptotic normality of statistics T_n and $\widetilde{T}_n$ both under the null-hypothesis and the alternative, i.e. by condition (1.4.1). If this condition fails, the considerable difficulties arise when calculating the Pitman ARE as the latter may not at all exist or may depend on α and β. The results by Rothe (1981) give an example of such situation for the limiting χ^2-distribution. In the case of the Kolmogorov–Smirnov and ω^2-statistics or their variants which, as it is well-known, have nonnormal limiting distributions under the null-hypothesis, the values of the Pitman ARE are either unknown or very crude bounds are obtained for them. See Yu (1971), Mikulski (1976), Archambault and Mikulski (1979), Neuhaus (1982).

The important progress in this field had been made by Wieand (1976), who has established the correspondence between the limit of the Pitman efficiency as $\alpha \to 0$ and the limit as $\theta \to \theta_0$ of the approximate Bahadur efficiency which is easy to calculate. The results of such nature for asymptotically normal statistics had been obtained by Bahadur (1960). The merit of the paper by Wieand (1976) consists in the possibility of considering much more general situations.

To formulate the main result by Wieand (1976) we need some strengthening of condition (1.2.6) in Theorem 1.2.2. Let us say that the sequence of statistics $\{T_n\}$ is a *Wieand sequence* if there exists θ^* such that for any $\varepsilon > 0$ and $\delta \in (0, 1)$ there exists a constant $\mathcal{C}$ with the following property: for any $\theta \in (\theta_0, \theta^*)$ and $n > \mathcal{C}\, b^{-2}(\theta)$

$$\mathbf{P}_\theta\Big(\big|\, T_n - b(\theta)\,\big| > \varepsilon\, b(\theta)\Big) < \delta\,. \tag{1.4.3}$$

Therefore for any Wieand's sequence the convergence T_n to $b(\theta)$ in probability takes place at a certain rate.

We quote an auxiliary statement by Wieand (1976) that often simplifies the verification of the above condition.

Theorem 1.4.2. *Let* $\{U_{n,\theta}\}$ *be a family of sequences of statistics satisfying for all* $\theta \in \Theta$ *and* $z \in \mathbf{R}^1$ *the relation*

$$\lim_{n\to\infty} \mathbf{P}_\theta\,(U_{n,\,\theta} < z) = \mathcal{Q}(z) \tag{1.4.4}$$

where $\mathcal{Q}$ *is continuous distribution function and the rate of convergence in* (1.4.4) *is independent on* θ. *Let* $d(\theta) \in (0, 1)$ *be an arbitrary function on* (θ_0, θ^*). *Then* *for any* $\varepsilon > 0$ *and* $\delta \in (0,\,1)$ *there exists a number* $\mathcal{C}'$ *such that for all* $\theta_0 < \theta < \theta^*$ *and* $n > \mathcal{C}' d^{-2}(\theta)$ *the following estimate holds:*

$$\mathbf{P}_\theta\Big(|\,U_{n,\theta}\,|/\sqrt{n} < \varepsilon\, d(\theta)\Big) > 1 - \delta\,.$$

Recall that the Pitman ARE $\mathbf{e}^{\,\mathrm{P}}_{V,\,T}(\alpha, \beta)$ of the sequence $\{V_n\}$ with respect to the sequence $\{T_n\}$ has been defined as the limit of the relative efficiency $\mathbf{e}_{V,\,T}(\alpha, \beta, \theta)$ as $\theta \to \theta_0$ (see (1.1.5)). As this limit may not necessarily exist, let us introduce the *upper* and *lower Pitman* ARE's as

$$\mathbf{e}^{+\,\mathrm{P}}_{V,\,T}(\alpha, \beta) := \sup_{(\mathbf{\Pi})} \ \overline{\lim_{j\to\infty}}\ \mathbf{e}_{V,\,T}(\alpha, \beta, \theta_j)\,,$$

$$\mathbf{e}^{-\,\mathrm{P}}_{V,T}(\alpha, \beta) := \inf_{(\mathbf{\Pi})} \ \underline{\lim}_{j\to\infty}\ \mathbf{e}_{V,T}(\alpha, \beta, \theta_j)$$

where the symbol $(\mathbf{\Pi})$ denotes the set of all sequences $\{\theta_j\}$, $\theta_j \in \Theta_1$, $\theta_j \to \theta_0$. The main result by Wieand (1976) is as follows.

Theorem 1.4.3. *Let* $\{V_n\}$ *and* $\{T_n\}$ *be two Wieand's sequences of statistics having under* H *the limiting distribution functions satisfying* (1.2.18) *and being strictly increasing for all large values of their arguments. If there exists the finite limit* $\lim_{\theta\to\theta_0} \mathbf{e}^{*\,\mathrm{B}}_{V,\,T}(\theta)$, *then*

$$\lim_{\theta\to\theta_0} \mathbf{e}^{*\,\mathrm{B}}_{V,\,T}(\theta) = \lim_{\alpha\to 0} \mathbf{e}^{+\,\mathrm{P}}_{V,\,T}(\alpha, \beta) = \lim_{\alpha\to 0} \mathbf{e}^{-\,\mathrm{P}}_{V,\,T}(\alpha, \beta)\,. \tag{1.4.5}$$

The common value of finite limits in (1.4.5) is called the *limiting* (as $\alpha \to 0$) *Pitman* ARE. Under conditions of Theorem 1.4.3 it is independent on β.

The absolute majority of statistics considered in this book satisfies the conditions of the theorems of present Section and this enables one to calculate their Pitman or limiting Pitman ARE's.

The examples of such calculations may be found in Kendall and Stuart (1967), Lehmann (1975), Wieand (1976), Pitman (1979), Pratt and Gibbons (1981). A survey of results concerning the Pitman ARE for parametric statistics, taking into account the effects of the "second order", has been presented by Chibisov (1983).

In conclusion we touch upon the problem of an upper bound for the Pitman ARE. As stated above, under certain conditions including the asymptotic normality under the null-hypothesis and the alternative, the Pitman ARE depends on the so-called "*efficacy*" $\big(\mu'(\theta_0)/\sigma(\theta_0)\big)^2$. Under some additional conditions Rao (1963) proved that

$$\big(\mu'(\theta_0)/\sigma(\theta_0)\big)^2 \le \mathbf{I}\,(\theta_0) \tag{1.4.6}$$

where $\mathbf{I}\,(\theta_0)$ is the Fisher information at the point θ_0, corresponding to the distribution of initial observations. If the equality in (1.4.6) takes place, the sequence of statistics is said to be *asymptotically optimal in the Pitman sense.* Any variants of (1.4.6) for statistics having nonnormal limit distribution are unknown.

• 1.5 Different Approaches to the Definition of the ARE

In this Section we shall describe briefly the different approaches to the definition and the calculation of ARE's. Some authors believe that the Bahadur and Hodges–Lehmann approaches contain a certain "lack of balance" as when determining the corresponding efficiencies one of error probabilities is kept fixed and the other tends to zero. The Chernoff ARE introduced by Chernoff (1952) is free from this drawback as here both error probabilities are tending to zero. The theory of the Chernoff efficiency was later developed by Kallenberg (1982).

Let, as in Section 1.3, $\beta_n(\alpha,\theta)$ be the power at the point $\theta \in \Theta_1$ of the test based on a sequence of statistics $\{T_n\}$ with critical large values and the size not exceeding $\alpha \in [0,1]$. Put

$$\rho_n(\alpha,\theta) := \max\big\{\alpha,\, 1-\beta_n(\alpha,\theta)\big\}$$

and further

$$\rho_n(\theta) := \inf\big\{\rho_n(\alpha,\theta)\colon\ 0 \le \alpha \le 1\big\}\,.$$

If there exists a function $\rho_T(\theta)$, $0 < \rho_T(\theta) < \infty$, such that for each $\theta \in \Theta_1$

$$\lim_{n\to\infty} n^{-1} \ln\, \rho_n(\theta) = -\rho_T(\theta)\,, \tag{1.5.1}$$

it is called the *Chernoff index* of a given sequence $\{T_n\}$. It is clear that $\rho_T(\theta)$ is describing the rate of the exponential decreasing for the maximum of the probabilities of the first and second kind errors.

