Anomaly Identification with Limited Sampling Budget

Julia Kuhn*†, Michel Mandjes†, and Thomas Taimre*
*School of Mathematics and Physics
The University of Queensland
St. Lucia, QLD, Australia
Email: j.kuhn@uq.edu.au, t.taimre@uq.edu.au
†Korteweg-de Vries Institute for Mathematics
University of Amsterdam
The Netherlands
Email: m.r.h.mandjes@uva.nl

***Abstract*—We consider a network of data streams from which an anomalous process with known target distribution is to be identified. Motivated by the realisation that in practice obtaining observations may be expensive, we assume that there is a constraint on the total number of observations based on which the decision has to be made. We derive a sufficient condition on the sampling budget such that the error probability is kept below some desired level. Furthermore, we show how to obtain a sampling allocation that can improve upon equal sampling allocation and achieves the desired accuracy.**

I. Introduction

Consider a network of d processes which are to be monitored with the objective to identify the anomalous process, that is, the process that stems from a given target distribution G rather than the reference distribution F. We may, for example, be interested in identifying an idle channel in a network of communication channels [1], the presence of a certain animal species in one of a number of monitored habitats, the drug that is efficient in curing a certain disease, and many other possible applications.

It is known that if only a single process is to be monitored and the problem is to decide as quickly as possible whether or not it stems from a target distribution subject to a constraint on the error probability, Wald's sequential probability ratio test (SPRT) is optimal [2]. In such a sequential setting, target identification in networks of multiple processes has been considered in [3], [4] under the assumption that all processes are observed at every time point.

In practice, it may often be the case that obtaining an observation is expensive. For example, in decentralized sensor networks there is usually a cost associated with the communication between sensors and the fusion center [5]. In many other application areas, such as biology, physics, medicine, psychology, geology, and ecology, observations can often only be obtained from experiments that involve human intervention, in which case an observation is particularly costly. In medical applications it often occurs that only a limited number of samples are available, e.g. for diagnostic testing or for testing the effectiveness of a certain medication. Such considerations motivate the investigation of how an anomalous process or target can be identified efficiently based on only a limited number of observations.

In a sequential setting, where one samples until the test statistic exceeds a certain threshold, [5] investigated the question of efficient sampling allocation in the context of change point detection, where the anomaly is not present initially but may appear at some unknown time point. In [6] a sequential procedure for sampling allocation was proposed subject to the constraint that only a number $k < d$ of all data streams can be observed at every time point. In contrast, in this paper we do not consider a sequential testing problem but instead assume that the decision has to be made as accurately as possible with a given sampling budget.

It seems plausible that a good sampling allocation should explicitly take into account the specific characteristics of each process. For example, processes with a high variance should be sampled more often; and more samples should be taken from processes that are more similar to the anomalous process. Motivated by such considerations we are interested in determining an allocation $\boldsymbol{\rho} := (\rho_1, \ldots, \rho_d) \in (0,1)^d$ with $\sum_{i=1}^d \rho_i = 1$ such that the decision can be made with the desired accuracy based on $\rho_i n$ samples from process $i \in \{1, \ldots, d\}$. (We neglect the minor technical issues arising when $\rho_i n$ is not integer-valued.)

We assume no prior knowledge as to which of the data streams in the network is anomalous (as opposed to a Bayesian setting where one has a prior belief about the characteristics of the data streams). In this framework, initially all processes have to be observed for a certain amount of time so as to collect information about their nature. We therefore first derive a bound on the number of samples needed to obtain the desired minimal security about which process is the anomalous one. Then, in a second step, we show how the remaining sampling budget can be allocated so as to optimize the accuracy of the identification.

As is common in the literature on anomaly identification [3], [4] we assume that we know which behaviour should be characterised as anomalous and which behaviour is normal, meaning we know the distribution of the data streams in both cases. This is a reasonable assumption for example for the problem of searching the idle channel in a communication network [6].

