МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО
ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Международная лаборатория статистики случайных
процессов и количественного финансового анализа

Международная научная конференция «Робастная статистика и финансовая математика – 2022»

(04–05 июля 2022 г.)

Сборник статей

Под редакцией
д-ра физ.-мат. наук, профессора С.М. Пергаменщикова,
канд. физ.-мат. наук, доцента Е.А. Пчелинцева

Томск
Издательский Дом Томского государственного университета
2022

Early detection of epidemics [1]

Tenzin R., Pchelintsev E., Pergamenchtchikov S.

Tomsk State University, Tomsk, Russia
e-mail: 4ubarok@gmail.com

Abstract

The paper considers the problem of constructing an optimal procedure for detecting a disruption in a Markov process, which is the epidemiological dynamics of the number of infected people. A similar statement of the problem is also considered in [1].

Keywords: Disruption in Markov chain; Optimal stopping time; Bayesian risk; Epidemics detection.

Let us consider a Markov process $(X_n)_{0\le n\le N}$ on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with values in a measurable space $(\mathcal{X}, \mathcal{B}, \mu)$, where μ is a σ - finite measure on $\mathcal{B}$. In this case, we denote the natural filtering of the process $(X_n)_{0\le n\le N}$, which is all available information by the time $0 \le n \le N$, as $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma\{X_1, \ldots, X_n\}$. Disruption occurs in the process at some unknown point in time ν. That means before the time moment ν (inclusive) there is one homogeneous Markov process $(X_n)_{0\le n\le \nu}$ with a transition conditional distribution density $f^*(y\,|\,x)$, and after this moment there is another homogeneous Markov process $(X_n)_{\nu+1\le n\le N}$ with a transition conditional density $f(y\,|\,x)$.

Note that in our formulation, the observations of the process $(X_n)_{0\le n\le N}$ at each moment of time n represent the number of people presumably infected with a certain virus. At the same time, the number of observations N is the duration of the epidemiological season under consideration and the moment of disruption is the last moment of time before the start of the epidemic. Transitional distribution densities in this case have the following form

$$f^*(y|x) = \binom{x}{y} (\theta_*)^{x-y} (1-\theta_*)^y \, \mathbf{1}_{\{x\ge y\}} \,,$$

$$f(y|x) = \binom{x}{y} \theta^{x-y} (1-\theta)^y \, \mathbf{1}_{\{x\ge y\}} \,,$$

where the parameters θ^* and θ represent the probabilities of infection before and after the moment of disruption, and also, as you can see, exactly this parameter changes after disruption moment.

[1] The work was supported by the RSF, project no 22-21-00302.

In this case, for any $A \in \mathcal{B}$ we can use the following transition probabilities:

$$\mathbf{P}\left(X_{n+1} \in A | X_n = x\right) = \int_A f^*(y \,|\, x)\mu(\mathrm{d}y) \quad \text{for} \quad 0 \le n \le \nu$$

and

$$\mathbf{P}\left(X_{n+1} \in A | X_n = x\right) = \int_A f(y \,|\, x)\mu(\mathrm{d}y) \quad \text{for} \quad n \ge \nu + 1\,.$$

This way, we can also set the probabilities of joint distributions, which for $A \in \mathcal{B}_n$, where $\mathcal{B}_n = \underbrace{\mathcal{B} \otimes \cdots \otimes \mathcal{B}}_{n}$, for $0 \le n \le \nu$ will have the following form

$$\mathbf{P}^*\left((X_1, \ldots, X_n) \in A\right) = \int_A \mathbf{q}^*(y_1, \ldots, y_n)\mu(\mathrm{d}y_1)\ldots\mu(\mathrm{d}y_n)\,,$$

where

$$\mathbf{q}^*(y_1, \ldots, y_n) = \prod_{i=1}^{n} f^*(y_i | y_{i-1})\,;$$

and for $n \ge \nu + 1$

$$\mathbf{P}_\nu\left((X_1, \ldots, X_n) \in A\right) = \int_A \mathbf{q}_\nu(y_1, \ldots, y_n)\mu(\mathrm{d}y_1)\ldots\mu(\mathrm{d}y_n)\,,$$

where

$$\mathbf{q}_\nu(y_1, \ldots, y_n) = \prod_{i=1}^{\nu} f^*(y_i | y_{i-1}) \prod_{i=\nu+1}^{n} f(y_i | y_{i-1})\,.$$

