



HAL
open science

Sequential Analysis and its Applications

Pergamenshchikov Serguei, Pchelintsev Evgeny

► **To cite this version:**

Pergamenshchikov Serguei, Pchelintsev Evgeny. Sequential Analysis and its Applications. 2022. hal-03649695

HAL Id: hal-03649695

<https://hal.science/hal-03649695v1>

Preprint submitted on 22 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sequential Analysis and its Applications ^{*}

Pergamenshchikov Serguei[†] and Pchelintsev Evgeny, [‡]

April 22, 2022

Abstract

This course is devoted to the main problems of the sequential analysis: sequential estimation and sequential hypothesis testing. Firstly we construct the least square estimator for the scalar regression model and then we propose the sequential least square estimator for the autoregression models. Finally, we study the non asymptotic properties for the sequential estimation procedures. Then in the second part of this course we construct and study the sequential Wald procedure for hypothesis testing. We study its main properties: the mean times and the optimality properties in the sense of minimal mean time. Then we consider some examples of the Wald procedures.

^{*}This work was done under financial support of the Russian Federal Professor program (project no. 1.472.2016/1.4, Ministry of Education and Science) and under the Ministry of Education and Science of the Russian Federation in the framework of the research project no. 2.3208.2017/4.6,

[†]Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS- Université de Rouen Normandie, France and International Laboratory of Statistics of Stochastic Processes and Quantitative Finance of National Research Tomsk State University, e-mail: Serge.Pergamenshchikov@univ-rouen.fr

[‡]International Laboratory of Statistics of Stochastic Processes and Quantitative Finance of National Research Tomsk State University

1 Basic Stochastic Calculus

Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbf{P})$ be a probability space with a filtration $(\mathcal{F}_n)_{n \geq 0}$ which is an increasing sequence of fields, i.e. $\mathcal{F}_{n-1} \subseteq \mathcal{F}_n$ for $n \geq 1$. Moreover, we assume that the initial fields is trivial, i.e. $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Definition 1.1. A random variable $\tau \in \mathbb{N} = \{0, 1, 2, \dots\}$ is called Markov moment, if $\{\tau = n\} \in \mathcal{F}_n$ for all $n \geq 0$. If $\mathbf{P}(\tau < \infty) = 1$, then τ is stopping time.

For any Markov moment τ we set

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq n\} \in \mathcal{F}_n \quad \forall n \geq 0\}. \quad (1.1)$$

Main properties of stopping times.

1. If τ is a constant, i.e. $\tau = \tau(\omega) = k$ for some $k \in \mathbb{N}$ and for all $\omega \in \Omega$, then it is stopping time.
2. The set $\{\omega \in \Omega : \tau(\omega) \geq n\} \in \mathcal{F}_{n-1}$ for all $n \geq 1$.
3. A random variable τ with the values in \mathbb{N} is stopping time if and only if $\{\omega \in \Omega : \tau(\omega) \leq n\} \in \mathcal{F}_n$ for all $n \geq 0$.
4. If τ and σ are stopping times, then $\tau \wedge \sigma$ and $\tau \vee \sigma$ are stopping times.
5. \mathcal{F}_τ is field.
6. If τ is a constant, i.e. $\tau = \tau(\omega) = k$ for some $k \in \mathbb{N}$ and for all $\omega \in \Omega$, then $\mathcal{F}_\tau = \mathcal{F}_k$.
7. τ is measurable with respect to \mathcal{F}_τ , i.e. $\{\omega \in \Omega : \tau(\omega) = m\} \in \mathcal{F}_\tau$ for any $m \in \mathbb{N}$.

Definition 1.2. A sequence of random variables $\xi = (\xi_n)_{n \geq 0}$ is called adapted, if ξ_n is \mathcal{F}_n - measurable for any $n \geq 0$ and it is called predictable, if ξ_0 is \mathcal{F}_0 - measurable and ξ_n is \mathcal{F}_{n-1} - measurable for any $n \geq 1$.

Definition 1.3. A sequence of random variables $M = (M_n)_{n \geq 0}$ is called martingale, if the following properties hold:

- it is adapted;
- it is integrable, i.e. $\mathbf{E}|M_n| < \infty$ for any $n \geq 1$;
- $\mathbf{E}(M_{n+1} | \mathcal{F}_n) = M_n$ for any $n \geq 0$.

Definition 1.4. A martingale $M = (M_n)_{n \geq 0}$ is called the square integrable martingale if $\mathbf{E} M_n^2 < \infty$ for all $n \geq 1$. In this case the sequence

$$\langle M \rangle = (\langle M \rangle_n)_{n \geq 1}$$

defined for all $n \geq 1$ as

$$\langle M \rangle_n = \sum_{j=1}^n \mathbf{E} ((M_j - M_{j-1})^2 | \mathcal{F}_{j-1}) \quad (1.2)$$

is called quadratic characteristic.

Definition 1.5. Let $v = (v_k)_{k \geq 1}$ be a predictable bounded sequence and $\mathbf{m} = (\mathbf{m}_n)_{n \geq 0}$ be a square integrable martingale. Then the process $M = (M_n)_{n \geq 0}$ defined as

$$M_n = M_0 + \sum_{j=1}^n v_j (\mathbf{m}_j - \mathbf{m}_{j-1}) \quad (1.3)$$

is called martingale transformation.

Theorem 1.1. If M is square integrable martingale and τ is bounded stopping times, then $\mathbf{E} M_\tau = M_0$ and $\mathbf{E} M_\tau^2 = M_0^2 + \mathbf{E} \langle M \rangle_\tau$.

Proof. First for any $n \geq 1$ we defined the stopping martingale $\widetilde{M} = (\widetilde{M}_n)_{n \geq 1}$ as

$$\widetilde{M}_n = M_{\tau \wedge n} = M_0 + \sum_{j=1}^n v_j (M_j - M_{j-1}) \quad \text{and} \quad v_j = \mathbf{1}_{\{j \leq \tau\}}.$$

According to Property 2 the sequence $(v_j)_{j \geq 1}$ is predictable and, therefore, in view of Exercice 4, the sequence $\widetilde{M} = (\widetilde{M}_n)_{n \geq 0}$ is square integrable martingale with $\widetilde{M}_0 = M_0$ and

$$\langle \widetilde{M} \rangle_n = \sum_{j=1}^n \mathbf{1}_{\{j \leq \tau\}} (\langle M \rangle_j - \langle M \rangle_{j-1}) = \langle M \rangle_{\tau \wedge n}.$$

Moreover, note, that in our case the stopping time τ is bounded, i.e. there exists some integer $N \geq 1$ such that $\mathbf{P}(\tau \leq N) = 1$, i.e. $M_\tau = \widetilde{M}_N$. Therefore, $\mathbf{E} M_\tau = \mathbf{E} \widetilde{M}_N = \widetilde{M}_0 = M_0$ and

$$\mathbf{E} M_\tau^2 = \mathbf{E} \widetilde{M}_N^2 = M_0^2 + \mathbf{E} \langle \widetilde{M} \rangle_N = M_0^2 + \mathbf{E} \langle M \rangle_\tau.$$

Hence Theorem 1.1. \square

The following result is known as the Wald identity.

Theorem 1.2. Assume that $(\xi_j)_{j \geq 1}$ are i.i.d. integrable random variables, i.e. $\mathbf{E}|\xi_j| < \infty$ and τ is a integrable stopping time with respect the filtration $(\mathcal{F}_j)_{j \geq 0}$ with $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_j = \sigma\{\xi_1, \dots, \xi_j\}$ for $j \geq 1$. Then

$$\mathbf{E} S_\tau = \mathbf{E} \tau \mathbf{E} \xi_1, \quad (1.4)$$

where $S_n = \sum_{j=1}^n \xi_j$.

Proof. First assume, that $\xi_j \geq 0$ a.s. In this case we can represent S_τ as

$$S_\tau = \lim_{m \rightarrow \infty} S_{\tau_m} \quad \text{and} \quad \tau_m = \tau \wedge m.$$

Note here, that for any $m \geq 1$, in view of Property 2,

$$\mathbf{E} S_{\tau_m} = \sum_{j=1}^m \mathbf{E} \mathbf{1}_{\{j \leq \tau\}} \xi_j = \sum_{j=1}^m \mathbf{E} \mathbf{1}_{\{j \leq \tau\}} \mathbf{E} (\xi_j | \mathcal{F}_{j-1}) = \mathbf{E} \xi_1 \mathbf{E} \tau \wedge m.$$

Taking into account, that the sequence $(S_{\tau_m})_{m \geq 1}$ is increasing, we get through monotone convergence theorem, that

$$\mathbf{E} S_\tau = \lim_{m \rightarrow \infty} \mathbf{E} S_{\tau_m} = \mathbf{E} \xi_1 \lim_{m \rightarrow \infty} \mathbf{E} \tau \wedge m = \mathbf{E} \xi_1 \mathbf{E} \tau.$$

In the general case, we obtain the equality (1.4) using the representation $\xi_j = (\xi_j)_+ - (\xi_j)_-$, where $(x)_+ = \max(x, 0)$ and $(x)_- = -\min(x, 0)$. Hence Theorem 1.2. \square

Exercises 1

1. Show, that if M is a square integrable martingale, then

$$\text{Var}(M_n) = \mathbf{E} \langle M \rangle_n.$$

2. Show, that if $(\xi_n)_{n \geq 1}$ is i.i.d sequence of random variables with $\mathbf{E} \xi_1 = 0$ and $\mathbf{E} \xi_1^2 = \sigma^2, \infty$, then for any constant M_0 the sequence

$$M_n = M_0 + \sum_{j=1}^n \xi_j$$

is square integrable martingale with respect to the filtration

$$(\mathcal{F}_n = \sigma\{\xi_1, \dots, \xi_n\})_{n \geq 1}$$

and its quadratic characteristic $\langle M \rangle_n = n\sigma^2$.