The methodology we use is motivated by the work of [7], [8] on ordinal optimization. They consider the problem of finding the process with the largest sample mean from a given set of independent and identically distributed (i.i.d.) observations, subject to a constraint on the available sampling budget.

The paper is organized as follows. In Section II we explain the problem and propose an algorithm for sampling allocation and target identification that provably does better than a pre-specified accuracy. In Sections III and IV we explain how the steps of the algorithms can be carried out. We conclude in Section V.

II. Problem Formulation and Sampling Algorithm

Denote the observation of process i at time n by $X_i(n)$. We assume that the sequences $\big(X_i(n)\big)_n$ and $\big(X_j(n)\big)_n$ are independent for $i \neq j$. Unless otherwise stated, the observations need not be independent over time. Without loss of generality, we assume that the anomalous process is the process labelled by 1. Suppose that $X_2(n), \ldots, X_d(n)$ have a distribution with density f, while $X_1(n)$ has a distribution with density g. Motivated by log-likelihood ratio (LLR) hypothesis testing, we declare process i to be anomalous based on a total of N observations sampled according to $\boldsymbol{\rho}$ if

$$\mathscr{L}_i(N\rho_i) := \frac{1}{N\rho_i} \sum_{n=1}^{N\rho_i} \ell_i(n) \geq \mathscr{L}_j(N\rho_j)\,,$$

with $i, j \in \{1, \ldots, d\}$, where

$$\ell_i(n) := \log \frac{g\big(x_i(n) \big| x_i(1), \ldots, x_i(n-1)\big)}{f\big(x_i(n) \big| x_i(1), \ldots, x_i(n-1)\big)}$$

denote the LLR increments. We assume that $\mathbb{E}\big[\ell_1(1)\big] > \mathbb{E}\big[\ell_j(1)\big]$, $j \in \{2, \ldots, d\}$, so that the anomalous process is indeed distinguishable. This assumption was also imposed in [7].

Suppose the total sampling budget is N. Then, since at the beginning of the testing it is not known which process is from the target distribution, the naive approach is to allocate an equal number of samples to each process. Later we will see, however, that we can easily improve upon this simple approach.

The event of a false selection (FS), i.e. of declaring the wrong process to be anomalous, based on n samples is given by

$$\mathrm{FS}(n, \boldsymbol{\rho}) := \left\{ \mathscr{L}_1(n\rho_1) < \max_{i \in \{2, \ldots, d\}} \mathscr{L}_i(n\rho_i) \right\}\,.$$

We propose the following "algorithm" for determining the sampling allocation $\boldsymbol{\rho}$ such that $\mathrm{FS}(N, \boldsymbol{\rho})$ is small and guaranteed to be below a chosen level α.

(A) Fix α. Observe all processes until time $\min\{\bar{n}/d, N/d\}$, where $\bar{n}$ is such that $\mathbb{P}(\mathrm{FS}(\bar{n}, \boldsymbol{\rho}_d)) \leq \alpha$ for $\boldsymbol{\rho}_d := (1/d, \ldots, 1/d)$.

(B) If $\bar{n} \leq N$, distribute the remaining sampling budget according to the allocation $\boldsymbol{\rho}^*$ that solves

$$\begin{aligned} &\min_{\boldsymbol{\rho}} \mathbb{P}\big(\mathrm{FS}(N, \boldsymbol{\rho})\big) \\ &\text{s. t.} \quad \sum_{i=1}^{d} \rho_i = 1\,, \quad \rho_i N \geq \rho_i \bar{n}\ \forall\, i\,. \end{aligned} \tag{1}$$

Provided that the LLR increments satisfy the conditions of the law of large numbers, it is clear that $\bar{n}$ from Step (A) of the algorithm indeed exists. If the false selection probability decreases monotonically in n for all $n \geq \bar{n}$, then the false selection probability achieved using the above algorithm based on N samples is guaranteed to be below α. This motivates that we investigate the monotonicity of the false selection probability in the remainder of this section.