For two sequences of statistics $\{T_n\}$ and $\{V_n\}$ with corresponding Chernoff indices $\rho_T(\theta)$ and $\rho_V(\theta)$ the Chernoff ARE is defined by

$$\mathbf{e}^{\,C}_{T,V}(\theta) := \rho_T(\theta)\,/\rho_V(\theta)\,. \tag{1.5.2}$$

The following theorem proved by Kallenberg (1982) shows that the computation of the Chernoff indices needs rough large deviation asymptotics for test statistics both under the null-hypothesis and the alternative.

Theorem 1.5.1. *Let for some* $c^* \in \mathbf{R}^1$ *and* $\theta \in \Theta_1$

$$\lim_{n\to\infty} n^{-1} \ln \sup\Big\{\mathbf{P}_{\theta_0}(T_n > c^*):\ \theta_0 \in \Theta_0\Big\} = \lim_{n\to\infty} n^{-1} \ln \mathbf{P}_\theta\left(T_n \le c^*\right). \tag{1.5.3}$$

Then $\rho_T(\theta) = -\,u(c^*)$ *where* $u(c^*)$ *is the common value of two limits standing in* (1.5.3).

There exists an upper bound for the Chernoff indices, too. We describe it now in the simplest case. Assume we are testing the simple hypothesis $H\colon\ \theta = \theta_0$ against the alternative $A\colon\ \theta = \theta_1$ and the distributions $\mathbf{P}_{\theta_0}$ and $\mathbf{P}_{\theta_1}$ have the common support in $\mathbf{R}^1$. Denote by $f(x;\theta_0)$ and $f(x;\theta_1)$ their densities with respect to the measure

$$\mu = \tfrac{1}{2}\left(\mathbf{P}_{\theta_0} + \mathbf{P}_{\theta_1}\right)$$

and consider the function

$$\psi(t) := \ln \int_{\mathbf{R}^1} \exp\left[\, t\,\ln\left\{f(x;\theta_1)/f(x;\theta_0)\right\}\right] d\,\mathbf{P}_{\theta_0}(x)\,.$$

It is obvious that ψ is strictly convex on $[0,1]$ and $\psi(0) = \psi(1) = 0$. Hence, there exists a unique point $t^* \in [0,1]$ such that

$$\psi(t^*) = \min\left\{\psi(t):\ t \in [0,1]\right\}.$$

It turns out (see Kallenberg (1982)) that the Chernoff index $\rho_T(\theta_1)$ of any sequence of statistics $\{T_n\}$ used for this problem satisfies the inequality

$$\rho_T(\theta_1) \le -\psi(t^*)\,. \tag{1.5.4}$$

More general variants of this inequality may also be found in Kallenberg (1982). In the mentioned paper it has been proved that the likelihood ratio test is Chernoff asymptotically optimal if the distribution of observations belongs to an exponential family.

Ronzhin (1985) introduced a certain generalization of the Chernoff efficiency where the error probabilities α and β are decreasing in such manner that

$$\ln\alpha \,/\, \ln\beta \longrightarrow \mu \in [0, +\infty]\,.$$

Kallenberg (1983a) presented another approach to measuring the ARE, taking some intermediate position between the Bahadur and Pitman approaches. Under this approach the level α_n tends to zero, but not too fast. The alternative value of parameter θ_n also tends to θ_0 at a controlled rate. Besides, θ_n and α_n are coordinated in such a way that the power of the test at the point θ_n is separated from 0 and 1. Kallenberg (1983a) has shown that for calculating this "intermediate", or Kallenberg ARE, one needs the information about large deviations of moderate or Cramér's type. Moderate deviations have been used also by Rubin and Sethuraman (1965b) for the definition of the Bayes risk efficiency. (See also Serfling (1980).)

Recently Borovkov and Mogulskii (1992) proposed another definition of efficiency where the probabilities of errors had been used in a more symmetric way. Denote for any $\alpha \in \mathbf{R}^1$ and any sequence of suitably normed test statistics $\{T_n\}$

$$\varepsilon_{n,\alpha}(T) := \sup\Big\{\mathbf{P}_\theta\big(T_n > \alpha\sqrt{n}\big)\colon\ \theta\in\Theta_0\Big\}\,,$$

$$\delta_{n,\alpha}(T) := \sup\Big\{\mathbf{P}_\theta\big(T_n \le \alpha\sqrt{n}\big)\colon\ \theta\in\Theta_1\Big\}\,,$$

and let us introduce for any given $\varepsilon > 0$ and $\delta \in (0,\,1)$ a new characteristic as

$$n(\varepsilon,\delta) := \min\Big\{\, n \;\Big|\; \exists\, \alpha_0\colon\ \varepsilon_{n,\,\alpha_0}(T) \le \varepsilon\,,\ \delta_{n,\,\alpha_0}(T) \le \delta \,\Big\}\,. \tag{1.5.5}$$

It is clear that $n(\varepsilon,\delta)$ is the minimal sample size, ensuring for $\{T_n\}$ the given probabilities of errors ε and δ. By applying (1.5.5) it makes possible to define the relative efficiency and further the ARE with the aid of the appropriate limit.

Three last types of the ARE are of indisputable interest, however their study and calculation are beyond the framework of this book.

• 1.6 Some Results on Probabilities of Large Deviations

As we have seen in the previous Sections the calculation of the Bahadur, Hodges–Lehmann and Chernoff ARE's is closely related with the evaluation of rough asymptotics for the large deviation probabilities both under the null-hypothesis and the alternative. Besides we are interested in the first place in the deviations of order $O(\sqrt{n})$. In the classification by Wentzell (1990, Sec. 4.2) these are "very large" deviations. Sometimes they are called "Chernoff's type" to distinguish them from "Cramér's type" deviations of order $o(\sqrt{n})$ and moderate deviations of order $O(\sqrt{\ln n})$ for which the analysis requires somewhat different technique. The literature connected with large deviations is extremely vast and numbers hundreds positions. We point out among them the fundamental papers by Chernoff (1952) and Sanov (1957) as well as the papers by Bahadur and Ranga Rao (1960), Hoeffding (1965), Borovkov (1967), Petrov (1972), Donsker and Varadhan (1975a, 1975b, 1976), Bahadur and Zabell (1979), Groeneboom, Oosterhoff and Ruymgaart (1979), Borovkov and Mogulskii (1978, 1980), Azencott (1980), Bolthausen (1984), Varadhan (1984), Ellis (1984), Mogulskii (1984), Stroock (1984), Deuschel and Stroock (1989), Wentzell (1990), Saulis and Statulevicius (1991).

We quote in this Section some most frequently used results in this field necessary for the next Chapters. They may be symbolically divided in following two groups: one is connected with the Chernoff theorem and the other is connected with the Sanov theorem.

Let Y be a real random variable with d.f. F. Denote by ψ the moment generating function of Y :

$$\psi(t) := \mathbf{E}\exp\{t\,Y\} = \int_{-\infty}^{\infty} e^{ty}\,dF(y)\,, \qquad -\infty < t < \infty\,, \tag{1.6.1}$$

and put

$$\rho := \inf\left\{\psi(t)\colon\ t \geq 0\right\}. \tag{1.6.2}$$

It is clear that $0 \leq \rho \leq 1$.

The following elementary result ascending to S.N. Bernstein appears very useful in the future.

Theorem 1.6.1.

$$\mathbf{P}\,(Y \ge 0) \le \rho\,.$$

For any sequence $\{Y_j\}$ of independent identically distributed (i.i.d.) random variables having the common distribution with Y this theorem yields

$$\mathbf{P}\big(Y_1 + \cdots + Y_n \ge 0\big) \le \rho^n\,, \qquad n \ge 1\,. \tag{1.6.3}$$

But in this respect profound results have been obtained by the following theorem.

Theorem 1.6.2. *Let* $\{u_n\}$ *be a real sequence such that* $u_n \to u$, $-\infty < u < +\infty$, *and*

$$\mathbf{P}\,(Y > u) > 0\,. \tag{1.6.4}$$

Then

$$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}\big(Y_1 + Y_2 + \cdots + Y_n \ge n\,u_n\big) = -f(u)\,, \tag{1.6.5}$$

$$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}\big(Y_1 + Y_2 + \cdots + Y_n > n\,u_n\big) = -f(u) \tag{1.6.6}$$

where $f(u)$ *is determined by the equality*

$$\exp\left[-f(u)\right] = \inf\left\{e^{-tu}\,\psi(t):\ t \ge 0\right\}. \tag{1.6.7}$$

If one assumes additionally that $\psi(t) < \infty$ *in a neighborhood of zero,* $\mathbf{E}\,Y_i = 0$ *and* $\operatorname{Var} Y_i = \sigma^2 > 0$, *then*

$$f(u) = \frac{u^2}{2\,\sigma^2}\,\big(1 + o(1)\big) \qquad \text{as} \qquad u \to 0\,.$$

Theorem 1.6.2 belongs to Chernoff (1952) who used as a base the earlier results by Cramér (1937). Variants of the proof may be found, e.g., in Bahadur (1971) and Steinebach (1980).

