Albert N. SHIRYAEV

Steklov Mathematical Institute

Lectures on some specific topics of STOCHASTIC CALCULUS and its METHODS

SDE, Optimal Stopping and Quickest Detection Problems with application to Finance and Control

e-mail: albertsh@mi.ras.ru

TOPIC I: Classical and recent results on the stochastic differential equations

TOPIC II: Optimal Stopping (OS) theory. Basic formulations, concepts and methods of solutions

TOPIC III: Quickest detection problems: Discrete time

TOPIC IV: Quickest detection problems: Continuous time (Brownian model)

TOPIC V: Applications-1: Detection of spontaneously appearing effects

TOPIC VI: Applications-2: Testing two and three hypotheses for Brownian motion with drift

TOPIC VII: Applications-3: Optimal stopping in some problems of the one-time portfolio rebalancing

TOPIC I: Classical and recent results on the stochastic differential equations

TOPIC II: Optimal Stopping (OS) problems. Basic formulations, concepts and methods of solutions

§ 1. Standard and nonstandard optimal stopping problems

§ 2. OS-lecture 1: Introduction

§ 3. OS-lecture 2-3: Theory of optimal stopping for discrete time (finite and infinite horizons)

A) Martingale approach

B) Markovian approach

§ 4. OS-lecture 4-5: Theory of optimal stopping for continuous time (finite and infinite horizons)

A) Martingale approach

B) Markovian approach

TOPIC III: Quickest detection problems

§ 1. W. Shewhart and E. Page's works

§ 2. Definition of the θ-model and Bayesian G-model in the quickest detection

§ 3. Four basic formulations (Variants A, B, C, D) of the quickest detection problems for the Brownian case

§ 4. The reduction of Variants A and B to the standard form

§ 5. Variant C and D: lower estimates for the risk functions

§ 6. Recurrent equations for statistics π_n, φ_n, ψ_n, γ_n

§ 7. Variants A and B: solving the optimal stopping problem

§ 8. Variants C and D: around the optimal stopping times

TOPIC IV: Quickest detection problems: Continuous time (Brownian model)

§ 1. INTRODUCTION

§ 2. Four basic formulations (VARIANTS A, B, C, D) of the quickest detection problems for the Brownian case

§ 3. VARIANT A

§ 4. VARIANT B

§ 5. VARIANT C

§ 6. VARIANT D

TOPIC V: Applications-1: Detection of spontaneously appearing effects

TOPIC I: Classical and recent results on the stochastic differential equations

§ 1. Introduction: Kolmogorov and Itô's papers

1931 A. N. Kolmogorov. "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung", Math. Ann., 104, 415-458.

Dynamic system: $P(s,x;t,A)$ is a transition probability for the system which starts at time s from point x to be at time $t > s$ in a set A.

It is assumed that ("Chapman–Kolmogorov equation")

$$\boxed{P(s,x;t,A) = \int P(s,x;u,dy)\,P(u,y;t,A)} \quad \text{for } 0 \le s \le u < t.$$

Kolmogorov did not use the words “a system is described by a Markov provess $X = (X_t)_{t \geq 0}$ such that

$$\mathsf{P}(X_t \in A \,|\, X_s = x) = P(s, x; t, A)\text{''}.$$

The reason is the following: in 1931 there was no theorem about possibility to construct a process with given system of transition probability, or finite dimensional distributions. It was done in

|| 1933 A. N. Kolmogorov. Grundbegriffe der Wahrscheinlichkeisrechnung, Berlin (Chapter III, § 4).

(English translation: *Foundations of the Theory of Probability*, Chelsea, New York, 1950.)

From what we call now the Chapman–Kolmogorov equation, Kolmogorov deduced

backward differential equation

$$\boxed{-\frac{\partial}{\partial s}f(s,x;t,y) = A(s,x)\,\frac{\partial}{\partial x}f(s,x;t,y) + \frac{B^2(s,x)}{2}\,\frac{\partial^2}{\partial x^2}f(s,x;t,y)}$$

forward differential equation

$$\boxed{\frac{\partial}{\partial t}f(s,x;t,y) = -\frac{\partial}{\partial y}\Big[A(t,y)f(s,x;t,y)\Big] + \frac{1}{2}\,\frac{\partial^2}{\partial y^2}\Big[B^2(t,y)\,f(s,x;t,y)\Big]}$$

[for $f(s,x;t,y) = \dfrac{\partial P(s,x;t,(-\infty,y])}{\partial y}$ with

$$A(s,x) = \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\int_{-\infty}^{\infty}(y-x)f(s,x;s+\Delta,y)\,dy,$$

$$B^2(s,x) = \lim_{\Delta\downarrow 0}\frac{1}{\Delta}\int_{-\infty}^{\infty}(y-x)^2 f(s,x;s+\Delta,y)\,dy,$$

]

In the 1940–50s K. Itô considered the problem of

> an **EXPLICIT CONSTRUCTION** of a Markov process $X = (X_t)_{t \geq 0}$ whose **local drift** and **local variance** coincide with a given function $\boldsymbol{A(s,x)}$ and $\boldsymbol{B^2(s,x)}$

The Itô's idea was the following:

- To take a Wiener (Brownian) process $W = (W_t)_{t\geq 0}$ that can be constructed explicitly

  for example: $W_t = \xi_0 t + \sum_{n=1}^{\infty} \sum_{k=2^{n-1}}^{2^n-1} \xi_k \sqrt{2}\,\frac{\sin(\pi k t)}{\pi k}$ with i.i.d., $N(0,1)$, random variables $\xi_0, \xi_1, \ldots$ (this is a **Paley–Wiener construction)**

  and after this

- to define the process $X = (X_t)_{t\geq 0}$ as a solution of the stochastic differential equation

$$\boxed{dX_t = A(t, X_t)\,dt + B(t, X_t)\,dW_t} \qquad \text{(SDE)}$$

  in the sense that $X_t = X_0 + \int_0^t A(s, X_s)\,ds + \int_0^t B(s, X_s)\,dW_s$ for all $t \geq 0$.

Here we have a "stochastic integral" $\int_0^t B(s, X_s)\, dW_s$ which was defined in

> 1944 K. Itô. "Stochastic integral", Proc. Imp. Acad. Tokyo, 20, 519-524.

In a series of papers (1942–1951) K. Itô

- gave a definition of the notion of the solution of the stochastic differential equation (SDE), and
- gave condition on $A(s,x)$ and $B(s,x)$ which implies existence and uniqueness of the solution of SDE with coefficients $A(s,x)$ and $B(s,x)$.

THEOREM 1. Let $A(s,x)$ and $B(s,x)$ satisfy

- the **local Lipschitz condition**: for any $n \geq 1$

$$\boxed{|A(s,x) - A(s,x')| + |B(s,x) - B(s,x')| \leq C_n}$$

for $|x| \leq n$, $|x'| \leq n$, and $s \leq n$,

- the **condition of linear growth**: for any $n \geq 1$

$$\boxed{|A(s,x)| + |B(s,x)| \leq C_n(1 + |x|)} \quad \text{for } |x| \leq n \text{ and } s \leq n,$$

where C_n are constants.

Then (on any stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P})$ on which the Wiener process is defined) a **strong** solution exists and is unique (up to indistinguishability).

A **strong** solution is a solution $X = (X_t)_{t \geq 0}$ such that X_t is $\mathcal{F}_t$-measurable for each $t \geq 0$.

The Wiener process $W = (W_t)_{t \geq 0}$ is assumed to be defined on a given stochastic basis; in particular, W_t is $\mathcal{F}_t$-measurable for each $t \geq 0$.

$\mathcal{F}_t$ stands for the σ-algebra generated by W_s, $s \leq T$, and by all P-null sets from the σ-algebra $\sigma(W_s,\, s \geq 0)$.

Itô's result about existence of a **strong** solution of the stochastic differential equation (SDE) was extended in

|| 1961 A. V. Skorokhod, Issledovania po teorii sluchajnykh protsessov, Kiev, Izd-vo KGU

(English translation: *Studies in the Theory of Random Processes*, Addison-Wesley, Reading, 1965.)

THEOREM 2 [Skorokhod, Chap. 3, § 3, and Chap. 5, § 4]. Let the coefficients $A(t, x)$ and $B(t, x)$ be bounded and continuous. Then there exists a "solution" * of the (SDE).

---

* We put the word *solution* in the inverted commas, since the Skorokhod solution is not a strong solution. His solution is the so-called weak solution.

STRONG solution of the (SDE).

$$\left.\begin{array}{l}\bullet \text{ a filtered probability space } (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P}) \\ \bullet \text{ an } \mathbb{F} = (\mathcal{F}_t)\text{-Wiener process } W \\ \bullet \text{ an } \mathcal{F}_0\text{-measurable random variable } \xi \end{array}\right\} \quad \text{are given}$$

A strong solution is then defined as an $\mathbb{F}$-adapted process X with $X_0 = \xi$ a.s. which satisfies (SDE).

WEAK solution of the (SDE).

We assume that only the initial distribution μ and coefficients $A(s,x)$ and $B(s,x)$ are given. The weak solution consists of

- a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P})$ (a priori not given),
- an $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$-measurable Wiener process $W = (W_t)_{t\geq 0}$,
- an $\mathbb{F}$-adapted process $X = (X_t)_{t\geq 0}$ with $\mathsf{P} \circ X_0^{-1} = \mu$ which satisfies (SDE), i.e., for all $t \geq 0$

$$X_t = X_0 + \int_0^t A(s, X_s)\, ds + \int_0^t B(s, X_s)\, dW_s \quad \mathsf{P}\text{-a.s.}$$

$\boxed{\text{I-1-10}}$

§ 2. WEAK AND STRONG SOLUTIONS. 1: EXAMPLES OF EXISTENCE AND NONEXISTENCE

EXAMPLE 1 (**M. Barlow** *). There exists a continuous bounded function $B(x)\colon \mathbb{R} \to (0, \infty)$ such that the following SDE has a weak solution but does NOT have any strong solution:

$$\boxed{dX_t = B(X_t)\, dW_t}, \qquad X_0 = x_0$$

Weak solution	**S**trong solution
$+$	$-$

---

* ‖ 1982 M. Barlow, "One-dimensional stochastic differential equations with no strong solution", J. Lond. Math. Soc., 26, 335-347.

EXAMPLE 2. Deterministic SDE:

$$\boxed{dX_t = -\operatorname{sgn} X_t\, dt}, \qquad X_0 = 0,$$

$$\text{where} \quad \operatorname{sgn} x = \begin{cases} 1, & x > 0, \\ -1, & x \le 0, \end{cases}$$

W	S
—	—

Indeed, our equation has the form

$$f(t) = -\int_0^t \operatorname{sgn} f(s)\, ds, \qquad t \ge 0. \tag{$*$}$$

Let us show that this integral equation has no solution.

Let f be a solution. Assume that there exist $a > 0$ and $t > 0$ such that $f(t) = a$. Set

$$\tau = \inf\{t \geq 0\colon f(t) = a\}, \qquad \sigma = \sup\{t \leq \tau\colon f(t) = 0\}.$$

Using the equation $(*)$ we get

$$a = f(\tau) = f(\sigma) = -(\tau - \sigma).$$

The above construction shows that $f \leq 0$. In a similar way we prove that $f \geq 0$. Thus, $f \equiv 0$, but then f is not a solution of $(*)$.

As a result, $(*)$ has no solution.

EXAMPLE 3.

$$\boxed{dX_t = -\frac{1}{2X_t} I(X_t \neq 0)\,dt + dW_t}$$

W	S
—	—

Indeed, suppose that (X, W) is a solution (weak or strong). Then $X_t = -\int_0^t \frac{1}{2X_t} I(X_t \neq 0)\,ds + W_t$. By Itô's formula (that holds for both strong and weak solutions):

$$\begin{aligned} X_t^2 &= -\int_0^t 2X_s \frac{1}{2X_s} I(X_s \neq 0)\,ds + \int_0^t 2X_s\,dB_s + t \\ &= \int_0^t I(X_s = 0)\,ds + \int_0^t 2X_s\,dB_s. \end{aligned}$$

---

* Results of Examples 3 and 4 was obtained in

‖ 1974 A. Zvonkin, "A transformation of the phase space of a process that removes the drift", Math. Sb., 93:1, 129-149.

The process X is a continuous semimartingale with $\langle X \rangle_t = t$. Hence, by the "occupation times formula",

$$\int_0^t I(X_s = 0)\, ds = \int_{\mathbb{R}} I(x = 0) L_t^x(X)\, dx = 0,$$

where $L_t^x(X)$ is the local time of the process X spent at the point x by the time t.

So, X^2 is a positive local martingale and hence a supermartingale. Since $X^2 \geq 0$ and $X_0^2 = 0$, we conclude that $X^2 = 0$ a.s. Hence $X_t^2 = \int_0^t 2X_s\, dB_s$. Then our SDE takes the form $dX_t = dB_t$, a contradiction (since $X_t = 0$ while B_t ia a Brownian motion).

EXAMPLE 4. The following SDE has a (**unique**) strong (and thus weak) solution:

$$\boxed{dX_t = -\,\mathrm{sgn}\, X_t\, dt + dW_t}$$

W	S
+	+

This example is a particular case of the following statement.

Let $a(x)$ be a bounded Borel measurable function. The the following SDE has a (**unique**) strong (and thus weak) solution:

$$\boxed{dX_t = a(X_t)\, dt + dW_t}, \qquad X_0 = x_0$$

W	S
+	+

Our interest to the equation $dX_t = -\operatorname{sgn} X_t\,dt + dW_t$ comes from the following optimization problem.

Consider SDE

$$dX_t^u = u(X_t^u)\,dt + dW_t, \qquad X_0^u = 0,$$

with the Lipschitzian "control" $u = u(x)$, $|u(x)| \le 1$. Put

$$\tau(u) = \inf\{t\colon |X_t^u| \ge 1\}.$$

We want to find an optimal control $u^* = u^*(x)$ such that

$$\mathsf{E}\tau(u^*) = \sup_{u \in \mathcal{L}\mathrm{ip}} \mathsf{E}\tau(u),$$ where $\mathcal{L}$ip is a class of Lipschitzian "controls" $u = u(x)$ with $|u(x)| \le 1$.

Intuitively, it is evident that $u^*(x) = -\operatorname{sgn} x$. However, this function $u^*(x) \notin \mathcal{L}\mathrm{ip}$ and there is a natural question: does exist a solution of the SDE $dX_t = -\operatorname{sgn} X_t\,dt + dW_t$? It turned out, this SDE has a strong solution.

EXAMPLE 5. The "two-sided" Tanaka equation:

$$\boxed{dX_t = \operatorname{sgn} X_t \, dW_t}, \qquad X_0 = 0$$

W	S
$+$	$-$

If X is a solution, then by the Lévy characterization theorem, X is a Wiener process. We find

$$\int_0^t \operatorname{sgn} X_s \, dX_s = \int_0^t \operatorname{sgn}^2 X_s \, dW_s = W_s. \qquad (*)$$

By the Tanaka formula,

$$|X_t| = \int_0^t \operatorname{sgn} X_s \, dX_s + L_t(X), \qquad (**)$$

where $L_t(X)$ is the local time of the process X at zero.

From $(*)$ and $(**)$ if follows that

$$W_t = |X_t| - L_t(X).$$

But $L_t(X) = L_t(|X|)$, since $L_t(X) = \lim\limits_{\varepsilon \downarrow 0} = \frac{1}{2\varepsilon} \int_0^t I(|X_s| \le \varepsilon)\, ds$. Hence

$$\mathcal{F}_t^W \subseteq \mathcal{F}_t^{|X|}. \tag{$***$}$$

if X is a strong solution of the SDE $dX_t = \operatorname{sgn} X_t\, dW_t$, then (by definition of a strong solution)

$$\mathcal{F}_t^X \subseteq \mathcal{F}_t^W. \tag{$****$}$$

From $(***)$ and $(****)$ we get

$$\mathcal{F}_t^X \subseteq \mathcal{F}_t^{|X|},$$

which cannot be true because X is a Wiener process (σ-algebra $\mathcal{F}_t^X$ contains also σ-algebra $\mathcal{F}_t^{\operatorname{sgn} X}$).

To prove the existence of a weak solution, take a Wiener process $\widetilde{W} = (\widetilde{W}_t)_{t\geq 0}$ and a process $W_t = \int_0^t \operatorname{sgn} \widetilde{W}_s \, d\widetilde{W}_s$, $t \geq 0$. Put $X_t = \widetilde{W}_t$, then

$$\operatorname{sgn} X_t \, dW_t = \operatorname{sgn} \widetilde{W}_t \cdot \left(\operatorname{sgn} \widetilde{W}_t \; d\widetilde{W}_t\right) = d\widetilde{W}_t = dX_t.$$

So, the pair $(W_t, X_t)_{t\geq}$, where $(W_t)_{t\geq 0}$ is a Wiener process, solves the equation

$$dX_t = \operatorname{sgn} X_t \, dW_t.$$

This pair $(W_t, X_t)_{t\geq}$ is a weak solution of this equation.

In Examples 6 and 7 we consider the stochastic differential "one-sided" Tanaka equation

$$dX_t = \lambda\, dt + I(X_t > 0)\, dt$$

with $\lambda = 0$ and $\lambda \neq 0$, respectively.

EXAMPLE 6. The "one-sided" Tanaka equation:

$$\boxed{dX_t = I(X_t > 0)\, dW_t}, \qquad X_0 = \xi$$

W	S
+	+

where ξ and W are independent.

PROOF of Example 6. Consider the stopping time $\sigma = \inf\{t \geq 0 \colon \xi + W_t \leq 0\}$ and set

$$X_t = \xi + W_{t\wedge\sigma}. \qquad (\bullet)$$

This process X is a strong solution of SDE $dX_t = I(X_t > 0)\,dW_t$, $X_0 = \xi$. Indeed, it is $(\mathcal{F}_t^{(\xi,W)})_{t\geq 0}$-adapted, $X_0 = \xi$, and

$$X_t - X_0 = \int_0^t I(\sigma > s)\,dW_s = \int_0^t I(\xi + W_{\sigma\wedge s} > 0)\,dW_s = \int_0^t I(X_s > 0)\,dW_s,$$

i.e., $dX_t = I(X_t > 0)\,dW_t$, $X_0 = \xi$.

So, the qualitative properties of

the "one-sided" Tanaka equation $dX_t = I(X_t > 0)\,dW_t$

differs markedly from those of

the "original" Tanaka equation $dX_t = \operatorname{sgn} X_t\,dW_t$.

EXAMPLE 7 (Karatzas–Shiryaev–Shkolnikov). The one-dimensional "one-sided" Tanaka equation with drift:

$$\boxed{dX_t = \lambda\, dt + I(X_t > 0)\, dW_t}, \quad X_0 = \xi,$$

where ξ and W are independent

$\lambda < 0:$

W	S
+	+

$\lambda > 0:$

W	S
+	−

For $\lambda > 0$ the weak solution is **unique**.

⟨For the proof see the recent paper:

I.Karatzas, A.N.Shiryaev, M.Shkolnikov, "On the one-sided Tanaka equation with drift", Electronic Comm. in Probab., 16 (2011), 664–677.⟩

In connection with Examples 6 and 7, the following theorem is interesting.

THEOREM. For any real constant λ, and with W and V standard Brownian motions, the **PERTURBED** one-sided Tanaka equation

$$dX(t) = \lambda\, dt + I(X(t) > 0)\, dW(t) + \frac{\eta}{2} dV(t), \qquad 0 \le t < \infty,$$

has a pathwise unique strong solution, provided either

(i) $\eta \notin [-1, 1]$ and $\langle W, V\rangle(t) = -t/\eta$, $0 \le t < \infty$, or

(ii) $\eta \neq 0$ and W, V are independent.

EXAMPLE 8 **(I.V.Girsanov)**. Let $0 < \alpha < 1/2$.

$$\boxed{dX_t = |X_t|^\alpha \, dB_t}, \qquad X_0 = 0,$$

W	S
+	+

where $B = (B_t)_{t \geq 0}$ is a Brownian motion (Wiener process).

It is clear that the process $X \equiv 0$, $t \geq 0$, solves the SDE. This trivial solution is both strong and weak. The interesting question:

Does a **nontrivial** solution exist **?** The answer is **"YES"**.

The method of the proof suggested by I.V.Girsanov is based on the ideas of the

CHANGE of TIME.

Take $(\Omega, \mathcal{F}) = (C, \mathcal{B})$, where C is the space of continuous functions $\omega = (\omega_t)_{t \geq 0}$. Let $W = (W_t(\omega))_{t \geq 0}$ with $W_t(\omega) = \omega_t$ is a (canonical) Wiener process (w.r.t. the Wiener measure P^W on $(C, \mathcal{B})$ which exists as we know).

Let $A = (A_t(\omega))_{t \geq 0}$ be a process defined by

$$A_t(\omega) = \int_0^t |W_s(\omega)|^{-2\alpha} \, ds, \qquad 0 < \alpha < 1/2. \qquad (\bullet)$$

Define $\widehat{T}(\theta) = \inf\{t \geq 0 \colon A_t > \theta\}$. Since $0 < \alpha < 1/2$, the process A has P^W-a.s. continuous nondecreasing trajectories, $\mathsf{P}^W(A_t < \infty) = 1$ for each $t \geq 0$ and $A_t \to \infty$ P^W-a.s. as $t \to \infty$.

Starting from W and $(\widehat{T}(\theta))_{\theta\geq 0}$, construct

$$\widehat{W}_\theta(\omega) = W_{\widehat{T}(\theta)}(\omega) \quad \text{and} \quad \widehat{\mathcal{F}}_\theta = \mathcal{F}_{\widehat{T}(\theta)}.$$

The process $\widehat{W} = (\widehat{W}_\theta)_{\theta\geq 0}$ is a continuous local martingale w.r.t. $(\widehat{\mathcal{F}}_\theta)_{\theta\geq 0}$ with $\langle \widehat{W}_\theta \rangle = \widehat{T}(\theta)$.

From $(\bullet)$, by the change-of-variable formula, we get

$$\begin{aligned} \widehat{T}(\theta) &= \int_0^{\widehat{T}(\theta)} |W_s|^{-2\alpha}\, dA_s \\ &= \int_0^{A_{\widehat{T}(\theta)}} |W_{\widehat{T}(s)}|^{-2\alpha}\, ds = \int_0^\theta |\widehat{W}_s|^{-2\alpha}\, ds. \end{aligned} \qquad (\bullet\bullet)$$

Introduce the process

$$B_\theta = \int_0^\theta |\widehat{W}_s|^{-\alpha}\, d\widehat{W}_s. \qquad (\bullet\bullet\bullet)$$

(as a stochastic integral w.r.t. a local martingale $\widehat{W}$). This process is an $((\widehat{\mathcal{F}}(\theta))_{\theta\ge 0}, \mathrm{P}^W)$-local martingale with the quadratic characteristic

$$\langle B\rangle_\theta = \int_0^\theta |\widehat{W}_s|^{-2\alpha}\, d\langle\widehat{W}\rangle_s = \int_0^\theta |\widehat{W}_s|^{-2\alpha}\, d\widehat{T}(s)$$
$$\overset{\text{by } (\bullet\bullet)}{=} \int_0^\theta |\widehat{W}_s|^{-2\alpha}|\widehat{W}_s|^{2\alpha}\, ds = \theta.$$

Thus, the local martingale $B = (B_\theta)_{\theta\ge 0}$ has the quadratic characteristic $\langle B\rangle_\theta = \theta$; therefore, by the Lévy characterization theorem, it is a P^W-Brownian motion.

If we change notation: $X_\theta = \widehat{W}_\theta$, then, by $(\bullet\bullet\bullet)$,

$$B_\theta = \int_0^\theta |X_s|^{-\alpha}\, dX_s \qquad \text{and thus} \qquad dX_\theta = |X_\theta|^\alpha\, dB_\theta.$$

Therefore the pair of processes

$$\boxed{(X_\theta, B_\theta)_{\theta\ge 0}, \quad \text{where} \quad \begin{aligned} X_\theta &= W_{\widehat{T}(\theta)} = \widehat{W}_\theta, \\ B_\theta &= \int_0^\theta |\widehat{W}_s|^{-2\alpha}\, d\langle \widehat{W}\rangle_s, \end{aligned}}$$

provides a weak solution of the SDE

$$dX_t = |X_t|^\alpha\, dB_t \qquad (0 < \alpha < 1/2).$$

Return to Example 4.

Let $a(x)$ be a bounded Borel measurable function. The the following SDE has a (**unique**) strong (and thus weak) solution:

$$\boxed{dX_t = \boldsymbol{a(X_t)}\, dt + dW_t}, \qquad X_0 = x_0$$

W	S
+	+

There have been many attempts to extend this result about existence of the strong solution to the "functional" case:

$$dX_t = \boldsymbol{a(t, X)}\, dt + dW_t, \qquad (\star)$$

where for each t the functional $a(t, X)$ depends on the past $\{X_s, s \le t\}$ of the process X.

More exactly, let $a = a(t, x)$ be a bounded (predictable) functional $a(t, x) \colon \mathbb{R}_+ \times C(\mathbb{R}_+) \to \mathbb{R}$. The following example shows that $(\star)$ with past-depending drift term **can have no strong solution**.

§ 3. The TSIRELSON EXAMPLE

Let $x = (x_t)_{0 \le t \le 1} \in C([0,1])$ with $x_0 = 0$. Define the numbers t_k, $k = 0, -1, -2, \dots$ such that $0 < \cdots < t_{-2} < t_{-1} < t_0 = 1$ and consider the function $a(t, x)$ such that $a(0, x) = 0$ and

$$a(t,x) = \left\{ \frac{x_{t_k} - x_{t_{k-1}}}{t_k - t_{k-1}} \right\}, \qquad t_k \le t < t_{k+1},$$

where $\{\alpha\}$ stands for the fractional part of α.

For such a function the SDE $dX_t = a(t, X_t)\,dt + dW_t$ takes the form

$$X_{t_{k+1}} - X_{t_k} = \left\{ \frac{X_{t_k} - X_{t_{k-1}}}{t_k - t_{k-1}} \right\} (t_{k+1} - t_k) + (W_{t_{k+1}} - W_{t_k})$$

which is equivalent to the recurrent equations

$$\eta_{k+1} = \{\eta_k\} + \varepsilon_{k+1}, \quad \text{where} \quad \eta_k = \frac{X_{t_k} - X_{t_{k-1}}}{t_k - t_{k-1}}, \quad \varepsilon_k = \frac{W_{t_k} - W_{t_{k-1}}}{t_k - t_{k-1}}.$$

From this recurrent equation we have

$$e^{2\pi i\eta_{k+1}} = e^{2\pi i\{\eta_{k+1}\}}e^{2\pi i\varepsilon_{k+1}} = e^{2\pi i\eta_k}e^{2\pi i\varepsilon_{k+1}}.$$

Denote $m_k := \mathsf{E}e^{2\pi i\eta_k}$. If our SDE $dX_t = a(t,X)\,dt + dW_t$ with the function $a(t,X)$ defined above has a strong solution then η_k must be $\mathcal{F}^W_{t_k} = \sigma\{W_s, s \le t_k\}$-measurable. Consequently,

$$m_{k+1} = m_k\mathsf{E}e^{2\pi i\varepsilon_{k+1}} = m_k e^{-2\pi^2/(t_{k+1}-t_k)}.$$

Therefore,

$$m_{k+1} = m_{k+1-n}\exp\left\{-2\pi^2\left(\frac{1}{t_{k+1}-t_k} + \cdots + \frac{1}{t_{k+2-n}-t_{k+1-n}}\right)\right\}.$$

Thus, $|m_{k+1}| \le e^{-2\pi^2 n}$ for any n. Hence $m_k = 0$ for any $k = 0, -1, -2, \ldots$.

Further, from the same recurrent equation we have:

$$e^{2\pi i\eta_{k+1}} = e^{2\pi i\eta_k}\, e^{2\pi i\varepsilon_{k+1}} = \dots = e^{2\pi i\eta_{k-n}}\, e^{2\pi i(\varepsilon_{k+1-n}+\dots+\varepsilon_{k+1})}.$$

Assume there exists a strong solution to our SDE. Under this assumption η_{k-n} should be $\mathcal{F}^W_{t_{k-n}}$-measurable. Consequently, if we denote $\mathcal{G}^W_{t_{k-n},t_{k+1}} = \sigma\Big\{\omega\colon W_t - W_s,\ \ t_{k-n} \le s \le t_{k+1}\Big\}$, then the independence of the σ-algebras $\mathcal{F}^W_{t_{k-n}}$ and $\mathcal{G}^W_{t_{k-n},t_{k+1}}$ implies that

$$\mathsf{E}\Big(e^{2\pi i\eta_{k+1}} \,|\, \mathcal{G}^W_{t_{k-n},t_{k+1}}\Big) = e^{2\pi i(\varepsilon_{k+1-n}+\dots+\varepsilon_{k+1})}\, \underbrace{\mathsf{E}e^{2\pi i\eta_{k-n}}}_{=\,\mathbf{0}} = 0.$$

Since $\mathcal{G}^W_{t_{k-n},t_{k+1}} \uparrow \mathcal{F}^W_{t_{k+1}}$ as $n \uparrow \infty$, we conclude ("Lévy theorem") that

$$\mathsf{E}(e^{2\pi i\eta_{k+1}} \,|\, \mathcal{F}^W_{t_{k+1}}) = 0.$$

If a strong solution then existed, then the variables η_{k+1} would be $\mathcal{F}^W_{t_{k+1}}$-measurable implying the identity $e^{2\pi i\eta_{k+1}} = 0$ which is clearly impossible.

The contradiction obtained shows that the equation $dX_t = a(t, X)\,dt + dW_t$ with the above function $a(t, X)$ does not have a strong solution.

In connection with this example of nonexistence it is reasonable to give general sufficient conditions for existence of s strong solution of the SDE with path-dependent nonanticipating coefficients:

$$dX_t = A(t, X)\,dt + B(t, X)\,dW_t.$$

The following result * is a straightforward extension of the result of Theorem 1.

---

* R. Liptser, A. Shiryaev, *Statistics of random processes*, Springer, 1977-78, 2001.

THEOREM 3. Suppose that $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,1]}, \mathsf{P})$ is a given filtered probability space and $W = (W_t)_{t\in[0,1]}$ is a $(\mathcal{F}_t)$-Wiener process. Let the nonanticipating functionals $A(t,X)$, $B(t,X)$, $t \in [0,1]$, $x \in C([0,1])$, satisfy the global Lipschitz condition:

$$|A(t,x) - A(t,y)|^2 + |B(t,x) - B(t,y)|^2 \le L_1 \int_0^t |x_s - y_s|^2 \, dK(s) + L_2 |x_t - y_t|^2,$$

$$A^2(t,x) + B^2(t,x) \le L_1 \int_0^t (1 + x_s^2) \, dK(s) + L_2(1 + x_t^2),$$

where L_1 and L_2 are constants, $K(s)$ is a nondecreasing right-continuous functions, $0 \le K(s) \le 1$, $x, y \in C([0,1])$. Let ξ be an $\mathcal{F}_0$-measurable random variable. Then the SDE

$$dX_t = A(t,X)\, dt + B(t,X)\, dW_t$$

has a unique strong solution $X = (X_t, \mathcal{F}_t)_{t\ge0}$, $0 \le t \le 1$.

Let us give also some results related with the SDE $dX_t = a(X_t)\,dt + dW_t$ considered in Example 4.

THEOREM 4 (Zvonkin, 1974). Suppose that for a one-dimensional SDE

$$dX_t = A(t,X_t)\,dt + B(t,X_t)\,dW_t, \qquad X_0 = 0,$$

the coefficient $A(t,x)$ is Borel measurable and bounded, the coefficient $A(t,x)$ is continuous and bounded, and there exists constants $c > 0$, $\varepsilon > 0$ such that

$$|B(t,x) - B(t,y)| \le c\sqrt{|x-y|}, \quad t \ge 0, \quad x,y \in \mathbb{R},$$
$$|B(t,x)| \ge \varepsilon, \quad t \ge 0, \quad x \in \mathbb{R}.$$

Then the SDE has a unique strong solution.

For homogeneous SDEs there exists a stronger result.

THEOREM 5 (Engelbert, Schmidt, 1985).
Suppose that, for a one-dimensional SDE

$$dX_t = A(X_t)\,dt + B(X_t)\,dW_t, \qquad X_0 = x_0,$$

the coefficient B does not vanish, $AB^{-2} \in L'_{\mathrm{loc}}(\mathbb{R})$, i.e.,

$$\int_{|x|\le\varepsilon} A(x)B^{-2}(x)\,dx < \infty, \qquad \varepsilon > 0,$$

and there exists a constant $c > 0$ such that

$$|B(x) - B(y)| \le c\sqrt{|x-y|}, \qquad x, y \in \mathbb{R},$$
$$|A(x)| + |B(x)| \le c(1+|x|), \qquad x \in \mathbb{R}.$$

Then the SDE has a unique strong solution.

§ 4. GIRSANOV's CHANGE of MEASURES

$(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\ge 0}, \mathsf{P})$ is a filtered probability space, $\mathcal{F}_0 = \{\varnothing, \Omega\}$,

$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots)$ is a sequence of i.i.d., $N(0,1)$, r.v., ε_n is $\mathcal{F}_n$-measurable.

$$\left.\begin{aligned} \mu &= (\mu_n)_{n\ge 1} \\ \sigma &= (\sigma_n)_{n\ge 1} \end{aligned}\right\}$$ are given predictable sequences (i.e., μ_n and σ_n are $\mathcal{F}_{n-1}$-measurable). σ_n are assumed > 0.

Set $\xi_n = \mu_n + \sigma_n \varepsilon_n$. Evidently, $\boxed{\mathrm{Law}(\xi \,|\, \mathcal{F}_{n-1}; \mathsf{P}) = N(\mu_n, \sigma_n^2)}$ which allows one to call $\xi = (\xi_n)_{n\ge 0}$

a CONDITIONALLY GAUSSIAN sequence (w.r.t. P)

with

conditional expectation $\mathbf{E}(\xi_n \,|\, \mathcal{F}_{n-1}) = \mu_n$ and **conditional variance** $\mathbf{D}(\xi_n \,|\, \mathcal{F}_{n-1}) = \sigma_n$

We have $\mathrm{Law}(\xi \,|\, \mathcal{F}_{n-1}; \mathsf{P}) = N(\mu_n, \sigma_n^2)$ and we want to construct a new measure P^μ such that

$$\boxed{\mathrm{Law}(\xi \,|\, \mathcal{F}_{n-1}; \mathsf{P}^\mu) = N(0, \sigma_n^2)}.$$

In particular, if $\sigma_n = 1$, then r.v. $\xi = \mu_n + \varepsilon$, $1 \le n \le N$, are i.i.d., $N(0,1)$, w.r.t. the measure $\mathsf{P}^\mu_N = \mathsf{P}^\mu | \mathcal{F}_N$:

$$\mathrm{Law}(\xi_1, \dots, \xi_N \,|\, \mathsf{P}^\mu) = \mathrm{Law}(\varepsilon_1, \dots, \varepsilon_N \,|\, \mathsf{P}),$$

i.e.,

$$\boxed{\mathrm{Law}(\mu_1 + \varepsilon_1, \dots, \mu_N + \varepsilon_N \,|\, \mathsf{P}^\mu) = \mathrm{Law}(\varepsilon_1, \dots, \varepsilon_N \,|\, \mathsf{P})} \qquad (\bullet)$$

The simplest version of corresponding continuous analog of this result is the following.

Let $B = (B_t)_{t \le T}$ be a Brownian motion (w.r.t. a measure P) and let $B^\mu = (B^\mu_t)_{t \le T}$ be defined by $B^\mu_t = \mu t + B_t$ (= a Brownian motion with drift). Then one can construct the new measure P^μ such that

$$\boxed{\mathrm{Law}(B^\mu_t, t \le T \,|\, \mathsf{P}^\mu_T) = \mathrm{Law}(B_t, t \le T \,|\, \mathsf{P}_T)}$$

where $\mathsf{P}_T = \mathsf{P}|\mathcal{F}_T$ and $\mathsf{P}^\mu_T = \mathsf{P}^\mu|\mathcal{F}_T$.

Let us prove the result $(\bullet)$.

For $N \geq 1$ define $Z_n = \exp\Big\{-\sum_{k=1}^{n} \frac{\mu_k}{\sigma_k}\,\varepsilon_k - \frac{1}{2}\sum_{k=1}^{n}\Big(\frac{\mu_k}{\sigma_k}\Big)^2\Big\}$.

LEMMA. 1) The sequence $Z = (Z_n)_{n\geq 1}$ is a $(\mathsf{P}, (\mathcal{F}_n))$-martingale with $\mathsf{E}Z_n = 1$.

2) Let $\mathcal{F}_\infty = \sigma(\bigcup \mathcal{F}_n)$ and assume that

$$\mathsf{E}\exp\Big\{\frac{1}{2}\sum_{k=1}^{\infty}\Big(\frac{\mu_k}{\sigma_k}\Big)^2\Big\} < \infty \qquad \textbf{(Novikov's condition).}$$

Then $Z = (Z_n)_{n\geq 1}$ is a uniformly integrable martingale with limit (P-a.s.) $Z_\infty = \lim Z_n$ such that

$$Z_\infty = \exp\Big\{-\sum_{k=1}^{\infty} \frac{\mu_k}{\sigma_k}\,\varepsilon_k - \frac{1}{2}\sum_{k=1}^{\infty}\Big(\frac{\mu_k}{\sigma_k}\Big)^2\Big\}$$

and $Z_n = \mathsf{E}(Z_\infty \,|\, \mathcal{F}_n)$.

PROOF. 1) This is obvious, since for each $n \geq 1$

$$\mathsf{E}\left[\exp\left\{\frac{\mu_k}{\sigma_k}\,\varepsilon_k - \frac{1}{2}\left(\frac{\mu_k}{\sigma_k}\right)^2\right\}\;\Big|\;\mathcal{F}_{n-1}\right] = 1$$

by the $\mathcal{F}_{n-1}$-measurability of μ_n/σ_n and the conditional Gaussian property $\mathrm{Law}(\varepsilon_n \,|\, \mathcal{F}_{n-1}; \mathsf{P}) = N(0,1)$.

2) This will be proved below.

Now turn to the proof of the **GIRSANOV THEOREM**:

$$\boxed{\mathsf{Law}(\mu_1+\varepsilon_1,\dots,\mu_N+\varepsilon_N\,|\,\widetilde{\mathsf{P}})=\mathsf{Law}(\varepsilon_1,\dots,\varepsilon_N\,|\,\mathsf{P})} \qquad (\bullet)$$

with $d\widetilde{\mathsf{P}}_N = Z_N\, d\mathsf{P}_N$.

First of all, recall the well-known **Bayes formula**:

Let $\widetilde{\mathsf{P}}_N \ll \mathsf{P}_N$, and let Y be a bounded $\mathcal{F}_N$-measurable r.v. Then for each $m \le N$

$$\widetilde{\mathsf{E}}(Y\,|\,\mathcal{F}_m)=\frac{1}{Z_m}\mathsf{E}(YZ_N\,|\,\mathcal{F}_m),$$

where in our case $\widetilde{\mathsf{E}}=\mathsf{E}_{\widetilde{\mathsf{P}}_N}$ and $Z_N=\exp\Big\{-\sum_{k=1}^N\frac{\mu_k}{\sigma_k}\varepsilon_k-\frac12\sum_{k=1}^N\Big(\frac{\mu_k}{\sigma_k}\Big)^2\Big\}$.

By this Bayes formula we get

$$\widetilde{\mathsf{E}}\Big(e^{i\lambda\xi_n}\,\big|\,\mathcal{F}_{n-1}\Big) = \mathsf{E}\bigg[\exp\bigg\{\Big(i\lambda\sigma_n - \frac{\mu_n}{\sigma_n}\Big)\varepsilon_n + i\lambda\mu_n - \frac{1}{2}\Big(\frac{\mu_n}{\sigma_n}\Big)^2\bigg\}\,\bigg|\,\mathcal{F}_{n-1}\bigg]$$

$$= \mathsf{E}\bigg[\overbrace{\exp\Big\{\Big(i\lambda\sigma_n - \tfrac{\mu_n}{\sigma_n}\Big)\varepsilon_n - \tfrac{1}{2}\Big(i\lambda\sigma_n - \tfrac{\mu_n}{\sigma_n}\Big)^2\Big\}}^{=:\, I,\ \ \mathsf{E}[I\,|\,\mathcal{F}_{n-1}]=1}$$

$$\times \underbrace{\exp\Big\{\tfrac{1}{2}\Big(i\lambda\sigma_n - \tfrac{\mu_n}{\sigma_n}\Big)^2 + i\lambda\mu_n - \tfrac{1}{2}\Big(\tfrac{\mu_n}{\sigma_n}\Big)^2\Big\}}_{=\exp\{-\lambda^2\sigma_n^2/2\}}\,\bigg|\,\mathcal{F}_{n-1}\bigg]$$

$$= \exp\Big\{-\frac{\lambda^2\sigma_n^2}{2}\Big\} \quad \widetilde{\mathsf{P}}_N\text{-a.s.}$$

$$\Downarrow \quad \Downarrow \quad \Downarrow$$

$$\boxed{\mathrm{Law}(\mu_1+\varepsilon_1,\dots,\mu_N+\varepsilon_N\,|\,\widetilde{\mathsf{P}}) = \mathrm{Law}(\varepsilon_1,\dots,\varepsilon_N\,|\,\mathsf{P})} \qquad (\bullet)$$

To extend this result to the case $N=\infty$, consider again the martingale $Z=(Z_n)_{n\ge1}$ with

$$Z_n=\exp\biggl\{-\sum_{k=1}^{n}\frac{\mu_k}{\sigma_k}\,\varepsilon_k-\frac{1}{2}\sum_{k=1}^{n}\Bigl(\frac{\mu_k}{\sigma_k}\Bigr)^2\biggr\}.$$

Since $\mathsf{E}Z_N=1$, we have had possibility to define the measure $\widetilde{\mathsf{P}}_N$:

$$d\widetilde{\mathsf{P}}_N=Z_N\,d\mathsf{P}_N,$$

w.r.t. which the discrete version of the Girsanov theorem holds.

For case $N=\infty$ it is reasonable to put

$$\boxed{d\widetilde{\mathsf{P}}_\infty=Z_\infty\,d\mathsf{P}_\infty},\qquad\text{where}\;\;\mathsf{P}_\infty=\mathsf{P}|\mathcal{F}_\infty,\;\;\mathcal{F}_\infty=\sigma\bigl(\bigcup\mathcal{F}_n\bigr).$$

Generally speaking, we cannot guarantee that $\mathsf{E} Z_\infty = 1$. That is true if $Z = (Z_n)_{n \ge 1}$ is a uniformly integrable martingale, since for such martingales there exists $\lim_{n\to\infty} Z_n$ $(= Z_\infty)$ and $\mathsf{E} Z - \infty = 1$.

The most famous sufficient condition for the uniform integrability of the family $(Z_n)_{n\ge 1}$ is the **Novikov criterion**:

$$\boxed{\mathsf{E} \exp\left\{\frac{1}{2}\sum_{k=1}^{\infty}\left(\frac{\mu_n}{\sigma_n}\right)^2 < \infty\right\}}$$

The proof of this criterion will be provided below (§ 5) for more general case of continuous time, which we consider now.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \ge 0}, \mathsf{P})$ be a filtered probability space, and $B = (B_t)_{t \ge 0}$ a Brownian motion (Wiener process).

Consider the question on existence of a **weak** solution of the SDE

$$\boxed{dX_t = \mu(t, X)\, dt + dW_t} \qquad \boxed{\boldsymbol{\mu}}$$

where for each $t \ge 0$ the functional $\mu(t,x) \equiv \mu(t; x_s, s \le t)$ is $\mathcal{B}_t$-measurable, where $\mathcal{B}_t = \sigma(x_s \colon s \le t)$, $x = (x_s) \in C[0,\infty)$, a space of continuous functions defined on $[0,\infty)$.

The basic method of the proof is based on the ideas of the change of measure launched in

1960 I. V. Girsanov. "On transformation of one class of random process with the help of absolutely continuous change of measure", Theory Probab. Appl., 5:1, 314-330.

Girsanov theorem claims the following.

Suppose $\Omega = C[0,T]$ and let P_T^W be a Wiener measure on $(C[0,T], \mathcal{B}_T)$; hence the canonical process $W = (W_t(x))_{t\geq T}$ with $W_t(x) = x_t$ is a P^W-Wiener process.

Let $Z_t = \exp\Big\{-\int_0^T \mu(s,W)\,dW_s - \frac{1}{2}\int_0^T \mu^2(s,W)\,ds\Big\}$ and $\mathsf{E}^W Z_T = 1$, where E^W stands for expectation w.r.t. the measure P_T^W.

Define a new measure $\widetilde{\mathsf{P}}_T$ such that

$$d\widetilde{\mathsf{P}}_T = Z_T\, d\mathsf{P}_T^W.$$

Then the process $B_t = W_t + \int_0^T \mu(s,W)\,ds$ is $\widetilde{\mathsf{P}}_T$-Wiener process:

$$\mathrm{Law}\Big(W_t + \int_0^T \mu(s,W)\,ds,\ t \leq T \,\Big|\, \widetilde{\mathsf{P}}_T\Big) = \mathrm{Law}(W_t,\ t \leq T \,|\, \mathsf{P}_T^W).$$

Put $X_t = W_t$. Then

$$\boxed{dX_t = -\mu(t, X)\, dt + dB_t} \qquad (\bullet\bullet)$$

where B is a $\widetilde{\mathsf{P}}_T$-Wiener process.

So, the set $(B, X, \widetilde{\mathsf{P}}_T)$ forms a weak solution of equation $(\bullet\bullet)$.

Consider a particular case $\mu(t,W) = -\mu$, where $\mu = \text{const}$. Then

$$Z_T = e^{-\mu B_T - \mu^2 T/2}, \quad d\widetilde{\mathsf{P}}_T^{\mu} = Z_T \, d\mathsf{P}_T, \quad B_t = W_t + \mu t,$$

and $\mathrm{Law}(W_t + \mu t,\ t \le T \,|\, \mathsf{P}_T^{\mu}) = \mathrm{Law}(W_t,\ t \le T \,|\, \mathsf{P}_T)$. From here we find that if $G_T(x) = G_T(x_t,\ t \le T)$ is a bounded functional on space $C[0,T]$, then

$$\boxed{\mathrm{Law}(G_T(W^{\mu}) \,|\, \mathsf{P}_T^{\mu}) = \mathrm{Law}(G_T(W) \,|\, \mathsf{P}_T)}$$

where $W_t^{\mu} = \mu t + W_t$ and $\mathsf{P}_T^{\mu} = \widetilde{\mathsf{P}}_T$ with

$$\boxed{d\mathsf{P}_T^{\mu} = e^{-\mu W_T - \mu^2 T/2} \, d\mathsf{P}_T} \quad \left(= e^{-\mu W_T^{\mu} - \mu^2 T/2} \, d\mathsf{P}_T\right).$$

As a result we get

$$\mathsf{E}G_T(W^\mu) = \mathsf{E}^\mu \frac{d\mathsf{P}_T}{d\mathsf{P}_T^\mu} G_T(W^\mu) = \mathsf{E}^\mu\, e^{\mu W_T^\mu - \mu^2 T/2}\, G_T(W^\mu)$$

$$\overset{\textbf{Girsanov's theorem}}{=} \quad \mathsf{E}\, e^{\mu W_T - \mu^2 T/2}\, G_T(W).$$

So, we have the following useful formula:

$$\boxed{\mathsf{E}G_T(W^\mu) = \mathsf{E}\, e^{\mu W_T - \mu^2 T/2}\, G_T(W)}$$

For many problems of the stochastic analysis, optimal stopping, it is very useful to extend the above results

$$\mathsf{Law}\Big(W_t + \int_0^T \mu(s,W)\,ds,\ t \le T \,\Big|\, \widetilde{\mathsf{P}}_T\Big) = \mathsf{Law}(W_t,\ t \le T \,|\, \mathsf{P}_T),$$

$$\mathsf{E}G_T(W^\mu) = \mathsf{E}\, e^{\mu W_T - \mu^2 T/2}\, G_T(W)$$

for case when T is replaced by stopping time τ:

$$(\blacklozenge) \qquad \mathsf{Law}\Big(W_t + \mu(s,W)\,ds,\ t \le \tau \,\Big|\, \mathsf{P}^\mu_\tau\Big) = \mathsf{Law}(W_t,\ t \le \tau \,|\, \mathsf{P}_\tau),$$

$$(\blacklozenge\blacklozenge) \qquad \mathsf{E}G_\tau(W^\mu) = \mathsf{E}\, e^{\mu W_\tau - \mu^2 \tau/2}\, G_\tau(W).$$

Here $d\mathsf{P}^\mu_T = e^{\mu W_\tau - \mu^2\tau/2} = d\mathsf{P}_\tau$ and, of course, we must assume that τ is such that

$$\boxed{\mathbf{E} e^{\mu W_\tau - \mu^2 \tau/2} = 1}\,.$$

If instead of

$$(\blacklozenge)\qquad \mathsf{Law}\Big(W_t + \mu(s,W)\,ds,\ \ t \le \tau \;\Big|\; \mathsf{P}^{\mu}_{\tau}\Big) = \mathsf{Law}(W_t,\ \ t \le \tau \,|\, \mathsf{P}_{\tau}),$$

$$(\blacklozenge\blacklozenge)\qquad \mathsf{E}G_{\tau}(W^{\mu}) = \mathsf{E}\, e^{\mu W_{\tau} - \mu^2\tau/2}\, G_{\tau}(W).$$

we consider

$$(\blacklozenge)\qquad \mathsf{Law}\Big(W_t + \int_0^t \mu(s,W)\,ds,\ \ t \le \tau \;\Big|\; \mathsf{P}^{\mu}_{\tau}\Big) = \mathsf{Law}(W_t,\ \ t \le \tau \,|\, \mathsf{P}_{\tau}),$$

$$(\blacklozenge\blacklozenge)\qquad \mathsf{E}G_{\tau}\Big(W_t + \int_0^t \mu(s,W)\,ds;\ \ t \le \tau\Big)$$

$$= \mathsf{E}\ \exp\Big\{-\int_0^{\tau} A(s,W)\,dW_s - \frac{1}{2}\int_0^{\tau} A^2(s,W)\,ds\Big\} \cdot G_{\tau}(W),$$

then it is necessary to replace the condition $\mathsf{E}e^{\mu W_{\tau} - \mu^2\tau/2} = 1$ by

$$\boxed{\mathsf{E}\ \exp\Big\{-\int_0^{\tau} \mu(s,W)\,dW_s - \frac{1}{2}\int_0^{\tau} \mu^2(s,W)\,ds\Big\} = 1}.$$

Put $\lambda = -\mu$ and denote

$$\mathcal{E}(\lambda)_t = e^{\lambda W_t - \lambda^2 t/2}, \qquad t \geq 0.$$

The process $(\mathcal{E}(\lambda)_t)_{t\geq 0}$ is the so-called stochastic exponential which solves the linear SDE

$$\boxed{d\mathcal{E}(\lambda)_t = \lambda \mathcal{E}(\lambda)_t \, dW_t}, \quad \mathcal{E}(\lambda)_0 = 1.$$

The process $(\mathcal{E}(\lambda)_t)_{t\geq 0}$ is an $(\mathcal{F}_t^W)_{t\geq 0}$-martingale with $\mathcal{E}(\lambda)_t = 1$. This martingale is not uniformly integrable and, generally speaking, $\mathcal{E}(\lambda)_\tau \leq 1$.

Our aim now is to find conditions on τ which imply $\mathcal{E}(\lambda)_\tau = 1$.

§ 5. CRITERIA of UNIFORM INTEGRABILITY for STOCHASTIC EXPONENTIALS

Recall first some famous **WALD's IDENTITIES** for a Wiener process (Brownian motion) $W = (W_t)_{t \geq 0}$.

Assume that the stopping time $\tau = \tau(\omega)$ is a stopping time w.r.t. the σ-algebras $(\mathcal{F}_t^W)_{t \geq 0}$ (i.e., $\{\tau(\omega) \leq t\} \in \mathcal{F}_t^W$, $t \geq 0$).

It is well known that:

$$\mathsf{E}\sqrt{\tau} < \infty \quad \Rightarrow \quad \mathsf{E}W_\tau = 0$$

$$\mathsf{E}\tau < \infty \quad \Rightarrow \quad \boxed{\mathsf{E}W_\tau = 0, \quad \mathsf{E}W_\tau^2 = \mathsf{E}\tau}$$

(Wald's identities)

Introduce the following conditions:

Gikhman & Skorokhod: $\boxed{\mathsf{E}e^{\lambda^2\tau} < \infty}$ $\mathbf{I(1;\boldsymbol{\lambda})}$

Liptser & Shiryaev: $\boxed{\mathsf{E}e^{(\frac{1}{2}+\varepsilon)\lambda^2\tau} < \infty}$ $\mathbf{II(\frac{1}{2}+\boldsymbol{\varepsilon};\boldsymbol{\lambda})}$

Novikov: $\boxed{\mathsf{E}e^{\frac{1}{2}\lambda^2\tau} < \infty}$ $\mathbf{III(\frac{1}{2};\boldsymbol{\lambda})}$

Yan; Krylov: $\boxed{\lim_{\varepsilon\downarrow 0} \varepsilon \log \mathsf{E}e^{(\frac{1}{2}-\varepsilon)\lambda^2\tau} = 0}$ $\mathbf{IV(\frac{1}{2}-;\boldsymbol{\lambda})}$

We begin with the demonstration of the following implication:

$\boxed{\mathbf{II}(\tfrac{1}{2}+\varepsilon;\lambda)\text{: } \mathsf{E}e^{(\frac{1}{2}+\varepsilon)\lambda^2\tau} < \infty \text{ for some } 0 < \varepsilon < 1} \Rightarrow \boxed{\mathsf{E}\mathcal{E}(\lambda)_\tau = 1}$

For simplicity take $\lambda = 1$ and fix a $\varepsilon > 0$.

It is sufficient (for the uniform integrability) to check that

$$\boxed{\sup_t \mathsf{E}\Big(Z_{t\wedge\tau}(1)\Big)^{1+\delta} < \infty \quad \text{for some } \delta > 0,}$$

where $Z_t(\lambda) = \mathcal{E}(\lambda)_t = e^{\lambda W_t - \lambda^2 t/2}$.

We have (with $p = 1 + \varepsilon$ and $q = (1 + \varepsilon)/\varepsilon$):

$$
\begin{aligned}
(Z_t(1))^{1+\delta} &= \psi_t^{(1)} \cdot \psi_t^{(2)} \\
&= e^{(1+\delta)W_t - p(1+\delta)^2 t/2} \cdot e^{p(1+\delta)^2 t/2 - (1+\delta)t/2}.
\end{aligned}
$$

By the Hölder inequality,

$$
\begin{aligned}
\mathsf{E}(Z_t(1))^{1+\delta} &= \Big(\mathsf{E}(\psi_t^{(1)})^p\Big)^{1/p} \cdot \Big(\mathsf{E}(\psi_t^{(2)})^q\Big)^{1/q} \\
&= \Big(\mathsf{E} e^{p(1+\delta)W_t - p^2(1+\delta)^2 t/2}\Big)^{1/p} \cdot \Big(\mathsf{E}(\psi_t^{(2)})^q\Big)^{1/q} \\
&= 1 \cdot \Big(\mathsf{E}(\psi_t^{(2)})^q\Big)^{1/q}.
\end{aligned}
$$

Take $\delta > 1$ such that for the given $0 < \varepsilon < 1$

$$\delta(1+\varepsilon) \le \frac{\varepsilon^2}{(1+\varepsilon)(1+2\varepsilon)}.$$

Then $(\psi_t^{(2)})^q = e^{pq(1+\delta)^2 t/2 - q(1+\delta)t/2} \le e^{(1/2+\varepsilon)t}$ and

$$\mathsf{E}(\psi_{t\wedge\tau}^{(2)})^q \le \mathsf{E}e^{(1/2+\varepsilon)(t\wedge\tau)} \le \mathsf{E}e^{(1/2+\varepsilon)\tau} < \infty.$$

So, $\qquad \mathsf{E}(Z_{t\wedge\tau}(1))^{1+\delta} \le \Big(\mathsf{E}(\psi_{t\wedge\tau}^{(2)})^q\Big)^{1/q} \le (\mathsf{E}e^{(1/2+\varepsilon)\tau})^{1/q} < \infty.$

The same is true if instead of $\lambda = 1$ we take any $\lambda > 0$. Therefore

$\boxed{\mathbf{II}(\tfrac{1}{2}+\boldsymbol{\varepsilon};\boldsymbol{\lambda})\text{: } \mathsf{E}e^{(\frac{1}{2}+\varepsilon)\lambda^2\tau} < \infty \text{ for some } 0 < \varepsilon < 1} \;\Rightarrow$

$\Rightarrow\; \boxed{\mathsf{E}Z_\tau(\lambda) = 1} \;\Rightarrow\; \boxed{\mathsf{E}\mathcal{E}(\lambda)_\tau = 1}$

Using the obtained result, let us prove the **Novikov criterion**:

$$\boxed{\mathbf{III}(\tfrac{1}{2};\boldsymbol{\lambda})\text{: } \mathsf{E}e^{\frac{1}{2}\lambda^2\tau} < \infty} \quad \Rightarrow \quad \boxed{\mathsf{E}\mathcal{E}(\lambda)_\tau = 1}$$

For the proof take $0 < \varepsilon < 1$ and $\lambda_\varepsilon = (1-\varepsilon)\lambda$. Then

$$\mathsf{E}e^{(1+\varepsilon)\lambda_\varepsilon^2\tau} = \mathsf{E}e^{(1+\varepsilon)(1-\varepsilon)^2\lambda^2\tau} \le \mathsf{E}e^{\frac{1}{2}\lambda^2\tau},$$

since $(1+\varepsilon)(1-\varepsilon)^2 = (1-\varepsilon^2)(1-\varepsilon) < 1$.

We know from the criterion $\mathbf{II}(\frac{1}{2}+\varepsilon;\boldsymbol{\lambda})$ that

$$\boxed{\mathsf{E}e^{(\frac{1+\varepsilon}{2})\lambda_\varepsilon^2\tau} < \infty \;\Rightarrow\; \mathsf{E}Z_\tau(\lambda_\varepsilon) = \boxed{\mathbf{1}} =}$$

$$= \mathsf{E}Z_\tau((1-\varepsilon)\lambda) = \mathsf{E}e^{\lambda(1-\varepsilon)W_\tau - \lambda^2(1-\varepsilon)^2\tau/2}$$

$$= \mathsf{E}\Big[e^{(1-\varepsilon)(\lambda W_\tau - \lambda^2\tau/2} \cdot e^{(1-\varepsilon)\varepsilon\lambda^2\tau/2}\Big] \qquad \left(\begin{array}{l}\textbf{Hölder's ineq. with} \\ \mathbf{1/p = 1-\varepsilon,\ 1/q = \varepsilon}\end{array}\right)$$

$$\le \Big(\mathsf{E}Z_\tau(\lambda)\Big)^{1-\varepsilon}\Big(\mathsf{E}e^{(1-\varepsilon)\lambda^2\tau/2}\Big)^{\varepsilon} \boxed{\le \Big(\mathsf{E}Z_\tau(\lambda)\Big)^{1-\varepsilon}\Big(\mathsf{E}e^{\lambda^2\tau/2}\Big)^{\varepsilon}}.$$

Taking $\varepsilon \downarrow 0$, we get $1 \le \mathsf{E}Z_\tau(\lambda)$. But $\mathsf{E}Z_\tau(\lambda) \le 1$ for any stopping time. Hence, $\mathsf{E}Z_\tau(\lambda) = \mathsf{E}\mathcal{E}(\lambda)_\tau = 1$.

Finally, let us show that

$$\boxed{\mathbf{IV}(\tfrac{1}{2}-;\boldsymbol{\lambda})\textbf{:}\ \lim_{\varepsilon\downarrow 0}\varepsilon\log \mathsf{E}e^{(\frac{1}{2}-\varepsilon)\lambda^2\tau}=0} \quad\Rightarrow\quad \boxed{\mathsf{E}Z_\tau(\lambda)=1}$$

Condition $\lim_{\varepsilon\downarrow 0}\varepsilon\log \mathsf{E}e^{(\frac{1}{2}-\varepsilon)\lambda^2\tau}=0$ is equivalent to

$$\lim_{\varepsilon\downarrow 0}\mathsf{E}\Big[e^{(1-\varepsilon)\lambda^2\tau/2}\Big]^{\varepsilon}=1. \qquad (*)$$

From the latter condition it follows that for sufficiently small $\varepsilon>0$ $\mathsf{E}e^{(1-\varepsilon)\lambda^2\tau/2}<\infty$. So, for $\lambda_\varepsilon=(1-\varepsilon)\lambda$

$$\mathsf{E}e^{\frac{1+\varepsilon}{2}\lambda_\varepsilon^2\tau}=\mathsf{E}e^{\frac{(1+\varepsilon)(1-\varepsilon)^2}{2}\lambda^2\tau}=\mathsf{E}e^{\frac{(1-\varepsilon^2)(1-\varepsilon)}{2}\lambda^2\tau}<\infty.$$

Then applying the criterion $\mathbf{II}(\frac{1+\varepsilon}{2};\boldsymbol{\lambda_\varepsilon})$ we get that

$$\mathsf{E}Z_\tau(\lambda_\varepsilon)=1.$$

Just as above we obtain

$$\boxed{\mathbf{1 = E}\boldsymbol{Z_\tau(\lambda_\varepsilon)}}$$

$$= \mathsf{E} Z_\tau((1-\varepsilon)\lambda) = \mathsf{E} e^{\lambda(1-\varepsilon)W_\tau - \lambda^2(1-\varepsilon)^2\tau/2}$$

$$= \mathsf{E}\Big[e^{(1-\varepsilon)(\lambda W_\tau - \lambda^2\tau/2} \cdot e^{(1-\varepsilon)\varepsilon\lambda^2\tau/2}\Big] \qquad \left(\begin{matrix}\textbf{Hölder's ineq. with} \\ \boldsymbol{1/p = 1-\varepsilon,\ 1/q = \varepsilon}\end{matrix}\right)$$

$$\boxed{\boldsymbol{\le \big(\mathbf{E} Z_\tau(\lambda)\big)^{1-\varepsilon}\big(\mathbf{E} e^{(1-\varepsilon)\lambda^2\tau/2}\big)^{\varepsilon}}}.$$

Put here $\varepsilon \downarrow 0$. Then taking into account the condition $(*)$ we obtain that $\mathsf{E} Z_\tau(\lambda) \ge 1$. But at the same time $\mathsf{E} Z_\tau(\lambda) \le 1$. Hence $\mathsf{E} Z_\tau(\lambda) = 1$, and criterion $\mathbf{IV}(\tfrac{1}{2}-; \boldsymbol{\lambda})$ is proved.

§ 6. WEAK and STRONG SOLUTIONS. 2: GENERAL REMARKS on the EXISTENCE and UNIQUENESS

1. The development of the general theory of SDE of the type

$$\boxed{dX_t = B(X,t)\,dt + A(X,t)\,dB_t} \qquad (\diamond)$$

where $B = (B_t)_{t \geq 0}$ is a Brownian motion, and given above different examples of existence/nonexistence of weak and strong solutions led to the following general concept of

the SOLUTION of SDE $(\diamond)$.

DEFINITION 1. A **SOLUTION** of SDE ($\diamond$) is a pair (X, B) of processes $X = (X_t)_{t \geq 0}$, $B = (B_t)_{t \geq 0}$ defined on some (not given apriori) filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathsf{P})$ and such that (X, B) are adapted, B is a P-Brownian motion, and P-a.s.

$$X_t = X_0 + \int_0^t B(X, s)\, ds + \int_0^t A(X, s)\, dB_s, \qquad t \geq 0$$

(integrals are assumed to be well-defined).

Such a pair is very often called a **weak solution** of the SDE ($\diamond$).

If the solution (X, B) is such that X_t is $\mathcal{F}_t^B$-measurable for all t, then we say that (X, B) is a **strong solution**.

DEFINITION 2. We say that the SDE ($\diamond$) has the property of **uniqueness in law** if for any solutions (X, B) and (X', B') (which may be defined on different filtered probability spaces $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathsf{P})$ and $(\Omega', \mathcal{F}', (\mathcal{F}'_t), \mathsf{P}'))$ one has

$$\boxed{\mathrm{Law}(X \mid \mathsf{P}) = \mathrm{Law}(X' \mid \mathsf{P}')}.$$

DEFINITION 3. We say that the SDE ($\diamond$) has the property of **pathwise uniqueness** if for any solutions (X, B) and (X', B) (which are defined on the same filtered probability spaces) one has

$$\boxed{\mathsf{P}(X_t = X'_t,\ t \geq 0) = 1}.$$

NOTE. Stroock and Varadhan developed a little bit different approach to the notion of the solution of SDE. They treat a solution not as a pair of processes but rather as a single object, namely, a measure on the path space which is a solution of the martingale problem.

Implications which follow directly **from the definitions**:

Strong existence	$\Longrightarrow$	Weak existence
Pathwise uniqueness	$\Longrightarrow$	Uniqueness in law

The well-known **Yamada–Watanabe theorem**:

Weak existence Pathwise uniqueness	$\Longrightarrow$	Strong existence

H.-J. Engelbert and A. Cherny established the counterpart of :

Strong existence Uniqueness in law	$\Longrightarrow$	Pathwise uniqueness

2. We conclude our lectures on the SDEs by the following notes (proposed by A. Cherny, H.-J. Engelbert et al.):

(1) It may happen that there exists no solution to the SDE (on any filtered probability space): $\begin{array}{|c|c|}\hline \mathbf{W} & \mathbf{S} \\ \hline - & - \\ \hline \end{array}$.

Examples:

(α) $\boxed{dX_t = -\operatorname{sgn} X_t \, dt}, \quad X_0 = 0,$

(deterministic equation)

here $\operatorname{sgn} x = \begin{cases} 1, & x > 0, \\ -1, & x \le 0; \end{cases}$

(β) $\boxed{dX_t = -\dfrac{1}{2X_t} I(X_t \neq 0) \, dt + dB_t}, \quad X_0 = 0,$

(2) It may happen that there exists a weak solution

but there is no strong solution:

W	S
+	−

.

<u>Examples:</u>

(α) $\boxed{dX_t = \operatorname{sgn} X_t \, dB_t}$, $\quad X_0 = 0$,

("two-sided" Tanaka equation)

(β) $\boxed{dX_t = A(X_t) \, dB_t}$, $\quad X_0 = 0$,

(Barlow's case)

with $A(x)\colon \mathbb{R} \to (0, \infty)$ bounded continuous

(3) If there exists a strong solution of the SDE on **some** filtered probability space, then there exists a strong solution on **any other** probability space with Brownian motion.

(4) If there exists a solution and if pathwise uniqueness holds, then on any filtered probability space with a Brownian motion there exists exactly one solution and this solution is strong (by the Yamada–Watanabe theorem).

(5) It may happen that there are several solutions with the same Brownian motion.

Examples:

(α) $\boxed{dX_t = I(X_t \neq 0)\, dB_t}$, $\quad X_0 = 0,$

W	S
+	+

here (B, B) and $(0, B)$ are solutions;

(β) $\boxed{dX_t = -\dfrac{\delta - 1}{2X_t} I(X_t \neq 0)\, dt + dB_t}$, $\quad X_0 = x_0 \neq 0, \quad \delta \geq 2.$

here we have the weak existence and pathwise uniqueness, so there exists a unique strong solution (the Yamada–Watanabe theorem).

§ 7. WEAK and STRONG SOLUTIONS. 3: SUFFICIENT CONDITIONS for EXISTENCE and UNIQUENESS. SUMMARY

$$\boxed{dX_t = A(t, X_t)\, dt + B(t, X_t)\, dW_t}$$

Itô: If $|A(t,x) - A(t,y)| + |B(t,x) - B(t,y)| \le c|x-y|$
$|A(t,x)| + |B(t,x)| \le c(1+|x|)$,
then there exists a unique strong solution

Skorokhod: If A and B are continuous bounded,
then there exists a weak solution

Zvonkin: If A is measurable bounded, B is continuous bounded
and $|B(t,x) - B(t,y)| \le c\sqrt{|x-y|}$, $|B(t,x)| \ge \varepsilon > 0$,
then strong existence and pathwise uniqueness hold

$$\boxed{dX_t = A(X_t)\,dt + B(X_t)\,dW_t} \qquad X_0 = 0$$

Engelbert & Schmidt:

If $B \neq 0$, $AB^{-2} \in L^1_{\text{loc}}(\mathbb{R})$, $|B(x) - B(y)| \leq c|x - y|$ and $|A(x)| + |B(x)| \leq c(1 + |x|)$,

then there exists a (pathwise) unique strong solution

$$\boxed{dX_t = A(t, X_t)\, dt + B(t, X_t)\, dW_t}$$

Yamada & Watanabe:

If there exist a constant $c > 0$ and a strictly increasing function $h\colon \mathbb{R}_+ \to \mathbb{R}_+$ with $\int_0^{0+} h^{-2}(x)\, dx = \infty$ such that $|A(t,x) - A(t,y)| \le c|x-y|$, $|B(t,x) - B(t,y)| \le h|x-y|$,
then pathwise uniqueness hold

Stroock & Varadhan:

If A is measurable bounded, B is continuous bounded and there exists a constant $\varepsilon(t,x) > 0$ such that $|A(t,x)\lambda| \ge \varepsilon(t,x)|\lambda|$, $\lambda \in \mathbb{R}$,
then weak existence and uniqueness in law hold

Portenko:

If $|B(t,x)| \ge \varepsilon > 0$ and $\int_0^t \int_{\mathbb{R}} |A(S,x)|^{n+\delta}\, ds\, dx < \infty$, $\delta > 0$,
then the previous Stroock–Varadhan result holds

$$\boldsymbol{dX_t = A(X_t)\,dt + B(X_t)\,dW_t}$$

Krylov:

> If A and B are measurable bounded
> and there exists a constant $\varepsilon > 0$ such that $|B(x)| \geq \varepsilon$,
> then weak existence and uniqueness in law hold

TOPIC II: Optimal stopping problems. Basic formulations, concepts and methods of solutions

§ 1. Standard and Nonstandard Optimal Stopping Problems

1. Optimal stopping theory is a part of the stochastic optimization theory with a wide set of applications and well-developed methods of solution.