This means that we can also set the corresponding Radon-Nikodym derivative for $\mathbf{P}^*$ and $\mathbf{P}_\nu$ on $\mathcal{B}_n$:

$$h_{\nu,n} = \frac{\mathrm{d}\mathbf{P}_\nu}{\mathrm{d}\mathbf{P}^*} = \frac{\mathbf{q}_\nu(y_1, \ldots, y_n)}{\mathbf{q}^*(y_1, \ldots, y_n)} = \prod_{i=\nu+1}^{n} \frac{f(y_i | y_{i-1})}{f^*(y_i | y_{i-1})}\,,$$

where $h_{\nu,n} = 1$ for $\nu \ge n$.

In this paper, we consider the Bayesian setting of the problem, this means that we assume that the moment of disruption is a random variable with values in the set $\mathcal{I}_N = \{0, \ldots, N\}$ belonging to some distribution $\mathbf{P}(\nu = n)$, which is usually called the prior distribution of the disruption moment. In this article, we consider the uniform prior distribution, i.e.

$$\pi_j = \pi_* = \mathbf{P}(\nu = n) = \frac{1}{N+1} \quad \text{for} \quad 0 \le n \le N\,.$$

Also in this setting we need the following Bayesian probability measure $\widetilde{\mathbf{P}}$, given on the σ-algebra $\mathcal{I}_N \otimes \mathcal{B}_N$ and for any $I \subseteq \mathcal{I}_N\,, \quad A \in \mathcal{B}_N$ defined as

$$\widetilde{\mathbf{P}}(I \times A) = \sum_{i \in I} \pi_i \mathbf{P}_i(A)\,.$$

As well as we need to define a posterior probability measure of the disruption moment ν as

$$\widetilde{\mathbf{P}}(\nu \leq n \,|\, \mathcal{F}_n) \,.$$

Moreover, in this subject a different representation of a posterior probability measure of the disruption moment ν was found.

Lemma 1. *A posterior probability measure of the disruption moment ν can be represented by the following way*

$$\widetilde{\mathbf{P}}(\nu \leq n \,|\, \mathcal{F}_n) = \frac{R_n}{R_n + \overline{\pi}_n} \,,$$

where $R_n = \sum\limits_{i=0}^{n} \pi_i h_{i,n}$ *and* $\overline{\pi}_n = \sum\limits_{i=n+1}^{N} \pi_i$.

Also let us note that Roberts statistics R_n can be calculated recursively.

Proposition 2. *For the calculating Roberts statistics R_n one can use the following formula*

$$R_n = \eta_n \sum_{i=0}^{n-1} \pi_i \, h_{i,n-1} + \pi_n = \eta_n R_{n-1} + \pi_n \,,$$

where $\eta_j = \eta(X_j, X_{j-1})$, $\eta(y,x) = f(y,x)/f^*(y,x)$ *and* $R_0 = \pi_*$.

The main goal is to find the disruption moment (starting of the epidemic) as soon as possible with fixed probability of false alarm, i.e.

$$\widetilde{\mathbf{P}}(\tau < \nu) \leq \alpha \,,$$

where $0 \leq \alpha \leq 1$ and τ is the stopping time from the set $\mathcal{M}_\alpha$ of all stopping times which satisfy the fixed false alarm probability condition.

Thus, one needs to solve the following optimization problem

$$\inf_{\tau \in \mathcal{M}_\alpha} \widetilde{\mathbf{E}}(\tau - \nu)_+ \,, \tag{1}$$

where $(x)_+ = \max(x, 0)$.

Using the Lagrange method we can obtain the following problem statement

$$\min_{\tau \in \mathcal{M}} \left(\lambda \widetilde{\mathbf{E}}(\tau - \nu)_+ + \widetilde{\mathbf{P}}(\tau < \nu) \right) \,.$$

Moreover, this statement has another representation.