There exists a lot of generalizations of the Chernoff theorem (Borovkov and Mogulskii (1978, 1980), Bahadur and Zabell (1979), Bretagnolle (1979) and others). We confine ourselves to the following result by Sethuraman (1964) (see also Rao (1972)) dealing with random variables taking on values in a separable Banach space

$\mathcal{Y}$ with the corresponding norm $\|\cdot\|$. Denote by $\mathcal{Y}_1^*$ the space of all continuous linear functionals y^* on $\mathcal{Y}$ with the unit norm.

Theorem 1.6.3. *Let $\{Y_n(\omega)\}$ be a sequence of independent identically distributed random variables defined on a probability space $(\Omega, \mathfrak{F}, \mathbf{P})$ with the values in $\mathcal{Y}$. Suppose that*

$$\int_\Omega y^*\big(Y_1(\omega)\big)\, d\,\mathbf{P}(\omega) = 0\,, \qquad \forall\ y^* \in \mathcal{Y}_1^*\,, \tag{1.6.8}$$

$$\int_\Omega \exp\big\{\, z\, \|\, Y_1(\omega)\,\|\,\big\}\, d\,\mathbf{P}(\omega) < \infty\,, \qquad \forall\ z \in \mathbf{R}^1\,. \tag{1.6.9}$$

Then for any $\varepsilon > 0$ one has

$$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}\big\{\omega\colon\ \|\, Y_1(\omega) + \cdots + Y_n(\omega)\,\| \ge n\,\varepsilon\big\} = \ln \rho(\mathcal{Y}_1^*, \varepsilon)$$

where

$$\rho(\mathcal{Y}_1^*, \varepsilon) := \sup\Big\{\rho(y^*, \varepsilon)\colon\ y^* \in \mathcal{Y}_1^*\Big\}\,,$$

and

$$\begin{aligned}\rho(y^*, \varepsilon) := \max\Big[\rho_1(y^*, \varepsilon) &\equiv \min\Big\{e^{-t\,\varepsilon}\,\mathbf{E}\,\exp\Big[t\,y^*(Y_1(\omega))\Big]\colon\ t \ge 0\Big\}\,,\\ \rho_2(y^*, \varepsilon) &\equiv \min\Big\{e^{t\,\varepsilon}\,\mathbf{E}\,\exp\Big[t\,y^*(Y_1(\omega))\Big]\colon\ t \le 0\Big\}\Big]\,.\end{aligned}$$

Moreover,

$$\rho(\mathcal{Y}_1^*, \varepsilon) = -\frac{\varepsilon^2}{2\,\sigma^2}\,\big(1 + o(1)\big) \qquad \text{as} \qquad \varepsilon \to 0$$

where

$$\sigma^2 = \sup\Big\{\,\mathrm{Var}\ y^*(Y_1(\omega))\colon\ y^* \in \mathcal{Y}_1^*\,\Big\}\,.$$

Note that the function $\rho\big(\mathcal{Y}_1^*, \varepsilon\big)$ is continuous in ε in a neighborhood of zero.

Other generalizations of the Chernoff theorem are connected with the transition from sums of random variables to U-statistics. Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables and $\mathbf{\Phi}\colon \mathbf{R}^m \to \mathbf{R}^1$ be a symmetric function of m variables, $m < n$. We remind that U-statistic is a random variable looking as

$$U_n := \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} \mathbf{\Phi}(Y_{i_1}, \ldots, Y_{i_m})\,.$$

The number m and $\mathbf{\Phi}$ are defined as the *degree* of U-statistic and its *kernel* respectively. The problem of the Chernoff type large deviations of U-statistics has been solved in the following theorem by Dasgupta (1984). Denote first

$$\varphi(t) := \mathbf{E}\left\{\Phi(Y_1, \dots, Y_m) \,|\, Y_1 = t\right\}$$

and suppose that

$$\delta^2 := \mathbf{E}\,\varphi^2(Y_1) > 0\,. \tag{1.6.10}$$

Usually U-statistics satisfying (1.6.10) are said to be *nondegenerate.*

Theorem 1.6.4. *Let* (1.6.10) *and the conditions*

$$\mathbf{E}\,\mathbf{\Phi}(Y_1, \dots, Y_m) = 0\,, \qquad \mathbf{E}\,\exp\left\{t\,\mathbf{\Phi}^2(Y_1, \dots, Y_m)\right\} < \infty \quad \textit{for any} \quad t \in \mathbf{R}^1\,,$$

and

$$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}\Big(\sum_{i=1}^n \varphi(Y_i) > n\,\varepsilon_0\Big) < \infty \qquad \textit{for some} \qquad \varepsilon_0 > 0$$

be fulfilled. Then for a sequence $\{\gamma_n\}$ *such that* $\gamma_n \to 0$ *as* $n \to \infty$ *we have*

$$\mathbf{P}\Big(\frac{U_n}{2\,\delta} \ge \varepsilon + \gamma_n\Big) = \mathbf{P}\Big(\sum_{i=1}^n \varphi(Y_i) > n\,\varepsilon\Big)\big(1 + o(1)\big)\,.$$

Combining Theorems 1.6.4 and 1.6.2 we come to

Corollary. *Under conditions of Theorems* 1.6.4 *and* 1.6.2 *with respect to the random variables* $\varphi(Y_i)$, $i = 1, \dots, n$, *we have*

$$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}\big(U_n \ge a + \gamma_n\big) = -\frac{a^2}{8\,\delta^2}\big(1 + o(1)\big) \qquad \text{as} \qquad a \to 0\,. \tag{1.6.11}$$

Some variants of such results were described in Korolyuk and Borovskikh (1993).

Numerous nonparametric statistics could not be represented in a form of U-statistics but in a form of von Mises functionals, intimately connected with them. Remind that these r.v.'s look as

$$V_n := n^{-m} \sum_{i_1=1}^n \cdots \sum_{i_m=1}^n \mathbf{\Phi}\big(Y_{i_1}, \dots, Y_{i_m}\big)$$

where the kernel $\mathbf{\Phi}$ is defined as above. Consider U-statistics

$$U_{nk} := \binom{n}{k}^{-1} \sum_{1\le i_1<\cdots<i_k\le n} \mathbf{\Phi}_{mk}\big(Y_{i_1}, \dots, Y_{i_k}\big)\,, \qquad\qquad k = 1,\, 2,\, \dots,\, m\,,$$

with the symmetric kernels

$$\boldsymbol{\Phi}_{mk}(x_1, \dots, x_k) := \sum_{\substack{\nu_1+\cdots+\nu_k=m \\ \nu_j \geq 1}} \frac{m\,!}{\nu_1\,!\ \cdots\ \nu_k\,!}\ \boldsymbol{\Phi}\big(\mathbf{x}_1^{\nu_1}, \dots, \mathbf{x}_k^{\nu_k}\big)$$

where $\mathbf{x}_j^{\nu_j} = (x_j, \dots, x_j)$ is a vector with ν_j identical coordinates.

The following Bönner–Kirschner (1977) representation is valid:

$$V_n = n^{-m} \sum_{k=1}^{m} \binom{n}{k} U_{nk}\,. \tag{1.6.12}$$

This formula enables one in many cases to reduce large deviations for von Mises functionals to large deviations of U-statistics. For example, if the kernels of U-statistics U_{nk} are bounded then the main asymptotic contribution to (1.6.12) belongs to the summand $n^{-m}\binom{n}{m} U_{nm}$ whereas the others may be neglected. To study the large deviation asymptotics of the random variables $n^{-m}\binom{n}{m} U_{nm}$ one may apply Theorem 1.6.4.