3. Show, that the sequence (1.2) is predictable.
4. Show, that if $\mathbf{m} = (\mathbf{m}_n)_{n \geq 0}$ is a square integrable martingale, then the martingale transformation (1.3) is square integrable martingale and its quadratic characteristic is

$$\langle M \rangle_n = \sum_{j=1}^n v_j^2 (\langle \mathbf{m} \rangle_j - \langle \mathbf{m} \rangle_{j-1}) .$$

2 Least Square Method

First, we consider the scalar linear regression model, i.e.

$$y_j = \theta x_j + \varepsilon_j, \quad 1 \leq j \leq n, \quad (2.1)$$

where θ is unknown parameter, $(x_j)_{1 \leq j \leq n}$ are non random regression variables and $(\varepsilon_j)_{1 \leq j \leq n}$ is unobservable white noise, i.e. $\mathbf{E}\varepsilon_j = 0$ and $\mathbf{E}\varepsilon_j^2 = \sigma^2$ for any $1 \leq j \leq n$ and $\mathbf{E}\varepsilon_j\varepsilon_l = 0$ for $j \neq l$.

The identification problem for the model (2.1) is to estimate the parameter θ on, the basis of the observations $(y_j)_{1 \leq j \leq n}$. To this end we will use the Least Square Estimator (LSE) method according to which one needs to minimize over unknown parameter the integral noise intensity, i.e.

$$\sum_{j=1}^n (y_j - \theta x_j)^2 \rightarrow \min_{\theta \in \mathbb{R}} . \quad (2.2)$$

Therefore, if

$$\sum_{j=1}^n x_j^2 > 0 \quad (2.3)$$

then we obtain immediately that least square estimator is

$$\hat{\theta}_n = \frac{\sum_{j=1}^n y_j x_j}{\sum_{j=1}^n x_j^2} . \quad (2.4)$$

From the model (2.1) it is easy to deduce that

$$\hat{\theta}_n = \theta + \frac{\sum_{j=1}^n x_j \varepsilon_j}{\sum_{j=1}^n x_j^2} .$$

Therefore,

$$\mathbf{E} \widehat{\theta}_n = \theta + \frac{\sum_{j=1}^n x_j \mathbf{E} \varepsilon_j}{\sum_{j=1}^n x_j^2} = 0$$

and, moreover, the mean square estimation accuracy in this case can be calculated as

$$\mathbf{V}(\widehat{\theta}_n) = \mathbf{E} (\widehat{\theta}_n - \theta)^2 = \frac{\mathbf{E} \left(\sum_{j=1}^n x_j \varepsilon_j \right)^2}{\left(\sum_{j=1}^n x_j^2 \right)^2} = \frac{\sigma^2}{\sum_{j=1}^n x_j^2}. \quad (2.5)$$

From this we can obtain immediately the necessary and sufficient condition for the convergence in \mathbf{L}_2 as $n \rightarrow \infty$.

Proposition 2.1. *The least square estimator (2.4) tends to θ in \mathbf{L}_2 if and only if*

$$\lim_{n \rightarrow \infty} \sum_{l=1}^n x_l^2 = +\infty. \quad (2.6)$$

For this estimator one can show the following theorem.

Theorem 2.1. *(Gauss - Markov) The least square estimator (2.5) is the best estimator in the class of all linear unbiased estimators of the non zero parameter θ in the model (2.1) with the condition (2.3) in the means square accuracy sense*

$$\mathbf{E} (\widetilde{\theta}_n - \theta)^2 \geq \mathbf{E} (\widehat{\theta}_n - \theta)^2, \quad (2.7)$$

where $\widetilde{\theta}_n$ is an arbitrary linear estimator, i.e. an estimator of the form

$$\widetilde{\theta}_n = \sum_{j=1}^n \mathbf{g}_j y_j$$

and $(\mathbf{g}_j)_{1 \leq j \leq n}$ are non random coefficients.

Proof. Indeed, note that for unbiased estimators we have

$$\theta = \mathbf{E} \widetilde{\theta}_n = \sum_{j=1}^n \mathbf{g}_j \mathbf{E} y_j = \theta \sum_{j=1}^n \mathbf{g}_j \mathbf{x}_j,$$

i.e. $\sum_{j=1}^n \mathbf{g}_j \mathbf{x}_j = 1$. Using here the Cauchy Bunyakovsky Schwarz we get

$$1 = \left(\sum_{j=1}^n \mathbf{g}_j \mathbf{x}_j \right)^2 \leq \sum_{j=1}^n \mathbf{g}_j^2 \sum_{j=1}^n \mathbf{x}_j^2.$$

Therefore,

$$\mathbf{E}(\tilde{\theta}_n - \theta)^2 = \mathbf{E} \left(\sum_{j=1}^n \mathbf{g}_j \varepsilon_j \right)^2 = \sigma^2 \sum_{j=1}^n \mathbf{g}_j^2 \geq \frac{\sigma^2}{\sum_{j=1}^n \mathbf{x}_j^2}.$$

Now, the property (2.5) implies directly (2.7). Hence Theorem 2.1. \square

Exercises 2

1. Check, if the LSE will be convergence in \mathbf{L}_2 in following cases :
 - $x_j = j/n$ for $1 \leq j \leq n$
 - $x_j = \sin(2\pi j/n)$ for $1 \leq j \leq n$
2. Write the convergence criteria in \mathbf{L}_2 for LSE for multidimensional case.

3 Sequential Estimation

In this section we study the auto-regression process defined as

$$y_j = \theta_1 y_{j-1} + \dots + \theta_p y_{j-p} + \varepsilon_j, \quad j \geq 1, \quad (3.1)$$

where $(\varepsilon_j)_{j \geq 1}$ are i.i.d. random variables with $\mathbf{E} \varepsilon_j = 0$ and $\mathbf{E} \varepsilon_j^2 = 1$ which are unobserved. The initial values y_0, \dots, y_{-p+1} are nonrandom known constants. The problem is to estimate the unknown parameters $(\theta_1, \dots, \theta_p)$ on the basis of observations $(y_j)_{j \geq 1}$. For this problem we use the filtration generated by the observations, i.e. $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_j = \sigma\{y_1, \dots, y_j\}$ for $j \geq 1$. First we study this model, when $p = 1$, i.e.

$$y_j = \theta y_{j-1} + \varepsilon_j. \quad (3.2)$$

This model can be represented as a particularly case of the random regression model (2.1) with $x_j = y_{j-1}$. So, we will study the model (2.1) with predictable sequence $(x_j)_{j \geq 1}$. Note, that in this case the least square estimator (2.4) is a non linear function of the observations $(y_j)_{j \geq 1}$ and we can't study it by the usual methods in the non asymptotic setting, i.e. for the finite n . By this reason, for the non asymptotic analysis we use the sequential procedure (τ_H, θ_H^*) proposed in [2], where τ_H is a stopping time defining the number of

the observations, θ_H^* is the sequential estimator and $H > 0$ is non negative non random parameter, which will be specified later. In this case we have

$$\tau_H = \inf\{n \geq 1 : \sum_{l=1}^n x_l^2 \geq H\} \quad (3.3)$$

and

$$\theta_H^* = \frac{\sum_{j=1}^{\tau_H-1} x_j y_j + \beta_{\tau_H} x_{\tau_H} y_{\tau_H}}{H}, \quad (3.4)$$

where β_{τ_H} is the corrected coefficient defined from the following condition

$$\sum_{k=1}^{\tau_H-1} x_k^2 + \beta_{\tau_H} x_{\tau_H}^2 = H, \quad \text{i.e.} \quad \beta_{\tau_H} = \frac{H - \sum_{k=1}^{\tau_H-1} x_k^2}{x_{\tau_H}^2}. \quad (3.5)$$

According to Exercice 2 the function τ_H is the stopping time for any $H > 0$, i.e. $\tau_H < \infty$ a.s. Therefore, the coefficient $0 < \beta \leq 1$ a.s.