Lemma 3.

$$\min_{\tau \in \mathcal{M}} \left(\lambda \widetilde{\mathbf{E}}(\tau - \nu)_+ + \widetilde{\mathbf{P}}(\tau < \nu) \right) = 1 - \max_{\tau \in \mathcal{M}} \mathbf{E}^* G_\tau \,,$$

where $G_n = R_n - \lambda \sum\limits_{i=0}^{n-1} R_i$ *and* $\mathbf{E}^*$ *is the expectation over the probability* $\mathbf{P}^*$.

Thus, we got the opportunity to pass from a Bayesian probability measure to a measure in case if the disruption has not yet occurred. Moreover, thanks to this, now instead of considering the disruption problem in a Markov process $(X_n)_{0\le n\le N}$, we can consider the problem of optimal stopping problem for a homogeneous Markov process $(Z_n)_{0\le n\le N}$, where $Z_n = (R_n, X_n)$.

Remark 1. *In fact, the process $(Z_n)_{0\le n\le N}$ will not always be homogeneous Markov process, however, in order to achieve this property, this paper considers the binomial model proposed by M. Baron in [1], in which the necessary property of the process is proved.*

To solve the problem of optimal stopping for a homogeneous Markov process, this paper uses the approach developed by A. Shiryaev in [2], which consists in solving the following problems for $0 \le n \le N$

$$\sup_{\tau\in\mathcal{T}_n} \mathbf{E}^*_z \mathbf{G}_\tau \,, \tag{2}$$

where $\mathbf{E}^*_z$ is the mathematical expectation under condition that at the initial moment of time the value of the process $Z_0 = z$, that is, more specifically

$$\mathbf{E}^*_z = \mathbf{E}^* \left(h(\cdot) \,|\, Z_0 = z\right) \,,$$

and also

$$\mathcal{T}_n = \{\tau \in \mathcal{M}_1 \;:\; \tau \le n \quad \mathbf{P}^* - \text{a.s.}\} \,.$$

According to the Shiryaev's approach, to solve the optimal stopping problem (2) it is necessary to construct the Snell envelope for the process $(G_k)_{0\le k\le n}$. To do this for some function h for which the condition $\mathbf{E}^*_z \,|h(Z)|$ is true, where $z = (r, x)$, we define an operator of the following form

$$\mathbf{Q}_\lambda(h)(z) = \max\left(h(z)\,,\, \mathbf{T}(h)(z) - \lambda r\right) \,,$$

where $\mathbf{T}(h)(Z) = \mathbf{E}^*_z\, h(Z)$.

Lemma 4. *Let condition $\mathbf{E}^*_z \,|h(Z)|$ holds true for some function h : $\mathbb{R}_+ \times \mathcal{X} \to \mathbb{R}$. Then for $n \ge 1$*

$$\mathbf{Q}^n_\lambda(h)(z) = \max\left(h(z), \mathbf{T}(\mathbf{Q}^{n-1}_\lambda(h))(z) - \lambda\, r\right) \,.$$

Lemma 5. *The following process*

$$Y_k = \mathbf{Q}^{n-k}_\lambda(h)(Z_k) - \lambda \sum_{i=0}^{k-1} R_i \,.$$

is the Sell envelope for the process $(G_k)_{0\le k\le n}$.

In this case, according to the approach under consideration, we can define the following sequential detection procedure

$$\tau^*_n = \min\left\{0 \le k \le n :\, \mathbf{Q}^{n-k}_\lambda(g)(R_k, X_k) = R_k\right\} \,, \tag{3}$$

where $g(r, x) = r$.

The most significant results of this paper are presented in the following theorems

Theorem 1. *For any* $\lambda > 0$ *stopping moment* τ_n^* *gained by the procedure* (3) *is optimal, i.e. it is the solution of the problem* (2)

$$\sup_{\tau \in \mathcal{T}_n} \mathbf{E}_z^* \mathbf{G}_\tau = \mathbf{E}_z^* \mathbf{G}_{\tau_n^*} .$$

Theorem 2. *If exists such* λ_α *which satisfies the fixed false alarm probability condition* $\tilde{\mathbf{P}}(\tau < \nu) < \alpha$, *then the result of the following sequential procedure*

$$\mathbf{t}_\lambda^* = \min\left\{k \geq 0 : \mathbf{Q}_\lambda^{N-k}(g)(R_k, X_k) = R_k\right\} \tag{4}$$

is the optimal (non-asymptotic!) solution of the main problem (1) *of disruption detection for Markov process with fixed probability of false alarm.*