The Sanov theorem presents another fundamental result in the theory of large deviations. Let again $\{Y_j\}$ be a sequence of i.i.d. random variables with the common d.f. F_0 and

$$F_n(t) := n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{Y_i < t\}} \tag{1.6.13}$$

be the empirical d.f. based on the sample $Y_1, Y_2, \dots, Y_n$. The main result by Sanov (1957) is as follows: *for a sufficiently regular set* $\boldsymbol{\Omega}$ *in the space of d.f.'s on the real line* (F_0*-distinguishable* in his terminology)

$$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}\,(F_n \in \boldsymbol{\Omega}) = -\inf \left\{ \int_{-\infty}^{\infty} \ln \frac{dF}{dF_0}\ dF\ :\ F \in \boldsymbol{\Omega} \right\}. \tag{1.6.14}$$

The integral standing in the right-hand side coincides with the Kullback–Leibler information introduced in (1.2.9).

As it is hard to verify the condition of F_0-distinguishability, numerous papers have been connected with sufficient conditions for validity of (1.6.14). We present now some results in this field.

Let us denote, in agreement with (1.2.10), the right-hand side of (1.6.14) by $K(\mathbf{\Omega}, F_0)$. Let $\mathcal{E}$ be the set of d.f.'s g on $\mathbf{R}^1$ being absolutely continuous with respect to F_0 and possessing such property that *there exists a finite division in intervals* $\Delta_1, \ldots, \Delta_n$ *of the real line on which the derivative* dg/dF_0 *is monotone and bounded.* We denote by ρ the uniform metric in the space $\mathbf{\Lambda}_1$ of d.f.'s on $\mathbf{R}^1$ and by Cl_ρ the closure in this metric.

Theorem 1.6.5 (Borovkov (1967)). *Suppose* F_0 *is a continuous distribution function and* $\mathbf{\Omega}$ *is a* ρ*-open set in* $\mathbf{\Lambda}_1$ *such that*

$$K\big(\mathbf{\Omega} \cap \mathcal{E}, F_0\big) = K\big(\mathrm{Cl}_\rho(\mathbf{\Omega}) \cap \mathcal{E}, F_0\big).$$

Then equality (1.6.14) *is valid.*

When studying large deviations of concrete statistics they may be often represented as functionals $T(F_n)$ of empirical d.f.'s. In this connection let us introduce for any $r \in \mathbf{R}^1$ the sets

$$\mathbf{\Omega}_r^T := \big\{F \in \Lambda_1 \colon T(F) \geq r\big\}.$$

Theorem 1.6.6 (Hoadley (1967). *Let* F_0 *be continuous distribution function,* $T(F)$ *be a uniformly continuous in the* ρ*-topology functional and* $\{u_n\}$ *be a real sequence such that* $u_n \to 0$ *as* $n \to \infty$*. If the function* $t \mapsto K(\mathbf{\Omega}_t^T, F_0)$ *is continuous at the point* $t = a$ *then*

$$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}\big(T(F_n) \geq a + u_n\big) = -K(\mathbf{\Omega}_a^T, F_0). \tag{1.6.15}$$

Hoadley (1967) also presented the generalization of this theorem for several independent samples.

Later Groeneboom, Oosterhoff and Ruymgaart (1979) gave generalizations of Theorems 1.6.6 and 1.6.5 simultaneously in three directions: they considered the space of measures on a Hausdorff space $\mathbf{S}$ instead of the space of d.f.'s on the real line; the uniform continuity of the functional $T(F)$ was replaced by continuity in some suitable topology; finally, they showed that the distribution of the initial sample could be not only continuous, but also atomic.

Let $\mathbf{S}$ be a Hausdorff space, $\mathfrak{B}$ be the Borel σ-algebra of sets in $\mathbf{S}$ and $\mathbf{\Lambda}$ be the space of all probability measures on $\mathfrak{B}$. Define for any $P,\, Q \in \mathbf{\Lambda}$, in agreement with (1.2.9),

$$K(Q,P) := \begin{cases} \int\limits_{\mathbf{S}} \ln \dfrac{d\,Q}{d\,P}\, d\,Q\,, & \text{if } Q \ll P\,, \\ +\infty\,, & \text{otherwise}\,, \end{cases} \tag{1.6.16}$$

and put for any subset $\mathbf{\Omega}$ of $\mathbf{\Lambda}$

$$K(\mathbf{\Omega},P) := \inf\big\{K(Q,P)\colon\ Q \in \mathbf{\Omega}\big\}\,, \tag{1.6.17}$$

keeping in mind that $K(\varnothing,P) = +\infty$.

Let $\{Y_j\}$ be a sequence of i. i. d. random variables with values in $\mathbf{S}$ and having the distribution $P \in \mathbf{\Lambda}$. Denote by $\mathcal{P}_n$ the empirical distribution based on $Y_1,\, Y_2,\, \dots\,,\, Y_n$, i.e., for any $B \in \mathfrak{B}$ put

$$\mathcal{P}_n(B) := n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \in B\}}\,.$$

Consider, at last, in the space $\mathbf{\Lambda}$ the topology τ of convergence on all Borel sets, i.e., the "coarsest" topology for which the application $Q \mapsto Q(B),\ Q \in \mathbf{\Lambda}$, is continuous for all $B \in \mathfrak{B}$.

The main result by Groeneboom at al. (1979) is as follows.

Theorem 1.6.7. *Let $P \in \mathbf{\Lambda}$ and $T\colon\ \mathbf{\Lambda} \mapsto \bar{\mathbf{R}}^1$ be a τ-continuous functional on each*

$$Q \in \mathbf{\Gamma} := \big\{R \in \mathbf{\Lambda}\colon\ K(R,P) < \infty\big\}.$$

If the function $t \mapsto K\big(\mathbf{\Omega}_t^T,P\big)$ where $t \in \mathbf{R}^1$ and $\mathbf{\Omega}_t^T := \big\{Q \in \mathbf{\Lambda}\colon\ T(Q) \ge t\big\}$ is right-continuous at the point $t = r$, then

$$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}\,\big(\,T(\mathcal{P}_n) \ge r + u_n\,\big) = -K\big(\mathbf{\Omega}_r^T,P\big) \tag{1.6.18}$$

for a real sequence $\{u_n\}$ such that $u_n \to 0$ as $n \to \infty$.

To compute the Hodges–Lehmann ARE one needs the asymptotics for the probability that $\mathcal{P}_n$ hits the sets of the form

$$\mathbf{\Delta}_z^T := \big\{\,Q \in \mathbf{\Lambda}:\ T(Q) \le z\big\}\,.$$

But for many important examples the function $z \mapsto K\big(\mathbf{\Delta}_z^T, P\big)$ is right-continuous, due to Lemma 3.3 by Groeneboom et al. (1979), and is not left-continuous. This property does not allow one to obtain any analogue of (1.6.18) for $\mathbf{\Delta}_z^T$. In this connection we formulate here an asymptotic inequality that may be drawn easily from Lemmas 2.4 and 3.1 by Groeneboom et al. (1979).

Theorem 1.6.8. *Let* $P \in \mathbf{\Lambda}$, $\mathbf{\Omega} \subset \mathbf{\Lambda}$ *and*

$$K(\mathbf{\Omega}, P) = K\big(\mathrm{Cl}_\tau(\mathbf{\Omega}), P\big). \tag{1.6.19}$$

Then

$$\varlimsup_{n\to\infty} n^{-1} \ln \mathbf{P}\big(\mathcal{P}_n \in \mathbf{\Omega}\big) \le -K\big(\mathbf{\Omega}, P\big). \tag{1.6.20}$$

It is well-known (see Groeneboom et al. (1979)) that the τ-topology is stronger than the ρ-topology. Consequently if a functional T is ρ-continuous then the corresponding set $\mathbf{\Delta}_z^T$ is ρ-closed and a fortiori τ-closed. This enables one to apply Theorem 1.6.8 as most functionals encountered in nonparametric statistics are usually ρ-continuous.

Other conditions supporting the validity of Sanov's theorem may be found in Stone (1974), Sievers (1976) and Fu (1985).