Theorem 3.1. *The sequential procedure (3.4) - (3.4) for any $H > 0$ satisfies the following properties:*

1. $\tau_H < \infty$ a.s.;
2. the estimator θ_H^* is unbiased, i.e. for any $\theta \in \mathbb{R}$

$$\mathbf{E}_\theta \theta_H^* = \theta; \quad (3.6)$$

3. the estimator θ_H^* has a fixed mean square accuracy, i.e.

$$\sup_{\theta \in \mathbb{R}} \mathbf{E}_\theta (\theta_H^* - \theta)^2 \leq \frac{1}{H}. \quad (3.7)$$

Proof. The first property is shown in the exercice 1. As to the second property, note that the deviation of the estimator (3.4) can be represented as

$$\theta_H^* - \theta = \frac{\sum_{j=1}^{\tau_H-1} x_j \varepsilon_j + \beta_{\tau_H} x_{\tau_H} \varepsilon_{\tau_H}}{H} = \frac{1}{H} M_{\tau_H},$$

where

$$M_n = \sum_{j=1}^n v_j \varepsilon_j \quad \text{and} \quad v_j = x_j \mathbf{1}_{\{\tau_H > j\}} + \beta_{\tau_H} x_{\tau_H} \mathbf{1}_{\{\tau_H = j\}}.$$

It should be noted here, that the definition (3.3) implies, that the sets $\{\tau_H > j\}$ and $\{\tau_H = j\}$ belong to \mathcal{F}_{j-1} for any $j \geq 1$. Note also, that the from (3.5) we

can conclude, that β_k is \mathcal{F}_{k-1} measurable for any $k \geq 1$. Therefore, taking into account, that

$$\beta_{\tau_H} \mathbf{1}_{\{j=\tau_H\}} = \beta_j \mathbf{1}_{\{j=\tau_H\}},$$

we get, that the sequence $(v_j)_{j \geq 1}$ is predictable, i.e. v_j is \mathcal{F}_{j-1} - measurable for any $j \geq 1$. Moreover, it is bounded, i.e. $\sup_{j \geq 1} |v_j| \leq \sqrt{H}$ a.s. Therefore, by Exercice 1.4 we get, that the sequence $M = (M_n)_{n \geq 0}$ is a square integrable martingale with

$$\langle M \rangle_n = \sum_{j=1}^n v_j^2 \leq H.$$

Therefore, in view of Theorem 1.1 for any $N > 1$ we get, that

$$\mathbf{E}_\theta M_{\tau_H \wedge N} = M_0 = 0 \quad \text{and} \quad \mathbf{E}_\theta M_{\tau_H \wedge N}^2 = \mathbf{E}_\theta \langle M \rangle_{\tau_H \wedge N} \leq H.$$

This implies immediately, that the sequence $(M_{\tau_H \wedge N})_{N \geq 1}$ is uniformly integrable, i.e.

$$\mathbf{E}_\theta M_{\tau_H} = \mathbf{E}_\theta \lim_{N \rightarrow \infty} M_{\tau_H \wedge N} = \lim_{N \rightarrow \infty} \mathbf{E}_\theta M_{\tau_H \wedge N} = 0.$$

This implies the property (3.6). The property (3.7) can be obtained through Fatou's lemma, i.e.

$$\mathbf{E}_\theta M_{\tau_H}^2 = \mathbf{E}_\theta \liminf_{N \rightarrow \infty} M_{\tau_H \wedge N}^2 \leq H. \quad (3.8)$$

Hence Theorem 3.1. \square

Lets consider now the estimation problem for the process (3.1). To this end we set $X_j = (y_{j-1}, \dots, y_{j-p+1})'$, where ' denotes the transposition. Then, we can represent the model (3.1) as

$$y_j = X_j' \theta + \varepsilon_j, \quad j \geq 1, \quad (3.9)$$

where in this case $\theta = (\theta_1, \dots, \theta_p)'$ is the unknown vector in \mathbb{R}^p . We will study the general random regression model with the predictable sequence $(X_j)_{j \geq 1}$. It should be noted, that the least square estimator for the unknown vector $\theta \in \mathbb{R}^p$ is defined as

$$\hat{\theta}_n = G_n^{-1} \sum_{j=1}^n X_j y_j \quad \text{and} \quad G_n = \sum_{j=1}^n X_j X_j'. \quad (3.10)$$

Note, that this estimator can be defined only under the condition, that the matrix G_n^{-1} exists. To calculate these estimators for sufficiently large n , we assumes that

$$\lim_{n \rightarrow \infty} \frac{1}{n} G_n = F \quad \text{a.s.}, \quad (3.11)$$

where F is a positive defined non random matrix. It should be noted, that for the model (3.1) (see, for example, in [1]) this condition holds with

$$F = \sum_{j \geq 0} A^j B (A')^j, \quad (3.12)$$

where

$$A = \begin{pmatrix} \theta_1 & , \dots, \theta_p \\ 1 & , \dots, 0 \\ \dots & \dots \\ 0 & , \dots, 1, 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & , \dots, 0 \\ \dots & \dots \\ 0 & , \dots, 0 \end{pmatrix}.$$

Now, to estimate this parameter through we use the two - step sequential estimation method developed in [6]. To this end, first we fixe the increasing sequence of positive numbers $(c_n)_{n \geq 1}$ such that

$$\rho = \sum_{n \geq 1} \frac{1}{c_n} < \infty. \quad (3.13)$$

In the first step we set the following sequence of stopping times defined as

$$\tau_n = \inf\{k \geq 1 : \text{tr} G_k \geq c_n\}, \quad (3.14)$$

where $\inf\{\emptyset\} = +\infty$. Now on the set $\{\tau_n < \infty\}$ we define the following correction for the matrix G_n defined in (3.10)

$$\tilde{G}_n = \sum_{j=1}^{\tau_n-1} X_j X_j' + \beta_{\tau_n} X_{\tau_n} X_{\tau_n}', \quad (3.15)$$

where the correction coefficient β_{τ_n} is defined as

$$\text{tr} G_{\tau_n-1} + \beta_{\tau_n} |X_{\tau_n}|^2 = c_n, \quad (3.16)$$

where $|\cdot|$ is the Euclidean norm in \mathbb{R}^p , i.e. $|X|^2 = X'X$. Now on the set

$$\Gamma_n = \{\tau_n < \infty\} \cap \{\det \tilde{G}_n > 0\} \quad (3.17)$$

we modify the least square estimator (3.10) as

$$\tilde{\theta}_n = \tilde{G}_n^{-1} \left(\sum_{j=1}^{\tau_n-1} X_j y_j + \beta_{\tau_n} X_{\tau_n} y_{\tau_n} \right). \quad (3.18)$$

Now setting the weight coefficients

$$\mathbf{b}_n = \frac{1}{|\tilde{G}_n^{-1}| c_n} \mathbf{1}_{\Gamma_n}, \quad (3.19)$$

we define for any $H > 0$ the stopping time as

$$\sigma_H = \inf\{n \geq 1 : \sum_{k=1}^n \mathbf{b}_k^2 \geq H\}. \quad (3.20)$$

Finally, we define the sequential procedure (N_H, θ_H^*) as

$$N_H = \tau_{\sigma_H} \quad \text{and} \quad \theta_H^* = \frac{1}{\sum_{k=1}^{\sigma_H} \mathbf{b}_k^2} \sum_{k=1}^{\sigma_H} \mathbf{b}_k^2 \tilde{\theta}_k \mathbf{1}_{\{\sigma_H < \infty\}}. \quad (3.21)$$

The properties of this procedure are given in this theorem from [6].

Theorem 3.2. *If the condition (3.11) holds, then the sequential procedure (3.21) for any $H > 0$ satisfies the following properties:*

1. $N_H < \infty$ a.s.;
2. the estimator θ_H^* has a fixed mean square accuracy, i.e.

$$\mathbf{E}_\theta |\theta_H^* - \theta|^2 \leq \frac{\rho}{H}, \quad (3.22)$$

where the coefficient ρ is defined in (3.13).