GENERAL FORMULATION.

We have a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathsf{P})$ and a family of the stochastic processes $G = (G_t)_{t \geq 0}$, where G_t is interpreted as the **gain** if the observation is stopped at time t.

The optimal stopping problems consist in finding the value functions

$$\boxed{V = \sup_{\tau \in \mathfrak{M}} \mathsf{E} G_\tau} \quad \text{or} \quad \boxed{\overline{V} = \sup_{\tau \in \overline{\mathfrak{M}}} \mathsf{E} G_\tau I(\tau < \infty)}$$

Here $\overline{\mathfrak{M}}$ is the class of Markov times $\tau = \tau(\omega)$ (i.e., random variables τ with values on $[0, \infty]$ such that $\{\omega \colon \tau(\omega) \le t\} \in \mathcal{F}_t$ for all $t \ge 0$),

$\mathfrak{M}$ is a subclass of $\overline{\mathfrak{M}}$, namely, the random variables τ such that $\tau(\omega) < \infty$ for all $\omega \in \Omega$, or sometimes $\mathsf{P}(\tau(\omega) < \infty) = 1$.

Note that we do not make any measurability assumption on the gain functions G_t, $t \geq 0$ except for $\mathcal{F}$-measurability.

If G_t is $\mathcal{F}_t$-measurable for each $t \geq 0$, then we say that the problems of finding V and $\overline{V}$ are **STANDARD** problems.

If G_t is $\mathcal{F} = \sigma(\bigcup \mathcal{F}_t)$-measurable or $\mathcal{F}$-measurable, then we say that the problems of finding V and $\overline{V}$ are **NONSTANDARD** problems.

2. The general Optimal Stopping Theory (OST) is well-developed for **STANDARD** problems. For this case there are two main approaches to solve the problems V and $\overline{V}$:

$\mathbb{A}$. Martingale approach

operates with $\mathcal{F}_t$-measurable functions G_t and is based on two methods:

a) Method of **backward induction** (for case of discrete tine $t = n \le N$)

b) Method of **essential supremum** (for discrete and continuous time and finite or infinite horizon)

$\mathbb{B}$. Markovian approach

assumes that functions $G_t(\omega)$ have the Markovian representation, i.e. there exists a Markov process $X = (X_t)_{t \ge 0}$ such that $G_t(\omega) = G(t, X_t(\omega))$ with some measurable functions $G(t, x)$, where $x \in E$ and E is a phase space of X

The technique of reduction to the Markovian representation will be illustrated in Topic III where we consider some quickest detection problems formulated a priori as nonstandard stopping problems.

3. Before going to the results of the general theory for standard problems, let us consider the **procedures of reduction** of the nonstandard problems to the standard ones.

Assume $G_t(\omega)$ is (t,ω)-measurable positive (or bounded) functions, $t \geq 0$, $\omega \in \Omega$. From the General Theory of stochastic processes we know * that there exists an **optional** process (or optional projection) $G'_t(\omega)$ such that for any Markov time $\tau = \tau(\omega)$

$$\mathsf{E} G_\tau I(\tau < \infty) = \mathsf{E} G'_\tau I(\tau < \infty).$$

---

* See, e.g., C. Dellacherie, P.-A. Meyer. Probabilités et potentiel. Théorie des martingales (Ch. VI, § 2: Projections et projections duales), Hermann, 1980, 113–119.

NOTE. Process $G' = (G'_t(\omega))_{t\geq 0}$, $\omega \in \Omega$, is called **optional** if it is measurable w.r.t. the σ-field on $\Omega \times \mathbb{R}_+$, generated by all càdlàg adapted processes considered as a mappings on $\Omega \times \mathbb{R}_+$. If the process G' is optional, then G'_t is $\mathcal{F}_t$-measurable and for any Markov time τ the variable $G_\tau I(\tau < \infty)$ is $\mathcal{F}_\tau$-measurable, where the σ-algebra is defined as $\mathcal{F}_\tau = \{A \in \mathcal{F} \colon A \cap \{\tau \leq t\} \in \mathcal{F}_t,\ t \geq 0\}$.

DISCRETE TIME: there exists a very simple construction of G':

$$G'_n = \mathsf{E}(G_n \,|\, \mathcal{F}_n), \qquad n \geq 0.$$

CONTINUOUS TIME: the property

there exists an optional process $G'_t(\omega)$, $t \geq 0$, such that

$$\boxed{\mathbf{E} G_\tau I(\tau < \infty) = \mathbf{E} G'_\tau I(\tau < \infty)}$$

for any * stopping time τ

is EQUIVALENT to the property

there exists an optional process $G'_t(\omega)$, $t \geq 0$, such that the **(conditional)** identity hold:

$$\boxed{\mathbf{E}[G_\tau I(\tau < \infty) \,|\, \mathcal{F}_\tau] = G'_\tau I(\tau < \infty) \ \textbf{a.s.}}$$

for any * stopping time τ

---

* It is NOT sufficient to take only bounded stopping times τ.

§ 2. *OS-Lecture 1: INTRODUCTION*

1. Connections of the **Optimal stopping theory** and the **Mathematical analysis** (especially PDE-theory) can be illustrated by

the **Dirichlet problem** for the Laplace equation:
to find a harmonic function $u = u(x)$ in the class C^2 in the bounded open domain $C \subseteq \mathbb{R}^d$, i.e., to find a function $u \in C^2$ that satisfies the equation

$$\Delta u = 0, \quad x \in C, \tag{$*$}$$

and the boundary condition

$$u(x) = G(x), \quad x \in \partial D, \quad \text{where} \quad D = \mathbb{R}^d \setminus C. \tag{$**$}$$

Let

$$\tau_D = \mathrm{inf}\{t : B_t^x \in D\},$$

where

$$B_t^x = x + B_t$$

and $B = (B_t)_{t \geq 0}$ is a d-dimensional standard Brownian motion.

Then the probabilistic solution of the Dirichlet problem

$$\begin{aligned} \Delta u &= 0, & x &\in C, \\ u(x) &= G(x), & x &\in \partial D, \end{aligned}$$

is given by the formula

$$u(x) = \mathsf{E} G(B_{\tau_D}^x), \quad x \in C \cup \partial D$$

$$\Big(u(x) = \mathsf{E}_x G(B_{\tau_D})\Big).$$

The **optimal stopping theory** operates with the **optimization** problems, where

- we have a **set of domains** $\mathcal{C} = \left\{C : C \subseteq \mathbb{R}^d\right\}$ and
- we want to find the function

$$\boxed{U(x) = \sup_{\tau_D} \mathsf{E}_x G(B_{\tau_D})}, \quad \text{where} \quad G = G(x) \text{ is given for all } x \in \mathbb{R}^d, \; D \in \mathcal{D} = \{D = \bar{C} : C \in \mathcal{C}\}$$

or, generally, to find the function

$$\boxed{V(x) = \sup_{\tau} \mathsf{E}_x G(B_\tau)},$$

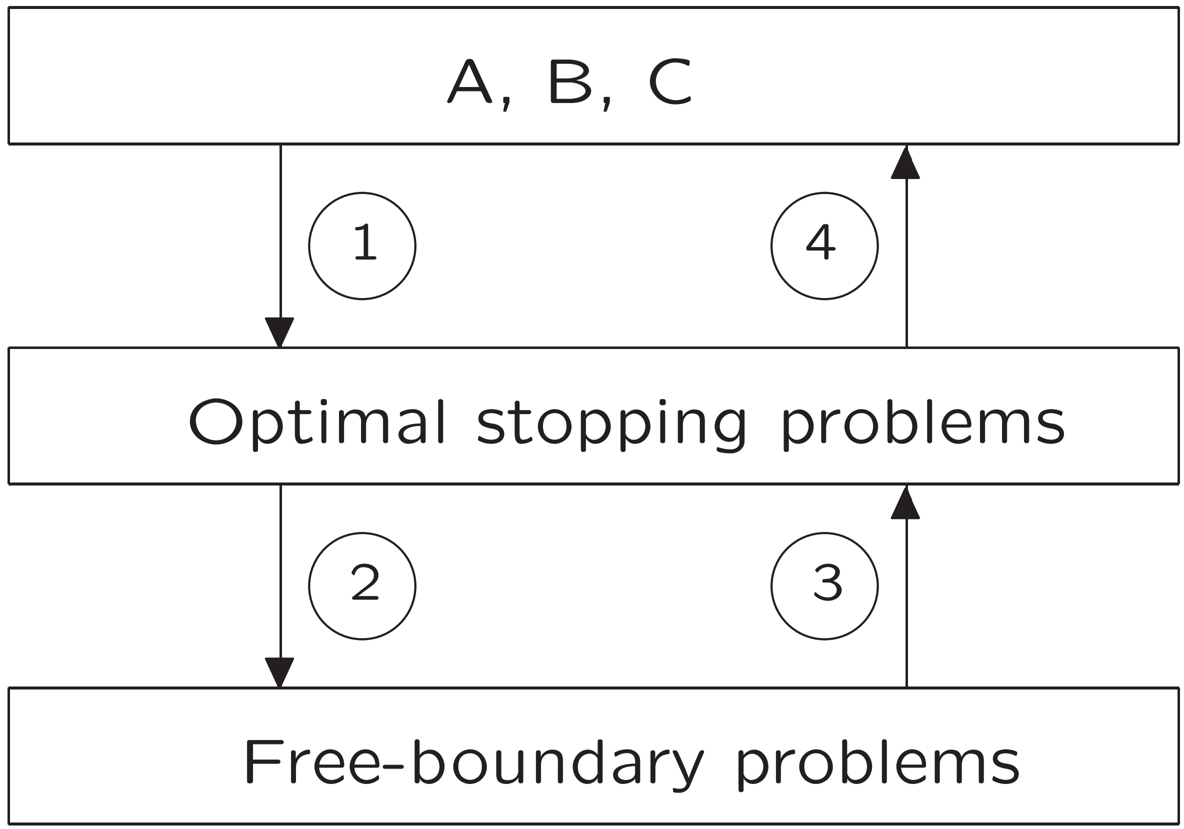
where τ is an **arbitrary finite stopping time** defined by the process B.

2. The following scheme illustrates the kind of **concrete** problems of **general interest** that will be studied in the courses of lectures:

A. Theory of probability sharp inequalities	B. Mathematical statistics sequential analysis

C. Financial mathematics stochastic equilibria

The solution method for problems **A, B, C** consists in **reformulation** to an optimal stopping problem and **reduction** to a free-boundary problem as stated in the diagram:


A, B, C
1
4
Optimal stopping problems
2
3
Free-boundary problems

3. To get some idea of the character of problems **A, B, C** that will be studied, let us begin with the following remarks.

(A) Let $B = (B_t)_{t \ge 0}$ be a standard Brownian motion. Then

Wald identities:

$$\boxed{\begin{array}{llll} \mathsf{E}B_T = 0 & \text{and} & \mathsf{E}B_\tau = 0 & \text{if } \mathsf{E}\sqrt{\tau} < \infty \\ \mathsf{E}B_T^2 = T & \text{and} & \mathsf{E}B_\tau^2 = \mathsf{E}\tau & \text{if } \mathsf{E}\tau < \infty \end{array}}$$

From Jensen's inequality and $\mathsf{E}|B_\tau|^2 = \mathsf{E}\tau$ we get

$$\boxed{\begin{array}{llll} \mathsf{E}|B_\tau|^p & \le & (\mathsf{E}\tau)^{p/2} & \text{for} \quad 0 < p \le 2 \\ \mathsf{E}|B_\tau|^p & \ge & (\mathsf{E}\tau)^{p/2} & \text{for} \quad 2 \le p < \infty \end{array}}$$

B. Davis (1976):

$$\boxed{\mathsf{E}|B_\tau| \le z_1^* \mathsf{E}\sqrt{\tau}, \qquad z_1^* = 1.30693\ldots}$$

Now our main interest relates with the estimation of the expectations

$$\mathsf{E}\max_{t\le\tau} B_t \quad \text{and} \quad \mathsf{E}\max_{t\le\tau} |B_t|.$$

We have

$$\max B \overset{\text{law}}{=} |B|.$$

So,

$$\mathsf{E}\max_{t\le T} B_t \;=\; \mathsf{E}|B_T| \;=\; \sqrt{\frac{2}{\pi}T}$$

and

$$\mathsf{E}\max_{t\le\tau} B_t \;=\; \mathsf{E}|B_\tau| \;\le\; \begin{cases} \sqrt{\mathsf{E}\tau}, & \\ z_1^*\mathsf{E}\sqrt{\tau}, & z_1^* = 1.30993\ldots \end{cases}$$

The case of $\max|B|$ is more difficult. We know that

$$\mathsf{P}\left(\max_{t\le T}|B_t|\le x\right)=\frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}\exp\left(-\frac{\pi^2(2n+1)^2}{8x^2}\right).$$

From here it is possible to obtain (but it is not easy!) that

$$\mathsf{E}\max_{t\le T}|B_t|=\sqrt{\frac{\pi}{2}T}\quad\Big(=1.25331\ldots\Big).$$

(Recall that $\mathsf{E}|B_T|=\sqrt{\frac{2}{\pi}T}\quad(=0.79788\ldots)$.)

SIMPLE PROOF:

$$(B_{at}; t \geq 0) \overset{\text{law}}{=} (\sqrt{a} B_t; t \geq 0).$$

Take $\sigma = \inf\{t > 0 : |B_t| = 1\}$. Then

$$\begin{aligned} \mathrm{P}\Big(\sup_{0 \leq t \leq 1} |B_t| \leq x\Big) &= \mathrm{P}\Big(\sup_{0 \leq t \leq 1} |B_{t/x^2}| \leq 1\Big) \\ &= \mathrm{P}\Big(\sup_{0 \leq t \leq 1/x^2} |B_t| \leq 1\Big) = \mathrm{P}\Big(\sigma \geq \frac{1}{x^2}\Big) = \mathrm{P}\Big(\frac{1}{\sqrt{\sigma}} \leq x\Big), \end{aligned}$$

that is,

$$\boxed{\sup_{0 \leq t \leq 1} |B_t| \overset{\text{law}}{=} \frac{1}{\sqrt{\sigma}}}$$

The **normal distribution** property:

$$\boxed{\sqrt{\frac{2}{\pi}}\int_0^\infty \mathsf{E}e^{-\frac{x^2}{2a^2}}\,dx = a}, \qquad a > 0. \qquad (*)$$

So,

$$\mathsf{E}\sup_{0\le t\le 1}|B_t| = \mathsf{E}\frac{1}{\sqrt{\sigma}} \overset{(*)}{=} \sqrt{\frac{2}{\pi}}\int_0^\infty \mathsf{E}e^{-\frac{x^2\sigma}{2}}\,dx.$$

Since $\mathsf{E}e^{-\lambda\sigma} = \dfrac{1}{\cosh\sqrt{2\lambda}}$, we get

$$\mathsf{E}\sup_{0\le t\le 1}|B_t| = \sqrt{\frac{2}{\pi}}\int_0^\infty \frac{dx}{\cosh x} = 2\sqrt{\frac{2}{\pi}}\int_0^\infty \frac{e^x\,dx}{e^{2x}+1} = \sqrt{\frac{2}{\pi}}\int_1^\infty \frac{dy}{1+y^2}$$

$$= 2\sqrt{\frac{2}{\pi}}\arctan(x)\Big|_1^\infty = 2\sqrt{\frac{2}{\pi}}\cdot\frac{\pi}{4} = \sqrt{\frac{\pi}{2}}.$$

$$\boxed{\mathsf{E}\sup_{0\le t\le 1}|B_t| = \sqrt{\frac{\pi}{2}}} \qquad \boxed{\mathsf{E}\sup_{0\le t\le T}|B_t| = \sqrt{\frac{\pi}{2}T}}$$

In connection with MAX the following can be interesting. In his speech delivered in 1856 before a grand meeting at the St.-Petersburg University the great mathematician

P. L. Chebyshev (1821–1894)

has formulated some statements about the "unity of theory and practice". In particular he emphasized that

> "a large portion of the practical questions can be stated in the form of problems of MAXIMUM and MINIMUM... Only the solution of these problems can satisfy the requests of practice which is always in search of the best and the most efficient."

4. Suppose that instead of $\max_{t \le T} |B_t|$, where, as already known,

$$\boxed{\mathsf{E} \max_{0 \le t \le T} |B_t| = \sqrt{\frac{\pi}{2} T}},$$

we have some **random** time τ and we want to find

$$\boxed{\mathsf{E} \max_{0 \le t \le \tau} |B_t| = \textbf{?}}$$

It is clear that it is **virtually impossible**

- **to compute** this expectation for every stopping time τ of B.

Thus, as the second best thing, one **can try**

- **to bound** it with a quantity which is easier computed.

A natural candidate for the latter is $\mathsf{E}\tau$ at least when finite.
In this way a $\boxed{PROBLEM \quad A}$ has appeared.

Problem A leads to the following **maximal inequality**:

$$\boxed{\mathsf{E}\Big(\max_{0\le t\le\tau}|B_t|\Big)\le C\sqrt{\mathsf{E}\tau}} \tag{1}$$

which is valid for all stopping times τ of B with the best constant C equal to $\sqrt{2}$.

We will see that the problem A can be solved in the form (1) by **REFORMULATION** to the following **optimal stopping problem**:

$$\boxed{V_* = \sup_{\tau}\mathsf{E}\Big(\max_{0\le t\le\tau}|B_t| - c\tau\Big)}, \tag{2}$$

where

- the supremum is taken over all stopping times τ of B satisfying $\mathsf{E}\tau<\infty$, and
- the constant $c>0$ is given and fixed.

It constitutes **Step 1** in the diagram above.

If $V_* = V_*(c)$ can be computed, then from (2) we get

$$\mathsf{E}\Big(\max_{0\le t\le\tau}|B_t|\Big) \le V_*(c) + c\,\mathsf{E}\tau \tag{3}$$

for all stopping times τ of B and all $c > 0$. Hence we find

$$\mathsf{E}\Big(\max_{0\le t\le\tau}|B_t|\Big) \le \inf_{c>0}\big(V_*(c) + c\,\mathsf{E}\tau\big). \tag{4}$$

for all stopping times τ of B. The RHS in (4) defines a function of $\mathsf{E}\tau$ that, in view of (2), provides a **sharp** bound of the LHS.

Our lectures demonstrate that the

optimal stopping problem (2)	can be reduced to a	**free-boundary problem**

This constitutes **Step 2** in the diagram above.

Solving the free-boundary problem one finds that $V_*(c) = 1/2c$. Inserting this into (4) yields

$$\inf_{c>0} \mathsf{E}\Big(V_*(c) + c\,\mathsf{E}\tau\Big) = \sqrt{2\,\mathsf{E}\tau} \tag{5}$$

so that the inequality (4) reads as follows:

$$\boxed{\mathsf{E}\Big(\max_{0\le t\le\tau} |B_t|\Big) \le \sqrt{2\,\mathsf{E}\tau}} \tag{6}$$

for all stopping times τ of B.

This is exactly the inequality (1) above with $C = \sqrt{2}$.
The constant $\sqrt{2}$ is **the best possible** in (6).

In the lectures we consider similar sharp inequalities for other stochastic processes using ramifications of the method just exposed.

Apart from being able to

- derive sharp versions of **known** inequalities

the method can also be used to

- derive some **new** inequalities.

(B) Classic examples of problems in **SEQUENTIAL ANALYSIS**:

- **WALD's problem** ("Sequential analysis", 1947) of sequential testing of two statistical hypotheses

$$H_0\colon \mu = \mu_0 \quad \text{and} \quad H_1\colon \mu = \mu_1 \tag{7}$$

about the **drift** parameter $\mu \in \mathbb{R}$ of the observed process

$$\boxed{X_t = \mu t + B_t}, \qquad t \geq 0, \qquad \text{where } B = (B_t)_{t\geq 0} \text{ is a standard } \textbf{Brownian motion}. \tag{8}$$

- The problem of sequential testing of two statistical hypotheses

$$H_0\colon \lambda = \lambda_0 \quad \text{and} \quad H_1\colon \lambda = \lambda_1 \tag{9}$$

about the **intensity** parameter $\lambda > 0$ of the observed process

$$\boxed{X_t = N_t^\lambda}, \qquad t \geq 0, \qquad \text{where } N = (N_t)_{t\geq 0} \text{ is a s tandard } \textbf{Poisson process}. \tag{10}$$

The basic problem in both cases seeks to find the

optimal decision rule (τ_*, d_*)

in the class $\Delta(\alpha, \beta)$ consisting of decision rules

(d, τ), where τ is the time of stopping and
accepting H_1 if $d = d_1$ or
accepting H_0 if $d = d_0$,

such that the probability errors of the first and second kind satisfy:

$$\mathsf{P}(\text{accept } H_1 \mid \text{true } H_0) \le \alpha \tag{11}$$

$$\mathsf{P}(\text{accept } H_0 \mid \text{true } H_1) \le \beta \tag{12}$$

and the mean times of observation $\mathsf{E}_0\tau$ and $\mathsf{E}_1\tau$ are as small as possible.

It is assumed that $\alpha > 0$ and $\beta > 0$ with $\alpha + \beta < 1$.

It turns out that with this $\boxed{(\textit{variational})\text{ problem}}$

$$\Downarrow$$

one may associate an $\boxed{\text{optimal stopping }(\textit{Bayesian})\text{ problem}}$

$$\Downarrow$$

which in turn can be reduced to a $\boxed{\text{free-boundary problem}}$.

This constitutes Steps 1 and 2 in the diagram above.

Solving the free-boundary problem leads to an optimal decision rule (τ_*, d_*) in the class $\Delta(\alpha, \beta)$ satisfying (11) and (12) as well as the following two identities:

$$\mathsf{E}_0\tau = \inf_{(\tau,d)} \mathsf{E}_0\tau, \qquad \mathsf{E}_1\tau = \inf_{(\tau,d)} \mathsf{E}_1\tau$$

where the infimum is taken over all decision rules (τ, d) in $\Delta(\alpha, \beta)$.

This constitutes Steps 3 and 4 in the diagram above.

In our lectures we study these as well as closely related problems of

QUICKEST DETECTION.

(The story of creating of the quickest detection problem of randomly appearing signal, its mathematical formulation, and the route of solving the problem (1961) are also interesting.)

Two of the prime findings, which also reflect the historical development of these ideas, are the

principles of SMOOTH and CONTINUOUS FIT

respectively.

C) One of the best-known specific problems of

MATHEMATICAL FINANCE,

that has a direct connection with optimal stopping problems, is the problem of determining the

arbitrage-free price of the **American put option**.

Consider the Black–Scholes model, where the stock price $X = (X_t)_{t \geq 0}$ is assumed to follow a geometric Brownian motion:

$$X_t = x \exp\Big(\sigma B_t + (r - \sigma^2/2)\, t\Big), \tag{13}$$

where $x > 0$, $\sigma > 0$, $r > 0$ and $B = (B_t)_{t \geq 0}$ is a standard Brownian motion. By Itô's formula one finds that the process X solves

$$dX_t = r\, X_t\, dt + \sigma\, X_t\, dB_t \quad \text{with} \quad X_0 = x. \tag{14}$$

General theory of financial mathematics makes it clear that the initial problem of determining the arbitrage-free price of the American put option can be reformulated as the following optimal stopping problem:

$$\boxed{V_* = \sup_{\tau} \mathsf{E} e^{-r\tau}(K - X_\tau)^+} \tag{15}$$

where the supremum is taken over all stopping times τ of X.

This constitutes **Step 1** in the diagram above.

The constant $K > 0$ is called the **strike price**. It has a certain financial meaning which we set aside for now.

It turns out that the optimal stopping problem (15):

$$V_* = \sup_\tau \mathsf{E} e^{-r\tau}(K - X_\tau)^+$$

can be reduced again to a free-boundary problem which can be solved explicitly. It yields the existence of a constant b_* such that the stopping time

$$\tau_* = \inf \{ \, t \geq 0 \mid X_t \leq b_* \, \} \tag{16}$$

is optimal in (15).

This constitutes **Steps 2** and **3** in the diagram above.

Both the optimal stopping point b_* and the arbitrage-free price V_* can be expressed explicitly in terms of the other parameters in the problem. A financial interpretation of these expressions constitutes **Step 4** in the diagram above.

In the formulation of the problem (15) above:

$$V_* = \sup_{\tau} \mathsf{E} e^{-r\tau}(K - X_\tau)^+$$

no restriction was imposed on the class of admissible stopping times, i.e. for certain reasons of simplicity it was assumed there that

τ belongs to the class of stopping times

$$\mathfrak{M} = \{\, \tau \mid 0 \le \tau < \infty \,\} \qquad (17)$$

without any restriction on their size.

A **more realistic** requirement on a stopping time in search for the arbitrage-free price leads to the following optimal stopping problem:

$$\boxed{V_*^T = \sup_{\tau \in \mathfrak{M}^T} \mathsf{E} e^{-r\tau} (K - X_\tau)^+} \tag{18}$$

where the supremum is taken over all τ belonging to the class of stopping times

$$\mathfrak{M}^T = \{\, \tau \mid 0 \le \tau \le T \,\} \tag{19}$$

with the horizon T being finite.

The optimal stopping problem (18) can be also reduced to a free-boundary problem that apparently **cannot be solved explicitly**.

Its study yields that the stopping time

$$\tau_* = \inf \{ \, 0 \le t \le T \mid X_t \le b_*(t) \, \} \tag{20}$$

is optimal in (18), where $b_* \colon [0, T] \to \mathbb{R}$ is an increasing continuous function.

A nonlinear Volterra integral equation can be derived which characterizes the optimal stopping boundary $t \mapsto b_*(t)$ and can be used to compute its values numerically as accurate as desired.

The comments on Steps 1–4 in the diagram above made in the infinite horizon case carry over to the finite horizon case without any change.

In our lectures we study these and other similar problems that arise from various financial interpretations of options.

5. So far we have only discussed problems A, B, C and their reformulations as optimal stopping problems. Now we want to address the methods of solution of optimal stopping problems and their reduction to free-boundary problems.

There are essentially two equivalent approaches to finding a solution of the optimal stopping problem. The first one deals with the problem

$$\boxed{V_* = \sup_{\tau \in \mathfrak{M}} \mathsf{E} G_\tau} \quad \text{in the case of } \textbf{infinite horizon}, \tag{21}$$

or the problem

$$\boxed{V_*^T = \sup_{\tau \in \mathfrak{M}^T} \mathsf{E} G_\tau} \quad \text{in the case of } \textbf{finite horizon}, \tag{22}$$

where $\mathfrak{M} = \{\, \tau \mid 0 \le \tau \le \infty \}$, and $\mathfrak{M}^T = \{\, \tau \mid 0 \le \tau \le T \,\}$.

In this formulation it is important to realize that

> $G = (G_t)_{t \geq 0}$ is an arbitrary stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathsf{P})$, where it is assumed that G is adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$ which in turn makes each τ from $\mathfrak{M}$ or $\mathfrak{M}^T$ a **stopping time**.

Since the method of solution to the problems (21) and (22) is based on results from the theory of martingales (**Snell's envelope**, 1952), the method itself is often referred to as the

MARTINGALE METHOD.

On the other hand, if we are to take a state space $(E, \mathcal{B})$ large enough, then one obtains the

$$\text{“Markov representation”} \quad G_t = G(X_t)$$

for some measurable function G, where $X = (X_t)_{t \geq 0}$ is a Markov process with values in E. Moreover, following the contemporary theory of Markov processes it is convenient to adopt the definition of a Markov process X as the **family** of Markov processes

$$((X_t)_{t \geq 0}, (\mathcal{F}_t)_{t \geq 0}, (\mathsf{P}_x)_{x \in E}) \tag{23}$$

where $\mathsf{P}_x(X_0 = x) = 1$, which means that the process X starts at x under P_x. Such a point of view is convenient, for example, when dealing with the Kolmogorov forward or backward equations, which presuppose that the process can start at any point in the state space.

Likewise, it is a profound attempt, developed in stages, to study optimal stopping problems through functions of initial points in the state space.

In this way we have arrived to the second approach which deals with the problem

$$V(x) = \sup_{\tau} \mathsf{E}_x G(X_\tau) \tag{24}$$

where the supremum is taken over $\mathfrak{M}$ or $\mathfrak{M}^T$ as above (**Dynkin's formulation**, 1963).

Thus, if the Markov representation of the initial problem is valid, we will refer to the

MARKOVIAN METHOD of solution.

6. To make the exposed facts more transparent, let us consider the optimal stopping problem

$$V_* = \sup_{\tau} \mathsf{E}\Big(\max_{0 \le t \le \tau} |B_t| - c\tau \Big)$$

in more detail.

Denote

$$X_t = |x + B_t| \tag{25}$$

for $x \ge 0$, and enable the maximum process to start at any point by setting for $s \ge x$

$$S_t = s \vee \Big(\max_{0 \le r \le t} X_r \Big). \tag{26}$$

$$\boxed{S_t = s \vee \Big(\max_{0 \le r \le t} X_r \Big)}$$

The process $S = (S_t)_{t \ge 0}$ is **not Markov**, but the pair $(X, S) = (X_t, S_t)_{t \ge 0}$ forms a **Markov process** with the state space $E = \{ (x, s) \in \mathbb{R}^2 \mid 0 \le x \le s \}$.

The value V_* from (2) above: $V_* = \sup_\tau \mathsf{E}\Big(\max_{0 \le t \le \tau} |B_t| - c\tau \Big)$ coincides with the value function

$$V_*(x, s) = \sup_\tau \mathsf{E}_{x,s}\Big(S_\tau - c\tau \Big) \tag{27}$$

when $x = s = 0$. The problem thus needs to be solved in this more general form.

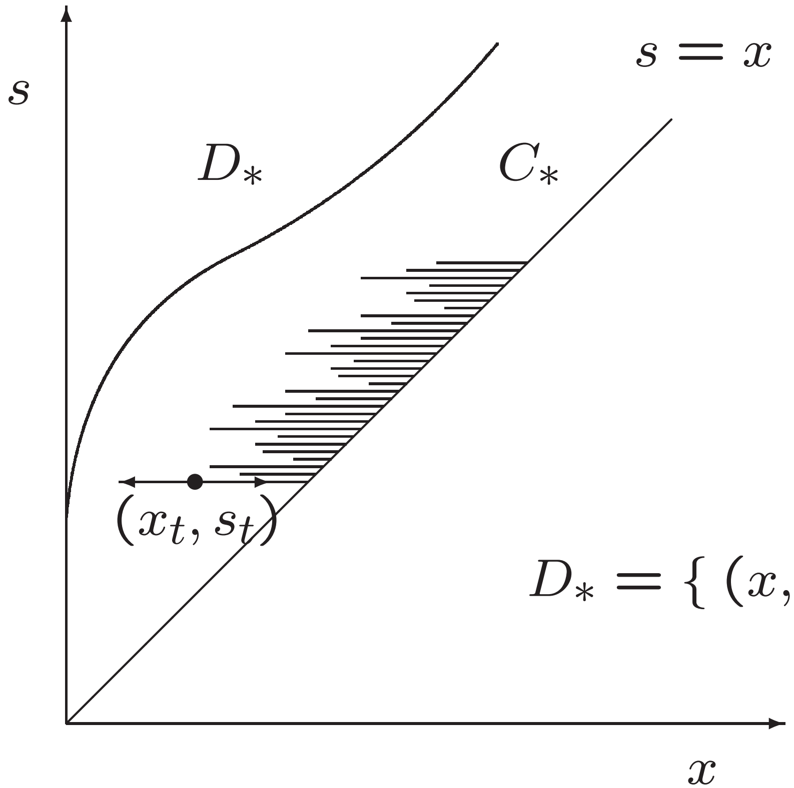
The general theory of optimal stopping for Markov processes makes it clear that the optimal stopping time in (27) can be written in the form

$$\tau_* = \inf \{ t \geq 0 \mid (X_t, S_t) \in D_* \} \tag{28}$$

where D_* is a **stopping set**, and
$C_* = E \setminus D_*$ is the **continuation set**.

In other words,

- if the observation of X was not stopped before time t since $X_s \in C_*$ for all $0 \leq s < t$, and we have that $X_t \in D_*$, then it is optimal to stop the observation at time t,
- if it happens that $X_t \in C_*$ as well, then the observation of X should be continued.



Heuristic considerations on the shape of the sets C_* and D_* make it plausible to guess that there exist a point $s_* \geq 0$ and a continuous increasing function $s \mapsto g_*(s)$ with $g_*(s_*) = 0$ such that

$$D_* = \{ (x, s) \in \mathbb{R}^2 \mid 0 \leq x \leq g_*(s)\,,\ s \geq s_* \} \qquad (29)$$

Note that such a guess about the shape of the set D_* can be made using the following intuitive arguments. If the process (X, S) starts from a point (x, s) with small x and large s, then it is reasonable to stop immediately because to increase the value s one needs a large time τ which in the formula (27) appears with a minus sign.

At the same time it is easy to see that

> if x is close or equal to s then it is reasonable to continue the observation, at least for small time Δ, because s will increase for the value $\sqrt{\Delta}$ while the cost for using this time will be $c\Delta$, and thus $\sqrt{\Delta} - c\Delta > 0$ if Δ is small enough.

Such an a priori analysis of the shape of the boundary between the stopping set C_* and the continuation set D_* is typical to the act of finding a solution to the optimal stopping problem. The

art of GUESSING

in this context very often plays a crucial role in solving the problem.

Having guessed that the stopping set D_* in the optimal stopping problem $V_*(x,s) = \sup_\tau \mathsf{E}_{x,s}(S_\tau - c\tau)$ takes the form

$$D_* = \{\, (x,s) \in \mathbb{R}^2 \mid 0 \le x \le g_*(s)\,,\ s \ge s_* \,\},$$

it follows that τ_* attains the supremum, i.e.,

$$V_*(x,s) = \mathsf{E}_{x,s}\Big(S_{\tau_*} - c\tau_*\Big) \quad \text{for all } (x,s) \in E. \tag{30}$$

Consider $V_*(x,s)$ for (x,s) in the continuation set

$$C_* = C_*^1 \cup C_*^2 \tag{31}$$

where the two subsets are defined as follows:

$$C_*^1 = \{\, (x,s) \in \mathbb{R}^2 \mid 0 \le x \le s < s_* \,\} \tag{32}$$

$$C_*^2 = \{\, (x,s) \in \mathbb{R}^2 \mid g_*(s) < x \le s\,,\ s \ge s_* \,\}. \tag{33}$$

Denote by

$$\mathbb{L}_X = \frac{1}{2} \frac{\partial^2}{\partial x^2}$$

the infinitesimal operator of the process X. By the strong Markov property one finds that V_* solves

$$\boxed{\mathbb{L}_X V_*(x,s) = c \quad \text{for } (x,s) \text{ in } C_*.} \tag{34}$$

If the process (X,S) starts at a point (x,s) with $x < s$, then during a positive time interval the second component S of the process remains equal to s.

This explains why the infinitesimal operator of the process (X,S) reduces to the infinitesimal operator of the process X in the interior of C_*.

On the other hand, from the structure of the process (X, S) it follows that at the diagonal in $\mathbb{R}^2_+$

- the condition of **normal reflection** holds:

$$\frac{\partial V_*}{\partial s}(x,s)\Big|_{x=s-} = 0. \tag{35}$$

Moreover, it is clear that for $(x,s) \in D_*$

- the condition of **instantaneous stopping** holds:

$$V_*(x,s) = s. \tag{36}$$

Finally, either by guessing or providing rigorous arguments, it is found that at the optimal boundary g_*

- the condition of **smooth fit** holds:

$$\frac{\partial V_*}{\partial x}(x,s)\Big|_{x=g_*(s)+} = 0. \tag{37}$$

This analysis indicates that the value function V_* and the optimal stopping boundary g_* can be obtained by searching for the **pair of functions** (V, g) solving the following **free-boundary problem**:

$$\mathbb{L}_X V(x,s) = c \quad \text{for } (x,s) \text{ in } C_g \tag{38}$$

$$\frac{\partial V}{\partial s}(x,s)\Big|_{x=s-} = 0 \quad \text{(normal reflection)} \tag{39}$$

$$V(x,s) = s \quad \text{for } (x,s) \text{ in } D_g \quad \text{(instantaneous stopping)} \tag{40}$$

$$\frac{\partial V}{\partial x}(x,s)\Big|_{x=g(s)+} = 0 \quad \text{(smooth fit)} \tag{41}$$

where the two sets are defined as follows ($g(s_0) = 0$):

$$C_g = \{\, (x,s) \in \mathbb{R}^2 \mid 0 \le x \le s < s_0 \text{ or } g(s) < x \le s,\ s \ge s_0 \,\} \tag{42}$$

$$D_g = \{\, (x,s) \in \mathbb{R}^2 \mid 0 \le x \le g(s)\,,\ s \ge s_0 \,\} \tag{43}$$

It turns out that this system does not have a unique solution so that an additional criterion is needed to make it unique in general.

Let us show how to solve the free-boundary problem (38)–(41) by picking the right solution (more details will be given in the lectures).

From (38) one finds that for (x, s) in C_g we have

$$V(x,s) = cx^2 + A(s)\,x + B(s) \tag{44}$$

where A and B are some functions of s. To determine A and B as well as g we can use the three conditions

$$\frac{\partial V}{\partial s}(x,s)\Big|_{x=s-} = 0 \qquad \text{(normal reflection)}$$

$$V(x,s) = s \quad \text{for} \quad (x,s) \quad \text{in} \quad D_g \qquad \text{(instantaneous stopping)}$$

$$\frac{\partial V}{\partial x}(x,s)\Big|_{x=g(s)+} = 0 \qquad \text{(smooth fit)}$$

which yield

$$g'(s) = \frac{1}{2(s - g(s))}, \quad \text{for} \quad s \geq s_0. \tag{45}$$

It is easily verified that the linear function

$$g(s) = s - \frac{1}{2c} \tag{46}$$

solves (45). In this way a candidate for the optimal stopping boundary g_* is obtained.

For $(x, s) \in E$ with $s \geq \frac{1}{2c}$ one can determine $V(x, s)$ explicitly using

$$V(x, s) = cx^2 + A(s)\,x + B(s)$$

and

$$g(s) = s - \frac{1}{2c}.$$

This in particular gives that $V(1/2c, 1/2c) = 3/4c$.

For other points $(x, s) \in E$ when $s < 1/2c$ one can determine $V(x, s)$ using that the observation must be continued. In particular for $x = s = 0$ this yields that

$$V(0,0) = V(1/2c, 1/2c) - c\,\mathsf{E}_{0,0}(\sigma) \tag{47}$$

where σ is the first hitting time of the process (X, S) to the point $(1/2c, 1/2c)$.

Because $\mathsf{E}_{0,0}(\sigma) = \mathsf{E}_{0,0}(X_\sigma^2) = (1/2c)^2$ and $V(1/2c, 1/2c) = 3/4c$, we find that

$$V(0,0) = \frac{1}{2c} \tag{48}$$

as already indicated prior to (5) above. In this way a candidate for the value function V_* is obtained.

The key role in the proof of the fact that

$$V = V_* \quad \text{and} \quad g = g_*$$

is played by

> **Itô's formula** (stochastic calculus) and the
> **optional sampling theorem** (martingale theory).

This step forms a **VERIFICATION THEOREM** that makes it clear that

> the solution of the free-boundary problem coincides with the solution of the optimal stopping problem

7. The important point to be made in this context is that the verification theorem is usually not difficult to prove in the cases when a candidate solution to the free-boundary problem is obtained **explicitly**.

This is quite typical for one-dimensional problems with **infinite horizon**, or some simpler two-dimensional problems, as the one just discussed.

In the case of problems with **finite horizon**, however, or other multidimensional problems, the situation can be radically different.

In these cases, in a manner quite opposite to the previous ones, the general results of optimal stopping can be used to prove the existence of a solution to the free-boundary problem, thus providing an alternative to analytic methods.

8. From the material exposed above it is clear that our basic interest concerns the case of **continuous** time.

The theory of optimal stopping in the case of continuous time is considerably more complicated than in the case of **discrete** time.

However, since the former theory uses many basic ideas from the latter, we have chosen to present the case of discrete time first, both in the **martingale** and **Markovian** setting, which is then likewise followed by the case of continuous time. The two theories form several my lectures.

§ 3. *LECTURES 2–3:*

Theory of optimal stopping for discrete time.

A. Martingale approach.

1. Definitions

$(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\ge 0}, \mathsf{P})$, $\quad \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \cdots \subseteq \mathcal{F}$, $\quad G = (G_n)_{n\ge 0}$.

Gain G_n is $\mathcal{F}_n$-measurable

Stopping (*Markov*) *time* $\tau = \tau(\omega)$:

$$\tau\colon \Omega \to \{0, 1, \dots, \infty\}, \quad \{\tau \le n\} \in \mathcal{F}_n \text{ for all } n \ge 0.$$

$\mathfrak{M}$ is the family of all **finite** stopping times

$\overline{\mathfrak{M}}$ is the family of **all** stopping times

$\mathfrak{M}_n^N = \{\tau \in \mathfrak{M} \,|\, n \le \tau \le N\}$

For simplicity we will set $\quad \mathfrak{M}^N = \mathfrak{M}_0^N \quad$ and $\quad \mathfrak{M}_n = \mathfrak{M}_n^\infty$.

The **optimal stopping problem** to be studied seeks to solve

$$\boxed{V_* = \sup_{\tau} \mathsf{E}\, G_\tau}\,. \tag{49}$$

For the existence of $\mathsf{E}\, G_\tau$ suppose (for simplicity) that

$$\mathsf{E} \sup_{0 \le k < \infty} |G_k| < \infty \tag{50}$$

(then $\mathsf{E}\, G_\tau$ is well defined for all $\tau \in \mathfrak{M}_n^N$, $n \le N < \infty$).

In the class $\mathfrak{M}_n^N$ we consider

$$\boxed{V_n^N = \sup_{\tau \in \mathfrak{M}_n^N} \mathsf{E}\, G_\tau}\,, \qquad 0 \le n \le N. \tag{51}$$

Sometimes we admit that τ in (49) takes the value ∞ ($\mathsf{P}(\tau = \infty) > 0$), so that $\tau \in \overline{\mathfrak{M}}$. We put $G_\tau = 0$ on $\{\tau = \infty\}$.

Sometimes it is useful to set $G_\infty = \limsup_{n \to \infty} G_n$.

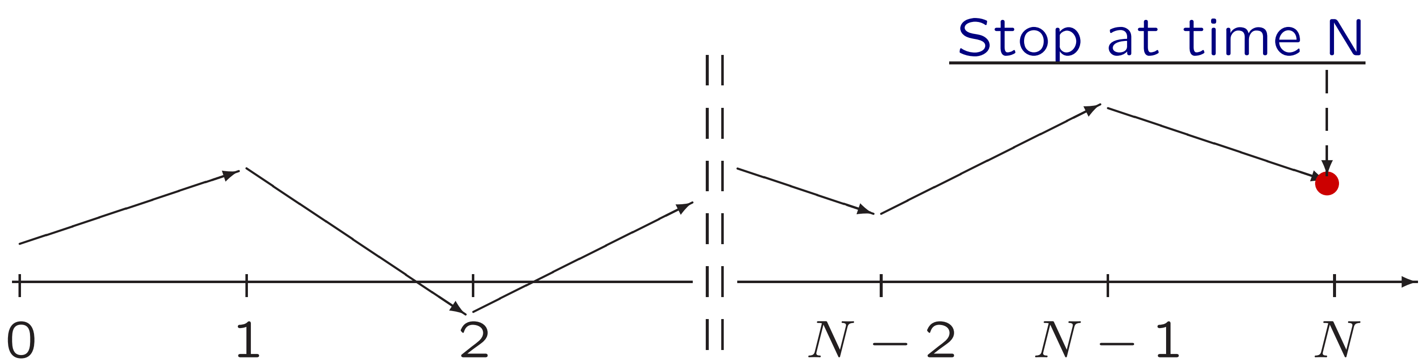
2. The method of backward induction.

$$V_n^N = \sup_{n \le \tau \le N} \mathsf{E}\, G_\tau$$

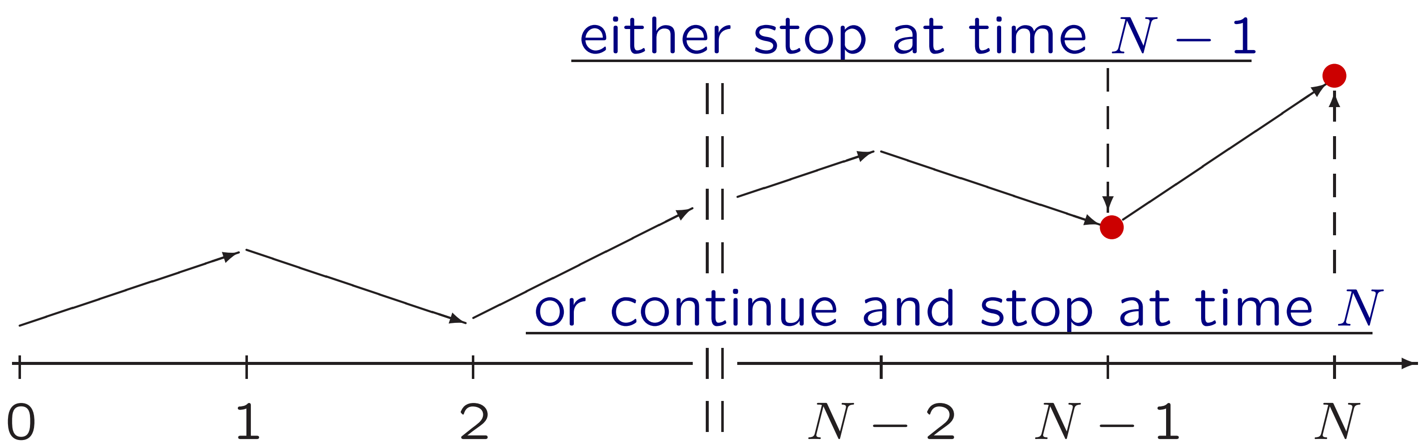
To solve this problem we introduce (by backward induction) a special stochastic sequence $S_N^N, S_{N-1}^N, \ldots, S_0^N$:

$$S_N^N = G_N, \qquad S_n^N = \max\{G_n, \mathsf{E}(S_{n+1}^N \mid \mathcal{F}_n)\},$$
$$n = N-1, \ldots, 0.$$

If $n = N$ we have to stop and our stochastic gain S_N^N, equals G_N.



For $n = N-1$ we can either stop or continue. If we stop, our gain S^N_{N-1}, equals G_{N-1}, and if we continue our gain S^N_{N-1} will be equal to $\mathsf{E}(S^N_N \,|\, \mathcal{F}_{N-1})$.



So,

$$S^N_{N-1} = \max\{G_{N-1}, \mathsf{E}(S^N_N \,|\, \mathcal{F}_{N-1})\}$$

and optimal stopping time is

$$\tau^N_{N-1} = \min\{N-1 \le k \le N \colon S^N_k = G_k\}.$$

Define now a sequence $(S_n^N)_{0\le n\le N}$ recursively as follows:

$$\begin{aligned} S_n^N &= G_N, && n = N, \\ S_n^N &= \max\{G_n, \mathsf{E}(S_{n+1}^N \mid \mathcal{F}_n)\}, && n = N-1, \dots, 0. \end{aligned}$$

The described method suggests to consider the following stopping time:

$$\tau_n^N = \inf\{n \le k \le N : S_k^N = G_k\} \quad \text{for} \quad 0 \le n \le N.$$

The first part of the following theorem shows that S_n^N and τ_n^N solve the problem in a stochastic sense.

The second part of the theorem shows that this leads also to a solution of the initial problem

$$V_n^N = \sup_{n\le\tau\le N} \mathsf{E}\, G_\tau \quad \text{for each} \quad n = 0, 1, \dots, N.$$

Theorem 1. ***(Finite horizon)***

I. *For all* $0 \le n \le N$ *we have*:

$$\text{(a)} \quad S_n^N \ge \mathsf{E}(G_\tau \,|\, \mathcal{F}_n), \qquad \forall \tau \in \mathfrak{M}_n^N;$$

$$\text{(b)} \quad S_n^N = \mathsf{E}(G_{\tau_n^N} \,|\, \mathcal{F}_n).$$

II. *Moreover, if* $0 \le n \le N$ *is given and fixed, then we have*:

(c) τ_n^N *is optimal in* $\quad V_n^N = \sup\limits_{n \le \tau \le N} \mathsf{E}\, G_\tau$*;*

(d) *if* τ_* *is also optimal then* $\tau_n^N \le \tau_*$*;*

(e) *the sequence* $(S_k^N)_{n \le k \le N}$ *is the smallest supermartingale which dominates* $(G_k)_{n \le k \le N}$ *(Snell's envelope)*

(f) *the stopped sequence* $(S_{k \wedge \tau_n^N}^N)_{n \le k \le N}$ *is a martingale.*

Proof of Theorem 1.

I. Induction over $n = N, N-1, \dots, 0$.

Conditions

$$\text{(a)} \quad S_n^N \geq \mathsf{E}(G_\tau \,|\, \mathcal{F}_n), \quad \forall \tau \in \mathfrak{M}_n^N,$$

and

$$\text{(b)} \quad S_n^N = \mathsf{E}(G_{\tau_n^N} \,|\, \mathcal{F}_n)$$

are trivially satisfied for $n = N$.

Suppose that (a) and (b) are satisfied for $n = N, N-1, \dots, k$, where $k \geq 1$, and let us show that they must then also hold for $n = k-1$.

(a) $\left(S_n^N \geq \mathsf{E}(G_\tau \,|\, \mathcal{F}_n), \quad \forall \tau \in \mathfrak{M}_n^N\right)$: Take $\tau \in \mathfrak{M}_{k-1}^N$ and set $\bar{\tau} = \tau \vee k$; then $\bar{\tau} \in \mathfrak{M}_k^N$, and since $\{\tau \geq k\} \in \mathcal{F}_{k-1}$ it follows that

$$\begin{aligned} \mathsf{E}(G_\tau \,|\, \mathcal{F}_{k-1}) &= \mathsf{E}[I(\tau = k-1) G_{k-1} \,|\, \mathcal{F}_{k-1}] + \mathsf{E}[I(\tau \geq k) G_{\bar{\tau}} \,|\, \mathcal{F}_{k-1}] \\ &= I(\tau = k-1) G_{k-1} + I(\tau \geq k)\, \mathsf{E}[\mathsf{E}(G_{\bar{\tau}} \,|\, \mathcal{F}_k) \,|\, \mathcal{F}_{k-1}]. \end{aligned} \tag{52}$$

By the induction hypothesis, (a) holds for $n = k$. Since $\bar{\tau} \in \mathfrak{M}_k^N$ this implies that

$$\mathsf{E}(G_{\bar{\tau}} \,|\, \mathcal{F}_k) \leq S_k^N. \tag{53}$$

From $S_n^N = \max(G_n, \mathsf{E}(S_{n+1}^N \,|\, \mathcal{F}_n))$ for $n = k-1$ we have

$$G_{k-1} \leq S_{k-1}^N, \tag{54}$$

$$\mathsf{E}(S_k^N \,|\, \mathcal{F}_{k-1}) \leq S_{k-1}^N. \tag{55}$$

Using (53)–(55) in (52) we get

$$\begin{aligned}
\mathsf{E}(G_\tau \,|\, \mathcal{F}_{k-1}) &\le I(\tau\!=\!k\!-\!1)\, S^N_{k-1} + I(\tau \ge k)\ \mathsf{E}(S^N_k \,|\, \mathcal{F}_{k-1}) \\
&\le I(\tau\!=\!k\!-\!1)\, S^N_{k-1} + I(\tau \ge k)\, S^N_{k-1} = S^N_{k-1}.
\end{aligned} \tag{56}$$

This shows that

$$S^N_n \ge \mathsf{E}(G_\tau \,|\, \mathcal{F}_n), \quad \forall \tau \in \mathfrak{M}^N_n$$

holds for $n = k-1$ as claimed.

(b) $\Big(S^N_n = \mathsf{E}(G_{\tau^N_n} \,|\, \mathcal{F}_n)\Big)$: To prove (b) for $n = k-1$ it is enough to check that all inequalities in (52) and (56) remain equalities when $\tau = \tau^N_{k-1}$. For this, note that

$$\begin{aligned}
\tau^N_{k-1} &= \tau^N_k && \text{on } \{\tau^N_{k-1} \ge k\}; \\
G_{k-1} &= S^N_{k-1} && \text{on } \{\tau^N_{k-1} = k-1\}; \\
\mathsf{E}(S^N_k \,|\, \mathcal{F}_{k-1}) &= S^N_{k-1} && \text{on } \{\tau^N_{k-1} \ge k\}.
\end{aligned}$$

Then we get

$$\begin{aligned}
\mathsf{E}\Big[G_{\tau^N_{k-1}} \,|\, \mathcal{F}_{k-1}\Big] &= I(\tau^N_{k-1} = k-1)\, G_{k-1} \\
&\quad + I(\tau^N_{k-1} \ge k)\ \mathsf{E}\Big[\,\mathsf{E}(G_{\tau^N_k} \,|\, \mathcal{F}_k) \,|\, \mathcal{F}_{k-1}\Big] \\
&= I(\tau^N_{k-1} = k-1)\, G_{k-1} + I(\tau^N_{k-1} \ge k)\ \mathsf{E}(S^N_k \,|\, \mathcal{F}_{k-1}) \\
&= I(\tau^N_{k-1} = k-1)\, S^N_{k-1} + I(\tau^N_{k-1} \ge k)\, S^N_{k-1} = S^N_{k-1}.
\end{aligned}$$

Thus

$$S^N_n = \mathsf{E}\Big(G_{\tau^N_n} \,|\, \mathcal{F}_n\Big)$$

holds for $n = k-1$. (We supposed by induction that (b) holds for $n = N, \ldots, k$.)

(c) $\Big(\tau_n^N$ is optimal in $V_n^N = \sup\limits_{n \le \tau \le N} \mathsf{E}\, G_\tau\Big)$:

Take expectation E in $S_n^N \ge \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)$, $\tau \in \mathfrak{M}_n^n$. Then

$$\mathsf{E}\, S_n^N \ge \mathsf{E}\, G_\tau \quad \text{for all} \quad \tau \in \mathfrak{M}_n^N$$

and by taking the supremum over all $\tau \in \mathfrak{M}_n^N$ we see that

$$\mathsf{E}\, S_n^N \ge V_n^N \quad \Big(= \sup_{\tau \in \mathfrak{M}_n^N} \mathsf{E}\, G_\tau \Big).$$

On the other hand, taking the expectation in $S_n^N = \mathsf{E}(G_{\tau_n^N} \,|\, \mathcal{F}_n)$ we get

$$\mathsf{E}\, S_n^N = \mathsf{E}\, G_{\tau_n^N}$$

which shows that

$$\mathsf{E}\, S_n^N \le V_n^N \quad \Big(= \sup_{\tau \in \mathfrak{M}_n^N} \mathsf{E}\, G_\tau \Big).$$

So,

$$\mathsf{E}\, S_n^N = V_n^N$$

and since $\mathsf{E}\, S_n^N = \mathsf{E}\, G_{\tau_n^N}$, we see that

$$V_n^N = \mathsf{E}\, G_{\tau_n^N}$$

implying the claim (c): "The stopping time τ_n^N is optimal".

(d) $\Big($if τ_* is also optimal then $\tau_n^N \le \tau_*\Big)$:

If we suppose that τ_* is also optimal, then $\tau_n^N \le \tau_*$. We claim that the optimality of τ_* implies that $S_{\tau_*}^N = G_{\tau_*}$ (P-a.s.). Indeed,

$$\text{for all } n \le k \le N \quad S_k^N \ge G_k, \qquad \text{thus} \quad S_{\tau_*}^N \ge G_{\tau_*}.$$

If $S_{\tau_*}^N \ne G_{\tau_*}$ (P-a.s.), then

$$\mathsf{P}(S_{\tau_*}^N > G_{\tau_*}) > 0.$$

It thus follows that

$$\mathsf{E}\, G_{\tau_*} < \mathsf{E}\, S^N_{\tau_*} \overset{(\alpha)}{\leq} \mathsf{E}\, S^N_n \overset{(\beta)}{=} V^N_n,$$

where

- **(α)** follows by the supermartingale property of $(S^N_k)_{n \le k \le N}$ (see (e)) and the optional sampling theorem, and
- **(β)** was obtained in (c).

The strict inequality $\mathsf{E}\, G_{\tau_*} < V^N_n$, however, contradicts the fact that τ_* is optimal.

Hence $S^N_{\tau_*} = G_{\tau_*}$ (P-a.s.) and the fact that $\tau^N_n \le \tau_*$ (P-a.s.) follows from the definition

$$\tau^N_n = \inf\{n \le k \le N \colon S^N_k = G_k\}.$$

(e) $\big($the sequence $(S_k^N)_{n\le k\le N}$ is the smallest supermartingale which dominates $(G_k)_{n\le k\le N}\big)$:

From

$$S_k^N = \max\{G_k, \mathsf{E}(S_{k+1}^N \,|\, \mathcal{F}_k)\}, \quad k = N-1, \dots, n,$$

we see that $(S_k^N)_{n\le k\le N}$ is a supermartingale:

$$S_k^N \ge \mathsf{E}(S_{k+1}^N \,|\, \mathcal{F}_k).$$

Also we have $S_k^N \ge G_k$. It means that $(S_k^N)_{n\le k\le N}$ is a supermartingale which dominates $(G_k)_{n\le k\le N}$.

Suppose that $(\widetilde{S}_k)_{n\le k\le N}$ is another supermartingale which dominates $(G_k)_{n\le k\le N}$, then the claim that $\widetilde{S}_k \ge S_k^N$ (P-a.s.) can be verified by induction over $k = N, N-1, \dots, l$.

Indeed, if $k = N$ then the claim follows by $S_n^N = G_N$ for $n = N$.

Assuming that $\widetilde{S}_k \geq S_k^N$ for $k = N, N-1, \ldots, l$ with $l \geq n+1$ it follows that

$$\begin{aligned} S_{l-1}^N &= \max(G_{l-1}, \mathsf{E}(S_l^N \,|\, \mathcal{F}_{l-1})) \\ &\leq \max(G_{l-1}, \mathsf{E}(\widetilde{S}_l \,|\, \mathcal{F}_{l-1})) \leq \widetilde{S}_{l-1} \quad \text{(P-a.s.)} \end{aligned}$$

using the supermartingale property of $(\widetilde{S}_k)_{n \leq k \leq N}$. So, $(S_k^N)_{n \leq k \leq N}$ is the smallest supermartingale which dominates $(G_k)_{n \leq k \leq N}$ (Snell's envelop).

(f) $\Big($the stopped sequence $(S^N_{k\wedge\tau^N_n})_{n\le k\le N}$ is a martingale$\Big)$:

To verify the martingale property

$$\mathsf{E}\left[S^N_{(k+1)\wedge\tau^N_n}\,|\,\mathcal{F}_k\right] = S^N_{k\wedge\tau^N_n}$$

with $n \le k \le N-1$ given and fixed, note that

$$\begin{aligned}\mathsf{E}\left[S^N_{(k+1)\wedge\tau^N_n}\,|\,\mathcal{F}_k\right] &= \mathsf{E}\left[I(\tau^N_n \le k)\, S^N_{k\wedge\tau^N_n}\,|\,\mathcal{F}_k\right] \\ &\quad + \mathsf{E}\left[I(\tau^N_n \ge k+1)\, S^N_{k+1}\,|\,\mathcal{F}_k\right] \\ &= I(\tau^N_n \le k) S^N_{k\wedge\tau^N_n} + I(\tau^N_n \ge k+1)\,\mathsf{E}(S^N_{k+1}\,|\,\mathcal{F}_k) \\ &= I(\tau^N_n \le k) S^N_{k\wedge\tau^N_n} + I(\tau^N_n \ge k+1) S^N_k = S^N_{k\wedge\tau^N_n}\end{aligned}$$

where we used that

$$S^N_k = \mathsf{E}(S^N_{k+1}\,|\,\mathcal{F}_k) \quad \text{on } \{\,\tau^N_n \ge k+1\,\}$$

and $\{\,\tau^N_n \ge k+1\,\} \in \mathcal{F}_k$ since τ^N_n is a stopping time.

Summary

1) The optimal stopping problem

$$V_0^N = \sup_{\tau \in \mathfrak{M}_0^N} \mathsf{E}\, G_\tau$$

is solved inductively by solving the problems

$$V_n^N = \sup_{\tau \in \mathfrak{M}_n^N} \mathsf{E}\, G_\tau \quad \text{for} \quad n = N, N-1, \ldots, 0.$$

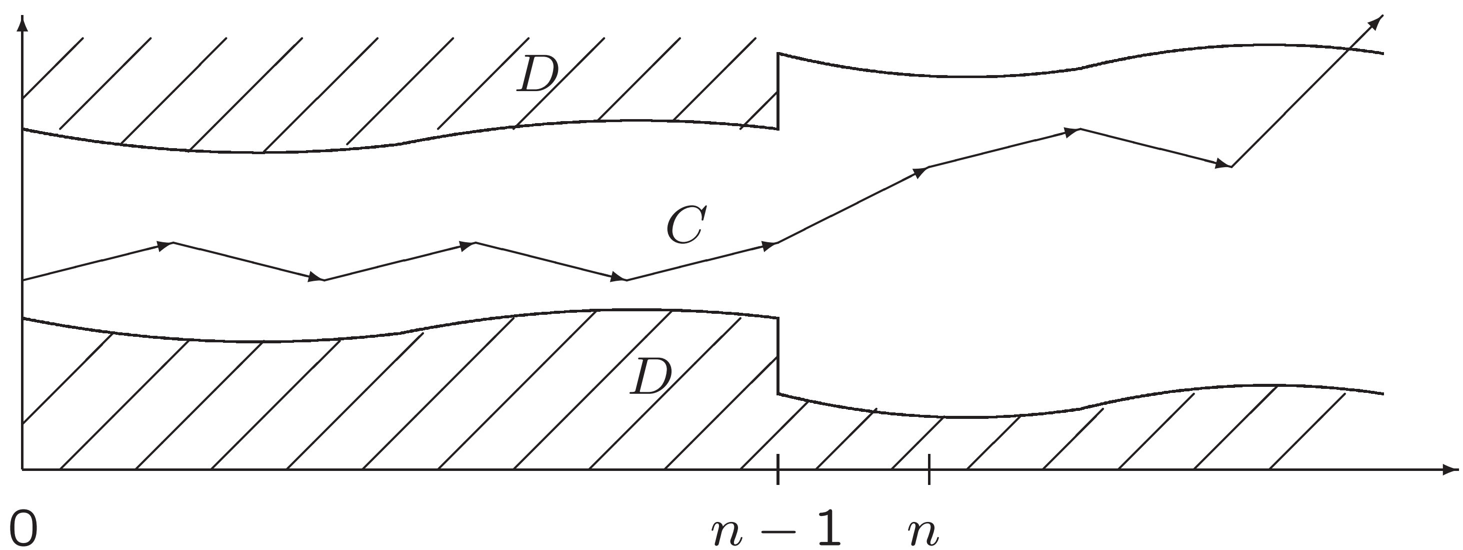
2) The optimal stopping rule τ_n^N for V_n^N satisfies

$$\tau_n^N = \tau_k^N \quad \text{on} \quad \{\tau_n^N \geq k\}$$

for $0 \leq n \leq k \leq N$ when τ_k^N is the optimal stopping rule for V_k^N. In other words, this means that if it was not optimal to stop within the time set $\{n, n+1, \ldots, k-1\}$ then the same optimality rule for V_n^N applies in the time set $\{k, k+1, \ldots, N\}$.

3) In particular, when specialized to the problem V_0^N, the following general principle (of dynamic programming) is obtained:

> if the stopping rule τ_0^N is optimal for V_0^N and it was not optimal to stop within the time set $\{0, 1, \dots, n-1\}$, then starting the observation at time n and being based on the information $\mathcal{F}_n$, the same stopping rule is still optimal for the problem V_n^N.



3. The method of ESSENTIAL SUPREMUM

The method of backward induction by its nature requires that the horizon N be FINITE so that the case of infinite horizon remains uncovered.

It turns out, however, that the random variables S_n^N defined by the recurrent relations

$$
\begin{aligned}
S_n^N &= G_N, && n = N, \\
S_n^N &= \max\{G_n, \mathsf{E}(S_{n+1}^N \,|\, \mathcal{F}_n)\}, && n = N-1, \dots, 0,
\end{aligned}
$$

admit a different characterization which can be directly extended to the case of infinite horizon N.

This characterization forms the base of the SECOND method that will now be presented.

Note that the relations

$$\begin{aligned} &\text{(a)} \quad S_n^N \geq \mathsf{E}(G_\tau \,|\, \mathcal{F}_n) \quad \forall \tau \in \mathfrak{M}_n^N; \\ &\text{(b)} \quad S_n^N = \mathsf{E}(G_{\tau_n^N} \,|\, \mathcal{F}_n) \end{aligned}$$

from Theorem 1 suggest that the following identity should hold:

$$\boxed{S_n^N = \sup_{\tau \in \mathfrak{M}_n^N} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)}\,.$$

(!) Difficulty: $\sup_{\tau \in \mathfrak{M}_n^N} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)$ need not define a measurable function.

To overcome this difficulty it turns out that the concept of

ESSENTIAL SUPREMUM

proves useful.

Lemma (about Essential Supremum)**.**

Let $\{Z_\alpha, \alpha \in \mathfrak{A}\}$ *be a family of random variables defined on* $(\Omega, \mathcal{F}, \mathsf{P})$ *where the index set* $\mathfrak{A}$ *can be arbitrary.*

I. *Then there exists a countable subset* J *of* $\mathfrak{A}$ *such that the random variable* $Z^*\colon \Omega \to \overline{\mathbb{R}}$ *defined by*

$$Z^* = \sup_{\alpha \in J} Z_\alpha$$

satisfies the following two properties:

(a) $\mathsf{P}(Z_\alpha \le Z^*) = 1$, $\forall \alpha \in \mathfrak{A}$;

(b) *If* $\tilde{Z}\colon \Omega \to \overline{\mathbb{R}}$ *is another random variable satisfying* $\mathsf{P}(Z_\alpha \le Z^*) = 1$, $\forall \alpha \in \mathfrak{A}$, *then* $\mathsf{P}(Z^* \le \tilde{Z}) = 1$.

II. *Moreover, if the family* $\{Z_\alpha, \alpha \in \mathfrak{A}\}$ *is upwards directed in the sense that*

$$\text{\textit{for any } } \alpha \text{ \textit{and} } \beta \text{ \textit{in} } \mathfrak{A} \text{ \textit{there exists} } \gamma \text{ \textit{in} } \mathfrak{A}$$
$$\text{\textit{such that} } \max(Z_\alpha, Z_\beta) \le Z_\gamma \text{ (P-\textit{a.s.}),}$$

then the countable set $J = \{\alpha_n, n \ge 1\}$ *can be chosen so that*

$$Z^* = \lim_{n\to\infty} Z_{\alpha_n} \quad \text{(P-\textit{a.s.})}$$

where $Z_{\alpha_1} \le Z_{\alpha_2} \le \cdots$ (P-*a.s.*).

Proof. (1) Since $x \mapsto \frac{2}{\pi}\arctan(x)$ is a strictly increasing function from $\overline{\mathbb{R}}$ to $[-1, 1]$, it is no restriction to assume that $|Z_\alpha| \le 1$.

(2) Let $\mathcal{C}$ denote the family of all countable subsets C of $\mathfrak{A}$. Choose an increasing sequence $\{C_n, n \ge 1\}$ in $\mathcal{C}$ such that

$$a \overset{\text{def}}{=} \sup_{C\in\mathcal{C}} \mathsf{E}\Big(\sup_{\alpha\in C} Z_\alpha\Big) = \sup_{n\ge1} \mathsf{E}\Big(\sup_{\alpha\in C_n} Z_\alpha\Big).$$

Then $J \stackrel{\text{def}}{=} \bigcup_{n=1}^{\infty} C_n$ is a countable subset of $\mathfrak{A}$ and we claim that

$$Z^* \stackrel{\text{def}}{=} \sup_{\alpha \in J} Z_\alpha$$

satisfies the properties (a) and (b).

(3) To verify these claims take $\alpha \in \mathfrak{A}$ arbitrarily.

(a): If $\alpha \in J$ then $Z_\alpha \leq Z^*$ so that (a) holds. If $\alpha \notin J$ and we assume that $\mathsf{P}(Z_\alpha > Z^*) > 0$, then

$$a < \mathsf{E}(Z^* \vee Z_\alpha) \leq a$$

since $a = \mathsf{E}\, Z^* \in [-1, 1]$ (by the monotone convergence theorem) and $J \cup \{\alpha\}$ belongs to $\mathcal{C}$. As the strict inequality is impossible, we see that $\mathsf{P}(Z_\alpha \leq Z^*) = 1$, $\forall \alpha \in \mathfrak{A}$ as claimed.

(b): follows from $Z^* = \sup_{\alpha \in J} Z_\alpha$ and (a): $\mathsf{P}(Z_\alpha \leq Z^*) = 1$, $\forall \alpha \in \mathfrak{A}$, since J is countable.