Ermakov (1990) formulated a "uniform" variant of Theorem 1.6.7. Put for any sets $\mathbf{\Omega}$, $\mathbf{\Pi} \subseteq \mathbf{\Lambda}$

$$K\big(\mathbf{\Omega}, \mathbf{\Pi}\big) := \inf\big\{K(Q,P)\colon\ Q \in \mathbf{\Omega},\ P \in \mathbf{\Pi}\big\}.$$

Theorem 1.6.9. *Assume* $\mathbf{\Pi}$ *is compact in the* τ*-topology. If*

$$K\big(\mathrm{Int}_\tau(\mathbf{\Omega}), P\big) = K\big(\mathrm{Cl}_\tau(\mathbf{\Omega}), P\big) \qquad \text{for all} \qquad P \in \mathbf{\Pi},$$

then

$$\lim_{n\to\infty} \sup_{P\in\mathbf{\Pi}} \Big| n^{-1} \ln \mathbf{P}\big(\mathcal{P}_n \in \Omega\big) + K(\Omega, P) \Big| = 0.$$

If, however,

$$K\big(\mathrm{Int}_\tau(\mathbf{\Omega}), \mathbf{\Pi}\big) = K\big(\mathrm{Cl}_\tau(\mathbf{\Omega}), \mathbf{\Pi}\big)$$

then

$$\lim_{n\to\infty} \sup_{P\in\mathbf{\Pi}} n^{-1} \ln \mathbf{P}\left(\mathcal{P}_n \in \mathbf{\Omega}\right) = -K(\mathbf{\Omega}, P)\,.$$

This result was applied by Ermakov (1990) to the analysis of asymptotic minimaxity of some nonparametric tests.

The theorems on large deviations for the Markov processes with close formulations are presented in the book of Wentzell (see Wentzell (1990), Sec. 4.3). However they do not imply the Sanov theorem and its generalizations.

We emphasize also that there exists a very close connection between the results of the Chernoff and Sanov types allowing to derive one from the other. (See Bahadur and Zabell (1979), Bretagnolle (1979), Groeneboom et al. (1979), Azencott (1980), Fu (1985), Deuschel and Stroock (1989).)

Inglot and Ledwina (1990) proposed another method to obtain the information on asymptotics of large deviation probabilities based on the idea of the strong approximation.

Let $G_n(t)$ be the empirical d.f. based on a sample of size n from the uniform distribution on $[0,\,1]$ and

$$\alpha_n(t) = \sqrt{n}\left(G_n(t) - t\right), \qquad t \in [0,\,1]\,,$$

be the usual empirical process. The well-known result by Komlós, Major and Tusnády (1975, 1976) is as follows:

there exist a probability space and the sequences of processes $\{\alpha_n^*\} \overset{\mathbf{D}}{=} \{\alpha_n\}$ *and Brownian bridges* $\{B_n\}$ *defined on this probability space such that the following inequality*

$$\mathbf{P}\Big(\sup_{0\le t\le 1} \big|\,\alpha_n^*(t) - B_n(t)\,\big| > n^{-1/2}\left(C \ln n + x\right)\Big) \le L\, e^{-lx} \tag{1.6.21}$$

holds for all n *and* x *with some positive constants* C, L *and* l.

Consider now a sequence of statistics $\{T_n\}$ representable in the form $T_n = T(\alpha_n^*)$ where T: $\mathbf{D}\,[0,1] \mapsto \bar{\mathbf{R}}^{\,1}$ is some functional satisfying the following conditions:

there exists a constant c *such that*

$$\big|\,T(\alpha_n^*) - T(B_n)\,\big| \le c \sup_{0\le t\le 1} \big|\,\alpha_n^*(t) - B_n(t)\,\big| \qquad \text{a.s.}\,, \tag{1.6.22}$$

(1.6.23) $$\ln \mathbf{P}\left(T(B_1) \geq y\right) = -\tfrac{1}{2}\, a\, y^2 \left(1 + o(1)\right), \qquad y \to \infty,$$

for some constant $a > 0$.

Theorem 1.6.10 (Inglot and Ledwina (1990)). *Suppose the conditions* (1.6.22) *and* (1.6.23) *are fulfilled. Then for any* $d > 1$, $K = ac/(2l)$ *and any* $x \in \left(0, (\sqrt{d} - 1)/(dK)\right)$ *the following inequalities are valid:*

(1.6.24) $$\begin{aligned} -\tfrac{1}{2}\, a\, x^2 (1 + d\, K\, x)^2 &\leq \varliminf_{n\to\infty} n^{-1} \ln \mathbf{P}\left(T(\alpha_n) \geq x\sqrt{n}\right) \\ &\leq \varlimsup_{n\to\infty} n^{-1} \ln \mathbf{P}\left(T(\alpha_n) \geq x\sqrt{n}\right) \\ &\leq -\tfrac{1}{2}\, a\, x^2 (1 - K\, x)^2 . \end{aligned}$$

Proof. Denote
$$H_n := \sup_{0 \leq t \leq 1} \left|\alpha_n^*(t) - B_n(t)\right| .$$
Using (1.6.21)–(1.6.23) we obtain

$$\begin{aligned} \mathbf{P}\left(T(\alpha_n) \geq x\sqrt{n}\right) &\leq \mathbf{P}\left(T(B_n) \geq x\,(1 - K\, x)\sqrt{n}\right) + \mathbf{P}\left(H_n \geq c^{-1} K\, x^2 \sqrt{n}\right) \\ &\leq \exp\left\{-\tfrac{1}{2}\, a\, x^2 (1 - K\, x)^2 n \left(1 + o(1)\right)\right\} \\ &\quad + L \exp\left\{-l\left(c^{-1} K\, x^2 n - C \ln n\right)\right\} . \end{aligned}$$

The lower estimate may be found in the same way:

$$\begin{aligned} \mathbf{P}\left(T(\alpha_n) \geq x\sqrt{n}\right) &\geq \mathbf{P}\left(T(B_n) \geq x\,(1 + d\, K\, x)\sqrt{n}\right) - \mathbf{P}\left(H_n \geq c^{-1} d\, K\, x^2 \sqrt{n}\right) \\ &\geq \exp\left\{-\tfrac{1}{2}\, a\, x^2 (1 + d\, K\, x)^2 n \left(1 + o(1)\right)\right\} \\ &\quad - L \exp\left\{-l\left(c^{-1} d\, K\, x^2 n - C \ln n\right)\right\} . \end{aligned}$$

Taking logarithms, using the definition of the constant K and passing to the limit we get (1.6.24).

If for one reason or another it is known that the usual limit

$$I(x) := \lim_{n\to\infty} n^{-1} \ln \mathbf{P}\left(T(\alpha_n) \geq x\sqrt{n}\right)$$

exists then (1.6.24) yields

(1.6.25) $$\lim_{x\to 0} \lim_{n\to\infty} (n\, x^2)^{-1} \ln \mathbf{P}\left(T(\alpha_n) \geq x\sqrt{n}\right) = -\tfrac{1}{2}\, a .$$

Formula (1.6.25) gives the useful information about the leading term of large deviation asymptotics. It is important that the problem of the calculation of the constant a is connected with the well-studied problem of finding the tail asymptotics for the distributions of Brownian bridge functionals (see Fernique (1971), Marcus and Shepp (1972), Borell (1975), Kallianpur and Oodaira (1978), Aki and Kashiwagi (1990)).

It should be emphasized, however, that the proof of the existence of the mentioned above limit $I(x)$ and its continuity in x that is important for the calculation of the Bahadur efficiency (cf. Theorem 1.2.2) is a serious problem itself. Therefore the possibilities of the method under discussion are seriously limited when studying the Chernoff's type deviations.

On the contrary this method proved his utility for the study of Cramér's type and moderate deviations (Inglot, Kallenberg and Ledwina (1992), Inglot and Ledwina (1993)) where the problem of existence of this limit is unessential.

