Robustness Analysis in Sequential Statistical Decisions

Alexey Kharin
Belarusian State University

Abstract

The sequential statistical decision making is considered. Performance characteristics (error probabilities and expected sample sizes) for the sequential statistical decision rules (tests) are analysed. Both cases of simple and composite hypotheses are considered. Asymptotic expansions under distortions are constructed for the performance characteristics enabling robust sequential test construction.

Keywords: sequential statistical decision, test, hypothesis, distortion, robustness.

1. Introduction

In many fields of statistical methodology applications, especially in medicine, finance, quality control, problems of statistical decision making are topical (Ghosh and Sen 1991), (Jennison and Turnbull 2000). In statistical decision making, one of the most important problems is to construct decisions providing the requested accuracy on the basis of the minimal information (minimal number of observations). To solve this problem, the sequential approach (Wald 1947), (Lai 2001) is used. Within this approach, the number of observations required to provide the prescribed performance is not fixed a priori, but is considered as a random variable that depends on random observations. To find the optimal dependency is a complicated task, but this scheme of decision making requires a number of observations that is essentially less than what is required by the approach based on fixed sample sizes (see Aivazian (1959)).

The assumed hypothetical probability models used in the sequential approach are quite often distorted in practice (Huber and Ronchetti 2009), (Maevskii and Kharin 2002), (Kharin 2011), (Kharin 2005). Therefore, the problem of robustness analysis (Kharin and Zhuk 1998), (Kharin 1997), (Kharin and Vecherko 2013), (Galinskij and Kharin 1999) for sequential statistical decision making under distortions is an important one.

In (Quang 1985) the robustness analysis is given for a special hypothetical model in case of simple hypotheses. An empirical study of robustness is performed in (Pandit and Gudaganavara 2009) for the scale parameter of gamma and exponential distributions. Here we systemize some of the results developed by the author for quantitative robustness analysis of sequential decision rules.
2. Case of simple hypotheses

Let on a probability space \((\Omega, \mathcal{F}, \mathcal{P})\) random variables \(x_t \in U = \{u_1, u_2, \ldots, u_M\}\) be observed, \(t \in \mathbb{N}\), independent in total and identically distributed. The probability distribution of each random variable depends on parameter value \(\theta \in \Theta = \{\theta_0, \theta_1\}\), \(\theta_0, \theta_1 \in \mathbb{R}\), \(\theta_0 \neq \theta_1\), and satisfy the grid condition:

\[
P(u; \theta) = P_{\theta}\{x_t = u\} = a^{-J(u; \theta)}, \ t \in \mathbb{N}, \ u \in U, \tag{1}
\]

where \(a \in \mathbb{Q}\), \(a > 1\); \(J(u; \theta): U \times \Theta \rightarrow \mathbb{Z}_+\) is a function, satisfying the normalization condition

\[
\sum_{u \in U} a^{-J(u; \theta)} = 1. \tag{2}
\]

Concerning the parameter value \(\theta\) of the probability distribution (1) the following two simple hypotheses are considered:

\[
\mathcal{H}_0: \ \theta = \theta_0, \ \mathcal{H}_1: \ \theta = \theta_1. \tag{3}
\]

Denote the statistic

\[
\Lambda_n = \Lambda_n(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \lambda_i, \tag{4}
\]

where \(\lambda_i = \log_a (P(x_i; \theta_1)/P(x_i; \theta_0)) = J(x_i; \theta_0) - J(x_i; \theta_1) \in \mathbb{Z}\) is the logarithm of the likelihood ratio, calculated by the observation \(x_t\).

In the sequential probability ratio test (SPRT, Wald test) for hypotheses (3) testing after \(n\) observations \((n = 1, 2, \ldots)\) the decision is made according to:

\[
d = 1_{[c_+, +\infty)}(\Lambda_n) + 2 \cdot 1_{[c_-, c_+)}(\Lambda_n). \tag{5}
\]

The values \(d = 0\) and \(d = 1\) correspond to stopping of the observation process and acceptance of the hypothesis \(\mathcal{H}_0\) (if \(d = 0\)) or \(\mathcal{H}_1\) (if \(d = 1\)) by \(n\) observations. If \(d = 2\), the observation number \((n + 1)\) should be made. In (5) parameters \(c_-, c_+ \in \mathbb{Z}\) are calculated as follows:

\[
c_− = \lceil \log_a (\beta_0/(1 - \alpha_0)) \rceil, \quad c_+ = \lceil \log_a ((1 - \beta_0)/\alpha_0) \rceil, \tag{6}
\]

where \(\alpha_0, \beta_0\) are the requested values of the error type I and II probabilities respectively; \([\cdot]\) means the integer part of an argument.