If $\mathbb{P}\big(\mathrm{FS}(n, \boldsymbol{\rho})\big)$ is not everywhere decreasing in n, we can still ensure that $\mathbb{P}\big(\mathrm{FS}(N, \boldsymbol{\rho})\big) \leq \alpha$ by deriving $\bar{n}$ in Step (A) from an upper bound on the false selection probability that does decrease monotonically in n; we provide such a bound in Section III.

In the remainder of this section we focus on the case of i.i.d. observations. Denote the mean and variance of $\ell_i(1)$ by μ_i and σ_i^2, respectively. We introduce the random process

$$Z_j(n) := \mathscr{L}_1(n\rho_1) - \mathscr{L}_j(n\rho_j)\,, \quad j \in \{2, \ldots, d\}\,.$$

Then the false selection probability is everywhere monotonically decreasing in n if

$$\mathbb{P}\big(Z_j(n+1) \leq 0\big) \leq \mathbb{P}\big(Z_j(n) \leq 0\big) \tag{2}$$

holds for all $n \in \mathbb{N}$, $j \in \{2, \ldots, d\}$. The mean of $Z_j(n)$ is $a_j := \mu_1 - \mu_j > 0$ and the variance is $n^{-1} v_j^2(\boldsymbol{\rho})$, where $v_j^2(\boldsymbol{\rho}) := \sigma_1^2/\rho_1 + \sigma_j^2/\rho_j$. We further define $v_j^2 := \sigma_1^2 + \sigma_j^2$.

We can check that (2) holds for all $n \in \mathbb{N}$ in the case of Gaussian random variables (Lemma 1). The distribution function of the standard normal distribution is denoted by Φ.

Lemma 1. *For each i assume that $X_i(n)$ are i.i.d. Gaussian random variables. Then the false selection probability decreases monotonically as a function of n.*

Proof: By assumption the processes $Z_j(n)$ are Gaussian as a convolution of independent Gaussian random variables. Then (2) is equivalent to

$$\Phi\left(-\frac{\sqrt{n}a_j}{v_j(\boldsymbol{\rho})}\right) \geq \Phi\left(-\frac{\sqrt{n+1}a_j}{v_j(\boldsymbol{\rho})}\right)\,,$$

which clearly holds for any $n \in \mathbb{N}$ since $a_j > 0$. ∎

In general, the false selection probability need not be decreasing. We can show, however, that $\mathbb{P}(Z_j(n) \leq 0)$

lies within an interval the center of which is decreasing in n, and with bounds that become increasingly tight as n grows.

To this end, note that $Z_j(n)$ is a sample mean of n independent random variables

$$Y_j(t) := \frac{\ell_1(t)}{\rho_1}\mathbb{1}\{t \leq n\rho_1\} - \frac{\ell_j(t)}{\rho_j}\mathbb{1}\{t \leq n\rho_j\}\,,$$

with mean $a_j(t) > 0$ and variance $v_j(t)^2$. Then the following lemma readily follows from the Berry-Esséen theorem [9].

Lemma 2. *Let $X_i(t)$ be i.i.d. random variables such that $a_j(t) > 0$, $v_j^2(t) < \infty$ for $t \in \mathbb{N}$, and*

$$\sup_{t\in\mathbb{N}} \mathbb{E}\left[|Y_j(t) - a_j(t)|^3\right] < B\,,$$

for some $B < \infty$. Then $\mathbb{P}(Z_j(n) \leq 0)$ is bounded by

$$\Phi\left(-\sqrt{n}\,\frac{a_j}{v_j}\right) \pm \frac{3\,B}{5v_j^3\sqrt{n}}\,. \tag{3}$$

The condition on the absolute third moment is very mild; for example, for i.i.d. Gaussian observations the absolute third moments are bounded by $2v_j^3\sqrt{2/\pi}$.

In Sections III and IV we discuss how to carry out Steps (A) and (B) of the algorithm, respectively.