In conclusion we present some results on large deviations under null-hypothesis of linear rank statistics. Let $R_1, R_2, \dots, R_n$ be the ranks of n random variables $Z_1, Z_2, \dots, Z_n$, i.e., R_i be the number of Z_i in the variational series based on $Z_1, Z_2, \dots, Z_n$. Consider as the null-hypothesis that $(R_1, R_2, \dots, R_n)$ is equally likely to be any of $n!$ permutations of $(1, 2, \dots, n)$. This supposition is in agreement with special hypotheses considered in the sequel. A linear rank statistic is one of the form

$$T_n = \sum_{i=1}^{n} J_n\left(\frac{R_i}{n+1}\, , \, \frac{i}{n+1} \right) \tag{1.6.26}$$

where $J_n(u,v)$ is a function on a unit square. Let

$$\mathcal{H} := \left\{ h\colon\ h \ge 0,\ \int_0^1 h(u,v)\,du = \int_0^1 h(u,v)\,dv = 1 \right\}$$

be the set of all bivariate densities on a unit square with uniform marginals. It will be assumed that the sequence of functions $\{J_n\}$ satisfies Woodworth's (1970) property **A**, i.e.,

1) *for each* n J_n *is constant on the rectangles*

$$\left\{ i-1 \le n\,u < i,\ j-1 \le n\,v < j \right\}, \qquad 1 \le i, j \le n\,;$$

2) *there exists a function* J *on a unit square such that*

$$\sup_{h\in\mathcal{H}} \left\{ \Big| \int_0^1\!\!\int_0^1 (J_n - J)\, h\, du\, dv \Big| :\ h \in \mathcal{H} \right\} \longrightarrow 0\,, \qquad n \to \infty\,.$$

Woodworth (1970) noted that property A is fulfilled for almost all known linear rank statistics. There were given also simple sufficient conditions for its validity.

Without loss of generality we may assume

$$\int_0^1\!\!\int_0^1 J(u,v)\, du\, dv = 0\,. \tag{1.6.27}$$

First papers on large deviations of linear rank statistics belong to Klotz (1965), Stone (1967, 1968) and Hoadley (1967), who considered only most known representatives of this class: the Wilcoxon and normal scores statistics. The first general result had been obtained by Woodworth (1970).

Theorem 1.6.11. *Let* $\{T_n\}$ *be a sequence of linear rank statistics* (1.6.26) *satisfying property* A *and* $\{\varepsilon_n\}$ *be a sequence of real numbers such that* $\varepsilon_n \to \varepsilon$ *as* $n \to \infty$. *Then for* $0 < \varepsilon < \varepsilon(J)$

$$\lim_{n\to\infty} n^{-1} \ln \mathbf{P}\big(T_n \ge n\, \varepsilon_n\big) = -I(\varepsilon, J) \tag{1.6.28}$$

where

$$\varepsilon(J) := \sup\left\{ \int_0^1\!\!\int_0^1 J\, h\, du\, dv:\ h \in \mathcal{H} \right\},$$

$$I(\varepsilon, J) := \inf\left\{ \int_0^1\!\!\int_0^1 h\, \ln h\, du\, dv:\ \int_0^1\!\!\int_0^1 J\, h\, du\, dv \ge \varepsilon,\ h \in \mathcal{H} \right\}, \tag{1.6.29}$$

and the function $\varepsilon \mapsto I(\varepsilon, J)$ *is continuous for sufficiently small* $\varepsilon > 0$.

The further efforts have been directed to the computation of $I(\varepsilon, J)$ for small ε and under appropriate conditions imposed on a score function J. The papers of Woodworth (1970), Hwang (1976), Kremer (1979a, 1981, 1982) and Ledwina (1987) have been connected with this problem. The most general result belongs to Kallenberg and Ledwina (1987).

Theorem 1.6.12. *Let*

$$\widetilde{J}(u,v) := J(u,v) - \int_0^1 J(u,y)\,dy - \int_0^1 J(x,v)\,dx + \int_0^1\!\!\int_0^1 J(u,v)\,du\,dv\,,$$

$$b^2 := \int_0^1\!\!\int_0^1 \widetilde{J}^{\,2}(u,v)\,du\,dv\,.$$

Suppose that $b > 0$ *and for some* $r > 0$

$$\int_0^1\!\!\int_0^1 \exp\left\{r\,\widetilde{J}(u,v)\right\} du\,dv \;<\; \infty\,. \tag{1.6.30}$$

Then

$$I(\varepsilon, J) \;=\; \frac{\varepsilon^2}{2\,b^2} \;+\; o(\varepsilon^2) \qquad \text{as} \qquad \varepsilon \to 0\,.$$

We note that condition (1.6.30) is much less restrictive than the conditions in Woodworth (1970), Kremer (1979a, 1981) and permits us to use unbounded score functions.

• 1.7 On Large Deviations of Empirical Measures in the Case of Several Independent Samples

Many problems of statistics are in natural way connected with the simultaneous use of several independent samples. Statistical tests proposed for the solution of these problems (two-sample tests of homogeneity or, say, multisample tests of symmetry) are usually based on functionals of corresponding empirical measures. This Section contains some results on large deviations of such functionals from Groeneboom et al. (1979) and Nikitin (1987a). We preserve the notations introduced in Section 1.6.

Denote by $\mathbf{\Lambda}^c$ the c-fold Cartesian product of spaces $\mathbf{\Lambda}$ (c is arbitrary positive integer). The space $\mathbf{\Lambda}^c$ is endowed with the product-topology generated by the τ-topology. Put for brevity

$$\mathbb{P} := (P_1,\, P_2,\, \dots\,,\, P_c) \in \mathbf{\Lambda}^c\,, \qquad\qquad \boldsymbol{\rho} := (\rho_1,\, \rho_2,\, \dots\,,\, \rho_c)\,,$$

where $0 < \rho_i < 1$ and $\sum_{i=1}^{c} \rho_i = 1$. Denote also for any $\mathbb{P} = (P_1, P_2, \ldots, P_c)$ and $\mathbb{Q} = (Q_1, Q_2, \ldots, Q_c)$ from $\mathbf{\Lambda}^c$ and any $\mathbf{\Omega} \subset \mathbf{\Lambda}^c$

$$J_{\boldsymbol{\rho}}(\mathbb{Q}, \mathbb{P}) := \sum_{i=1}^{c} \rho_i K(Q_i, P_i), \tag{1.7.1}$$

$$J_{\boldsymbol{\rho}}(\mathbf{\Omega}, \mathbb{P}) := \inf \Big\{ J_{\boldsymbol{\rho}}(\mathbb{Q}, \mathbb{P}) \colon \ \mathbb{Q} \in \mathbf{\Omega} \Big\}. \tag{1.7.2}$$

Consider now $c \geq 1$ independent samples $Y_{i,1}, \ldots, Y_{i,n_i}$, $1 \leq i \leq c$, taking on values in $\mathbf{S}$ and having distributions $P_1, \ldots, P_c$. Assume that sample sizes n_i tend to infinity in such a way that

$$\lim_{N \to \infty} \frac{n_i}{N} = \rho_i > 0 \tag{1.7.3}$$

where $N = n_1 + \cdots + n_c$. Let $P_{1,n_1}, \ldots, P_{c,n_c}$ be the empirical measures based on these samples and put for brevity

$$\mathcal{P}_N := (P_{1,n_1}, \ldots, P_{c,n_c}).$$

For any functional $T \colon \mathbf{\Lambda}^c \mapsto \bar{\mathbf{R}}^1$ and any $z \in \mathbf{R}^1$ consider the set

$$\mathbf{\Omega}_z^T := \Big\{ \mathbb{Q} \in \mathbf{\Lambda}^c \colon \ T(\mathbb{Q}) \geq z \Big\}. \tag{1.7.4}$$

Theorem 1.7.1 (Groeneboom et al. (1979)). *Let $\mathbb{P} \in \mathbf{\Lambda}^c$ and $T \colon \mathbf{\Lambda}^c \mapsto \bar{\mathbf{R}}^1$ be a τ-continuous functional on each $\mathbb{Q} \in \Gamma := \big\{ \mathbb{R} \in \mathbf{\Lambda}^c \colon \ J_{\boldsymbol{\rho}}(\mathbb{R}, \mathbb{P}) < \infty \big\}$. If the function $z \mapsto J\big(\mathbf{\Omega}_z^T, \mathbb{P}\big)$ is right-continuous at the point $z = a$ and $\{u_N\}$ is a real sequence, such that $u_N \to 0$ as $N \to \infty$, then*

$$\lim_{N \to \infty} N^{-1} \ln \mathbf{P}\big(T(\mathcal{P}_N) \geq a + u_N \big) = -J_{\boldsymbol{\rho}}(\mathbf{\Omega}_a^T, \mathbb{P}).$$

As in the case of one sample, the τ-continuity follows from the ρ-continuity that may be verified in a much easier way and takes place for the majority of functionals encountered in nonparametric statistics.