Let the model described above is distorted – instead of (1) observations \(x_1, x_2, \ldots\) are obtained from the mixture of discrete distributions:

\[
P(u; \theta) = P_{\theta}\{x_t = u\} = (1 - \varepsilon)P(u; \theta) + \varepsilon \tilde{P}(u; \theta), \ t \in \mathbb{N}, \ u \in U, \tag{7}
\]

where \(\varepsilon \in [0, \frac{1}{2}]\) is the “contamination” probability; the “contaminating” probability distribution has the form

\[
\tilde{P}(u; \theta) = a^{-J(u; \theta)}, \ u \in U, \ \theta \in \Theta, \tag{8}
\]

where \(\tilde{J}(u; \theta): U \times \Theta \rightarrow \mathbb{Z}_+\) is a function, different from \(J(\cdot)\), and satisfying \(\sum_{u \in U} a^{-\tilde{J}(u; \theta)} = 1\).

For \(i \in (C_-, C_+)\) denote

\[
\tilde{P}_{ij}^{(k)}(g) = \begin{cases} 
\sum_{u \in U} \delta_{g(J(u; \theta_0) - J(u; \theta_1))} \tilde{P}(u; \theta_j), & j \in (C_-, C_+), \\
\sum_{u \in U} 1_{(-\infty, C_-)}(g(J(u; \theta_0)) - J(u; \theta_1)) + \varepsilon \tilde{P}(u; \theta_j), & j = C_-,
\end{cases} \tag{9}
\]

\[
\tilde{P}(g) = \left(\tilde{P}_{ij}^{(k)}(g)\right), \ \tilde{\pi}^{(k)}(g) = (\tilde{\pi}_{ij}^{(k)}(g)), \ i \in (C_-, C_+); \tag{10}
\]

where \(\delta_{g(J(u; \theta_0) - J(u; \theta_1))} \tilde{P}(u; \theta_j)\) is a function, different from \(\tilde{P}(\cdot)\), and satisfying \(\sum_{u \in U} a^{-\tilde{J}(u; \theta)} = 1\).
The case of Markov chains is considered in details in (Kharin 2013).

The case of inhomogeneous data – the model of time series with a trend – is analyzed in sequential tests that makes possible to construct the robust sequential test by the minimax theory, the performance characteristics can be calculated for tests from a family of distorted models differ from the correspondent characteristics calculated for the hypothetical model, by the values of the order of the risk criterion. The detailed proof is given in (Kharin 2013).

Proof. The detailed proof is given in (Kharin 2013).

Using this theory, the performance characteristics can be calculated for tests from a family of sequential tests that makes possible to construct the robust sequential test by the minimax theory, the performance characteristics can be calculated for tests from a family of distorted models differ from the correspondent characteristics calculated for the hypothetical model, by the values of the order of the risk criterion. The detailed proof is given in (Kharin 2013).

3. Case of composite hypotheses

3.1. Mathematical model and notation

Let on a probability space $(\Omega, F, \mathcal{P})$ a random sequence of independent inhomogeneous variables $x_1, x_2, \ldots \in \mathbb{R}$ be observed, with p.d.f $p_1(x|\theta)$, $p_2(x|\theta), \ldots$, where $\theta \in \Theta \subseteq \mathbb{R}^k$ is an unknown value of the random parameters vector. The prior probability density function $p(\theta)$ of this vector in the Bayesian setting is supposed to be known. There are two composite hypotheses on the value of $\theta$:

$$H_0 : \theta \in \Theta_0, \quad H_1 : \theta \in \Theta_1; \quad \Theta_0 \cup \Theta_1 = \Theta, \quad \Theta_0 \cap \Theta_1 = \emptyset. \quad (9)$$