Finally, assume that the condition in II is satisfied. Then the initial countable set

$$J = \{\alpha_1, \alpha_2, \ldots\}$$

can be replaced by a new countable set $J^\circ = \{\alpha_1^\circ, \alpha_2^\circ, \ldots\}$ if we initially set $\alpha_1^\circ = \alpha_1$, and then inductively choose $\alpha_{n+1}^\circ \geq \alpha_n^\circ \vee \alpha_{n+1}$ for $n \geq 1$, where $\gamma \geq \alpha \vee \beta$ corresponds to Z_α, Z_β and Z_γ such that $Z_\gamma \geq Z_\alpha \vee Z_\beta$ (P-a.s.). The concluding claim $Z^* = \lim_{n\to\infty} Z_{\alpha_n}$ in II is then obvious, and the proof of the lemma is complete. $\square$

With the concept of essential supremum we may now rewrite

$$S_n^N \geq \mathsf{E}(G_\tau \,|\, \mathcal{F}_n) \quad \forall \tau \in \mathfrak{M}_n^N; \qquad\qquad S_n^N = \mathsf{E}(G_{\tau_n^N} \,|\, \mathcal{F}_n)$$

in Theorem 49 above as follows:

$$\boxed{S_n^N = \operatorname*{ess\,sup}_{n\leq\tau\leq N} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n) \quad \text{for all} \;\; 0 \leq n \leq N.}$$

This $\operatorname{ess\,sup}$ identity provides an additional characterization of the sequence of r.v.'s $(S_n^N)_{0\le n\le N}$ introduced initially by means of the recurrent relations

$$\begin{aligned} S_n^N &= G_N, && n = N, \\ S_n^N &= \max\{G_n, \mathsf{E}(S_{n+1}^N \,|\, \mathcal{F}_n)\}, && n = N-1, \dots, 0. \end{aligned}$$

Its advantage in comparison with these recurrent relations lies in the fact that the identity

$$S_n^N = \operatorname*{ess\,sup}_{n\le\tau\le N} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)$$

can naturally be extended to the case of **INFINITE** horizon N. This programme will now be described.

Consider (instead of $V_n^N = \sup_{\tau\in\mathfrak{M}_n^N} \mathsf{E}\, G_\tau$)

$$V_n = \sup_{\tau\in\mathfrak{M}_n^\infty} \mathsf{E}\, G_\tau.$$

To solve this problem we will consider the sequence of r.v.'s $(S_n)_{n\geq 0}$ defined as follows:

$$S_n = \operatorname*{ess\,sup}_{\tau \geq n} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)$$

as well as the following stopping time:

$$\tau_n = \inf\{k \geq n \,|\, S_k = G_k\} \quad \text{for} \quad n \geq 0,$$

where $\inf \varnothing = \infty$ by definition.

The first part **(I)** of the next theorem shows that $(S_n)_{n\geq 0}$ satisfies the same recurrent relations as $(S_n^N)_{0\leq n\leq N}$.

The second part **(II)** of the theorem shows that S_n and τ_n solve the problem in a stochastic sense.

The third part **(III)** shows that this leads to a solution of the initial problem $V_n = \sup_{\tau\geq n} \mathsf{E}\, G_\tau$.

The fourth part **(IV)** provides a supermartingale characterization of the solution.

Theorem 2 (Infinite horizon).

Consider the optimal stopping problems

$$V_n = \sup_{\tau \ge n} \mathsf{E}\, G_\tau, \qquad \tau \in \mathfrak{M}_n^\infty, \quad n \ge 0$$

assuming that the condition $\mathsf{E} \sup_{0 \le k < \infty} |G_k| < \infty$ *holds.*

I. *The following recurrent relations hold*:

$$S_n = \max\{G_n, \mathsf{E}(S_{n+1} \,|\, \mathcal{F}_n)\}, \quad \forall n \ge 0.$$

II. *Assume moreover if required below that*

$$\mathsf{P}(\tau_n < \infty) = 1.$$

Then for all $n \ge 0$ *we have*:

$$S_n \ge \mathsf{E}(G_\tau \,|\, \mathcal{F}_n) \quad \forall \tau \in \mathfrak{M}_n, \qquad S_n = \mathsf{E}(G_{\tau_n} \,|\, \mathcal{F}_n).$$

III. *Moreover, if* $n \geq 0$ *is given and fixed, then we have*:

The stopping time $\tau_n = \inf\{k \geq n : S_k = G_k\}$ *is optimal in* $V_n = \sup_{\tau \geq n} \mathsf{E}\, G_\tau$.

If τ_* *is an optimal stopping time for* $V_n = \sup_{\tau \geq n} \mathsf{E}\, G_\tau$ *then* $\tau_n \leq \tau_*$ (P-*a.s.*).

IV. *The sequence* $(S_k)_{k \geq n}$ *is the smallest supermartingale which dominates* $(G_k)_{k \geq n}$ *(Snell's envelope).*

The stopped sequence $(S_{k \wedge \tau_n})_{k \geq n}$ *is a martingale.*

Finally, if the condition $\mathsf{P}(\tau_n < \infty) = 1$ *fails so that* $\mathsf{P}(\tau_n = \infty) > 0$, *then there is NO optimal stopping time in* $V_n = \sup_{\tau \geq n} \mathsf{E}\, G_\tau$.

Proof. I. We need prove the recurrent relations

$$S_n = \max\{G_n, \mathsf{E}(S_{n+1} \,|\, \mathcal{F}_k)\}, \qquad n \geq 0.$$

Let us first show that

$$S_n \leq \max\{G_n, \mathsf{E}(S_{n+1} \,|\, \mathcal{F}_k)\}.$$

For this, take $\tau \in \mathfrak{M}_n$ and set $\bar{\tau} = \tau \vee (n+1)$.
Then $\bar{\tau} \in \mathfrak{M}_{n+1}$, and since $\{\tau \geq n+1\} \in \mathcal{F}_n$ we have

$$\begin{aligned}
\mathsf{E}(G_\tau \,|\, \mathcal{F}_n) &= \mathsf{E}[I(\tau = n)G_n \,|\, \mathcal{F}_n] + \mathsf{E}[I(\tau \geq n+1)G_{\bar{\tau}} \,|\, \mathcal{F}_n] \\
&= I(\tau = n)G_n + I(\tau \geq n+1)\,\mathsf{E}(G_{\bar{\tau}} \,|\, \mathcal{F}_n) \\
&= I(\tau = n)G_n + I(\tau \geq n+1)\,\mathsf{E}[\mathsf{E}(G_{\bar{\tau}} \,|\, \mathcal{F}_{n+1}) \,|\, \mathcal{F}_n] \\
&\leq I(\tau = n)G_n + I(\tau \geq n+1)\,\mathsf{E}\, S_{n+1} \,|\, \mathcal{F}_n) \\
&\leq \max\{G_n, \mathsf{E}(S_{n+1} \,|\, \mathcal{F}_n)\}.
\end{aligned}$$

From this inequality it follows that

$$S_n = \operatorname*{ess\,sup}_{\tau \ge n} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n) \le \max\{G_n, \mathsf{E}(S_{n+1} \,|\, \mathcal{F}_n)\}$$

which is the desired inequality.

For the reverse inequality, let us first note that $S_n \ge G_n$ (P-a.s.) by the definition of S_n, so that it is enough to show (and it is the **most difficult part** of the proof) that

$$S_n \ge \mathsf{E}(S_{n+1} \,|\, \mathcal{F}_n)$$

which is the supermartingale property of $(S_n)_{n\ge 0}$. To verify this inequality, let us first show that the family $\{\mathsf{E}(G_\tau \,|\, \mathcal{F}_{n+1}); \tau \in \mathfrak{M}_{n+1}\}$ is upwards directed in the sense that

for any α and β in $\mathfrak{A}$ there exists γ in $\mathfrak{A}$ such that $Z_\alpha \vee Z_\beta \le Z_\gamma$. **(∗)**

For this, note that if σ_1 and σ_2 are from $\mathfrak{M}_{n+1}$ and we set $\sigma_3 = \sigma_1 I_A + \sigma_2 I_{\bar{A}}$ where

$$A = \{\mathsf{E}(G_{\sigma_1} \,|\, \mathcal{F}_{n+1}) \geq \mathsf{E}(G_{\sigma_2} \,|\, \mathcal{F}_{n+1})\},$$

then $\sigma_3 \in \mathfrak{M}_{n+1}$ and we have

$$\begin{aligned}\mathsf{E}(G_{\sigma_3} \,|\, \mathcal{F}_{n+1}) &= \mathsf{E}(G_{\sigma_1} I_A + G_{\sigma_2} I_{\bar{A}} \,|\, \mathcal{F}_{n+1})\\ &= I_A\, \mathsf{E}(G_{\sigma_1} \,|\, \mathcal{F}_{n+1}) + I_{\bar{A}}\, \mathsf{E}(G_{\sigma_2} \,|\, \mathcal{F}_{n+1})\\ &= \mathsf{E}(G_{\sigma_1} \,|\, \mathcal{F}_{n+1}) \vee \mathsf{E}(G_{\sigma_2} \,|\, \mathcal{F}_{n+1})\end{aligned}$$

implying **(∗)** as claimed. Hence by Lemma there exists a sequence $\{\sigma_k, k \geq 1\}$ in $\mathfrak{M}_{n+1}$ such that

$$\operatorname*{ess\,sup}_{\tau \geq n+1} \mathsf{E}(G_\tau \,|\, \mathcal{F}_{n+1}) = \lim_{k\to\infty} \mathsf{E}(G_{\sigma_k} \,|\, \mathcal{F}_{n+1})$$

where

$$\mathsf{E}(G_{\sigma_1} \,|\, \mathcal{F}_{n+1}) \leq \mathsf{E}(G_{\sigma_2} \,|\, \mathcal{F}_{n+1}) \leq \cdots \quad \text{(P-a.s.)}.$$

Since

$$S_{n+1} = \operatorname*{ess\,sup}_{\tau \geq n+1} \mathsf{E}(G_\tau \,|\, \mathcal{F}_{n+1}),$$

by the conditional monotone convergence theorem we get

$$\begin{aligned}
\mathsf{E}(S_{n+1} \,|\, \mathcal{F}_n) &= \mathsf{E}\Big[\lim_{k\to\infty} \mathsf{E}(G_{\sigma_k} \,|\, \mathcal{F}_{n+1}) \,|\, \mathcal{F}_n\Big] \\
&= \lim_{k\to\infty} \mathsf{E}\Big[\mathsf{E}(G_{\sigma_k} \,|\, \mathcal{F}_{n+1}) \,|\, \mathcal{F}_n\Big] \\
&= \lim_{k\to\infty} \mathsf{E}(G_{\sigma_k} \,|\, \mathcal{F}_n) \leq S_n.
\end{aligned}$$

So, $S_n = \max\{G_n, \mathsf{E}(S_{n+1} \,|\, \mathcal{F}_n)\}$ and the proof if I is complete.

II. The inequality $S_n \geq \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)$, $\forall \tau \in \mathfrak{M}_n$, follows from the definition $S_n = \operatorname{ess\,sup}_{\tau \geq n} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)$.

For the proof of the equality $S_n = \mathsf{E}(G_{\tau_n} \,|\, \mathcal{F}_n)$ we use the fact stated below in IV that the stopped sequence $(S_{k\wedge\tau_n})_{k\geq n}$ is a martingale.

Setting $G_n^* = \sup_{k \geq n} |G_k|$ we have

$$|S_k| \leq \operatorname*{ess\,sup}_{\tau \geq k} \mathsf{E}\left(|G_\tau| \,|\, \mathcal{F}_k\right) \leq \mathsf{E}(G_n^* \,|\, \mathcal{F}_k) \tag{$*$}$$

for all $k \geq n$. Since G_n^* is integrable due to $\mathsf{E} \sup_{k \geq n} |G_k| < \infty$, it follows from $(*)$ that $(S_k)_{k \geq n}$ is uniformly integrable.

Thus the optional sampling theorem can be applied to the martingale $(M_k)_{k \geq n} = (S_{k \wedge \tau_n})_{k \geq n}$ and we get

$$M_n = \mathsf{E}(M_{\tau_n} \,|\, \mathcal{F}_n). \tag{$**$}$$

Since $M_n = S_n$ and $M_{\tau_n} = S_{\tau_n}$ we see that $(**)$ is the same as $S_n = \mathsf{E}(G_{\tau_n} \,|\, \mathcal{F}_n)$.

III: "The stopping time τ_n is optimal in $V_n = \sup_{\tau \geq n} \mathsf{E}\, G_\tau$."

The proof uses II and is similar to the corresponding proof in Theorem 1 ($N < \infty$).

IV. "The sequence $(S_k)_{k\ge n}$ is the smallest supermartingale which dominates $(G_k)_{k\ge n}$"   (Snell's envelop).

We proved in I that $(S_k)_{k\ge n}$ is a supermartingale. Moreover, from the definition

$$S_n = \operatorname*{ess\,sup}_{\tau\ge n} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)$$

it follows that $S_k \ge G_k$, $k \ge n$, which means that $(S_k)_{k\ge n}$ dominates $(G_k)_{k\ge n}$. Finally, if $(\tilde S_k)_{k\ge n}$ is another supermartingale which dominates $(G_k)_{k\ge n}$, then from $S_n = \mathsf{E}(G_{\tau_n} \,|\, \mathcal{F}_n)$ (Part II) we find

$$S_k = \mathsf{E}(G_{\tau_k} \,|\, \mathcal{F}_k) \le \mathsf{E}(\tilde S_{\tau_k} \,|\, \mathcal{F}_k) \le \tilde S_k, \quad \forall k \ge n.$$

(The last inequality follows by the optional sampling theorem being applicable since $\tilde S_k^- \le G_k^- \le G_n^*$ $(= \sup_{k\ge n} |G_k|)$ with G_n^* integrable.)

The statement

“The stopped sequence $(S_{k\wedge\tau_n})_{k\ge n}$ is a martingale”

is proved in exactly the same way as for case $N<\infty$.

Finally, note that the final claim

“If the condition $\mathsf{P}(\tau_n<\infty)=1$ fails so that $\mathsf{P}(\tau_n=\infty)>0$, then there is **NO** optimal stopping time in the problem $V_n=\sup_{\tau\ge n}\mathsf{E}\,G_\tau$”

follows directly from III (“If τ_n is optimal stopping tome then $\tau_n\le\tau_*$ (P-a.s.) for the problem $V_n=\sup_{\tau\ge n}\mathsf{E}\,G_\tau$”).

Remark. From the definition

$$S_n = \operatorname*{ess\,sup}_{n \le \tau \le N} \mathsf{E}(G_\tau \mid \mathcal{F}_n)$$

it follows that

$$N \mapsto S_n^N \quad \text{and} \quad N \mapsto \tau_n^N$$

are increasing. So,

$$S_n^\infty = \lim_{N \to \infty} S_n^N \quad \text{and} \quad \tau_n^\infty = \lim_{N \to \infty} \tau_n^N$$

exist P-a.s. for each $n \ge 0$.

Note also that from

$$V_n^N = \sup_{n \le \tau \le N} \mathsf{E}\, G_\tau$$

it follows that $N \mapsto V_n^N$ is increasing, so that $V_n^\infty = \lim_{N\to\infty} V_n^N$ exists for each $n \ge 0$.

From $S_n^N = \operatorname{ess\,sup}_{n \le \tau \le N} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)$ and $S_n = \operatorname{ess\,sup}_{\tau \ge n} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)$ we see that

$$S_n^\infty \le S_n \quad \text{and} \quad \tau_n^\infty \le \tau_n. \tag{$*$}$$

Similarly,

$$V_n^\infty \le V_n \quad \Big(= \sup_{\tau \ge n} \mathsf{E}\, G_\tau \Big). \tag{$**$}$$

If condition $\mathsf{E} \sup_{n \le k < \infty} |G_k| < \infty$ does not hold then the inequalities in $(*)$ and $(**)$ can be strict.

Theorem 3 (From finite to infinite horizon)**.**

If $\mathsf{E}\sup_{0\le k<\infty}|G_k| < \infty$ *then in* $S_n^\infty \le S_n$*,* $\tau_n^\infty \le \tau_n$ *and* $V_n^\infty \le V_n$ *we have equalities for all* $n \ge 0$*.*

Proof. From

$$S_n^N = \max\{G_n, \mathsf{E}(S_{n+1}^N \,|\, \mathcal{F}_n)\}, \quad n \ge 0,$$

we get

$$S_n^\infty = \max\{G_n, \mathsf{E}(S_{n+1}^\infty \,|\, \mathcal{F}_n)\}, \quad n \ge 0.$$

So, $(S_n^\infty)_{n\ge 0}$ is a supermartingale.

Since $S_n^\infty \ge G_n$ we see that

$$(S_n^\infty)^- \le G_n^- \le \sup_{n\ge 0} G_n^-, \quad n \ge 0.$$

So, $((S_n^\infty)^-)_{n\ge 0}$ is uniformly integrable.

Then by the optional sampling theorem we get

$$S_n^\infty \geq \mathsf{E}(S_\tau^\infty \,|\, \mathcal{F}_n) \quad \text{for all} \quad \tau \in \mathfrak{M}_n. \tag{$*$}$$

Moreover, since $S_k^\infty \geq G_k$, $k \geq n$, it follows that $S_\tau^\infty \geq G_\tau$ for all $\tau \in \mathfrak{M}_n$, and hence

$$\mathsf{E}(S_\tau^\infty \,|\, \mathcal{F}_n) \geq \mathsf{E}(G_\tau \,|\, \mathcal{F}_n) \tag{$**$}$$

for all $\tau \in \mathfrak{M}_n$. From ($*$), ($**$) and

$$S_n = \operatorname*{ess\,sup}_{\tau \geq n} \mathsf{E}(G_\tau \,|\, \mathcal{F}_n)$$

we see that $S_n^\infty \geq S_n$.

Since the reverse inequality holds in general as shown above, this establishes that $S_n^\infty = S_n$ (P-a.s.) for all $n \geq 0$. From this it also follows that $\tau_n^\infty = \tau_n$ (P-a.s.), $n \geq 0$. Finally, the third identity $V_n^\infty = V_n$ follows by the monotone convergence theorem.

B. Markovian approach.

We will present basic results of optimal stopping when

the time is discrete and **the process is Markovian**.

1. We consider a time-homogeneous Markov chain $X = (X_n)_{n\geq 0}$

- defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathsf{P}_x)$
- taking values in a measurable space $(E, \mathcal{B})$

where for simplicity we will assume that

(a) $E = \mathbb{R}^d$ for some $d \geq 1$

(b) $\mathcal{B} = \mathcal{B}(\mathbb{R}^d)$ is the Borel σ-algebra on $\mathbb{R}^d$.

It is assumed that the chain X starts at x under P_x for $x \in E$.

It is also assumed that the mapping $x \mapsto \mathsf{P}_x(F)$ is measurable for each $F \in \mathcal{F}$.

It follows that the mapping $x \mapsto \mathsf{E}_x(Z)$ is measurable for each random variable Z.

Finally, without loss of generality we will assume that $(\Omega, \mathcal{F})$ equals the canonical space $(E^{\mathbb{N}_0}, \mathcal{B}^{\mathbb{N}_0})$ so that the shift operator $\theta_n \colon \Omega \to \Omega$ is well defined by

$$\theta_n(\omega)(k) = \omega(n+k) \quad \text{for} \quad \omega = (\omega(k))_{k \geq 0} \in \Omega \quad \text{and} \quad n, k \geq 0.$$

(Recall that $\mathbb{N}_0$ stands for $\mathbb{N} \cup \{0\}$.)

Given a measurable function $G\colon E \to \mathbb{R}$ satisfying the following condition (with $G(X_N) = 0$ if $N = \infty$):

$$\mathsf{E}_x \Big(\sup_{0 \le n \le N} |G(X_n)| \Big) < \infty$$

for all $x \in E$, we consider the optimal stopping problem

$$V^N(x) = \sup_{0 \le \tau \le N} \mathsf{E}_x\, G(X_\tau)$$

where $x \in E$ and the supremum is taken over all stopping times τ of X. The latter means that τ is a stopping time w.r.t. the natural filtration of X given by

$$\mathcal{F}_n^X = \sigma(X_k; 0 \le k \le n) \quad \text{for} \quad n \ge 0.$$

Since the same results remain valid if we take the supremum in

$$V^N(x) = \sup_{0 \le \tau \le N} \mathsf{E}_x\, G(X_\tau) \qquad (*)$$

over stopping times τ w.r.t. $(\mathcal{F}_n)_{n\geq 0}$, and this assumption makes final conclusions more powerful (at least formally), we will assume in the sequel that the supremum in $(*)$ is taken over this larger class of stopping times.

Note also that in $(*)$ we admit that N can be $+\infty$ as well.

In this case, however, we still assume that the supremum is taken over stopping times τ, i.e. over Markov times τ satisfying $0 \leq \tau < \infty$. In this way any specification of $G(X_\infty)$ becomes irrelevant for the problem $(*)$.

To solve

$$V^N(x) = \sup_{0 \le \tau \le N} \mathsf{E}_x\, G(X_\tau) \qquad (*)$$

when $N < \infty$, we may note that by setting $G_n = G(X_n)$ for $n \ge 0$ the problem reduces to the problem

$$\boxed{V_n^N = \sup_{n \le \tau \le N} \mathsf{E}_x\, G_\tau}. \qquad (**)$$

Having identified $(*)$ as $(**)$, we can apply the method of backward induction which leads to a sequence of r.v.'s $(S_n^N)_{0 \le n \le N}$ and a stopping time $\tau_n^N = \inf\{n \le k \le N : S_k^N = G_k\}$.

The key identity is

$$\boxed{S_n^N = V^{N-n}(X_n)} \quad \text{for} \quad 0 \le n \le N, \quad \mathsf{P}_x\text{-a.s.}; \ x \in E \qquad (***)$$

Once $(***)$ is known to hold, the results of the Theorem 1 (finite horizon) from the Martingale theory translate immediately into the present Markovian setting and get a more transparent form.

To get formulation, let us define

$$C_n^N = \{ x \in E : V^{N-n}(x) > G(x) \}$$
$$D_n^N = \{ x \in E : V^{N-n}(x) = G(x) \}$$

for $0 \le n \le N$. We also define stopping time

$$\tau_D = \inf \{ 0 \le n \le N \colon X_n \in D_n^N \}.$$

and the transition operator T of X

$$TF(x) = \mathsf{E}_x\, F(X_1)$$

for $x \in E$ whenever $F \colon E \to \mathbb{R}$ is a measurable function so that $F(X_1)$ is integrable w.r.t. P_x for all $x \in E$.

Theorem 4 (Finite horizon: The time-homogeneous case)

Consider the optimal stopping problems

$$V^n(x) = \sup_{0 \le \tau \le n} \mathsf{E}_x\, G(X_\tau) \qquad (*)$$

assuming that $\mathsf{E}_x \sup_{0 \le k \le N} |G(X_k)| < \infty$. *Then*

I. *Value functions* V^n *satisfy the "Wald–Bellman equation"*

$$V^n(x) = \max(G(x), TV^{n-1}(x)) \qquad (x \in E)$$

for $n = 1, \ldots, N$ *where* $V^0 = G$.

II. *The stopping time* $\tau_D = \inf\{\, 0 \le n \le N \colon X_n \in D_n^N \,\}$ *is optimal in* $(*)$ *for* $n = N$.

III. *If* τ_* *is an optimal stopping time in* $(*)$ *then* $\tau_D \le \tau_*$ (P_x-*a.s.*) *for every* $x \in E$.

IV. *The sequence* $(V^{N-n}(X_n))_{0 \le n \le N}$ *is the smallest supermartingale which dominates* $(G(X_n))_{0 \le n \le N}$ *under* P_x *for* $x \in E$ *given and fixed.*

V. *The stopped sequence* $(V^{N-n}(X_{n \wedge \tau_D}))_{0 \le n \le N}$ *is a martingale under* P_x *for every* $x \in E$*.*

Proof. To verify the equality $S_n^N = V^{N-n}(X_n)$ recall that

$$S_n^N = \mathsf{E}_x(G(X_{\tau_n^N}) \,|\, \mathcal{F}_n) \tag{i}$$

for $0 \le n \le N$. Since $S_k^{N-n} \circ \theta_n = S_{n+k}^N$ we get that τ_n^N satisfies

$$\tau_n^N = \inf\{n \le k \le N : S_k^N = G(X_k)\} = n + \tau_0^{N-n} \circ \theta_n \tag{ii}$$

for $0 \le n \le N$ $(\theta_n \omega(k) = \omega(k+n))$.

Inserting (ii) into (i) and using the Markov property we obtain

$$
\begin{aligned}
S_n^N &= \mathsf{E}_x \Big[G(X_{n+\tau_0^{N-n}\circ\theta_n}) \,|\, \mathcal{F}_n \Big] = \mathsf{E}_x \Big[G(X_{\tau_0^{N-n}}) \circ \theta_n \,|\, \mathcal{F}_n \Big] \\
&= \mathsf{E}_{X_n} G(X_{\tau_0^{N-n}}) \overset{(\alpha)}{=} V^{N-n}(X_n)
\end{aligned} \qquad \text{(iii)}
$$

where (α) follows by (i): $S_n^N = \mathsf{E}_x(G(X_{\tau_n^N}) \,|\, \mathcal{F}_n)$, which imply

$$
\mathsf{E}_x \, S_0^{N-n} = \mathsf{E}_x \, G(X_{\tau_0^{N-n}}) = \sup_{0 \le \tau \le N-n} \mathsf{E}_x \, G(X_\tau) = V^{N-n}(x) \qquad \text{(iv)}
$$

for $0 \le n \le N$ and $x \in E$.

Thus $S_n^N = V^{N-n}(X_n)$ holds as claimed.

To verify the “Wald–Bellman equation”, note that the equality

$$S_n^N = \max\{G_n, \mathsf{E}(S_{n+1}^N \,|\, \mathcal{F}_n)\},$$

using the Markov property, reads as follows:

$$\begin{aligned} V^{N-n}(X_n) &= \max\Big\{G(X_n), \mathsf{E}_x\Big[V^{N-n-1}(X_{n+1}) \,|\, \mathcal{F}_n\Big]\Big\} \\ &= \max\Big\{G(X_n), \mathsf{E}_x\Big[V^{N-n-1}(X_1)\circ\theta_n \,|\, \mathcal{F}_n\Big]\Big\} \\ &= \max\Big\{G(X_n), \mathsf{E}_{X_n} V^{N-n-1}(X_1)\Big\} \\ &= \max\Big\{G(X_n), TV^{N-n-1}(X_n)\Big\} \end{aligned} \qquad (*)$$

for all $0 \le n \le N$. Letting $n=0$ and using that $X_0 = x$ under P_x we see that $(*)$ yields $V^n(x) = \max\{G(x), TV^{n-1}(x)\}$.

The remaining statements of the theorem follow directly from the Martingale Theorem (1). The proof is complete. □

The "Wald–Bellman equation" can be written in a more compact form as follows. Introduce the operator Q by setting

$$QF(x) = \max(G(x), TF(x))$$

for $x \in E$ where $F\colon E \to \mathbb{R}$ is a measurable function for which $F(X_1) \in L^1(\mathsf{P}_x)$ for $x \in E$. Then the "Wald–Bellman equation" reads as follows:

$$V^n(x) = Q^n G(x)$$

for $1 \leq n \leq N$ where Q^n denotes the n-th power of Q. These recursive relations form a constructive method for finding V^N when $\mathrm{Law}(X_1 \,|\, \mathsf{P}_x)$ is known for $x \in E$.

TIME-INHOMOGENEOUS MARKOV CHAINS $X = (X_n)_{n \geq 0}$

Put $Z_n = (n, X_n)$.

$Z = (Z_n)_{n \geq 0}$ is a time-homogeneous Markov chain.

Optimal stopping problem:

$$(*) \qquad \boxed{V^N(n,x) = \sup_{0 \leq \tau \leq N-n} \mathsf{E}_{n,x}\, G(n+\tau, X_{n+\tau})}\,, \qquad 0 \leq n \leq N.$$

We assume

$$(**) \qquad \mathsf{E}_{n,x}\Big(\sup_{0 \leq k \leq N-n} |G(n+k, X_{n+k})|\Big) < \infty, \qquad 0 \leq n \leq N.$$

Theorem 5 (Finite horizon: The time-inhomogeneous case)

Consider the optimal stopping problem $(*)$ *upon assuming that the condition* $(**)$ *holds. Then*:

I. *The function* V^n *satisfies the* "Wald–Bellman equation"

$$V^N(n,x) = \max(G(n,x), TV^N(n,x))$$

for $n = N-1, \dots, 0$ *where*

$$TV^N(n,x) = \mathsf{E}_{n,x}\, V^N(n+1, X_{n+1}), \quad n = N-1, \dots, 0,$$

and

$$TV^N(N-1,x) = \mathsf{E}_{N-1,x}\, G(N, X_N);$$

II. *The stopping time*

$$\tau_D^N = \inf\{n \le k \le N : (n+k, X_{n+k}) \in D\}$$

with

$$D = \Big\{(n,x) \in \{0,1,\dots,N\} \times E : V(n,x) = G(n,x)\Big\}$$

is optimal in the problem $(*)$:

$$V^N(n,x) = \sup_{0 \le \tau \le N-n} \mathsf{E}_{n,x}\, G(n+\tau, X_{n+\tau});$$

III. *If* τ_*^N *is an optimal stopping time in* $(*)$ *then* $\tau_D^N \le \tau_*^N$ ($\mathsf{P}_{n,x}$-*a.s.*) *for every* $(n,x) \in \{0,1,\dots,N\} \times E$;

IV. *The value function* V^N *is the smallest superharmonic function which dominates the gain function* G *on* $\{0,\dots,N\}\times E$,

$$TV^N(n,x) \le V^N(n,x), \quad V^N(n,x) \ge G(n,x);$$

V. *The stopped sequence*

$$\Big(V^N((n+k)\wedge\tau_D^N, X_{(n+k)\wedge\tau_D^N})\Big)_{0\le k\le N-n}$$

is a martingale under $\mathsf{P}_{n,x}$ *for every* $(n,x)\in\{0,1,\dots,N\}\times E$*;*

The proof is carried out in exactly the same way as the proof of Theorem 4.

Optimal stopping for infinite horizon ($N = \infty$):

$$\boxed{V(x) = \sup_{\tau} \mathsf{E}_x\, G(X_\tau)}$$

Theorem 6

Assume $\mathsf{E}_x \sup_{n \ge 0} |G(X_n)| < \infty$, $x \in E$.

I. *The value function* V *satisfies the* "Wald–Bellman equation"

$$V(x) = \max(G(x), TV(x)), \quad x \in E.$$

II. *Assume moreover when required below that* $\mathsf{P}_x(\tau_D < \infty) = 1$ *for all* $x \in E$, *where*

$$\tau_D = \inf\{t \ge 0 : X_t \in D\}$$

with $D = \{x \in E : V(x) = G(x)\}$. *Then the stopping time* τ_D *is optimal.*

III. *If* τ_* *is an optimal stopping time then* $\tau_D \le \tau_*$ (P_x*-a.s. for every* $x \in E$).

IV. *The value function* V *is the smallest superharmonic function (Dynkin's characterization)* ($TV \le V$) *which dominates the gain function* G *on* E, *or, equivalently,* $(V(X_n))_{n\ge 0}$ *is the smallest supermartingale (under* P_x, $x \in E$) *which dominates* $(G(X_n))_{n\ge 0}$.

V. *The stopped sequence* $(V(X_{n\wedge\tau_D}))_{n\ge 0}$ *is a martingale under* P_x *for every* $x \in E$.

VI. *If the condition* $\mathsf{P}_x(\tau_D{<}\infty) = 1$ *fails so that* $\mathsf{P}_x(\tau_D{=}\infty) > 0$ *for some* $x \in E$, *then there is no optimal stopping time in the problem* $V(x) = \sup_\tau \mathsf{E}_x\, G(X_\tau)$ *for all* $x \in E$.

Corollary (Iterative method). *We have*

$$\boxed{V(x) = \lim_{n\to\infty} Q^n G(x)}$$

(*a constructive method for finding the value function* V).

Uniqueness in the Wald–Bellman equation

$$\boxed{F(x) = \max(G(x), TF(x))}$$

Suppose $\mathsf{E}\sup_{n\geq 0} F(X_n) < \infty$.

Then F equals the value function V if and only if the following "*boundary condition at infinity*" holds:

$$\limsup_{n\to\infty} F(X_n) = \limsup_{n\to\infty} G(X_n) \quad \mathsf{P}_x\text{-a.s.} \;\; \forall x \in E.$$

2. Given $\alpha \in (0,1]$ and bounded $g\colon E \to \mathbb{R}$ and $c\colon E \to \mathbb{R}_+$, consider the optimal stopping problem

$$V(x) = \sup_{\tau} \mathsf{E}_x \left(\alpha^{\tau} g(X_\tau) - \sum_{k=1}^{\tau} \alpha^{k-1} c(X_{k-1}) \right).$$

Let $\widetilde{X} = (\widetilde{X}_n)_{n \geq 0}$ denote the Markov chain X killed at rate α. It means that

$$\widetilde{T}F(x) = \alpha\, TF(x).$$

Then

$$V(x) = \sup_{\tau} \mathsf{E}_x \left(g(\widetilde{X}_\tau) - \sum_{k=1}^{\tau} c(\widetilde{X}_{k-1}) \right).$$

The "Wald–Bellman equation" takes the following form:

$$V(x) = \max\Big\{ g(x), \alpha T V(x) - c(x) \Big\}.$$

§ 4. *LECTURES 4–5.*

Theory of optimal stopping for continuous time

A. Martingale approach

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P})$ be a stochastic basis (a filtered probability space with right-continuous family $(\mathcal{F}_t)_{t\geq 0}$ where each $\mathcal{F}_t$ contains all P-null sets from $\mathcal{F}$.

Let $G = (G_t)_{t\geq 0}$ be a gain process. (We interpret G_t as the *gain* if the observation of G is stopped at time t.)

DEFINITION.

A random variable $\tau\colon \Omega \to [0, \infty]$ is called a **Markov time** if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

A Markov time is called a **stopping time** if $\tau < \infty$ P-a.s.

We assume that $G = (G_t)_{t \ge 0}$ is right-continuous and left-continuous over stopping times (if $\tau_n \uparrow \tau$ then $G_{\tau_n} \to G_\tau$ P-a.s.).

We also assume that

$$\mathsf{E}\Big(\sup_{0 \le t \le T} |G_t|\Big) < \infty \qquad (G_T = 0 \ \text{ if } \ T = \infty).$$

BASIC OPTIMAL STOPPING PROBLEM:

$$\boxed{V_t^T = \sup_{t \le \tau \le T} \mathsf{E}\, G_\tau}\,.$$

We shall admit that $T = \infty$. In this case the supremum is still taken over stopping times τ, i.e. over Markov times τ satisfying $t \le \tau < \infty$.

Two ways to tackle the problem $V_t^T = \sup_{t \le \tau \le T} \mathsf{E}\, G_\tau$:

(1) Discrete time approximation

$[0,T] \longrightarrow \mathbb{T}^{(n)} = \left\{t_0^{(n)}, t_1^{(n)}, \ldots, t_n^{(n)}\right\} \uparrow \mathbb{T}$ is a dense subset of $[0,T]$

$G \longrightarrow G^{(n)} = (G_{t_i^{(n)}})$

with applying previous discrete-time results and then passing to the limit $n \to \infty$;

(2) Straightforward extension of the method of essential supremum. This programme will now be addressed.

We denote for simplicity of the notation

$$V_t = V_t^T \qquad (T < \infty \quad \text{or} \quad T = \infty).$$

Consider the process $S = (S_t)_{t \ge 0}$ defined as follows:

$$S_t = \operatorname*{ess\,sup}_{\tau \ge t} \mathsf{E}(G_\tau | \mathcal{F}_t).$$

The process S is the **Snell's envelope** of G.

Introduce

$$\tau_t = \inf \{u \ge t \,|\, S_u = G_u\} \quad \text{where } \inf \varnothing = \infty \text{ by definition.}$$

We shall see below that

$$S_t \ge \max\{G_t, \mathsf{E}(S_u \,|\, \mathcal{F}_t)\} \quad \text{for} \quad u \ge t.$$

The reverse inequality is not true generally.

However,

$$S_t = \max\{G_t, \mathsf{E}(S_{\sigma \wedge \tau_t} \,|\, \mathcal{F}_t)\}$$

for every stopping time $\sigma \ge t$ and τ_t given above.

Theorem 1. *Consider the optimal stopping problem*

$$V_t = \sup_{\tau \geq t} \mathsf{E}\, G_\tau, \quad t \geq 0,$$

upon assuming $\mathsf{E} \sup_{t \geq 0} |G_t| < \infty$*. Assume moreover when required below that*

$$\mathsf{P}(\tau_t < \infty) = 1, \quad t \geq 0.$$

(*Note that this condition is automatically satisfied when the horizon* T *is finite.*) *Then*:

I. *For all* $t \geq 0$ *we have*

$$S_t \geq \mathsf{E}(G_\tau \,|\, \mathcal{F}_t) \quad \textit{for each } \tau \in \mathfrak{M}_t$$
$$S_t = \mathsf{E}(G_{\tau_t} \,|\, \mathcal{F}_t)$$

where $\mathfrak{M}_t = \{\tau : \tau \leq T\}$ *if* $T < \infty$,
$\mathfrak{M}_t = \{\tau : \tau < \infty\}$ *if* $T = \infty$.

II. *The stopping time* $\tau_t = \inf\{u \geq t : S_u = G_u\}$ *is optimal (for the problem* $V_t = \sup_{\tau \geq t} \mathsf{E}\, G_\tau$*).*

III. *If* τ_t^* *is an optimal stopping time as well then* $\tau_t \leq \tau_t^*$ P-*a.s.*

IV. *The process* $(S_u)_{u \geq t}$ *is the smallest right-continuous supermartingale which dominates* $(G_s)_{s \geq t}$.

V. *The stopped process* $(S_{u \wedge \tau_t})_{u \geq t}$ *is a right-continuous martingale.*

VI. *If the condition* $\mathsf{P}(\tau_t < \infty) = 1$ *fails so that* $\mathsf{P}(\tau_t = \infty) > 0$, *then there is no optimal stopping time.*

Proof. 1°. Let us first prove that $S = (S_t)_{t\ge 0}$ defined by

$$S_t = \operatorname*{ess\,sup}_{\tau \ge t} \mathsf{E}(G_\tau \,|\, \mathcal{F}_t)$$

is a supermartingale.

Show that the family $\{\mathsf{E}(G_\tau \,|\, \mathcal{F}_t) : \tau \in \mathfrak{M}_t\}$ is upwards directed in the sense that if σ_1 and σ_2 are from $\mathfrak{M}_t$ then there exists $\sigma_3 \in \mathfrak{M}_t$ such that

$$\mathsf{E}(G_{\sigma_1} \,|\, \mathcal{F}_t) \vee \mathsf{E}(G_{\sigma_2} \,|\, \mathcal{F}_t) \le \mathsf{E}(G_{\sigma_3} \,|\, \mathcal{F}_t).$$

Put $\sigma_3 = \sigma_1 I_A + \sigma_2 I_{\bar{A}}$ where

$$A = \{\mathsf{E}(G_{\sigma_1} | \mathcal{F}_t) \ge \mathsf{E}(G_{\sigma_2} \,|\, \mathcal{F}_t)\}.$$

Then $\sigma_3 \in \mathfrak{M}_t$ and

$$\begin{aligned}
\mathsf{E}(G_{\sigma_3} \,|\, \mathcal{F}_t) &= \mathsf{E}(G_{\sigma_1} I_A + G_{\sigma_2} I_{\bar{A}} \,|\, \mathcal{F}_t) = I_A \,\mathsf{E}(G_{\sigma_1} \,|\, \mathcal{F}_t) + I_{\bar{A}} \,\mathsf{E}(G_{\sigma_2} \,|\, \mathcal{F}_t) \\
&= \mathsf{E}(G_{\sigma_1} \,|\, \mathcal{F}_t) \vee \mathsf{E}(G_{\sigma_2} \,|\, \mathcal{F}_t).
\end{aligned}$$

Hence there exists a sequence $\{\sigma_k; k \geq 1\}$ in $\mathfrak{M}_t$ such that

$$(*) \qquad \operatorname*{ess\,sup}_{\tau \in \mathfrak{M}_t} \mathsf{E}(G_\tau | \mathcal{F}_t) = \lim_{k\to\infty} \mathsf{E}(G_{\sigma_k} | \mathcal{F}_t)$$

where

$$\mathsf{E}(G_{\sigma_1} | \mathcal{F}_t) \leq \mathsf{E}(G_{\sigma_2} | \mathcal{F}_t) \leq \cdots \quad \mathsf{P}\text{-a.s.}$$

From $(*)$ and the conditional monotone convergence theorem (using $\mathsf{E} \sup_{t\geq 0} |G_t| < \infty$) we find that for $0 \leq s < t$

$$\begin{aligned} \mathsf{E}(S_t \,|\, \mathcal{F}_s) &= \mathsf{E}\Big(\lim_{k\to\infty} \mathsf{E}(G_{\sigma_k} | \mathcal{F}_t) \,|\, \mathcal{F}_s \Big) \\ &= \lim_{k\to\infty} \mathsf{E}[\mathsf{E}(G_{\sigma_k} | \mathcal{F}_t) \,|\, \mathcal{F}_s] \\ &= \lim_{k\to\infty} \mathsf{E}(G_{\sigma_k} | \mathcal{F}_s) \leq S_s \quad \Big(= \operatorname*{ess\,sup}_{\tau \geq s} \mathsf{E}(G_\tau \,|\, \mathcal{F}_s) \Big). \end{aligned}$$

Thus $(S_t)_{t\geq 0}$ is a supermartingale as claimed.

Note that from $\mathsf{E}\sup_{t\ge 0}|G_t| < \infty$ and

$$S_t = \operatorname*{ess\,sup}_{\tau\ge t} \mathsf{E}(G_\tau \,|\, \mathcal{F}_t),$$
$$\operatorname*{ess\,sup}_{\tau\ge t} \mathsf{E}(G_\tau \,|\, \mathcal{F}_t) = \lim_{k\to\infty} \mathsf{E}(G_{\sigma_k} \,|\, \mathcal{F}_t)$$

it follows that

$$\boxed{\mathsf{E}\, S_t = \sup_{\tau\ge t} \mathsf{E}\, G_\tau}\,.$$

2°. Let us next show that the supermartingale S admits a right-continuous modification $\widetilde{S} = (\widetilde{S}_t)_{t\ge 0}$.

From the general martingale theory it follows that it suffices to check that

$$t \rightsquigarrow \mathsf{E}\, S_t \quad \text{is right-continuous on } \mathbb{R}_+.$$

By the supermartingale property of S

$$\mathsf{E}\, S_t \geq \cdots \geq \mathsf{E}\, S_{t_2} \geq \mathsf{E}\, S_{t_1}, \qquad t_n \uparrow t.$$

So, $L := \lim_{n\to\infty} \mathsf{E}\, S_{t_n}$ exists and

$$\mathsf{E}\, S_t \geq L.$$

To prove the reverse inequality, fix $\varepsilon > 0$ and by means of $\mathsf{E}\, S_t = \sup_{\tau \geq t} \mathsf{E}\, G_\tau$ choose $\sigma \in \mathfrak{M}_t$ such that

$$\mathsf{E}\, G_\sigma \geq \mathsf{E}\, S_t - \varepsilon.$$

Fix $\delta > 0$ and note that there is no restriction to assume that $t_n \in [t, t+\delta]$ for all $n \geq 1$. Define

$$\sigma_n = \begin{cases} \sigma & \text{if } \sigma > t_n, \\ t + \sigma & \text{if } \sigma \leq t_n. \end{cases}$$

Then for all $n \geq 1$ we have

$$(*) \qquad \mathsf{E}\, G_{\sigma_n} = \mathsf{E}\, G_\sigma I(\sigma > t_n) + \mathsf{E}\, G_{t+\delta} I(\sigma \leq t_n) \leq \mathsf{E}\, S_{t_n}$$

since $\sigma_n \in \mathfrak{M}_{t_n}$ and $\mathsf{E}\, S_t = \sup_{\tau \geq t} \mathsf{E}\, G_\tau$. Letting $n \to \infty$ in $(*)$ and assuming that $\mathsf{E} \sup_{0 \leq t \leq T} |G_t| < \infty$ we get

$$\mathsf{E}\, G_\sigma I(\sigma > t) + \mathsf{E}\, G_{t+\delta} I(\sigma = t) \leq L \quad (= \lim_n \mathsf{E}\, S_{t_n}).$$

Letting now $\delta \downarrow 0$ and using that G is right-continuous we obtain

$$\mathsf{E}\, G_\sigma I(\sigma > t) + \mathsf{E}\, G_t I(\sigma = t) = \mathsf{E}\, G_\sigma \leq L.$$

From here and $\mathsf{E}\, G_\sigma \geq \mathsf{E}\, S_t - \varepsilon$ we see that $L \geq \mathsf{E}\, S_t - \varepsilon$ for all $\varepsilon > 0$. Hence $L \geq \mathsf{E}\, S_t$ and thus

$$\lim_{n \to \infty} \mathsf{E}\, S_{t_n} = L = \mathsf{E}\, S_t, \qquad t_n \uparrow t,$$

showing that S admits a right-continuous modification $\widetilde{S} = (\widetilde{S}_t)_{t \geq 0}$ which we also denote by S throughout.

Let us prove property IV:

> The process $(S_u)_{u\ge t}$ is the smallest right-continuous supermartingale which dominates $(G_s)_{s\ge t}$.

For this, let $\widehat{S}=(\widehat{S}_u)_{u\ge t}$ be another right-continuous supermartingale which dominates $G=(G_u)_{u\ge t}$. Then by the optional sampling theorem (using $\mathsf{E}\sup_{t\ge 0}|G_t|<\infty$) we have

$$\widehat{S}_u \ge \mathsf{E}(\widehat{S}_\tau \,|\, \mathcal{F}_u) \ge \mathsf{E}(G_\tau \,|\, \mathcal{F}_u)$$

for all $\tau\in\mathfrak{M}_u$ when $u\ge t$. Hence by the definition $S_u=\operatorname*{ess\,sup}_{\tau\ge u}\mathsf{E}(G_\tau\,|\,\mathcal{F}_u)$ we find that $S_u\le\widehat{S}_u$ (P-a.s.) for all $u\ge t$. By the right-continuity of S and $\widehat{S}$ this further implies that

$$\mathsf{P}(S_u\le\widehat{S}_u \text{ for all } u\ge t)=1$$

as claimed.

Property I: for all $t \geq 0$

$$(*) \qquad S_t \geq \mathsf{E}(G_\tau \,|\, \mathcal{F}_t) \quad \text{for each } \tau \in \mathfrak{M}_t,$$

$$(**) \qquad S_t = \mathsf{E}(G_{\tau_t} \,|\, \mathcal{F}_t).$$

The inequality $(*)$ follows from the definition $S_t = \operatorname{ess\,sup}_{\tau \geq t} \mathsf{E}(G_\tau \,|\, \mathcal{F}_t)$. The proof of $(**)$ is the most difficult part of the proof of the Theorem.

The sketch of the proof is as follows.

Assume that $G_t \geq 0$ for all $t \geq 0$.

(α) Introduce, for $\lambda \in (0,1)$, the stopping time

$$\tau_t^\lambda = \inf\{s \geq t : \lambda S_s \leq G_s\}$$

(Then $\lambda S_{\tau_t^\lambda} \leq G_{\tau_t^\lambda}$, $\tau_{t+}^\lambda = \tau_t$.)

(β) We show that

$$S_t = \mathsf{E}(S_{\tau_t^\lambda} \,|\, \mathcal{F}_t) \quad \text{for all} \quad \lambda \in (0,1).$$

So $S_t \leq (1/\lambda)\, \mathsf{E}(G_{\tau_t^\lambda} \,|\, \mathcal{F}_t)$ and letting $\lambda \uparrow 1$ we get

$$S_t \leq \mathsf{E}(G_{\tau_t^1} | \, \mathcal{F}_t)$$

where $\tau_t^1 = \lim_{\lambda \uparrow 1} \tau_t^\lambda$ ($\tau_t^\lambda \uparrow$ when $\lambda \uparrow$).

(γ) Verify that $\tau_t^1 = \tau_t$. Then $S_t \leq \mathsf{E}(G_{\tau_t} | \, \mathcal{F}_t)$ and evidently $S_t \geq \mathsf{E}(G_{\tau_t} | \, \mathcal{F}_t)$. Thus $S_t = \mathsf{E}(G_{\tau_t} | \, \mathcal{F}_t)$.

For the proof of property V:

> The stopped process $(S_{u\wedge\tau_t})_{u\geq t}$ is a right-continuous martingale

it is enough to prove that

$$\mathsf{E}\, S_{\sigma\wedge\tau_t} = \mathsf{E}\, S_t$$

for all bounded stopping times $\sigma \geq t$.

The optional sampling theorem implies

$$\mathsf{E}\, S_{\sigma\wedge\tau_t} \leq \mathsf{E}\, S_t. \tag{57}$$

On the other hand, from $S_t = \mathsf{E}(G_{\tau_t}\,|\,\mathcal{F}_t)$ and $S_{\tau_t} = G_{\tau_t}$ we see that

$$\mathsf{E}\, S_t = \mathsf{E}\, G_{\tau_t} = \mathsf{E}\, S_{\tau_t} \leq \mathsf{E}\, S_{\sigma\wedge\tau_t}.$$

Thus, $\mathsf{E}\, S_{\sigma\wedge\tau_t} = \mathsf{E}\, S_t$ and $(S_{u\wedge\tau_t})_{u\geq t}$ is a martingale. $\square$

B. Markovian approach

Let $X = (X_t)_{t\geq 0}$ be a strong Markov process defined on a filtered probability space

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathsf{P}_x)$$

where $\quad x \in E \ (= \mathbb{R}^d), \quad \mathsf{P}_x(X_0 = x) = 1,$

$\quad x \to \mathsf{P}_x(A)$ is measurable for each $A \in \mathcal{F}$.

Without loss of generality we will assume that

$$(\Omega, \mathcal{F}) = (E^{[0,\infty)}, \mathcal{B}^{[0,\infty)}) \quad \text{(canonical space)}$$

Shift operator $\theta_t = \theta_t(\omega)\colon \Omega \to \Omega$ is well defined by

$$\theta_t(\omega)(s) = \omega(t+s) \quad \text{for} \quad \omega = (\omega(s))_{s\geq 0} \in \Omega \quad \text{and} \quad t, s \geq 0.$$

We consider the optimal stopping problem

$$\boxed{V(x) = \sup_{0 \le \tau \le T} \mathsf{E}_x\, G(X_\tau)}$$

$$G(X_T) = 0 \quad \text{if} \quad T < \infty; \qquad \mathsf{E}_x \sup_{0 \le t \le T} |G(X_t)| < \infty.$$

Here $\tau = \tau(\omega)$ is a stopping time w.r.t.

$$(\mathcal{F}_t)_{t \ge 0} \quad (\mathcal{F}_t^X \subseteq \mathcal{F}_t,\ \ \mathcal{F}_t^X = \sigma(X_s; 0 \le s \le t)).$$

G is called the **gain function**,
V is called the **value function**.

CASE $T = \infty$:

$$\boxed{\begin{array}{c} V(x) = \sup_{\tau} \mathsf{E}_x \, G(X_\tau) \\ \mathsf{P}_x(X_0 = x) = 1 \end{array}}$$

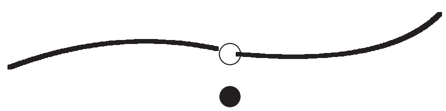
Introduce

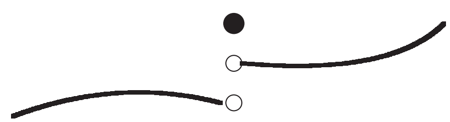
the ***continuation set*** $C = \{x \in E : V(x) > G(x)\}$ and

the ***stopping set*** $D = \{x \in E : V(x) = G(x)\}$

<u>**NOTICE!**</u> If

V is lsc (lower semicontinuous) & G is usc (upper semicontinuous)





then

$$\boxed{C \text{ is open} \quad \text{and} \quad D \text{ is closed}}$$

The first entry time

$$\tau_D = \inf\{t \geq 0 : X_t \in D\}$$

for *closed* D is a stopping time since both X and $(\mathcal{F}_t)_{t\geq 0}$ are right-continuous.

DEFINITION. A measurable function $F = F(x)$ is said to be *superharmonic* (for X) if

$$\mathsf{E}_x\, F(X_\sigma) \leq F(x)$$

for all stopping times σ and all $x \in E$. (It is assumed that $F(X_\sigma) \in L^1(\mathsf{P}_x)$ for all $x \in E$ whenever σ is a stopping time.)

We have:

$\boxed{F \text{ is superharmonic}}$ iff $\boxed{\begin{array}{l}(F(X_t))_{t\geq 0} \text{ is a supermartingale} \\ \text{under } \mathsf{P}_x \text{ for every } x \in E.\end{array}}$

The following theorem presents

NECESSARY CONDITIONS

for the existence of an optimal stopping time.

Theorem. *Let us assume that there exists an optimal stopping time τ_* in the problem*

$$V(x) = \sup_{\tau} \mathsf{E}_x\, G(X_\tau)$$

i.e. $V(x) = \mathsf{E}_x\, F(X_{\tau_*})$. *Then*

(I) *The value function V is the smallest superharmonic function (**Dynkin's characterization**) which dominates the gain function G on E.*

Let us in addition to "$V(x) = \mathsf{E}_x\, F(X_{\tau_*})$" assume that

$$V \text{ is lsc and } G \text{ is usc.}$$

Then

(II) *The stopping time* $\tau_D = \inf\{t \geq 0 : X_t \in D\}$ *satisfies*

$$\tau_D \leq \tau_* \quad (\mathsf{P}_x\text{-a.s., } \; x \in E)$$

and is optimal;

(III) *The stopped process* $(V(X_{t \wedge \tau_D}))_{t \geq 0}$ *is a right-continuous martingale under* P_x *for every* $x \in E$.

Now we formulate

SUFFICIENT CONDITIONS

for the existence of an optimal stopping time.

Theorem. *Consider the optimal stopping problem*

$$V(x) = \sup_{\tau} \mathsf{E}_x\, G(X_\tau)$$

upon assuming that the condition

$$\mathsf{E}_x \sup_{t \ge 0} |G(X_t)| < \infty, \quad x \in E,$$

is satisfied.

Let us assume that there exists the smallest superharmonic function $\widehat{V}$ *which dominates the gain function* G *on* E.

Let us in addition assume that

$$\widehat{V} \text{ is lsc } \text{ and } G \text{ is usc.}$$

Set $D = \{x \in E : \widehat{V}(x) = G(x)\}$ *and let* $\tau_D = \inf\{t : X_t \in D\}$.

We then have:

(a) *If* $\mathsf{P}_x(\tau_D < \infty) = 1$ *for all* $x \in E$, *then* $\widehat{V} = V$ *and* τ_D *is optimal in* $V(x) = \sup_\tau \mathsf{E}_x\, G(X_\tau)$;

(b) *If* $\mathsf{P}_x(\tau_D < \infty) < 1$ *for some* $x \in E$, *then there is no optimal stopping time in* $V(x) = \sup_\tau \mathsf{E}_x\, G(X_\tau)$.

Corollary (The existence of an optimal stopping time).

Infinite horizon $(T = \infty)$. *Suppose that* V *is lsc and* G *is usc. If* $\mathsf{P}_x(\tau_D < \infty) = 1$ *for all* $x \in E$*, then* τ_D *is optimal. If* $\mathsf{P}_x(\tau_D < \infty) < 1$ *for some* $x \in E$*, then there is no optimal stopping time.*

Finite horizon $(T < \infty)$. *Suppose that* V *is lsc and* G *is usc. Then* τ_D *is optimal.*

Proof for $T = \infty$. (The case $T < \infty$ can be proved in exactly the same way as the case $T = \infty$ if the process (X_t) is replaced by the process (t, X_t).)

The key is to show that V is SUPERHARMONIC.

If so, then evidently V is the **smallest superharmonic function** which dominates G on E. Then the claims of the corollary follow directly from the Theorem (on sufficient conditions) above.

For this, note that V is measurable (since it is lsc) and thus so is the mapping

$$(*) \qquad V(X_\sigma) = \sup_\tau \mathsf{E}_{X_\sigma} G(X_\tau)$$

for any stopping time σ which is given and fixed.

On the other hand, by the strong Markov property we have

$$(**) \qquad \mathsf{E}_{X_\sigma} G(X_\tau) = \mathsf{E}_x \left[\, G(X_{\sigma+\tau\circ\theta_\sigma}) \,|\, \mathcal{F}_\sigma \right]$$

for every stopping time τ and $x \in E$. From $(*)$ and $(**)$ we see that

$$V(x_\sigma) = \operatorname*{ess\,sup}_\tau \mathsf{E}_x \left[\, G(X_{\sigma+\tau\circ\theta_\sigma}) \,|\, \mathcal{F}_\sigma \right]$$

under P_x where $x \in E$ is given and fixed.

We can show that the family

$$\Big\{ \mathsf{E}[X_{\sigma+\tau\circ\theta_\sigma} \,|\, \mathcal{F}_\sigma] : \tau \text{ is a stopping time} \Big\}$$

is upwards directed: if $\rho_1 = \sigma + \tau_1 \circ \theta_\sigma$ and $\rho_2 = \sigma + \tau_2 \circ \theta_\sigma$ then there is $\rho = \sigma + \tau \circ \theta_\sigma$ such that

$$\mathsf{E}[G(X_\rho) \,|\, \mathcal{F}_\sigma] = \mathsf{E}[G(X_{\rho_1}) \,|\, \mathcal{F}_\sigma] \vee \mathsf{E}[G(X_{\rho_2}) \,|\, \mathcal{F}_\sigma].$$

From here we can conclude that there exists a sequence of stopping times $\{\tau_n; n \geq 1\}$ such that

$$V(X_\sigma) = \lim_n \mathsf{E}_x[G(X_{\sigma+\tau_n\circ\theta_\sigma}) \,|\, \mathcal{F}_n]$$

where the sequence $\{\mathsf{E}_x[G(X_{\sigma+\tau_n\circ\theta_\sigma}) \,|\, \mathcal{F}_n]\}$ is *increasing* P_x-a.s.

By the monotone convergence theorem using $\mathsf{E} \sup_{t \ge 0} |G_t| < \infty$ we can conclude

$$\mathsf{E}_x V(X_\sigma) = \lim_n \mathsf{E}_x G(X_{\sigma + \tau_n \circ \theta_\sigma}) \le V(x)$$

for all stopping times σ and all $x \in E$. This proves that V is superharmonic.

REMARK 1. If the function

$$x \mapsto \mathsf{E}_x G(X_\tau)$$

is continuous (or lsc) for every stopping time τ, then $x \mapsto V(x)$ is lsc and the results of the Corollary are applicable. This yields a powerful existence result by simple means.

REMARK 2. The above results have shown that the optimal stopping problem

$$V(x) = \sup_{\tau} \mathsf{E}_x\, G(X_\tau)$$

is equivalent to the problem of finding the **smallest superharmonic function** $\widehat{V}$ which dominates G on E. Once $\widehat{V}$ is found it follows that $V = \widehat{V}$ and $\tau_D = \inf\{t : G(X_t) = \widehat{V}(X_t)\}$ is optimal.

There are two traditional ways for finding $\widehat{V}$:

(i) **Iterative procedure** (constructive but non-explicit)

(ii) **Free-boundary problem** (explicit or non-explicit).

For (i), e.g., it is known that if G is lsc and

$$\mathsf{E}_x \inf_{t \ge 0} G(X_t) > -\infty \quad \text{for all} \quad x \in E,$$

then $\widehat{V}$ can be computed as follows:

$$\widehat{V}(x) = \lim_{n \to \infty} \lim_{N \to \infty} Q_n^N G(x)$$

where

$$Q_n G(x) := G(x) \vee \mathsf{E}_x\, G(X_{1/2^n})$$

and Q_n^N is the N-th power of Q_n.

The basic idea (ii) is that

$$\widehat{V} \quad \text{and} \quad C \text{ (or } D\text{)}$$

should solve the free-boundary problem:

$$(*) \quad \mathbb{L}_X \widehat{V} \leq 0$$

$$(**) \quad \widehat{V} \geq G \quad (\widehat{V} > G \text{ on } C \;\;\&\;\; \widehat{V} = G \text{ on } D)$$

where $\mathbb{L}_X$ is the characteristic (infinitesimal) operator of X.

Assuming that G is smooth in a neighborhood of ∂C the following "rule of thumb" is valid.

If X after starting at ∂C enters immediately into int(D) (e.g. when X is a diffusion process and ∂C is sufficiently nice) then the condition $\mathbb{L}_X \widehat{V} \le 0$ under $(**)$ splits into the two conditions:

$$\mathbb{L}_X \widehat{V} = 0 \quad \text{in } C$$

$$\frac{\partial \widehat{V}}{\partial x}\bigg|_{\partial C} = \frac{\partial G}{\partial x}\bigg|_{\partial C} \quad (\textit{smooth fit}).$$

On the other hand, if X after starting at ∂C does not enter immediately into int(D) (e.g. when X has jumps and no diffusion component while ∂C may still be sufficiently nice) then the condition $\mathbb{L}_X \widehat{V} \le 0$ (i.e. $(*)$) under $(**)$ splits into the two conditions:

$$\mathbb{L}_X \widehat{V} = 0 \quad \text{in } C$$

$$\widehat{V}\Big|_{\partial C} = G\Big|_{\partial C} \quad (\textit{continuous fit}).$$

Proof of the Theorem on *NECESSARY* conditions

Basic lines

(I) The value function V is the smallest superharmonic function which dominated the gain function G on E.

We have by the strong Markov property:

$$\begin{aligned} \mathsf{E}_x\, V(X_\sigma) &= \mathsf{E}_x\, \mathsf{E}_{X_\sigma}\, G(X_{\tau_*}) = \mathsf{E}_x\, \mathsf{E}_x[G(X_{\tau_*}) \circ \theta_\sigma \,|\, \mathcal{F}_\sigma] \\ &= \mathsf{E}_x\, G(X_{\sigma+\tau_*\circ\theta_\sigma}) \le \sup_\tau \mathsf{E}_x\, G(X_\tau) = V(x) \end{aligned}$$

for each stopping time σ and all $x \in E$.

Thus V is superharmonic.

Let F be a superharmonic function which dominates G on E. Then

$$\mathsf{E}_x\, G(X_\tau) \le \mathsf{E}_x\, F(X_\tau) \le F(x)$$

for each stopping time τ and all $x \in E$. Taking the supremum over all τ we find that $V(x) \le F(x)$ for all $x \in E$. Since V is superharmonic itself, this proves that V is the smallest superharmonic function which dominated G.

(II) Let us show that the stopping time

$$\tau_D = \inf\{t : V(X_t) = G(X_t)\}$$

is optimal (if V is lsc and G is usc).

We assume that there exists an optimal stopping time τ_*:

$$V(x) = \mathsf{E}_x\, G(X_{\tau_*}), \quad x \in E.$$

We claim that $V(X_{\tau_*}) = G(X_{\tau_*})$ P_x-a.s. for all $x \in E$.

Indeed, if $\mathsf{P}_x\{V(X_{\tau_*}) > G(X_{\tau_*})\} > 0$ for some $x \in E$, then

$$\mathsf{E}_x\, G(X_{\tau_*}) < \mathsf{E}_x\, V(X_{\tau_*}) \le V(x)$$

since V is superharmonic, leading to a contradiction with the fact that τ_* is optimal. From the identity just verified it follows that

$$\tau_D \le \tau_* \quad \mathsf{P}_x\text{-a.s.} \quad \text{for all} \quad x \in E.$$

By (I) the value function V is the superharmonic ($\mathsf{E}_x\, V(X_\sigma) \le V(X)$ for all stopping time σ and $x \in E$). Setting $\sigma \equiv s$ and using the Markov property we get for all $t, s \ge 0$ and all $x \in E$

$$V(X_t) \ge \mathsf{E}_{X_t}\, V(X_s) = \mathsf{E}_x\, [V(X_{t+s}) \,|\, \mathcal{F}_t].$$

This shows that

The process $(V(X_t))_{t\ge 0}$ is a supermartingale under P_x for each $x \in E$.

Suppose for the moment that V is **continuous**. Then obviously it follows that $(V(X_t))_{t\ge 0}$ is **right-continuous**. Thus, by the optional sampling theorem (using $\mathsf{E} \sup_{t\ge 0} |G(X_t)| < \infty$), we see that

$$\mathsf{E}_x\, V(X_\tau) \le \mathsf{E}_x\, V(X_\sigma) \quad \text{for} \quad \sigma \le \tau.$$

In particular, since $\tau_D \le \tau_*$ we get

$$\begin{aligned} V(x) &= \mathsf{E}_x\, G(X_{\tau_*}) = \mathsf{E}_x\, V(X_{\tau_*}) \\ &\le \mathsf{E}_x\, V(X_{\tau_D}) = \mathsf{E}_x\, G(X_{\tau_D}) \le V(x), \end{aligned}$$

where we used that

$$V(X_{\tau_D}) = G(X_{\tau_D})$$

Now it is easy to show that τ_D is optimal if V is continuous.

If V is only lsc, then again (see the lemma below) the process $(V(X_t))_{t \geq 0}$ is right-continuous (P_x-a.s. for each $x \in E$), and the proof can be completed as above.

This shows that τ_D is optimal if V is lsc as claimed.

Lemma. *If a superharmonic function $F : E \to \mathbb{R}$ is lsc, then the supermartingale $(F(X_t))_{t \geq 0}$ is right-continuous (P_x-a.s. for each $x \in E$).*

We omit the proof.

(III) The stopped process $(V(X_{t\wedge\tau_D}))_{t\geq 0}$ is a right-continuous martingale under P_x for every $x \in E$.

PROOF. By the strong Markov property we have

$$\begin{aligned}
\mathsf{E}_x[V(X_{t\wedge\tau_D}) \,|\, \mathcal{F}_{s\wedge\tau_D}] &= \mathsf{E}_x\Big[\,\mathsf{E}_{X_{t\wedge\tau_D}} G(X_{\tau_D}) \,|\, \mathcal{F}_{s\wedge\tau_D}\Big] \\
&= \mathsf{E}_x\Big(\mathsf{E}_x[G(X_{\tau_D}) \circ \theta_{t\wedge\tau_D} \,|\, \mathcal{F}_{t\wedge\tau_D}] \,|\, \mathcal{F}_{s\wedge\tau_D}\Big) \\
&= \mathsf{E}_x\Big(\mathsf{E}_x[G(X_{\tau_D}) \,|\, \mathcal{F}_{t\wedge\tau_D}] \,|\, \mathcal{F}_{s\wedge\tau_D}\Big) = \mathsf{E}_x[G(X_{\tau_D}) \,|\, \mathcal{F}_{s\wedge\tau_D}] \\
&= \mathsf{E}_{X_{s\wedge\tau_D}} G(X_{\tau_D}) = V(X_{s\wedge\tau_D})
\end{aligned}$$

for all $0 \leq s \leq t$ and all $x \in E$ proving the martingale property. The right-continuity of $\Big(V(X_{t\wedge\tau_D})\Big)_{t\geq 0}$ follows from the right-continuity of $(V(X_t))_{t\geq 0}$ that we proved above.

The proof of the theorem on necessary conditions is complete.

REMARK. The result and proof of the Theorem extend in exactly the same form (by slightly changing the notation only) to the *finite* horizon problem

$$V_T(X) = \sup_{0 \le \tau \le T} \mathsf{E}_x\, G(X_\tau).$$

Now we formulate the theorem which provides

sufficient condition

for the existence of an optimal stopping time.

THEOREM. *Consider the optimal stopping problem*

$$V(x) = \sup_{\tau} \mathsf{E}_x\, G(X_\tau)$$

upon assuming that $\mathsf{E}_x \sup_{t\ge 0} |G(X_t)| < \infty$, $x \in E$. *Let us assume that*

(a) *there exists the smallest superharmonic function* $\widehat{V}$ *which dominates the gain function* G *on* E;

(b) $\widehat{V}$ *is lsc and* G *is usc.*

Set $D = \{x \in E : \widehat{V}(x) = G(x)\}$ *and* $\tau_D = \inf\{t : X_t \in D\}$.

We then have:

(I) *If* $\mathsf{P}_x(\tau_D < \infty) = 1$ *for all* $x \in E$, *then* $\widehat{V} = V$ *and* τ_D *is optimal*;

(II) *If* $\mathsf{P}_x(\tau_D < \infty) < 1$ *for some* $x \in E$, *then there is no optimal stopping time.*

SKETCH OF THE PROOF.

(I) Since $\widehat{V}$ is superharmonic majorant for G, we have

$$\mathsf{E}_x\, G(X_\tau) \le \mathsf{E}_x\, \widehat{V}(X_\tau) \le V(x)$$

for all stopping times τ and all $x \in E$. So

$$G(x) \le V(x) = \sup_\tau \mathsf{E}_x\, G(X_\tau) \le \widehat{V}(x)$$

for all $x \in E$.

Next step (difficult!): assuming that $\mathsf{P}_x(\tau_D < \infty) = 1$ for all $x \in E$, we prove the inequality

$$\widehat{V}(x) \le V(x)$$

and optimality of time τ_D.

(II) If $\mathsf{P}_x(\tau_D < \infty) < 1$ for some $x \in E$ then there is no optimal stopping time.

Indeed, by "necessary-condition theorem" if there exists optimal optimal τ_* then $\tau_D \le \tau_*$.

But τ_D takes value ∞ with positive probability for some $x \in E$.

So, for this state x we have $\mathsf{P}_x(\tau_* = \infty) > 0$ and τ_* cannot be optimal (in the class $\mathfrak{M} = \{\tau : \tau < \infty\}$). $\square$

TOPIC III: Quickest detection problems

§ 1. W. Shewhart and E. Page's works

1. Let us describe the (chronologically) first approaches to the Quickest Detection (QD) problems initiated in the 1920-30s by W. Shewhart who proposed – to control industrial products – the so-called **control charts** (which are used till now). The next step in this direction was made in the 1950s by E. Page who invented the so-called **CUSUM method**, which became very popular in the statistical practice. None of these approaches was underlain by any deep stochastic analysis.

In the late 1950s A. N. Kolmogorov and the author gave precise mathematical formulation of two QD-problems. The basic problem was a **multistage** problem of quickest detection of the random target which appears in the steady stationary regime under assumption that the mean time between false alarms is large (see Topic V).

The second problem was a Bayesian problem whose solution became a crucial step in solving the first problem.

2. W. Shewart approach * supposes that $x_1, x_2, \ldots$ are observations on the random variables $X_1, X_2, \ldots$ and θ is an unknown parameter ("hidden parameter") which takes values in the set $\{1, 2, \ldots, \infty\}$.