Proof. First, note, that the condition (3.11) directly implies

$$\lim_{n \rightarrow \infty} \frac{\tau_n}{c_n} = \frac{1}{\text{tr}F} \quad \text{and} \quad \lim_{n \rightarrow \infty} \mathbf{b}_n = \frac{1}{|F^{-1}| \text{tr}F} \quad \text{a.s.}$$

Therefore, $N_H < \infty$ a.s. for any $H > 0$. Moreover, note, that on the set (3.17) the estimator (3.18) can be represented as

$$\tilde{\theta}_k = \theta + \tilde{G}_k^{-1} \xi_k \quad \text{and} \quad \xi_k = \sum_{j=1}^{\tau_k-1} X_j \varepsilon_j + \beta_{\tau_k} X_{\tau_n} \varepsilon_{\tau_k}.$$

Then, from (3.19) and (3.21) we obtain, that

$$|\theta_H^* - \theta| \leq \frac{1}{\sum_{k=1}^{\sigma_H} \mathbf{b}_k^2} \sum_{k=1}^{\sigma_H} \mathbf{b}_k^2 |\tilde{G}_k^{-1}| |\xi_k| \leq \frac{1}{\sum_{k=1}^{\sigma_H} \mathbf{b}_k^2} \sum_{k=1}^{\sigma_H} \mathbf{b}_k \frac{|\xi_k|}{c_k}.$$

Note here, that for any $k \geq 1$ similarly to (3.8) the variance of ξ_k can be estimated from above as

$$\mathbf{E}_\theta |\xi_k|^2 \leq c_k.$$

Therefore, using the Cauchy–Bunyakovsky–Schwarz inequality and the definition of σ_H we obtain that

$$\begin{aligned} \mathbf{E}_\theta |\theta_H^* - \theta|^2 &\leq \mathbf{E}_\theta \left(\frac{1}{\sum_{k=1}^{\sigma_H} \mathbf{b}_k^2} \sum_{k=1}^{\sigma_H} \mathbf{b}_k \frac{|\xi_k|}{c_k} \right)^2 \\ &\leq \frac{1}{H} \sum_{k \geq 1} \frac{\mathbf{E}_\theta |\xi_k|^2}{c_k^2} \leq \frac{1}{H} \sum_{k \geq 1} \frac{1}{c_k}, \end{aligned}$$

i.e. we obtain the bound (3.22). Hence Theorem 3.2. \square

Exercises 3

1. Let us consider the random regression model (2.1) with the predictable sequence $(x_j)_{j \geq 1}$ with respect to the filtration generated by the observations, i.e. $\mathcal{F}_j = \sigma\{y_1, \dots, y_j\}$ for $j \geq 1$. Show, that for this model $\sigma\{y_1, \dots, y_j\} = \sigma\{\varepsilon_1, \dots, \varepsilon_j\}$ for $j \geq 1$.
2. Show, that the rule (3.3) is stopping time for any $H > 0$ with respect to the filtration generated by the observations, defined in Exercice 3.1.
3. Show, that the coefficient β defined in (3.5) is \mathcal{F}_τ measurable.
4. Show, that if in the model (3.2) the parameter $|\theta| < 1$, then

$$\mathbf{P} - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n y_j^2 = \frac{1}{1 - \theta^2}.$$

To conclude from this, that

$$\mathbf{P} - \lim_{H \rightarrow \infty} \frac{\tau_H}{H} = 1 - \theta^2.$$

5. Show, that the random moment N_H defined in (3.21) is a stopping time.
6. Construct the sequential estimator (3.21) for the parameter μ in the model

$$y_j = \mu + \lambda y_{j-1} + \varepsilon_j,$$

where $|\lambda| < 1$ and $(\varepsilon_j)_{j \geq 1}$ is i.i.d random variables with $\mathbf{E} \varepsilon_j = 0$ and $\mathbf{E} \varepsilon_j^2 = 1$.

4 Hypothesis Testing

In this section we consider the testing problem for the observations which are i.i.d. random variables $(X_j)_{j \geq 1}$ defined on some probability space $(\Omega, \mathcal{F}, \tilde{\mathbf{P}})$ with values in the sample space $(\mathcal{X}, \mathcal{A}, \mu)$, where μ is a σ -finite measure. In this case the filtration $(\mathcal{F}_j)_{j \geq 0}$ is generated by the observations $(X_j)_{j \geq 1}$, i.e.

$$\mathcal{F}_0 = \{\emptyset, \Omega\} \quad \text{and} \quad \mathcal{F}_j = \sigma\{X_1, \dots, X_j\} \quad \text{for } j \geq 1. \quad (4.1)$$

We assume, that the random variables $(X_j)_{j \geq 1}$ have a density f with respect to the measure μ , i.e. for any $\Gamma \in \mathcal{A}$

$$\tilde{\mathbf{P}}(X_j \in \Gamma) = \int_{\Gamma} f(x) d\mu.$$

The density f is unknown, it is only known that it is either f_0 or f_1 , where f_0 and f_1 are $\mathcal{X} \rightarrow \mathbb{R}_+$ known probability densities. The problem is to decide which density of observations $(X_j)_{j \geq 1}$ is true f_0 or f_1 , i.e. one has to study the following hypothesis testing problem

$$\begin{cases} \mathbf{H}_0 : & f = f_0; \\ \mathbf{H}_1 : & f = f_1. \end{cases} \quad (4.2)$$

In the classical setting one needs to accept or reject \mathbf{H}_0 or \mathbf{H}_1 on the basis of the observations (X_1, \dots, X_n) .

Definition 4.1. We call any measurable $\mathcal{X}^n \rightarrow \{0, 1\}$ mapping \mathbf{d}_n a statistical test for testing between hypotheses \mathbf{H}_0 and \mathbf{H}_1 .

We denote by $\mathbf{P}_0^{(n)}$ and $\mathbf{P}_1^{(n)}$ the probability measures on $\mathcal{A}^n = \mathcal{A} \otimes \dots \otimes \mathcal{A}$ corresponding to the densities f_0 and f_1 , which for any $\Gamma \in \mathcal{A}^n$ are defined as

$$\mathbf{P}_\nu^{(n)}(\Gamma) = \int_{\Gamma} f_\nu^{(n)}(x) \mu^{(n)}(dx), \quad \nu = \overline{0, 1}, \quad (4.3)$$

where the density $f_\nu^{(n)}(x) = \prod_{j=1}^n f_\nu(x_j)$, $\mu^{(n)}(dx) = \mu(dx_1) \cdots \mu(dx_n)$ and the vector $x = (x_1, \dots, x_n) \in \mathcal{X}^n$. According to the statistical test we accept \mathbf{H}_0 if $\mathbf{d}_n = 0$ and \mathbf{H}_1 for $\mathbf{d}_n = 1$. The quality of a hypothesis test \mathbf{d}_n can be measured by the following error probabilities

$$\mathbf{P}_0^{(n)}(\mathbf{d}_n = 1) \quad \text{and} \quad \mathbf{P}_1^{(n)}(\mathbf{d}_n = 0). \quad (4.4)$$

The probability $\mathbf{P}_0^{(n)}(\mathbf{d}_n = 1)$ is called the *error probability of type I* or *size of test* and $\mathbf{P}_1^{(n)}(\mathbf{d}_n = 0)$ is called the *error probability of type II*. Moreover, the probability $\mathbf{P}_1^{(n)}(\mathbf{d}_n = 1) = 1 - \mathbf{P}_1^{(n)}(\mathbf{d}_n = 0)$ is called *power* of the test.

Now, to construct the test function we need the following condition.

C₀) The density f_0 and f_1 are positive on \mathcal{X} , i.e. $f_0(x) > 0$ and $f_1(x) > 0$ for all $x \in \mathcal{X}$ and $\mu(x \in \mathcal{X} : f_1(x) \neq f_0(x)) > 0$.

Remark 4.1. Note, that we can always to reduce the sample space \mathcal{X} to the set defined as $\{x \in \mathcal{X} : \min(f_0(x), f_1(x)) > 0\}$. Moreover, it should be noted also, that if $\mu(x \in \mathcal{X} : f_1(x) \neq f_0(x)) = 0$, then $\mathbf{P}_0^{(n)} = \mathbf{P}_1^{(n)}$ and the hypotheses \mathbf{H}_0 and \mathbf{H}_1 are the same. Therefore, Condition \mathbf{C}_0 is not restrictable.

For the problem (4.2) we construct the Neyman – Pearson test procedure (see, for example, in [4, 7]). To this end we set the likelihood as

$$\Lambda_n = \prod_{j=1}^n \frac{f_1(X_j)}{f_0(X_j)} \quad \text{and} \quad \lambda_n = \ln \Lambda_n = \sum_{j=1}^n z_j, \quad (4.5)$$

where i.i.d. variables $z_j = \ln f_1(X_j)/f_0(X_j)$. Now we set

$$\mathbf{d}_n^* = \begin{cases} 1, & \text{if } \Lambda_n \geq \mathbf{c}; \\ 0, & \text{if } \Lambda_n < \mathbf{c}, \end{cases} \quad (4.6)$$

where $\mathbf{c} > 0$ is threshold which will be specified later.

Theorem 4.1. Assume, that for $0 < \alpha < 1$ there exists $\mathbf{c} = \mathbf{c}_\alpha$ such, that

$$\mathbf{P}_0^{(n)}(\Lambda_n \geq \mathbf{c}_\alpha) = \alpha. \quad (4.7)$$

Then the test function (4.6) with the threshold defined in (4.7) has the maximal test power in the class of all test procedures having the error probability of type I less than α .