To calculate the Hodges–Lehmann ARE of a number of nonparametric statistics we are interested of special sets of the form

$$\mathbf{\Delta}_z^T := \Big\{ \mathbb{Q} \in \mathbf{\Lambda}^c \colon \ T(\mathbb{Q}) \leq z \Big\} \tag{1.7.5}$$

where T is a functional on $\mathbf{\Lambda}^c$ and $z \in \mathbf{R}^1$. The attempt to obtain for such sets the precise analogue of Theorem 1.7.1 fails as the function $z \mapsto J_{\boldsymbol{\rho}}(\mathbf{\Delta}_z^T, \mathbb{P})$ is not obliged to be left-continuous (though the right-continuity takes place and may be proved in the same way as in Lemma 3.3 from Groeneboom et al. (1979)).

In this connection we formulate and outline the proof of the c-sample variant of Theorem 1.6.8.

Theorem 1.7.2 (Nikitin (1987a)). *Let* $\mathbb{P} \in \mathbf{\Lambda}^c$, $\mathbf{\Omega} \subset \mathbf{\Lambda}^c$ *and*

$$J_{\boldsymbol{\rho}}(\mathbf{\Omega}, \mathbb{P}) = J_{\boldsymbol{\rho}}(\mathrm{Cl}_{\tau}(\mathbf{\Omega}), \mathbb{P}) \,. \tag{1.7.6}$$

Then the estimate

$$\overline{\lim_{N\to\infty}}\, N^{-1} \ln \mathbf{P}\big\{ \mathcal{P}_N \in \mathbf{\Omega} \big\} \le -J_{\boldsymbol{\rho}}(\mathbf{\Omega}, \mathbb{P}) \tag{1.7.7}$$

is valid.

The proof of Theorem 1.7.2 basically repeats the arguments of Groeneboom et al. (1979) and we give only the main stages of the proof.

For any $\mathbb{P}, \mathbb{Q} \in \mathbf{\Lambda}$ and any finite partition $\boldsymbol{\pi}$ of the space $\mathbf{S}$ consisting of the Borel sets $B_1, B_2, \ldots, B_m$ put

$$K_{\boldsymbol{\pi}}(\mathbb{Q}, \mathbb{P}) := \sum_{j=1}^{m} Q(B_j) \ln \frac{Q(B_j)}{P(B_j)} \,.$$

For c partitions $\pi_1, \pi_2, \ldots, \pi_c$ of $\mathbf{S}$ define by $\boldsymbol{\pi} := \pi_1 \times \cdots \times \pi_c$ a decomposition of $\mathbf{S}^c$ consisting of direct products of elements of corresponding partitions. Denote for any $\mathbb{P} = (P_1, P_2, \ldots, P_c)$ and $\mathbb{Q} = (Q_1, Q_2, \ldots, Q_c)$ from $\mathbf{\Lambda}^c$

$$J_{\boldsymbol{\rho}, \boldsymbol{\pi}}(\mathbb{Q}, \mathbb{P}) := \sum_{i=1}^{c} K_{\pi_i}(Q_i, P_i) \,.$$

Lemma 1.7.1. *The function* $\mathbb{Q} \mapsto J_{\boldsymbol{\rho}}(\mathbb{Q}, \mathbb{P})$ *is* τ*-lower semicontinuous on* $\mathbf{\Lambda}^c$.

This lemma is a generalization of Lemma 2.2 from Groeneboom et al. (1979) and has been proved by Nikitin (1987a).

Lemma 1.7.2. *Let* $\mathbb{P} \in \mathbf{\Lambda}^c$ *and* $\Gamma := \big\{\mathbb{Q} \in \mathbf{\Lambda}^c \colon J_{\boldsymbol{\rho}}(\mathbb{Q}, \mathbb{P}) \le \alpha\big\}$ *for some finite constant* α. *The set* Γ *is compact and sequentially compact in the product-topology* τ *on* $\mathbf{\Lambda}^c$.

This lemma generalizes again the corresponding statement by Groeneboom et al. (1979), namely Lemma 2.3; the proof has been given by Nikitin (1987a).

Lemma 1.7.3. *Let* $\boldsymbol{\pi}$ *be an arbitrary partition of* $\mathbf{S}^c$. *Under the conditions of Theorem* 1.7.2 *we have*

$$J_{\boldsymbol{\rho}}\left(\boldsymbol{\Omega},\mathbb{P}\right)=\sup_{\text{all }\boldsymbol{\pi}}\ J_{\boldsymbol{\rho},\boldsymbol{\pi}}\left(\boldsymbol{\Omega},\mathbb{P}\right). \tag{1.7.8}$$

The *proof* relies on Lemma 1.7.2 and has no essential changes in the comparison with the proof of Lemma 2.4 by Groeneboom et al. (1979).

Lemma 1.7.4. *Let conclusion* (1.7.8) *of Lemma* 1.7.3 *be fulfilled, then*

$$\overline{\lim_{N\to\infty}}\ N^{-1}\ln\mathbf{P}\left\{\mathcal{P}_N\in\boldsymbol{\Omega}\right\}\leq -J_{\boldsymbol{\rho}}\left(\boldsymbol{\Omega},\mathbb{P}\right). \tag{1.7.9}$$

This lemma generalizes Lemma 3.1 from Groeneboom et al. (1979) and uses at the proof the similar technique based on the asymptotical analysis of polynomial distribution.

Combining Lemmas 1.7.3 and 1.7.4 we establish the conclusion of Theorem 1.7.2.

Let us quote some more auxiliary statements from this class of problems necessary for the sequel.

Lemma 1.7.5. *Let* $\boldsymbol{\Omega}$ *be a nonempty* τ*-closed set of probability measures in* $\boldsymbol{\Lambda}^c$ *and* $\mathbb{P}\in\boldsymbol{\Lambda}^c$. *Then there exists such* $\mathbb{Q}\in\boldsymbol{\Omega}$ *that*

$$J_{\boldsymbol{\rho}}\left(\mathbb{Q},\mathbb{P}\right)=J_{\boldsymbol{\rho}}\left(\boldsymbol{\Omega},\mathbb{P}\right).$$

The *proof* is similar to the proof of Lemma 3.2 from Groeneboom et al. (1979) and is based on Lemmas 1.7.1 and 1.7.2.

Lemma 1.7.6. *Let* $\mathbb{P}\in\boldsymbol{\Lambda}^c$ *and* T *be a* τ*-continuous functional on* $\boldsymbol{\Lambda}^c$. *Consider for any real* z *the sets* $\boldsymbol{\Omega}_z^T$ *and* $\boldsymbol{\Delta}_z^T$ *introduced by* (1.7.4) *and* (1.7.5). *Then the function* $z\mapsto J_{\boldsymbol{\rho}}\left(\boldsymbol{\Omega}_z^T,\mathbb{P}\right)$ *is left-continuous and the function* $z\mapsto J_{\boldsymbol{\rho}}\left(\boldsymbol{\Delta}_z^T,\mathbb{P}\right)$ *is right-continuous.*

The *proof* is again analogous to the proof of Lemma 3.3 from Groeneboom et al. (1979), but one should use Lemmas 1.7.1–1.7.5 instead of those results that these lemmas generalize.

• 1.8 Some Results from Theory of Extremal Problems and Theory of Implicit Operators

As we have seen in Sections 1.5–1.7 studying the large deviation probabilities entails extremal problems connected with the minimization of the Kulback–Leibler information on the sets of special type that depend on the structure of the sequences of statistics under consideration. We present below some results being useful by solution of such problems in subsequent chapters.

Let $\mathcal{X}$ and $\mathcal{Y}$ be normed spaces, f_i, $i = 0, 1, \ldots, m$, be smooth functions on $\mathcal{X}$ and F be a smooth application of $\mathcal{X}$ into $\mathcal{Y}$. Consider the following extremal problem:

$$\begin{aligned} &f_0(x) \to \text{extr}\,, \\ &F(x) = 0\,, \\ &f_i(x) \gtrless 0, \qquad i = 1, \ldots, m\,, \end{aligned} \tag{1.8.1}$$

where the symbol $f_i(x) \gtrless 0$ is meaning that the ith limitation looks as $f_i(x) = 0$, as $f_i(x) \geq 0$ or as $f_i(x) \leq 0$.