The case $\boldsymbol{\theta = \infty}$ is interpreted as a "normal" run of the inspected industrial process. In this case $X_1, X_2, \ldots$ are i.i.d. random variables with the density $f^{\infty}(x)$.

If $\boldsymbol{\theta = 0}$ or $\boldsymbol{\theta = 1}$, then $X_1, X_2, \ldots$ are again i.i.d. with density $f^{0}(x)$.

If $\mathbf{1 < \boldsymbol{\theta} < \infty}$, then $X_1, \ldots, X_{\theta-1}$ are i.i.d. with the density $f^{\infty}(x)$ and $X_{\theta}, X_{\theta+1}, \ldots$ run with the density $f^{0}(x)$: $\boxed{f^{\infty}(x) \xrightarrow{\theta} f^{0}(x).}$

---

* W. A. Shewhart, "The application of statistics as an aid in maintaining quality of manufactured product", J. Amer. Statist. Assoc., 138 (1925), 546-548.

W. A. Shewhart, Economic Control of Manufactured Product, Van Nostrand Rainhold, N.Y., 1931. (Republished in 1981 by the Amer. Soc. for Quality Control, Milwaukee.)

Alarm signal is a random stopping time $\tau = \tau(x)$, $x = (x_1, x_2, \ldots)$ such that $\tau(x) = \inf\{n \geq 1 \colon x_n \in D\}$, where D is some set in the space of states x.

For $f^0(x) \sim \mathcal{N}(\mu_0, \sigma)$, $f^\infty(x) \sim \mathcal{N}(\mu_\infty, \sigma)$ Shewart proposes to take

$$\tau(x) = \inf\{n \geq 1 \colon |x_n - \mu_\infty| \geq 3\sigma\}.$$

It is easy to find the probability of false alarm (on each step)

$$\alpha \equiv \mathsf{P}_{(\mu_\infty,\sigma)}\{|X_1 - \mu_\infty| \geq 3\sigma\} \approx 0.0027.$$

For $\mathsf{E}_{(\mu_\infty,\sigma)}\tau$ we find

$$\mathsf{E}_{(\mu_\infty,\sigma)}\tau = \sum_{k=1}^{\infty} k\alpha(1-\alpha)^{k-1} = \frac{1}{\alpha} \approx 370.$$

Similarly we can find the probability of the correct alarm $\beta = \mathsf{P}_{(\mu_0,\sigma)}\{|X_1 - \mu_\infty| \geq 3\sigma\}$ and $\mathsf{E}_{(\mu_0,\sigma)}\tau$.

W. Shewhart did not formulated optimization problems.

A possible approach can be as follows.

Let $\mathfrak{M}_T = \{\tau\colon \mathsf{E}^\infty \tau \ge T\}$, where T is a fixed constant. The stopping time τ_T^* is called a minimax if

$$\boxed{\sup_\theta \mathsf{E}^\theta\Big(\tau_T^* - \theta \,|\, \tau_T^* \ge \theta\Big) = \inf_{\tau \in \mathfrak{M}_T} \sup_\theta \mathsf{E}^\theta(\tau - \theta \,|\, \tau \ge \theta)}$$

(here P^θ is the distribution on $(\mathbb{R}^\infty, \mathcal{B})$ generated by $X_1, \dots, X_{\theta-1}, X_\theta, X_{\theta+1}, \dots$).

Another possible formulation is the following:

θ is a random variable and $\tau_{\alpha,h}^*$ is optimal if

$$\boxed{\inf_{\tau \in \mathfrak{M}_{(\alpha)}} \mathsf{P}\Big((\tau - \theta)^+ \ge h\Big) = \mathsf{P}\Big((\tau_{\alpha,h}^* - \theta)^+ \ge h\Big)}$$

where $\mathfrak{M}_{(\alpha)} = \{\tau\colon \mathsf{P}(\tau \le \theta) \le \alpha\}$.

By Chebyshev's equality $\mathsf{P}\Big((\tau-\theta)^+ \ge h\Big) \le \frac{1}{h}\mathsf{E}(\tau-\theta)^+$ and

$$\mathsf{P}\Big((\tau-\theta)^+ \ge h\Big) = \mathsf{P}(e^{k(\tau-\theta)^+} \ge e^{kh}) \le \frac{\mathsf{E}e^{k(\tau-\theta)^+}}{e^{kh}}.$$

So, $\mathsf{P}\Big((\tau-\theta)^+ \ge h\Big) \le \inf_{k>0} \dfrac{\mathsf{E}e^{k(\tau-\theta)^+}}{e^{kh}}$ and we have the problems:

$$\boxed{\inf_{\tau\in\mathfrak{M}_{(\alpha)}} \mathsf{E}(\tau-\theta)^+ = \mathsf{E}(\tau^*-\theta)^+, \qquad \inf_{\tau\in\mathfrak{M}_{(\alpha)}} \mathsf{E}e^{k(\tau-\theta)^+} = \mathsf{E}e^{k(\tau^*-\theta)^+}}.$$

It is interesting to solve the problems

$$\boxed{\inf \mathsf{E}|\tau-\theta|, \qquad \inf \mathsf{E}e^{k|\tau-\theta|}}$$

where inf is taken over the class $\mathfrak{M}$ of all stoping times τ.

Solutions of these problems will be discussed later.

Now only note that for all Bayesian problems we need to know the distributions of θ.

For the problem

$$\sup_{\tau} \mathsf{P}(|\tau-\theta| \geq h) = \mathsf{P}(|\tau_j^* - \theta| \leq h)$$

with $h = 0$, i.e., for the problem

$$\sup_{\tau} \mathsf{P}(\tau = \theta) = \mathsf{P}(\tau_0^* = \theta),$$

under the assumption that θ has the geometric distribution, the optimal time τ_0^* has the following simple structure:

$$\boxed{\tau_0^* = \inf\left\{n \geq 1 \colon \frac{f^0(x_n)}{f^\infty(x_n)} \in D_0^*\right\}}$$

- For Gaussian distributions $f^0(x) \sim \mathcal{N}(\mu_0, \sigma)$ and $f^\infty(x) \sim \mathcal{N}(\mu_\infty, \sigma)$ we find that

$$\boldsymbol{\tau_0^* = \inf\{n \geq 1 \colon x_n \in A_0^*\}.}$$

- If $f^0(x) = \frac{1}{2}\lambda_0 e^{-\lambda_0 |x|}$ and $f^\infty(x) = \frac{1}{2}\lambda_\infty e^{-\lambda_\infty |x|}$, then

$$\boldsymbol{\tau_0^* = \inf\{n \geq 1 \colon x_n \in B_0^*\}.}$$

Generally, if f^0 and f^∞ belong to the exponential family:

$$f^a(x) = \widetilde{c}_a g(x) \exp\{c_a \varphi_a(x)\}, \quad \text{where } \widetilde{c}_a,\ c_a,\ g(x) \geq 0,$$

then $\dfrac{f^0(x)}{f^\infty(x)} = \dfrac{\widetilde{c}_0}{\widetilde{c}_\infty} \exp\Big\{c_0\varphi_0(x) - c_\infty\varphi_\infty(x)\Big\}$, thus

$$\boldsymbol{\tau_0^* = \inf\{n \geq 1 \colon c_0\varphi_0(x_n) - c_\infty\varphi_\infty(x_n) \in C_0^*\}.}$$

3. E. Page's approach. * Below we will consider in details the CUSUM method initiated by E. Page. Now we give only definition of this procedure.

The **SHEWHART method** is based on the statistics $S_n = \frac{f^0(x_n)}{f^\infty(x_n)}$, which for Gaussian densities $f^0(x) \sim \mathcal{N}(\mu_0, 1)$ and $f^\infty(x) \sim \mathcal{N}(0,1)$ takes the form

$$\boxed{S_n = \exp\{\boldsymbol{\mu_0(x_n - \tfrac{\mu_0}{2})}\}} = \exp\{\Delta Z_n\}, \quad \text{where } Z_n = \sum_{k=1}^{n} \mu_0(x_k - \tfrac{\mu_0}{2}).$$

The **CUSUM method** is based on the statistics (see details in § 5)

$$\boxed{\gamma_n = \max_{1\le k\le n} \exp\Big\{\sum_{i=n-k+1}^{n} \boldsymbol{\mu_0(x_i - \tfrac{\mu_0}{2})}\Big\}} = \max_{1\le k\le n} \exp\{Z_n - Z_{n-k+1}\}.$$

* E.S.Page, "Continuous inspection schemes", Biometrika, 41 (1954), 100-114.
E.S.Page, "Control charts with warning lines", Biometrika, 42 (1955), 243-257.

The CUSUM stopping time is $\tau^* = \inf\{n \geq 1 \colon \gamma_n \geq d\}$.

It is important to emphasize that to construct the CUSUM statistics γ_n we must know the densities $f^\infty(x)$ and $f^0(x)$. Instead of $Z_n = \sum_{k=1}^n \mu_0(x_k - \frac{\mu_0}{2})$ we can use the following interesting statistics. Define

$$\widetilde{Z}_n = \sum_{k=1}^n |x_k|(x_k - \tfrac{|x_k|}{2}) \quad \text{and} \quad \widetilde{T}_n = \widetilde{Z}_n - \min_{1 \leq k \leq n} \widetilde{Z}_k.$$

Note that

$$|x_k|(x_k - \tfrac{|x_k|}{2}) = \begin{cases} x^2/2, & x \geq 0, \\ -3x^2/2, & x < 0. \end{cases}$$

So, for negative x_k, $k \leq n$, the statistics $\widetilde{T}$ is close to 0. But if x_k become positive, then the values $\widetilde{T}_n$ will increase. Thus, the statistics $\widetilde{T}_n$ help us to discover the appearing of the positive values x_k.

§ 2. Definition of the θ-model and Bayesian G-model in the quickest detection

1. We consider now the case of discrete time $n = 0, 1, \dots$ and assume that $(\Omega, \mathcal{F}, (\mathcal{F})_{n\ge 0}, \mathsf{P}^0, \mathsf{P}^\infty)$ is the binary statistical experiments. The measure P^∞ corresponds to the situation $\theta = \infty$, the measure P^0 corresponds to the situation $\theta = 0$.

Assume first that $\Omega = \mathbb{R}^\infty = \{x\colon (x_1, x_2, \dots), x_i \in \mathbb{R}\}$.

Let $f_n^0(x_1, \dots, x_n)$ and $f_n^\infty(x_1, \dots, x_n)$ be the densities of P_n^0 and P_n^∞ (w.r.t. the measure $\mathsf{P}_n = \frac{1}{2}(\mathsf{P}_n^0 + \mathsf{P}_n^\infty)$).

We denote by $f_n^0(x_n \,|\, x_1, \dots, x_{n-1})$ and $f_n^\infty(x_n \,|\, x_1, \dots, x_{n-1})$ the corresponding conditional densities.

How to define the conditional density $f_n^\theta(x_n \,|\, x_1, \dots, x_{n-1})$ for the value $0 < \theta < \infty$?

The meaning of the value θ as a change-point (disorder, disruption, in Russian "razladka") suggests that it is reasonable to define

$$f_n^\theta(x_n \,|\, x_1,\dots,x_{n-1}) = \begin{cases} f_n^{\infty}(x_n \,|\, x_1,\dots,x_{n-1}), & n<\theta, \\ f_n^{0}(x_n \,|\, x_1,\dots,x_{n-1}), & n\ge\theta, \end{cases}$$

or

$$\boxed{\begin{aligned} f_n^\theta(x_n \,|\, x_1,\dots,x_{n-1}) = I(n<\theta) f_n^\infty(x_n \,|\, x_1,\dots,x_{n-1}) \\ + I(n\ge\theta) f_n^0(x_n \,|\, x_1,\dots,x_{n-1}) \end{aligned}} \qquad (*)$$

Since it should be

$$f_n^\theta(x_1,\dots,x_{n-1}) = f_{n-1}^\theta(x_1,\dots,x_{n-1}) f_n^\theta(x_n \,|\, x_1,\dots,x_{n-1}),$$

we find that

$$\boxed{\begin{aligned} f_n^\theta(x_1,\dots,x_{n-1}) = I(n<\theta) f_n^\infty(x_1,\dots,x_{n-1}) \\ + I(n\ge\theta) f_{\theta-1}^\infty(x_1,\dots,x_{\theta-1}) \,\frac{f_n^0(x_1,\dots,x_{n-1})}{f_{\theta-1}^0(x_1,\dots,x_{\theta-1})} \end{aligned}} \qquad (**)$$

The formula $(**)$ can be taken as a definition of the density $f_n^\theta(x_1,\dots,x_{n-1})$.

It should be emphasized that, vice versa, from $(**)$ we obtain $(*)$.

Note that from $(*)$ we get the formula

$$\begin{aligned} f_n^\theta(x_n \,|\, x_1,\dots,x_{n-1}) - 1 = I(n<\theta)[f_n^\infty(x_n \,|\, x_1,\dots,x_{n-1}) - 1] \\ + I(n\ge\theta)[f_n^0(x_n \,|\, x_1,\dots,x_{\theta-1}) - 1] \end{aligned}$$

which explains the following general definition of the measures P_n^θ which is based on some martingale reasoning.

2. General stochastic θ-model. The previous considerations show how to define measures P^θ for case of general binary statistical experiment $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\ge 0}, \mathsf{P}^0, \mathsf{P}^\infty)$, $\mathcal{F}_0 = \{\varnothing, \Omega\}$.

Introduce the notation:

$$\mathsf{P} = \frac{1}{2}(\mathsf{P}^0 + \mathsf{P}^\infty), \qquad \mathsf{P}_n^0 = \mathsf{P}^0|\mathcal{F}_n, \qquad \mathsf{P}_n^\infty = \mathsf{P}^\infty|\mathcal{F}_n, \qquad \mathsf{P}_n = \frac{1}{2}(\mathsf{P}_n^0 + \mathsf{P}_n^\infty),$$

$$L^0 = \frac{d\mathsf{P}^0}{d\mathsf{P}}, \qquad L^\infty = \frac{d\mathsf{P}^\infty}{d\mathsf{P}}, \qquad L_n^0 = \frac{d\mathsf{P}_n^0}{d\mathsf{P}_n}, \qquad L_n^\infty = \frac{d\mathsf{P}_n^\infty}{d\mathsf{P}_n}$$

($\dfrac{d\mathsf{Q}}{d\mathsf{P}}$ is the Radon–Nikodým derivative).

Since for $A \in \mathcal{F}_n$

$$\int_A \mathsf{E}(L^0 \,|\, \mathcal{F}_n)\, d\mathsf{P} = \int_A L^0\, d\mathsf{P} = \mathsf{P}^0(A) = \mathsf{P}^0_n(A) = \int_A \frac{d\mathsf{P}^0_n}{d\mathsf{P}_n}\, d\mathsf{P}_n = \int_A L^0_n\, d\mathsf{P},$$

we have the martingale property

$$L^0_n = \mathsf{E}(L^0 \,|\, \mathcal{F}_n) \quad \text{and similarly} \quad L^\infty_n = \mathsf{E}(L^\infty \,|\, \mathcal{F}_n).$$

Note that $\mathsf{P}(L^0_n = 0) = \mathsf{P}(L^\infty_n = 0) = 0$.

Associate with the martingales $L^0 = (L^0_n)_{n\ge 0}$ and $L^\infty = (L^\infty_n)_{n\ge 0}$ their stochastic logarithms

$$M^0_n = \sum_{k=1}^{n} \frac{\Delta L^0_k}{L^0_{k-1}} I(L^0_{k-1} > 0), \qquad M^\infty_n = \sum_{k=1}^{n} \frac{\Delta L^\infty_k}{L^\infty_{k-1}} I(L^\infty_{k-1} > 0),$$

where $\Delta L^0_k = L^0_k - L^0_{k-1}$ and $\Delta L^\infty_k = L^\infty_k - L^\infty_{k-1}$.

The processes $(M_n^0, \mathcal{F}_n, \mathsf{P})_{n\ge 0}$, $(M_n^\infty, \mathcal{F}_n, \mathsf{P})_{n\ge 0}$ are P-local martingales and

$$\Delta L_n^0 = L_{n-1}^0 \Delta M_n^0, \qquad \Delta L_n^\infty = L_{n-1}^\infty \Delta M_n^\infty.$$

In case of the coordinate space $\Omega = \mathbb{R}^\infty$, we find that (P-a.s.)

$$\begin{aligned} \Delta M_n^0 &= \frac{\Delta L_n^0}{L_{n-1}^0} = \frac{L_n^0}{L_{n-1}^0} - 1 = \frac{f_n^0(x_1,\dots,x_n)}{f_{n-1}^0(x_1,\dots,x_n)} - 1 \\ &= f_n^0(x_n \,|\, x_1,\dots,x_{n-1}) - 1 \end{aligned}$$

and similarly

$$\Delta M_n^\infty = f_n^\infty(x_n \,|\, x_1,\dots,x_{n-1}) - 1. \qquad (\bullet)$$

Above we defined $f_n^\theta(x_n \,|\, x_1, \dots, x_{n-1})$ as

$$f_n^\theta(x_n \,|\, x_1, \dots, x_{n-1}) = I(n < \theta) f_n^\infty(x_n \,|\, x_1, \dots, x_{n-1}) \\ + I(n \geq \theta) f_n^0(x_n \,|\, x_1, \dots, x_{n-1}).$$

Thus, if we take into account (•), then for general case it is reasonable to define ΔM_n^θ as

$$\Delta M_n^\theta = I(n < \theta)\Delta M_n^\infty + I(n \geq \theta)\Delta M_n^0. \qquad (\bullet\bullet)$$

We have

$$L_n^0 = \mathcal{E}(M^0)_n, \quad L_n^\infty = \mathcal{E}(M^\infty)_n,$$

where $\mathcal{E}$ is the stochastic exponential:

$$\mathcal{E}(M)_n = \prod_{k=1}^{n} (1 + \Delta M_k) \qquad (\bullet\bullet\bullet)$$

with $\Delta M_k = M_k - M_{k-1}$. Thus it is reasonable to define L_n^θ by

$$L_n^\theta = \mathcal{E}(M^\theta)_n.$$

From formulas (••) and (• • •) it follows that

$$\mathcal{E}(M^\theta)_n = \mathcal{E}(M^\infty)_n, \qquad n < \theta,$$

$$\mathcal{E}(M^\theta)_n = \mathcal{E}(M^\infty)_{\theta-1} \frac{\mathcal{E}(M^0)_n}{\mathcal{E}(M^0)_{\theta-1}}, \qquad 1 \le \theta \le n.$$

So, for L_n^θ we find that (P-a.s.)

$$L_n^\theta = I(n < \theta) L_n^\infty + I(n \ge \theta) L_{\theta-1}^\infty \cdot \frac{L_n^0}{L_{\theta-1}^0},$$

$$\text{or} \quad L_n^\theta = I(n < \theta) L_n^\infty + I(n \ge \theta) L_n^0 \cdot \frac{L_{\theta-1}^\infty}{L_{\theta-1}^0},$$

$$\text{or} \quad L_n^\theta = L_{(\theta-1)\wedge n}^\infty \cdot \frac{L_n^0}{L_{(\theta-1)\wedge n}^0}.$$

So, we have
$$\boxed{L_n^\theta = \begin{cases} L_n^\infty, & \theta > n \\ L_{\theta-1}^\infty \cdot \dfrac{L_n^0}{L_{\theta-1}^0}, & \theta \le n \end{cases}}$$

Define now for $A \in \mathcal{F}_n$

$$\mathsf{P}_n^\theta(A) = \mathsf{E}[I(A)\mathcal{E}(M^\theta)_n], \quad \text{or} \quad \mathsf{P}_n^\theta(A) = \mathsf{E}[I(A)L_n^\theta].$$

The family of measures $\{\mathsf{P}_n^\theta\}_{n\geq 1}$ is consistent and we can expect that there exists a measure P^θ on $\mathcal{F}_\infty = \bigvee \mathcal{F}_n$ such that

$$\mathsf{P}^\theta | \mathcal{F}_n = \mathsf{P}_n^\theta.$$

Without special assumptions on L_n^θ, $n \geq 1$, we cannot guarantee existence of such a measure*. It will be so, if, for example, the martingale $(L_n^\theta)_{n\geq 0}$ is uniformly integrable. In this case there exists an $\mathcal{F}_\infty$-measurable random variable L^θ such that

$$L_n^\theta = \mathsf{E}(L^\theta \,|\, \mathcal{F}_n) \quad \text{and} \quad \mathsf{P}^\theta(A) = \mathsf{E}[I(A)L^\theta].$$

---

* See, e.g., the corresponding example in: `A.N.Shiryaev, Probability, Chapter II, §3`.

Another way to construct the measure P^θ is based on the famous Kolmogorov theorem on the extension of measures on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$. This theorem states that

if $\mathsf{P}_1^\theta, \mathsf{P}_2^\theta, \dots$ is a sequence of probability measures on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ which have the consistency property

$$\mathsf{P}_{n+1}^\theta(B \times \mathbb{R}) = \mathsf{P}_n(B), \qquad B \in \mathcal{B}(\mathbb{R}_n),$$

then there is a unique probability measure P^θ on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ such that

$$\mathsf{P}^\theta(J_n(B)) = \mathsf{P}_n(B)$$

for $B \in \mathcal{B}(\mathbb{R}_n)$, where $J_n(B)$ is the cylinder in $\mathbb{R}^\infty$ with base $B \in \mathcal{B}(\mathbb{R}_n)$.

Note that, for the case of continuous time, the measures P^θ based on the measures P^θ_t, $t \ge 0$, can be constructed in a similar way.

The measures P^θ constructed for all $0 \le \theta \le \infty$ from the measures P^0 and P^∞ have the following characteristic properties:

$$\boxed{\mathsf{P}^\theta(A) = \mathsf{P}^\infty(A), \quad \text{if} \;\; n < \theta}.$$

The constructed filtered statistical experiment

$$(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\ge 0}; \mathsf{P}^\theta, 0 \le \theta \le \infty)$$

will be called a ***θ*-model** constructed via measures P^0 (“change-point”, “disorder” time θ equals 0) and P^∞ (“change-point”, “disorder” time θ equals ∞).

3. General stochastic G-models. Let

$$(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}; \mathsf{P}^\theta, 0 \leq \theta \leq \infty)$$

be the θ-model. Now we shall consider θ as a random variable (given on some probability space $(\Omega', \mathcal{F}', \mathsf{P}'))$ with the distribution function $G = G(h)$, $h \geq 0$. Define

$$\overline{\Omega} = \Omega \times \Omega', \qquad \overline{\mathcal{F}} = \mathcal{F}_\infty \otimes \mathcal{F}'$$

and put for $A \in \mathcal{F}_\infty$ and $B' \in \mathcal{F}'$

$$\mathsf{P}^G(A \times B') = \sum_{\theta \in B'} \mathsf{P}^\theta(A) \Delta G(\theta),$$

where $\Delta G(\theta) = G(\theta) - G(\theta - 1)$, $\Delta G(0) = G(0)$.

The extension of this function of sets $A \times B'$ onto $\overline{\mathcal{F}} = \mathcal{F}_\infty \otimes \mathcal{F}'$ will be denoted by P^G.

It is clear that for $\mathsf{P}^G(A) = \mathsf{P}^G(A \times N')$ with $A \in \mathcal{F}_n$, where $N' = \{0, 1, \dots, \infty\}$, we get

$$\mathsf{P}^G(A) = \sum_{\theta=0}^{n} \mathsf{P}_n^\theta(A) \Delta G(\theta) + (1 - G(n)) \mathsf{P}_n^\infty(A),$$

where we have used that $\mathsf{P}_n^\theta(A) = \mathsf{P}_n^\infty(A)$ for $A \in \mathcal{F}_n$ and $\theta > n$. Denote

$$\mathsf{P}_n^G = \mathsf{P}^G | \mathcal{F}_n \quad \text{and} \quad L_n^G = \frac{d\mathsf{P}_n^G}{d\mathsf{P}_n}.$$

Then we see that

$$L_n^G = \sum_{\theta=0}^{\infty} L_n^\theta \Delta G(\theta).$$

Taking into account that

$$L_n^\theta = \begin{cases} L_n^\infty, & \theta > n \\ L_{\theta-1}^\infty \cdot \dfrac{L_n^0}{L_{\theta-1}^0}, & \theta \le n, \end{cases} \qquad \text{with} \quad L_{-1}^0 = L_{-1}^\infty = 1,$$

we find the following representation:

$$L_n^G = \sum_{\theta=0}^{n} L_{\theta-1}^\infty \frac{L_n^0}{L_{\theta-1}^0} \, \Delta G(\theta) + L_n^\infty (1 - G(n)),$$

where $L_{-1}^\infty = L_{-1}^0 = 1$.

EXAMPLE. Geometrical distribution:

$$G(0) = \pi, \qquad \Delta G(n) = (1-\pi)q^{n-1}p, \quad n \geq 1.$$

Here

$$L_n^G = \pi L_n^0 + (1-\pi)L_n^0 \sum_{k=0}^{n-1} pq^k \frac{L_k^\infty}{L_k^0} + (1-\pi)q^n L_n^\infty.$$

If $f_n^0 = f_n^0(x_1, \dots, x_n)$ and $f_n^\infty = f_n^\infty(x_1, \dots, x_n)$ are densities of P_n^0 and P_n^∞ w.r.t. the Lebesgue measure, then we find

$$\begin{aligned} f_n^G(x_1, \dots, x_n) &= \pi f_n^0(x_1, \dots, x_n) \\ &\quad + (1-\pi) f_n^0(x_1, \dots, x_n) \sum_{k=0}^{n-1} pq^k \frac{f_k^\infty(x_1, \dots, x_k)}{f_k^0(x_1, \dots, x_k)} \\ &\quad + (1-\pi) q^n f_n^\infty(x_1, \dots, x_n). \end{aligned}$$

Thus

$$
\begin{aligned}
f_n^G(x_1, \ldots, x_n) &= \pi f_n^0(x_1, \ldots, x_n) \\
&\quad + (1-\pi) \sum_{k=0}^{n-1} p q^k f_k^\infty(x_1, \ldots, x_k) f_{n,k}^0(x_{k+1}, \ldots, x_n \mid x_1, \ldots, x_k) \\
&\quad + (1-\pi) q^n f_n^\infty(x_1, \ldots, x_n),
\end{aligned}
$$

where

$$
f_{n,k}^0(x_{k+1}, \ldots, x_n \mid x_1, \ldots, x_k) = \frac{f_n^0(x_1, \ldots, x_n)}{f_k^0(x_1, \ldots, x_k)}.
$$

§ 3. Four basic formulations (VARIANTS A, B, C, D) of the quickest detection problems for the Brownian case

1. **VARIANT A**. We assume that G-model is given and $\mathfrak{M}_\alpha = \{\tau \colon \mathsf{P}^G(\tau < \theta) \le \alpha\}$, where $\alpha \in (0,1)$, $\mathfrak{M}$ is the class of all finite stoping times.

- **Conditionally extremal formulation:**
  To find an optimal stopping time $\tau^*_\alpha \in \mathfrak{M}_\alpha$ for which

$$\boxed{\mathsf{E}^G(\tau^*_\alpha - \theta)^+ = \inf_{\tau \in \mathfrak{M}_\alpha} \mathsf{E}^G(\tau - \theta)^+}\,.$$

- **Bayesian formulation:** To find an optimal stopping time $\tau^*_{(c)} \in \mathfrak{M}$ for which

$$\boxed{\mathsf{P}^G(\tau^*_{(c)} < \theta) + c\,\mathsf{E}^G(\tau^*_{(c)} - \theta)^+ = \inf_{\tau \in \mathfrak{M}} \Big[\mathsf{P}(\tau < \theta) + c\,\mathsf{E}(\tau - \theta)^+\Big]}\,.$$

2. **VARIANT B** **(Generalized Bayesian formulation)**. We assume that θ-model is given an:

$\mathfrak{M}_T = \{\tau \in \mathfrak{M} \colon \mathsf{E}^\infty \tau \ge T\}$ [the class of stopping times τ for which the mean time $\mathsf{E}_\infty \tau$ of τ, under assumption that there was no change point (disorder) at all, equals a given a priori constant $T > 0$].

The problem is to find the value

$$\boxed{\mathbb{B}(T) = \inf_{\tau \in \mathfrak{M}_T} \sum_{\theta \ge 1} \mathsf{E}^\theta (\tau - \theta)^+}$$

and the optimal stopping time τ_T^* for which

$$\boxed{\sum_{\theta \ge 1} \mathsf{E}^\theta (\tau - \theta)^+ = \mathbb{B}(T)}\,.$$

3. **VARIANT C** **(the first minimax formulation)**.

The problem is to find the value

$$\boxed{\mathbb{C}(T) = \inf_{\tau \in \mathfrak{M}_T} \sup_{\theta \geq 1} \mathsf{E}^{\theta}(\tau - \theta \,|\, \tau \geq \theta)}$$

and the optimal stopping time $\widetilde{\tau}_T$ for which

$$\boxed{\sup_{\theta \geq 1} \mathsf{E}^{\theta}(\widetilde{\tau}_T - \theta \,|\, \widetilde{\tau}_T \geq \theta) = \mathbb{C}(T)}$$

VARIANT D (the second minimax formulation).

The problem is to find the value

$$\boxed{\mathbb{D}(T) = \inf_{\tau \in \mathfrak{M}_T} \sup_{\theta \geq 1} \operatorname*{ess\,sup}_{\omega} \mathsf{E}^{\theta}((\tau - \theta)^{+} \mid \mathcal{F}_{\theta-1})}$$

and the optimal stopping time $\overline{\tau}_T$ for which

$$\boxed{\sup_{\theta \geq 1} \operatorname*{ess\,sup}_{\omega} \mathsf{E}^{\theta}((\overline{\tau}_T - \theta)^{+} \mid \mathcal{F}_{\theta-1})(\omega) = \mathbb{D}(T)}$$

[Essential supremum w.r.t. the measure P of the nonnegative function $f(\omega)$ (notation: $\operatorname{ess\,sup} f$, or $\|f\|_\infty$, or $\operatorname{vraisup} f$) is defined as follows:

$$\operatorname*{ess\,sup}_{\omega} f(\omega) = \inf\{0 \leq c \leq \infty : \mathsf{P}(|f| > c) = 0\}.$$
]

4. There are many works, where, instead of the described penalty functions, the following functions are investigated:

$$W(\theta,\tau) = \begin{cases} W_1(\tau), & \tau < \theta, \\ W_2(\tau-\theta), & \tau \ge \theta, \end{cases},$$

$$W(\theta,\tau) = W_1((\tau-\theta)^+) + W_2((\tau-\theta)^+),$$

in particular, $W(\theta,\tau) = \mathsf{E}|\tau-\theta|$

$$W(\theta,\tau) = \mathsf{P}(|\tau-\theta| \ge h), \quad \text{etc.}$$

§ 4. The reduction of VARIANTS A and B to the standard form

1. Denote

$$\mathbb{A}_1(c) = \inf_{\tau \in \mathfrak{M}} \Big[\mathsf{P}^G(\tau < \theta) + c\,\mathsf{E}^G(\tau - \theta)^+ \Big],$$

$$\mathbb{A}_2(c) = \inf_{\tau \in \mathfrak{M}} \Big[\mathsf{P}^G(\tau < \theta) + c\,\mathsf{E}^G(\tau - \theta + 1)^+ \Big].$$

THEOREM 1. Let $\pi_n = \mathsf{P}^G(\theta \le n \,|\, \mathcal{F}_n)$. Then

$$\mathbb{A}_1(c) = \inf_{\tau \in \mathfrak{M}} \mathsf{E}^G \Big[(1 - \pi_\tau) + c \sum_{k=0}^{\tau-1} \pi_k \Big],$$

$$\mathbb{A}_2(c) = \inf_{\tau \in \mathfrak{M}} \mathsf{E}^G \Big[(1 - \pi_\tau) + c \sum_{k=0}^{\tau} \pi_k \Big].$$

PROOF follows from the formulae $\mathsf{P}^G(\tau < \theta) = \mathsf{E}^G(1-\pi_\tau)$ and

$$\underbrace{(\tau-\theta)^+ = \sum_{k=0}^{\tau-1} I(\theta \le k),}_{} \qquad (\tau-\theta+1)^+ = \sum_{k=0}^{\tau} I(\theta \le k). \qquad (\bullet)$$

This follows from the property $\xi^+ = \sum_{k\ge 1} I(\xi \ge k)$:

$$(\tau-\theta)^+ = \sum_{k\ge 1} I(\tau-\theta \ge k) = \sum_{k\ge 1} I(\theta \le \tau - k) = \sum_{l=0}^{\tau-1} I(\theta \le l)$$

$$\begin{aligned}
\mathsf{E}^G(\tau-\theta)^+ &= \mathsf{E}^G \sum_{k=0}^{\tau-1} I(\theta \le k) \\
&= \mathsf{E}^G \sum_{k=0}^{\infty} I(k \le \tau-1) I(\theta \le k) \\
&= \mathsf{E}^G \sum_{k=0}^{\infty} \mathsf{E}^G\big[I(k \le \tau-1) I(\theta \le k) \,|\, \mathcal{F}_k\big] \\
&= \mathsf{E}^G \sum_{k=0}^{\infty} \mathsf{E}^G\big[I(\tau \ge k+1) I(\theta \le k) \,|\, \mathcal{F}_k\big] \\
&= \mathsf{E}^G \sum_{k=0}^{\infty} I(k \le \tau-1) \mathsf{E}^G\big[I(\theta \le k) \,|\, \mathcal{F}_k\big] = \mathsf{E}^G \sum_{k=0}^{\tau-1} \pi_k.
\end{aligned}$$

Representations $(\bullet)$ imply

$$\mathsf{E}^G(\tau-\theta+1)^+ = \mathsf{E}^G \sum_{k=0}^{\tau} \pi_k, \qquad \overbrace{\mathsf{E}^G(\tau-\theta)^+ = \mathsf{E}^G \sum_{k=0}^{\tau-1} \pi_k}$$

THEOREM 2. Define

$$\mathbb{A}_3 = \inf_\tau \mathsf{E}^G|\tau - \theta|, \qquad \mathbb{A}_4 = \inf_\tau \mathsf{E}^G|\tau - \theta + 1|.$$

Then

$$\mathbb{A}_3 = \inf_\tau \mathsf{E}^G\Big[\theta + \sum_{k=0}^{\tau-1}(2\pi_k - 1)\Big], \quad \mathbb{A}_4 = \inf_\tau \mathsf{E}^G\Big[\theta + \sum_{k=0}^{\tau}(2\pi_k - 1)\Big].$$

PROOF is based on the formulae

$$|\tau - \theta| = \theta + \sum_{k=0}^{\tau-1}(2I(\theta \le k) - 1),$$
$$|\tau - \theta + 1| = \theta + \sum_{k=0}^{\tau}(2I(\theta \le k) - 1).$$

THEOREM 3. Let $G = G(n)$, $n \geq 0$, be the geometrical distribution:

$$\Delta G(n) = pq^{n-1}, \quad n \geq, \qquad G(0) = 0.$$

Then for $\mathbb{A}_3 = \mathbb{A}_3(p)$ we have

$$\mathbb{A}_3(p) = \tfrac{1}{p}\mathbb{A}_1(p),$$

i.e.,

$$\inf_\tau \mathsf{E}^G|\tau - \theta| = \frac{1}{p} \inf_\tau \Big[\mathsf{P}^G(\tau < \theta) + p\mathsf{E}^G(\tau - \theta)^+\Big].$$

2. Consider now the criterion

$$\mathbb{A}(W) = \inf_{\tau} \mathsf{E}^G W(\theta,\tau) \quad \text{with} \quad W(\theta,\tau) = \begin{cases} W_1(\tau), & \tau < \theta, \\ W_2((\tau-\theta)^+), & \tau \ge \theta, \end{cases}$$

where $\quad W_2(n) = \sum_{k=1}^{n} f(k), \quad W_2(0) = 0. \quad$ Then

$$\mathsf{E}^G W(\theta,\tau) = \mathsf{E}^G \Bigg[(1-\pi_\tau) \Bigg(W_1(\tau) + \frac{L_\tau}{1-G(\tau)} \sum_{k=0}^{\tau-1} W_2(\tau-k) \, \frac{\Delta G(k)}{L_{k-1}} \Bigg) \Bigg].$$

For example, for $W_1(n) \equiv 1$, $W_2(n) = cn^2$ we get

$$\mathsf{E}^G W(\theta,\tau) = \mathsf{E}^G \Bigg[(1-\pi_\tau) \Bigg(1 + \sum_{k=0}^{\tau-1} (\tau-k)^2 \, \frac{L_\tau}{L_{k-1}} \Bigg) \Bigg].$$

3. In Variant B: $\mathbb{B}(T) = \inf\limits_{\tau \in \mathfrak{M}_T} \sum\limits_{\theta \geq 1} \mathsf{E}^{\theta}(\tau - \theta)^{+}$.

THEOREM 4. For any finite stopping time τ we have

$$\sum_{\theta=1}^{\infty} \mathsf{E}^{\theta}(\tau - \theta)^{+} = \mathsf{E}^{\infty} \sum_{n=1}^{\tau-1} \psi_n, \qquad (*)$$

where $\psi_n = \sum\limits_{\theta=1}^{n} \dfrac{L_n}{L_{\theta-1}}$, $L_n = \dfrac{L_n^0}{L_n^\infty}$, $L_n^0 = \dfrac{d\mathsf{P}_n^0}{d\mathsf{P}_n}$, $L_n^\infty = \dfrac{d\mathsf{P}_n^\infty}{d\mathsf{P}_n}$.

Therefore,

$$\mathbb{B}_T = \inf_{\tau \in \mathfrak{M}_T} \mathsf{E}^{\infty} \sum_{k=1}^{\tau-1} \psi_n, \quad \text{where} \quad \mathfrak{M}_T = \{\tau \in \mathfrak{M} \colon \mathsf{E}^{\infty}\tau \geq T\}.$$

Some generalization:

$$\mathbb{B}_F(T) = \inf_{\tau \in \mathfrak{M}_T} \sum_{\theta \geq 1} \mathsf{E}^{\theta} F((\tau - \theta)^{+}), \quad \text{where} \quad \begin{aligned} & F(n) = \textstyle\sum_{k=1}^{n} f(k), \\ & F(0) = 0, \\ & f(k) \geq 0 \end{aligned}$$

PROOF of $(*)$. Since $(\tau-\theta)^+ = \sum_{k=1}^{\infty} I(\tau-\theta \ge k) = \sum_{k\ge\theta+1} I(\tau \ge k)$ and $\mathsf{P}^k(A) = \mathsf{P}^\infty(A)$ for $A \equiv \{\tau \ge k\} \in \mathcal{F}_{k-1}$, we find that

$$
\begin{aligned}
\mathsf{E}^\theta(\tau-\theta)^+ &= \sum_{k\ge\theta+1} \mathsf{E}^\theta I(\tau\ge k) \\
&= \sum_{k\ge\theta+1} \mathsf{E}^k\left[I(\tau\ge k)\,\frac{d(\mathsf{P}^\theta|\mathcal{F}_{k-1})}{d(\mathsf{P}^k|\mathcal{F}_{k-1})}\right] \\
&= \sum_{k\ge\theta+1} \mathsf{E}^k\left[I(\tau\ge k)\,\frac{L^\theta_{k-1}}{L^\infty_{k-1}}\right] \\
&= \sum_{k\ge\theta+1} \mathsf{E}^k\left[I(\tau\ge k)\,\frac{L^\theta_{k-1}}{L^\infty_{k-1}}\right] \\
&= \sum_{k\ge\theta+1} \mathsf{E}^k\left[I(\tau\ge k)\,\frac{L^0_{k-1}L^\infty_{\theta-1}}{L^0_{\theta-1}L^\infty_{k-1}}\right] \\
&= \sum_{k\ge\theta+1} \mathsf{E}^k\left[I(\tau\ge k)\,\frac{L_{k-1}}{L_{\theta-1}}\right].
\end{aligned}
$$

Therefore,
$$\sum_{\theta=1}^{\infty} \mathsf{E}^{\theta}(\tau-\theta)^{+} = \mathsf{E}^{\infty} \sum_{\theta=1}^{\infty} \left[\sum_{k=\theta+1}^{\infty} I(\tau \geq k)\, \frac{L_{k-1}}{L_{\theta-1}} \right]$$
$$= \mathsf{E}^{\infty} \sum_{\theta=1}^{\infty} \sum_{k=2}^{\infty} I(\theta+1 \leq k \leq \tau)\, \frac{L_{k-1}}{L_{\theta-1}}$$
$$= \mathsf{E}^{\infty} \sum_{k=2}^{\infty} \left[\sum_{\theta=1}^{k-1} \frac{L_{k-1}}{L_{\theta-1}} \right]$$
$$= \mathsf{E}^{\infty} \sum_{k=2}^{\tau} \psi_k = \mathsf{E}^{\infty} \sum_{k=1}^{\tau-1} \psi_k.$$

We find here that

$$\mathbb{B}_F(T) = \inf_{\tau \in \mathfrak{M}_T} \mathsf{E}^{\infty} \sum_{n=0}^{\tau-1} \Psi_n(f),$$

where

$$\Psi_n(f) = \sum_{\theta=0}^{n} f(n+1-\theta)\, \frac{L_n}{L_{\theta-1}}.$$

§ 5. VARIANT C and D: lower estimates for the risk functions

1. In Variant C the risk function is

$$\mathbb{C}(T) = \inf_{\tau \in \mathfrak{M}_T} \sup_{\theta \geq 1} \mathsf{E}^{\theta}(\tau - \theta \,|\, \tau \geq \theta).$$

THEOREM 5. For any stopping time τ with $\mathsf{E}^{\infty}\tau < \infty$,

$$\boxed{\sup_{\theta \geq 1} \mathsf{E}^{\theta}(\tau - \theta \,|\, \tau \geq \theta) \geq \frac{1}{\mathsf{E}^{\infty}\tau} \mathsf{E}^{\infty} \sum_{n=1}^{\tau-1} \psi_n} \qquad \text{where} \quad \psi_n = \sum_{\theta=1}^{n} \frac{L_n}{L_{\theta-1}}.$$

Thus, in the class $\widehat{\mathfrak{M}}_T = \{\tau \in \mathfrak{M}_T \colon \mathsf{E}^{\infty}\tau = T\}$,

$$\widehat{\mathbb{C}}(T) = \inf_{\tau \in \widehat{\mathfrak{M}}_T} \sup_{\theta \geq 1} \mathsf{E}^{\theta}(\tau - \theta \,|\, \tau \geq \theta) \geq \frac{1}{T} \widehat{\mathbb{B}}(T),$$

where $\widehat{\mathbb{B}}(T) = \inf_{\tau \in \widehat{\mathfrak{M}}_T} \sum_{\theta \geq 1} \mathsf{E}^{\theta}(\tau - \theta)^{+} = \inf_{\tau \in \widehat{\mathfrak{M}}_T} \mathsf{E}^{\infty} \sum_{k=1}^{\tau-1} \psi_k.$

PROOF. We have

$$\sum_{\theta\ge 1} \mathsf{E}^\theta(\tau-\theta)^+ = \sum_{\theta\ge 1} \mathsf{E}^\theta(\tau-\theta \,|\, \tau\ge\theta)\mathsf{P}^\theta(\tau\ge\theta)$$

$$\le \sup_{\theta\ge 1} \mathsf{E}^\theta(\tau-\theta \,|\, \tau\ge\theta) \sum_{\theta\ge 1} \underbrace{\mathsf{P}^{\boldsymbol{\theta}}(\tau\ge\theta)}$$

by definition of the θ-model,
since $\{\tau\ge\theta\}\in\mathcal{F}_{\theta-1}$ $\quad \Big\|$

$$= \sup_{\theta\ge 1} \mathsf{E}^\theta(\tau-\theta \,|\, \tau\ge\theta) \underbrace{\sum_{\theta\ge 1} \overbrace{\mathsf{P}^{\infty}(\tau\ge\theta)}}_{=\mathsf{E}^\infty\tau}.$$

Thus,

$$\sup_{\theta\ge 1} \mathsf{E}^\theta(\tau-\theta \,|\, \tau\ge\theta) \ge \frac{1}{\mathsf{E}^\infty\tau}\sum_{\theta\ge 1}\mathsf{E}^\theta(\tau-\theta)^+ = \frac{1}{\mathsf{E}^\infty\tau}\mathsf{E}^\infty\sum_{n=1}^{\tau-1}\psi_n.$$

2. Now we consider Variant D, where

$$\mathbb{D}(T) = \inf_{\tau \in \mathfrak{M}_T} \sup_{\theta \ge 1} \operatorname*{ess\,sup}_{\omega} \mathsf{E}^{\theta}\Big[(\tau - (\theta - 1))^{+} \,|\, \mathcal{F}_{\theta-1}\Big].$$

THEOREM 6. For any stopping time τ with $\mathsf{E}^{\infty}\tau < \infty$,

$$\sup_{\theta \ge 1} \operatorname*{ess\,sup}_{\omega} \mathsf{E}^{\theta}\Big[(\tau - (\theta - 1))^{+} \,|\, \mathcal{F}_{\theta-1}\Big] \ge \frac{\mathsf{E}^{\infty} \sum_{k=0}^{\tau-1} \gamma_k}{\mathsf{E}^{\infty} \sum_{k=0}^{\tau-1} (1 - \gamma_k)^{+}}, \qquad (*)$$

where $(\gamma_k)_{k \ge 0}$ is the CUSUM-statistics:

$$\gamma_k = \max_{1 \le \theta \le n} \frac{L_n^{\theta}}{L_n^{\infty}} \qquad \Big(L_n^{\theta} = \frac{d\mathsf{P}_n^{\theta}}{d\mathsf{P}_n}, \quad L_n^{\infty} = \frac{d\mathsf{P}_n^{\infty}}{d\mathsf{P}_n}\Big).$$

Thus,

$$\mathbb{D}(T) \ge \inf_{\tau \in \mathfrak{M}_T} \frac{\mathsf{E}^{\infty} \sum_{k=0}^{\tau-1} \gamma_k}{\mathsf{E}^{\infty} \sum_{k=0}^{\tau-1} (1 - \gamma_k)^{+}}.$$

Recall, first of all, some useful facts about CUSUM-statistics $(\gamma_n)_{n\ge 0}$:

$$\gamma_n = \max_{1\le\theta\le n} \frac{L_n}{L_{\theta-1}} = \max_{1\le k\le n} \frac{L_n}{L_{n-k}};$$

$$\gamma_n = \frac{L_n}{\boldsymbol{L_{n-1}}} \underbrace{\max(1, \gamma_{n-1})}_{\substack{= \sum_{\theta=1}^{n} (1-\gamma_{\theta-1})^+ \frac{\boldsymbol{L_{n-1}}}{L_{\theta-1}} \\ \text{by induction}}} = \sum_{\theta=1}^{n} (1-\gamma_{\theta-1})^+ \frac{L_n}{L_{\theta-1}} \quad \Rightarrow$$

$$\Rightarrow \quad \boldsymbol{\gamma_n = \frac{L_n}{L_{n-1}}\big[(1-\gamma_{n-1})^+ + \gamma_{n-1}\big]} = \frac{L_n}{L_{n-1}} \begin{cases} 1, & \gamma_{n-1} < 1, \\ \gamma_{n-1}, & \gamma_{n-1} \ge 1 \end{cases}$$

(cf. $\boldsymbol{\psi_n} = \sum_{\theta=1}^{n} \frac{L_n}{L_{n-1}} \boldsymbol{= \frac{L_n}{L_{n-1}}[1+\psi_{n-1}]}$).

PROOF of the basic inequality $(*)$:

$$\sup_{\theta\ge 1} \operatorname*{ess\,sup}_{\omega} \mathsf{E}^{\theta}\Big[(\tau-(\theta-1))^{+} \,|\, \mathcal{F}_{\theta-1}\Big] \ge \frac{\mathsf{E}^{\infty}\sum_{k=0}^{\tau-1}\gamma_k}{\mathsf{E}^{\infty}\sum_{k=0}^{\tau-1}(1-\gamma_k)^{+}}. \qquad (*)$$

For $\tau \in \mathfrak{M}_T$

$$d^{\theta}(\tau) \stackrel{\text{def}}{=} \mathsf{E}^{\theta}\Big[\ \underbrace{(\tau-(\theta-1))^{+}}\,|\,\mathcal{F}_{\theta-1}\Big](\omega)$$

$$= \sum_{k\ge 1} I(\tau-(\theta-1)\ge k) = \sum_{k\ge 1} I(\tau \ge k+(\theta-1)) = \sum_{k\ge\theta} I(\tau\ge k),$$

$$= \sum_{k\ge\theta}\mathsf{E}^{\theta}[I(\tau\ge k)\,|\,\mathcal{F}_{\theta-1}] = \sum_{k\ge\theta}\mathsf{E}^{\theta}\Big[I(\tau\ge k)\frac{L_{k-1}}{\theta-1}\,\Big|\,\mathcal{F}_{\theta-1}\Big] \quad (\mathsf{P}^{\infty}\text{-a.s.}).$$

[Here we used the fact that if r.v. ξ is ≥ 0 and $\mathcal{F}_{k-1}$-measurable, then

$$\mathsf{E}^{\theta}(\xi\,|\,\mathcal{F}_{\theta-1}) = \mathsf{E}^{\infty}\Big[\xi\frac{L_{k-1}}{L_{\theta-1}}\,\Big|\,\mathcal{F}_{\theta-1}\Big], \qquad L_n = \frac{L_n^0}{L_n^{\infty}}.$$
]

Denote $d(\tau) = \sup_{\theta\ge 1} \operatorname{ess\,sup}_\omega d^\theta(\tau)$. For each τ and each $\theta \ge 1$

$$d(\tau) \ge d^\theta(\tau) \quad (\mathsf{P}^\theta\text{-a.s.})$$

and for any nonnegative $\mathcal{F}_\theta$-measurable function $f = f_\theta(\omega)$ (all r.v.'s here are $\mathcal{F}_\theta$-measurable)

$$f_{\theta-1} I(\tau \ge \theta) d(\tau) \ge f_{\theta-1} I(\tau \ge \theta) d^\theta(\tau) \quad (\mathsf{P}^\theta\text{-a.s. } \underbrace{\text{and } \mathsf{P}^\infty\text{-a.s.}}_{\substack{\text{since } \mathsf{P}^\theta(A) = \mathsf{P}^\infty(A) \\ \text{for every } A \in \mathcal{F}_{\theta-1} \\ \text{by definition of the } \theta\text{-model}}}\ ) \quad (**)$$

Taking into account $(**)$, we get

$$\begin{aligned}
\mathsf{E}^\infty\Big\{f_{\theta-1} I(\tau\ge\theta)\Big\} d(\tau) &= \mathsf{E}^\infty\Big\{f_{\theta-1} I(\tau\ge\theta) d(\tau)\Big\} \ge \mathsf{E}^\infty\Big\{f_{\theta-1} I(\tau\ge\theta) d^\theta(\tau)\Big\} \\
&= \mathsf{E}^\infty\Big\{f_{\theta-1} I(\tau \ge \theta) \sum_{k\ge\theta} \mathsf{E}^\infty\Big(I(\tau \ge k)\, \frac{L_{k-1}}{L_{\theta-1}} \,\Big|\, \mathcal{F}_{\theta-1}\Big)\Big\} \\
&= \mathsf{E}^\infty\Big\{I(\tau \ge \theta) \sum_{k\ge\theta} f_{\theta-1} \frac{L_{k-1}}{L_{\theta-1}} I(\tau \ge \theta)\Big\} \\
&= \mathsf{E}^\infty\Big\{I(\tau \ge \theta) \sum_{k=\theta}^{\tau} f_{\theta-1} \frac{L_{k-1}}{L_{\theta-1}}\Big\}.
\end{aligned}$$

Taking summation over θ we find that

$$d(\tau)\sum_{\theta=1}^{\infty}\mathsf{E}^{\infty}f_{\theta-1}I(\tau\ge\theta)\ge\mathsf{E}^{\infty}\sum_{\theta=1}^{\tau}\sum_{k=\theta}^{\tau}f_{\theta-1}\frac{L_{k-1}}{L_{\theta-1}}$$
$$=\mathsf{E}^{\infty}\sum_{k=1}^{\tau}\sum_{\theta=1}^{\tau}f_{\theta-1}\frac{L_{k-1}}{L_{\theta-1}}.$$

From this we get $d(\tau)\ge\dfrac{\mathsf{E}^{\infty}\sum_{k=1}^{\tau}\sum_{\theta=1}^{\tau}f_{\theta-1}\frac{L_{k-1}}{L_{\theta-1}}}{\mathsf{E}^{\infty}\sum_{\theta=1}^{\tau}f_{\theta-1}}$.

Take $f_\theta=(1-\gamma_\theta)^+$. Then

$$\sum_{\theta=1}^{k}(1-\gamma_{\theta-1})^+\frac{L_{k-1}}{L_{\theta-1}}=\frac{L_{k-1}}{L_k}\sum_{\theta=1}^{k}(1-\gamma_{\theta-1})^+\frac{L_k}{L_{\theta-1}}=\frac{L_{k-1}}{L_k}\gamma_k.$$

Since $\gamma_k=\frac{L_k}{L_{k-1}}\max\{1,\gamma_{k-1}\}$, we have

$$\mathsf{E}^{\infty}\sum_{k=1}^{\tau}\sum_{\theta=1}^{\tau}f_{\theta-1}\frac{L_{k-1}}{L_{\theta-1}}=\sum_{k=1}^{\tau}\max\{1,\gamma_{k-1}\}=\sum_{k=1}^{\tau-1}\max\{1,\gamma_k\}.$$

Thus, inequality $(*)$ of Theorem 6 is proved.

§ 6. Recurrent equations for statistics π_n, φ_n, ψ_n, γ_n

1. We know from § 4 that in Variant A the value

$$\mathbb{A}_1(c) = \inf_{\tau \in \mathfrak{M}} \Big[\mathsf{P}^G(\tau < \theta) + c\mathsf{E}^G(\tau - \theta)^+ \Big]$$

can be represented in the form

$$\mathbb{A}_1(c) = \inf_{\tau \in \mathfrak{M}} \mathsf{E}^G \Big[(1 - \pi_\tau) + c \sum_{k=0}^{\tau-1} \pi_k \Big],$$

where $\boxed{\pi_n = \mathsf{P}^G(\theta \le n \,|\, \mathcal{F}_n), \quad n \ge 1, \quad \pi_0 = \pi \equiv G(0)}$. (If we have observations $X_0, X_1, \ldots$, then $\mathcal{F}_n = \mathcal{F}_n^X = \sigma(X_0, \ldots, X_n)$.) Using the Bayes formula, we find

$$\boxed{\pi_n = \frac{\sum_{\theta \le n} L_n^\theta \, \Delta G(\theta)}{L_n^G}}, \qquad \text{where} \quad L_n^\theta = \frac{d\mathsf{P}_n^\theta}{d\mathsf{P}_n}, \quad L_n^G = \frac{d\mathsf{P}_n^G}{d\mathsf{P}_n}.$$

Introduce $\varphi_n = \pi_n/(1-\pi_n)$. For the statistics φ_n one find that

$$\varphi_n = \frac{\sum_{\theta \le n} L_n^\theta \, \Delta G(\theta)}{\sum_{\theta > n} L_n^\theta \, \Delta G(\theta)} = \frac{L_n^0 \sum_{\theta \le n} \frac{L_{\theta-1}^\infty}{L_{\theta-1}^0} \, \Delta G(\theta)}{(1-G(n)) L_n^\infty}.$$

Since $\pi_n = \varphi_n/(1+\varphi_n)$, we get

$$\pi_n = \frac{\sum_{\theta \le n} \frac{L_n}{L_{\theta-1}} \, \Delta G(\theta)}{\frac{L_n}{L_{\theta-1}} \, \Delta G(\theta) + (1-G(n))},$$

and therefore

$$\pi_n = \frac{\frac{L_n}{L_{n-1}}\Big\{(1-\pi_n)\frac{\Delta G(n)}{1-G(n-1)} + \pi_{n+1}\Big\}}{\frac{L_n}{L_{n-1}}\Big\{(1-\pi_n)\frac{\Delta G(n)}{1-G(n-1)} + \pi_{n+1}\Big\} + (1-\pi_{n-1})\frac{1-G(n)}{1-G(n-1)}},$$

$$1-\pi_n = \frac{(1-\pi_{n-1})\frac{1-G(n)}{1-G(n-1)}}{\frac{L_n}{L_{n-1}}\Big\{(1-\pi_n)\frac{\Delta G(n)}{1-G(n-1)} + \pi_{n+1}\Big\} + (1-\pi_{n-1})\frac{1-G(n)}{1-G(n-1)}}.$$

Thus,

$$\varphi_n = \frac{\pi_n}{1-\pi_n} = \frac{\frac{L_n}{L_{n-1}}\left\{(1-\pi_n)\frac{\Delta G(n)}{1-G(n-1)} + \pi_{n+1}\right\}}{(1-\pi_{n-1})\frac{1-G(n)}{1-G(n-1)}}.$$

Finally,

$$\boxed{\varphi_n = \frac{L_n}{L_{n-1}}\left\{\frac{\Delta G(n)}{1-G(n)} + \varphi_{n-1}\frac{1-G(n-1)}{1-G(n)}\right\}}.$$

From the formulas $\pi_n = \dfrac{\sum_{\theta\le n} L_n^\theta \, \Delta G(\theta)}{L_n^G}$ and $\varphi_n = \dfrac{\pi_n}{1-\pi_n}$ we find also that

$$\boxed{\varphi_n = \frac{1}{G(n)} \sum_{\theta\le n} \frac{L_n}{L_{\theta-1}} \, \Delta G(\theta)}.$$

EXAMPLE. If G is geometrical distribution: $\Delta G(0) = G(0) = \pi$, $\Delta G(n) = (1-\pi)q^{n-1}p$, then

$$\frac{\Delta G(n)}{1-G(n)} = \frac{p}{q}, \quad \frac{\Delta G(n-1)}{1-G(n)} = \frac{1}{q}, \quad \text{and} \quad \varphi_n = \frac{L_n}{qL_{n-1}}(p+\varphi_{n-1}).$$

For $\widehat{\psi}_n(p) := \dfrac{\varphi_n}{p}$, $p > 0$: $\qquad \widehat{\psi}_n(p) = \dfrac{L_n}{qL_{n-1}}(1+\widehat{\psi}_{n-1}(p))$;

For $\widehat{\psi}_n := \lim\limits_{p\downarrow 0} \widehat{\psi}_n(p)$: $\qquad \boxed{\widehat{\psi}_n = \dfrac{L_n}{L_{n-1}}(1+\widehat{\psi}_{n-1})}$.

If $\varphi_0 = 0$, then $\widehat{\psi}_0 = 0$ and $\boxed{\widehat{\psi}_n = \sum\limits_{\theta=1}^{n} \dfrac{L_n}{L_{n-1}}}$.

We see that $\widehat{\psi}_n = \psi_n$ $(= \sum_{\theta=1}^{n} \frac{L_n}{L_{\theta-1}})$, where the statistics has appeared in Variant B:

$$\sum_{\theta=1}^{\infty} \mathsf{E}^{\theta}(\tau-\theta)^{+} = \mathsf{E}^{\infty} \sum_{n=1}^{\tau-1} \psi_n.$$

So, we conclude that statistics ψ_n (in Variant B) can be obtained from the statistic $\varphi_n(p)$ (which appeared in Variant A).

2. Consider the term L_n/L_{n-1} in the above formulas. Let σ-algebras $\mathcal{F}_n$ be generated by the **independent** (w.r.t. both P^0 and P^∞) observations $x_0, x_1, \dots$ with densities $f^0(x)$ and $f^\infty(x)$ for x_n, $n \geq 1$. Then

$$\frac{L_n}{L_{n-1}} = \frac{f^0(x_n)}{f^\infty(x_n)}.$$

So, in this case

$$\varphi_n = \frac{f^0(x_n)}{f^\infty(x_n)} \left\{ \frac{\Delta G(n)}{1 - G(n)} + \varphi_{n-1} \frac{1 - G(n-1)}{1 - G(n)} \right\}.$$

If $x_0, \ldots, x_n$ has the densities $f^0(x_0, \ldots, x_n)$ and $f^\infty(x_0, \ldots, x_n)$, then

$$\frac{L_n}{L_{n-1}} = \frac{f^0(x_n|x_0, \ldots, x_n)}{f^\infty(x_n|x_0, \ldots, x_n)}$$

and

$$\varphi_n = \frac{f^0(x_n|x_0, \ldots, x_{n-1})}{f^\infty(x_n|x_0, \ldots, x_{n-1})}\left\{\frac{\Delta G(n)}{1-G(n)} + \varphi_{n-1}\frac{1-G(n-1)}{1-G(n)}\right\}.$$

In the case of **Markov** observations

$$\varphi_n = \frac{f^0(x_n|x_{n-1})}{f^\infty(x_n|x_{n-1})}\left\{\frac{\Delta G(n)}{1-G(n)} + \varphi_{n-1}\frac{1-G(n-1)}{1-G(n)}\right\}.$$

From the above representations we see that

- in the case of independent observations $x_0, x_1, \ldots$ (w.r.t. P^0 and P^∞) the statistics φ_n and π_n form a **Markov sequences** (w.r.t. P^G);

- in the case of Markov sequences $x_0, x_1, \ldots$ (w.r.t. P^0 and P^∞) the **PAIRS** (φ_n, x_n) and (π_n, x_n) form **Markov sequences** (w.r.t. P^G).

§ 7. VARIANTS A and B: Solving the optimal stopping problem

1. We know that $\mathbb{A}_1(c) = \inf\limits_{\tau\in\mathfrak{M}} \Big[\mathsf{P}^G(\tau < \theta) + c\mathsf{E}^G(\tau-\theta)^+\Big]$ can be represented in the form

$$\mathbb{A}_1(c) = \inf_{\tau\in\mathfrak{M}} \mathsf{E}_\pi\left[(1-\pi_\tau) + c\sum_{n=0}^{\tau-1}\pi_n\right],$$

here E_π is the expectation E^G under assumption $G(0)=\pi$ ($\in [0,1]$).

Denote $\quad V^*(\pi) = \inf\limits_{\tau\in\mathfrak{M}} \mathsf{E}_\pi\left[(1-\pi_\tau) + c\sum\limits_{n=0}^{\tau-1}\pi_n\right], \quad$ for **fixed $c>0$**,

$Tg(\pi) = \mathsf{E}_\pi g(\pi_1)$ for any nonnegative (or bounded) function $g=g(\pi)$, $\pi\in[0,1]$,

$Qg(\pi) = \min\{g(\pi), Tg(\pi)\}$.

We assume that in our G-model

$$G(0) = \pi, \quad \Delta G(n) = (1-\pi)q^{n-1}p, \quad 0 \le \pi < 1, \quad 0 < p < 1.$$

In this case for $\varphi_n = \pi_n/(1-\pi_n)$ we have

$$\varphi_n = \frac{L_n}{qL_{n-1}}(p + \varphi_{n-1}).$$

In case of the P^0- and P^∞-i.i.d. observations $x_1, x_2, \dots$ with the densities $f^0(x)$ and $f^\infty(x)$ we have

$$\varphi_n = \frac{f^0(x_n)}{qf^\infty(x_n)}(p + \varphi_{n-1}).$$

From here it follows that (φ_n) is an homogeneous Markov sequence (w.r.t. P^G). Since $\pi_n = \varphi_n/(1+\varphi_n)$, we see that (π_n) is also P^G-Markov sequence. So, to solve the optimal stopping problem $V^*(\pi)$ one can use the General Markovian Optimal Stopping Theory.

From this theory it follows that

a) $V^*(\pi) = \lim Q^n g(\pi)$, where $g(\pi) = 1 - \pi$,
b) optimal stopping time has the form $\tau^* = \inf\{n \ge 0 \colon V^*(\pi) = 1 - \pi\}$.

Note that $\tau^* < \infty$ (P_π-a.s.) and $V^*(\pi)$ is a concave function. So,

$\boxed{\tau^* = \inf\{n \colon \pi_n \ge \pi^*\}}$ where π^* is a (unique) root of the equation $V^*(\pi) = 1 - \pi$.

We have

$$\boxed{V^*(\pi) = \lim_n Q^n(1-\pi)} \quad \text{with} \quad Q(1-\pi) = \min\Big\{(1-\pi),\ c\pi + \underbrace{\mathsf{E}_\pi(1-\pi_1)}_{=(1-\pi)(1-p)}\Big\}.$$

Since $V^*(\pi) = \lim_n Q^n(1-\pi) \le Q(1-\pi) \le 1 - \pi$, we find that

$$Q(1-\pi^*) \le 1 - \pi^*, \quad \text{or} \quad \min\Big\{(1-\pi^*),\ c\pi^* + (1-\pi^*)(1-p)\Big\} \le 1 - \pi^*.$$

From here we obtain
the **LOWER ESTIMATE** for π^*:

$$\boxed{\pi^* \ge \frac{p}{c+p}}.$$

In Topic IV (§ 3) we shall consider Variant A for the continuous (diffusion) case. In this case for π^* we shall obtain the explicit formula for π^*.

2\. In Variant B

$$\boxed{\mathbb{B}(T) = \inf_{\tau \in \mathfrak{M}_T} \sum_{\theta \geq 1} \mathsf{E}^{\theta}(\tau - \theta)^+, \quad \text{where} \quad \mathfrak{M}_T = \{\tau\colon \mathsf{E}^{\infty}\tau \geq T\}, \quad T > 0.}$$

We know that $\displaystyle\sum_{\theta \geq 1} \mathsf{E}^{\theta}(\tau - \theta)^+ = \mathsf{E}^{\infty} \sum_{n=1}^{\tau-1} \psi_n,$ where $\displaystyle\psi_n = \sum_{\theta=1}^{n} \frac{L_n}{L_{\theta-1}}.$

From here

$$\psi_n = \underbrace{\boxed{\frac{L_n}{L_{n-1}}}}_{= \frac{f^0(x_n)}{f^\infty(x_n)}} (1 + \psi_{n-1}), \qquad \psi_{-1} = 0.$$

$= \dfrac{f^0(x_n)}{f^\infty(x_n)}$ in the case of (P^0- and P^∞-) i.i.d. observations

For the more general case

$$\boxed{\mathbb{B}_F(T) = \inf_{\tau \in \mathfrak{M}_T} \sum_{\boldsymbol{\theta} \geq 1} \mathsf{E}^{\boldsymbol{\theta}} \boldsymbol{F}((\boldsymbol{\tau} - \boldsymbol{\theta})^+) \quad \text{with} \quad \begin{array}{l} F(n) = \sum\limits_{k=1}^{n} f(k), \\ F(0) = 0, \quad f(k) \geq 0 \end{array}}$$

we find that

$$\sum_{\boldsymbol{\theta} \geq 1} \mathsf{E}^{\boldsymbol{\theta}} \boldsymbol{F}((\boldsymbol{\tau} - \boldsymbol{\theta})^+) = \mathsf{E}^{\infty} \sum_{n=0}^{\tau - 1} \Psi_n(f),$$

where

$$\Psi_n(f) = \sum_{\theta=0}^{n} f(n + 1 - \theta) \frac{L_n}{L_{\theta-1}}.$$

- If $f(t) = \sum_{m=0}^{M} c_{m0} e^{\lambda_m t}$, then $\boxed{\Psi_n(f) = c_{00}\psi_n + \sum_{m=1}^{M} c_{m0}\psi_n^{(m,0)}}$

  with
$$\begin{cases} \psi_n = \dfrac{L_n}{L_{n-1}}(1+\psi_{n-1}), & \psi_{-1} = 0, \\ \psi_n^{(m,0)} = e^{\lambda_m}\dfrac{L_n}{L_{n-1}}(1+\psi_{n-1}^{(m,0)}), & \psi_{-1}^{(m,0)} = 0. \end{cases}$$

- If $f(t) = \sum_{k=0}^{K} c_{0k} t^k$, then $\boxed{\Psi_n(f) = c_{00}\psi_n + \sum_{k=1}^{K} c_{0k}\psi_n^{(0,k)}}$

  with
$$\begin{cases} \psi_n = \dfrac{L_n}{L_{n-1}}(1+\psi_{n-1}), & \psi_{-1} = 0, \\ \psi_n^{(0,k)} = \dfrac{L_n}{L_{n-1}}\left(1+\sum_{i=0}^{k} c_n^i \psi_{n-1}^{(0,i)}\right), & \psi_{-1}^{(0,k)} = 0. \end{cases}$$

- For the general case $f(t) = \sum_{m=0}^{M} \sum_{k=0}^{K} c_{mk} e^{\lambda_m t} t^k$, $\lambda_0 = 0$, we have

$$\boxed{\Psi_n(f) = \sum_{m=0}^{M} \sum_{k=0}^{K} c_{mk} \psi_n^{(m,k)}}$$

with $\quad \psi_n^{(m,k)} = \sum_{\theta=0}^{n} e^{\lambda_m (n+1-\theta)} (n+1-\theta)^k \frac{L_n}{L_{\theta-1}}.$

The statistics $\psi_n^{(m,k)}$ satisfy the system

$$\psi_n^{(m,k)} = e^{\lambda_m} \frac{L_n}{L_{n-1}} \left(\sum_{i=0}^{K} c_k^i \psi_{n-1}^{(m,i)} + 1 \right), \qquad \begin{array}{l} 0 \le m \le M, \\ 0 \le k \le K. \end{array}$$

EXAMPLE 1. If $F(n) = n$, then

$$f(n) \equiv 1, \quad \Psi_n(f) = \psi_n, \quad \text{where} \quad \psi_n = \frac{L_n}{L_{n-1}}(1 + \psi_n), \quad \psi_{-1} = 0.$$

EXAMPLE 2. If $F(n) = n^2 + n$, then

$$f(n) = 2n.$$

In this case

$$\Psi_n(f) = 2\psi_n^{(0,1)} \quad \text{with} \quad \psi_n^{(0,1)} = \frac{L_n}{L_{n-1}}(1 + \psi_{n-1} + \psi_{n-1}^{(0,1)}).$$

Thus

$$\Psi_n(f) = \Psi_{n-1}(f) + 2\psi_n.$$

3. We know that

$$\boxed{\mathbb{B}(T) = \inf_{\tau\in\mathfrak{M}_T} \sum_{\theta\geq 1} \mathsf{E}^{\theta}(\tau-\theta)^{+} = \inf_{\tau\in\mathfrak{M}_T} \mathsf{E}^{\infty} \sum_{n=1}^{\tau-1} \Psi_n}, \qquad \Psi_n = \sum_{\theta=1}^{n} \frac{L_n}{L_{\theta-1}}.$$

By the Lagrange method, to find $\inf_{\tau\in\mathfrak{M}_T} \mathsf{E}^{\infty} \sum_{n=1}^{\tau-1} \psi_n$ we need to solve the problem

$$\inf_{\tau\in\mathfrak{M}} \mathsf{E}^{\infty} \sum_{n=1}^{\tau-1} (\psi_n + c) \quad (c \text{ is a Lagrange multiplier}). \qquad (*)$$

For i.i.d. case and geometric distribution the statistics (ψ_n) form a homogeneous Markov chain.