Proof. Indeed, let \mathbf{d}_n test function for which $\mathbf{P}_0^{(n)}(\mathbf{d}_n = 1) \leq \alpha$. We set

$$\mathcal{S}_1 = \{x \in \mathcal{X}^n : \mathbf{d}_n > \mathbf{d}_n^*\} \quad \text{and} \quad \mathcal{S}_2 = \{x \in \mathcal{X}^n : \mathbf{d}_n < \mathbf{d}_n^*\}.$$

Now, setting $\Delta_n = \mathbf{P}_1^{(n)}(\mathbf{d}_n^* = 1) - \mathbf{P}_1^{(n)}(\mathbf{d}_n = 1)$ and using the definitions (4.3), we can represent this difference as

$$\begin{aligned} \Delta_n &= \mathbf{E}_1^{(n)}(\mathbf{d}_n^* - \mathbf{d}_n) = \int_{\mathcal{S}_1 \cup \mathcal{S}_2} (\mathbf{d}_n^* - \mathbf{d}_n) d\mathbf{P}_1^{(n)} \\ &= \int_{\mathcal{S}_1 \cup \mathcal{S}_2} (\mathbf{d}_n^*(x) - \mathbf{d}_n(x)) f_1^{(n)}(x) \mu^{(n)}(dx). \end{aligned}$$

Moreover, note here, that the function

$$(\mathbf{d}_n^*(x) - \mathbf{d}_n(x)) (f_1^n(x) - \mathbf{c}_\alpha f_0^n(x)) \geq 0$$

for any $x \in \mathcal{S}_1 \cup \mathcal{S}_2$. Therefore,

$$\begin{aligned} \Delta_n &\geq \mathbf{c}_\alpha \int_{\mathcal{S}_1 \cup \mathcal{S}_2} (\mathbf{d}_n^*(x) - \mathbf{d}_n(x)) f_0^{(n)}(x) \mu^{(n)}(dx) \\ &= \mathbf{c}_\alpha \int_{\mathcal{X}^n} (\mathbf{d}_n^*(x) - \mathbf{d}_n(x)) f_0^{(n)}(x) \mu^{(n)}(dx) \\ &= \mathbf{c}_\alpha (\mathbf{E}_0^{(n)} \mathbf{d}_n^* - \mathbf{E}_0^{(n)} \mathbf{d}_n). \end{aligned}$$

Taking into account the condition (4.7), we obtain that the last term in the right side of this inequality can be estimated from below as

$$\mathbf{E}_0^{(n)} \mathbf{d}_n^* - \mathbf{E}_0^{(n)} \mathbf{d}_n = \alpha - \mathbf{P}_0^{(n)}(\mathbf{d}_n = 1) \geq 0.$$

Hence Theorem 4.1. \square

Exercises 5

1. Show, that under Condition \mathbf{C}_0) the probability measure $\mathbf{P}_1^{(n)}$ is absolutely continuous with respect to the measure $\mathbf{P}_0^{(n)}$ and the random variable Λ_n defined in (4.5) is the Radon - Nikodym derivative $d\mathbf{P}_1^{(n)}/\mathbf{P}_0^{(n)}$.
2. Write the the Neyman - Pearson test for

$$\mathbf{H}_0 : \mathcal{N}(0, 1) \quad \text{and} \quad \mathbf{H}_1 : \mathcal{N}(m, 1),$$

where $\mathcal{N}(\mu, \sigma^2)$ is the gaussian distribution with the parameters (μ, σ^2) and $m \neq 0$.

3. Write the the Neyman - Pearson test for

$$\mathbf{H}_0 : \mathcal{E}(\lambda_0) \quad \text{and} \quad \mathbf{H}_1 : \mathcal{E}(\lambda_1),$$

where $\mathcal{E}(\lambda)$ is the exponential distribution with the parameter $\lambda > 0$ and $\lambda_0 \neq \lambda_1$.

5 Sequential Hypothesis Testing

In this section we consider the problem (4.2) in the sequential analysis setting proposed by Wald in [12], i.e. we don't fix the number of observations in advance the number of observations, but we chose it as a stopping time with respect to filtration (4.1). Therefore, for this problem we need to use in the sample space the field generated by all observations, i.e. $\mathcal{A}_\infty = \sigma \{ \cup_{n \geq 1} \mathcal{A}^n \}$. Moreover, in this case we denote by \mathbf{P}_0 and \mathbf{P}_1 the probability measures on \mathcal{A}_∞ defined by its finite dimension distributions $(\mathbf{P}_0^{(n)})_{n \geq 1}$ and $(\mathbf{P}_1^{(n)})_{n \geq 1}$. In this case we need to construct a sequential procedure $\delta = (T, \mathbf{d}_T)$ in which the number of observations T is stopping time with respect to the filtration (4.1) and \mathbf{d}_T is a test function, i.e. \mathcal{F}_T measurable random variable with value in the set $\{0, 1\}$. The problem is to find a sequential procedure with the minimal mean observations and the bounded I and II type errors probabilities. To

this end for some fixed $0 < \alpha_0, \alpha_1 < 1$ we introduce the class of sequential procedures

$$\begin{aligned} \mathcal{C}(\alpha_0, \alpha_1) = \{ \delta : \mathbf{P}_0(d_T = 1) \leq \alpha_0, \quad \mathbf{P}_1(d_T = 0) \leq \alpha_1 \\ \mathbf{P}_0(T < \infty) = \mathbf{P}_1(T < \infty) = 1 \} . \end{aligned} \quad (5.1)$$

The Wald procedure $(\tau^*, \mathbf{d}_{\tau^*}^*)$ for fixed thresholds $0 < A_0 < 1 < A_1$ is defined as

$$\tau^* = \inf \{ n \geq 0 : \Lambda_n \notin (A_0, A_1) \} = \inf \{ n \geq 0 : \lambda_n \notin (-a_0, a_1) \} , \quad (5.2)$$

where $\inf\{\emptyset\} = +\infty$, the sequences Λ_n and λ_n are defined in (4.5) and the thresholds $a_0 = -\ln A_0$, $a_1 = \ln A_1$. In this case the test decision rule is defined as

$$\mathbf{d}_{\tau^*}^* = \begin{cases} 1 & \text{if } \lambda_{\tau^*} \geq a_1 ; \\ 0 & \text{if } \lambda_{\tau^*} \leq -a_0 . \end{cases} \quad (5.3)$$

It should be noted, that under Condition \mathbf{C}_0)

$$\mathbf{P}_0(z_1 = 0) = \int_{\{x \in \mathcal{X} : f_0(x) = f_1(x)\}} f_0(x) \mu(dx) < 1$$

and

$$\mathbf{P}_1(z_1 = 0) = \int_{\{x \in \mathcal{X} : f_0(x) = f_1(x)\}} f_1(x) \mu(dx) < 1 .$$

Therefore, in view of Lemma A.1 we obtain that for any a_0 and a_1

$$\mathbf{E}_0 \tau^* < \infty \quad \text{and} \quad \mathbf{E}_1 \tau^* < \infty .$$

Now we study the type I and II error probabilities

$$q_0^* = \mathbf{P}_0(\mathbf{d}_{\tau^*}^* = 1) \quad \text{and} \quad q_1^* = \mathbf{P}_1(\mathbf{d}_{\tau^*}^* = 0) . \quad (5.4)$$

As we already seen, the random variables Λ_n is the Radon - Nikodym density $d\mathbf{P}_1/d\mathbf{P}_0$ on the field \mathcal{F}_n . This means, that for any \mathcal{F}_n measurable bounded random variable η_n , i.e. $\eta_n = g_n(X_1, \dots, X_n)$ and g_n is a $\mathcal{X}^n \rightarrow \mathbb{R}$ function, we get using the definitions (4.3), that

$$\mathbf{E}_1 \eta_n = \int_{\mathcal{X}^n} g_n(x) f_1^n(x) \mu^{(n)}(dx) = \mathbf{E}_0 \eta_n \Lambda_n .$$

Therefore, for any stopping time T any bounded random variable $\eta_T = g_T(X_1, \dots, X_T)$ we can conclude that

$$\mathbf{E}_1 \eta_T = \sum_{n=0}^{\infty} \mathbf{E}_1 \eta_n \mathbf{1}_{\{T=n\}} = \sum_{n=0}^{\infty} \mathbf{E}_0 \eta_n \Lambda_n \mathbf{1}_{\{T=n\}} = \mathbf{E}_0 \eta_T \Lambda_T .$$