Introduce the notation:

$$1_S(s) = \begin{cases} 1, & s \in S, \\ 0, & s \notin S; \end{cases}$$
\[
W_i = \frac{1}{W_i} \cdot p(\theta) = \frac{1}{W_i} \cdot p(\theta) \cdot 1_{\Theta_i}(\theta), \ \theta \in \Theta, \ i = 0, 1.
\]
Denote by
\[
\Lambda_n = \Lambda_n(x_1, \ldots, x_n) = \log \frac{\int_{\Theta} w_1(\theta)p_1(x_1 | \theta) d\theta \cdots \int_{\Theta} w_1(\theta)p_n(x_n | \theta) d\theta}{\int_{\Theta} w_0(\theta)p_1(x_1 | \theta) d\theta \cdots \int_{\Theta} w_0(\theta)p_n(x_n | \theta) d\theta}
\]  
the logarithm of the generalized likelihood ratio statistic, that is calculated by \( n \) observations \( x_1, \ldots, x_n \).

To test the hypotheses (9), the following parametric family of Bayesian sequential tests is used:
\[
N = \min \{ n \in \mathbb{N} : \Lambda_n \notin (C_-, C_+) \},
\]
\[
d = 1_{[C_+, +\infty]}(\Lambda_N),
\]
where \( N \) is the random number of the observation that determines the stopping time, after that observation the decision \( d \) is made according to the decision rule (12). The decision \( d = i \) means that the hypothesis \( \mathcal{H}_i \) is accepted, \( i = 0, 1; \ C_- < 0, \ C_+ > 0 \) are parameters of the test (11), (12):
\[
C_- = \ln(\beta_0/(1 - \alpha_0)), \ C_+ = \ln((1 - \beta_0)/\alpha_0),
\]
where \( \alpha_0, \beta_0 \in (0, \frac{1}{2}) \) are some values close to maximal admissible levels of error type I and II probabilities (Wald 1947). The actual values \( \alpha, \beta \) of the error type I and II probabilities may deviate from \( \alpha_0, \beta_0 \).

For calculation of \( \alpha, \beta \) and mathematical expectations of the random variable \( N \) determined by (11), let us use a stochastic approximation of the statistic \( \Lambda_n, n \in \mathbb{N} \). Let \( m \in \mathbb{N} \) be a parameter of the approximation, \( h = (C_+ - C_-)/m \). Let \( p_{\Lambda_n}(u) \) be the probability density function of the statistic (10); \( p_{\Lambda_n \mid \Lambda_n}(u | y) \) be the conditional probability density function, \( n \in \mathbb{N} \); let \( R^{(n)}(\theta) \) and \( Q^{(n)}(\theta) \) be the blocks of the sizes \( m \times 2 \) and \( m \times m \) respectively for the approximating Markov chain, \( I_k \) is the identity matrix of the size \( k \), \( \mathbf{0}_{(2 \times m)} \) is the matrix of the size \( (2 \times m) \), with all elements equal to 0. Let \( \pi(\theta) = (\pi_1(\theta)) \) be the vector of initial probabilities of the states 1, \ldots, \( m \) for the approximating random sequence; \( \pi_0(\theta), \pi_{m+1}(\theta) \) be the initial probabilities of the absorbing states 0 and \( m + 1; \mathbf{1}_m \) be the vector of size \( m \), with all components equal to 1. Denote:
\[
S(\theta) = \mathbf{1}_m + \sum_{i=1}^{\infty} \prod_{j=1}^{i} Q^{(j)}(\theta);
\]
\[
B(\theta) = R^{(1)}(\theta) + \sum_{i=1}^{\infty} \prod_{j=1}^{i} Q^{(j)}(\theta) R^{(i+1)}(\theta).
\]
Let \( B^{(j)}(\theta) \) be the column number \( j \) of the matrix \( B(\theta), j = 1, 2; t_i = E\{N | \theta \in \Theta_i\}, i = 0, 1; t = E\{N\} \).