By the Markovian optimal stopping theory, the optimal stopping time $\tau^* = \tau^*(c)$ for the problem $(*)$ has the form

$$\boxed{\tau^*(c) = \inf\{n \colon \psi_n \geq b^*(c)\}}$$

Suppose that for given $T > 0$ we can find $c = c(T)$ such that

$$\mathsf{E}^{\infty}\tau^*(c(T)) = T.$$

Then the stopping time

$$\tau^*(c(T)) = \inf\{n\colon \psi_n \geq b^*(c(T))\}$$

will be optimal stopping time in the problem

$$\mathbb{B}(T) = \inf_{\tau \in \mathfrak{M}_T} \sum_{\theta \geq 1} \mathsf{E}^{\infty}(\tau - \theta)^+$$

§ 8. VARIANTS C and D: around the optimal stopping times

1. In Variant C the risk function is

$$\mathbb{C}(T) = \inf_{\tau \in \mathfrak{M}_T} \underbrace{\sup_{\theta \geq 1} \mathsf{E}^\theta(\tau - \theta \,|\, \tau \geq \theta)}$$

$$\geq \frac{1}{\mathsf{E}^\infty \tau} \underbrace{\mathsf{E}^\infty \sum_{n=1}^{\tau-1} \psi_n}, \text{ where } \psi_n = \sum_{\theta=1}^{n} \frac{L_n}{L_{\theta-1}} \quad \textbf{(§ 6, Thm. 5)}$$

$$= \sum_{\theta=1}^{\infty} \mathsf{E}^\theta(\tau - \theta)^+ \quad \textbf{(Thm. 4)}$$

So, in the class $\widehat{\mathfrak{M}}_T = \{\tau \in \mathfrak{M}_T \colon \mathsf{E}^\infty \tau = T\}$ we have $\boxed{\mathbb{C}(T) \geq \frac{1}{T}\mathbb{B}(T)}$.

For any stopping time $\tau^0 \in \widehat{\mathfrak{M}}_T$

$$\sup_{\theta \geq 1} \mathsf{E}^\theta(\tau - \theta \,|\, \tau \geq \theta) \geq \mathbb{C}(T) \geq \frac{1}{T}\mathbb{B}(T).$$

Thus, if we take a "good" time τ^0 and find $\mathbb{B}(T)$, then we can obtain a "good" estimate for $\mathbb{C}(T)$. In Topic IV (§ 6) we consider this procedure for the case of continuous time (Brownian model).

2. Finally, in § 5 (Theorem 6) it was demonstrated that for any stopping time τ with $\mathsf{E}^\infty \tau < \infty$

$$\sup_{\theta \ge 1} \operatorname*{ess\,sup}_{\omega} \mathsf{E}^\theta \Big[(\tau - (\theta - 1))^+ \,|\, \mathcal{F}_{\theta-1} \Big] \ge \frac{\mathsf{E}^\infty \sum\limits_{k=0}^{\tau-1} \gamma_k}{\mathsf{E}^\infty \sum\limits_{k=0}^{\tau-1} (1-\gamma_k)^+}.$$

So

$$\mathbb{D}(T) \ge \inf_{\tau \in \mathfrak{M}_T} \frac{\mathsf{E}^\infty \sum\limits_{k=0}^{\tau-1} \gamma_k}{\mathsf{E}^\infty \sum\limits_{k=0}^{\tau-1} (1-\gamma_k)^+} \ge \frac{\inf\limits_{\tau \in \mathfrak{M}_T} \mathsf{E}^\infty \sum\limits_{k=0}^{\tau-1} \gamma_k}{\sup\limits_{\tau \in \mathfrak{M}_T} \mathsf{E}^\infty \sum\limits_{k=0}^{\tau-1} (1-\gamma_k)^+}.$$

G. Lorden * proved that CUSUM stopping time

$$\sigma^*(T) = \inf\{n \geq 0 \colon \gamma_n \geq d^*(T)\}$$

is asymptotically optimal as $T \to \infty$ (for i.i.d. model).

In 1986 G.V.Moustakides ** proved that CUSUM statistics (γ_n) is optimal for all $T < \infty$. We consider these problems in § 6 of Topic IV for the case of continuous time (Brownian model).

---

* "Procedures for reacting to a change in distribution", Ann. Math. Statist., 42:6 (1971), 1897–1908.

** "Optimal stopping times for detecting changes in distributions", Ann. Statist., 14:4 (1986), 1379–1387.

TOPIC IV: Quickest detection problems: Continuous time (Brownian model)

§ 1. Introduction

1.1. In the talk we intend to present the basic—from our point of view—aspects of the problems of the quickest detection of disorders in the observed data, with accent on

the **MINIMAX approaches**.

As a model of the observed process $X = (X_t)_{t \geq 0}$ with a disorder, we consider the scheme of the Brownian motion with changing drift. More exactly, we assume that

$$X_t = \mu(t-\theta)^+ + \sigma B_t, \qquad \text{or} \qquad dX_t = \begin{cases} \sigma dB_t, & t < \theta, \\ \mu\, dt + \sigma dB_t, & t \geq \theta, \end{cases}$$

where θ is a hidden parameter which can be a random variable or some parameter with values in $\overline{\mathbb{R}^+} = [0, \infty]$.

1.2. A discrete-time analogue of the process $X = (X_t)_{t \geq 0}$ is a model

$$X = (X_1, X_2, \dots, X_{\theta-1}, X_\theta, X_{\theta+1}, \dots),$$

where, for a given θ,

> $X_1, X_2, \dots, X_{\theta-1}$ are i.i.d. with the distribution F_∞ and
> $X_\theta, X_{\theta+1}, \dots$ are i.i.d. with the distribution F_0.

Walter A. Shewart was the first to use—in 1920–30s—this model for the description of the quality of manufactured product. The so-called ‘control charts’, proposed by him, are widely used in the industry until now.

The idea of his method of control can be illustrated by his own

EXAMPLE: Suppose that

for $n < \theta$ the r.v.'s X_n are $\mathcal{N}(\mu_\infty, \sigma^2)$ and
for $n \ge \theta$ the r.v.'s X_n are $\mathcal{N}(\mu_0, \sigma^2)$, where $\mu_0 > \mu_\infty$.

Shewart proposed to declare alarm about appearing of a disorder ('change-point') at a time which is

$$\tau = \inf\{n \ge 1\colon X_n - \mu_0 \ge 3\sigma\}.$$

He did not give explanations whether this (stopping, Markov) time is optimal, and much later it was shown that:

If θ has geometric distribution:

$$\mathsf{P}(\theta = 0) = \pi, \qquad \mathsf{P}(\theta = k \mid \theta > 0) = q^{k-1} p,$$

then in the problem

$$\tau \rightsquigarrow \inf_{\tau} \mathsf{P}(\tau = \theta) \tag{$*$}$$

the time $\boxed{\tau^* = \inf\Big\{n \geq 1 \colon X_n \geq c^*(\mu_0, \mu_\infty, \sigma^2, p)\Big\}}$, where $c^* = c^*(\mu_0, \mu_\infty, \sigma^2, p) = \text{const}$, is optimal for criterion ($*$).

Here the optimal decision about declaring of alarm at time n depends only on X_n. However, for the more complicated models the optimal stopping time will depend not only on the last observation X_n but on the whole past history $(X_1, \ldots, X_n)$.

1.3. This remark was used in the 1950s by E. S. Page who proposed new control charts, well known now as a CUSUM (CUmulative SUMs) method. In view of a great importance of this method, we recall its construction (for the discrete-time case).

NOTATION:

$\mathbf{P}_n^0$ and $\mathbf{P}_n^\infty$ are the distributions of the sequences $(X_1, \dots, X_n)$, $n \ge 1$, under assumptions that $\underline{\underline{\theta = 0}}$ and $\underline{\underline{\theta = \infty}}$, resp.; $\mathbf{P}_n = \frac{1}{2}(\mathsf{P}_n^0 + \mathsf{P}_n^\infty)$;

$$L_n^0 = \frac{d\mathsf{P}_n^0}{d\mathsf{P}_n}, \quad L_n^\infty = \frac{d\mathsf{P}_n^\infty}{d\mathsf{P}_n}, \quad L_n = \frac{L_n^0}{L_n^\infty} \quad \text{(the likelihood ratios);}$$

$$L_n^\theta = I(n < \theta) L_n^\infty + I(n \ge \theta) L_n^0 \frac{L_{\theta-1}^\infty}{L_{\theta-1}^0}$$

$$(L_{-1}^0 = L_{-1}^\infty = 1, \quad L_0^0 = L_0^\infty = 1).$$

For the **GENERAL DISCRETE-TIME SCHEMES**:

the **Shewart method of control charts** is based on the statistics

$$\boxed{S_n = \frac{L_n^n}{L_n^\infty}}, \qquad n \geq 0$$

the **CUSUM method** is based on the statistics

$$\boxed{\gamma_n = \max_{\theta \geq 0} \frac{L_n^\theta}{L_n^\infty}}, \qquad n \geq 0$$

It is easy to find that since $L_n^\theta = L_n^\infty$ for $\theta > n$, we have

$$\gamma_n = \max\left(1, \max_{0 \leq \theta \leq n} \frac{L_n^\theta}{L_n^\infty}\right),$$

or

$$\gamma_n = \max\left(1, \max_{0 \leq \theta \leq n} \left(\frac{L_n^0}{L_n^\infty} \frac{L_{\theta-1}^\infty}{L_{\theta-1}^0}\right)\right) = \max\left(1, \max_{0 \leq \theta \leq n} \frac{L_n}{L_{\theta-1}}\right),$$

where $L_{-1} = 1$.

Let $Z_n = \log L_n$ and $T_n = \log \gamma_n$. Then we see that

$$\boxed{T_n = \max\Big(0, Z_n - \min_{0\le\theta\le n-1} Z_\theta\Big)},$$

whence we find that

$$\boxed{T_n = Z_n - \min_{0\le\theta\le n} Z_\theta} \quad \text{and} \quad \boxed{T_n = \max(0, T_{n-1} + \Delta Z_n)},$$

where $T_0 = 0$, $\Delta Z_n = Z_n - Z_{n-1} = \log(L_n/L_{n-1})$.

(In § 6, we shall discuss the corresponding formulas for the continuous-time Brownian model. In § 6, the question about optimality of the CUSUM method will also be considered.)

§ 2. Four basic formulations (VARIANTS A, B, C, D) of the quickest detection problems for the Brownian case

Recall our basic model

$$\boxed{dX_t = \begin{cases} \sigma\, dB_t, & t < \theta, \\ \mu\, dt + \sigma\, dB_t, & t \geq \theta, \end{cases}}$$

where

- $\mu \neq 0$, $\sigma > 0$,
- $B = (B_t)_{t \geq 0}$ is a standard ($\mathsf{E}B_t = 0$, $\mathsf{E}B_t^2 = t$) Brownian motion, and
- θ is a time of appearing of a disorder.

VARIANT A. Here

$\theta = \theta(\omega)$ is a random variable with the values from $\overline{\mathbb{R}^+} = [0, \infty]$ and $\tau = \tau(\omega)$ are stopping (Markov) times (w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$, where $\mathcal{F}_t = \sigma(X_s, s \leq t)$.

- **Conditionally variational formulation:**
  In the class $\mathfrak{M}_\alpha = \{\tau \colon \mathsf{P}(\tau \leq \theta) \leq \alpha\}$, where α is a given number from $(0,1)$, to find an optimal stopping time τ_α^* for which

$$\boxed{\mathsf{E}(\tau_\alpha^* - \theta \,|\, \tau_\alpha^* \geq \theta) = \inf_{\tau \in \mathfrak{M}_\alpha} \mathsf{E}(\tau - \theta \,|\, \tau \geq \theta)}\,.$$

- **Bayesian formulation:** To find

$$\boxed{\mathbb{A}^*(c) = \inf_\tau \Big[\mathsf{P}(\tau < \theta) + c\,\mathsf{E}(\tau - \theta)^+\Big]}$$

  and an optimal stopping time $\tau_{(c)}^*$ (if it exists) for which

$$\boxed{\mathsf{P}(\tau_{(c)}^* < \theta) + c\,\mathsf{E}(\tau_{(c)}^* - \theta)^+ = \mathbb{A}^*(c)}\,.$$

VARIANT B (Generalized Bayesian). Notation:

$$\mathfrak{M}_T = \{\tau : \mathsf{E}_\infty \tau = T\}$$

[the class of stopping times τ for which the mean time $\mathsf{E}_\infty \tau$ of τ, under assumption that there was no change point (disorder) at all, equals a given constant T].

The problem is to find a stopping time τ_T^* in the class $\mathfrak{M}_T$ for which

$$\boxed{\inf_{\tau \in \mathfrak{M}_T} \frac{1}{T} \int_0^\infty \mathsf{E}_\theta (\tau - \theta)^+ \, d\theta = \frac{1}{T} \int_0^\infty \mathsf{E}_\theta (\tau_T^* - \theta)^+ \, d\theta}\,.$$

We call this variant of the quickest detection problem

generalized Bayesian

because the integration w.r.t. $d\theta$ can be considered as the integration w.r.t. the "generalized uniform" distribution on $\mathbb{R}^+$.

In § 4 we describe the structure of the optimal stopping time τ_T^* and calculate the value

$$\mathbb{B}(T) = \frac{1}{T}\int_0^T \mathsf{E}_\theta(\tau_T^* - \theta)^+ \, d\theta.$$

These results will be very useful for the description of the asymptotically ($T \to \infty$) optimal method for the **minimax**

VARIANT C. To find

$$\boxed{\mathbb{C}(T) = \inf_{\tau \in \mathfrak{M}_T} \sup_{\theta \geq 0} \mathsf{E}(\tau - \theta \,|\, \tau \geq \theta)}$$

and an optimal stopping time $\overline{\tau}_T$ if it exists.

Notice that the problem of finding $\mathbb{C}(T)$ and $\overline{\tau}_T$ is not solved yet. We do not even know whether $\overline{\tau}_T$ exists. However, for large T it is possible to find an asymptotically optimal stopping time and an asymptotic expansion of $\mathbb{C}(T)$ up to terms which vanish as $T \to \infty$.

The following minimax criterion is well known as a Lorden criterion.

VARIANT D. To find

$$\boxed{\mathbb{D}(T) = \inf_{\tau \in \mathfrak{M}_T} \sup_{\theta \ge 0} \operatorname*{ess\,sup}_{\omega} \mathsf{E}_{\theta}\big((\tau - \theta)^+ \mid \mathcal{F}_{\theta}\big)(\omega)}\,.$$

this value can be interpreted as the "worst-case" mean detection delay

Below we give the sketch of the proof that the optimal stopping time is of the form

$$\widehat{\tau}_T = \inf\{t \ge 0 \colon \gamma_t \ge d(T)\},$$

where $(\gamma_t)_{t \ge 0}$ is the corresponding CUSUM statistics.

§ 3. VARIANT A

3.1. Under assumption that $\theta = \theta(\omega)$ is a random variable with exponential distribution,

$$\mathsf{P}(\theta = 0) = \pi, \qquad \mathsf{P}(\theta > t \,|\, \theta > 0) = e^{-\lambda t}$$
$$(\pi \in [0,1) \text{ and } \lambda > 0 \text{ is known}),$$

the problems of finding both the optimal stopping times $\tau^*_{(c)}$, τ^*_α and the values

$$\mathbb{A}^*(c) = \mathsf{P}\Big(\tau^*_{(c)} < \theta\Big) + c\,\mathsf{E}\Big(\tau^*_{(c)} - \theta\Big)^+, \qquad \mathbb{A}^*(\alpha) = \mathsf{E}(\tau^*_\alpha - \theta \,|\, \tau^*_\alpha \ge \theta)$$

were solved by the author a long time ago.

We recall here the main points, since they will be useful for solving the problem of Variant B.

Introducing the *a posteriori* probability

$$\pi_t = \mathsf{P}(\theta \le t \,|\, \mathcal{F}_t), \qquad \pi_0 = \pi,$$

we find that for any stopping time τ

$$\mathsf{P}(\tau \le \theta) + c\mathsf{E}(\tau - \theta)^+ = \mathsf{E}_\pi\Big[(1 - \pi_\tau) + c\int_0^\tau \pi_t\, dt\Big],$$

where E_π stands for the expectation w.r.t. the distribution P_π of the process X with $\pi_0 = \pi$. The process $(\pi_t)_{t\ge 0}$ has the stochastic differential

$$d\pi_t = \Big(\lambda - \frac{\mu^2}{\sigma^2}\pi_t^2\Big)(1 - \pi_t)\, dt + \frac{\mu}{\sigma^2}\pi_t(1 - \pi_t)\, dX_t.$$

The process $X = (X_t)_{t \ge 0}$ admits the **innovation** representation

$$X_t = r \int_0^t \pi_s \, ds + \sigma \overline{B}_t, \quad \text{i.e.,} \quad dX_t = r\pi_t \, dt + \sigma \, d\overline{B}_t,$$

where

$$\overline{B}_t = B_t + \frac{\mu}{\sigma} \int_0^t (\theta_s - \pi_s) \, ds$$

is a Brownian motion w.r.t. $(\mathcal{F}_t^X)_{t \ge 0}$. So, one can find that

$$d\pi_t = \lambda(1 - \pi_t) \, dt + \frac{r}{\sigma} \pi_t (1 - \pi_t) \, d\overline{B}_t.$$

Consequently, the process $(\pi_t, \mathcal{F}_t^X)_{t \ge 0}$ is a diffusion Markov process and the problem

$$\tau \in \mathfrak{M}_T \quad \rightsquigarrow \quad \inf_\tau \Big[\mathsf{P}(\tau < \theta) + c \, \mathsf{E}(\tau - \theta)^+ \Big]$$

$$= \inf_\tau \mathsf{E}_\pi \Big[(1 - \pi_\tau) + c \int_0^\tau \pi_t \, dt \Big] \quad (\equiv V^*(\pi))$$

is a problem of optimal stopping for a diffusion Markov process $(\pi_t, \mathcal{F}_t)_{t \ge 0}$.

To solve this problem, we consider the corresponding

STEFAN (free-boundary) PROBLEM :

$$\boxed{\begin{aligned} V(\pi) &= 1-\pi, & \quad \pi \geq A,\\ \mathcal{A}V(\pi) &= -c\pi, & \quad \pi < A, \end{aligned}}$$

where $\mathcal{A}$ is the infinitesimal operator of the process $(\pi_t, \mathcal{F}_t)_{t\geq 0}$:

$$\mathcal{A} = \lambda(1-\pi)\frac{d}{d\pi} + \frac{1}{2}\Big(\frac{\mu}{\sigma}\Big)^2 \pi^2(1-\pi)^2 \frac{d}{d\pi^2},$$

The general solution of the equation

$$\mathcal{A}V(\pi) = -c\pi$$

for $\pi < A$ contains two undetermined constants (say, C_1 and C_2). So, we have **three unknown** constants A, C_1, C_2 and only **one** additional condition:

$\boxed{\mathbf{1}}$ $V(\pi) = 1 - \pi$ for $\pi \geq A$.

It turned out that **two** other conditions are:

$\boxed{\mathbf{2}}$ **(smooth-fit):** $\left.\dfrac{dV(\pi)}{d\pi}\right|_{\pi \uparrow A} = \left.\dfrac{dV_0(\pi)}{d\pi}\right|_{\pi \downarrow A}$ with $V_0(\pi) = 1 - \pi$,

$\boxed{\mathbf{3}}$ $\left.\dfrac{dV}{d\pi}\right|_{\pi \uparrow 0} = 0.$

These three conditions allow us to find a unique solution $V(\pi)$ of the free-boundary problem.

After that we prove that $V(\pi) = V^*(\pi)$. Thus, we find

$$V(\pi) = \begin{cases} (1-A) - \int_\pi^A y(x)\,dx, & \pi \in [0,A), \\ 1-\pi, & \pi \in [A,1], \end{cases}$$

where

$$y(x) = -C\int_0^x e^{\Lambda[G(x)-G(u)]}\,\frac{du}{u(1-u)^2}, \qquad G(u) = \log\frac{u}{1-u} - \frac{1}{u},$$

$$\Lambda = \frac{\lambda}{\rho}, \qquad C = \frac{c}{\rho}, \qquad \rho = \frac{\mu^2}{2\sigma^2}.$$

The optimal boundary point $A^* = A^*(c)$ can be found from the equation

$$C\int_0^{A^*} e^{\Lambda[G(A^*)-G(u)]}\,\frac{du}{u(1-u)^2} = 1.$$

The optimal stopping time is

$$\tau^*(c) = \inf\{t \geq 0 \colon \pi_t \geq A^*(c)\}$$

For the conditionally extremal problem

$$\tau \quad \rightsquigarrow \quad \inf_{\tau \in \mathfrak{M}_\alpha} \mathsf{E}_\pi(\tau - \theta \,|\, \tau \geq \theta)$$

the optimal stopping time τ_α^* has a very simple structure:

$$\tau_\alpha^* = \inf\{t \colon \pi_t \geq 1 - \alpha\}.$$

Indeed, $\mathsf{P}_\pi(\tau < \theta) = \mathsf{E}_\pi(1 - \pi_\tau)$ if $\pi \leq 1 - \alpha$. So, from the continuity of the process $(\pi_t)_{t \geq 0}$ it follows that we must have $1 - \pi_{\tau_\alpha^*} = \alpha$, or $\pi_{\tau_\alpha^*} = 1 - \alpha$. This proves optimality of the stopping time τ_α^*. (Note that if $\pi \geq 1 - \alpha$, then $\tau_\alpha^* = 0$.)

The corresponding value $A^*(\alpha) = \mathsf{E}(\tau^*_\alpha - \theta \,|\, \tau^*_\alpha \geq \theta)$ can be found, say, for $\pi = 0$, from the previous expression for

$$V^*(0) = \mathsf{P}(\tau^*(c) < \theta) + c\mathsf{E}(\tau^*(c) - \theta)^+.$$

Indeed, in the expression

$$\tau^*(c) = \inf\{t \geq 0 \colon \pi_t \geq A^*(c)\}$$

the threshold $A^*(c)$ depends continuously on c, and we can find $c = c_\alpha$ such that $A^*(c_\alpha) = 1 - \alpha$. So

$$\begin{aligned} V^*(0) &= \mathsf{P}(\tau^*_\alpha < \theta) + c_\alpha \mathsf{E}(\tau^*(c_\alpha) - \theta)^+ \\ &= \alpha + c_\alpha \mathsf{E}\Big(\tau^*(c_\alpha) - \theta \,|\, \tau^*(c_\alpha) \geq \theta\Big)(1 - \alpha). \end{aligned}$$

From here it follows that

$$\mathbb{A}^*(\alpha) = \mathsf{E}\Big(\tau^*(c_\alpha) - \theta \,|\, \tau^*(c_\alpha) \geq \theta\Big) = \frac{V^*(0) - \alpha}{(1 - \alpha)c_\alpha}$$

with c_α such that $A^*(\alpha) = 1 - \alpha$.

3.2. Let us turn again to the process $(\pi_t)_{t\ge 0}$ which plays a role of a sufficient statistics for the problems in Variant A. Put $\varphi_t = \pi_t/(1-\pi_t)$. Then we can find that

$$\varphi_t = \varphi_0 e^{\lambda t} L_t + \lambda e^{\lambda t} \int_0^t e^{-\lambda s} \frac{L_t}{L_s}\, ds, \tag{58}$$

where

$L_t = \dfrac{d\mathsf{P}_0}{d\mathsf{P}_\infty}(t, X)$ is the Radon–Nykodým derivative of the measure $\mathsf{P}_0(t, A) = \mathsf{Law}[(X_s, s \le t) \in A \,|\, \theta = 0]$ w.r.t. the measure $\mathsf{P}_\infty(t, A) = \mathsf{Law}[(X_s, s \le t) \in A \,|\, \theta = \infty]$.

For L_t we have the representation

$$L_t = e^{H_t}, \quad \text{where} \quad H_t = \frac{\mu}{\sigma^2} X_t - \frac{\mu^2}{2\sigma^2} t.$$

From here by the Itô formula we find that $dL_t = \frac{\mu}{\sigma^2} L_t\, dX_t$, which, together with (58) and after using the Itô formula again, gives

$$d\varphi_t = \lambda(1+\varphi_t)\, dt + \frac{\mu}{\sigma^2} \varphi_t\, dX_t.$$

The process $(\varphi_t)_{t\geq 0}$ depends on λ, of course.
Consider $\psi_t = \lim_{\lambda\downarrow 0}(\varphi_t/\lambda)$.
If $\lambda \to 0$ and $\pi_0 \to 0$ in such a way that $\pi_0/\lambda \to m$, then the formula for φ_t yields the following representation for ψ_t:

$$\boxed{\psi_t = mL_t + \int_0^t \frac{L_t}{L_s}\,ds}.$$

From here or from the equation for φ_t we find that

$$\boxed{d\psi_t = dt + \frac{\mu}{\sigma^2}\psi_t\,dX_t, \qquad \psi_0 = m}.$$

In the next section we will see that the process $(\psi_t)_{t\geq 0}$ plays a crucial role in solving the optimization problem in Variant B.

§ 4. VARIANT B

4.1. We want to solve the following optimal stopping problem: To find

$$\mathbb{B}(T) = \inf_{\tau \in \mathfrak{M}_T} \frac{1}{T} \int_0^\infty \mathsf{E}_\theta(\tau - \theta)^+ \, d\theta,$$

where θ is a parameter with values in $\mathbb{R}^+$ and $\mathfrak{M}_t = \{\tau \colon \mathsf{E}_\infty \tau = T\}$. The key point is the following representation:

$$\boxed{\int_0^\infty \mathsf{E}_\theta(\tau - \theta)^+ \, d\theta = \mathsf{E}_\infty \int_0^\tau \psi_u \, du}, \qquad (59)$$

where $d\psi_u = du + \mu\sigma^{-2}\psi_u \, dX_u$.

To prove representation (59), we note first of all that $(\tau-\theta)^+ = \int_\theta^\infty I(u \le \tau)\,du$. Using change of measure, we get

$$\mathsf{E}_\theta(\tau-\theta)^+ = \int_\theta^\infty \mathsf{E}_\theta I(u\le\tau)\,du = \int_\theta^\infty \mathsf{E}_\infty \frac{L_u}{L_\theta} I(u\le\tau)\,du = \mathsf{E}_\infty \int_0^\tau \frac{L_u}{L_\theta}\,du$$

and

$$\begin{aligned}\int_0^\infty \mathsf{E}_\theta(\tau-\theta)^+\,d\theta &= \mathsf{E}_\infty \int_0^\infty \left[\int_0^\tau \frac{L_u}{L_\theta}\,du\right] d\theta \\ &= \mathsf{E}_\infty \int_0^\tau \left[\int_0^\infty \frac{L_u}{L_\theta}\,du\right] d\theta \\ &= \mathsf{E}_\infty \int_0^\tau \psi_u\,du.\end{aligned}$$

The process $(\psi_t)_{t\ge 0}$ is a P_∞-diffusion Markov process with the differential $d\psi_t = dt + \frac{\mu}{\sigma}\psi_t\,dB_t$, $\psi_0 = m$. We see that

$$\inf_{\tau\in\mathfrak{M}_T} \int_0^\tau \mathsf{E}_\theta(\tau-\theta)^+\,d\theta = \inf_{\tau\in\mathfrak{M}_T} \mathsf{E}_\infty \int_0^\tau \psi_u\,d\theta.$$

From the general theory of optimal stopping for Markov processes it follows that an optimal stopping time in the problem

$$\tau \quad \rightsquigarrow \quad \inf_{\tau \in \mathfrak{M}_T} \mathsf{E}_\infty \int_0^\tau \psi_u \, du$$

has the following form:

$$\tau_T^* = \inf\{t \geq 0\colon\ \psi_t \geq b(T)\},$$

where $b(T)$ is such that $\mathsf{E}_\infty \tau_T^* = T$. Since $\psi_t = t + (\mu/\sigma) \int_0^t \psi_u \, dB_u$, we find that

$$\mathsf{E}_\infty \psi_{\tau_T^*} = \mathsf{E}_\infty \tau_T^*.$$

But $\psi_{\tau_T^*} = b(T)$, so that $b(T) = \mathsf{E}_\infty \tau_T^* = T$. We have got, for optimal stopping time τ_T^* in Variant B, the very simple formula:

$$\tau_T^* = \inf\{t \geq 0\colon\ \psi_t \geq T\}.$$

4.2. For this stopping time τ_T^*, the quantity $\mathsf{E}_\infty \int_0^{\tau_T^*} \psi_u \, du$ is easy to find. Indeed, consider the process $(\psi_t)_{t\ge 0}$ with $\psi_0 = x \ge 0$. The corresponding function

$$\boxed{U(x) = \mathsf{E}_\infty^{(x)} \int_0^{\tau_T^*} \psi_u \, du}$$

[$\mathsf{E}_\infty^{(x)}$ stands for averaging w.r.t. the P_∞-distribution of $(\psi_t)_{t\ge 0}$ when $\psi_0 = x$]

satisfies the backward equation

$$\boxed{L_\infty U(x) = -x, \quad \text{where} \quad L_\infty \equiv \frac{\partial}{\partial x} + \rho x^2 \frac{\partial^2}{\partial x^2} = -x, \quad \rho = \frac{\mu^2}{2\sigma^2}}.$$

Put for simplicity $\rho = 1$, then it is easy to find that

$$\boxed{U(x) = G\Big(\frac{1}{T}\Big) - G\Big(\frac{1}{x}\Big)}, \quad \text{where} \quad G(x) = \int_x^\infty \boldsymbol{F(u)} u^{-2} \, du,$$

$$\boldsymbol{F(u)} = e^u(-\,\mathbf{Ei}(-\boldsymbol{u})),$$

$$-\,\mathbf{Ei}(-\boldsymbol{u}) \equiv \int_u^\infty \frac{e^{-t}}{t} \, dt.$$

These formulae imply that

$$\mathbb{B}(\boldsymbol{T}) = \inf_{\tau\in\mathfrak{M}_T} \frac{1}{T}\int_0^\infty \mathsf{E}_\theta(\tau-\theta)^+\,d\theta = \inf_{\tau\in\mathfrak{M}_T} \frac{1}{T}\mathsf{E}_\infty \int_0^\tau \psi_u\,du$$

$$= \frac{1}{T}\mathsf{E}_\infty \int_0^{\tau_T^*} \psi_u\,du = \frac{1}{T}U(0) = \frac{1}{T}G\Big(\frac{1}{T}\Big) \overset{\text{straightforward calculations}}{=\!=}$$

$$= \boldsymbol{F}\Big(\frac{1}{\boldsymbol{T}}\Big) - \boldsymbol{\Delta}\Big(\frac{1}{\boldsymbol{T}}\Big), \quad \text{where} \quad \boxed{\Delta(b) = 1 - b\int_0^\infty e^{-bu}\,\frac{\log(1+u)}{u}\,du}.$$

Thus, $\boxed{\mathbb{B}(T) = \frac{1}{T}G\Big(\frac{1}{T}\Big) = F\Big(\frac{1}{T}\Big) - \Delta\Big(\frac{1}{T}\Big)}$ and we have the following asymptotics for small and large T:

$$\boxed{\mathbb{B}(T) = \begin{cases} \frac{T}{2} + O(T^2), & T \to 0, \\ \log T - (1+\mathbf{C}) + O(T^{-1}\log^2 T), & T \to \infty, \end{cases}}$$

where $\mathbf{C} = 0.577\ldots$ is the Euler constant.

§ 5. VARIANT C

5.1. It is clear that

$$\frac{1}{T}\int_0^\infty \mathsf{E}_\infty(\tau-\theta)^+\,d\theta =$$

$$= \frac{1}{T}\int_0^\infty \mathsf{E}_\theta(\tau-\theta\,|\,\tau\ge\theta)\,\mathsf{P}_\theta(\tau\ge\theta)\,d\theta$$

$$\left[\begin{array}{l}\text{since } \{\tau\ge\theta\}\in\mathcal{F}_\theta \text{ and}\\ \mathsf{P}_\theta(A)=\mathsf{P}_\infty(A) \text{ for } A\in\mathcal{F}_\theta\end{array}\right]$$

$$= \frac{1}{T}\int_0^\infty \mathsf{E}_\theta(\tau-\theta\,|\,\tau\ge\theta)\,\mathsf{P}_\infty(\tau\ge\theta)\,d\theta$$

$$\le \frac{1}{T}\int_0^\infty \sup_\theta \mathsf{E}_\theta(\tau-\theta\,|\,\tau\ge\theta)\,\mathsf{P}_\infty(\tau\ge\theta)\,d\theta$$

$$\left[\begin{array}{l}\text{because}\\ \frac{1}{T}\int_0^\infty \mathsf{P}_\infty(\tau\ge\theta)\,d\theta=\frac{1}{T}\mathsf{E}_\infty\tau=1\\ \text{for } \tau\in\mathfrak{M}_T=\{\tau\colon \mathsf{E}_\infty\tau=T\}\end{array}\right]$$

$$= \sup_\theta \mathsf{E}_\theta(\tau-\theta\,|\,\tau\ge\theta),$$

As a result, we find that

$$\boxed{\mathbb{B}(T)\le\mathbb{C}(T)}$$

It is clear that

$$\mathbb{C}(T) = \inf_{\tau \in \mathfrak{M}_T} \sup_{\theta \geq 0} \mathsf{E}_\theta(\tau - \theta \,|\, \tau \geq \theta) \leq \sup_{\theta \geq 0} \mathsf{E}_\theta(\tau_T^* - \theta \,|\, \tau_T^* \geq \theta) = \mathsf{E}_0 \tau_T^*.$$

The value $\mathsf{E}_0 \tau_T^*$ is easy to find from the backward equation for $\mathsf{E}_0^{(x)} \tau_T^*$: $\boxed{\mathsf{E}_0 \tau_T^* = F(1/T)}$. Taking into account the lower estimate for $\mathbb{C}(T)$, we obtain the following result:

$$F\Big(\frac{1}{T}\Big) - \Delta\Big(\frac{1}{T}\Big) \leq \mathbb{C}(T) \leq F\Big(\frac{1}{T}\Big),$$

which implies that **for large T**

$$\boxed{\log T - (1+\mathbf{C}) + O\Big(\frac{\log^2 T}{T}\Big) \leq \mathbb{C}(T) \leq \log T - \mathbf{C} + O\Big(\frac{\log^2 T}{T}\Big)}$$

For small T we have $\boxed{T/2 + O(T^2) \leq \mathbb{C}(T) \leq T + O(T^2)}$.

5.2. From the last three inequalities it follows that there exists a **GAP** between the left-hand and right-hand sides. One possibility to eliminate this gap consists in

considering a WIDER class of stopping times.

This idea, launched by M. Pollak in the discrete-time case, leads us

To consider a class of **RANDOMIZED** stopping times $\overline{\tau} = \overline{\tau}(\omega, \varpi)$, where randomization is defined by the randomness ϖ in the initial value of the process $(\psi_t)_{t \geq 0}$:

$$d\psi_t(\omega, \varpi) = dt + \rho\psi_t(\omega, \varpi)\, dX_t(\omega)$$

with $\psi_0(\omega, \varpi) = \xi(\varpi)$, where a random variable $\xi(\varpi)$ does not depend on the Brownian motion $(B_t(\omega))_{t \geq 0}$.

Denote by $\overline{\mathbb{B}}(T)$ and $\overline{\mathbb{C}}(T)$ the corresponding analogues of $\mathbb{B}(T)$ and $\mathbb{C}(T)$ when the class $\mathfrak{M}_T = \{\tau = \tau(\omega)\colon \mathsf{E}_\infty \tau = T\}$ is replaced by the wider class $\overline{\mathfrak{M}}_T = \{\overline{\tau} = \overline{\tau}(\omega, \overline{\omega})\colon \mathsf{E}_\infty \overline{\tau} = T\}$. We have

$$\mathbb{B}(T) = \overline{\mathbb{B}}(T) \le \overline{\mathbb{C}}(T)$$

and to get the estimate from above

$$\overline{\mathbb{C}}(T) \le \mathsf{E}_0 \overline{\tau}(\omega, \overline{\omega})$$

we construct the special stopping time $\overline{\tau}(\omega, \overline{\omega})$ in the following way. We take a random variable $\xi(\overline{\omega})$ with a special density $g_A(y)$ concentrated on $[0, A]$, where A will be defined later.

This density is defined from the ideas of the quasi-stationary distribution. More specifically, we take $g_A(y)$ as a solution of the forward equation

$$(y^2 g_A(y))'' - g_A'(y) = 0 \quad \text{with} \quad g_A(A) = 0, \ \int_0^A g_A(y)\,dy = 1.$$

Solving this equation we get

$$g_A(y) = f(y)\,\frac{h(A) - h(y)}{A},$$

where

$$f(y) = \frac{1}{y}\,e^{-1/y}, \qquad h(y) = ye^{1/y} - \mathrm{Ei}\left(\frac{1}{y}\right)$$

with

$$-\mathrm{Ei}(z) = \begin{cases} \displaystyle\int_{-z}^{\infty} \frac{e^{-t}}{t}\,dt, & z < 0, \\ \displaystyle\lim_{\varepsilon\to 0}\left[\int_{-z}^{-\varepsilon} \frac{e^{-t}}{t}\,dt + \int_{\varepsilon}^{\infty} \frac{e^{-t}}{t}\,dt\right], & z > 0. \end{cases}$$

Let

$$\begin{aligned}\tau_A^y &= \inf\{t\colon \psi_t(\omega,\overline{\omega}) = A \text{ with } \psi_0(\omega,\overline{\omega}) = y\},\\ \overline{\tau}^*_{g_A} &= \tau_A^{\xi(\overline{\omega})}.\end{aligned}$$

Then

$$\begin{aligned}\overline{\mathsf{E}}_\infty \overline{\tau}^*_{g_A} &= \int_0^A \Big(\mathsf{E}_\infty \tau_A^y\Big) g_A(y)\,dy = A - \left[F\left(\frac{1}{A}\right) - \Delta\left(\frac{1}{A}\right)\right]\\ &= \begin{cases} A/2 + O(A^2), & A \to 0,\\ A - \log A + (1+\mathbb{C}) + O(A^{-1}\log^2 A), & A \to \infty.\end{cases}\end{aligned}$$

Take $A = A(T)$ such that $A - [F(1/A) - \Delta(1/A)] = T$. For such a choice we have (for $\rho = 1$)

$$A(T) = \begin{cases} 2T + O(T^2), & T \to 0,\\ T + \log T - (1+\mathbb{C}) + O(T^{-1}\log^2 T), & T \to \infty.\end{cases}$$

We find that $\mathsf{E}_0\overline{\tau}^*_{g_A} = F(1/A) - \Delta(1/A)$. Taking $A = A(T)$, we see that since $F(1/T) - \Delta(1/T) = \mathbb{B}(T) = \overline{\mathbb{B}}(T) \le \overline{\mathbb{C}}(T) \le \mathsf{E}_0\overline{\tau}^*_{g_{A(T)}}$, the following inequality hold:

$$F\Big(\frac{1}{T}\Big) - \Delta\Big(\frac{1}{T}\Big) \le \overline{\mathbb{C}}(T) \le F\Big(\frac{1}{A(T)}\Big) - \Delta\Big(\frac{1}{A(T)}\Big).$$

Then

for small T: $\quad T/2 + O(T^2) \le \overline{\mathbb{C}}(T) \le T + O(T^2)$,

for large T: $\quad \overline{\mathbb{C}}(T) = \log T - (1+\mathbb{C}) + O(T^{-1}\log^2 T)$.

The existence of the optimal stopping times belonging to the classes $\mathfrak{M}_T$ and $\overline{\mathfrak{M}}_T$ for Variant C and Variant $\overline{\mathrm{C}}$ is still an open problem.

§ 6. VARIANT D

6.1. In this variant we are interested to find

$$\mathbb{D}(T) = \inf_{\tau \in \mathfrak{M}_T} \sup_{\theta \geq 0} \operatorname*{ess\,sup}_{\omega} \mathsf{E}_\theta\Big((\tau - \theta)^+ \,|\, \mathcal{F}_\theta^X\Big)(\omega).$$

We had already mentioned that here an optimal stopping exists and has the form

$$\widehat{\tau}_T = \inf\{t \geq 0 \colon \ \gamma_t \geq d(T)\},$$

where $(\gamma_t)_{t\geq 0}$ is the CUSUM-process: $\gamma_t = \sup\limits_{\theta \leq t} \dfrac{L_t}{L_\theta}$ with

$$L_t = \frac{d\mathsf{P}_0}{d\mathsf{P}_\infty}(t,X) = \frac{d\mathsf{P}_0}{d\mathsf{P}_t}(t,X) = \exp\Big\{\frac{\mu}{\sigma^2}\,X_t - \frac{\mu^2}{2\sigma^2}\,t\Big\}$$

and

$$dL_t = \frac{\mu}{\sigma^2}\,L_t\,dX_t.$$

As in the discrete-time case, γ_t can be defined also by

$$\gamma_t = \sup_{\theta \le t} \frac{d\mathsf{P}_\theta}{d\mathsf{P}_\infty}(t, X),$$

since $\mathsf{P}_\theta(\cdot) = \mathsf{Law}(X \,|\, \theta)$ and

$$\frac{d\mathsf{P}_\theta}{d\mathsf{P}_\infty}(t, X) = \frac{d\mathsf{P}_\theta}{d\mathsf{P}_t}(t, X) = \frac{d\mathsf{P}_0}{d\mathsf{P}_t}(t, X) \cdot \frac{1}{\dfrac{d\mathsf{P}_0}{d\mathsf{P}_\theta}(t, X)} = \frac{L_t}{L_\theta}.$$

Consider in more details the structure of $(\gamma_t)_{t\geq 0}$. We have

$$\gamma_t = \frac{L_t}{\inf_{\theta \leq t} L_\theta} \equiv \frac{L_t}{N_t}.$$

By the Itô formula,

$$d\gamma_t = d\Big(\frac{L_t}{N_t}\Big) = \frac{dL_t}{N_t} - \frac{L_t\,dN_t}{(N_t)^2} = \frac{\mu}{\sigma^2}\gamma_t\,dX_t - \gamma_t\,\frac{dN_t}{N_t}.$$

Note that $\gamma_t \geq 1$ and (N_t) changes values on the set $\{\gamma_t = 1\}$. It leads to the representation

$$d\gamma_t = \frac{\mu}{\sigma^2}\gamma_t\,dX_t - \gamma_t\,I(\gamma_t = 1)\frac{dN_t}{N_t}.$$

Put $H_t = -\int_0^t \gamma_s\,I(\gamma_s = 1)\frac{dN_s}{N_s}$. Then

$$d\gamma_t = dH_t + \frac{\mu}{\sigma^2}\gamma_t\,dX_t, \qquad \gamma_0 = 1.$$

Therefore,

$$\gamma_t = 1 + H_t + \int_0^t \frac{\mu}{\sigma^2}\,\gamma_s\,dX_s,$$

which is the non-homogeneous Doléans-Dade equation whose solution can be written in the form

$$\gamma_t = L_t + \int_0^t \frac{L_t}{L_s}\,dH_s. \qquad (*)$$

Recall that the process $(\psi_t)_{t\geq 0}$ satisfies the equation

$$d\psi_t = dt + \frac{\mu}{\sigma^2}\,\psi_t\,dX_t.$$

A solution of this equation with $\psi_0 = 1$ is given by

$$\psi_t = L_t + \int_0^t \frac{L_t}{L_s}\,dH_s. \qquad (**)$$

Comparing (*) and (**) reveals the similarity of the equations for $(\psi_t)_{t\geq 0}$ and $(\gamma_t)_{t\geq 0}$. The both processes are Markov processes.

Define for $\tau \in \mathfrak{M}_T$

$$D_\theta(\tau;\omega) = \mathsf{E}_\theta\Big((\tau-\theta)^+ \,|\, \mathcal{F}_\theta\Big)(\omega).$$

Since $(\tau-\theta)^+ = \int_0^\infty I(u \le \tau)\,du$, we have

$$D_\theta(\tau;\omega) = \int_0^\infty \mathsf{E}_\theta\Big\{I(u \le \tau) \,|\, \mathcal{F}_\theta\Big\}du.$$

The random variable $\xi = I(u \le \tau)$ is $\mathcal{F}_u$-measurable, therefore, by change of measures in conditional expectations,

$$\mathsf{E}_\theta(\xi \,|\, \mathcal{F}_\theta) = \mathsf{E}_\infty\left(\xi\,\frac{L_u}{L_\theta}\,\middle|\, \mathcal{F}_\theta\right)$$

and

$$\begin{aligned} D_\theta(\tau;\omega) &= \int_0^\infty \mathsf{E}_\infty\left(\frac{L_u}{L_\theta}\,I(u \le \tau)\,\middle|\,\mathcal{F}_\theta\right)du \\ &= \mathsf{E}_\infty\left[\int_0^\infty \frac{L_u}{L_\theta}\,I(u \le \tau)\,du\,\middle|\,\mathcal{F}_\theta\right] = \mathsf{E}_\infty\left[\int_0^\tau \frac{L_u}{L_\theta}\,du\,\middle|\,\mathcal{F}_\theta\right]. \end{aligned}$$

Define $\boxed{D(\tau) = \sup_{\theta \ge 0} \operatorname*{ess\,sup}_{\omega} \mathsf{E}_\theta\Big((\tau-\theta)^+ \,|\, \mathcal{F}_\theta\Big)(\omega) = \sup_{\theta \ge 0} \operatorname*{ess\,sup}_{\omega} D_\theta(\tau;\omega)}$.

This definition and the previous formulae imply that

$$\begin{aligned} D(\tau)\,\mathsf{E}_\infty H_\tau &= \mathsf{E}_\infty(D(\tau) H_\tau) = \mathsf{E}_\infty \int_0^\infty D(\tau) I(\theta \le \tau)\, dH_\theta \\ &\ge \mathsf{E}_\infty \int_0^\infty D_\theta(\tau;\omega) I(\theta \le \tau)\, dH_\theta \\ &= \mathsf{E}_\infty \int_0^\infty I(\theta \le \tau) \mathsf{E}_\infty\left[\int_0^\tau \frac{L_u}{L_\theta}\, du \,\middle|\, \mathcal{F}_\theta\right] dH_\theta \\ &= \mathsf{E}_\infty \int_0^\tau \mathsf{E}_\infty\left[\int_0^\tau \frac{L_u}{L_\theta}\, du \,\middle|\, \mathcal{F}_\theta\right] dH_\theta. \end{aligned}$$

This inequality, together with the property $\mathsf{E}_\infty \gamma_\tau = 1 + \mathsf{E}_\infty H_\tau$, leads (after some transformations) to the following important estimate:

$$D(\tau) \ge \frac{\mathsf{E}_\infty \int_0^\tau \gamma_t\, dt}{\mathsf{E}_\infty \gamma_\tau}.$$

Therefore, $$\boxed{\mathbb{D}(T) = \inf_{\tau \in \mathfrak{M}_T} D(\tau) \ge \frac{\inf_{\tau \in \mathfrak{M}_T} \mathsf{E}_\infty \int_0^\tau \gamma_t\, dt}{\sup_{\tau \in \mathfrak{M}_T} \mathsf{E}_\infty \gamma_\tau}}.$$ IV-1-40

Define $\boxed{\tau(d) = \inf\{t \colon \gamma_t = d\}}$ and **find $\mathsf{E}_\infty \tau(d)$ and $\mathsf{E}_0 \tau(d)$**.

With respect to the measure P_∞ the process $(\gamma_t)_{t \ge 0}$ is a diffusion process with values in $[1, \infty)$, where the point $\{1\}$ is a reflection boundary (with $f'(1+) = 0$). If $V(x) = \mathsf{E}_\infty(\tau(d) \,|\, \gamma_0 = x)$, then, taking into account that $d\gamma_t = (\mu/\sigma)\gamma_t \, dB_t + dH_t$, we find (for $\rho \equiv \mu^2/(2\sigma^2) = 1$) that

$$x^2 V''(x) = -1, \quad x > 0, \qquad V'(1+) = 0, \quad V(d) = 0.$$

So, $V(x) = d - x + \log(x/d)$ and

$$\boxed{\mathsf{E}_\infty(\tau(d) \,|\, \gamma_0 = 1) = V(1) = d - 1 - \log d}\,.$$

Similarly, for $U(x) = \mathsf{E}_0(\tau(d) \,|\, \gamma_0 = x)$ we have $d\gamma_t = (\mu/\sigma)\gamma_t \, dB_t + dH_t$ and we find (for $\rho = 1$) that

$$\frac{\mu^2}{\sigma^2} x U'(x) + \frac{\mu^2}{2\sigma^2} x^2 U''(x) = -1 \quad \text{with} \quad U'(1+) = 0, \quad U(d) = 0.$$

It gives $U(x) = \dfrac{1}{d} - \dfrac{1}{x} - \log \dfrac{d}{x}$. So,

$$\boxed{\mathsf{E}_0(\tau(d) \,|\, \gamma_0 = 1) = U(1) = \tfrac{1}{d} + \log d - 1}.$$

Define

$$\tau(d(T)) = \inf\{t \geq 0 \colon \gamma_t = d(T)\}.$$

Then we get the following formula for $d(T)$:

$$\mathsf{E}_\infty\Big(\tau(d(T)) \,|\, \gamma_0 = 1\Big) = d(T) - 1 - \log d(T) = T.$$

Solving the optimal stopping problems

$$\tau \rightsquigarrow \sup_{\tau \in \mathfrak{M}_T} \mathsf{E}_\infty \gamma_\tau \quad \text{and} \quad \tau \rightsquigarrow \inf_{\tau \in \mathfrak{M}_T} \mathsf{E} \int_0^\tau \gamma_t \, dt,$$

we find that for both of them the optimal stopping time in the class $\mathfrak{M}_T$ is $\tau_T^* = \tau(d(T))$. Hence

$$\mathsf{E}_0 \tau_T^* = \mathsf{E}_0 \tau(d(T)) \geq \mathbb{D}(T) \geq \frac{\mathsf{E}_\infty \int_0^{\tau_T^*} \gamma_t \, dt}{\mathsf{E}_\infty \gamma_{\tau_T^*}} = \frac{1}{d(T)} \mathsf{E}_\infty \int_0^{\tau_T^*} \gamma_t \, dt. \quad (***)$$

The calculations give that $\mathsf{E}_\infty \int_0^{\tau_T^*} \gamma_t \, dt = d(T) \log d(T) + 1 - d(T)$. So, from (***), taking into account that $\mathsf{E}_0 \tau_T^* = 1/d(T) + \log d(T) - 1$, we find

$$\frac{1}{d(T)} + \log d(T) - 1 \geq \mathbb{D}(T) \geq \frac{d(T) \log d(T) + 1 - d(T)}{d(T)}.$$

Here left-hand and right-hand sides coincide. Therefore,

$$\mathbb{D}(T) = \log d(T) - 1 + \frac{1}{d(T)}.$$

Since $d(T) - 1 - \log d(T) = T$, we find the following asymptotics:

$$\mathbb{D}(T) = \begin{cases} T + O(T^2), & T \to 0, \\ \log T - 1 + O(T^{-1}), & T \to \infty. \end{cases}$$

Recall that

$$\mathbb{B}(T) = \begin{cases} T/2 + O(T^2), & T \to 0, \\ \log T - 1 - \mathbb{C} + O(T^{-1} \log T), & T \to \infty, \end{cases}$$

and

$$T/2 + O(T^2) \le \mathbb{C}(\boldsymbol{T}) \le T + O(T^2), \quad T \to 0,$$

$$\log T - (1 + \mathbf{C}) + O\left(\frac{\log^2 T}{T}\right) \le \mathbb{C}(\boldsymbol{T}) \le \log T - \mathbf{C} + O\left(\frac{\log^2 T}{T}\right), \quad T \to \infty.$$

For $\overline{\mathbb{C}}(T)$ we have got ($\mathbf{C} = 0.577\ldots$ is the Euler constant)

$$T/2 + O(T^2) \le \overline{\mathbb{C}}(\boldsymbol{T}) \le T + O(T^2), \quad T \to 0,$$
$$\overline{\mathbb{C}}(\boldsymbol{T}) = \log T - (1 + \mathbf{C}) + O(T^{-1} \log^2 T), \quad T \to \infty.$$

TOPIC V: Applications-1: Detection of spontaneously appearing effects

§ 1. INTRODUCTION

1.1. Fifty years ago, near the end of 1958, Andrei Nikolaevich Kolmogorov held several conversations with Yurii Borisovich Kobzarev, an expert in statistical radio engineering and the founder of the Soviet school of radiolocation. One of the prime interest of Kobzarev was related to

> the correct formulation of the problem of quickest detection of the reflected signal arriving from a target appearing at a **RANDOM** time which is not known in advance.

He said that in real systems one often uses detection methods based on the

Neyman–Pearson criterion and **Wald criterion**.

However, these methods are optimal **only** in problems of distinguishing the following **TWO** hypotheses:

$\boldsymbol{H_\infty}$:	the target appears at time $\boldsymbol{\theta = \infty}$, i.e., there is no target during the whole period of observation;
$\boldsymbol{H_0}$:	the target is present from the very beginning of observation, i.e., from the time $\boldsymbol{\theta = 0}$.

At that it is assumed that the observed process $X = (X_t)_{t \geq 0}$ is of the form

$$X_t = \begin{cases} N_t, & \text{i.e., only ``}\underline{\underline{\text{noise}}}\text{'',} \quad \text{under hypothesis } H_\infty, \\ N_t + S_t, & \text{i.e., ``}\underline{\underline{\text{noise+signal}}}\text{'',} \quad \text{under hypothesis } H_0. \end{cases}$$

From the point of view of statistics, the key role in distinguishing these two hypotheses is played by the **likelihood ratio**

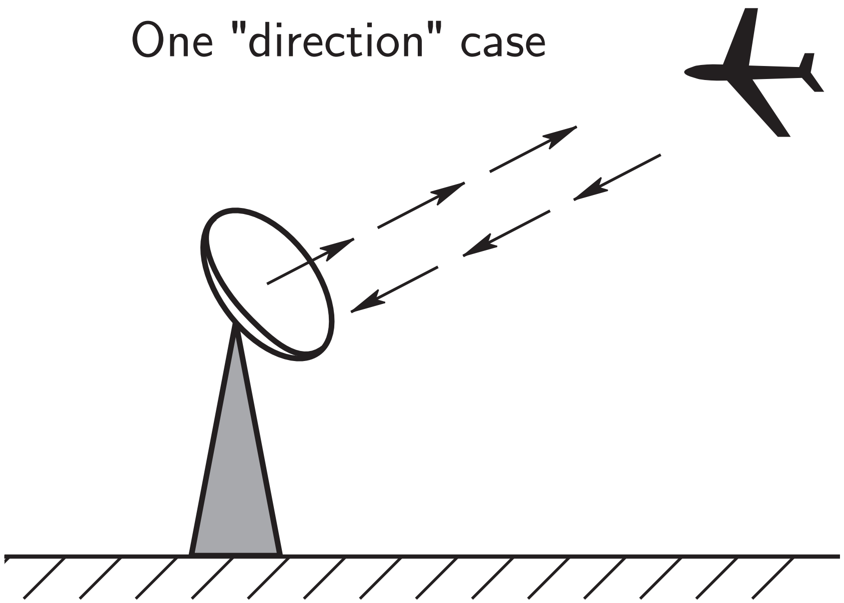
$$L_t = \frac{d\mathsf{P}_0}{d\mathsf{P}_\infty}(t, X)$$

which is the Radon–Nykodým derivative of the measure $\mathsf{P}_0 = \mathrm{Law}(X \mid H_0)$ w.r.t. the measure $\mathsf{P}_\infty = \mathrm{Law}(X \mid H_\infty)$, $t \ge 0$.

For example, the **Neyman–Pearson method** states that the optimal procedure of distinguishing the hypotheses H_0 and H_∞ by the criterion $\inf(\alpha + \beta)$, where α and β are the first and second kind errors, under assumption that we know $(X_t)_{t \le T}$, is the following one:

$$\boxed{\text{we accept} \quad \begin{array}{l} H_0, \text{ if } L_T \ge K, \\ H_\infty, \text{ if } L_T < K, \end{array} \quad \text{where } K \text{ is a certain constant.}}$$

At that $\boxed{\inf(\alpha + \beta) = \frac{1}{2}(1 - \|\mathsf{P}_0 - \mathsf{P}_\infty\|)}$, where $\|\mathsf{P}_0 - \mathsf{P}_\infty\|$ is the distance in variation between the measures $\mathsf{P}_0 = \mathrm{Law}(X_s, s \le T \mid H_0)$ and $\mathsf{P}_\infty = \mathrm{Law}(X_s, s \le T \mid H_\infty)$.



Kobzarev emphasized that in reality a target (signal) can appear not only at time 0 or ∞ but at any time θ. Thus, we actually have not **TWO** hypotheses H_0 and H_∞ but **CONTINUUM** of hypotheses H_θ if $\theta \in [0, \infty]$, at that the observed process is $\boxed{X_t = \begin{cases} N_t, & t < \theta, \\ N_t + S_{t-\theta}, & t \geq \theta. \end{cases}}$

[Change-point θ was called a time of appearing of a **disorder** (**disruption**).] In that way we arrived at a problem of

<u>mathematical formulation of the problem of detection of θ</u>

<u>taking into account the specific features of</u>

<u>real systems like a radiolocator</u>.

1.2. In January of 1959 A. N. Kolmogorov said me that we should engage ourselves into the problems which Kobzarev was interested in.

We had many consultations with experts which gave us a lot of useful information about properties which are demanded from "optimal", or "good", systems of detection of spontaneously appearing targets (signals), about statistical data on radar noise, etc.

The considered systems had the following important feature: in contrast to both Neyman–Pearson and Wald procedures they were **multistage**, i.e., after one makes a decision that a target had appeared, the whole system does not stop working but starts anew.

We soon understood that the

CORRECT FORMULATION of the PROBLEM

must take into account the following two requirements:

(A) if a target does not appears during a long period of observation, then this should result in sounding a “signal of alarm” **as rarely as possible**;

(B) if a target appears, then the “signal of alarm” should follow it with **minimal possible delay**.

It is natural to seek **numerical characteristics** of the quality of observation systems according to requirements (A) and (B). To this end one could—by analogy with the theory of testing two hypotheses—consider the following quantities:

$\alpha =$ **"probability of a false alarm",**

$\beta =$ **"probability of false tranquillity".**

Here we should note that introducing these quantities implies that the whole observation process is **divided into separate stages**, each one being finished with one of two decisions:

either $\boxed{\text{"}\textit{there is a target}\text{"}}$ or $\boxed{\text{"}\textit{there is no target}\text{"}}$.

The first decision means sounding of alarm, the second decision means merely passing to the next stage of observation.

It is clear that the

probabilities α and β

in themselves, without taking into account the duration of individual stages of observation,

are NOT reasonable characteristics

of an observation system.

For example: in practice one often meets with the probability of false tranquillity β which is close to 1, and this does not discredit the system of observation provided that the duration of individual stages is sufficiently small.

These considerations suggested that it is reasonable **to characterize the quality of an observation system**

- **from the point of view of requirement (A)** by the expectation

$$\boxed{T = \mathsf{E}_\infty \tau}$$

of the interval τ between two false alarms under condition that there is no target (i.e., under hypothesis H_∞);

- **from the point of view of requirement (B)** by

$$\boxed{\mathbb{R} = \mathsf{E}_* \sigma}\,{}^*,$$

the expectation of the time σ between appearance of a target and sounding the alarm under assumption that a target appears when the stationary regime have been already established.

---

* E_* stands for averaging w.r.t. the measure P_* which corresponds to the established stationary regime (the necessary details will be given later).

Requirements (A) and (B) “conflict” with each other.

Notation: If the expectation $\mathsf{E}_\infty \tau$ equals T, then we denote the minimal delay time $\mathbb{R}$ by $\mathbb{R}(T)$.

The above discussion shows that to tackle the problem of finding

“optimal observation system with minimal delay time $\mathbb{R}(T)$”

one should determine the exact meaning of the terms

- the **observed process** $X = (X_t)_{t \ge 0}$,
- the **detection system**,
- the **established stationary regime** which is a background for the appearing target.

Numerous consultations with experts in radio engineering made it clear that it is very difficult to cover all arising requirements on and all types of the process $X = (X_t)_{t \geq 0}$ which arrives to the entrance of a decision device and which contains the noise component and—possibly—a signal component.

Kolmogorov emphasized that it is very hard to embrace **ALL** the requirements and we could succeed by concentrating upon

the "MOST DIFFICULT" case of

detection of a signal hidden in a noise.

This case corresponds to the model in which

- the noise is simulated by the (Gaussian) white noise (with zero mean) and
- the appearing signal is a constant.

Thus, if θ is a time of appearing of a signal, then the observed process $\xi = (\xi_t)_{t \geq 0}$ is

$$\xi_t = \begin{cases} \sigma\delta_t, & t < \theta, \\ r + \sigma\delta_t, & t \geq \theta, \end{cases} \tag{60}$$

where δ_t is white noise.

In the 1950–60s processes of type of "white noise", "color noise"—though without sufficiently exact mathematical definitions—were very popular in the literature on radio engineering.

It was quickly understood that to get an exact mathematical model for the observed process $X = (X_t)_{t\ge 0}$ one should "integrate" the preceding relation (60) for $(\xi_t)_{t\ge 0}$. Then for $X_t = \int_0^t \xi_s\, ds$ we get the formula

$$\boxed{X_t = r\int_0^t I(s>\theta)\, ds + \sigma B_t}, \tag{61}$$

where $B_t = \int_0^t \delta_s\, ds$ is a Brownian motion (Wiener process).

Rewriting (61) in differentials, we find that

$$\boxed{dX_t = rI(t>\theta)\, dt + \sigma\, dB_t}. \tag{62}$$

That is mainly this model that we investigate below.

Certainly, discrete-time models are also of interest. We consider such general models (without assumption of independency) in § 10.

In the following two sections we will discuss the special systems of detection (**W- and NP-systems**) which are based on ideas of the **Wald and Neyman–Pearson methods**. These cases will make clear the meaning of the notions **detection system** and **established stationary regime** in multistage detection problems. The main results of these sections are formulae for the corresponding delay times $\mathbb{R}_{\mathsf{W}}(T)$ and $\mathbb{R}_{\mathsf{NP}}(T)$.

§ 2. Multistage cyclic-return detection $\mathbb{W}$-system

2.1. From (62) it follows that if $\theta = 0$, then

$$dX_t = r\,dt + \sigma\,dB_t, \qquad X_0 = 0, \tag{63}$$

and if $\theta = \infty$, then

$$dX_t = \sigma\,dB_t, \qquad X_0 = 0. \tag{64}$$

According to A. Wald, the sequential likelihood criterion in the problem of distinguishing two hypotheses $H_0\colon \theta = 0$ and $H_\infty\colon \theta = \infty$ is based on the statistics

$$L_t = \frac{d\mathsf{P}_0}{d\mathsf{P}_\infty}(t, X)$$

which is the Radon–Nykodým derivative of the measure $\mathsf{P}_0(\cdot) = \mathrm{Law}(X_s, s \le t \,|\, \theta = 0)$ w.r.t. the measure $\mathsf{P}_\infty(\cdot) = \mathrm{Law}(X_s, s \le t \,|\, \theta = \infty)$.

The logarithm

$$Z_t = \log L_t \tag{65}$$

is easy to find:

$$Z_t = \frac{r}{\sigma^2} X_t - \frac{r^2}{2\sigma^2}\, t. \tag{66}$$

Consequently,

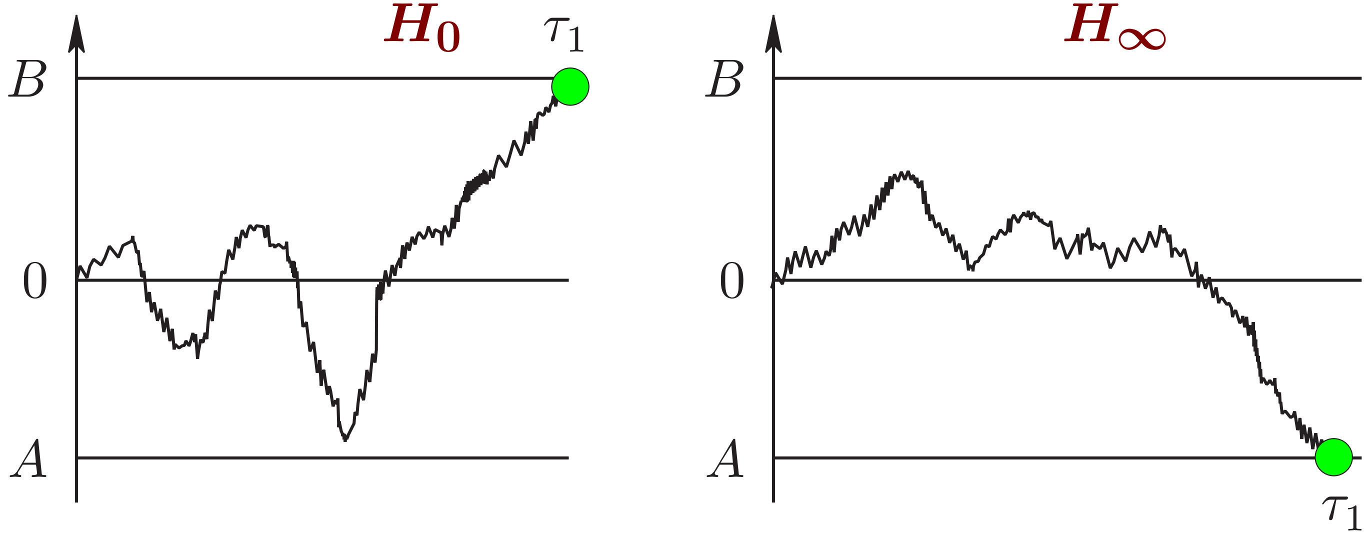
$$Z_t = \begin{cases} \dfrac{r^2}{2\sigma^2}\, t + \dfrac{r}{\sigma} B_t, & \text{if } \theta = 0, \\ -\dfrac{r^2}{2\sigma^2}\, t + \dfrac{r}{\sigma} B_t, & \text{if } \theta = \infty. \end{cases} \tag{67}$$

The **Wald (one-stage) method** of distinguishing between two hypotheses consists in

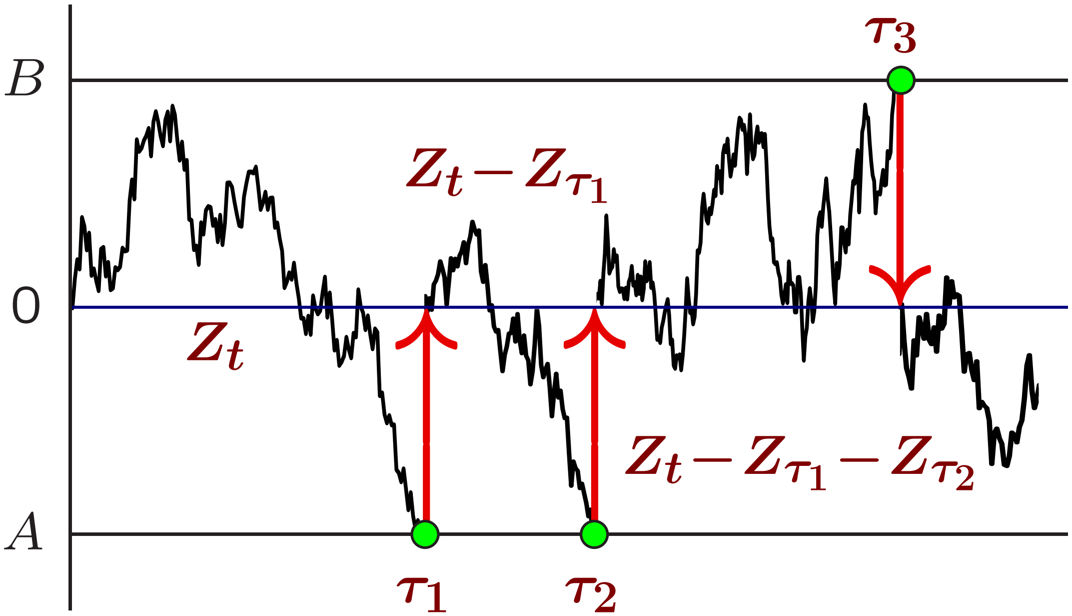
choosing two constants A and B, $A < 0 < B$, and observing the process Z_t, $t \geq 0$, up to the time

$$\tau(A,B) = \inf\{t \geq 0 \colon Z_t \notin (A,B)\};$$

if $Z_{\tau(A,B)} = A$, then we accept the hypothesis H_∞, and if $Z_{\tau(A,B)} = B$, then we accept the hypothesis H_0:



The considered **multistage W(Wald)-system** of observation/detection can be illustrated as follows:



When the process Z_t reaches the level A (at time τ_1), the process returns to zero, so that just after the time τ_1 the observed process Z_t^* is $Z_t - A$.

When $Z_t - A$ in turn reaches the level A (at time τ_2), it drops to zero, and we observe the process $Z_t^* = Z_t - 2A$. The times τ_1, τ_2 and other times of reaching the lower bound suggest that, to all appearance, there is only noise, and we decide that there is no disorder.