Similarly, we can obtain, that $\mathbf{E}_0 \eta_T = \mathbf{E}_1 \eta_T \Lambda_T^{-1}$. Therefore, asymptotically as $a_0, a_1 \rightarrow \infty$,

$$q_1^* = \mathbf{E}_1 \mathbf{1}_{\{\mathbf{d}_{\tau^*}^* = 0\}} = \mathbf{E}_0 e^{\lambda_{\tau^*}} \mathbf{1}_{\{\mathbf{d}_{\tau^*}^* = 0\}} \approx e^{-a_0} \mathbf{P}_0(\mathbf{d}_{\tau^*}^* = 0) = e^{-a_0} (1 - q_0^*)$$

and

$$q_0^* = \mathbf{E}_0 \mathbf{1}_{\{\mathbf{d}_{\tau^*}^* = 1\}} = \mathbf{E}_1 e^{-\lambda_{\tau^*}} \mathbf{1}_{\{\mathbf{d}_{\tau^*}^* = 1\}} \approx e^{-a_1} \mathbf{P}_1(\mathbf{d}_{\tau^*}^* = 1) = e^{-a_1} (1 - q_1^*).$$

This implies, that, asymptotically, as $a_0, a_1 \rightarrow \infty$,

$$q_0^* \approx \frac{e^{a_0} - 1}{e^{a_0 + a_1} - 1} \quad \text{and} \quad q_1^* \approx \frac{e^{a_1} - 1}{e^{a_0 + a_1} - 1}.$$

Therefore, setting in the Wald procedure (5.2),

$$a_0 = \ln \frac{1 - \alpha_0}{\alpha_1} \quad \text{and} \quad a_1 = \ln \frac{1 - \alpha_1}{\alpha_0}, \quad (5.5)$$

we obtain, that that the errors (5.4) satisfy the following properties

$$\lim_{\alpha_0 + \alpha_1 \rightarrow 0} \frac{q_0^*}{\alpha_0} = 1 \quad \text{and} \quad \lim_{\alpha_0 + \alpha_1 \rightarrow 0} \frac{q_1^*}{\alpha_1} = 1.$$

Therefore, asymptotically, as $\alpha_0 + \alpha_1 \rightarrow 0$, the Wald procedure $\delta^* = (\tau^*, \mathbf{d}_{\tau^*}^*)$ belongs to the class (5.1).

5.1 Properties of $\mathbf{E}_0 \tau^*$

Now we define the Kullback informations

$$I_0 = -\mathbf{E}_0 z_1 = -\mathbf{E}_0 \ln \frac{f_1(X_1)}{f_0(X_1)} \quad (5.6)$$

and

$$I_1 = \mathbf{E}_1 z_1 = \mathbf{E}_1 \ln \frac{f_1(X_1)}{f_0(X_1)}. \quad (5.7)$$

We assume that

$$I_0 > 0 \quad \text{and} \quad I_1 > 0. \quad (5.8)$$

Note, that the conditions (5.8) imply that $\mathbf{P}_0(z_0 = 0) < 1$ and $\mathbf{P}_1(z_1 = 0) < 1$. Therefore, by the Stein lemma we obtain that

$$\mathbf{E}_0 \tau^* < \infty \quad \text{and} \quad \mathbf{E}_1 \tau^* < \infty.$$

To study the properties of the mean time $\mathbf{E}_0 \tau^*$ note that in view of Wald's identity

$$\mathbf{E}_0 \lambda_{\tau^*} = \mathbf{E}_0 \sum_{j=1}^{\tau^*} z_j = \mathbf{E}_0 z_1 \mathbf{E}_0 \tau^* = -I_0 \mathbf{E}_0 \tau^* \quad (5.9)$$

and

$$\mathbf{E}_1 \lambda_{\tau^*} = \mathbf{E}_1 \sum_{j=1}^{\tau^*} z_j = \mathbf{E}_1 z_1 \mathbf{E}_0 \tau^* = I_1 \mathbf{E}_1 \tau^*. \quad (5.10)$$

Let us calculate now the expectation $\mathbf{E}_0 \lambda_{\tau^*}$. We have

$$\begin{aligned} \mathbf{E}_0 \lambda_{\tau^*} &= \mathbf{E}_0 \lambda_{\tau^*} \mathbf{1}_{\{\lambda_{\tau^*} \leq -a_0\}} + \mathbf{E}_0 \lambda_{\tau^*} \mathbf{1}_{\{\lambda_{\tau^*} \geq a_1\}} \\ &\approx -a_0 \mathbf{P}_0(\lambda_{\tau^*} \leq -a_0) + a_1 \mathbf{P}_1(\lambda_{\tau^*} \geq a_1) \\ &= -a_0(1 - q_0^*) + a_1 q_0^* \approx -a_0(1 - \alpha_0) + a_1 \alpha_0 \end{aligned}$$

Using here, that

$$a_0 = \ln \frac{1 - \alpha_0}{\alpha_1} \quad \text{and} \quad a_1 = \ln \frac{1 - \alpha_1}{\alpha_0},$$

we obtain

$$\mathbf{E}_0 \lambda_{\tau^*} \approx -(1 - \alpha_0) \ln \frac{1 - \alpha_0}{\alpha_1} + \alpha_0 \ln \frac{1 - \alpha_1}{\alpha_0} = -\beta(\alpha_0, \alpha_1),$$

where

$$\beta(x, y) = (1 - x) \ln \frac{1 - x}{y} + x \ln \frac{x}{1 - y}. \quad (5.11)$$

Note here that, the function $V(x) = -\ln x$ is convex, i.e. for any $0 < \theta < 1$, b_1 and b_2

$$V((1 - \theta)b_1 + \theta b_2) \leq (1 - \theta)V(b_1) + \theta V(b_2).$$

So, using this property with

$$\theta = x, \quad b_1 = \frac{y}{1 - x} \quad \text{and} \quad b_2 = \frac{1 - y}{x}.$$

we obtain

$$\begin{aligned} \beta(x, y) &= (1 - \theta)V(b_1) + \theta V(b_2) \\ &\geq V((1 - \theta)b_1 + \theta b_2) \\ &= V(1 - y + y) = -\ln 1 = 0. \end{aligned}$$

Finally, asymptotically, as $\alpha_0 + \alpha_1 \rightarrow 0$, we obtain, that

$$\mathbf{E}_0 \tau^* = \frac{\mathbf{E}_0 \lambda_{\tau^*}}{I_0} \approx \frac{\beta(\alpha_0, \alpha_1)}{I_0}. \quad (5.12)$$

5.2 Properties of $\mathbf{E}_1 \tau^*$

Let us calculate now the expectation $\mathbf{E}_1 \lambda_{\tau^*}$. We have

$$\begin{aligned} \mathbf{E}_1 \lambda_{\tau^*} &= \mathbf{E}_1 \lambda_{\tau^*} \mathbf{1}_{\{\lambda_{\tau^*} \leq -a_0\}} + \mathbf{E}_1 \lambda_{\tau^*} \mathbf{1}_{\{\lambda_{\tau^*} \geq a_1\}} \\ &\approx -a_0 \mathbf{P}_1(\lambda_{\tau^*} \leq -a_0) + a_1 \mathbf{P}_1(\lambda_{\tau^*} \geq a_1) \\ &= -a_0 q_1^* + a_1(1 - q_1^*) \approx -a_0 \alpha_1 + a_1(1 - \alpha_1). \end{aligned}$$

Using here, that

$$a_0 = \ln \frac{1 - \alpha_0}{\alpha_1} \quad \text{and} \quad a_1 = \ln \frac{1 - \alpha_1}{\alpha_0},$$

we obtain

$$\mathbf{E}_1 \lambda_{\tau^*} \approx (1 - \alpha_1) \ln \frac{1 - \alpha_1}{\alpha_0} + \alpha_1 \ln \frac{\alpha_1}{1 - \alpha_0} = \beta(\alpha_1, \alpha_0),$$

where the function β is defined in (5.11). Finally, we obtain that, asymptotically, as $\alpha_0 + \alpha_1 \rightarrow 0$,

$$\mathbf{E}_1 \tau^* = \frac{\mathbf{E}_1 \lambda_{\tau^*}}{I_1} \approx \frac{\beta(\alpha_1, \alpha_0)}{I_1}. \quad (5.13)$$

Now we have to study the lower bounds for arbitrary procedure.

5.3 Optimality properties

First we show the following lemma.

Lemma 5.1. *Let $(\mathcal{X}, \mathcal{A})$ be a measurable space with two equivalent probability measures \mathbf{P} and \mathbf{Q} , i.e. $\mathbf{P} \sim \mathbf{Q}$. Then for any set $\Gamma \in \mathcal{A}$*

$$\mathbf{E} \ln \rho \geq \mathbf{P}(\Gamma) \ln \frac{\mathbf{P}(\Gamma)}{\mathbf{Q}(\Gamma)} + (1 - \mathbf{P}(\Gamma)) \ln \frac{\mathbf{P}(\Gamma^c)}{\mathbf{Q}(\Gamma^c)} \quad (5.14)$$

where

$$\mathbf{E} g = \int_{\mathcal{X}} g(x) d\mathbf{P} \quad \text{and} \quad \rho(x) = \frac{d\mathbf{P}}{d\mathbf{Q}}(x)$$

is the Radon - Nykodym derivative.