Moreover, in this paper was found the method of calculating such λ_α. During investigation was noted that

Proposition 6.

$$\tilde{\mathbf{P}}\left(\mathbf{t}_\lambda^* \leq \nu\right) = \pi_* \sum_{m=1}^{N-1} \mathbf{E}^* U_{m,\lambda}(k_1, \ldots, k_m) ,$$

where

$$U_{m,\lambda}(k_1, \ldots, k_m) = \mathbf{1}_{\left\{\min_{1 \leq j \leq m}\left(\mathbf{Q}_\lambda^{N-j}(g)(r_j, k_j) - r_j\right) = 0\right\}} .$$

This way, by defining the function of the following view

$$\mathbf{F}(\lambda) = \tilde{\mathbf{P}}\left(\mathbf{t}_\lambda^* \leq \nu\right) ,$$

we can calculate the value λ_α^* that satisfies to the condition $\tilde{\mathbf{P}}\left(\tau < \nu\right) < \alpha$ from the following procedure

$$\lambda_\alpha^* = \sup\{\lambda \geq 0 : \mathbf{F}(\lambda) \leq \alpha\} . \tag{5}$$

Then, by using the value λ_α^*, we can gain non-asymptotic optimal solution of the problem (1), i.e.

$$\inf_{\tau \in \mathcal{M}_\alpha} \widetilde{\mathbf{E}}(\tau - \nu)_+ = \widetilde{\mathbf{E}}(\lambda_\alpha^* - \nu)_+ .$$

To apply this theory to a real epidemiological problem, according to the considered binomial model, the process $(X_n)_{0 \leq n \leq N}$ can be represented as a Galton-Watson process (see [3]), that is, the initial value of the process is fixed $X_0 = D$, where D is the number of susceptible people at the initial time, and the remaining observations X_n for $1 \leq n \leq N$ are calculated using the following formula

$$X_n = S_{n, X_{n-1}}$$

and

$$S_{n,m} = \sum_{i=1}^{m} \xi_{n,i} ,$$

where $\left(\xi_{n,i}\right)_{1 \leq n \leq N, i \geq 1}$ is independent identically distributed Bernoulli's random values, for which $\mathbf{P}(\xi_{n,i} = 1) = 1 - \vartheta_n$ and $\vartheta_n = \theta_* \mathbf{1}_{n \leq \nu} + \theta \mathbf{1}_{\nu > n}$.

Thus, to detect the epidemic one needs, firstly, by using the procedure (5), calculate value λ^*_α, then calculate all requirement values from the sequence $\left(\mathbf{Q}^{N-k}_{\lambda^*_\alpha}(g)(Z_k)\right)_{0\le k\le n}$ and finally get the solution as an optimal stopping moment gained from the procedure (4).

References

[1] Baron M., Choudhary K., Yu X. Change-Point Detection in Binomial Thinning Processes, with Applications in Epidemiology // Sequential Analysis: Design Methods and Applications. 2013. V. 32. P. 350-367.

[2] Shiryaev A.N. Optimal Stopping Rules // New York: Springer. 1978.

[3] Baron M. Bayes and asymptotically pointwise optimal stopping rules for the de-tection of influenza epidemics. // In. in Case Studies in Bayesian Statistics, vol. 6, C. Gatsonis, R. E. Kass, A. Carriquiry, A. Gelman, D. Higdon, D. K. Pauler, and I. Verdinelli, eds. // New York: Springer. 2002. P. 153-163.

Теньзин Р., Пчелинцев Е., Пергаменщиков С. (Томский государственный университет, Томск, Россия, 2022)
 Раннее выявление эпидемий

Аннотация. В работе рассматривается задача построения оптимальной процедуры выявления разладки марковского процесса, представляющего собой эпидемиологическую динамику числа инфицированных. Аналогичная постановка задачи рассмотрена и в [1].

Ключевые слова: разладка в цепи Маркова, оптимальная остановка, байесовский риск, обнаружение эпидемий.