The time τ_3—which is the first time when the process reaches the level B—is identified with the time of alarm about the appearance of a "disorder". If in fact there was no disorder, then it is a false alarm and $\mathsf{E}_\infty \tau_3$ is the time till the first false alarm. However, in any case, just after reaching the level B the process is shifted to zero.

Suppose that such a procedure lasts infinitely long and $\theta = \infty$. Let $(Z_t^*)_{t \geq 0}$ be a cyclic-return process described above; it is generated by the process $(Z_t)_{t \geq 0}$ and is built upon the times $\tau_1, \tau_2, \tau_3, \ldots$ in such a way that, as soon as a level A or B is reached, the process returns instantaneously to zero.

Under the hypothesis H_∞ the process $(Z_t^*)_{t\ge 0}$ enters into a **stationary regime** in a sense that the limit

$$q_\infty^*(z)\,dz = \lim_{t\to\infty} \mathsf{P}_\infty\Big(z \le Z_t^* < z + dz \,|\, Z_0^* = z_0\Big), \qquad A \le z \le B,$$

exists for all $z_0 \in (A, B)$ (see, e.g., **(Karlin, Taylor, 1981, p. 260–261)**).

By definition, the **mean time of delay** in detection of a disorder is

$$\boxed{\mathbb{R}_{\mathsf{W}}(A,B) = \int_A^B \mathsf{E}_0(\tau^*(B)\,|\,Z_0^* = z) q_\infty^*(z)\,dz}, \tag{68}$$

where $\tau^*(B) = \inf\{t\colon Z_t^* = B\}$ is the first-exit time of the process $(Z_t^*)_{t\ge 0}$ to the level B.

The expectation $T_\infty(z) = \mathsf{E}_\infty[\tau^*(B)\,|\,Z_0^* = z]$ is the **mean time till the first false alarm** under assumption that $Z_0^* = z$ and the hypothesis H_∞ is true.

The problem of the quickest detection of a disorder, provided that the disorder appears against a background of the established stationary regime and we use the cyclic-return W-procedure, consists in finding

$$\boxed{\mathbb{R}_{\mathsf{W}}(T) = \inf_{\{(A,B)\colon T(A,B)=T\}} \mathbb{R}_{\mathsf{W}}(A,B), \quad \text{where} \quad T(A,B) = T_\infty(0)}\,.$$

Since all these quantities depend on r and σ only through the ratio signal/noise $\rho = r^2/(2\sigma^2)$, for simplicity of notation we can take $r = \sqrt{2}$, $\sigma = 1$. Then the process $(Z_t)_{t\ge 0}$ is

$$\begin{aligned} Z_t &= z - t + \sqrt{2}W_t \quad \text{under hypothesis } H_\infty, \\ Z_t &= z + t + \sqrt{2}W_t \quad \text{under hypothesis } H_0. \end{aligned}$$

LEMMA 1. The **mean time between two false alarms** is

$$\boxed{T_\infty(0) = \frac{\omega(\alpha,\beta)}{\alpha}}, \qquad (69)$$

where $\omega(\alpha,\beta) = (1-\alpha)\log\dfrac{1-\alpha}{\beta} + \alpha\log\dfrac{\alpha}{1-\beta}$ and

$\alpha = \mathsf{P}(Z_{\tau_1} = B \,|\, Z_0 = 0, H_\infty)$ (**probability of the false alarm**),

$\beta = \mathsf{P}(Z_{\tau_1} = A \,|\, Z_0 = 0, H_0)$ (**probability of the false tranquillity**)

are given by the "**Wald formulae**":

$$\boxed{\alpha = \frac{1-e^A}{e^B - e^A}, \qquad \beta = \frac{e^A(e^B-1)}{e^B-e^A}}. \qquad (70)$$

PROOF. "Wald's formulae" (70): Let $\alpha(x) = \mathsf{P}(Z_{\tau_1} = B \,|\, Z_0 = x, H_\infty)$ and $\beta(x) = \mathsf{P}(Z_{\tau_1} = A \,|\, Z_0 = x, H_\infty)$. It is clear that

$$\alpha(B) = 1, \quad \alpha(A) = 0 \qquad \text{and} \qquad \beta(B) = 0, \quad \beta(A) = 1.$$

Moreover, α and β solves the (backward) Kolmogorov equations

$$\alpha''(x) - \alpha'(x) = 0 \quad \text{and} \quad \beta''(x) + \beta'(x) = 0.$$

It is easy to find that

$$\alpha(x) = \frac{e^x - e^A}{e^B - e^A}, \qquad \beta(x) = \frac{e^A(e^{B-x} - 1)}{e^B - e^A}, \tag{71}$$

which implies, in particular, the formulae (70) for $\alpha = \alpha(0)$ and $\beta = \beta(0)$.

(70) is proved

Remark. From the Wald formulae (70) if follows that

$$A = \log \frac{\beta}{1-\alpha}, \qquad B = \log \frac{1-\beta}{\alpha}. \tag{72}$$

Proof of (69): Introduce the expectations

$$M_\infty(x) = \mathsf{E}_\infty(\tau_1 \,|\, Z_0 = x), \qquad M_0(x) = \mathsf{E}_0(\tau_1 \,|\, Z_0 = x).$$

For these expectations we have the equations

$$\begin{aligned} M''_\infty(x) - M'_\infty(x) &= -1 \quad \text{with } M_\infty(A) = M_\infty(B) = 0, \\ M''_0(x) + M'_0(x) &= -1 \quad \text{with } M_0(A) = M_0(B) = 0, \end{aligned}$$

whose solutions are given by

$$M_\infty(x) = \frac{(e^B - e^x)(B-A)}{e^B - e^A} - B + x, \tag{73}$$

$$M_0(x) = \frac{(e^B - e^{A+B-x})(B-A)}{e^B - e^A} + A - x, \tag{74}$$

respectively. In particular,

$$M_\infty(0) = \frac{B(e^A - 1) - A(e^B - 1)}{e^B - e^A} \qquad (= \omega(\alpha, \beta)),$$

$$M_0(0) = \frac{Be^B(1 - e^A) + Ae^A(e^B - 1)}{e^B - e^A} \qquad (= \omega(\beta, \alpha)).$$

Let $T_\infty(x)$ be the mean time till exit to the boundary B under assumptions that $Z_0 = x$ and that no disorder appears during the whole period of observation.

By the total probability formula,

$$T_\infty(x) = M_\infty(x) + (1 - \alpha(x))T_\infty(0), \tag{75}$$

whence

$$T_\infty(0) = \frac{M_\infty(0)}{\alpha} = \frac{\omega(\alpha, \beta)}{\alpha}. \qquad \boxed{\text{(69) \textbf{is proved}}}$$

Lemma 1 is proved.

2.2. Let us turn to formula (68). There is no difficulty to find $T_0(z) = \mathsf{E}_0(\tau^*(B) \,|\, Z_0^* = z)$. Indeed, by analogy with (75),

$$T_0(z) = M_0(z) + \beta(z) T_0(0),$$

whence

$$T_0(0) = \frac{M_0(0)}{1-\beta} = \frac{\omega(\beta,\alpha)}{1-\beta},$$

and, therefore,

$$T_0(z) = M_0(z) + \beta(z)\,\frac{M_0(0)}{1-\beta}, \tag{76}$$

where β, $\beta(z)$, and $M_0(z)$ are given in (70), (71), and (74).

The limiting stationary density $q^*_\infty(z)$ can be found by the methods based on the Green function (according to the formula (81) below).

Recall that if $(Z_t)_{t\geq 0}$ is a regular diffusion process with phase space $[A, B]$ and stochastic differential (in (A, B))

$$dZ_t = \mu(Z_t)\,dt + \sigma(Z_t)\,dB_t, \qquad Z_0 = x \quad (x \in (A, B)),$$

then the **Green function** $G(x, z)$ of this process is given by

$$G(x,z) = \begin{cases} 2\,\dfrac{(S(x)-S(A))(S(B)-S(z))}{S(B)-S(A)}\,\dfrac{1}{\sigma^2(z)s(z)}, & A \leq x \leq z \leq B, \\ 2\,\dfrac{(S(B)-S(x))(S(z)-S(A))}{S(B)-S(A)}\,\dfrac{1}{\sigma^2(z)s(z)}, & A \leq z \leq x \leq B, \end{cases} \tag{77}$$

where the **scale function** $S(x)$ is given by

$$S(x) = \int_c^x s(z)\,dz \qquad \text{with} \quad s(z) = \exp\left\{-\int_c^z \frac{2\mu(y)}{\sigma^2(y)}\,dy\right\} \tag{78}$$

(c is a fixed constant).

REMARK 1. **Regularity** of the process $(Z_t)_{t\geq 0}$ (on the interval (A,B)) means that for all $x \in (A,B)$ and $y \in (A,B)$

$$\mathsf{P}(\sigma(y) < \infty \,|\, \sigma(0) = x) > 0, \quad \text{where } \sigma(y) = \inf\{t > 0\colon Z_t = y\}.$$

REMARK 2. The function $2/[\sigma^2(z)s(z)]$ is called a **density of speed measure** $m(dz)$ given by

$$m(dz) = \frac{2}{\sigma^2(z)s(z)}\,dz.$$

REMARK 3. The relations defining the Green function $G(x,z)$ are

$$\mathsf{E}_x \int_0^{\sigma(A,B)} I_\Gamma(Z_t)\,dt = \int_\Gamma G(x,z)\,dz, \tag{79}$$

where Γ are Borel sets in $[A,B]$ and $\sigma(A,B) = \inf\{t\colon Z_t = A \text{ or } B\}$. This implies that $G(x,z)$ is the density of the mean time (up to $\tau(A,B)$) which the trajectory $(Z_t)_{t\geq 0}$ passes in a point z under assumption that $Z_0 = x$.

In the considered case of the process $(Z_t)_{t\geq 0}$ with the differential $dZ_t = -dt + \sqrt{2}\,dB_t$, we have $s(z) = e^{z-c}$ and $S(z) = e^{z-c} - 1$. Thus, from (77) we find that

$$G(x,z) = \begin{cases} \dfrac{(e^x - e^A)(e^B - e^z)}{(e^B - e^A)e^z}, & A \leq x \leq z \leq B, \\ \dfrac{(e^B - e^x)(e^z - e^A)}{(e^B - e^A)e^z}, & A \leq z \leq x \leq B, \end{cases}$$

and, therefore,

$$G_\infty(0,z) = G(0,z) = \begin{cases} \dfrac{(1 - e^A)(e^{B-z} - 1)}{e^B - e^A}, & 0 \leq z \leq B, \\ \dfrac{(e^B - 1)(1 - e^{A-z})}{e^B - e^A}, & A \leq z \leq 0. \end{cases}$$

This and the well-known formula **(Karlin, Taylor, 1981, p. 260–261)**

$$q^*_\infty(z) = G_\infty(0,z)\left(\int_A^B G_\infty(0,y)\,dy\right)^{-1}$$

gives the following result.

LEMMA 2. The stationary density $q^*_\infty(z)$ is given by

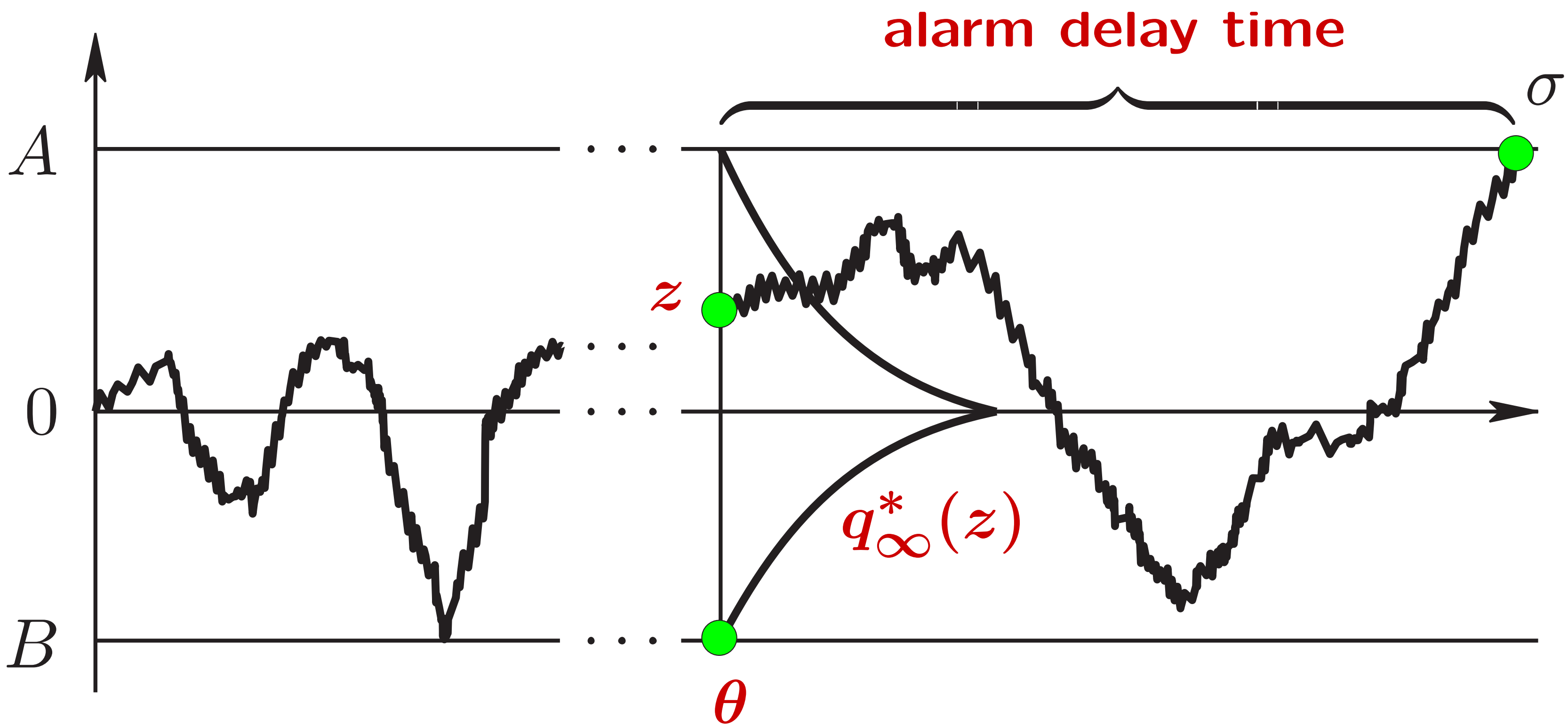
$$q^*(y) = \begin{cases} \dfrac{(1-e^B)(e^{A-y}-1)}{A(1-e^B)+B(e^A-1)}, & A \le y \le 0, \\ \dfrac{(e^A-1)(1-e^{B-y})}{A(1-e^B)+B(e^A-1)}, & 0 \le y \le B. \end{cases} \tag{81}$$

Inserting (76) and (81) into

$$\mathbb{R}_{\mathsf{W}}(A,B) = \int_A^B T_0(z) q^*_\infty(z)\, dz$$

(see (68)), we find the following representation for $\mathbb{R}_{\mathsf{W}}(A,B)$:

$$\mathbb{R}_{\mathsf{W}}(A,B) = \frac{B-A-2}{2} - \frac{Ae^{A-B}}{1-e^A} - \frac{A(e^B-e^A)(B-1+e^{-B})}{2[A(1-e^B)+B(e^A-1)]}. \tag{82}$$



V-2-17

The mean time between two false alarms equals (see (69))

$$T(A,B) = T_\infty(0) = \frac{\omega(\alpha,\beta)}{\alpha} = \frac{A(1-e^B) + B(e^A-1)}{1-e^A}. \tag{83}$$

Consider those pairs (A,B) for which

$$T(A,B) = T.$$

Among these pairs (A,B), it is natural to choose those for which $\mathbb{R}_\mathsf{W}(A,B)$ attains its minimum. In other words,

we should find

$$\boxed{\mathbb{R}_\mathsf{W}(T) = \inf \mathbb{R}_\mathsf{W}(A,B)}\,, \tag{84}$$

where infimum is taken over the pairs (A,B) for which $T(A,B) = T$.

THEOREM 1. Infimum in (84) is attained at the pair (A^*, B^*), where $A^* = 0$ and B^* solves the equation

$$T = e^{B^*} - B^* - 1. \tag{85}$$

The value $\mathbb{R}_{\mathsf{W}}(T)$ is determined by

$$\boxed{\mathbb{R}_{\mathsf{W}}(T) = \frac{1}{T}\left[B^*\left(e^{B^*} - e^{-B^*} - \frac{B^*}{2}\right) - \frac{3}{2}\left(e^{B^*} - 2 + e^{-B^*}\right)\right]}.$$

In particular,

$$\mathbb{R}_{\mathsf{W}}(T) = \begin{cases} \log T - \frac{3}{2} + O\left(T^{-1}\log^2 T\right), & T \to \infty, \\ \frac{5}{6}T + O(T^2), & T \to 0. \end{cases}$$

The proof consists in the direct analysis of (82) and (83).

REMARK 4. In our early works we refer to the cyclic-return W-method with levels (A^*, B^*), where $A^* = 0$ and B^* is determined from (85), as to

degenerated sequential analysis.

It worth noting that up to the time $\tau^*(B^*) = \inf\{t \geq 0\colon Z_t^* = B^*\}$ the process (Z_t^*) coincides with the CUSUM-process

$$\gamma_t = Z_t - \min_{s \leq t} Z_s,$$

which is known from the papers by Page*. In general, the cyclic-return process $(Z_t^*)_{t\geq 0}$ coincides with a cyclic-return CUSUM-process $(\gamma_t^*)_{t\geq 0}$, which is obtained from $(\gamma_t)_{t\geq 0}$ by shifting it to 0 each time that it reaches the boundary B^*. These considerations explain why—along with $\mathbb{R}_{\mathrm{W}}(T)$—one sometimes use the notation $\mathbb{R}_{\mathrm{CUSUM}}(T)$.

---

* E.S.Page, **Continuous inspection schemes**, Biometrika, **41** (1954), 100–114; **Control charts with warning lines**, Biometrika, **42** (1955), 243–257.

REMARK 5.

In the literature in English, for the quantities $T_\infty(z)$, $T_0(z)$, and $\mathbb{R}_\mathsf{W}(A,B)$ introduced above one often use also the following notation:

$$T_\infty(z) = \mathsf{ARL}_\infty \text{ — Average in-control Run-Length;}$$

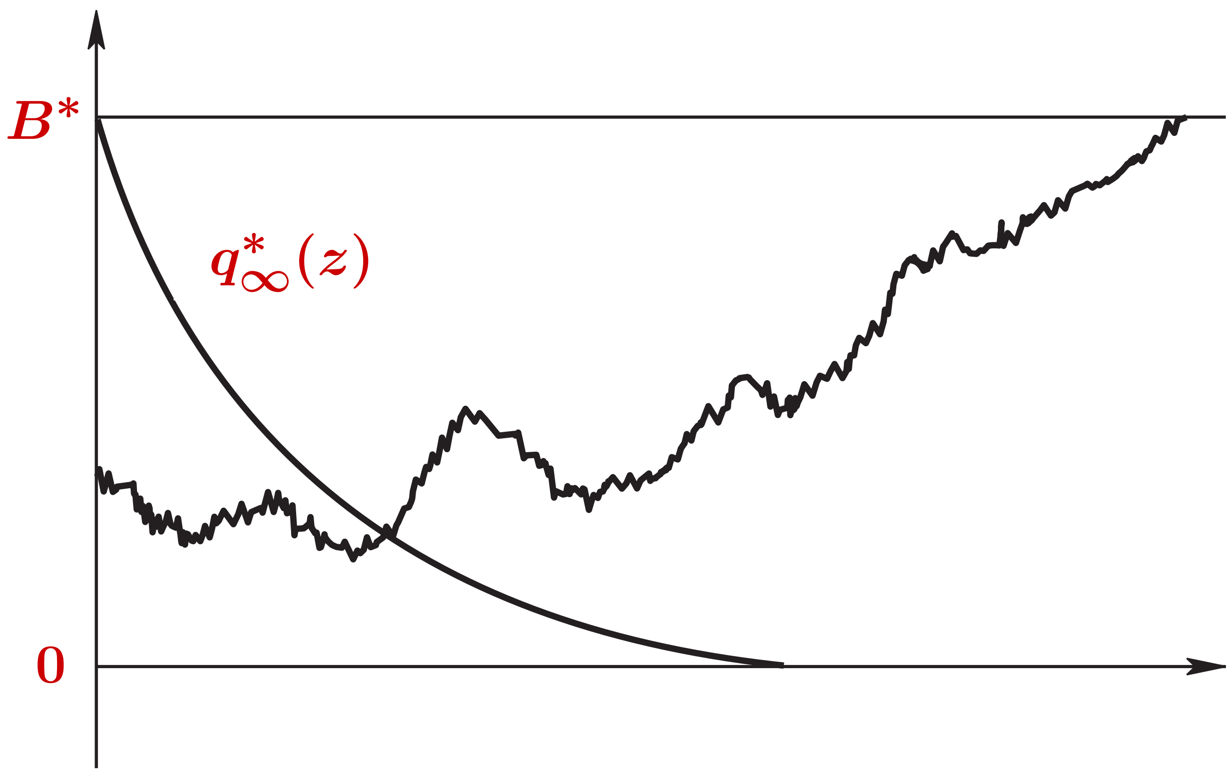
$$T_0(z) = \mathsf{ARL}_0 \text{ — Average out-of-control Run-Length;}$$

$$\mathbb{R}_\mathsf{W}(A,B) = \mathsf{SADT}(A,B) \text{ — Stationary Average Delay Time.}$$

REMARK 6.

For $A^* = 0$ and B^*determined from (85), the stationary density is given by

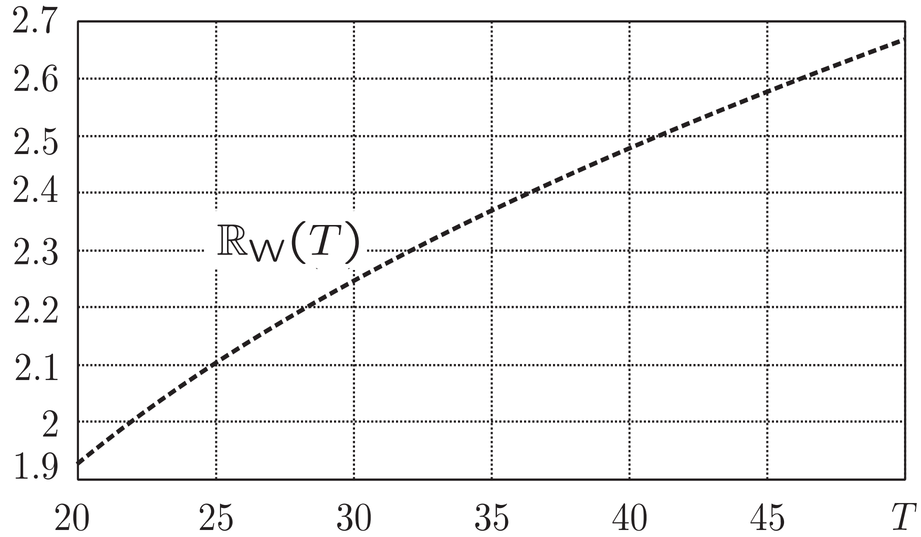
$$q^*_\infty(z) = \frac{1}{T}(e^{B^*-z} - 1).$$



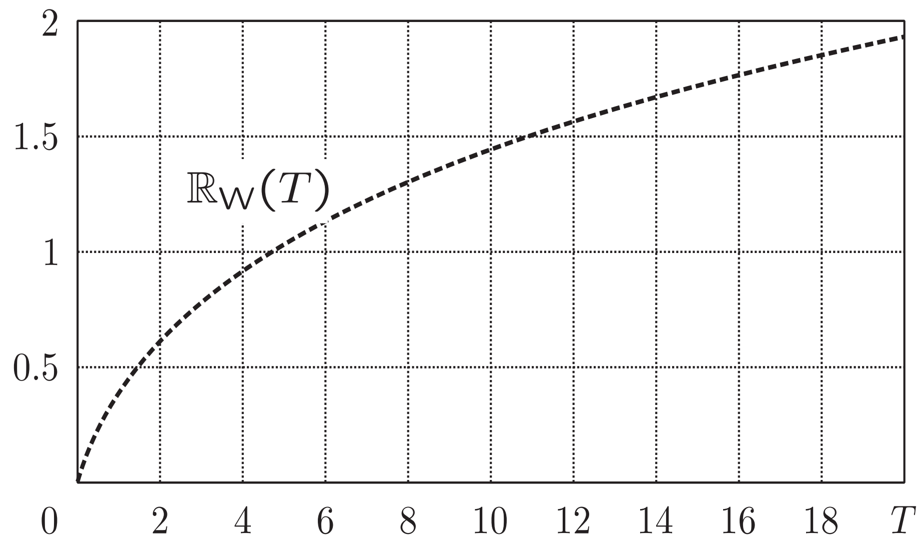
V-2-22

REMARK 7.
The following graphs give an idea of behavior of the function $\mathbb{R}_{\mathsf{W}}(T)$

for **large** T:



and for **small** T:



§ 3. Multistage cyclic-return detection NP-system

3.1. According to the Neyman–Pearson method, to distinguish between two hypotheses $H_0\colon \theta = 0$ and $H_\infty\colon \theta = \infty$ one should choose two numbers $m > 0$ and h and if $Z_m \geq h$, then one accepts the hypothesis H_0. But if $Z_m < h$, then one gives preference to the hypothesis H_∞.

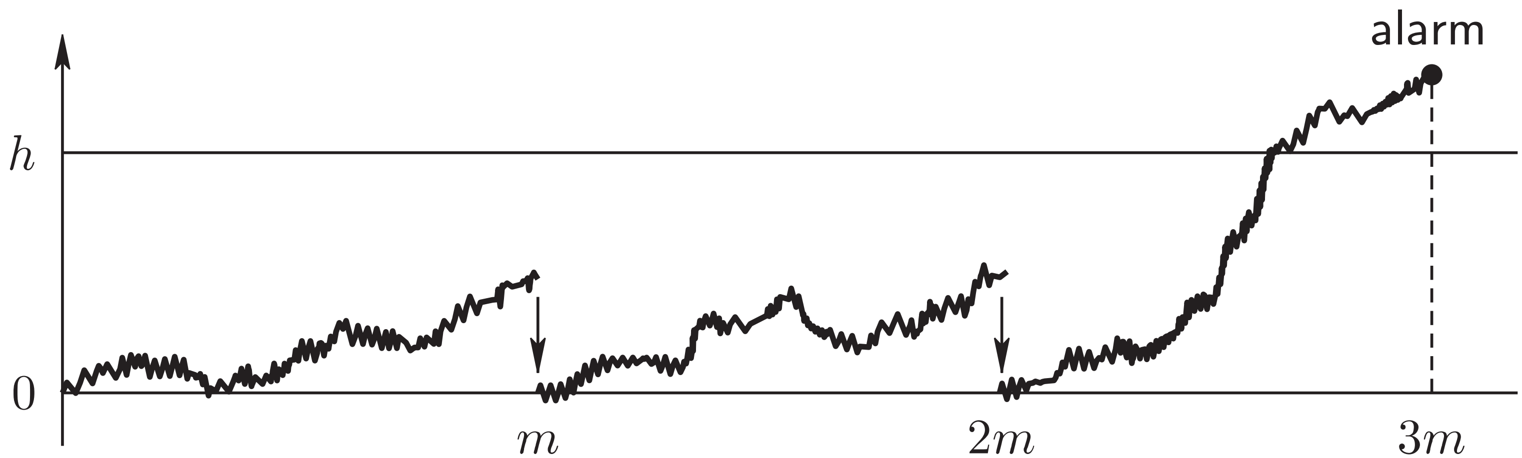
We use these element of the Neyman–Pearson test adapting them to the multistage cyclic-return detection problem in the following way.

We start observations with $Z_0 = 0$. If $Z_m < h$, then we decide that there is no disorder.

At the next step we observe the process $(\widetilde{Z}_t)_{m \le t < 2m}$ with $\widetilde{Z}_t = Z_t - Z_m$. If $\widetilde{Z}_{2m} < h$, we decide again that there is no disorder.

If at a certain step (on the figure this is the third step) $\widetilde{Z}_{3m} \ge h$, then we decide that the disorder have appeared.

After that the observations do not stop but start over again.



V-3-2

Consider the probabilities of errors of the first and second kind:

$$\alpha = \mathsf{P}(Z_m \geq h \,|\, H_\infty) \quad (= \mathsf{P}_\infty(Z_m \geq h)),$$
$$\beta = \mathsf{P}(Z_m < h \,|\, H_0) \quad (= \mathsf{P}_0(Z_m < h)).$$

Since Z_m are Gaussian, it is easy to find that

$$m = \frac{1}{2}(C_\alpha + C_\beta)^2, \qquad h = \frac{1}{2}(C_\alpha^2 - C_\beta^2),$$

where C_α and C_β solve

$$\Phi(C_\alpha) = 1 - \alpha \quad \text{and} \quad \Phi(C_\beta) = 1 - \beta$$

with $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-t^2/2}\, dt$.

Suppose that the disorder does not appear during the whole period of observation (i.e., $\theta = \infty$). Denote by T_∞ the mean time till the first false alarm. The signal of alarm can come at times km, $k \geq 1$, with probabilities $\alpha(1-\alpha)^{k-1}$.

Therefore

$$T_\infty = m/\alpha.$$

Recall that in the W-method the analogous quantity $T_\infty(0)$ is given by (69):

$$T_\infty(0) = \omega(\alpha, \beta)/\alpha.$$

To find the mean time of delay of disorder detection one should make some suggestions about the character of appearance of a disorder. In W-method we have assumed that the disorder occurs against the background of established stationary regime which is preceded by a long-lasting observation process.

3.2. If we take the Bayesian point of view, then we could assume that a priori distribution $\mathsf{P}(\theta \le t)$ for θ is the exponential distribution $\mathsf{P}(\theta \le t) = 1 - e^{-\lambda t}$ with $\mathsf{E}\theta = 1/\lambda$. The assumption in the W-method that a disorder appears at "infinity" (when the process enters into the stationary regime) can be interpreted as the condition $\mathsf{E}\theta \to \infty$, which is equivalent to $\lambda \to 0$.

For the exponential distribution $\mathbb{P}^{(\lambda)}$ (with parameter λ) we find that

$$\lim_{\lambda \downarrow 0} \mathbb{P}^{(\lambda)}\Big(\theta \in (a,b) \,|\, \theta \in (A,B)\Big) = \frac{b-a}{B-A}, \tag{86}$$

if $A < a < b < B$.

In other words, the limit (as $\lambda \to 0$) of the exponential distribution $\mathbb{P}^{(\lambda)}$ is conditionally uniform (in a sense that (86) is fulfilled).

From these considerations we assume that the probability distribution of appearance of disorder is **uniform** on each interval $(am, (a+1)m)$.

Therefore the mean time of delay $\mathbb{R}_{\mathsf{NP}}(m,h)$ is

$$\boxed{\mathbb{R}_{\mathsf{NP}}(m,h) = \frac{1}{m}\int_0^m \mathbf{R}_t\, dt}, \tag{87}$$

where $\mathbf{R}_t$ is the mean time up to detection of disorder provided that the disorder appeared at a time $t \in [0,m]$.

THEOREM 2. The mean time of delay $\mathbb{R}_{\mathsf{NP}}(m,h)$ is given by

$$\mathbb{R}_{\mathsf{NP}}(m,h) = \frac{m}{2} + \frac{m}{1-\beta}\left\{1 - \frac{1}{\sqrt{2m}}\int_{C_{1-\beta}}^{C_\alpha} (1-\Phi(x))\, dx\right\}, \tag{89}$$

where $C_\alpha = (h+m)/\sqrt{2m}$, $C_{1-\beta} = (h-m)/\sqrt{2m}$.

Proof of Theorem 2. It is clear that

$$\mathbf{R}_t = \mathsf{E}_\infty R_t(z) = \int R_t(z) p_\infty(0,0;t,z)\,dz, \tag{90}$$

where $R_t(z)$ is the mean time of detection of a disorder, which have appeared at time $t \in [0,m]$, when $Z_t = z$, and

$$p_\infty(0,0;t,z) = \frac{1}{\sqrt{2\pi t\sigma}} \exp\Big\{-\frac{(z+t)^2}{2\sigma^2 t}\Big\}$$

is the transition density of the process $(Z_s)_{s\geq 0}$ with $dZ_s = -ds + \sqrt{2}\,dB_s$.

We have

$$\begin{aligned} R_t(z) = (m-t)\mathsf{P}_t(Z_m \geq h \,|\, Z_t = z) \\ + [(m-t) + R_0(0)]\mathsf{P}_t(Z_m < h \,|\, Z_t = z), \end{aligned} \tag{91}$$

where $\mathsf{P}_t(\cdot)$ is the probability distribution provided that a disorder have appeared at time t.

From (91) we find that

$$R_t(z) = (m-t) + R_0(0)\mathsf{P}_t(Z_m < h \,|\, Z_t = z),$$

Thus, $R_0(0) = m/(1-\beta)$ and

$$\begin{aligned} R_t(z) &= (m-t) + \frac{m}{1-\beta}\mathsf{P}_t(Z_m < h \,|\, Z_t = z) \\ &= (m-t) + \frac{m}{1-\beta}\int_0^h p_0(0,z;m-t,y)\,dy, \end{aligned} \qquad (92)$$

where $p_0(0,z;m-t,y)$ is the density of the probability that the process $(Z_s)_{s\geq 0}$ with disorder, starting at time $s=0$ from the point y, will find itself at time $m-t$ in a point z. It is clear that

$$p_0(0,z;m-t,y) = \frac{1}{\sqrt{2\pi(m-t)\sigma}}\exp\left\{-\frac{[y-(z+m-t)]^2}{2\sigma^*(m-t)}\right\}, \qquad (93)$$

where $\sigma = \sqrt{2}$. Using (92), (93), (90), and (87), we get the required formula (89) for $\mathbb{R}_{\mathsf{NP}}(m,h)$. $\square$

Taking into account the formula $T_\infty = m/\alpha$, where $m = (C_\alpha + C_\beta)^2/2 = (C_\alpha - C_{1-\beta})^2/2$, we come, after straighforward transformations, to the following formula for $\mathbb{R}_{\mathsf{NP}}(m,h)$:

$$\mathbb{R}_{\mathsf{NP}}(m,h) = \frac{T_\infty \Psi(x)}{2}\Bigg\{1 + \frac{2}{\Psi(y)}\Bigg[1 - \frac{1}{\sqrt{2T_\infty\Psi(x)}} \times \Big(x\Psi(x) - y\Psi(y) - \varphi(x) + \varphi(y)\Big)\Bigg]\Bigg\},$$

where $x = C_\alpha$, $y = C_{1-\beta}$, $\Psi(x) = 1 - \Phi(x)$. This formula turns out to be useful for studying the asymptotic behavior of

$$\mathbb{R}_{\mathsf{NP}}(T) = \inf \mathbb{R}_{\mathsf{NP}}(m,h),$$

where infimum is taken over all pairs (m,h) for which $T_\infty = T$.

THEOREM 3. For $\mathbb{R}_{\mathsf{NP}}(T)$ the following asymptotics hold:

$$\boxed{\mathbb{R}_{\mathsf{NP}}(T) \sim \begin{cases} \frac{3}{2}\log T, & T \to \infty, \\ \frac{T}{2}, & T \to 0. \end{cases}}$$

For large T the pairs (m, h) and (α, β) have the following asymptotics:

$$m \sim \log T, \qquad h \sim \log T$$

and

$$\alpha \sim \frac{\log T}{T}, \qquad \beta \sim \frac{1}{\sqrt{2\log T \cdot \log\log T}}.$$

Comparing the formulae

$$\mathbb{R}_{\mathsf{W}}(T) \sim \log T, \qquad \mathbb{R}_{\mathsf{NP}}(T) \sim \frac{3}{2} \log T, \qquad T \to \infty, \tag{94}$$

and

$$\mathbb{R}_{\mathsf{W}}(T) \sim \frac{5}{6} T, \qquad \mathbb{R}_{\mathsf{NP}}(T) \sim \frac{T}{2}, \qquad T \to 0, \tag{95}$$

shows that for large T the W-method is "3/2 times" more effective than the NP-method. But for small T the NP-method is "5/6 times" more effective than the W-method.

§ 4. Cyclic-return systems of observation over N directions

The quantitative analysis of the W- and NP-systems of detection given in § 2 and § 3 was the **FIRST stage** of our project "The quickest detection problems". The main results, obtained in 1959–60, were welcomed by the concerned experts, who were particularly interested in the question whether the **logarithmic** asymptotic

$$\log T, \qquad T \to \infty,$$

for the delay time, obtained for the W-method, is unimprovable.

Furthermore, the question arose as whether the asymptotics

$$\frac{T}{2}, \qquad T \to 0,$$

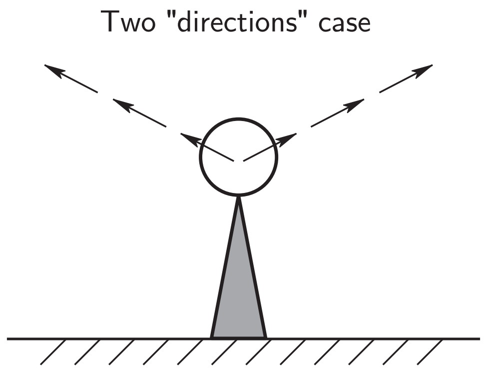
found for the Neyman–Pearson method, is also unimprovable.

All this suggested that we ought to go in for finding the **optimal** method.

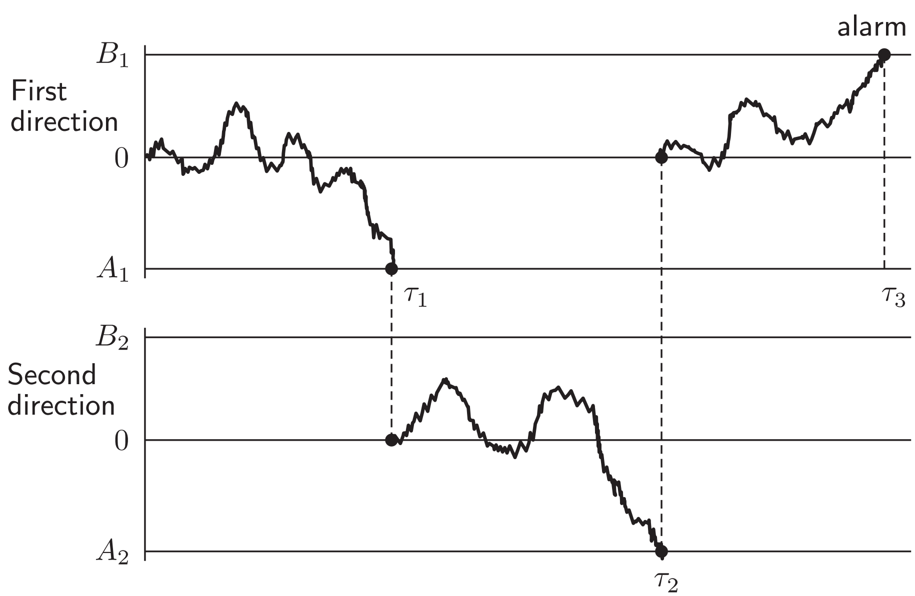
This became a subject of the **SECOND stage** of our project, but at the first stage we considered also a more general scheme of observations, which will be referred to as "cyclic-return system of observations over N directions". The core of this **practically** important scheme can be described as follows.

We have one radio-locator, which is able to inspect in a cyclic way one of N directions. It is assumed that the target—if it appears—arises equiprobably in one of N directions. The probabilistic-statistical characteristics of all these N directions are identical.

W-system, adapted to the case of N directions, can be illustrated in the following way ($N = 2$).



Let the observation start at the 1st direction. If the process exits from the interval (A, B) for the first time through the lower level A, then we decide that there is no target in the 1st direction, and the radio-locator switches over to observation in the 2nd direction, and so on in a cyclic way. The system signalizes the appearance of a target if, in one of the directions, the process reaches the upper level.



Let $T = T(A,B)$ be the mean time between two false alarms. For all N this time is the same as for $N = 1$:

$$T(A,B) = \frac{A(1-e^B) + B(e^A - 1)}{1 - e^A}.$$

The mean delay time $\mathbb{R}_{\mathsf{W}}^{(N)}(A,B)$ can be calculated in the same way as in the case $N = 1$, under assumption of established stationary regime.

Let $\mathbb{R}_{\mathsf{W}}^{(N)}(T) = \inf \mathbb{R}_{\mathsf{W}}^{(N)}(A,B)$, where infimum is taken over all A, B such that $T(A,B) = T$. It turned out that

$$\mathbb{R}_{\mathsf{W}}^{(N)}(T) = T\Big[N - \frac{5}{3} + \frac{3}{2N}\Big] + O(T^2), \qquad T \to 0, \tag{96}$$

and

$$\mathbb{R}_{\mathsf{W}}^{(N)}(T) = \log T + \Big(N - \frac{1}{N}\Big) - \frac{3}{2} + O\Big(\frac{N \log^2 T}{T}\Big), \qquad T \to \infty. \tag{97}$$

At that for $N > 1$ the optimal lower level $A^*(T, N)$ is not anymore zero, as for $N = 1$, and $A^*(T, N) \to 0$ as $T \to \infty$ which means the swift passing from one direction to another (заведомо when there is no target).

The NP-system can be evidently adapted for the case of N directions.

It turned out [see **(Shiryaev, 1963-II)**] that the corresponding time of delay is of the form

$$\mathbb{R}_{\mathsf{NP}}^{(N)}(T) \sim \begin{cases} \frac{1}{2}(N+2)\log T, & T \to \infty, \\ \frac{1}{2}NT, & T \to 0. \end{cases} \tag{98}$$

Comparing $\mathbb{R}_{\mathsf{W}}^{(N)}(T)$ with $\mathbb{R}_{\mathsf{NP}}^{(N)}(T)$ shows that for large T the W-method is more effective than the NP-method. For small T, to the contrary, the NP-method is more effective than W-method.

§ 5. Setup of the problem of optimal detection methods

5.1. The above results on behavior of mean time of delay, $\mathbb{R}_{\mathsf{W}}(T)$ and $\mathbb{R}_{\mathsf{NP}}(T)$, for large and small T concluded, as we already mentioned, the **first stage** of our work on quickest detection of randomly appearing targets.

At the **second stage** the main efforts were paid to the problem of formulation of optimization problem of quickest detection of a target (disorder) under the following assumptions:

(a) the appearance of the target is preceded by a long period of observation;

(b) a stationary regime of observation gets established (in the absence of a target);

(c) the target appears against a background of the established stationary regime.

Notation:

$\mathbb{R}(T)$ is the mean time of delay in the detection of the target under assumption that the mean time between two false alarms equals T (cf. $\mathbb{R}_{\mathsf{W}}(T)$ and $\mathbb{R}_{\mathsf{NP}}(T)$).

The key result given in two works "The problem of the most rapid detection of a disturbance of a stationary regime" **(Shiryaev, 1961-5)** and "On optimal methods in quickest detection problems" **(Shiryaev, 1963)**, is summarized in the following theorem.

(It is assumed that $\rho \equiv r^2/(2\sigma^2) = 1$.)

THEOREM 4.

1) For the optimal (multistage cyclic-return) method of observations the mean delay time is determined by

$$\boxed{\mathbb{R}(T) = e^b(-\,\mathrm{Ei}(-b)) - 1 + b\int_0^\infty e^{-bx}\,\frac{\log(1+x)}{x}\,dx}\,, \tag{99}$$

where $b = 1/T$ and $-\,\mathrm{Ei}(-b) = \int_0^\infty t^{-1}e^{-t}\,dt$.

In particular,

$$\boxed{\mathbb{R}(T) = \begin{cases} \frac{T}{2} + O(T^2), & T \to 0, \\ \log T - (1 + \mathbf{C}) + O(T^{-1}), & T \to \infty, \end{cases}} \tag{100}$$

where $\mathbf{C}$ is the Euler constant ($\mathbf{C} = 0.577\ldots$).

2) The optimal method is based on observations over the process $\psi = (\psi_t)_{t\ge 0}$ solving the stochastic differential equation

$$d\psi_t = dt + \sqrt{2}\psi_t \, dX_t.$$

Observations start with the value $\psi_0 = 0$. When the process $\psi = (\psi_t)_{t\ge 0}$ reaches the threshold T, it falls to zero, and the observations over this process $\psi = (\psi_t)_{t\ge 0}$ starts over again from zero. (If the initial process $X = (X_t)_{t\ge 0}$ is such that $dX_t = rI(t > \theta)\, dt + \sigma \, dB_t$, then the corresponding process $\psi = (\psi_t)_{t\ge 0}$ has the stochastic differential $d\psi_t = dt + r\sigma^{-2}\psi_t \, dX_t$.)

Let us trace the main steps of the proof of the theorem.

First of all we should determine what we mean under a "multistage" method of observations. Each method of observations of such type, say δ, is identified with a sequence $(\tau_1, \tau_2, \ldots)$ of stopping times with respect to the flow $(\mathcal{F}_t^B)_{t \geq 0}$ of σ-algebras $\mathcal{F}_t^B = \sigma(B_s, s \leq t)$, which characterize the duration of observations on the first stage, second stage, and so on. The times $\tau_1, \tau_2, \ldots$ are assumed independent. Under assumption that there is no disorder they have the same distribution. Denote by

$$T = \mathsf{E}_\infty \tau_1$$

the expectation of the time till the first false alarm.

We use the value T to guarantee that the requirement (A) in § 1, which demands that false alarms should occur "rarely", is fulfilled.

5.2. Let us discuss how to understand what is 'mean time of delay' under the assumption that the process have already entered the stationary regime. Assume that the disorder appears at time t. (In what follows we let $t \to \infty$.) We shall use the notation

$$\varkappa_t = \max\{k\colon \tau_1 + \cdots + \tau_k < t\}.$$

Then we shall understand the time of delay (when using the method δ) as the following quantity:

$$R_t^\delta = \mathsf{E}_t\Big[\sum_{i=1}^{\varkappa(t)+1} \tau_i - t\Big], \tag{101}$$

where $\mathsf{E}_t(\cdot)$ is averaging w.r.t. the measure P_t which corresponds to the case that the disorder appears at time t.

Let

$$\mathsf{P}_t^\delta(u) = \mathsf{P}_\infty\Big\{t - \sum_{i=1}^{\varkappa(t)} \tau_i \le u\Big\} \tag{102}$$

be the probability distribution of duration of time interval between t and $\sum_{i=1}^{\varkappa(t)} \tau_i$. Then

$$\begin{aligned} R_t^\delta &= \int_0^t \mathsf{E}_t\Big[\sum_{i=1}^{\varkappa(t)+1} \tau_i - t \,\Big|\, t - \sum_{i=1}^{\varkappa(t)} \tau_i = u\Big]\,\mathsf{P}_t^\delta(du) \\ &= \int_0^t \mathsf{E}_t\Big[\tau_{\varkappa(t)+1} - u \,\Big|\, \sum_{i=1}^{\varkappa(t)} \tau_i = t-u, \tau_{\varkappa(t)+1} > u\Big]\,\mathsf{P}_t^\delta(du) \\ &= \int_0^t \mathsf{E}_u(\tau - u \,|\, \tau > u)\,\mathsf{P}_t^\delta(du), \end{aligned} \tag{103}$$

where for simplicity of writing we denoted τ_1 by τ.

If we restrict ourselves to the detection methods for which the distribution $F_\infty(x) = \mathsf{P}_\infty(\tau \le x)$ is non-lattice, then, by the well-known renewal theorem, we find that

$$\lim_{t\to\infty} \mathsf{P}_t^\delta(u) = \lim_{t\to\infty} \mathsf{P}_\infty\Big\{t - \sum_{i=1}^{\varkappa(t)} \tau_i \le u\Big\} = \int_0^u (1 - F_\infty(x))\,\frac{dx}{\mathsf{E}_\infty \tau}. \quad (104)$$

Therefore (103) implies that if $\mathsf{E}_\infty \tau = T$, then

$$\begin{aligned} \lim_{t\to\infty} R_t^\delta &= \int_0^\infty \mathsf{E}_u(\tau - u \,|\, \tau > u)\,\frac{\mathsf{P}_\infty\{\tau > u\}}{T}\,du \\ &= \int_0^\infty \mathsf{E}_u(\tau - u \,|\, \tau > u)\,\frac{\mathsf{P}_u\{\tau > u\}}{T}\,du \\ &= \frac{1}{T}\int_0^\infty \mathsf{E}_u(\tau - u)^+\,du. \end{aligned} \quad (105)$$

For the δ-method, the quantity

$$R^{\delta}(T) = \lim_{t\to\infty} R_t^{\delta} \tag{106}$$

is called its mean time of delay in detection of disorder, which occurs against a background of established stationary regime.

From (105) we see that, in the scheme of multistage observations of the process

$X = (X_t)_{t\geq 0}$, where $dX_t = \sigma\, dB_t$ if there is no disorder, and where $dX_t = r\, dt + \sigma\, dB_t$ if the disorder have been occurred,

finding the optimal detection method reduces to the following variant of the optimal stopping problem.

VARIANT B. To find, in the class $\mathfrak{M}_T = \{\tau \colon \mathsf{E}_\infty \tau = T\}$, a time τ_T^* such that

$$\inf_{\tau \in \mathfrak{M}_T} \int_0^\infty \mathsf{E}_u(\tau - u)^+ \, du = \int_0^\infty \mathsf{E}_u(\tau_T^* - u)^+ \, du. \qquad (108)$$

Variant B is also called a **generalized Bayesian problem**.

This terminology will be clear from Variant A which we will consider later and which is a **Bayesian problem** of optimal stopping.

Now we will describe **TWO METHODS** of solving problem (108):

$$\boxed{\inf_{\tau\in\mathfrak{M}_T}\int_0^\infty \mathsf{E}_u(\tau-u)^+\,du = \int_0^\infty \mathsf{E}_u(\tau_T^*-u)^+\,du}\,.$$

The **first** one (**§ 6**), developed in our works of 1959–60, is based on considering a (one-stage) Bayesian quickest detection problem—which is constructed in a special way—by passing to the limit [as $\lambda \to 0$ in exponential distribution for the time of appearing of the target (disorder)].

The **second**—direct—method will be discussed in **§ 7**.

§ 6. BAYESIAN FORMULATION OF THE QUICKEST DETECTION PROBLEM

6.1. Let $\theta = \theta(\omega)$ be a random variable with exponential distribution

$$\mathsf{P}(\theta > t) = e^{-\lambda t}. \tag{109}$$

We consider the model

$$dX_t = rI(t \geq \theta)\,dt + \sigma\,dB_t, \tag{110}$$

where the time θ and the Brownian motion $B = (B_t)_{t\geq 0}$ are assumed to be independent.

Notation:

- $\mathsf{P}_u(\cdot)$ is the probability distribution of the process
  $X = (X_t)$ with $dX_t = rI(t \geq u)\,dt + \sigma\,dB_t$;
- $\mathbb{P}^{(\lambda)}(\cdot) = \lambda \int_0^\infty e^{-\lambda u}\,\mathsf{P}_u(\cdot)\,du$;
- $\mathbb{E}^{(\lambda)}$ and E_u are the averaging w.r.t. $\mathbb{P}^{(\lambda)}$ and P_u.

We find that

$$\mathbb{E}^{(\lambda)}(\tau-\theta)^+ = \lambda \int_0^\infty e^{-\lambda u}\,\mathsf{E}_u(\tau-u)^+\,du. \tag{111}$$

Consider the quantities $\varkappa(t) = \max\{k\colon\ \tau_1+\cdots+\tau_k < t\}$ introduced above. Let $T = \mathsf{E}_\infty\tau_1$ and $\alpha = \mathbb{P}^{(\lambda)}(\tau_1 \le \theta)$, where $\theta = \theta(\omega)$ has exponential distribution ($\mathsf{P}(\theta > t) = e^{-\lambda t}$).

As $\lambda \to 0$

$$T\mathbb{E}^{(\lambda)}\varkappa(\theta) \sim \mathbb{E}^{(\lambda)}\theta \quad (=\lambda^{-1}).$$

Since $\theta = \theta(\omega)$ is exponentially distributed, we have the equality $\mathbb{E}^{(\lambda)}\varkappa(\theta) = \alpha/(1-\alpha)$. Thus, $\alpha = \mathbb{P}^{(\lambda)}(\tau_1 \le \theta) \to 1$ as $\lambda \to 0$ and

$$\frac{1-\alpha}{\alpha\lambda} \sim T, \qquad \lambda \to 0.$$

Let us turn to representation (105):

$$\lim_{t\to\infty} R_t^{\delta} = \frac{1}{T}\int_0^{\infty} \mathsf{E}_u(\tau_1 - u)^+ \, du.$$

Let us transform the right-hand side:

$$\begin{aligned}
\frac{1}{T}\int_0^{\infty} \mathsf{E}_u(\tau_1 - u)^+ \, du &= \lim_{\lambda\to 0,\alpha\to 1:\ \frac{1-\alpha}{\alpha\lambda}=T} \frac{\int_0^{\infty} e^{-\lambda u}\mathsf{E}_u(\tau_1 - u)^+ \, du}{(1-\alpha)/(\alpha\lambda)} \\
&= \lim_{\lambda\to 0,\alpha\to 1:\ \frac{1-\alpha}{\alpha\lambda}=T} \frac{\lambda\int_0^{\infty} e^{-\lambda u}\mathsf{E}_u(\tau_1 - u)^+ \, du}{1-\alpha} \\
&= \lim_{\lambda\to 0,\alpha\to 1:\ \frac{1-\alpha}{\alpha\lambda}=T} \frac{\mathbb{E}^{(\lambda)}(\tau_1 - \theta)^+}{1-\alpha} \\
&= \lim_{\lambda\to 0,\alpha\to 1:\ \frac{1-\alpha}{\alpha\lambda}=T} \mathbb{E}^{(\lambda)}(\tau_1 - \theta \,|\, \tau_1 \geq \theta). \qquad (112)
\end{aligned}$$

Thus, the problem of finding

$$\inf \frac{1}{T}\int_0^{\infty} \mathsf{E}_u(\tau_1 - u)^+\,du,$$

where infimum is taken over the times τ_1 such that $\mathsf{E}_\infty\tau_1 = T$, is closely related with finding of conditional times of delay $\mathbb{E}^{(\lambda)}(\tau_1 - \theta \,|\, \tau_1 \ge \theta)$ in the following Bayesian problem of quickest detection.

VARIANT A. Let a random variable θ be exponentially distributed with mass at zero:

$$\mathsf{P}(\theta = 0) = \pi, \quad \mathsf{P}(\theta > t \mid \theta > 0) = e^{-\lambda t}. \tag{113}$$

Let the observed process $X = (X_t)_{t \ge 0}$ have the differential

$$dX_t = \begin{cases} \sigma\, dB_t, & t < \theta, \\ r\, dt + \sigma\, dB_t, & t \ge \theta. \end{cases} \tag{114}$$

Bayesian problem: To find the optimal time $\tau^* = \tau^*(c, \lambda)$ such that

$$\inf_{\tau} [\mathsf{P}(\tau \le \theta) + c\mathsf{E}(\tau - \theta)^+] = \mathsf{P}(\tau^* \le \theta) + c\mathsf{E}(\tau^* - \theta)^+. \tag{115}$$

Conditionally extremal problem: Given $\alpha \in (0, 1)$, to find in the class $\mathfrak{M}_\alpha = \{\tau \colon \mathsf{P}(\tau \le \theta) \le \alpha\}$ the time τ^*_α such that

$$\inf_{\tau \in \mathfrak{M}_\alpha} \mathsf{E}(\tau - \theta \mid \tau \ge \theta) = \mathsf{E}(\tau^*_\alpha - \theta \mid \tau^*_\alpha \ge \theta). \tag{116}$$

The Bayesian problem (115) was solved in the following way.

(a) We first show that for any τ with $\mathsf{E}\tau < \infty$

$$\mathsf{P}(\tau \le \theta) + c\mathsf{E}(\tau - \theta)^+ = \mathsf{E}\biggl[(1 - \pi_\tau) + c \int_0^\tau \pi_t \, dt\biggr], \tag{117}$$

where

$$\pi_t = \mathsf{P}(\theta \le t \,|\, \mathcal{F}_t^X)$$

is the a posteriori probability of appearing of disorder on the time interval $[0, t]$ under assumption that the observed data are $(X_s, s \le t)$, $\pi_0 = \pi$. (As usual, $\mathcal{F}_t^X = \sigma(X_s, s \le t)$.)

(b) We derive, for process $(\pi_t)_{t \ge 0}$, the following stochastic differential equation:

$$d\pi_t = \Bigl(\lambda - \frac{r^2}{\sigma^2}\pi_t^2\Bigr)(1 - \pi_t)\, dt + \frac{r}{\sigma^2}\pi_t(1 - \pi_t)\, dX_t. \tag{118}$$

(c) We show that the process $X = (X_t)_{t\ge 0}$ with

$$X_t = r\int_0^t \theta_s\,ds + \sigma B_t, \qquad \theta_s = I(s \ge \theta),$$

admits the "innovation" representation

$$X_t = r\int_0^t \pi_s\,ds + \sigma \overline{B}_t, \quad \text{i.e.,} \quad dX_t = r\pi_t\,dt + \sigma\,d\overline{B}_t, \tag{119}$$

where

$$\overline{B}_t = B_t + \frac{r}{\sigma}\int_0^t (\theta_s - \pi_s)\,ds$$

is a **Brownian motion** with respect to the filtration $(\mathcal{F}_t^X)_{t\ge 0}$.
From (118) and (119) it follows that

$$d\pi_t = \lambda(1-\pi_t)\,dt + \frac{r}{\sigma}\pi_t(1-\pi_t)\,d\overline{B}_t. \tag{120}$$

Consequently, the process $(\pi_t, \mathcal{F}_t^X)_{t\ge 0}$ is a **diffusion Markov process**.

(d) The process $\varphi_t = \pi_t/(1-\pi_t)$ admits the representation

$$\varphi_t = \varphi_0 e^{\lambda t} L_t + \lambda e^{\lambda t} \int_0^t e^{-\lambda s} \frac{L_t}{L_s}\, ds, \tag{121}$$

where

$L_t = \dfrac{d\mathsf{P}_0}{d\mathsf{P}_\infty}(t, X)$ is the Radon–Nykodým derivative of the measure $\mathsf{P}_0(\cdot) = \mathsf{Law}(X_s, s \le t \,|\, \theta = 0)$ w.r.t. the measure $\mathsf{P}_\infty(\cdot) = \mathsf{Law}(X_s, s \le t \,|\, \theta = \infty)$.

For L_t we have the representation

$$L_t = e^{H_t}, \quad \text{where} \quad H_t = \frac{r}{\sigma^2} X_t - \frac{r^2}{2\sigma^2} t. \tag{122}$$

By the Itô formula,

$$dL_t = \frac{r}{\sigma^2} L_t\, dX_t. \tag{123}$$

The simplest way to derive equation (118) for π_t is to get first the equation for $\varphi_t = \dfrac{\pi_t}{1-\pi_t}$ and then to derive from it (by the Itô formula) the equation for $\pi_t = \dfrac{\varphi_t}{1+\varphi_t}$.

By the Bayes formula,

$$\begin{aligned}\pi_t = \mathsf{P}(\theta \le t \,|\, \mathcal{F}_t^X) &= \int_0^t \frac{d\mu_s}{d\mu}(t,X) p_\theta(s)\,ds \\ &= \int_0^t \frac{d\mu_s}{d\mu_t}(t,X) p_\theta(s)\,ds \cdot \frac{d\mu_t}{d\mu}(t,X), \qquad (124)\end{aligned}$$

where μ_s is the measure of the process $(X_t)_{t\ge 0}$ with

$$dX_t = rI(t \ge s)\,dt + \sigma\,dB_t,$$

μ is the measure of the process $(X_t)_{t\ge 0}$ with

$$dX_t = rI(t \ge \theta)\,dt + \sigma\,dB_t$$

and $p_\theta(s) = \lambda e^{-\lambda s}$. (Now we assume for simplicity that $\pi_0 = 0$.)

In an analogous way, we find that

$$1 - \pi_t = \mathsf{P}(\theta > t \,|\, \mathcal{F}_t^X) = \int_t^\infty \frac{d\mu_s}{d\mu}(t, X) p_\theta(s)\, ds$$

$$= \int_t^\infty \frac{d\mu_t}{d\mu}(t, X) p_\theta(s)\, ds = e^{-\lambda t} \frac{d\mu_t}{d\mu}(t, X). \tag{125}$$

From (124) and (125) it follows that

$$\varphi_t = \frac{\pi_t}{1 - \pi_t} = \lambda e^{\lambda t} \int_0^t e^{-\lambda s} \frac{d\mathsf{P}_s}{d\mathsf{P}_t}(t, X)\, ds = \lambda e^{\lambda t} \int_0^t e^{-\lambda s} \frac{L_t}{L_s}\, ds.$$

If $\pi_0 \neq 0$, then we deduce in an analogous way that

$$\varphi_t = \varphi_0 e^{\lambda t} L_t + \lambda e^{\lambda t} \int_0^t e^{-\lambda s} \frac{L_t}{L_s}\, ds, \tag{126}$$

which yields

$$d\varphi_t = \lambda(1 + \varphi_t)\, dt + \frac{r}{\sigma^2} \varphi_t\, dX_t$$

with $\varphi_0 = \frac{\pi_0}{1 - \pi_0}$.

(e) If $\lambda \to 0$ and $\pi_0 \to 0$ in such a way that $\pi_0/\lambda \to m$, then (126) implies that $\psi_t = \lim_{\lambda \downarrow 0}(\varphi_t/\lambda)$ obeys the relation

$$\psi_t = mL_t + \int_0^t \frac{L_t}{L_s}\, ds, \tag{127}$$

which implies that

$$d\psi_t = dt + \frac{r}{\sigma^2}\psi_t\, dX_t, \qquad \psi_0 = m. \tag{128}$$

This process, introduced in 1959–60 in **(Shiryaev, 1961-5, 1963)**, plays a key role in solving the problem in Variant B. We shall discuss it a little bit late. Now we continue the description of the next steps of solving the Bayesian problem in Variant A.

6.2. (f) Let

$$V^*(\pi) = \inf_\tau \mathsf{E}_\pi \biggl((1-\pi_\tau) + c \int_0^\tau \pi_t \, dt \biggr), \tag{129}$$

where the upper index π of expectation means that $\pi_0 = \pi$. With respect to the corresponding measure P_π, the process $(\pi_t)_{t \ge 0}$ with $d\pi_t = \lambda(1-\pi_t)\, dt + \pi_t(1-\pi_t)\, dB_t$ is a diffusion process with infinitesimal operator

$$\mathcal{A} = a(\pi) \frac{d}{d\pi} + \frac{1}{2} b^2(\pi) \frac{d^2}{d\pi^2}, \tag{130}$$

where the local drift $a(\pi)$ local variance $b^2(\pi)$ are determined by

$$a(\pi) = \lambda(1-\pi), \tag{131}$$

$$b^2(\pi) = \Bigl(\frac{r}{\sigma}\Bigr)^2 \Bigl(\pi(1-\pi)\Bigr)^2, \tag{132}$$

respectively. To find the function $V(\pi)$ we reduced the optimal stopping problem (129) to a Stefan problem (free-boundary problem).

To this end let us consider two domain

$$C^* = \{\pi\colon V^*(\pi) < 1-\pi\} \quad \text{and} \quad D^* = \{\pi\colon V^*(\pi) \ge 1-\pi\};$$

$V_0^*(\pi) \equiv 1-\pi$ is the risk from the instantaneous stopping at state π.

Since the function $V^*(\pi)$ is concave, there exists a point A^* such that

$$C^* = \{\pi\colon \pi < A^*\} \quad \text{and} \quad D^* = \{\pi\colon \pi \ge A^*\}.$$

The Stefan problem mentioned above consists in finding a function $V = V(\pi)$, $\pi \in [0,1]$, and a boundary point A such that

$$V(\pi) = 1-\pi, \qquad \pi \ge A, \tag{133}$$

$$\mathcal{A}V(\pi) = -c\pi, \qquad \pi < A. \tag{134}$$

Equation (134) имеет вид

$$\lambda(1-\pi)V'(\pi) + \frac{1}{2}\left(\frac{r}{\sigma}\right)^2 \pi^2(1-\pi)^2 V''(\pi) = -c\pi. \tag{135}$$

The general solution of this equation contains two undetermined constants C_1 and C_2. Together with unknown constant A, we have three constants which are to be determined, so we need **THREE conditions** to determine A, C_1, and C_2.

One condition ia very natural, this is **condition** (133).

From a priori analysis of the properties of the function $V^*(\pi)$ if follows that other two conditions are the **smooth-fit condition**

$$\boxed{\frac{dV}{d\pi}\Big|_{\pi\uparrow A} = \frac{dV_0}{d\pi}\Big|_{\pi\downarrow A}}, \quad \text{i.e., the condition} \quad V'(A-) = -1, \tag{136}$$

and the condition

$$\boxed{\frac{dV}{d\pi}\Big|_{\pi\uparrow 0} = 0}, \tag{137}$$

which marks out that **solution of equation** (135) **which does not go away to** $+\infty$ **or** $-\infty$ as $\pi \uparrow 0$.

Having solved equation (135) with conditions (133), (136), and (137), we obtain that

$$\boxed{\boxed{V(\pi) = \begin{cases} (1-A) - \int_\pi^A y(x)\,dx, & \pi \in [0, A), \\ 1-\pi, & \pi \in [A, 1], \end{cases}}} \tag{138}$$

where

$$\boxed{y(x) = -C\int_0^x e^{\Lambda[G(x)-G(u)]}\,\frac{du}{u(1-u)^2}}, \qquad \boxed{(u) = \log\frac{u}{1-u} - \frac{1}{u}},$$

$$\boxed{\Lambda = \frac{\lambda}{r^2/(2\sigma^2)}}, \qquad \boxed{C = \frac{c}{r^2/(2\sigma^2)}}.$$

The constant A can be found from the equation

$$\boxed{C\int_0^A e^{\Lambda[G(A)-G(u)]}\,\frac{du}{u(1-u)^2} = 1}. \tag{139}$$

Application of the "verification theorem" allows us to prove (see **(Shiryaev, 1976)** that solutions $(V(\pi), A)$ of the Stefan problem found above give the solution $(V^*(\pi), A^*)$ of the initial optimal stopping problem (129). At that the optimal stopping time τ^* does exists and is given by

$$\tau^* = \inf\{t \geq 0 \colon \pi_t \geq A^*\}.$$

(The threshold A^* is determined as a unique root A of equation (139).)

Having the solution of the Bayesian problem we can find the solution in **conditionally extremal setup**:

To find—within the class $\mathfrak{M}_\alpha = \{\tau \colon \mathsf{P}_\pi(\tau < 0) \leq \alpha\}$—a time τ^*_α such that

$$\boxed{\inf_{\tau \in \mathfrak{M}_\alpha} \mathsf{E}_\pi(\tau - \theta \,|\, \tau \geq \theta) = \mathsf{E}_\pi(\tau^*_\alpha - \theta \,|\, \tau^*_\alpha \geq \theta)}$$

We claim that if, e.g., $\pi = 0$, then the optimal time τ_α^* is of the form

$$\tau_\alpha^* = \inf\{t\colon \pi_t \geq A_\alpha^*\},$$

where $A_\alpha^* = 1 - \alpha$.

For the proof it suffices to observe that $\mathsf{P}_\pi(\tau < \theta) = \mathsf{E}_\pi(1 - \pi_\tau)$ and for τ_α^*

$$\mathsf{E}_\pi(1 - \pi_{\tau_\alpha^*}) = 1 - A_\alpha^*.$$

(g) Let us turn to the formula

$$V^*(0) = \inf_\tau \Big\{ \mathsf{P}_0(\tau < \theta) + c\mathsf{E}_0(\tau - \theta)^+ \Big\}. \tag{140}$$

The optimal stopping time $\tau^* = \tau^*(c)$ is of the form

$$\tau^*(c) = \inf\{t \ge 0\colon \pi_t \ge A^*(c)\}.$$

The threshold $A^*(c)$ depends continuously on c and, for $0 < \alpha < 1$ given, one can choose $c = c_\alpha$ such that $A^*(c_\alpha) = A^*_\alpha$, where $A^*_\alpha = 1 - \alpha$. Let us use the notation $\tau^*_\alpha = \tau^*(c_\alpha)$. Then

$$\begin{aligned} V^*(0) &= \mathsf{P}_0(\tau^*_\alpha < \theta) + c_\alpha \mathsf{E}_0(\tau^*_\alpha - \theta \,|\, \tau^*_\alpha \ge \theta) \\ &= \alpha + c_\alpha \mathsf{E}_0\Big(\tau^*_\alpha - \theta \,|\, \tau^*_\alpha \ge \theta\Big)(1-\alpha). \end{aligned} \tag{141}$$

On the other hand, we already know from (138) that

$$V^*(0) = (1-A^*_\alpha)+\frac{c\alpha}{\rho}\int_0^{A^*_\alpha}\left[\int_0^x e^{-(\lambda/\rho)[G(x)-G(u)]}\,\frac{du}{u(1-u)^2}\right]dx. \quad (142)$$

Comparing (141) and (142) (with $A^*_\alpha = 1-\alpha$) for the conditional mean time of delay

$$\mathbb{R}(\alpha,\lambda) = \mathsf{E}_0(\tau^*_\alpha - \theta \,|\, \tau^*_\alpha > \theta) \quad (143)$$

under assumption that $\mathsf{P}_0(\tau^*_\alpha > \theta) = 1-\alpha$, the following formula:

$$\mathbb{R}(\alpha,\lambda) = \frac{1}{(1-\alpha)\rho}\int_0^{1-\alpha}\left[\int_0^x e^{-(\lambda/\rho)[G(x)-G(u)]}\,\frac{du}{u(1-u)^2}\right]dx.$$

If $\lambda \to 0$ and $\alpha \to 1$ in such a way that $(1-\alpha)/\lambda \to T$, then the limiting risk

$$\mathbb{R}(T) \equiv \lim \mathbb{R}(\alpha, \lambda) \tag{144}$$

is given by

$$\mathbb{R}(T) = \frac{1}{\rho}\left\{e^b\Big(-\operatorname{Ei}(-b)\Big) - 1 + b\int_0^\infty e^{-bu}\,\frac{\log(1+a)}{u}\,du\right\}, \tag{145}$$

where $-\operatorname{Ei}(-b) = \int_b^\infty e^{-t}t^{-1}\,dt$ and $b = (\rho T)^{-1}$. Letting for simplicity $\rho = 1$, we find from (144) that

$$\boxed{\mathbb{R}(T) = \begin{cases} \frac{T}{2} + O(T^2), & T \to 0, \\ \log T - (1 + \mathbf{C}) + O(T^{-1}), & T \to \infty, \end{cases}} \tag{146}$$

where $\mathbf{C}$ if the Euler constant ($\mathbf{C} = 0.577\ldots$).