3.2. Distortions
Let the hypothetical model described above be distorted, although the Bayesian sequential test (11), (12) is used. The test is constructed on the basis of the hypothetical probability density functions \( p(\theta), p_n(x_1, \ldots, x_n | \theta) \), but these probability density functions are simultaneously distorted. Actually, the parameters vector \( \theta \) has the distorted probability density function
\[
\tilde{p}(\theta) = (1 - \varepsilon_\theta) \cdot p(\theta) + \varepsilon_\theta \cdot \tilde{p}(\theta), \ \theta \in \Theta,
\]
where \( \varepsilon_\theta \in [0, \frac{1}{2}] \) is the probability of “contamination” w.r.t. the probability density of \( \theta \), and \( \tilde{p}(\theta) \) is a “contaminating” probability density function that differs from \( p(\theta) \). The distorted
conditional probability density function of observations is also a mixture of the hypothetical
probability density functions:

\[ \bar{p}_n(x_1, \ldots, x_n \mid \theta) = (1 - \varepsilon_x) \cdot p_n(x_1, \ldots, x_n \mid \theta) + \varepsilon_x \cdot \tilde{p}_n(x_1, \ldots, x_n \mid \theta), \]

\( \theta \in \Theta, \ x_1, \ldots, x_n \in \mathbb{R}, \ n \in \mathbb{N}, \)

where \( \varepsilon_x \in [0, \frac{1}{2}) \) can be interpreted as the probability of an “outlier” presence (see Huber and Ronchetti (2009)) w.r.t. the observations \( x_1, x_2, \ldots. \)

Let \( \tilde{\pi}(\theta), \tilde{\pi}_0(\theta), \tilde{\pi}_{m+1}(\theta), \tilde{Q}^{(n)}(\theta), \tilde{R}^{(n)}(\theta) \) be the elements calculated analogously to \( \pi(\theta), \pi_0(\theta), \pi_{m+1}(\theta), Q^{(n)}(\theta), R^{(n)}(\theta) \) by replacing the hypothetical p.d.f. \( p_n(x_1, \ldots, x_n \mid \theta) \) with the “contaminating” p.d.f. \( \tilde{p}_n(x_1, \ldots, x_n \mid \theta) \) in the probability distribution of the approximating random sequence; \( \Delta \pi_0(\theta) = \tilde{\pi}_0(\theta) - \pi_0(\theta), \Delta \pi_{1}(\theta) = \tilde{\pi}_{m+1}(\theta) - \pi_{m+1}(\theta) \); \( \bar{\ell}(\theta) \) and \( \bar{\gamma}_\ell(\theta), \)

\( i = 0, 1, \) be the conditional mathematical expectation of the sample size and the conditional probability of acceptance of the hypothesis \( H_i \) respectively, provided the parameters vector value is \( \theta \), for the distorted model (13), (14).

### 3.3. Robustness analysis via asymptotic expansions of the performance characteristics

Introduce the notation:

\[ \bar{W}_i = \int_{\Theta} \bar{p}(\theta)d\theta; \]

\[ A(\theta) = \left( \left( \bar{\pi}(\theta) - \pi(\theta) \right)' S(\theta) + \right) \]

\[ (\pi(\theta))' \cdot \sum_{i=1}^{\infty} \sum_{j=1}^{i-1} \prod_{k=1}^{j-1} Q^{(k)}(\theta) (\bar{Q}^{(j)}(\theta) - Q^{(j)}(\theta)) \prod_{k=j+1}^{i} Q^{(k)}(\theta) \cdot 1_m; \]

\[ F_i(\theta) = \Delta \pi_i(\theta) + (\bar{\pi}(\theta) - \pi(\theta))' B_{(i+1)}(\theta) + \tilde{R}^{(1)}(\theta) - R^{(1)}(\theta); \]

\[ \sum_{i=1}^{\infty} \left( \sum_{j=1}^{i-1} \prod_{k=1}^{j-1} Q^{(k)}(\theta) (\bar{Q}^{(j)}(\theta) - Q^{(j)}(\theta)) \prod_{k=j+1}^{i} Q^{(k)}(\theta) R^{(i+1)}(\theta) + \right) \]

\[ \left( \prod_{j=1}^{i} Q^{(j)}(\theta) (\bar{R}^{(i+1)}(\theta) - R^{(i+1)}(\theta)) \right), \ i = 0, 1. \]

**Theorem 2.** Let the random sequence (10) satisfies the Markov property, \( \forall \theta \in \Theta, \) the probability density functions \( p_{\Lambda_1}(u), p_{\Lambda_{n+1}}(u \mid y) \) be differentiable functions w.r.t. the variable \( u \in [C_-, C_+], \) and \( \exists C \in (0, +\infty): \)

\[ \left| \frac{dp_{\Lambda_1}(u)}{du} \right| \leq C, \left| \frac{dp_{\Lambda_{n+1}}(u \mid y)}{du} \right| \leq C, \ u, y \in [C_-, C_+], \ n \in \mathbb{N}. \]