III. Sufficient Sampling Budget

In this section, for a given allocation $\boldsymbol{\rho}$ we derive a bound on the total number of observations that ensures that the false selection probability is kept below a desired level α. That is, we want to determine $n_{\boldsymbol{\rho}}$ such that

$$\mathbb{P}\big(\text{FS}(n_{\boldsymbol{\rho}}, \boldsymbol{\rho})\big) \leq \alpha\,.$$

(Such an $n_{\boldsymbol{\rho}}$ exists because we assumed that $\mu_1 > \mu_j$ for $j \in \{2,\ldots,d\}$.)

In special cases, it may possible to compute the false selection probability explicitly by inducing independence via conditioning on the value of $\mathscr{L}_1(n\rho_1)$ (we leave this for future research). In general, however, it is difficult to evaluate $\mathbb{P}\big(\text{FS}(n,\boldsymbol{\rho})\big)$, and therefore we now show how it can be approximated using large deviations (LD) arguments when the given sampling budget n is large.

Let

$$I_i(x) := \sup_{\theta\in\mathbb{R}} \{\theta x - \Lambda_i(\theta)\}$$

denote the Fenchel-Legendre transformation of the limiting cumulant generating function Λ_i of the LLR of process i,

$$\Lambda_i(\theta) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{\theta\,\mathscr{L}_i(n)}\right]\,.$$

Assume that $\Lambda_i(\cdot)$ is everywhere differentiable and exists as a finite number for every $\theta \in \mathbb{R}$.

Proposition 1. *The false selection probability can be approximated as*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(\text{FS}(n,\boldsymbol{\rho})\big) = -\min_{j\in\{2,\ldots,d\}} G_j(\boldsymbol{\rho})\,, \tag{4}$$

where

$$G_j(\boldsymbol{\rho}) := \inf_x \{\rho_1 I_1(x) + \rho_j I_j(x)\}\,. \tag{5}$$

Proof: Using that $\lim_{n\to\infty} n^{-1}\log(d-1) = 0$, similar to [10, Lemma 1.2.15] it is readily obtained that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(\text{FS}(n,\boldsymbol{\rho})\big) = \max_{j\in\{2,\ldots,d\}} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\mathscr{L}_j(n\rho_j) > \mathscr{L}_1(n\rho_1)\right)\,. \tag{6}$$

From the Gärtner-Ellis theorem [10], we have that for $x > \mathbb{E}\ell_i(1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(\mathscr{L}_i(n\rho_i) > x\big) = -\rho_i I_i(x)\,.$$

Let $\mathcal{I}_j(\boldsymbol{x}) := \rho_1 I_1(x_1) + \rho_j I_j(x_j)$. Because $\mathscr{L}_j$ and $\mathscr{L}_1$ are independent, it follows that for $B \subset \mathbb{R}^2$ such that $\inf_{x\in B^o} \mathcal{I}_i(x) = \inf_{x\in\bar{B}} \mathcal{I}_i(x) =: \mathcal{I}_i(B)$ we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\big(\mathscr{L}_1(n\rho_1), \mathscr{L}_j(n\rho_j)\big)' \in B\right) = -I_j(B)\,.$$

Using the properties of the rate functions I_i, we can then argue as in [7, Section 2.2] that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\mathscr{L}_j(n\rho_j) > \mathscr{L}_1(n\rho_1)\right) = -G_j(\boldsymbol{\rho})\,.$$

This, together with (6) proves (4). ■

Let $D_{\Lambda_i} := \{\theta \in \mathbb{R} : \Lambda_i(\theta) < \infty\}$ and $\mathcal{F}_i := \{\Lambda_i'(\theta) : \theta \in D_{\Lambda_i^o}\}$, where A^o denotes the interior of set A. It is well known [10] that the Fenchel-Legendre transform is strictly convex and C^∞ for $x \in \mathcal{F}_i^o$, $I_i(\mu_i) = 0$ and $I_i(x) \geq 0$ for all $x \in \mathbb{R}$.