Proof. First, note that

$$\begin{aligned} \mathbf{E} \ln \rho &= \mathbf{E} \mathbf{1}_{\Gamma} \ln \rho + \mathbf{E} \mathbf{1}_{\Gamma^c} \ln \rho \\ &= \mathbf{P}(\Gamma) \int_{\Gamma} \ln \rho(x) d\mathbf{P}_1 + \mathbf{P}(\Gamma^c) \int_{\Gamma^c} \ln \rho(x) d\mathbf{P}_2, \end{aligned}$$

where

$$\mathbf{P}_1(A) = \frac{\mathbf{P}(A \cap \Gamma)}{\mathbf{P}(\Gamma)} = \mathbf{P}(A|\Gamma)$$

and

$$\mathbf{P}_2(A) = \frac{\mathbf{P}(A \cap \Gamma^c)}{\mathbf{P}(\Gamma^c)} = \mathbf{P}(A|\Gamma^c).$$

Moreover, note that

$$\int_{\Gamma} \ln \rho(x) d\mathbf{P}_1 = \mathbf{E}(\ln \rho|\Gamma) = -\mathbf{E}(\ln \rho^{-1}|\Gamma)$$

and, therefore,

$$\mathbf{E} \ln \rho = \mathbf{P}(\Gamma)\mathbf{E}(\ln \rho|\Gamma) + \mathbf{P}(\Gamma^c)\mathbf{E}(\ln \rho|\Gamma^c). \quad (5.15)$$

Note here, that the function \ln is concave, i.e. for any $0 \leq \alpha \leq 1$, $x > 0$ and $y > 0$

$$\ln(\alpha x + (1 - \alpha)y) \geq \alpha \ln x + (1 - \alpha) \ln y.$$

Therefore, by the Jensen inequality

$$\mathbf{E}(\ln \rho^{-1}|\Gamma) \geq \ln \mathbf{E}(\rho^{-1}|\Gamma) = \ln \frac{1}{\mathbf{P}(\Gamma)} \int_{\Gamma} \frac{d\mathbf{Q}}{d\mathbf{P}} d\mathbf{P} = \ln \frac{\mathbf{Q}(\Gamma)}{\mathbf{P}(\Gamma)},$$

i.e. for any $\Gamma \in \mathcal{A}$

$$\mathbf{E}(\ln \rho|\Gamma) \geq \ln \frac{\mathbf{P}(\Gamma)}{\mathbf{Q}(\Gamma)}.$$

Therefore, using this inequality in (5.15) with Γ^c we obtain the bound (5.14). Hence lemma 5.1.

Using this lemma we show now the following

Theorem 5.1. *Assume that the conditions (5.8) hold. Then, the class of the sequential decisions (5.1) for any $0 < \alpha_0 + \alpha_1 < 1$ admits the following lower bounds*

$$\inf_{\delta \in \mathcal{C}(\alpha_0, \alpha_1)} \mathbf{E}_0 T \geq \frac{\beta(\alpha_0, \alpha_1)}{I_0} \quad (5.16)$$

and

$$\inf_{\delta \in \mathcal{C}(\alpha_0, \alpha_1)} \mathbf{E}_1 T \geq \frac{\beta(\alpha_1, \alpha_0)}{I_1}, \quad (5.17)$$

where the function β is defined in (5.11).

Proof. Let $\delta = (T, d_T)$ be some fixed sequential procedure from $\mathcal{C}(\alpha_0, \alpha_1)$, i.e.

$$q_0(\delta) = \mathbf{P}_0(d_T = 1) \leq \alpha_0 \quad \text{and} \quad q_1(\delta) = \mathbf{P}_1(d_T = 0) \leq \alpha_1. \quad (5.18)$$

We start with the inequality (5.16). Assume that $\mathbf{E}_0 T < \infty$. If non, this inequality is obvious. We use now Lemma 5.1 on the probability space $(\Omega, \mathcal{F}_T, \mathbf{P}, \mathbf{Q})$ with

$$\mathbf{P} = \mathbf{P}_0|_{\mathcal{F}_T}, \quad \mathbf{Q} = \mathbf{P}_1|_{\mathcal{F}_T} \quad \text{and} \quad \Gamma = \{\omega \in \Omega : d_T = 0\},$$

where the probabilities $\mathbf{P}_1|_{\mathcal{F}_T}$ and $\mathbf{P}_0|_{\mathcal{F}_T}$ are reductions of these probabilities on the field \mathcal{F}_T . Note that, in this case

$$\rho = \frac{d\mathbf{P}}{d\mathbf{Q}} = \frac{d\mathbf{P}_0}{d\mathbf{P}_1}|_{\mathcal{F}_T} = \Lambda_T^{-1},$$

where the density λ_n is defined in (4.5). Therefore, in view of Lemma 5.14

$$\begin{aligned} \mathbf{E}_0 \lambda_T &= -\mathbf{E}_0 \ln \rho \leq -\mathbf{P}(\Gamma) \ln \frac{\mathbf{P}(\Gamma)}{\mathbf{Q}(\Gamma)} - \mathbf{P}(\Gamma^c) \ln \frac{\mathbf{P}(\Gamma^c)}{\mathbf{Q}(\Gamma^c)} \\ &= -\mathbf{P}_0(d_T = 0) \ln \frac{\mathbf{P}_0(d_T = 0)}{\mathbf{P}_1(d_T = 0)} - \mathbf{P}_0(d_T = 1) \ln \frac{\mathbf{P}_0(d_T = 1)}{\mathbf{P}_1(d_T = 1)} \\ &= -(1 - q_0(\delta)) \ln \frac{(1 - q_0(\delta))}{q_1(\delta)} - q_0(\delta) \ln \frac{q_0(\delta)}{1 - q_1(\delta)} = -\beta(q_0(\delta), q_1(\delta)), \end{aligned}$$

where the function $\beta(\cdot, \cdot)$ is defined in (5.11). Note here, that

$$\beta'_x(x, y) = \ln \frac{xy}{(1-x)(1-y)} < 0 \quad \text{and} \quad \beta'_y(x, y) = \frac{x+y-1}{y(1-y)} < 0$$

for $x + y < 1$. So, taking into account that $\alpha_0 + \alpha_1 < 1$ and the inequalities (5.18), we obtain that

$$\mathbf{E}_0 \lambda_T \leq -\beta(\alpha_0, \alpha_1). \quad (5.19)$$

Note here that, similar to (5.9) through the Wald identity we can get

$$\mathbf{E}_0 \lambda_T = -I_0 \mathbf{E}_0 T$$

and, therefore, from the inequality (5.19) it follows the lower bound (5.16). To show the bound (5.17) use again Lemma 5.1 on the probability space $(\Omega, \mathcal{F}_T, \mathbf{P}, \mathbf{Q})$ with

$$\mathbf{P} = \mathbf{P}_1, \quad \mathbf{Q} = \mathbf{P}_0 \quad \text{and} \quad \Gamma = \{\omega \in \Omega : d_T = 0\}.$$

Then, Lemma 1 yields the following lower bound

$$\begin{aligned} \mathbf{E}_1 \lambda_T &= \mathbf{E}_1 \ln \frac{d\mathbf{P}_1}{d\mathbf{P}_0} \geq \mathbf{P}_1(d_T = 0) \ln \frac{\mathbf{P}_1(d_T = 0)}{\mathbf{P}_0(d_T = 0)} \\ &\quad + \mathbf{P}_1(d_T = 1) \ln \frac{\mathbf{P}_1(d_T = 1)}{\mathbf{P}_0(d_T = 1)} = q_1(\delta) \ln \frac{q_1(\delta)}{1 - q_0(\delta)} \\ &\quad + (1 - q_1(\delta)) \ln \frac{(1 - q_1(\delta))}{q_0(\delta)} = \beta(q_1(\delta), q_0(\delta)). \end{aligned}$$

Taking into account the inequalities (5.18), we get

$$\mathbf{E}_1 \lambda_T \geq \beta(\alpha_1, \alpha_0).$$

Using here again the Wald identity we obtain

$$\mathbf{E}_1 \lambda_T = I_1 \mathbf{E}_1 T$$

and, therefore, the bound (5.17). Hence Theorem 5.1. \square

Theorem 5.2. *Assume that the conditions (5.8) hold. Then, the Wald procedure $\delta^* = (\tau^*, d_{\tau^*})$ defined in (5.3) and (5.4) is optimal in the minimum mean time sense among the sequential procedures defined in (5.1) as $\alpha_0 + \alpha_1 \rightarrow 0$, i.e.*

$$\lim_{\alpha_0 + \alpha_1 \rightarrow 0} \frac{\inf_{\delta \in \mathcal{C}(\alpha_0, \alpha_1)} \mathbf{E}_0 T}{\mathbf{E}_0 \tau^*} = 1 \quad (5.20)$$

and

$$\lim_{\alpha_0 + \alpha_1 \rightarrow 0} \frac{\inf_{\delta \in \mathcal{C}(\alpha_0, \alpha_1)} \mathbf{E}_1 T}{\mathbf{E}_1 \tau^*} = 1. \quad (5.21)$$