REMARK. Let

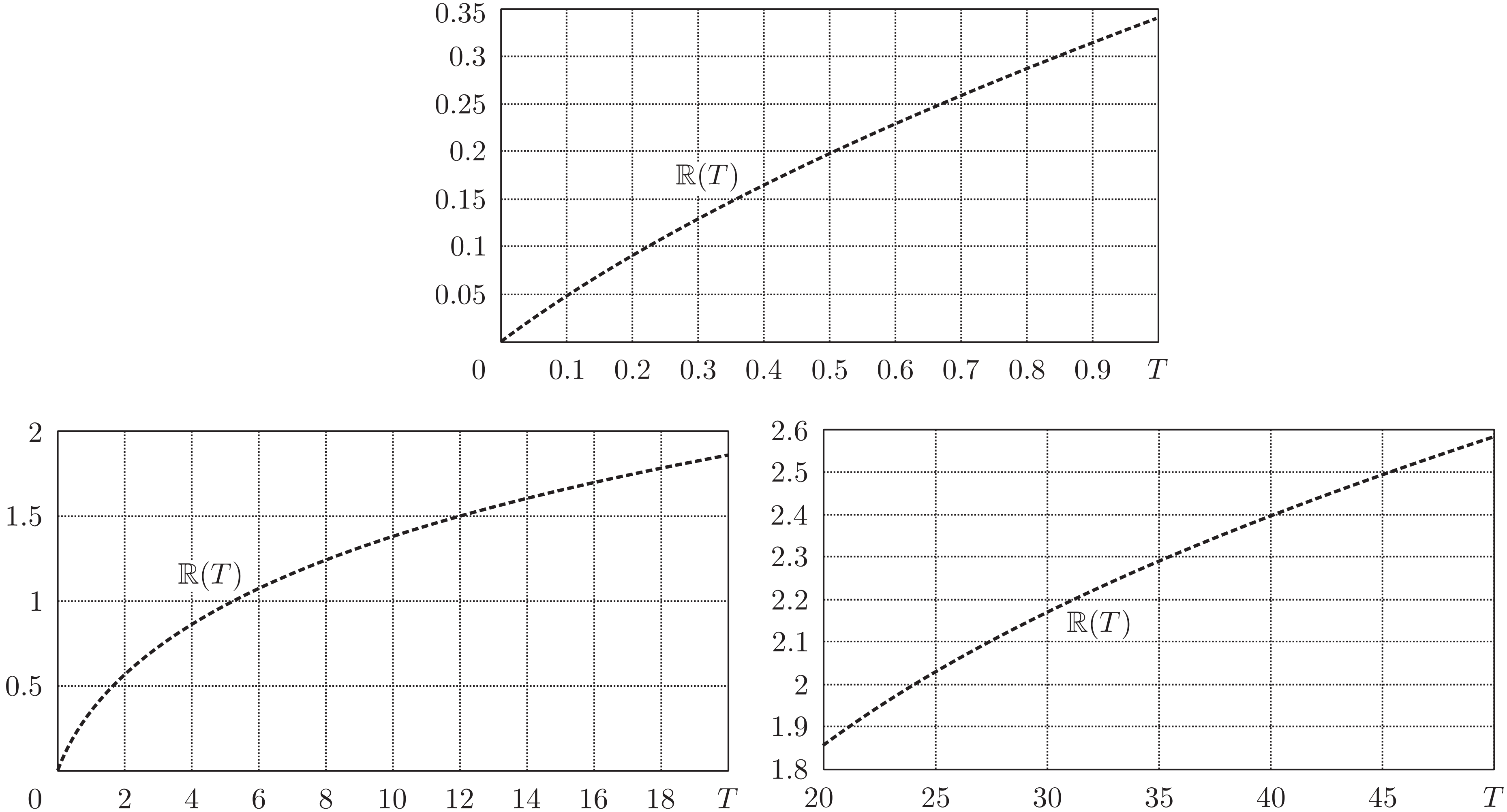
$$V(\tau; c) = \mathsf{E}_\pi\biggl((1-\pi_\tau) + c\int_0^\tau \pi_t\, dt\biggr)$$

and

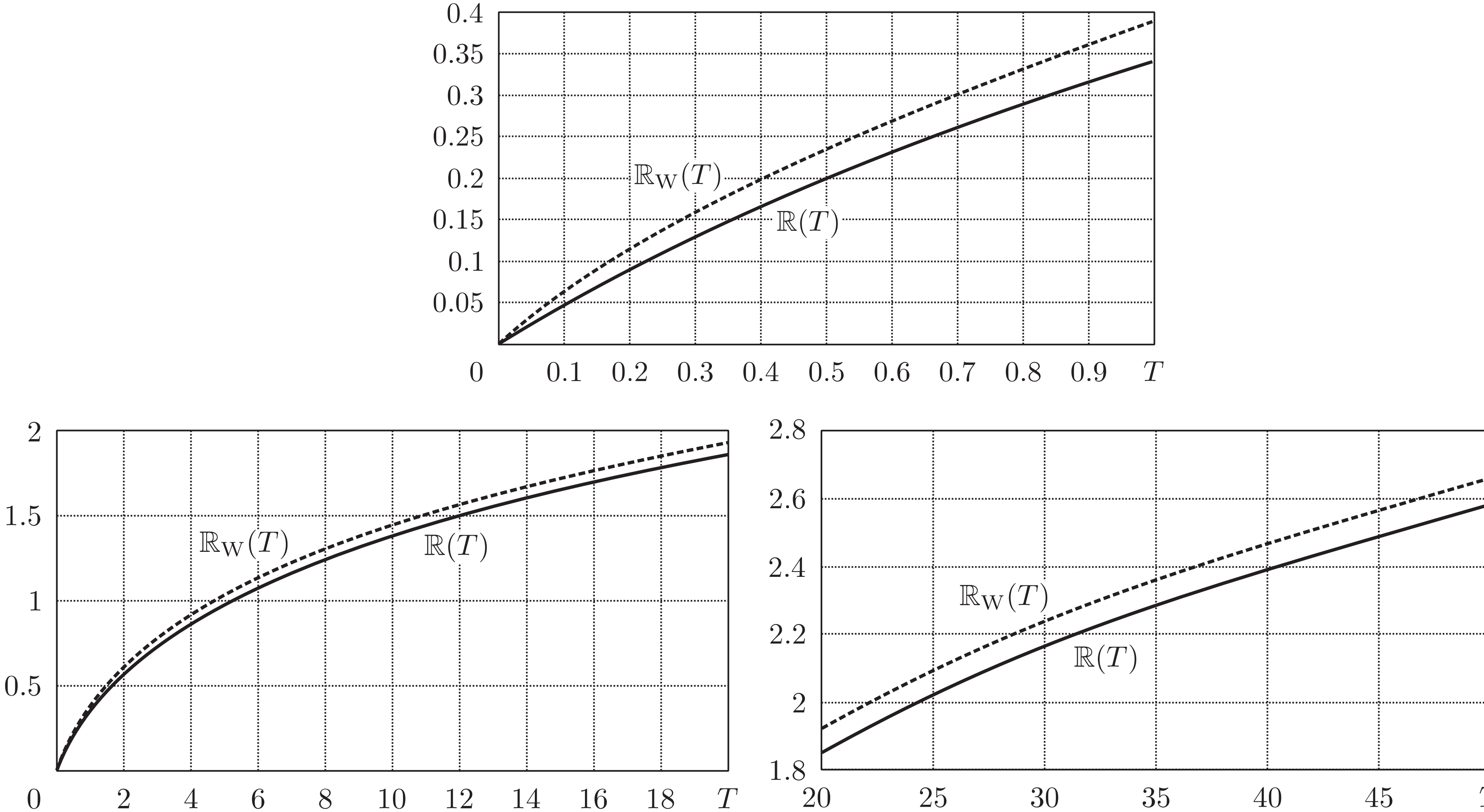
$$R(\tau) = \mathsf{E}_\pi|\tau - \theta|.$$

In **(Shiryaev, 2004)** it was shown that the criterion $R(\tau)$ is a particular case of the criterion $V(\tau; c)$. Moreover,

$$R(\tau) = \frac{1}{\lambda} V(\tau; \lambda).$$


0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 T
$\mathbb{R}(T)$
2
1.5
1
0.5
0
2 4 6 8 10 12 14 16 18 T
$\mathbb{R}(T)$
2.6
2.5
2.4
2.3
2.2
2.1
2
1.9
1.8
20 25 30 35 40 45 T
$\mathbb{R}(T)$

V-6-22



0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
T
$\mathbb{R}_{\mathrm{W}}(T)$
$\mathbb{R}(T)$
2
1.5
1
0.5
0
2
4
6
8
10
12
14
16
18
T
$\mathbb{R}_{\mathrm{W}}(T)$
$\mathbb{R}(T)$
2.8
2.6
2.4
2.2
2
1.8
20
25
30
35
40
45
T
$\mathbb{R}_{\mathrm{W}}(T)$
$\mathbb{R}(T)$

V-6-23

6.3. (h) As we have already seen, in conditionally extremal problem of finding

$$\inf_{\tau\in\mathfrak{M}_\alpha} \mathbb{E}^{(\lambda)}(\tau-\theta\,|\,\tau\ge\theta)$$

where $\mathfrak{M}_\alpha = \{\tau\colon \mathbb{P}^{(\lambda)}(\tau\le\theta)\le\alpha\}$ (c $\mathbb{P}^{(\lambda)}(\theta=0)=0$), the optimal rule τ_α^* is of the form

$$\tau_\alpha^* = \inf\{t\colon \pi_t = 1-\alpha\},$$

or, equivalently,

$$\tau_\alpha^* = \inf\Big\{t\colon \varphi_t = \frac{1-\alpha}{\alpha}\Big\}, \tag{147}$$

where

$$\varphi_t = \lambda\int_0^t e^{-\lambda s}\frac{L_t}{L_s}\,ds \tag{148}$$

with $L_t = \exp\{(r/\sigma^2)X_t - (r^2/2\sigma^2)t\}$. By the Itô formula,

$$d\varphi_t = \lambda(a-\varphi_t)\,dt + \frac{r}{\sigma^2}\varphi_t\,dX_t. \tag{149}$$

Letting $\psi_t = \varphi_t/\lambda$ and passing in (148), (149) to the limit, we find

$$\psi_t = \int_0^t \frac{L_t}{L_s}\, ds \tag{150}$$

and

$$d\psi_t = dt + \frac{r}{\sigma^2}\,\psi_t\, dX_t. \tag{151}$$

This process $\psi = (\psi_t)_{t\geq 0}$, introduced in **(Shiryaev, 1961-5, 1963)**, is a key to solving the problem

$$\inf_{\tau\colon\, \mathsf{E}_\infty\tau = T} \frac{1}{T}\int_0^\infty \mathsf{E}_u(\tau - u)^+\, du \qquad \textbf{(Variant B)}. \tag{152}$$

According to (147), we have

$$\begin{aligned}\tau^*_\alpha = \inf\Big\{t\colon \varphi_t = \frac{1-\alpha}{\alpha}\Big\} &= \inf\Big\{t\colon \frac{\varphi_t}{\lambda} = \frac{1-\alpha}{\alpha\lambda}\Big\}\\ \longrightarrow \tau^* = \inf\{t\colon \psi_t = T\},&\end{aligned} \tag{153}$$

if $\lambda \to 0$ and $\alpha \to 1$ in such a way that $(1-\alpha)/(\alpha\lambda) \to T$.

From (113) and (153) we obtain following result: the quantity

$$\mathbb{R}(T) = \inf_{\tau:\ \mathsf{E}_\infty \tau = T} \frac{1}{T} \int_0^\infty \mathsf{E}_u (\tau - u)^+ \, du, \tag{154}$$

which is the minimal mean time of delay in multistage return-cyclic system of observation, under the stationary regime, is given by (99). In particular,

$$\mathbb{R}(T) = \begin{cases} \frac{T}{2} + O(T^2), & T \to 0, \\ \log T - (1 + \mathbf{C}) + O(T^{-1} \log^2 T), & T \to \infty. \end{cases} \tag{155}$$

The optimal detection method is based on observation of the process $\psi = (\psi_t)_{t \ge 0}$, $\psi_0 = 0$ (which returns back to zero after reaching the level T).

6.4. From what we have said above, we see that the sufficient statistics for the problem (152) is the process $(\psi_t)_{t\geq 0}$ controlled by the stochastic differential equation (151). In **(Shiryaev, 2008b)** we considered a more general case, with some nonlinear function $\mathsf{E}_u G((\tau-u)^+)$ instead of the linear function of delay, $\mathsf{E}_u(\tau-u)^+$. Under some special assumptions on the functions $G(t)$, $t \geq 0$, we show that for any stopping time τ

$$\int_0^\infty \mathsf{E}_u G\Big((\tau-u)^+\Big)\,du = \mathsf{E}_\infty \int_0^\tau \Psi_u(G)\,du,$$

where $\Psi_u(G)$ is a diffusion process which can be represented as a linear combination of diffusion process which form the Markov family of sufficient statistics (see Theorem 1 in **(Shiryaev, 2008b)**).

In the discrete-time case the question about Markov family of sufficient statistics in nonadditive Bayesian and generalized Bayesian disorder problems was studied in **(Shiryaev, 1964TVP)**, **(Shiryaev, Zryumov, 2010)**.

§ 7. Direct method of finding the optimal stopping time in VARIANT B

Let us consider now the **direct** method of solving the problem

$$\inf_{\tau \in \mathfrak{M}_T} \int_0^\infty \mathsf{E}_u(\tau - u)^+ \, du,$$

where $\mathfrak{M}_T = \{\tau \colon \mathsf{E}_\infty \tau = T\}$.

The key point in this method is the following relation established in **(Feinberg, Shiryaev, 2006)**:

$$\boxed{\int_0^\infty \mathsf{E}_u(\tau - u)^+ \, du = \mathsf{E}_\infty \int_0^\tau \psi_u \, du},$$

where

$$d\psi_u = du + \frac{r}{\sigma^2} \psi_u \, dX_u.$$

Thus,

$$\mathbb{R}(T) = \inf_{\tau \in \mathfrak{M}_T} \frac{1}{T} \int_0^\infty \mathsf{E}_u(\tau - u)^+ \, du = \inf_{\tau \in \mathfrak{M}_T} \frac{1}{T} \mathsf{E}_\infty \int_0^\tau \psi_u \, du.$$

According to the general theory of stopping rules [see **(Shiryaev, 1976)**, **(Peskir, Shiryaev, 2005)**, the time

$$\tau_T^* = \inf\{t \colon \ \psi_t = T\}$$

is optimal in the problem

$$\tau \in \mathfrak{M}_T \qquad \rightsquigarrow \qquad \inf_\tau \frac{1}{T} \mathsf{E}_\infty \int_0^\tau \psi_u \, du.$$

To find $\boxed{\mathbb{R}(T) = \frac{1}{T} \mathsf{E}_\infty \int_0^{\tau_T^*} \psi_u \, du}$ we have to find $\boxed{U(x) = \mathsf{E}_\infty^{(x)} \int_0^{\tau_T^*} \psi_u \, du}$,

where $(\psi_u)_{u \ge 0}$ is a diffusion Markov process with $d\psi_u = du + \frac{r}{\sigma} \psi_u \, dB_u$ and $\psi_0 = x \ge 0$.

Solving the equation

$$U'(x) + \rho x^2 U''(x) = -x, \quad \text{where} \quad \rho = \frac{r^2}{2\sigma^2},$$

we find that

$$U(x) = G\left(\frac{1}{T}\right) - G\left(\frac{1}{x}\right),$$

where

$$G(x) = \int_x^\infty \frac{F(u)}{u^2}\, du, \qquad F(x) = e^x(-\operatorname{Ei}(-x))$$

$$(-\operatorname{Ei}(-x) = \int_x^\infty \frac{e^{-u}}{u}\, du).$$

Straightforward calculations show that

$$
\begin{aligned}
\mathbb{R}(T) &= \frac{1}{T}\mathsf{E}_{\infty}^{(0)}\int_0^{\tau_T^*}\psi_u\,du = \frac{1}{T}U(0) = \frac{1}{T}G\Big(\frac{1}{T}\Big)\\
&= F\Big(\frac{1}{T}\Big) - \left[1 - \frac{1}{T}\int_{1/T}^{\infty} e^{-u/T}\,\frac{\log(1+u)}{u}\,du\right],
\end{aligned}
$$

which coincides with (145).

In particular (as was already obtained in (146)),

$$
\mathbb{R}(T) = \begin{cases} \frac{T}{2} + O(T^2), & T \to 0,\\ \log T - (1+\mathbf{C}) + O(T^{-1}\log^2 T), & T \to \infty, \end{cases}
$$

where $\mathbf{C} = 0.577\ldots$ is the Euler constant.

§ 8. Minimax problems of quickest detection (Variants C and D)

8.1. For the scheme

$$dX_t = \begin{cases} \sigma\, dB_t, & t < \theta, \\ r\, dt + \sigma\, dB_t, & t \ge \theta, \end{cases}$$

where $\theta \in [0, \infty]$ is a parameter, the following two Variants (C and D) of minimax problems—proposed by **M. Pollak** and **G. Lorden**, respectively—are of a great interest.

VARIANT C. In the class $\mathfrak{M}_T = \{\tau\colon\ \mathsf{E}_\infty \tau = T\}$, to find

$$\boxed{\mathbb{C}(T) = \inf_{\tau \in \mathfrak{M}_T} \sup_{\theta \ge 0} \mathsf{E}_\theta(\tau - \theta \,|\, \tau \ge \theta)}. \tag{156}$$

VARIANT D. In the class $\mathfrak{M}_T = \{\tau\colon\ \mathsf{E}_\infty \tau = T\}$, to find

$$\boxed{\mathbb{D}(T) = \inf_{\tau \in \mathfrak{M}_T} \sup_{\theta \ge 0} \operatorname*{ess\,sup}_{\omega} \mathsf{E}_\theta(\tau - \theta \,|\, \mathcal{F}_\theta^X)}. \tag{157}$$

In **(Feinberg, Shiryaev, 2006)** it is shown that

$$\boxed{\mathbb{B}(T) \le \mathbb{C}(T) \le \mathbb{C}^*(T)}, \tag{158}$$

where

$$\mathbb{C}^*(T) = \mathsf{E}_0 \tau_T^* \quad \text{and} \quad \tau_T^* = \inf\{t \ge 0 \colon \psi_t \ge T\}.$$

Since $\mathbb{C}^*(T) = \mathsf{E}_0 \tau_T^* = F(b)$, $b = 1/T$, we have

$$F(b) - \Delta(b) \le \mathbb{C}(T) \le F(b), \tag{159}$$

where $\Delta(b) = 1 - b\int_0^\infty u^{-1} F(u)\,du$, $F(b) = -\operatorname{Ei}(-b)$.

From (159) we deduce that for large T

$$\boxed{\log T - (1+\mathbf{C}) + O(T^{-1}\log^2 T) \le \mathbb{C}(T) \le \log T - \mathbf{C} + O(T^{-1}\log^2 T)}. \tag{160}$$

For small T

$$\frac{T}{2} + O(T^2) \le \mathbb{C}(T) \le T + O(T^2).$$

The "gap" in second-order terms in (160) ($-(1+\mathbf{C})$ on the left-hand side and $-\mathbf{C}$ on the right-hand side) can be eliminated, if one turns to **RANDOMIZED rules** proposed by **M. Pollak**.

Let $\overline{\psi} = (\overline{\psi}_t)_{t\ge 0}$ be a process with

$$d\overline{\psi}_t = dt + \frac{r}{\sigma^2}\,\overline{\psi}_t\,dX_t$$

and a "random" initial state $\overline{\psi}_0$ whose probability density $g = g(y)$, $y \ge 0$, is chosen in a special way.

The results of **(Burnaev, Feinberg, Shiryaev,2008)** imply that, in the class $\overline{\mathfrak{M}}_T$ of randomized stopping times $\overline{\tau}$ with $\overline{\mathsf{E}}_\infty\overline{\tau} = T$, the following estimates hold for $\overline{\mathbb{C}}(T) = \inf_{\overline{\tau}\in\overline{\mathfrak{M}}_T} \sup_{\theta\geq 0} \overline{\mathsf{E}}_\theta(\overline{\tau}-\theta \,|\, \overline{\tau}\geq\theta)$:

$$\boxed{\mathbb{B}(T) \leq \overline{\mathbb{C}}(T) \leq \overline{\mathbb{C}}^*(T) = \overline{\mathsf{E}}_0\tau^*_{g^*(A)}}, \tag{161}$$

where $g^*(A)$ is the initial density which is chosen in a special way (on $[0, A]$) for $\overline{\psi}_0$ with a certain threshold A.

Calculation of $\overline{\mathsf{E}}_0\tau^*_{g^*(A)}$ in (161) shows that for large T

$$\begin{aligned}\log T-(1+\mathbf{C})+O(T^{-1}\log^2 T) &\le \\ \le \overline{\mathbb{C}}(T) &\le \\ &\le \log T-(1+\mathbf{C})+O(T^{-1}\log^2 T).\end{aligned}$$

This implies that the suggested randomized method, which is based on observation upon the process $(\overline{\psi}_t)_{t\ge 0}$ with certain special initial distribution for ψ_0, is asymptotically optimal $(T\to\infty)$ with the first two terms of asymptotics equal to $\log T-(1+\mathbf{C})$ and higher terms of order $O(T^{-1}\log^2 T)$.

The question about optimal (non-asymptotical) method in Variants C and $\overline{\mathsf{C}}$ is still open.

8.2. In the discrete-time case G. Lorden, who introduced the minimax criterion D, established [see **(Lorden, 1971)**] that the CUSUM-method is asymptotically optimal (as $\mathsf{E}_\infty \tau \ge T \to \infty$).

The optimality of this method was established in **(Moustakides, 1986)**.

In the continuous-time case, for the scheme $dX_t = rI(t > \theta)\,dt + \sigma\,dB_t$, the optimality of the CUSUM-method was established in **(Beibel, 1996)**, **(Shiryaev, 1996)**. See also **(Moustakides, 2004)**.

CUSUM method

discrete-time case

asymptotical optimality (as $\mathsf{E}_\infty\tau \geq T \to \infty$) is proved in **(Lorden, 1971)**

optimality is proved in **(Moustakides, 1986)**

continuous-time case

for the scheme $dX_t = rI(t > \theta)\,dt + \sigma\,dB_t$, optimality is proved in **(Beibel, 1996)** and **(Shiryaev, 1996)**

See also **(Moustakides, 2004)**

The essence of this method consists in the following.

Let $T_t = \sup_{\theta \le t} \dfrac{L_t}{L_\theta}$ be the CUSUM-process with $L_t = \dfrac{d\mathsf{P}_0}{d\mathsf{P}_\infty}(t,x)$.

In the class $\mathfrak{M}_T = \{\tau\colon\ \mathsf{E}_\infty \tau = T\}$, the **optimal time** for criterion D is

$$\boxed{\tau_T^* = \inf\{t\colon\ T_t \ge D\}}$$

where the threshold D is a root of the equation

$$D - 1 - \log D = T \quad (\text{in the case } \rho \equiv r^2/(2\sigma^2) = 1).$$

For large T $\boxed{\mathbb{D}(T) = \log T - 1 + O(T^{-1})}$.

The key to the proof of optimality of the time τ_T^* is the following inequality: for any time τ

$$\sup_\theta \operatorname*{ess\,sup}_\omega \mathsf{E}_\theta\Big[(\tau-\theta)^+ \,|\, \mathcal{F}_\theta^X\Big](\omega) \ge \frac{\mathsf{E}_\infty \int_0^\tau T_t\,dt}{\mathsf{E}_\infty T_\tau}.$$

§ 9. Summary of results in Variants A, B, C and D

The observed process $X = (X_t)_{t\ge 0}$ satisfies

$$dX_t = \begin{cases} \sigma\, dB_t, & t < \theta, \\ r\, dt + \sigma\, dB_t, & t \ge \theta. \end{cases}$$

VARIANT A. We assume that θ is a random variable with the exponential distribution $\exp\{\lambda\}$.

Bayesian setup: To find

$$\boxed{\mathbb{V}(T) = \inf_{\tau\in\mathfrak{M}} \Big[\mathsf{P}_\pi(\tau < \theta) + c\mathsf{E}_\pi(\tau - \theta)^+\Big]}.$$

Conditionally extremal setup: To find

$$\boxed{\mathbb{R}(\alpha, \lambda) = \inf_{\tau\in\mathfrak{M}_\alpha} \mathsf{E}_0(\tau - \theta \,|\, \tau \ge \theta)}.$$

If $\alpha \to 1$ and $\lambda \to 0$ in such a way that $(1-\alpha)/\lambda \to T$, then the **limiting risk** $\mathbb{R}(T) \equiv \mathbb{R}(\alpha, \lambda)$ is given by

$$\boxed{\mathbb{R}(T) = \begin{cases} T/2 + O(T^2), & T \to 0, \\ \log T - (1 + \mathbf{C}) + O(T^{-1}), & T \to \infty, \end{cases}}$$

where $\mathbf{C} = 0.577\ldots$ is the Euler constant (we put $\rho \equiv r^2/(2\sigma^2) = 1$).

The **optimal time τ^* in $\mathbb{V}(\pi)$-criterion** is given by

$$\boxed{\tau^* = \inf\{t \colon \pi_t \geq A^*\}}\,.$$

The **optimal time τ^*_α in $\mathbb{R}(\alpha, \lambda)$-criterion** is given by

$$\boxed{\tau^*_\alpha = \inf\{t \colon \pi_t \geq 1 - \alpha\}}\,.$$

VARIANT B. Here θ is assumed to be a parameter from $[0,\infty]$.

A multistage problem reduces to **generalized Bayesian setup**

$$\boxed{\mathbb{B}(T) = \inf_{\tau\in\mathfrak{M}_T} \frac{1}{T}\int_0^\infty \mathsf{E}_u(\tau-u)^+\,du}\,.$$

The **optimal stopping time**:

$$\boxed{\tau_T^* = \inf\{t\colon \psi_t = T\}}\,.$$

The **limiting risk**:

$$\boxed{\mathbb{B}(T) = \mathbb{R}(T) = \begin{cases} T/2 + O(T^2), & T\to 0,\\ \log T - (1+\mathrm{C}) + O(T^{-1}), & T\to\infty, \end{cases}}$$

VARIANT C. Here θ is a parameter from $[0,\infty]$ and the criteria of quality of detection are given by

$$\mathbb{C}(\boldsymbol{T}) = \inf_{\tau\in\mathfrak{M}_T} \sup_{\theta\geq 0} \mathsf{E}_\theta(\tau-\theta \,|\, \tau\geq\theta) \qquad (\mathfrak{M}_T = \{\tau\colon\ \mathsf{E}_\infty\tau = T\}),$$

$$\overline{\mathbb{C}}(\boldsymbol{T}) = \inf_{\overline{\tau}\in\overline{\mathfrak{M}}_T} \sup_{\theta\geq 0} \overline{\mathsf{E}}_\theta(\overline{\tau}-\theta \,|\, \overline{\tau}\geq\theta)$$

($\overline{\mathfrak{M}}_T$ is the class of randomized stopping times with $\overline{\mathsf{E}}_\infty\overline{\tau} = T$).

As $T\to\infty$

$$\log T - (1+\mathbf{C}) + O(T^{-1}\log^2 T) \leq \mathbb{C}(\boldsymbol{T}) \leq \log T - \mathbf{C} + O(T^{-1}\log^2 T),$$

$$\log T - (1+\mathbf{C}) + O(T^{-1}\log^2 T) = \overline{\mathbb{C}}(\boldsymbol{T}).$$

Asymptotically optimal methods (as $T\to\infty$) are based on observation upon the processes $(\psi_t)_{t\geq 0}$ and $(\overline{\psi}_t)_{t\geq 0}$ with

$$d\psi_t = dt + \psi_t\, dX_t, \qquad \psi_0 = 0$$

$$d\overline{\psi}_t = dt + \overline{\psi}_t\, dX_t, \qquad \overline{\psi}_0 \text{ is a random variable with a special distribution.}$$

VARIANT D. θ is a parameter from $[0,\infty]$ and the criterion of quality of detection is

$$\boxed{\mathbb{D}(T) = \inf_{\tau\in\mathfrak{M}_T} \sup_{\theta\geq 0} \operatorname*{ess\,sup}_{\omega} \mathsf{E}_\theta\Big[(\tau-\theta)^+ \,|\, \mathcal{F}_\theta^X\Big](\omega)}.$$

The **optimal stopping time** is

$$\boxed{\tau_T^* = \inf\{t\geq 0\colon\ T_t \geq D\}}$$

where

- $T_t = \sup_{\theta\leq t} \dfrac{L_t}{L_\theta}$ is the CUSUM-process and,
- D solves the equation $D - 1 - \log D = T$.

As $T\to\infty$ $\boxed{\mathbb{D}(T) = \log T - 1 + O(T^{-1})}$.

§ 10. Some remarks about θ-models and Bayesian G-models

10.1. We focused above on continuous-time models and, what is more, on the models of the form $dX_t = rI(t > \theta)\,dt + \sigma\,dB_t$, where $(B_t)_{t \geq 0}$ is a Brownian motion. The corresponding models with discrete time are certainly of a great interest. These models are generally considered in asymptotic variants. The ideas of invariance principle suggest that a great part of results for the discrete-time disorder problem (at least in asymptotical aspects) can be obtained from the above-exposed results for the model $dX_t = rI(t > \theta)\,dt + \sigma\,dB_t$. For more details see **(Shiryaev, 1963-I)**.

In the discrete-time case one considers, as a rule, models with independent observations. Let formulate now the *general model* [proposed in **(Shiryaev, 2008a)**] for disorder problems in the discrete-time case without independency assumption.

We start with a binary filtered probability-statistical experiment

$$(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}; \mathsf{P}^0, \mathsf{P}^\infty),$$

where $(\mathcal{F}_n)_{n\geq 0}$ is a filtration, $\mathcal{F}_0 = \{\varnothing, \Omega\} \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}$, $\mathcal{F}_\infty = \sigma(\bigcup \mathcal{F}_n) = \mathcal{F}$.

Letting

$$\mathsf{P}^0_n = \mathsf{P}^0|\mathcal{F}_n, \quad \mathsf{P}^\infty_n = \mathsf{P}^\infty|\mathcal{F}_n, \quad \mathsf{P} = \tfrac{1}{2}(\mathsf{P}^0 + \mathsf{P}^\infty), \quad \mathsf{P}_n = \mathsf{P}|\mathcal{F}_n,$$

introduce

$$L^0_n = \frac{d\mathsf{P}^0_n}{d\mathsf{P}_n}, \qquad L^\infty_n = \frac{d\mathsf{P}^\infty_n}{d\mathsf{P}_n}, \tag{162}$$

$$M^0_n = \sum_{k=1}^{n} \frac{\Delta L^0_k}{L^0_{k-1}}, \qquad M^\infty_n = \sum_{k=1}^{n} \frac{\Delta L^\infty_k}{L^\infty_{k-1}}. \tag{163}$$

Then, if $\mathcal{E}(M)_n = \prod(1+\Delta M_k)$, then L^0_n and L^∞_n admit the following representations:

$$L^0_n = \mathcal{E}(M^0)_n, \quad L^\infty_n = \mathcal{E}(M^\infty)_n. \tag{164}$$

For independent observations with densities $f^\infty(x)$, $f^0(x)$ we have $L_n^0 = f^0(x_1)\cdots f^0(x_n)$, $L_n^\infty = f^\infty(x_1)\cdots f^\infty(x_n)$. In this case

$$\Delta M_n^0 = \frac{\Delta L_n^0}{L_{n-1}^0} = \frac{L_n^0}{L_{n-1}^0} - 1 = f^0(x_n) - 1$$

and, evidently,

$$\mathcal{E}(M^0)_n = \prod_{k=1}^n (1 + \Delta M_k^0) = \prod_{k=1}^n f^0(x_k) = L_n^0.$$

Analogously, $\mathcal{E}(M^\infty)_n = L_n^\infty$.

If disorder occurs at time $\theta < n$, then the density of random variables $x_1, \dots, x_n$ is given by $f^\infty(x_1)\cdots f^\infty(x_{\theta-1}) f^0(x_\theta)\cdots f^0(x_n)$, which can be rewritten in the form

$$L_n^\theta = \mathcal{E}(M^\theta)_n \quad \text{if } \Delta M_k^\theta = I(k<\theta)\Delta M_k^\infty + I(k \geq \theta)\Delta M_k^0. \tag{165}$$

The above considerations show how to determine—for a given $\theta \in \{0, 1, \dots, \infty\}$—the corresponding measure P^θ: one should let

$$\mathsf{P}^\theta(A) = \mathsf{E}[I(A)\mathcal{E}(M^\theta)_\infty], \qquad A \in \mathcal{F}, \tag{166}$$

where

$$\mathcal{E}(M^\theta)_\infty = \lim_{n\to\infty} \mathcal{E}(M^\theta)_n, \qquad \mathcal{E}(M^\theta)_n = \prod_{k=1}^{n} (1 + \Delta M_k^\theta)$$

(ΔM_k^θ is given in (165)).

10.2. If θ is a r.v. with distribution function $G = G(n)$, $n \geq 0$, then the corresponding probability distribution $\mathsf{P}^G(A)$, $A \in \mathcal{F}_n$, is given by

$$\mathsf{P}^G(A) = \sum_{\theta=0}^{n} \mathsf{P}_n^\theta(A)\Delta G(\theta) + (1 - G(n))\mathsf{P}_n^\infty(A). \tag{167}$$

Let $\mathsf{P}_n^G = \mathsf{P}^G | \mathcal{F}_n$ and $L_n^G = \dfrac{d\mathsf{P}_n^G}{d\mathsf{P}_n}$. Then $L_n^G = \sum_{\theta=0}^{\infty} L_n^\theta \Delta G(\theta)$. Taking into account the representation

$$L_n^\theta = I(n < \theta)L_n^\infty + I(n \geq \theta)L_n^0 \frac{L_{\theta-1}^\infty}{L_{\theta-1}^0}, \tag{168}$$

we find that

$$L_n^G = L_n^0 \sum_{\theta=0}^{n} \frac{L_{\theta-1}^\infty}{L_{\theta-1}^0} \Delta G(\theta) + L_n^\infty (1 - G(n)) \tag{169}$$

(with $L_{-1}^\infty = L_{-1}^0 = 1$).

EXAMPLE. Let θ have the geometric distribution G with $G(0)=\pi$, $\Delta G(n)=(1-\pi)q^{n-1}p$, $n\ge 1$. In this case

$$L_n^G=\pi L_n^0+(1-\pi)\sum_{k=0}^{n-1}pq^k\,\frac{L_k^\infty}{L_k^0}+(1-\pi)q^nL_n^\infty.$$

If there exists the densities $f_\infty^G(x_1,\dots,x_k)$ and $f_k^0(x_1,\dots,k_n)$, then we find that

$$\begin{aligned}
f_n^G(x_1,\dots,x_n)&=\pi f_n^0(x_1,\dots,x_n)\\
&\quad+(1-\pi)\sum_{k=0}^{n-1}pq^kf_k^\infty(x_1,\dots,x_k)f_{n,k}^0(x_{k+1},\dots,x_n\,|\,x_1,\dots,x_k)\\
&\quad+(1-\pi)q^nf_n^\infty(x_1,\dots,x_n),
\end{aligned}$$

where $f_{n,k}^0(x_{k+1},\dots,x_n\,|\,x_1,\dots,x_k)$ are conditional densities.

Let us describe—for **θ-models** and **G-models**—some **BASIC STATISTICS** which are used in disorder problems.

θ-MODELS: one of the more important and frequently used is the CUSUM-statistics

$$\boxed{T_n = \max_{\theta \geq 0} \frac{L_n^\theta}{L_n^\infty}}.$$

Let $L_n := \dfrac{L_n^0}{L_n^\infty}$. Since $L_n^\theta = L_{\theta-1}^\infty \dfrac{L_n^0}{L_{(\theta-1)\wedge n}^0}$, we find that

$$T_n = \max\left(1, \max_{0 \leq \theta \leq n} \frac{L_n}{L_{\theta-1}}\right). \tag{170}$$

Let $\gamma_n := \log T_n$ and $Z_n := \log L_n$, then

$$\gamma_n = Z_n - \min_{0 \leq \theta \leq n} Z_\theta, \qquad \gamma_n = \max(0, \gamma_{n-1} + \Delta Z_n).$$

G-MODELS: the following statistics prove to be important:

$$\pi_n = \mathsf{P}^G(\theta \le n \,|\, \mathcal{F}_n), \qquad \varphi_n = \frac{\pi_n}{1-\pi_n}$$

and

$$\psi_n = \frac{L_n}{L_{n-1}}\,[1+\psi_{n-1}]. \tag{171}$$

From the Bayes formula,

$$\boxed{\pi_n = \frac{\sum_{\theta\le n} L_n^\theta \Delta G(\theta)}{L_n^G}}. \tag{172}$$

For φ_n with $L_n = L_n^0/L_n^\infty$ we find that

$$\boxed{\varphi_n = \frac{1}{1-G(n)} \sum_{\theta\le n} \frac{L_n}{L_{\theta-1}}\,\Delta G(\theta)}, \tag{173}$$

whence we deduce that

$$\varphi_n = \frac{L_n}{L_{n-1}}\left[\frac{\Delta G(n)}{1-G(n)} + \frac{1-G(n-1)}{1-G(n)}\,\varphi_{n-1}\right], \quad \varphi_0 = \frac{\pi}{1-\pi}. \tag{174}$$

If G is a **geometric distribution** with $\Delta G(0) = G(0) = \pi$ and

$$\Delta G(n) = (1 - \pi)q^{n-1}p, \qquad n \ge 1,$$

then it follows from (174) that

$$\varphi_n = \frac{L_n}{qL_{n-1}}[p + \varphi_{n-1}], \qquad \varphi_0 = \frac{\pi}{1-\pi}. \tag{175}$$

If $p \to 0$, $\pi \to 0$ in such a way that $\pi/p \to m \ge 0$, then (175) implies that $\psi_n = \lim_{p \downarrow 0}(\varphi_n/p)$ obeys the recurrent relations

$$\boxed{\psi_n = \frac{L_n}{L_{n-1}}[1 + \psi_{n-1}], \quad n \ge 1, \qquad \psi_0 = m.} \tag{176}$$

From (176) or (173) we find the following representation:

$$\psi_n = \sum_{\theta=1}^{n} \frac{L_n}{L_{\theta-1}}, \qquad n \ge 1.$$

(This statistics underlies the "**Shiryaev–Roberts procedure**" in disorder problems in the discrete-time case.)

10.3. The four variants (A, B, C, D) of disorder problems considered above play a dominating role also in the discrete-time case different assumptions on probabilistic character of observed data (see, e.g.,

(Pollak, 1985), (Roberts, 1966), (Wetherill, 1977), (Hawkins, Olwell, 1998), (Frisén, 2007), (Sequential Analysis, 2007), (Tartakovsky, 2008), (Tartakovsky, Veeravalli, 2005)

and bibliography therein).

REFERENCES

Beibel M. A note on Ritov's Bayes approach to the minimax property of the CUSUM procedure, *Ann. Statist.*, **24**:4 (1996), 1804–1812.

Burnaev E. V., Feinberg E. A., Shiryaev A. N. On the asymptotic optimality of second order in the minimax problem of the quickest detection of the time of change of drift for Brownian motion, *Theory Probab. Appl.*, **53**:3 (2008), 557–575.

Feinberg E. A., Shiryaev A. N. Quickest detection of drift change for Brownian motion in generalized Bayesian and minimax settings, *Statist. Decisions*, **24**:4 (2006), 445–470.

Frisén M. Properties of use of the Shewhart method and its followers, *Sequential Anal.*, **26**:2 (2007), 171–193.

Hawkins D. M., Olwell D. H. *Cumulative Sum Charts and Charting for Quality Improvement*, Springer-Verlag, New York, 1998.

Karlin S., Taylor H. M. *A second course in stochastic processes*, Academic Press, New York, 1981.

Lorden G. Procedures for reacting to a change in distribution, *Ann. Math. Statist.*, **42** (1971), 1897–1908.

Moustakides G. V. Optimal stopping times for detecting changes in distributions, *Ann. Statist.*, **14**:4 (1986), 1379–1387.

Moustakides G. V. Optimality of the CUSUM procedure in continuous time, *Ann. Statist.*, **32**:1 (2004), 302–315.

Page E. S. Continuous inspection schemes, *Biometrika*, **41** (1954), 100–114.

Page E. S. Control charts with warning lines, *Biometrika*, **42** (1955), 243–257.

Peskir G., Shiryaev A. N. *Optimal Stopping and Free-Boundary Problems*, Birkhäuser, Basel, 2006.

Pollak M. Optimal detection of a change in distribution, *Ann. Statist.*, **13**:1 (1985), 206–227.

Pollak M., Siegmund D. A diffusion process and its application to detecting a change in the drift of Brownian motion, *Biometrika*, **72**:2 (1985), 267–280.

Roberts S. W. Control charts based on geometric moving average, *Technometrics*, **1** (1959), 239–250.

Roberts S. W. A comparison of some control chart procedures, *Technometrics*, **8** (1966), 411–430.

Sequential Analysis. Design Methods and Applications, 26 : 2, 3 (2007). Celebrating eighty years of control charts. Walter A. Shewhart's Legacy. Taylor and Francis.

Shiryaev A. N. Обнаружение спонтанно возникающих эффектов, *Докл. АН СССР*, **138**:4 (1961), 799–801.

Shiryaev A. N. The problem of the most rapid detection of a disturbance of a stationary regime. (Russian) *Dokl. Akad. Nauk SSSR*, **138**:5 (1961), 1039–1042; Engl. transl. in *Soviet Math. Dokl.*, **2** (1961), 795–799.

Shiryaev A. N. On optimal methods in quickest detection problems. (Russian) *Teor. Veroyatnost. i Primenen.*, **8**:1 (1963), 26–51; Engl. transl. in *Theory Probab. Appl.*, **8**:1 (1963), 22–46.

Shiryaev A. N. On detecting of disorders in industrial processes. I. (Russian) *Teor. Veroyatnost. i Primenen.*, **8**:3 (1963), 264–281; Engl. transl. in *Theory Probab. Appl.*, **8**:3 (1963).

Shiryaev A. N. On detecting of disorders in industrial processes. II. (Russian) *Teor. Veroyatnost. i Primenen.*, **8**:4 (1963), 431–443; Engl. transl. in *Theory Probab. Appl.*, **8**:4 (1963).

Shiryaev A. N. Detection of randomly appearing target in a multichannel system. (Russian) *Trudy Mat. Inst. Steklova*, **71** (1964), 113–117.

Shiryaev A. N. On Markov sufficient statistics in nonadditive Bayes problems of sequential analysis. (Russian) *Teor. Veroyatnost. i Primenen.*, **9**:4 (1964), 670–686; Engl. transl. in *Theory Probab. Appl.*, **9**:4 (1964), 604–618.

Shiryaev A. N. Two problems of sequential analysis. (Russian) *Kibernetika* (*Kiev*) no. 2 (1967), 79–86; Engl. transl. in *Cybernetics*, **3**:2 (1967), 63–69.

Shiryaev A. N. Statistical sequential analysis. 2nd edn. (Russian) Nauka, Moscow, 1976; English transl.: *Optimal Stopping Rules.* Appl. Math., **8**, Springer-Verlag, New York–Heidelberg, 1978.

Shiryaev A. N. Минимаксная optimalость methoda кумулятивных сумм (CUSUM) в случае continuousого времени — Успехи матем. наук, 1996, т. 51, № 4, с. 173–174. English transl. in *Russian Math. Surveys*, **51**:4 (1996), 750–751.

Shiryaev A. N. Quickest detection problems in the technical analysis of the financial data, *Mathematical Finance — Bachelier Congress 2000: Selected papers from the First World Congress of the Bachelier Finance Society* (Paris, 2000). Ed. by H. Geman et al. Springer-Verlag, Springer Finance, Berlin, 2002, 487–521.

Shiryaev A. N. A remark on the quickest detection problems, *Statist. Decisions*, **22** (2004), 79–82.

Shiryaev A. N. On stochastic models and optimal methods in the quickest detection problems, *Theory Probab. Appl.*, **53**:3 (2008), 417–436.

Shiryaev A. N. Generalized Bayesian nonlinear quickest detection problems: On Markov family of sufficient statistics, *Mathematical Control Theory and Finance*, Springer, 2008, 377–386.

Shiryaev A. N., Zryumov P. Y. On the linear and nonlinear generalized Bayesian disorder problem (discrete-time case), *Optimality and Risk — Modern Trends in Mathematical Finance*, Springer, Berlin, 2010, 227–236.

Tartakovsky A. G. Asymptotic optimality in the Bayesian problems of the quickest detection under restriction on the global probability of false alarm, *Theory Probab. Appl.*, **53**:3 (2008), 472–499.

Tartakovsky A. G., Veeravalli V. General asymptotic Bayesian theory of quickest change detection, *Theory Probab. Appl.*, **49** (2005), 458–497.

Wetherill G. B. Sampling Inspection and Quality Control. London/New York: Chapman & Hall / Halsted Press, 1977, 146 p.

TOPIC VI. Applications-2: Testing two and three hypotheses for Brownian motion with drift

1°. On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathsf{P})$ we observe a process

$$X_t = \mu t + B_t, \qquad t \ge 0,$$

where $B = (B_t)_{t\ge 0}$ is a Brownian motion and μ takes one of the three values:

$$\begin{array}{lcr}
\mu = \mu^1, & & (\text{hypothesis } H^1), \\
\mu = \mu^0, & \text{with } \mu^1 < \mu^0 < \mu^2 & (\text{hypothesis } H^0), \\
\mu = \mu^2, & & (\text{hypothesis } H^2).
\end{array}$$

Some notation:

$$\mathcal{F}_t^X = \sigma(\omega\colon X_s,\, s \le t), \qquad \mathsf{P}^i = \mathsf{Law}(X \mid \mu = \mu^i),$$

$$\mathsf{P}_t^i = \mathsf{P}^i | \mathcal{F}_t^X, \quad L_t^i = \frac{d\mathsf{P}_t^i}{d\mathsf{P}_t^0}, \qquad \pi_t^i = \mathsf{P}(\mu = \mu^i \mid \mathcal{F}_t^X), \quad \varphi_t^i = \frac{\pi_t^i}{\pi_t^0}.$$

Known formulas:

$$L_t^i = \exp\Big\{(\mu^i - \mu^0)X_t - \frac{1}{2}\big[(\mu^i)^2 - (\mu^0)^2\big]t\Big\},$$

$$dL_t^i = L_t^i(\mu^i - \mu^0)(dX_t - \mu^0\,dt), \qquad L_0^i = 1,$$

$$d\varphi_t^i = \varphi_t^i(\mu^i - \mu^0)(dX_t - \mu^0\,dt) \qquad \text{(by Bayes' formula)},$$

$$\pi_t^1 = \frac{\varphi_t^1}{1 + \varphi_t^1 + \varphi_t^2}, \qquad \pi_t^2 = \frac{\varphi_t^2}{1 + \varphi_t^1 + \varphi_t^2}, \qquad \pi_t^0 = \frac{1}{1 + \varphi_t^1 + \varphi_t^2}.$$

The innovation representation of X:

$$dX_t = \mathsf{E}(\mu \,|\, \mathcal{F}_t^X)\,dt + d\overline{B}_t,$$

where $\quad \overline{B} = (\overline{B}_t, \mathcal{F}_t^X)$ is an innovation Brownian motion,

$\qquad\quad \mathsf{E}(\mu \,|\, \mathcal{F}_t^X) = \mu^1\pi_t^1 + \mu^0\pi_t^0 + \mu^2\pi_t^2.$

Thus,

$$dX_t = \frac{\mu^0 + \mu^1\varphi_t^1 + \mu^2\varphi_t^2}{1 + \varphi_t^1 + \varphi_t^2}\,dt + d\overline{B}_t,$$

$$d\varphi_t^i = (\mu^i - \mu^0)\varphi_t^i\Big(\frac{(\mu^1 - \mu^0)\varphi_t^1 + (\mu^2 - \mu^0)\varphi_t^2}{1 + \varphi_t^1 + \varphi_t^2}\,dt + d\overline{B}_t\Big).$$

Testing of 3 statistical hypotheses H^1, H^0, H^2: BAYESIAN FORMULATION.

Terminal risk:

$$\boxed{\begin{aligned} w(\mu^i, d^i) &= 0, \\ w(\mu^i, d^j) &= a_{ij}, \quad i \neq j \end{aligned}}$$

$(d^i = \text{accept } H^i)$.

Risk of the sequential decision $\delta = (\tau, d)$, where

- $\tau = \tau(\omega)$ is a $(\mathcal{F}_t^X)$-stopping time: $\{\tau \leq t\} \in \mathcal{F}_t^X$, $t \geq 0$;
- d is $\mathcal{F}_\tau^X$-measurable $(d = d^1, d^0, d^2)$:

$$\boxed{R_\delta(\pi) = \mathsf{E}_\pi(c\pi + w(\mu, d)),}$$

where $\mathsf{P}_\pi = \pi^1\mathsf{P}^1 + \pi^0\mathsf{P}^0 + \pi^2\mathsf{P}^2$.

It is easy to see that

$$R_\delta(\pi) \geq \mathsf{E}_\pi\Big(c\tau + \underbrace{\min\Big\{a_{10}\pi_\tau^1 + a_{20}\pi_\tau^2,\ a_{01}\pi_\tau^0 + a_{21}\pi_\tau^2,\ a_{02}\pi_\tau^0 + a_{12}\pi_\tau^1\Big\}}_{=:\, G(\pi_\tau^1,\, \pi_\tau^0,\, \pi_\tau^2)}\Big).$$

Then

$$\inf_{\delta=(\tau,d)} R_\delta(\pi) = \inf_\tau R_{(\tau,d^*)}(\pi) = \inf_\tau \mathsf{E}_\pi(c\tau + G(\pi_\tau^1, \pi_\tau^0, \pi_\tau^2)),$$

where d^* $(= d^1, d^0, d^2)$ is determined from the relationship

$$I_{\{d^0\}}(d^*)[a_{10}\pi^1 + a_{20}\pi^2] + I_{\{d^1\}}(d^*)[a_{01}\pi^0 + a_{21}\pi^2] \\ + I_{\{d^2\}}(d^*)[a_{02}\pi^0 + a_{12}\pi^1] = G(\pi^1, \pi^0, \pi^2).$$

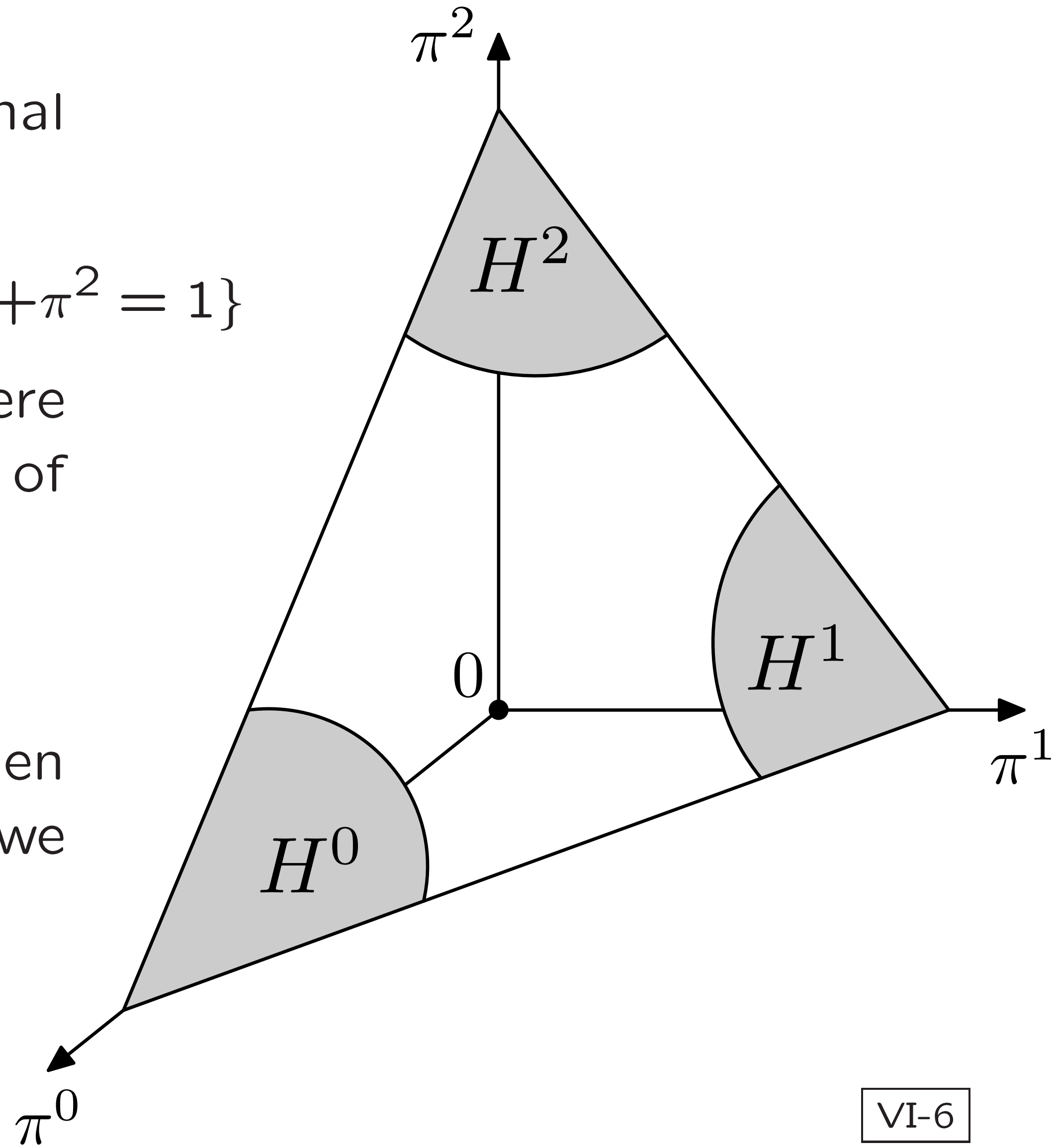
From general theory of optimal stopping it follows that the set

$$\Pi = \{0 \le \pi^1, \pi^0, \pi^2 \le 1 \colon \pi^1 + \pi^0 + \pi^2 = 1\}$$

is such that $\Pi = C^* + D^*$, where C^* is the set of continuation of observations and

$$D^* = D^1 + D^0 + D^2.$$

If $\pi_{\tau^*} \equiv (\pi^1_{\tau^*}, \pi^0_{\tau^*}, \pi^2_{\tau^*}) \in D^i$, then we make decision d_i^*, i.e., we accept hypothesis H^i.



For simplicity, assume that

$$\pi^1 = \pi^0 = \pi^2 = 1/3,$$
$$a_{ij} = 1, \quad i \neq j; \qquad a_{ii} = 0, \qquad \text{(symmetrical case).}$$
$$\mu^1 = -1, \quad \mu^0 = 0, \quad \mu^2 = 1$$

In this case $G(\pi^1, \pi^0, \pi^2) = \min\{\pi^1+\pi^2,\ \pi^0+\pi^2,\ \pi^0+\pi^1\}$, or, in terms of (φ^1, φ^2),

$$G(\varphi^1, \varphi^2) = \frac{\min\{\varphi^1+\varphi^2,\ 1+\varphi^2,\ 1+\varphi^1\}}{1 + \varphi^1 + \varphi^2}.$$

By symmetry, $d\varphi_t^1 = -\varphi_t^1\, dX_t$ and $d\varphi_t^2 = \varphi_t^2\, dX_t$. So, $\varphi_t^1 = e^{-X_t - t/2}$, $\varphi_t^2 = e^{X_t + t/2}$, and in terms of (t, x) the terminal risk function is

$$G(t, x) = \frac{\min\{1+e^{-x-t/2},\ 1+e^{x-t/2},\ e^{-x-t/2}+e^{x-t/2}\}}{1 + e^{-t/2}(e^{-x} + e^{x})}.$$

Our basic Bayesian problem of **sequential testing of three hypotheses** H_1 $(\mu = -1)$, H_0 $(\mu = 0)$, H_2 $(\mu = 1)$

IS REDUCED TO

the optimal stopping problem: to find a stopping time τ^* such that

$$R_{\tau^*} \equiv R_{\tau^*}(1/3, 1/3, 1/3) = \inf_{\tau} \mathsf{E}(c\tau + G(\tau, X_\tau)),$$

where

$$G(t,x) = \frac{\min\{1 + e^{-x-t/2}, 1 + e^{x-t/2}, e^{-x-t/2} + e^{x-t/2}\}}{1 + e^{-t/2}(e^{-x} + e^{x})}.$$

.

The infinitesimal operator of (t, X_t) is

$$L_{(t,x)} = \frac{\partial}{\partial t} + A(t,x)\frac{\partial}{\partial x} + \frac{1}{2}\frac{\partial^2}{\partial x^2},$$

where

$$A(t,x) = \frac{e^{x-t/2} - e^{-x-t/2}}{1 + e^{x-t/2} + e^{-x-t/2}} = \frac{e^{-t/2}(e^x - e^{-x})}{1 + e^{-t/2}(e^x + e^{-x})}.$$

It is interesting to compare these formulas with the corresponding formulas for problem of testing **TWO** hypotheses, say

$$\boxed{H_0\colon \mu = \mu_0 = 0} \quad \text{and} \quad \boxed{H_1\colon \mu = \mu_2 = 1}$$

For this case, denoting

$$R_\delta(\pi) = \mathsf{E}_\pi(c\tau + W(\mu, d)), \qquad \delta = (\tau, d),$$

with $\pi = (\pi^0, \pi^2)$, $\pi^0 = \mathsf{P}(\mu = 0) = 1/2$, $\pi^2 = \mathsf{P}(\mu = 1) = 1/2$, $\pi^0 + \pi^2 = 1$, and $W(\mu^i, d^i) = 0$ and $W(\mu^i, d^j) = 1$, $i \neq j$, we find that

$$\boxed{R_\delta(\pi) = \mathsf{E}_\pi\Big(c\tau + G(\pi_\tau^0, \pi_\tau^2)\Big)}$$

where

$$\pi_t^0 = \mathsf{P}(\mu = \mu^0 \,|\, \mathcal{F}_t^X), \quad \pi_t^2 = \mathsf{P}(\mu = \mu^2 \,|\, \mathcal{F}_t^X).$$

Introduce $\varphi_t^2 = \pi_t^2/\pi_t^0$. Then

$$\varphi_t^2 = e^{X_t - t/2}, \qquad G(\pi_\tau^1, \pi_\tau^2) = G(t, X_t) = \frac{\min(1, e^{X_t - t/2})}{1 + e^{X_t - t/2}}.$$

Innovation representation for $X = (X_t)_{t \ge 0}$ is

$$dX_t = A(t, X_t)\,dt + d\overline{B}_t, \quad \text{where} \quad A(t,x) == \frac{e^{x-t/2}}{1+e^{x-t/2}}.$$

For $G(t,x)$ we have

$$G(t,x) \;=\; \begin{cases} \dfrac{1}{1+e^{x-t/2}}, & x \ge \dfrac{t}{2}, \\[2ex] \dfrac{e^{x-t/2}}{1+e^{x-t/2}}, & x \le \dfrac{t}{2}. \end{cases}$$

If $x \neq t/2$, then we directly find that

$$G(t,x) \in C^{1,2} \quad \text{and} \quad L_{(t,x)}G(t,x) = 0,$$

where $L_{(t,x)}$ is the infinitesimal operator of X:

$$L_{(t,x)} = \frac{\partial}{\partial t} + A(t,x)\,\frac{\partial}{\partial x} + \frac{1}{2}\,\frac{\partial^2}{\partial x^2}.$$

Applying the generalized Itô formula, we get

$$\begin{aligned} G(t,X_t) = G(0,X_0) &+ \int_0^t L_{(s,x)}G(s,X_s)I\Big(X_s \neq \frac{s}{2}\Big)\,ds \\ &+ \int_0^t \frac{\partial G}{\partial x}(s,X_s)I\Big(X_s \neq \frac{s}{2}\Big)\,d\overline{B}_s \\ &+ \frac{1}{2}\int_0^t \left[\frac{\partial G}{\partial x}(s,X_{s+}) - \frac{\partial G}{\partial x}(s,X_{s-})\right] I\Big(X_s \neq \frac{s}{2}\Big)\,dL_s^+ \end{aligned}$$

where L_s^+ is a local time of X on the ray $x = t/2$, $t \leq s$.

For $x > t/2$: $\quad \dfrac{\partial G}{\partial x} = -\dfrac{e^{x-t/2}}{(1+e^{x-t/2})^2}$ and $\left.\dfrac{\partial G}{\partial x}\right|_{x\downarrow t/2} = -\dfrac{1}{4}.$

For $x < t/2$: $\quad \dfrac{\partial G}{\partial x} = \dfrac{e^{x-t/2}}{(1+e^{x-t/2})^2}$ and $\left.\dfrac{\partial G}{\partial x}\right|_{x\uparrow t/2} = \dfrac{1}{4}.$

Thus, $\left.\left[\dfrac{\partial G}{\partial x}(s, X_{s+}) - \dfrac{\partial G}{\partial x}(s, X_{s-})\right]\right|_{X_s = s/2} = -\dfrac{1}{2}$, and (with $\mathsf{E} = \mathsf{E}_{(1/2,1/2)}$) we have

$$\mathsf{E}G(\tau, X_\tau) = G(0, X_0) - \frac{1}{4}\mathsf{E}L_\tau^+,$$

$$\inf_\tau \mathsf{E}\Big[c\tau + G(\tau, X_\tau)\Big] = G(0, X_0) + \inf_\tau \mathsf{E}\Big[c\tau - \frac{1}{4}L_\tau^+\Big],$$

where L_s^+ is a local time on $x = t/2$, $t \le s$. From here we find that the set $\{(t,x)\colon x = t/2,\ t \ge 0,\ -\infty < x < \infty\}$ belongs to the set C of continuation of observations.

Note that $L_t^+ = \left|X_t - \frac{t}{2}\right| - \int_0^t \mathrm{sgn}\left(X_s - \frac{s}{2}\right) dX_s$. Thus,

$$\mathsf{E}L_\tau^+ = \mathsf{E}\left|X_\tau - \frac{\tau}{2}\right| - \mathsf{E}\int_0^\tau \mathrm{sgn}\left(X_s - \frac{s}{2}\right) \frac{e^{X_s - s/2}}{1 + e^{X_s - s/2}}\, ds$$

so that the problem of testing two hypotheses

$$H_0\colon \mu = 0 \quad \text{and} \quad H_2\colon \mu = 1$$

is reduced to the following optimal stopping problem:

$$\inf_\tau \mathsf{E}\Big[c\tau - \frac{1}{4}\mathsf{E}L_\tau^+\Big] = \inf_\tau \mathsf{E}\left[c\tau - \left|X_\tau - \frac{\tau}{2}\right| + \mathsf{E}\int_0^\tau \mathrm{sgn}\left(X_s - \frac{s}{2}\right) \frac{e^{X_s - s/2}}{1 + e^{X_s - s/2}}\, ds\right]$$

for the process $X = (X_t)_{t \geq 0}$ with

$$dX_t = \frac{e^{X_t - t/2}}{1 + e^{X_t - t/2}}\, ds + d\overline{B}_t.$$

If $\overline{X}_t = X_t - t/2$, then this stopping problem is reduced to the optimal stopping problem

$$\boxed{V_x = \sup_{\tau} \mathsf{E}_x\Bigg(|\overline{X}_\tau| - c\tau - \int_0^\tau \operatorname{sgn} \overline{X}_s \cdot \frac{e^{\overline{X}_s - s/2}}{1 + e^{\overline{X}_s - s/2}}\, ds\Bigg)}$$

where

$$d\overline{X}_s = \frac{e^{\overline{X}_s} - 1}{2(e^{\overline{X}_s} + 1)}\, ds + d\overline{B}_s, \qquad \overline{X}_0 = x.$$

Similarly, for the problem of testing two hypotheses

$$\boxed{H_1\colon \mu = -1} \quad \text{and} \quad \boxed{H_2\colon \mu = 1}$$

we find that the problem $\inf_\delta R_\delta(\pi)$ for case $\pi^1 = \pi^2 = 1/2$ is reduced to the problem

$$\inf_\tau \mathsf{E}\Big[c\tau - \frac{1}{2}L^0_\tau\Big],$$

where L^0_t is a local time in zero of the process $X = (X_s)_{s\ge 0}$ with

$$dX_s = \frac{e^{X_s - s/2} - e^{-X_s - s/2}}{e^{X_s - s/2} + e^{-X_s - s/2}}\, ds + d\overline{B}_s.$$

Note that

$$\overline{X}_t = \log \frac{\pi^2_t}{1-\pi^2_t}, \qquad \text{where} \quad \pi^2_t = \mathsf{P}\Big(\mu = \mu^2 \,|\, \mathcal{F}^X_t\Big).$$

Let us now return to the problem of testing **THREE** hypotheses.

2°. Properties of $G(t,x)$. For $x > t/2$ we have

$$G(t,x) = \frac{1 + e^{-x-t/2}}{1 + e^{-t/2}(e^x + e^{-x})}.$$

So,

$$\frac{\partial G}{\partial t} = \frac{e^{x-t/2}/2}{[1 + e^{-t/2}(e^x + e^{-x})]^2}, \qquad \frac{\partial G}{\partial x} = \frac{2e^{-t} + e^{x-t/2}}{[1 + e^{-t/2}(e^x + e^{-x})]^2},$$

$$\frac{\partial^2 G}{\partial x^2} = \frac{-e^{x-t/2}[(1+e^{-t/2})(e^x+e^{-x}) + (2e^{-t}+e^{x-t/2})e^{-t/2}(e^x+e^{-x})]}{[1 + e^{-t/2}(e^x + e^{-x})]^4}.$$

From these formulas we find that (for $x > t/2$)

$$\boxed{L_{(t,x)}G(t,x) = 0}.$$

In a similar way, we get the same relationship for $x < -t/2$ and for $|x| < t/2$.

Applying the generalized Itô formula to $G = G(t,x)$, we find that

$$
\begin{aligned}
G(t, X_t) = G(0, X_0) &+ \int_0^t L_{(s,x)} G(s, X_s) I\Big(X_s \neq \pm\frac{s}{2}\Big)\, ds \\
&+ \int_0^t \frac{\partial G}{\partial x}(s, X_s)\, I\Big(X_s \neq \pm\frac{s}{2}\Big)\, d\overline{B}_s \\
&+ \frac{1}{2}\int_0^t \Big[\frac{\partial G}{\partial x}(s, X_s+) - \frac{\partial G}{\partial x}(s, X_s-)\Big]\, I\Big(X_s = \frac{s}{2}\Big)\, dL_s^+ \\
&+ \frac{1}{2}\int_0^t \Big[\frac{\partial G}{\partial x}(s, X_s+) - \frac{\partial G}{\partial x}(s, X_s-)\Big]\, I\Big(X_s = -\frac{s}{2}\Big)\, dL_s^-,
\end{aligned}
$$

where L^+, L^- are local times of X on the rays $x = s/2$, $x = -s/2$.