The following lemma can be proven analogously to [7, Lemma 3]. Loosely speaking, assumption (7) of the lemma states that the LLRs can take any value in the interval $[\mu_d, \mu_1]$; in particular, it implies that $\mathbb{P}\big(Z_j(n\rho_j) \leq 0\big) > 0$. It holds for example if the distribution of $\ell_i(t)$ stems from the Normal or the Gamma family.

Lemma 3 (Glynn and Juneja [7])**.** *Assume that all observations are i.i.d., and that*

$$[\mu_d, \mu_1] \subset \bigcap_{i=1}^{d} \mathcal{F}_i^o\,. \tag{7}$$

Then for a given allocation $\boldsymbol{\rho}$ and sampling budget n we have

$$\mathbb{P}\big(\text{FS}(n,\boldsymbol{\rho})\big) \leq (d-1)\exp\left(-n \min_{j\in\{2,\ldots,d\}} G_j(\boldsymbol{\rho})\right)\,. \tag{8}$$

Thus, a lower bound for the minimal value of n can be achieved by putting the upper bound given in (8) equal to α, and solve for n. Note that this will yield a function of n that depends on the allocation $\boldsymbol{\rho}$.

In Step (A) of the algorithm proposed in the previous section we assume $\rho_i = 1/d$ for all i. We provide an example below.

a) Example: Suppose the observations of each stream are i.i.d. and normally distributed with $X_i(t) \sim \mathcal{N}(m_i, s^2)$ for $i \in \{1, \dots, d\}$, where $m_1 = \widetilde{m}$, whereas $m_j = m$ for $j \in \{2, \dots, d\}$. It is easy to see that the LLR increments are

$$\ell_i(t) = \frac{\widetilde{m} - m}{s^2} \left(X_i(t) - \frac{m + \widetilde{m}}{2} \right),$$

and hence $\ell_i(t) \sim \mathcal{N}\left(\mu_i, \sigma^2\right)$, where

$$\mu_i := \frac{\widetilde{m} - m}{s^2} \left(m_i - \frac{m + \widetilde{m}}{2} \right), \quad \sigma := \frac{\widetilde{m} - m}{s}.$$

Then the LLRs at time n have distribution $\mathscr{L}_i(\rho_i n) \sim \mathcal{N}\big(\mu_i, \sigma^2/(\rho_i n)\big)$. The cumulant generating function is

$$\Lambda_i(\theta) = \theta\, \mu_i + \frac{1}{2}\sigma^2\, \theta^2,$$

so that

$$I_i(x) = \frac{(x - \mu_i)^2}{2\sigma^2}.$$

Therefore, we obtain that

$$G_i(\boldsymbol{\rho}) = \frac{(\mu_i - \mu_1)^2}{2\sigma^2(1/\rho_1 + 1/\rho_j)}.$$

Consider the allocation $\boldsymbol{\rho} = (1/d, \dots, 1/d)$. From (8) we know that if n satisfies

$$\alpha = (d-1) \exp\left(-n G_j(\boldsymbol{\rho})\right)$$

for arbitrary $j \in \{2, \dots, d\}$, then $\mathbb{P}\big(\mathrm{FS}(n, \boldsymbol{\rho})\big) \leq \alpha$. Solving for n we obtain that under allocation $\boldsymbol{\rho}$ we can make a decision with the desired accuracy if

$$n \geq \bar{n} := \frac{2\sigma^2\left(1/\rho_1 + 1/\rho_j\right)}{(\mu_j - \mu_1)^2} \log\left(\frac{d-1}{\alpha}\right).$$

Fig. 1 shows the values of $\bar{n}$ obtained with equal allocation. Naturally, the value of $\bar{n}$ decreases as the difference between the anomalous process and the other processes increases. The achieved false selection probabilities are very conservative. The jumps are due to the rounding of $\bar{n}/d$.