Proof. This theorem directly follows from the asymptotic properties (5.12) – (5.13), Theorem 5.1 and the fact that the the function $\beta(\alpha_0, \alpha_1) \rightarrow +\infty$ and $\beta(\alpha_1, \alpha_0) \rightarrow +\infty$ as $\alpha_0 + \alpha_1 \rightarrow 0$. \square

Remark 5.1. *It should be noted that Theorem 5.2 means that the Wald rule gives for the Neyman-Pearson procedure the minimal mean observations number which provides errors (5.4) less than the sufficiently small fixed levels $0 < \alpha_0, \alpha_1 < 1$.*

Exercises 5

Write the Wald procedure and calculate the mean times $\mathbf{E}_0 \tau^*$ and $\mathbf{E}_1 \tau^*$ for the following problems

1.

$$\mathbf{H}_0 : \mathcal{N}(0, 1) \quad \text{and} \quad \mathbf{H}_1 : \mathcal{N}(m, 1),$$

where $\mathcal{N}(\mu, \sigma^2)$ is the gaussian distribution with the parameters (μ, σ^2) and $m \neq 0$.

2.

$$\mathbf{H}_0 : \mathcal{E}(\lambda_0) \quad \text{and} \quad \mathbf{H}_1 : \mathcal{E}(\lambda_1),$$

where $\mathcal{E}(\lambda)$ is the exponential distribution with the parameter $\lambda > 0$ and $\lambda_0 \neq \lambda_1$.

3.

$$\mathbf{H}_0 : \text{Bern}(p_0) \quad \text{and} \quad \mathbf{H}_1 : \text{Bern}(p_1),$$

where $\text{Bern}(p)$ is the Bernoulli distribution with $0 < p < 1$ and $p_0 \neq p_1$.

4.

$$\mathbf{H}_0 : \text{Bin}_m(p_0) \quad \text{and} \quad \mathbf{H}_1 : \text{Bin}_m(p_1),$$

where $\text{Bin}_m(p)$ is the Bernoulli distribution with the parameters $m \geq 1$ and $0 < p < 1$ and $p_0 \neq p_1$.

5.

$$\mathbf{H}_0 : \text{Gm}(p_0) \quad \text{and} \quad \mathbf{H}_1 : \text{Gm}(p_1),$$

where Gm is the geometric distribution with the parameter $0 < p < 1$ and $p_0 \neq p_1$.

A Appendix

A.1 Stein lemma

Let $(Y_j)_{j \geq 1}$ be i.i.d. random variables and

$$\tau_{a_0, a_1} = \inf\{n \geq 1 : S_n \notin [-a_0, a_1]\},$$

where $S_n = \sum_{j=1}^n Y_j$, $a_0 > 0$ and $a_1 > 0$ are some fixed constants.

Lemma A.1. *If $\mathbf{P}(Y_1 = 0) < 1$, then for any $a_0 > 0$ and $a_1 > 0$ there exist $0 < \varrho < 1$ and $\mathbf{c} > 0$ such that for any $n \geq 1$*

$$\mathbf{P}\left(\tau_{a_0, a_1} > n\right) < \mathbf{c} \varrho^n.$$

Proof. Note, that if $\mathbf{P}(Y_1 = 0) < 1$, then there exists $y_0 > 0$ such that $\mathbf{P}(Y_1 \geq y_0) = \epsilon > 0$. If non,, we can always to pass to the sequence $-Y_j$. Let now $m \geq 1$ such, that $m y_0 > a_0 + a_1$. Therefore,

$$\begin{aligned} \mathbf{P}(S_m > a_0 + a_1) &\geq \mathbf{P}(S_m \geq m y_0) \\ &\geq \mathbf{P}(Y_1 \geq y_0, \dots, Y_m \geq y_0) = \epsilon^m. \end{aligned}$$

Moreover, note, that for any $k \geq 1$

$$\mathbf{P}(\tau_{a_0, a_1} > m k) = \mathbf{P}\left(\bigcap_{n=1}^{m k} \{-a_0 \leq S_n \leq a_1\}\right) \leq \mathbf{P}(\Gamma_k),$$

where

$$\Gamma_k = \bigcap_{l=1}^k D_l \quad \text{and} \quad D_l = \{-a_0 \leq S_{ml} \leq a_1\}.$$

Note here, that for any $l \geq 2$ the sum S_{ml} can be represented as

$$S_{ml} = S_{(m-1)l} + \tilde{S}_{m,l} \quad \text{and} \quad \tilde{S}_{m,l} = \sum_{j=1}^m Y_{(l-1)m+j}.$$

This means, that for any $l \geq 2$ the intersection

$$D_{l-1} \cap D_l \subseteq D_{l-1} \cap \{|\tilde{S}_{m,l}| \leq a_0 + a_1\},$$

i.e. for any $k \geq 2$

$$\mathbf{P}(\Gamma_k) \leq \mathbf{P}\left(\Gamma_{k-1} \cap \{|\tilde{S}_{m,k}| \leq a_0 + a_1\}\right) = \mathbf{P}(\Gamma_{k-1}) \mathbf{P}\left(|\tilde{S}_{m,k}| \leq a_0 + a_1\right).$$

Taking into account here, that

$$\mathbf{P}\left(|\tilde{S}_{m,k}| \leq a_0 + a_1\right) \leq 1 - \mathbf{P}(|S_m| > a_0 + a_1) \leq 1 - \epsilon^m := q,$$

we get, that for any $k \geq 2$

$$\mathbf{P}(\Gamma_k) \leq \mathbf{P}(\Gamma_{k-1}) q,$$

i.e. for any $k \geq 1$

$$\mathbf{P}(\Gamma_k) \leq \mathbf{P}(\Gamma_1) q^{k-1} \leq q^k.$$

Note here, that, any $n > m$ can be represented as $n = k m + l$ with $k \geq 1$ and $0 \leq l < m$, i.e.

$$\mathbf{P}(\tau_{a_0, a_1} > n) \leq \mathbf{P}(\tau_{a_0, a_1} > k m) \leq q^k \leq q^{-1} \varrho^n,$$

where $\varrho = q^{1/m}$. Taking into account, that for $0 \leq n \leq m$ this upper bound $q^{-1} \varrho^n \geq q^{-1} \varrho^m = 1 \geq \mathbf{P}(\tau_{a_0, a_1} > n)$, we obtain that the last inequality holds true for any $n \geq 0$. Hence Lemma A.1. \square

References

- [1] Anderson, T. W. *The Statistical Analysis of Time Series*. Wiley, 1971
- [2] Borisov, V.Z., Konev, V.V. (1977) Sequential Estimation of Parameters of Discrete Processes. *Automat. and Remote Control*, **10**, p. 58 – 64.
- [3] Basseville, M. and Nikiforov, I.V. *Detection of Abrupt Changes - Theory and Applications*, Information and System Sciences Series, Prentice-Hall, Inc. Englewood Cliffs, NJ, USA, 1993.
- [4] Borovkov, A.A. *Mathematical Statistics*. Gordon and Breach Sciences Publishers, Amsterdam, NL, 1998.
- [5] Chow, Y.S., Robbins, H. and Siegmund, D. *Great Expectations: The Theory of Optimal Stopping*. HOUGHTON MIFFLIN COMPANY BOSTON, New York, Atlanta, Geneva, Illinois, Dallas, Palo Alto, 1971.
- [6] Konev, V.V. and Pergamenschikov, S.M. (1981) Sequential Identification Procedures for the Parameters of Dynamic Systems. *Automat. and Remote Control*, **7**, 84 – 92.
- [7] Lehmann, E. L. *Testing Statistical Hypotheses*. John Wiley & Sons, Inc., New York, USA, 1968.
- [8] Shiryaev, A. N. (1978). *Optimal Stopping Rules*, New York: Springer.
- [9] Shiryaev, A.N. *Probability*. Springer, Berlin, 1996.
- [10] Tartakovsky, A.G. *Sequential Change Detection and Hypothesis Testing: General Non-i.i.d. Stochastic Models and Asymptotically Optimal Rules*, Chapman & Hall/CRC Monographs on Statistics and Applied Probability) 1st Edition., 2019.
- [11] Tartakovsky, A., Nikiforov, I. and Basseville, M. *Sequential Analysis: Hypothesis Testing and Change-point Detection*. Chapman & Hall book, Taylor & Francis Group, 2015.
- [12] Wald, A. *Sequential Analysis*. John Wiley & Sons, Inc., New York, USA, 1947.