Then under simultaneous distortions (13), (14) the following asymptotic expansions hold for the error type I and II probabilities \( \bar{\alpha}, \bar{\beta} \) at \( \varepsilon_\theta \to 0, \varepsilon_x \to 0, \ h \to 0: \)

\[ \bar{\alpha} = \alpha + \varepsilon_x \cdot \frac{1}{W_0} \cdot \int_{\Theta} F_1(\theta)p(\theta)d\theta + \]

\[ \varepsilon_\theta \left( \frac{1}{W_0^2} \cdot \int_{\Theta} (\bar{p}(\theta) - p(\theta))d\theta \cdot \int_{\Theta} \gamma_{\ell_1}(\theta)p(\theta)d\theta + \frac{1}{W_0} \cdot \int_{\Theta} \gamma_{\ell_1}(\theta)(\bar{p}(\theta) - p(\theta))d\theta \right) + \]

\[ \mathcal{O}(\varepsilon_\theta^2) + \mathcal{O}(\varepsilon_x^2) + \mathcal{O}(h); \]

\[ \bar{\beta} = \beta + \varepsilon_x \cdot \frac{1}{W_1} \cdot \int_{\Theta} F_0(\theta)p(\theta)d\theta + \]

\[ \varepsilon_\theta \left( \frac{1}{W_1^2} \cdot \int_{\Theta} (\bar{p}(\theta) - p(\theta))d\theta \cdot \int_{\Theta} \gamma_{\ell_0}(\theta)p(\theta)d\theta + \frac{1}{W_1} \cdot \int_{\Theta} \gamma_{\ell_0}(\theta)(\bar{p}(\theta) - p(\theta))d\theta \right) + \]

\[ \mathcal{O}(\varepsilon_\theta^2) + \mathcal{O}(\varepsilon_x^2) + \mathcal{O}(h). \]
Robustness in Sequential Statistical Decisions

Proof. The detailed proof is given in (Kharin 2013).

Denote by $\bar{t}_i$, $i = 0, 1$, the conditional mathematical expectation of the random number of observations $N$ provided the hypothesis $H_i$ is true, if the hypothetical model is distorted according to (13), (14).

**Theorem 3.** Under the conditions of Theorem 2, the conditional expected sample sizes satisfy the asymptotic expansions:

$$
\bar{t}_i = t_i + \varepsilon_x \cdot \frac{1}{W_i} \cdot \int_{\Theta_i} A(\theta)p(\theta)d\theta + \varepsilon_\theta \cdot \left( t_i \cdot (\bar{W}_i - W_i) + \frac{1}{W_i} \cdot \int_{\Theta_i} t(\theta)(\bar{p}(\theta) - p(\theta))d\theta \right) + O(\varepsilon_x^2) + O(h), \ i = 0, 1.
$$

**Theorem 4.** If the conditions of Theorem 2 are satisfied, then the following asymptotic expansion holds for the expected sample size:

$$
\bar{t} = t + \varepsilon_x \cdot \int_{\Theta} A(\theta)p(\theta)d\theta + \varepsilon_\theta \cdot \int_{\Theta} t(\theta)(\bar{p}(\theta) - p(\theta))d\theta + O(\varepsilon_x^2) + O(h).
$$

Proof. Proofs of Theorems 3, 4 are presented in Kharin (2013).

This theory is used to calculate the performance characteristics of the sequential tests under distortions and to construct the robust sequential test (Kharin 2017). The approach is presented in (Kharin 2016) with some numerical results.

4. Conclusion

The problem of robustness analysis for sequential statistical decision rules is considered in the paper. The cases of simple and composite hypotheses are analyzed. Asymptotic expansions are constructed for the performance characteristics of the sequential statistical decision rules under distortion. Analyzing the constructed expansions and constructing the similar for sequential test from a generalized families, robust sequential statistical decision rules can be constructed. The results are also applied for the decision making in case of many hypotheses (Ton and Kharin 2019).

References


**Affiliation:**
Alexey Kharin
Department of Probability Theory and Mathematical Statistics
Belarusian State University
Independence av. 4, Minsk 220030, Belarus
E-mail: KharinAY@bsu.by
URL: [https://fpmi.bsu.by/?guid=32921](https://fpmi.bsu.by/?guid=32921)