Straightforward calculations show that

$$
\begin{aligned}
\frac{\partial G}{\partial x}(s, X_s+) - \frac{\partial G}{\partial x}(s, X_s-)\Big|_{X_s = s/2} &= -\frac{1}{2 + e^{-s}}, \\
\frac{\partial G}{\partial x}(s, X_s+) - \frac{\partial G}{\partial x}(s, X_s-)\Big|_{X_s = -s/2} &= -\frac{1}{2 + e^{-s}}.
\end{aligned}
$$

So, if $L_s = L_s^+ + L_s^-$, then for any stopping time τ such that $\mathsf{E}\tau < \infty$ we find that

$$\boxed{\mathsf{E}G(\tau, X_\tau) = G(0, X_0) - \frac{1}{2}\mathsf{E}\int_0^\tau \frac{dL_s}{2 + e^{-s}}}.$$

If $X_0 = 0$, then $G(0,0) = 2/3$. Since $1/3 \le 1/(2 + e^{-s}) \le 1/2$, we have

$$\frac{1}{4}\mathsf{E}(4c\tau - L_\tau) \le \mathsf{E}(c\tau + G(\tau, X_\tau)) \le \frac{1}{6}\mathsf{E}(6c\tau - L_\tau).$$

From here we conclude that solution of the problems

$$\inf_\tau \mathsf{E}(c\tau - L_\tau)$$

for different values of $c > 0$ will give a quite good approximation for our problem of finding

$$\inf_\tau \mathsf{E}\left(c\tau - \frac{1}{2}\int_0^\tau \frac{dL_s}{2 + e^{-s}}\right).$$

REMARK. The solutions of the problems

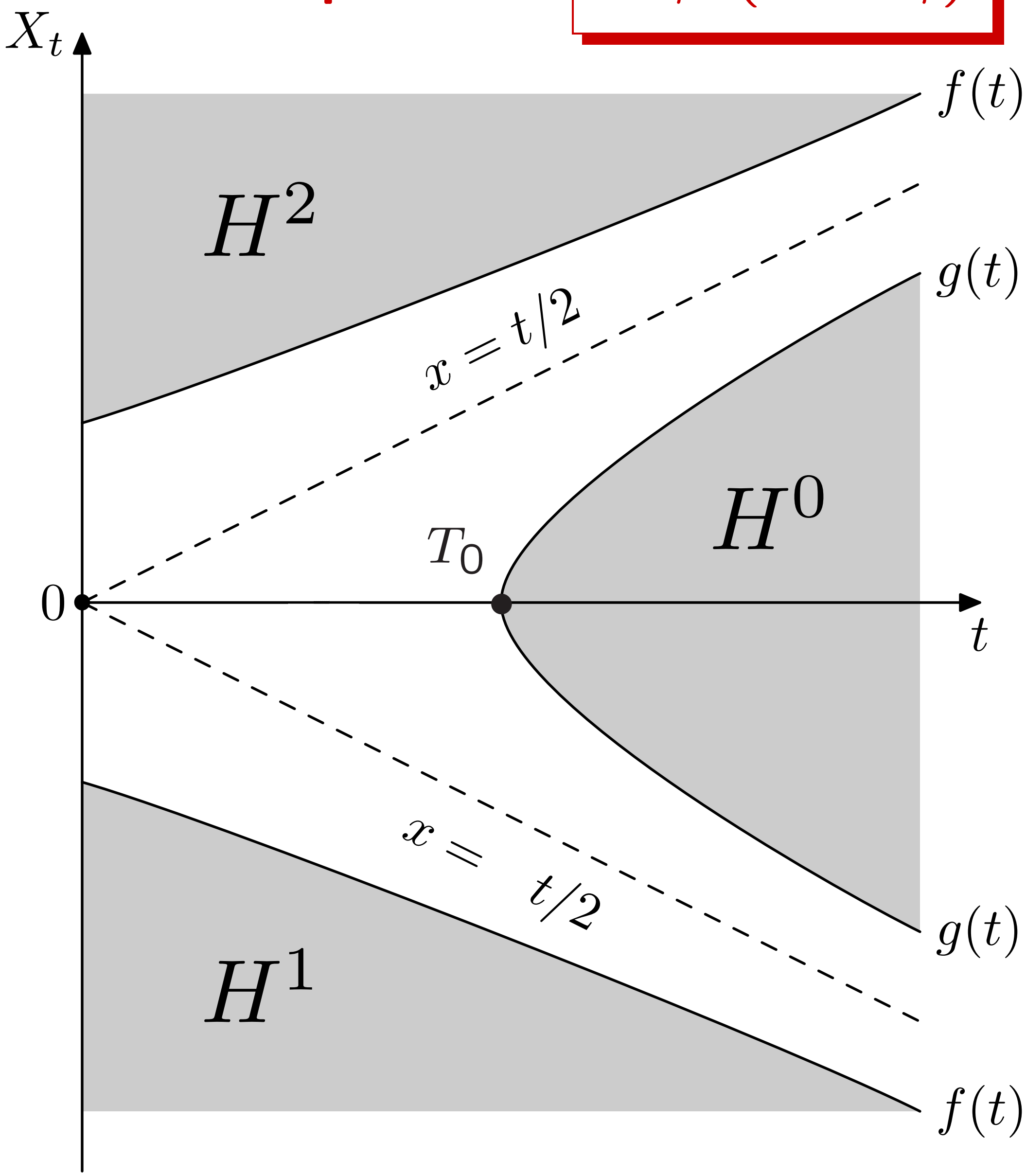
$$\inf_{\tau \geq t} \mathsf{E}\left(c(\tau - t) - \frac{1}{2} \int_0^\tau \frac{dL_s}{2 + e^{-s}} \right)$$

and

$$\inf_{\tau \geq t} \mathsf{E}\left(c(\tau - t) - \frac{1}{4} L_\tau \right)$$

are asymptotically $(t \to \infty)$ indistinguishable.

3°. The problem $\boxed{\inf_\tau \mathsf{E}(c\tau - L_\tau)}$. Here $L_t = L_t^+ + L_t^-$, where $L^\pm$ are local times on the rays $x = \pm t/2$. The presence of **local times** suggest the following picture (D^1, D^0, and D^2 are sets of stopping with accepting the hypotheses H^1, H^0, and H^2). The local times L^+ and L^- are local times accumulated along the rays $x = \pm t/2$. So, these rays and areas near these rays should belong to the set of continuation of observations.



THEOREM 1. Let $f = f(t)$ and $g = g(t)$ be upper and lower optimal stopping boundaries in the half-plane of positive x and $t > 0$. Then for **large** t

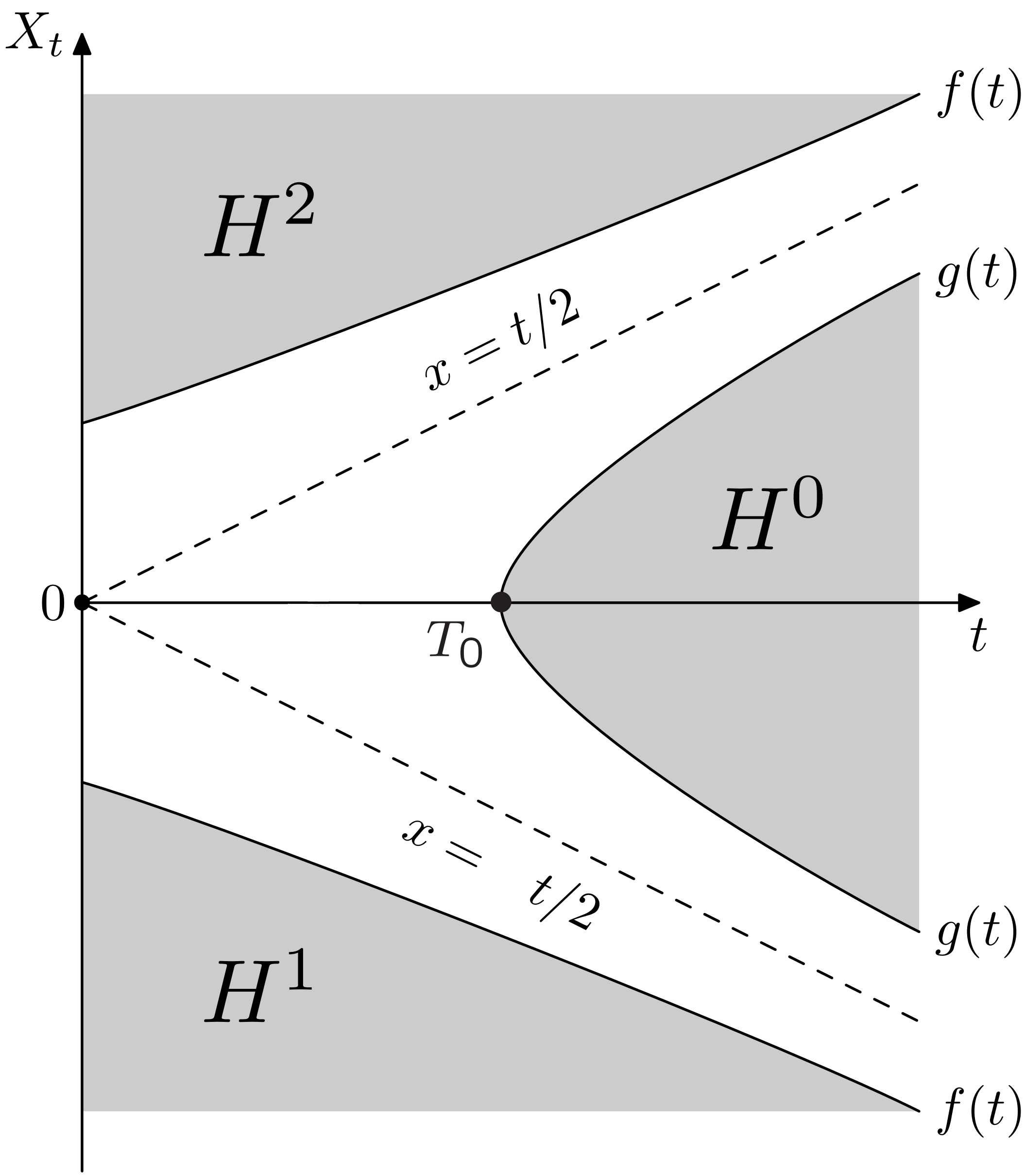
$$\boxed{f(t) = \frac{t}{2} + A + O(e^{-t}), \qquad g(t) = \frac{t}{2} - A + O(e^{-t})},$$

where constant A is a unique solution of the equation

$$e^{A} - e^{-A} + 2A = 2c^{-1}.$$

(So, $A \sim (2c)^{-1}$ for small c.)

For $x < 0$ and $t > 0$, the upper and lower functions are $-g(t)$ and $-f(t)$.



This result says that if we did not stop before time T_0, then the problem of testing of three hypotheses is split into two problems of testing of **two** statistical hypotheses

H^2 and H^0 (if $X_{T_0} > 0$)

and

H^1 and H^0 (if $X_{T_0} < 0$).

The following important result gives **integral** equations for optimal boundaries $f(t)$ and $(g(t)$.

THEOREM 2. There exists $T_0 > 0$ such that for all $t > T_0$

$$f(t) = \frac{t}{2} + \widehat{f}(t), \quad g(t) = \frac{t}{2} + \widehat{g}(t),$$

where $\widehat{f}(t)$ and $\widehat{g}(t)$ satisfy the following integral equations:

$$\boxed{\begin{aligned} &c\int_t^\infty G_1\Big(t, \widehat{f}(t); s, \widehat{g}(s), \widehat{f}(s)\Big)\, ds = G_2(t, \widehat{f}(t)), \\ &c\int_t^\infty G_1\Big(t, \widehat{g}(t); s, \widehat{g}(s), \widehat{f}(s)\Big)\, ds = G_2(t, \widehat{g}(t)) \end{aligned}}$$

with

$$\lim_{t\to\infty} \widehat{f}(t) = A, \quad \lim_{t\to\infty} \widehat{g}(t) = -A, \quad e^A - e^{-A} + 2A = \frac{2}{c}.$$

The functions G_1 and G_2 are given by

$$G_1(t,x;s,a,b) = \left[\sum_{i=0}^{2} \exp\left\{x\theta^i - \frac{(\theta^i)^2 t}{2}\right\}\right]^{-1} \sum_{i=0}^{2} \exp\left\{x\theta^i - \frac{(\theta^i)^2 t}{2}\right\}$$
$$\times \left[\Phi_{s-t}(b - x - \theta^i(s-t)) - \Phi_{s-t}(a - x - \theta^i(s-t))\right],$$

$$G_2(t,x) = \left[\sum_{i=0}^{2} \exp\left\{x\theta^i - \frac{(\theta^i)^2 t}{2}\right\}\right]^{-1} \sum_{i=0}^{2} \exp\left\{x\theta^i - \frac{(\theta^i)^2 t}{2}\right\} G_3(x,\theta^i),$$

where

$$G_3(x,\theta) = \theta \int_0^\infty \left(\theta\sqrt{r} - \frac{x}{\sqrt{r}}\right)\varphi\left(\theta\sqrt{r} + \frac{x}{\sqrt{r}}\right) dr - \begin{cases} |x| - x, & \theta > 0 \\ |x| + x, & \theta < 0 \end{cases},$$

$$\theta^i = \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}.$$

The critical point T_0 is determined from the equation $g(T_0) = 0$, i.e., $T_0 = \sup\{t \colon \widehat{g}(t) + t/2 = \theta\}$.

PROOF of Theorem 1: We have to prove that

$$f(t)=\frac{t}{2}+A+O(e^{-t}),\qquad g(t)=\frac{t}{2}-A+O(e^{-t}), \tag{177}$$

where $e^{A}-e^{-A}+2A=2/c$. Consider a new problem:

$$S^{+}(x)=\sup_{\tau}\mathsf{E}_{x}(L^{+}_{\tau}-c\tau),$$

where $L^{+}=(L^{+}_{t})_{t\ge 0}$ is a local time along the ray $x=t/2$ of the process $X=(X_t)_{t\ge 0}$ with

$$dX_t=\frac{e^{-t/2}(e^{x}-e^{-x})}{1+e^{-t/2}(e^{x}+e^{-x})}\,dt+d\overline{B}_t. \tag{178}$$

We also introduce

$$S^{+}(t,x)=\sup_{\tau\in\mathfrak{M}_t}\mathsf{E}_{t,x}\Big[L^{+}_{\tau}(t,X)-c(\tau-t)\Big],$$

where $\mathfrak{M}_t=\{\tau\colon \tau$ is a stopping time w.r.t. $X,\ \tau\ge t\}$, $L^{+}_{\tau}(t,X)$ is a local time of X on the ray $x=t/2$ on $[t,\tau]$.

By the Itô–Tanaka formula,

$$L_s^{+}(t,X) = \left|X_s - \frac{s}{2}\right| - |X_t| - \int_t^s \operatorname{sgn}\Big(X_u - \frac{u}{2}\Big)\,dX_u, \qquad s > t. \tag{179}$$

Transform our process X_t into the process $\widetilde{X}_t = X_t - t/2$, for which

$$d\widetilde{X}_t = \left(\frac{e^{\widetilde{X}_t} - e^{-\widetilde{X}_t - t}}{1 + e^{\widetilde{X}_t} + e^{-\widetilde{X}_t - t}} - \frac{1}{2}\right) dt + d\overline{B}_t. \tag{180}$$

The problem $S^{+}(t,X)$ for X is transformed into the problem

$$\widetilde{S}^{\circ}(t,x) = \sup_{\tau \in \mathfrak{M}_t} \mathsf{E}_{t,x}\Big[L_\tau^{\circ}(t,\widetilde{X}) - c(\tau - t)\Big],$$

where $L_\tau^{\circ}(t,\widetilde{X})$ is a local time of $\widetilde{X}$ in zero on $[t,\tau]$. It is clear that $S^{+}(t,x) = \widetilde{S}^{\circ}(t, x - t/2)$. Let us show that for this new problem the set $\widetilde{C}$ of continuation of observations has the form

$$\boxed{\widetilde{C} = \{(t,x) \colon \tilde{g}(t) < x < \tilde{f}(t)\}},$$

where $\tilde{g} = \tilde{g}(t)$ and $\tilde{f} = \tilde{f}(t)$ are unknown functions.

To prove this, note that points $(t,0)$ belong to the set $\widetilde{C}$ by the following "$\sqrt{\Delta}$-property" of a Brownian motion.

Indeed, for a Brownian motion $\overline{B}$ we have

$$\mathsf{Law}(L^{\circ}_{t+\Delta}(t,\overline{B})) = \mathsf{Law}(|\overline{B}_{\Delta}|).$$

Thus, $\mathsf{E}_{t,0} L^{\circ}_{t+\Delta}(t,\widetilde{X})$ has order $\sqrt{\Delta}$ for small Δ, moreover, for such small Δ

$$\mathsf{E}_{t,0}[L^{\circ}_{t+\Delta}(t,\widetilde{X}) - c\Delta] > 0.$$

So, $(t,0) \in \widetilde{C}$ (if we stop immediately, then our gain equals zero). Define

$$\tilde{f}(t) = \inf\{x > 0 \colon (t,x) \notin \widetilde{C}\}, \qquad \tilde{g}(t) = \sup\{x < 0 \colon (t,x) \notin \widetilde{C}\}.$$

Since $\widetilde{C}$ is open (it follows from the general theory of optimal stopping for Markov processes), we have

$$\tilde{f}(t) > 0, \quad \tilde{g}(t) < 0 \quad \text{for all } t > 0.$$

One can prove that $(t,x)\notin \widetilde{C}$ for all $x\ge \tilde f(t)$ and $x\le \tilde g$. From this it follows that $\widetilde{C}=\{(t,x)\colon \tilde g(t)<x<\tilde f(t)\}$.

Let us study the behavior of boundaries $\tilde f(t)$ and $\tilde g(t)$ as $t\to\infty$. For this purpose we approximate the process $\widetilde{X}$ with differential (180):

$$d\widetilde{X}_t=\left(\frac{e^{\widetilde{X}_t}-e^{-\widetilde{X}_t-t}}{1+e^{\widetilde{X}_t}+e^{-\widetilde{X}_t-t}}-\frac{1}{2}\right)dt+d\overline{B}_t,$$

by the process $\widetilde{X}^{\varepsilon}$ with differential

$$d\widetilde{X}^{\varepsilon}_t=\left(\frac{e^{\widetilde{X}^{\varepsilon}_t}}{1+e^{\widetilde{X}^{\varepsilon}_t}}-\frac{1}{2}-\varepsilon\,\mathrm{sgn}\,\widetilde{X}^{\varepsilon}_t\right)dt+d\overline{B}_t.$$

Fix (large) $T>0$ and put $\varepsilon_1=e^{T}$. Note that

$$\left(\frac{e^{x}}{1+e^{x}}-\frac{1}{2}-\varepsilon_1\,\mathrm{sgn}\,x\right)\mathrm{sgn}\,x\le\left(\frac{e^{x}-e^{-x-t}}{1+e^{x}+e^{-x-t}}-\frac{1}{2}\right)\mathrm{sgn}\,x. \qquad (181)$$

For the process $\widetilde{X}^{\varepsilon_1}$ consider the problem

$$\widetilde{S}^{\circ}_{\varepsilon_1}(t,x) = \sup_{\tau\in\mathfrak{M}_t} \mathsf{E}_{t,x}\Big[L^{\circ}_{\tau}(t,\widetilde{X}^{\varepsilon_1}) - c(\tau - t)\Big].$$

The process $\widetilde{X}^{\varepsilon_1}$ is **homogeneous** (!). Thus, the set of continuation of observations has the form

$$\widetilde{C}_{\varepsilon_1} = \{(t,x)\colon \tilde{g}_{\varepsilon_1} < x < \tilde{f}_{\varepsilon_1}\},$$

where $\tilde{g}_{\varepsilon_1}$ and $\tilde{f}_{\varepsilon_1}$ do not depend on t. From the majorization property (181) it follows that

$$\tilde{g}(t) \geq \tilde{g}_{\varepsilon_1}, \quad \tilde{f}(t) \leq \tilde{f}_{\varepsilon_1}. \tag{182}$$

For the problem of testing of **two** statistical hypotheses controlled by the processes $\widetilde{X}^{\varepsilon_1}$, we know that

$$\tilde{g}_{\varepsilon_1} = -A + O(\varepsilon_1), \quad \tilde{f}_{\varepsilon_1} = A + O(\varepsilon_1). \tag{183}$$

From (182) and (183) it follows that

$$\tilde{g}(t) \geq -A + O(e^{-T}), \quad \tilde{f}(t) \leq A + O(e^{-T}). \tag{184}$$

To obtain the inverse inequalities, we take the process $\widetilde{X}^{\varepsilon_2}$ with $\varepsilon_2 = e^{A-T}$. Then

$$\tilde{g}_{\varepsilon_2} < x < \tilde{f}_{\varepsilon_2}.$$

Similarly to (183), we find that

$$\tilde{g}_{\varepsilon_2} = -A + O(\varepsilon_2), \quad \tilde{f}_{\varepsilon_2} = A + O(\varepsilon_2).$$

Because

$$\left(\frac{e^x - e^{-x-t}}{1 + e^x + e^{-x-t}} - \frac{1}{2}\right) \operatorname{sgn} x \le \left(\frac{e^x}{1+e^x} - \frac{1}{2} + \varepsilon_2 \operatorname{sgn} x\right) \operatorname{sgn} x,$$

we find that

$$\tilde{g}(t) \le \tilde{g}_{\varepsilon_2}, \quad \tilde{f}(t) \ge \tilde{f}_{\varepsilon_2}, \qquad t \ge T.$$

So,

$$\tilde{g}(t) \le -A + O(e^{-t}), \quad \tilde{f}(t) \ge A + O(e^{-t}). \tag{185}$$

Together with (184) this gives

$$\boxed{\tilde{g}(t) = -A + O(e^{-t}), \quad \tilde{f}(t) = A + O(e^{-t}).}$$

From these properties it follows that for the process X we have

$$\boxed{g(t) = \frac{t}{2} - A + O(e^{-t}), \quad f(t) = \frac{t}{2} + A + O(e^{-t}), \qquad t \to \infty.}$$

The final step of the proof consists in demonstration that the "lower" boundaries $g_{\text{low}}(t)$ and $f_{\text{low}}(t)$ are such that

$$g_{\text{low}}(t) = -g(t), \quad f_{\text{low}}(t) = -f(t).$$

Theorem 1 is proved.

PROOF of Theorem 2. First, we get integral equations for boundaries $\tilde{g}(t)$ and $\tilde{f}(t)$ in case of the optimal stopping problem

$$S^{+}(t,x) = \sup_{\tau \in \mathfrak{M}_t} \mathsf{E}_{t,x}\Big[L_{\tau}^{+}(t,X) - c(\tau - t)\Big]. \tag{186}$$

Again put $\widetilde{X}_t = X_t - \dfrac{t}{2}$, $\quad \tilde{b}(t,x) = \dfrac{e^{x} - e^{-x-t}}{1 + e^{x} + e^{-x-t}} - \dfrac{1}{2}$. Then

$$d\widetilde{X}_t = \tilde{b}(t,\widetilde{X}_t)\,dt + d\overline{B}_t.$$

Denote also $\widetilde{H}(t,x) = \widetilde{S}^{\circ}(t,x)$, i.e.,

$$\widetilde{H}(t,x) = \sup_{\tau \in \mathfrak{M}_t} \mathsf{E}_{t,x}\Big[L_{\tau}^{\circ}(t,\widetilde{X}) - c(\tau - t)\Big]. \tag{187}$$

By the Itô–Tanaka formula,

$$\begin{aligned}
\widetilde{H}(t,x) &= \sup_{\tau \in \mathfrak{M}_t} \mathsf{E}_{t,x}\left[|\widetilde{X}_\tau| - |x| - \int_t^\tau \operatorname{sgn} \widetilde{X}_s\, d\widetilde{X}_s - c(\tau - t)\right] \\
&= \sup_{\tau \in \mathfrak{M}_t} \mathsf{E}_{t,x}\Bigg[|\widetilde{X}_\tau| - \int_t^\tau \operatorname{sgn} \widetilde{X}_s\, \tilde{b}(t,\widetilde{X}_t)\, ds \\
&\qquad\qquad - \int_t^\tau \operatorname{sgn} \widetilde{X}_s\, d\overline{B}_s - c(\tau - t)\Bigg].
\end{aligned}$$

We can prove that

$$\mathsf{E}_{t,x}\int_t^{\tau} \operatorname{sgn}\widetilde{X}_s\, d\overline{B}_s = 0 \quad \text{for } \tau \in \mathfrak{M}_t \text{ such that } \mathsf{E}_{t,x}\tau < \infty.$$

Denote $\widehat{H}(t,x) = \widetilde{H}(t,x) + |x|$. Itô's formula applied to $\widehat{H}(t,x) \in C^{1,1}$ gives

$$\widehat{H}(T,\widetilde{X}_T) = \widehat{H}(t,\widetilde{X}_t) + \int_t^T \Big(\frac{\partial \widehat{H}}{\partial s}(s,\widetilde{X}_s) + L_{\widetilde{X}}\widehat{H}(s,\widetilde{X}_s)\Big)\, ds + \text{Mart.},$$

where “Mart.” stands for a martingale with zero mean. Since

$$\frac{\partial \widehat{H}}{\partial s} + L_{\widetilde{X}}\widehat{H} = \begin{cases} \tilde{b}(s,x)\operatorname{sgn} x + c & \text{in the set } C \text{ of continuation of observations,} \\ \tilde{b}(s,x)\operatorname{sgn} x & \text{in the set of stopping } D \cap \{x \neq 0\}, \end{cases}$$

we have

$$\widehat{H}(T,\widetilde{X}_T) = \widehat{H}(t,\widetilde{X}_t) + \int_t^T \tilde{b}(s,\widetilde{X}_s)\operatorname{sgn}\widetilde{X}_s\, ds + c\int_t^T I\{\tilde{g}(s) < \widetilde{X}_s < \tilde{f}(s)\}\, ds + \text{Mart.}$$

Take $\widetilde{H}(t,x) = \widehat{H}(t,x) - |x|$. Then

$$\widetilde{H}(T,\widetilde{X}_T) = |\widetilde{X}_t| - |\widetilde{X}_T| + \widehat{H}(t,\widetilde{X}_t) + \int_t^T \tilde{b}(s,\widetilde{X}_s)\,\mathrm{sgn}\,\widetilde{X}_s\,ds$$

$$+ c\int_t^T I\{\tilde{g}(s) < \widetilde{X}_s < \tilde{f}(s)\}\,ds + \text{Mart.}$$

$$= \widehat{H}(t,\widetilde{X}_t) - L_T^\circ(t,\widetilde{X})$$

$$+ c\int_t^T I\{\tilde{g}(s) < \widetilde{X}_s < \tilde{f}(s)\}\,ds + \text{Mart.}$$

Taking $\mathsf{E}_{t,x}$ for $x = \tilde{f}(t)$ and $x = \tilde{g}(t)$, we get

$$\boxed{\begin{aligned}\mathsf{E}_{t,\tilde{f}(t)}\widetilde{H}(T,\widetilde{X}_T) &= -\mathsf{E}_{t,\tilde{f}(t)}L_T^{\circ}(t,\widetilde{X}) + c\int_t^T \mathsf{P}_{t,\tilde{f}(t)}\{\tilde{g}(s)<\widetilde{X}_s<\tilde{f}(s)\}\,ds\\ \mathsf{E}_{t,\tilde{g}(t)}\widetilde{H}(T,\widetilde{X}_T) &= -\mathsf{E}_{t,\tilde{g}(t)}L_T^{\circ}(t,\widetilde{X}) + c\int_t^T \mathsf{P}_{t,\tilde{g}(t)}\{\tilde{g}(s)<\widetilde{X}_s<\tilde{f}(s)\}\,ds\end{aligned}} \tag{188}$$

(we used that $\widetilde{H}(t,f(t)) = 0$, $\widetilde{H}(t,g(t)) = 0$, $\mathsf{E}_{t,x}(\text{Mart.}) = 0$). Since

$$\begin{aligned}|\mathsf{E}_{t,\tilde{f}(t)}\widetilde{H}(T,\widetilde{X}_T)| &= \Big|\mathsf{E}_{t,\tilde{f}(t)}\widetilde{H}(T,\widetilde{X}_T)I\{\tilde{g}(t) \le \widetilde{X}_T \le \tilde{f}(t)\}\Big|\\ &\le c\mathsf{P}\{\tilde{g}(t) \le \widetilde{X}_T \le \tilde{f}(t)\} \to 0, \quad T \to \infty,\end{aligned}$$

and similarly $\mathsf{E}_{t,\tilde{g}(t)}\widetilde{H}(T,\widetilde{X}_T) \to 0$, $T \to \infty$, we see that the left-hand sides in (188) tend to zero as $T \to \infty$.

Consider now other terms in (188).

Denote $G_1(t,x;s,a,b) = \mathsf{P}_{t,x}(a \le \widetilde{X}_s \le b)$, where $\widetilde{X}_t = X_t - t/2$ and $X_t = \mu t + B_t$. So, $\widetilde{X}_t = \theta t + B_t$, where $\mathsf{P}(\theta = \theta^i) = 1/3$,

$$\theta^1 = \mu^1 - \tfrac{1}{2} = \tfrac{1}{2}, \quad \theta^0 = \mu^0 - \tfrac{1}{2} = -\tfrac{1}{2}, \quad \theta^2 = \mu^2 - \tfrac{1}{2} = -\tfrac{3}{2}.$$

Using the Gaussian property of $\theta^i t + B_t$ and Bayes' formula, we find $\mathbf{P}(\boldsymbol{\theta} = \boldsymbol{\theta}^{\boldsymbol{i}} \mid \boldsymbol{X_t} = \boldsymbol{x}) = \exp\Big\{x\theta^i - \frac{(\theta^i)^2 t}{2}\Big\} \Big/ \sum\limits_{j=0}^{2} \exp\Big\{x\theta^j - \frac{(\theta^j)^2 t}{2}\Big\}$. Thus,

$$G_1(t,x;s,a,b) = \sum_{j=0}^{2} \mathsf{P}\Big(x + B_{s-t} + \theta^j(s-t) \in [a,b]\Big)\mathbf{P}(\boldsymbol{\theta} = \boldsymbol{\theta}^{\boldsymbol{j}} \mid \boldsymbol{X_t} = \boldsymbol{x})$$

$$= \frac{\sum\limits_{j=0}^{2} \Big\{\Phi_{s-t}(b - x - \theta^j(s-t)) - \Phi_{s-t}(a - x - \theta^j(s-t))\Big\} \exp\Big\{x\theta^j - \frac{(\theta^j)^2 t}{2}\Big\}}{\sum\limits_{j=0}^{2} \exp\Big\{x\theta^j - \frac{(\theta^j)^2 t}{2}\Big\}},$$

where $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt$, $\varphi(t) = (2\pi)^{-1/2} e^{-t^2/2}$.

Let us calculate $G_2(t,x,T) := \mathsf{E}_{t,x} L_T^{\circ}(t,\widetilde{X})$:

$$G_2(t,x,T) = \sum_{i=0}^{2} \underbrace{\mathsf{E} L_0^{\circ}(T-t, x + \overbrace{B_t^{\theta^i}}^{:=\,\theta^i t + B_t})}_{=:\, G_3(x,\theta^i,T-t)} \mathsf{P}(\theta = \theta^i \mid X_t = x)$$

$$\overset{\textbf{Bayes' formula}}{=} \frac{\sum_{i=0}^{2} G_3(x,\theta^i,T-t) \exp\Big\{x\theta^i - \frac{(\theta^i)^2 t}{2}\Big\}}{\sum_{i=0}^{2} \exp\Big\{x\theta^i - \frac{(\theta^i)^2 t}{2}\Big\}}. \tag{189}$$

Straightforward calculations for $G_3(x,\mu,s) = \lim_{\varepsilon\downarrow 0} \mathsf{E}\{(2\varepsilon)^{-1} \int_0^s I(|x + B_u^{\mu}| \le \varepsilon)\,du\}$ leads to the formula

$$G_3(x,\mu,s) = \sqrt{\frac{2s}{\pi}} e^{-(\mu s + x)^2/(2s)} + 2x\Phi\Big(\mu\sqrt{s} + \frac{x}{\sqrt{s}}\Big)$$
$$- x - |x| + \mu \int_0^s \Big(\mu\sqrt{r} - \frac{x}{\sqrt{r}}\Big)\varphi\Big(\mu\sqrt{r} + \frac{x}{\sqrt{r}}\Big)\,dr.$$

Let $G_3(x,\mu) := \lim_{s\to\infty} G_3(x,\mu,s)$. Then

$$G_3(x,\mu) = \mu \int_0^s \Big(\mu\sqrt{r} - \frac{x}{\sqrt{r}}\Big)\varphi\Big(\mu\sqrt{r} + \frac{x}{\sqrt{r}}\Big)\,dr - |x| + x\,\mathrm{sgn}\,\mu.$$

From existence of this limit and (189) we find that the limit $G_2(t,x) = \lim_{T\to\infty} G_2(t,x,T)$ exists and

$$G_2(t,x) = \frac{\sum\limits_{i=0}^{2} G_3(x,\theta^i)\exp\Big\{x\theta^i - \frac{(\theta^i)^2 t}{2}\Big\}}{\sum\limits_{i=0}^{2} \exp\Big\{x\theta^i - \frac{(\theta^i)^2 t}{2}\Big\}}.$$

The integral equations

$$\boxed{\begin{aligned} c\int_t^\infty G_1\Big(t,\widehat{f}(t);s,\widehat{g}(s),\widehat{f}(s)\Big)\,ds &= G_2(t,\widehat{f}(t)) \\ c\int_t^\infty G_1\Big(t,\widehat{g}(t);s,\widehat{g}(s),\widehat{f}(s)\Big)\,ds &= G_2(t,\widehat{g}(t)) \end{aligned}}$$

follow from (188) by passing to the limit as $T\to\infty$ and using above formulae.

TOPIC VII. Applications-3: Optimal stopping in some problems of the one-time portfolio rebalancing

In this presentation we investigate the structure of the

OPTIMAL ONE-TIME REBALANCING STRATEGY

on the (B, S)-market in the case of the possible spontaneous change of its parameter. We begin with the (B, S)-model (**Black–Scholes**)

$$\boxed{\begin{aligned} dB_t &= rB_t\,dt, & B_0 &= 1 \\ dS_t &= S_t\,(\mu\,dt + \sigma\,dW_t), & S_0 &= 1 \end{aligned}} \tag{190}$$

$[W = (W_t)_{t \le T}$ is a standard Wiener process (Brownian motion)].
Discounted prices $P_t = S_t / B_t$ solve the equation

$$dP_t = P_t\Big((\mu - r)\,dt + \sigma\,dW_t\Big), \qquad P_0 = 1,$$

and

$$P_t = \exp\{\nu t + \sigma W_t\}, \quad \text{where} \quad \nu = \mu - r - \tfrac{1}{2}\sigma^2.$$

Let $U = U(x)$ be a utility function (*e.g.*, $U(x) = \log x$ or $U(x) = x$).

In the paper:

A. Shiryaev, Z. Xu, X. Y. Zhou, **Thou Shalt Buy and Hold**,

presented by X. Y. Zhou at the

Conference on Quantitative Methods in Finance (Sydney, 2007)

[appeared in *Quantitative Finance*, December 2008],

the following problem was considered:

To find an optimal stopping time τ^* such that

$$\boxed{\mathsf{E}\, U\left(\frac{P_{\tau^*}}{M_T}\right) = \sup_{\tau \in \mathfrak{M}_T} \mathsf{E}\, U\left(\frac{P_\tau}{M_T}\right),}$$

where $\mathfrak{M}_T$ is the class of all stopping times taking values in $[0, T]$ and $M_T = \sup_{t \le T} P_t$.

The main result of our paper mentioned above implies that

for the **LINEAR** function $U(x) = x$ the (degenerated (!)) stopping time

$$\boxed{\tau^* = \begin{cases} T & \text{if } \ \nu > 0, \\ 0 & \text{if } \ \nu \le -\sigma^2/2, \end{cases}}$$

is **optimal**.

(In the case $-\sigma^2/2 < \nu \le 0$ the optimal stopping time is $\tau^* = 0$; this was shown by **J. du Toit, G. Peskir***.)

*"Selling a stock at the ultimate maximum", *Ann. Appl. Probab.*, 19 (2009), 983–1014.

In the case of the logarithmic function $U(x) = \log x$

$$\sup_{\tau \in \mathfrak{M}_T} \mathsf{E} \log \frac{P_\tau}{M_T} = \sup_{\tau \in \mathfrak{M}_T} \mathsf{E}\Big[\nu\tau + \sigma W_\tau - M_T\Big] = \sup_{\tau \in \mathfrak{M}_T} \mathsf{E}[\nu\tau + \sigma W_\tau] - \mathsf{E}M_T$$

$$= \sup_{\tau \in \mathfrak{M}_T} \mathsf{E}\nu\tau - \mathsf{E}M_T = \begin{cases} \nu T - \mathsf{E}M_T & \text{if} \quad \nu > 0, \\ -\mathsf{E}M_T & \text{if} \quad \nu \le 0. \end{cases}$$

So, in this logarithmic case an optimal stopping time is

$$\tau^* = \begin{cases} T & \text{if} \quad \nu > 0, \\ 0 & \text{if} \quad \nu \le 0, \end{cases}$$

which is the same as for the linear utility function.

In the present talk we consider the following generalization of the above problem using the model which has been proposed a long time ago by the author in the papers on the “quickest detection problem of the spontaneously appearing effects”.

More exactly, we shall consider the following two models (I and II):

I: $$\boxed{dS_t^{(\mathrm{I})} = S_t^{(\mathrm{I})} \left(\mu^{(\mathrm{I})}(t,\theta)\, dt + \sigma\, dW_t\right), \qquad S_0 = 1,}$$

and

II: $$\boxed{dS_t^{(\mathrm{II})} = S_t^{(\mathrm{II})} \left(\mu^{(\mathrm{II})}(t,\theta)\, dt + \sigma\, dW_t\right), \qquad S_0 = 1,}$$

where $\theta = \theta(\omega)$ is a random variable which is independent of W and assumes values in $\mathbb{R}_+ = [0,\infty)$ and

$$\mu^{(\mathrm{I})}(t,\theta) = \begin{cases} \mu_1, & t < \theta, \\ \mu_2, & t \geq \theta, \end{cases} \qquad \mu^{(\mathrm{II})}(t,\theta) = \begin{cases} \mu_2, & t < \theta, \\ \mu_1, & t \geq \theta; \end{cases}$$

the parameters μ_1 and μ_2 are assumed such that $\mu_1 > \mu_2$ and

$$\begin{aligned} \nu_1 &\equiv \mu_1 - r - \tfrac{1}{2}\sigma^2 > 0, \\ \nu_2 &\equiv \mu_2 - r - \tfrac{1}{2}\sigma^2 < 0, \end{aligned} \quad \text{so that} \quad \mu_2 - \tfrac{1}{2}\sigma^2 < r < \mu_1 - \tfrac{1}{2}\sigma^2.$$

Consider the **MODEL (I):** $(B, S^{(\mathrm{I})})$. In this case, starting from the initial time $t = 0$ the driving parameter is μ_1. If this value remains unchanged on the whole interval $[0, T]$ and $\nu_1 \equiv \mu_1 - r - \frac{1}{2}\sigma^2 > 0$, then by the previous result we should

hold the stock until time $t = T$ **and sell it at this time**.

But in fact the model (I) admits that at a certain random time θ the regime switches from μ_1 to μ_2, and if $\nu_2 \equiv \mu_2 - r - \frac{1}{2}\sigma^2 < 0$, then, again by the previous results, we should

sell the stock at this time θ.

However, this time is unobservable and so the time of selling must depend on the "correct" estimation of the time θ.

All these considerations lead to the following optimal problem:

To find "**one-time rebalancing**" stopping time τ_T^* such that

$$V_T^{(\mathrm{I})} \equiv \sup_{\tau \in \mathfrak{M}_T} \mathsf{E} \log \frac{P_\tau^{(\mathrm{I})}}{M_T^{(\mathrm{I})}} = \mathsf{E} \log \frac{P_{\tau_T^*}^{(\mathrm{I})}}{M_T^{(\mathrm{I})}}.$$

Since

$$P_t^{(\mathrm{I})} = \exp\Bigg\{ \overbrace{\int_0^t \underbrace{\nu(s,\theta)}_{:= \boldsymbol{\mu}^{(\mathrm{I})}(\boldsymbol{s},\boldsymbol{\theta}) - \boldsymbol{r} - \frac{1}{2}\boldsymbol{\sigma}^2} \, ds + \sigma W_t}^{(=: \boldsymbol{H}_t^{(\mathrm{I})})} \Bigg\}$$

and $\mathsf{E} W_\tau = 0$ for $\tau \in \mathfrak{M}_T$, we deduce that

$$V_T^{(\mathrm{I})} = \sup_{\tau \in \mathfrak{M}_T} \mathsf{E} \int_0^\tau \nu(s,\theta)\, ds.$$

It is interesting that this "$V_T^{(\mathrm{I})}$-**criterion**" is equivalent to the following "$W_T^{(\mathrm{I})}$-**criterion**":

$$\boxed{W_T^{(\mathrm{I})} = \sup_{\tau \in \mathfrak{M}_T} \mathsf{E} \log \left[S_\tau^{(\mathrm{I})} \frac{B_T}{B_\tau} \right]}.$$

The value $S_\tau^{(\mathrm{I})} B_T / B_\tau$ determines the capital at time T, if at time τ we sell the stock and put the gained value $S_\tau^{(\mathrm{I})}$ on the bank account. Since

$$S_\tau^{(\mathrm{I})} \frac{B_T}{B_\tau} = \frac{S_\tau^{(\mathrm{I})}}{B_\tau} B_T = P_\tau^{(\mathrm{I})} B_T,$$

we see that

$$W_T^{(\mathrm{I})} = \sup_{\tau \in \mathfrak{M}_T} \mathsf{E} \log P_\tau^{(\mathrm{I})} + \mathsf{E} \log B_T.$$

Hence

$$\boxed{W_T^{(\mathrm{I})} = V_T^{(\mathrm{I})} + \mathsf{E} \log M_T^{(\mathrm{I})} + \mathsf{E} \log B_T}.$$

Consider now the **MODEL (II)**, where $\mu_2 \to \mu_1$ at time θ. In this case, our "one-time rebalancing" problem can be reformulated as the problem of buying (after the time θ) the stock at minimal price:

$$\boxed{V_T^{(\mathrm{II})} = \inf_{\tau \in \mathfrak{M}_T} \mathsf{E} \log \frac{P_\tau^{(\mathrm{II})}}{\min_{t\in[0,T]} P_t^{(\mathrm{II})}}}.$$

The corresponding "$W^{(\mathrm{II})}$-**problem**" is defined by the criterion:

$$\begin{aligned} W_T^{(\mathrm{II})} &= \sup_{\tau \in \mathfrak{M}_T} \mathsf{E} \log \left[B_\tau \frac{S_T^{(\mathrm{II})}}{S_\tau^{(\mathrm{II})}} \right] = \sup_{\tau \in \mathfrak{M}_T} \mathsf{E} \log \left[\frac{B_\tau}{S_\tau^{(\mathrm{II})}} S_T^{(\mathrm{II})} \right] \\ &= \sup_{\tau \in \mathfrak{M}_T} \mathsf{E} \log \frac{1}{P_\tau^{(\mathrm{II})}} + \mathsf{E} \log S_T^{(\mathrm{II})} = - \inf_{\tau \in \mathfrak{M}_T} \mathsf{E} \log P_\tau^{(\mathrm{II})} + \mathsf{E} \log S_T^{(\mathrm{II})}. \end{aligned}$$

Hence

$$\boxed{W_T^{(\mathrm{II})} = -V_T^{(\mathrm{II})} + \mathsf{E} \log S_T^{(\mathrm{II})} - \mathsf{E} \log \min_{t \in [0,T]} P_t^{(\mathrm{II})}}.$$

Now we give the second proof of the key representation

$$\mathsf{E}\log P_\tau = \mathsf{E}\int_0^\tau \Big[\nu_1 - (\nu_1-\nu_2)\pi_t\Big]\,dt.$$

Recall that

$$P_t = \frac{S_t}{B_t} = e^{X_t},$$

where

$$X_t = \int_0^t \nu(s,\theta)\,ds + \sigma W_t, \qquad \nu(s,\theta) = \mu(s,\theta) - r - \tfrac{1}{2}\sigma^2.$$

Notation:

$$V_T = \sup_{\tau\in\mathfrak{M}_T} \mathsf{E}\log P_\tau = \sup_{\tau\in\mathfrak{M}_T} \mathsf{E}X_\tau = \sup_{\tau\in\mathfrak{M}_T} \mathsf{E}\int_0^\tau \nu(s,\theta)\,ds.$$

Fix some $\tau \in \mathfrak{M}_T$. Then

$$
\begin{aligned}
\mathsf{E}\int_0^\tau \nu(s,\theta)\,ds &= \mathsf{E}\Big\{I(\theta>\tau)\int_0^\tau \nu(s,\theta)\,ds + I(\theta\le\tau)\int_0^\tau \nu(s,\theta)\,ds\Big\} \\
&= \mathsf{E}\Big\{I(\theta>\tau)\nu_1\tau + I(\theta\le\tau)\Big[\nu_1\theta+\nu_2(\tau-\theta)\Big]\Big\} \\
&= \mathsf{E}\Big\{I(\theta>\tau)\nu_1\tau + I(\theta\le\tau)\Big[\nu_2\tau+\theta(\nu_1-\nu_2)\Big]\Big\} \\
&= \mathsf{E}\,\mathsf{E}\,(\{\ldots\}\,|\,\mathcal{F}_\tau) \\
&= \mathsf{E}\Big\{\nu_1\tau(1-\pi_\tau)+\nu_2\tau\pi_\tau+(\nu_1-\nu_2)\,\mathsf{E}\Big[\theta I(\theta\le\tau)\,|\,\mathcal{F}_\tau\Big]\Big\} \\
&= \mathsf{E}\Big\{\nu_1\tau+(\nu_2-\nu_1)\tau\pi_\tau+(\nu_1-\nu_2)\,\mathsf{E}\Big[\theta I(\theta\le\tau)\,|\,\mathcal{F}_\tau\Big]\Big\},
\end{aligned}
$$

where $\pi_t = \mathsf{P}(\theta \le t\,|\,\mathcal{F}_t)$ is *a posteriori* probability of the event $\{\theta \le t\}$ conditioned on $\mathcal{F}_t = \sigma(H_s^{(\mathrm{I})},\, s \le t)$; $\nu_i = \mu_i - r - \frac{1}{2}\sigma^2$, $i = 1, 2$;

$$
H_s^{(\mathrm{I})} = \int_0^s \Big(\mu^{(\mathrm{I})}(u,\theta) - r - \tfrac{1}{2}\sigma^2\Big)\,du + \sigma W_s, \qquad \mu^{(\mathrm{I})}(u,\theta) = \begin{cases} \mu_1, & u < \theta, \\ \mu_2, & u \ge \theta. \end{cases}
$$

To find an optimal stopping time in the problem

$$\sup_{\tau\in\mathfrak{M}_T} \mathsf{E}\int_0^\tau \nu(s,\theta)\,ds,$$

we have to solve the following optimal stopping problem:

$$\sup_{\tau\in\mathfrak{M}_T} \mathsf{E}\Big\{\nu_1\tau + (\nu_2-\nu_1)\Big[\tau\pi_\tau - \mathsf{E}\Big(\theta I(\theta\le\tau)\,|\,\mathcal{F}_\tau\Big)\Big]\Big\}.$$

Let us make the following assumption about the distribution of θ:

$$\boxed{\mathsf{P}(\theta=0)=\pi, \qquad \mathsf{P}(\theta>t\,|\,\theta>0)=e^{-\lambda t}, \quad \text{where } \lambda>0 \text{ is known}}\,.$$

Having this distribution it is easy to find the *a posteriori* probability $\pi_\tau = \mathsf{P}(\theta\le t\,|\,\mathcal{F}_\tau)$, where $\mathcal{F}_t=\sigma(H_s^{(\mathrm{I})},\ s\le t)$.

Indeed, introduce the processes

$$\varphi_t = \frac{\pi_t}{1-\pi_t} \quad \text{and} \quad L_t(H) = \frac{d(\mathsf{P}^{\nu_1} \,|\, \mathcal{F}_t^H)}{d(\mathsf{P}^{\nu_2} \,|\, \mathcal{F}_t^H)};$$

$L_t(H)$ is the likelihood process, *i.e.*, the Radon–Nikodým derivative of the measure $\mathsf{P}^{\nu_1}|\mathcal{F}_t^H$ w.r.t. the measure $\mathsf{P}^{\nu_2}|\mathcal{F}_t^H$, where P^{ν_i} is the distribution of the process $(H_t^i)_{t\ge 0}$ with

$$dH_t^i = \nu_i\,dt + \sigma\,dW_t.$$

It is well known that for each $i=1,2$

$$L_t(H^i) = \exp\left\{\frac{\nu_2-\nu_1}{\sigma^2}\,H_t^i - \frac{1}{2}\,\frac{\nu_2^2-\nu_1^2}{\sigma^2}\,t\right\} \quad (\mathsf{P}^{\nu_i}\text{-a.s.}).$$

By Itô's formula, this implies that

$$dL_t(H^i) = L_t(H^i)\frac{\nu_2-\nu_1}{\sigma^2}\,\{dH_t^i - \nu_1\,dt\}.$$

By Bayes' formula, taking for simplicity $\pi = 0$, we get

$$\varphi_t = \lambda \int_0^t \frac{e^{\lambda t} L_t(H^{(\mathrm{I})})}{e^{\lambda u} L_u(H^{(\mathrm{I})})}\, du.$$

From this representation, again by Itô's formula, we find

$$d\varphi_t = \left[\lambda(1+\varphi_t) - \varphi_t \nu_1 \frac{\nu_2 - \nu_1}{\sigma^2}\right] dt + \varphi_t \frac{\nu_2 - \nu_1}{\sigma^2}\, dH_t^{(\mathrm{I})}$$

and

$$\begin{aligned} d\pi_t = (1-\pi_t)&\left[\lambda - \nu_1 \frac{\nu_2 - \nu_1}{\sigma^2}\, \pi_t - \frac{(\nu_2 - \nu_1)^2}{\sigma^2}\, \pi_t^2\right] dt \\ &+ \frac{\nu_2 - \nu_1}{\sigma^2}\, \pi_t (1-\pi_t)\, dH_t^{(\mathrm{I})}. \end{aligned}$$

Let $\overline{W}_t^{(\mathrm{I})} = \dfrac{1}{\sigma}\Big[H_t^{(\mathrm{I})} - \nu t - (\nu_2 - \nu_1)\displaystyle\int_0^t \pi_s\,ds\Big]$. This (innovation) process is a Wiener process (w.r.t. $(\mathcal{F}_t^{H^{(\mathrm{I})}})_{t\ge 0}$). Taking into account this notation we find that

$$d\pi_t = \lambda(1-\pi_t)\,dt + \frac{\nu_2-\nu_1}{\sigma}\,\pi_t(1-\pi_t)\,d\overline{W}_t^{(\mathrm{I})}.$$

Return to the formula

$$\begin{aligned}\mathsf{E}\int_0^\tau \nu(s,\theta)\,ds = \mathsf{E}\Big\{&\nu_1\tau + (\nu_2-\nu_1)\tau\pi_\tau \\ &+ (\nu_1-\nu_2)\,\mathsf{E}\Big[\theta I(\theta\le\tau)\,|\,\mathcal{F}_\tau\Big]\Big\},\end{aligned}$$

where $\mathcal{F}_t = \sigma(H_s^{(\mathrm{I})},\, s\le t)$. Put

$$\rho_t = \mathsf{E}[\theta I(\theta\le t)\,|\,\mathcal{F}_t] \quad \text{and} \quad p_\theta(s) = \frac{\partial}{\partial s}\mathsf{P}(\theta\le s\,|\,\mathcal{F}_t), \quad s\le t.$$

Let us find $\mathsf{P}(\theta\le s\,|\,\mathcal{F}_t)$.

With that end in view we note that, again by Bayes' formula,

$$\frac{\mathsf{P}(\theta > s \,|\, \mathcal{F}_t)}{1 - \pi_t} = \frac{\mathsf{P}(\theta > s \,|\, \mathcal{F}_t)}{\mathsf{P}(\theta > t \,|\, \mathcal{F}_t)} = e^{\lambda t} \int_s^\infty e^{-\lambda u} \frac{d\mu_t^u}{d\mu_t^\infty}\, du$$

$$= 1 + \lambda e^{\lambda t} \int_s^t e^{-\lambda u} \frac{L_t}{L_u}\, du,$$

where

- μ_t^u and μ_t^∞ are measures on $[0, t]$ of the process $H^{(\mathrm{I})}$ under assumptions that $\theta = u$ and $\theta = \infty$, respectively, and
- $L_u = L_u(H^{(\mathrm{I})})$.

Hence, for $s \le t$

$$\mathsf{P}(\theta > s \,|\, \mathcal{F}_t) = (1-\pi_t)\left(1 + \lambda e^{\lambda t}\int_s^t e^{-\lambda u}\frac{L_t}{L_u}\,du\right),$$

$$\mathsf{P}(\theta \le s \,|\, \mathcal{F}_t) = \pi_t - (1-\pi_t)\left(\lambda e^{\lambda t}\int_s^t e^{-\lambda u}\frac{L_t}{L_u}\,du\right).$$

From here $\quad p_\theta(s) := \dfrac{\partial}{\partial s}\mathsf{P}(\theta \le s \,|\, \mathcal{F}_t) = (1-\pi_t)\lambda e^{\lambda t}e^{-\lambda s}\,\dfrac{L_t}{L_s},\quad$ thus

$$\rho_t := \mathsf{E}[\theta I(\theta \le t) \,|\, \mathcal{F}_t] = \int_0^t s p_\theta(s)\,ds = (1-\pi_t)\,\underbrace{\lambda e^{\lambda t}\int_0^t s e^{-\lambda s}\,\frac{L_t}{L_s}\,ds}_{=:\,\boldsymbol{\gamma_t}}\,.$$

Then

$$d\rho_t = d((1-\pi_t)\gamma_t) = (1-\pi_t)\,\underbrace{d\gamma_t}_{(*)} - \gamma_t\,\underbrace{d\pi_t}_{(**)} - \underbrace{d[\pi,\gamma]_t}_{(***)}\,. \qquad (191)$$

$$
(*): \quad d\gamma_t = \lambda\gamma_t\,dt + \lambda e^{\lambda t}\,d\left[\int_0^t s e^{-\lambda s}\,\frac{L_t}{L_s}\,ds\right]
$$

$$
= \lambda\gamma_t\,dt + \lambda e^{\lambda t}\left[t e^{-\lambda t}\,dt + dL_t \cdot \int_0^t \frac{s e^{-\lambda s}}{L_s}\,ds\right]
$$

$$
= \lambda(\gamma_t + t)\,dt + \gamma_t \frac{\nu_2 - \nu_1}{\sigma^2}\left[\underbrace{dH_t^{(\mathrm{I})} - \nu_1\,dt}_{= \sigma\, d\overline{W}_t^{(\mathrm{I})} + (\nu_2 - \nu_2)\pi_t\,dt}\right]
$$

$$
= \lambda(\gamma_t + t)\,dt + \gamma_t\left[\pi_t\left(\frac{\nu_2 - \nu_1}{\sigma}\right)^2 dt + \frac{\nu_2 - \nu_1}{\sigma}\,d\overline{W}_t^{(\mathrm{I})}\right]
$$

$$
(**): \quad d\pi_t = \lambda(1 - \pi_t)\,dt + \frac{\nu_2 - \nu_1}{\sigma}\,\pi_t(1 - \pi_t)\,d\overline{W}_t^{(\mathrm{I})}
$$

$$
(***): \quad d[\pi, \gamma]_t = \left(\frac{\nu_2 - \nu_1}{\sigma}\right)^2 \gamma_t \pi_t (1 - \pi_t)\,dt
$$

Inserting $(*)$–$(***)$ into (191) leads to

$$
\begin{aligned}
d\rho_t &= (1-\pi_t)\,\underbrace{d\gamma_t}_{(*)} - \gamma_t\,\underbrace{d\pi_t}_{(**)} - \underbrace{d[\pi,\gamma]_t}_{(***)} \\
&= (1-\pi_t)\Bigg\{\underbrace{\lambda(\gamma_t+t)\,dt + \gamma_t\Bigg[\pi_t\Big(\frac{\nu_2-\nu_1}{\sigma}\Big)^2 dt + \frac{\nu_2-\nu_1}{\sigma}\,d\overline{W}_t^{(\mathrm{I})}\Bigg]}_{(*)}\Bigg\} \\
&\quad - \gamma_t\Bigg\{\underbrace{\lambda(1-\pi_t)\,dt + \frac{\nu_2-\nu_1}{\sigma}\pi_t(1-\pi_t)\,d\overline{W}_t^{(\mathrm{I})}}_{(**)}\Bigg\} \\
&\quad - \underbrace{\Big(\frac{\nu_2-\nu_1}{\sigma}\Big)^2 \gamma_t\pi_t(1-\pi_t)\,dt}_{(***)} \\
&= \boxed{\lambda t(1-\pi_t)\,dt + \frac{\nu_2-\nu_1}{\sigma}\gamma_t(1-\pi_t)^2\,d\overline{W}_t^{(\mathrm{I})}}. \qquad (192)
\end{aligned}
$$

Using (∗∗), we find that

$$
\begin{aligned}
d(t\pi_t) &= \pi_t\,dt + t\,d\pi_t \\
&= \pi_t\,dt + t\left\{\lambda(1-\pi_t)\,dt + \frac{\nu_2-\nu_1}{\sigma}\,\pi_t(1-\pi_t)\,d\overline{W}_t^{(\mathrm{I})}\right\}.
\end{aligned}
$$

This expression and (192) imply that

$$
\begin{aligned}
\mathsf{E}\int_0^\tau \nu(s,\theta)\,ds &= \mathsf{E}\Big\{\nu_1\tau + (\nu_2-\nu_1)\tau\pi_\tau + (\nu_1-\nu_2)\rho_\tau\Big\} \\
&= \mathsf{E}\int_0^\tau \Big\{\nu_1 + (\nu_2-\nu_1)\Big[\pi_t + t\lambda(1-\pi_t)\Big] \\
&\qquad\qquad - (\nu_2-\nu_1)\lambda t(1-\pi_t)\Big\}\,dt \\
&= \mathsf{E}\int_0^\tau \Big(\nu_1 + (\nu_2-\nu_1)\pi_t\Big)\,dt.
\end{aligned}
$$

Thus, $\quad \mathsf{E}\int_0^\tau \nu(s,\theta)\,ds = \mathsf{E}\int_0^\tau \Big(\nu_1 + \underbrace{(\overbrace{\nu_2-\nu_1}^{<\,0})\pi_t}_{<\,0}\Big)\,dt.$

This implies that we need find a stopping time τ_T^* such that

$$\sup_{\tau\in\mathfrak{M}_T} \mathsf{E}\int_0^\tau \nu(s,\theta)\,ds = \sup_{\tau\in\mathfrak{M}_T} \mathsf{E}\int_0^\tau \Big[\nu_1 + (\nu_2-\nu_1)\pi_t\Big]\,dt$$
$$= \mathsf{E}\int_0^{\boldsymbol{\tau}_T^*} \Big[\nu_1 + (\nu_2-\nu_1)\pi_t\Big]\,dt,. \qquad (*)$$

$$\Downarrow \qquad \Downarrow \qquad \Downarrow$$

since the process π_t is continuous, then if $\pi_0 < \frac{\nu_1}{\nu_2-\nu_1}$, one should continue observations at least till the time

$\boxed{\widetilde{\boldsymbol{\tau}}_T = \min\Big\{0 \le t \le T\colon \pi_t \ge \widetilde{\boldsymbol{A}}_\infty := \dfrac{\nu_1}{\nu_1-\nu_2}\Big\}}$, i.e., $\boldsymbol{\tau}_T^* \ge \widetilde{\boldsymbol{\tau}}_T$.

Note also that

- if $\pi_t \equiv 0$ (i.e., from the beginning we have parameter $\nu_1 > 0$), then it follows from $(*)$ that $\tau_T^* = T$;
- If $\pi_t \equiv 1$ (i.e., from the beginning we have parameter $\nu_2 > 0$), then evidently $(*)$ implies that $\tau_T^* = 0$.

These deterministic Buy & Hold rules were already described above.

To find an optimal stopping time τ_T^*, let us consider more carefully the expression

$$I(t) = \int_0^t \Big[\nu_1 + (\nu_2 - \nu_1)\pi_t\Big]\, dt.$$

Since

$$d\pi_t = \lambda(1-\pi_t)\, dt + \frac{\nu_2 - \nu_1}{\sigma}\, \pi_t(1-\pi_t)\, d\overline{W}_t^{(\mathrm{I})},$$

we see (taking $\pi_0 = 0$) that

$$\lambda t = \pi_t + \lambda \int_0^t \pi_s\, ds - \int_0^t \frac{\nu_2 - \nu_1}{\sigma}\, \pi_s(1-\pi_s)\, d\overline{W}_s^{(\mathrm{I})}.$$

So,

$$t = \frac{\pi_t}{\lambda} + \int_0^t \pi_s\, ds - \int_0^t \frac{\nu_2 - \nu_1}{\lambda\sigma}\, \pi_s(1-\pi_s)\, d\overline{W}_s^{(\mathrm{I})}$$

and

$$\nu_1 t = \frac{\nu_1}{\lambda}\pi_t + \nu_1 \int_0^t \pi_s\, ds - \int_0^t \frac{(\nu_2 - \nu_1)\nu_1}{\lambda\sigma}\, \pi_s(1-\pi_s)\, d\overline{W}_s^{(\mathrm{I})}. \qquad (193)$$

Using the representation (193), we find that

$$\mathsf{E} I(\tau) = \mathsf{E}\int_0^\tau \Big[\nu_1 + (\nu_2 - \nu_1)\pi_t\Big]\,dt = \mathsf{E}\Big\{\frac{\nu_1}{\lambda}\pi_\tau + \nu_2\int_0^\tau \pi_t\,dt\Big\}$$

$$= \frac{\nu_1}{\lambda}\Big\{\pi_\tau + \frac{\nu_2}{\nu_1}\lambda\int_0^\tau \pi_t\,dt\Big\} = -\frac{\nu_1}{\lambda}\Big\{-\pi_\tau + \frac{|\nu_2|}{\nu_1}\lambda\int_0^\tau \pi_t\,dt\Big\}.$$

Hence, letting $c = (|\nu_2|/\nu_1)\lambda$, we get

$$\sup_{\tau\in\mathfrak{M}_T} \mathsf{E} I(\tau) = -\frac{\nu_1}{\lambda}\inf_{\tau\in\mathfrak{M}_T}\mathsf{E}\Big\{-\pi_\tau + c\int_0^\tau \pi_t\,dt\Big\}$$

$$= -\frac{\nu_1}{\lambda}\Big\{-1 + \inf_{\tau\in\mathfrak{M}_T}\mathsf{E}\Big\{(1-\pi_\tau) + c\int_0^\tau \pi_t\,dt\Big\}\Big\}$$

$$= \frac{\nu_1}{\lambda} - \frac{\nu_1}{\lambda}\underbrace{\inf_{\tau\in\mathfrak{M}_T}\mathsf{E}\Big\{(1-\pi_\tau) + c\int_0^\tau \pi_t\,dt\Big\}}_{=:\, \boldsymbol{V_T}}.$$

$$\boxed{V_T = \inf_{\tau \in \mathfrak{M}_T} \mathsf{E}\Big\{(1-\pi_\tau) + c\int_0^\tau \pi_t\,dt\Big\}}$$

The **case $T = \infty$** was investigated in an author's monograph *, from which it follows that an optimal stopping time is

$$\tau_\infty^* = \inf\{t\colon \pi_t \ge A_\infty^*\},$$

where A_∞^* is a unique root of the equation

$$\boxed{C\int_0^{A_\infty^*} \exp\Big\{-\Lambda[H(\pi) - H(y)]\Big\}\frac{dy}{y(1-y)^2} = 1,} \qquad \textbf{(**)}$$

where

$$C = \frac{c}{\rho}, \quad \Lambda = \frac{\lambda}{\rho}, \quad H(y) = \log\frac{y}{1-y} - \frac{1}{y}, \quad \rho = \frac{(\nu_1 - \nu_2)^2}{2\sigma^2}$$

$\Big[A_\infty^*$ can be shown to satisfy $A_\infty^* > \widetilde{A}_\infty \ \Big(= \dfrac{\nu_1}{\nu_1 - \nu_2} = \dfrac{\nu_1}{\nu_1 + |\nu_2|}\Big).\Big]$

---

* A. N. Shiryaev. Optimal Stopping Rules, Springer, 1978, 2008; Chap. 4, § 4.4 "The problem of disruption for a Wiener process".

In the **case $T < \infty$** the optimal stopping time τ_T^* has the following form:

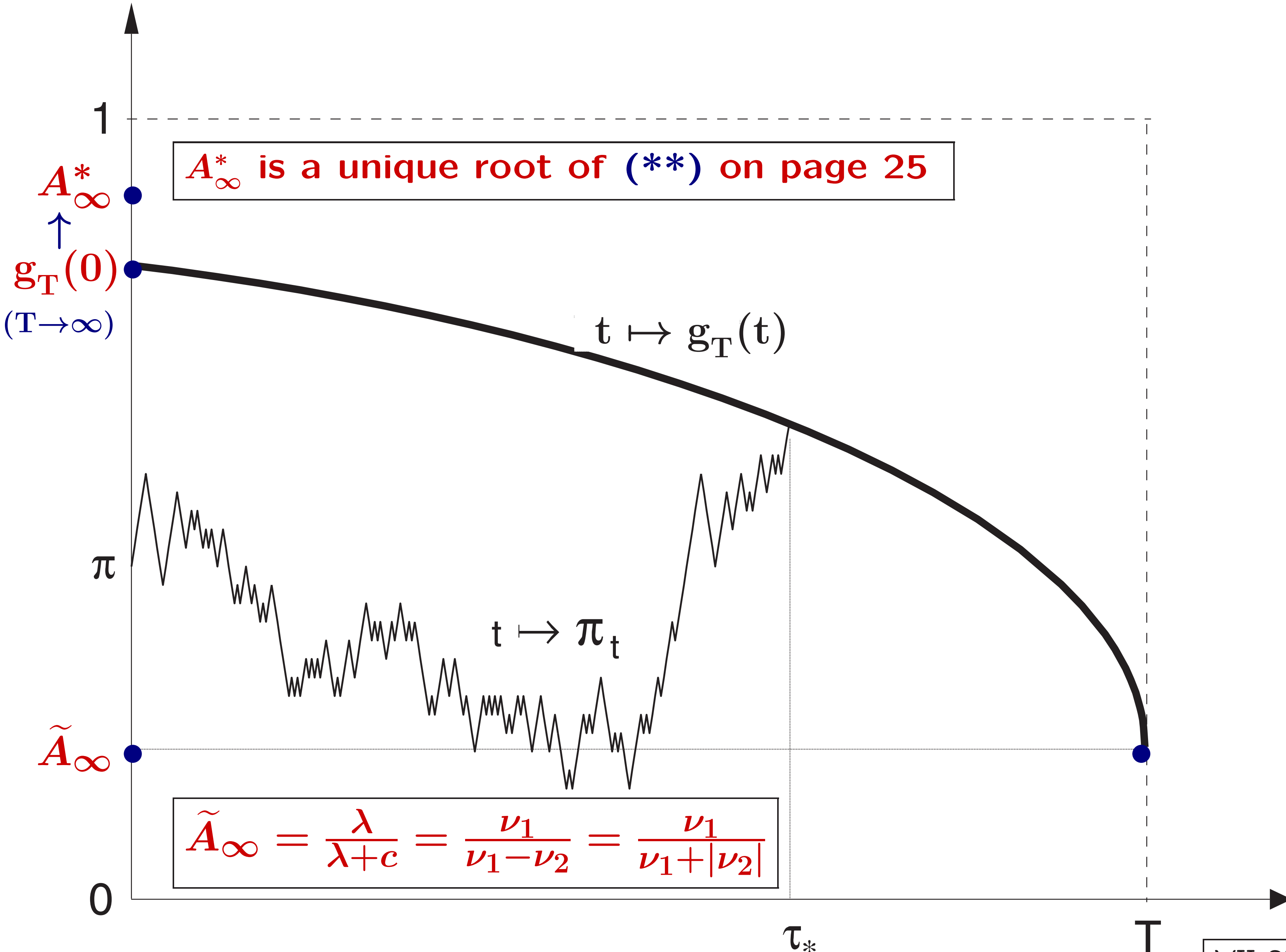
$$\tau_T^* = \inf\{0 \le t \le T\colon \pi_t \ge g_T(t)\},$$

where the optimal stopping boundary $g_T(t) =: g(t)$, $0 \le t \le T$, is a unique solution of the nonlinear integral equation **(G. Peskir)***:

$$\begin{aligned} \mathsf{E}_{t,g(t)}\pi_T = g(t) &+ c\int_0^{T-t} \mathsf{E}_{t,g(t)}\Big[\pi_{t+u}\, I(\pi_{t+u} < g(t+u)\Big]\, du \\ &+ \lambda \int_0^{T-t} \mathsf{E}_{t,g(t)}\Big[(1-\pi_{t+u})\, I(\pi_{t+u} < g(t+u)\Big]\, du. \end{aligned}$$

Note that $g_T(t) \uparrow A_\infty^*$ as $T \to \infty$ for all $t \ge 0$.

*see details in: G. Peskir, A. Shiryaev. Optimal Stopping and Free-Boundary problems, Birkhäuser, 2006; Chap. VI, Sec. 22, Fig. VI.2.



1
A^*_∞ is a unique root of (**) on page 25
A^*_∞
$\uparrow$
$g_T(0)$
$(T\to\infty)$
$t \mapsto g_T(t)$
π
$t \mapsto \pi_t$
$\widetilde{A}_\infty$
$\widetilde{A}_\infty = \frac{\lambda}{\lambda+c} = \frac{\nu_1}{\nu_1-\nu_2} = \frac{\nu_1}{\nu_1+|\nu_2|}$
0
τ_*
T

VII-27

$\boxed{W_T^{(\mathrm{I})} = \sup_{\tau\in\mathfrak{M}_T} \mathsf{E} \log\left[S_\tau^{(\mathrm{I})} \frac{B_t}{B_\tau}\right]}$: as we have seen, for this criterion the optimal stopping time is the same as for $V_T^{(\mathrm{I})}$-criterion. Hence

the above stopping time $\tau_T^* = \inf\{0 \le t \le T\colon \pi_t \ge g(t)\}$ is optimal.

- Let us mention that the $W_T^{(\mathrm{I})}$-criterion was considered by **Ch. Blanchet-Scalliet, A. Diop, R. Gibson, D. Talay, E. Taure**.* In particular, they proposed ad hoc to use for each $T > 0$ the stopping time $\tau_\infty^* = \inf\{t \le T\colon \pi_t \ge A^*\}$. From the above results it follows that their method is only **almost optimal**. (The **optimal** stopping time, as we have demonstrated, is the time τ_T^*.)

Criteria $V_T^{(\mathrm{II})}$ and $W_T^{(\mathrm{II})}$ can be investigated in a similar way.

*"Technical analysis compared to mathematical models based methods under parameters mis-specification", J. Banking Fin., 31 (2007), 1351–1373.

It is interesting to note that the process (π_t) proves to be optimal in other formulations of the problem "when sell the stock". For example, **M. Beibel and H. R. Lerche*** showed that the stopping time

$$\sigma_\infty^* = \inf\{t \geq 0 \colon \pi_t \geq B_\infty^*\},$$

where B_∞^* is a certain constant, is optimal in the following problem:

$$\sup_{\sigma \in \mathfrak{M}_\infty} \mathsf{E} \frac{S_\sigma}{B_\sigma}$$

provided that $\mu_2 < r < \mu_1$ and $\mu_1 < \lambda + r$.

*"A new look at optimal stopping problems related to mathematical finance", Statistica Sinica, 7:1 (1997), 93–108.