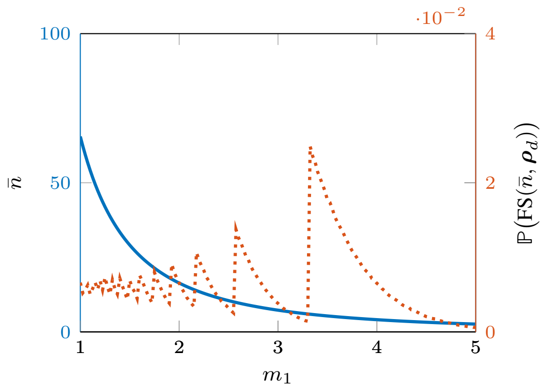

Figure 1. We plot the obtained values of $\bar{n}$ as a function of $m_1 = \mathbb{E}X_1(t)$ for a network with 4 processes under equal allocation (solid line). Other parameters are set as $m_j = 0$ for $j = 2, 3, 4$, $s = 1$, $\alpha = 0.01$. The dotted line shows the simulated values of $\mathbb{P}\big(\mathrm{FS}(\bar{n}, \boldsymbol{\rho}_d)\big)$, with values on the right vertical axis.

IV. Asymptotically Optimal Allocation

An asymptotically optimal allocation $\boldsymbol{\rho}^*$ that (approximately) solves (1) if N is large can be found by minimizing the rate function given in (4). Note that the rate function is concave in $\boldsymbol{\rho}$ as a minimum over affine functions. We can formulate the optimization problem as follows:

$$\begin{aligned} \max \quad z \quad & \text{s. t.} \\ G_j(\boldsymbol{\rho}) - z &\geq 0 \\ \sum_{i=1}^{d} \rho_i - 1 &= 0 && (9) \\ \rho_i N - \bar{n}/d &\geq 0\,. && (10) \end{aligned}$$

From the Karush-Kuhn-Tucker conditions [11] we know that there exists multipliers μ_j, η_i and λ such that

$$\sum_{j=2}^{d} \mu_j \frac{\partial G_j(\boldsymbol{\rho}^*)}{\partial \rho_1} + \eta_1 N = \lambda \tag{11}$$

$$\mu_j \frac{\partial G_j(\boldsymbol{\rho}^*)}{\partial \rho_j} + \eta_j N = \lambda\,, \quad j \in \{2, \dots, d\} \tag{12}$$

$$1 - \sum_{j=2}^{d} \mu_j = 0 \tag{13}$$

$$\mu_j\big(z - G_j(\boldsymbol{\rho}^*)\big) = 0\,, \quad j \in \{2, \dots, d\} \tag{14}$$

$$\eta_i\left(\bar{n}/d - \rho_i^* N\right) = 0\,, \quad i \in \{1, \dots, d\}\,. \tag{15}$$

This yields conditions on the asymptotically optimal allocation, which we formalize in the following proposition.

Proposition 2. *Let $N > \bar{n}$. If an allocation $\boldsymbol{\rho}^*$ with $\sum_{i=1}^{d} \rho_i^* = 1$ and $\rho_i^* \geq \bar{n}/(dN)\, \forall i$ minimizes $\mathbb{P}\big(\mathrm{FS}(N, \boldsymbol{\rho})\big)$, then*

$$G_j(\boldsymbol{\rho}^*) = G_k(\boldsymbol{\rho}^*) \tag{16}$$

for $k, j \in \{2, \dots, d\}$ such that $\rho_k^*, \rho_j^* \neq \bar{n}/(dN)$. If $\rho_i^* \neq \bar{n}/(dN)\, \forall i$, we also have

$$\sum_{j=2}^{d} \frac{\partial G_j(\boldsymbol{\rho}^*)/\partial \rho_1}{\partial G_j(\boldsymbol{\rho}^*)/\partial \rho_j} = 1\,. \tag{17}$$

Proof. From (13) we have that there exists $j \in \{2, \dots, d\}$ such that $\mu_j > 0$. Because $\partial G_j(\boldsymbol{\rho}^*)/\partial \rho_j > 0$, together with (12) this implies that that $\lambda > 0$.This means that for $j \geq 2$, if $\eta_j = 0$, then $\mu_j > 0$, in which case by (14) we have (16). By (15), $\eta_j = 0$ holds if $\rho_j^* \neq \bar{n}/(dN)$.