The problems of type (1.8.1) are usually called *smooth extremal problems* with the limitations of the equality and inequality type (Ioffe and Tikhomirov (1979), Alekseev et al. (1979)). For such problems the Lagrange multiplier rule is true. Let us call the *Lagrange function* of problem (1.8.1) the function

$$\mathcal{L}(x, y^*, \boldsymbol{\lambda}, \lambda_0) := \sum_{i=1}^{m} \lambda_i\, f_i(x) + \langle\, y^*, F(x)\rangle$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m) \in \mathbf{R}^m$, $\lambda_0 \in \mathbf{R}^1$, $y^* \in \mathcal{Y}^*$ and the symbol $\langle y^*, F(x)\rangle$ signifies the value of the functional y^* on the element $F(x) \in \mathcal{Y}$ and $\mathcal{Y}^*$ is the dual space with respect to $\mathcal{Y}$.

Theorem 1.8.1 (Alexeev et al. (1979), Sec. 3.2). *Let $\mathcal{X}$ and $\mathcal{Y}$ be the Banach spaces, $\mathcal{U}$ be an open set in $\mathcal{X}$, the functions f_i: $\mathcal{U} \mapsto \mathbf{R}^1$, $i = 0, 1, \ldots, m$, and the application F: $\mathcal{U} \mapsto \mathcal{Y}$ be strictly differentiable at the point $\hat{x}$. If $\hat{x}$ is a local extremum in problem* (1.8.1) *and if the image $\Im F'(\hat{x})$ is a closed subspace in $\mathcal{Y}$,* *then there exist such Lagrange multipliers* $\hat{y}^*$, $\hat{\boldsymbol{\lambda}}$, $\hat{\lambda}_0$, *not equal to* 0 *simultaneously, for which hold:*

a) *the condition of stationarity of the Lagrange function in* x*:*

$$\mathcal{L}_x(\hat{x}, \hat{y}^*, \hat{\boldsymbol{\lambda}}, \hat{\lambda}_0) = 0\,; \tag{1.8.2}$$

b) *the condition of co-ordination of signs:* $\hat{\lambda}_0 \ge 0$ *if one considers the problem on minimum,* $\hat{\lambda}_0 \le 0$, *if the problem is on maximum,*

$$\hat{\lambda}_i \gtrless 0\,, \qquad i = 1, 2, \ldots, m\,; \tag{1.8.3}$$

c) *the complementary condition:*

$$\hat{\lambda}_i\, f(\hat{x}) = 0\,, \qquad i = 1, 2, \ldots, m\,. \tag{1.8.4}$$

Conditions (1.8.3) mean that if in (1.8.1) $f_i(x) \ge 0$ then $\hat{\lambda}_i \le 0$; if $f_i(x) \le 0$ then $\hat{\lambda}_i \ge 0$; finally if $f_i(x) = 0$ then $\hat{\lambda}_i$ may have arbitrary sign.

We remind that the application F: $\mathcal{X} \mapsto \mathcal{Y}$ is said to be *strictly differentiable* at the point $\hat{x}$ if there exists a linear operator $\boldsymbol{\Upsilon}$: $\mathcal{X} \mapsto \mathcal{Y}$ such that for each $\varepsilon > 0$ exists such $\delta > 0$ that for any x_1 and x_2 satisfying the conditions

$$\| x_1 - \hat{x} \|_{\mathcal{X}} < \delta\,, \qquad \| x_2 - \hat{x} \|_{\mathcal{X}} < \delta\,,$$

the following inequality holds:

$$\Big\| F(x_1) - F(x_2) - \boldsymbol{\Upsilon}(x_1 - x_2) \Big\|_{\mathcal{Y}} < \varepsilon \, \| x_1 - x_2 \|_{\mathcal{X}}\,. \tag{1.8.5}$$

This condition may be verified using the result given by Alexeev et al. (1979, Sec. 2.2.3).

Theorem 1.8.2. *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be the normed spaces,* $\mathcal{U}$ *be a neighborhood of* $\hat{x}$ *in* $\mathcal{X}$ *and the application* F: $\mathcal{U} \mapsto \mathcal{Y}$ *be Gâteaux differentiable in each* $x \in \mathcal{U}$*. If the application* $x \mapsto F'(x)$ *is continuous* (*in the uniform operator topology of the space* $\mathcal{L}(\mathcal{X}, \mathcal{Y})$) *at the point* $\hat{x}$*, then the application* F *is strictly differentiable in* $\hat{x}$ (*and consequently Fréchet differentiable at the same point*). (Here $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ is the usual space of linear continuous mappings of $\mathcal{X}$ into $\mathcal{Y}$.)

Let also quote three results concerning the existence of implicit functions and operators and being of essential importance in the sequel.

Theorem 1.8.3 (Vainberg and Trenogin (1974)). *Let* $F_i(y_1, \dots, y_p; x_1, \dots, x_s)$, $i = 1, 2, \dots, p$, *be real continuous functions of real arguments turning into* 0 *at the point* $M(y_1^0, \dots, y_p^0; x_1^0, \dots, x_s^0)$ *that in some ball with centre in* M *admit the expansions in convergent series in the powers of* $y_i - y_i^0$ *and* $x_k - x_k^0$, $i = 1, 2, \dots, p$, $k = 1, 2, \dots, s$. *Suppose that these series do not contain free members. Then, if the functional determinant*

$$\mathcal{J} = \frac{D(F_1, \dots, F_p)}{D(y_1, \dots, y_p)}$$

is not equal to 0 *at the point* M, *the system of equations*

$$F_i(y_1, \dots, y_p; x_1, \dots, x_s) = 0, \qquad i = 1, 2, \dots, p,$$

has the unique solution $y_i = \varphi_i(x_1, \dots, x_s)$, *satisfying the condition*

$$\varphi_i(x_1^0, \dots, x_s^0) = y_i^0,$$

and, moreover, in some ball with centre at the point $(x_1^0, \dots, x_s^0)$ *the functions* φ_i *may be expanded in convergent series in the powers* of $x_k - x_k^0$, $k = 1, 2, \dots, s$.

Now denote by $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ three Banach spaces and by $\mathcal{D}_r(x_0, \mathbf{E})$ the ball of radius r in the space $\mathbf{E}$ with centre at the point x_0. Consider the problem of finding the solutions $x = x(y)$ (y plays a role of parameter) of the operator equation

$$F(x, y) = 0, \tag{1.8.6}$$

satisfying the condition

$$x(y) = x_0, \tag{1.8.7}$$

provided that

$$F(x_0, y_0) = 0. \tag{1.8.8}$$

Theorem 1.8.4 (Vainberg and Trenogin (1974), Sec. 22.3). *Let* $F(x, y)$ *be an analytic operator in* $\mathcal{D}_r(x_0, \mathbf{E}_1) \times \mathcal{D}_\rho(y_0, \mathbf{E}_2)$ *taking on values in* $\mathbf{E}_3$ *such that the operator*

$$B = F'_x(x_0, y_0) \tag{1.8.9}$$

(*the derivative in Fréchet sense*) *has a bounded inverse operator. Then there exist positive numbers* r_1 *and* ρ_1 *such that in the ball* $\mathcal{D}_{r_1}(x_0, \mathbf{E}_1)$ *there exists a unique solution*

$$x = f(y)$$

of equation (1.8.6). *This solution defined in* $\mathcal{D}_{\rho_1}(y_0, \mathbf{E}_2)$ *is analytic in the pointed ball and satisfies condition* (1.8.7).

Theorem 1.8.5 (Alexeev et al. (1979)). *Let* $\mathcal{W}$ *be a neighborhood in* $\mathbf{E}_1 \times \mathbf{E}_2$ *and* F: $\mathcal{W} \mapsto \mathbf{E}_3$ *be an application of the class* $\mathbf{C}^1(W)$. *If condition* (1.8.8) *is fulfilled and the operator* B *given by* (1.8.9) *is invertible, then there exist such* $\varepsilon > 0$ *and* $\delta > 0$ *and such application* φ: $\mathcal{D}_\delta(y_0, \mathbf{E}_2) \mapsto \mathbf{E}_1$ *of the class* $\mathbf{C}^1\big(\mathcal{D}_\delta(y_0, \mathbf{E}_2)\big)$ *that* $\varphi(y_0) = x_0$ *and that the condition* $\| y - y_0 \|_{\mathbf{E}_2} < \delta$ *implies* $\| \varphi(y_0) - x \|_{\mathbf{E}_1} < \varepsilon$ *and* $F\big(\varphi(y), y\big) = 0$. *This application* φ *is locally unique.*

We will constantly use these results in subsequent Chapters.