The latter also implies that if $\rho_j^* \neq \bar{n}/(dN)$ for all i, then $\eta_i = 0$ for all i, in which case (12) implies that $\mu_j = \lambda/\big[\partial G_j(\boldsymbol{\rho}^*)/\partial \rho_j\big]$. Substituting this into (11) then gives (17). $\square$

Note that since we assumed that the processes $2, \dots, d$ are stochastically identical, (16) implies that $\rho_2 = \dots = \rho_d =: \widetilde{\rho}$ (as one would expect). If (16) and (17) yield a feasible solution (i.e. an allocation that satisfies (9) and (10)), then this solution is optimal because the optimization problem is concave and thus the Karush-Kuhn-Tucker conditions are sufficient. Hence, in this case we assign $\rho_1^* N$ samples to the process that had the largest LLR after Step (A) of the algorithm, and ρ_j^* samples to all others. Otherwise, the optimal solution is $\rho_i^* = \bar{n}/(dN)$ for all i except the process that yielded the largest LLR with the first $\bar{n}/(dN)$ observations, which then has allocation $1-(d-1)\bar{n}/(dN)$.

b) Example: We assume that the observations are i.i.d. Gaussian so that the LLR increments have distribution $\mathcal{N}(\mu_i, \sigma_i^2)$, where processes $2, \ldots, d$ are assumed to be stochastically identical. We compute the allocation $\boldsymbol{\rho}^*$ obtained based on LD approximations as suggested above. For illustration purposes, we assume that $\bar{n} = 0$. Note that $\rho_i^* > 0$ because otherwise $\min_i G_i(\boldsymbol{\rho}^*) = 0$ while we know that $G_i(\boldsymbol{\rho}_d) > 0$ for every i. As in Example 1 in [7] we obtain from (16) and (17) that $\boldsymbol{\rho}^*$ satisfies the conditions

$$\rho_1^* = \sigma_1 \sqrt{\sum_{j=2}^{d} \frac{\rho_j^{*2}}{\sigma_j^2}}\,, \qquad \sum_{i=1}^{d} \rho_i = 1\,,$$

and

$$(\mu_j - \mu_1)^2 \left(\frac{\sigma_1^2}{\rho_1^*} + \frac{\sigma_k^2}{\rho_k^*} \right) = (\mu_k - \mu_1)^2 \left(\frac{\sigma_1^2}{\rho_1^*} + \frac{\sigma_j^2}{\rho_j^*} \right)$$

for $j, k \in \{2, \ldots, d\}$. Thus, the asymptotically optimal allocation $\boldsymbol{\rho}^*$ can be obtained by solving the resulting system of equations.

Since we assume that processes $j \in \{2, \ldots, d\}$ are identically distributed, we readily obtain

$$\begin{aligned} \rho_j &= \frac{\sigma_j}{\sigma\sqrt{d-1} + \sigma_j(d-1)}\,, \qquad j \in \{2, \ldots, d\} \\ \rho_1 &= \frac{\sigma_1}{\sigma_j}\sqrt{d-1}\,\rho_j\,. \end{aligned}$$

Note that the budget allocated to the anomalous process increases as σ_1 increases.

We now compare the false selection probabilities obtained numerically under $\boldsymbol{\rho}^*$ with those achieved under equal allocation, see Fig. 2. Naturally, the false selection probabilities overall decrease as m_1 and thus the difference between the means of the processes increases. We further note that $\mathbb{P}\big(\mathrm{FS}(10, \boldsymbol{\rho}_d)\big)$ is generally greater than $\mathbb{P}\big(\mathrm{FS}(10, \boldsymbol{\rho}^*)\big)$; the performance gain is more than 10%.

V. Conclusion

We proposed an algorithm for identifying a target process in a network of independent data streams subject to a constraint on the total number of observations. We showed that the probability of false selection can be decreased substantially compared to equal allocation of samples.

Future research should investigate the impact of the specific characteristics of each process on the optimal allocation. It should also generalize the results

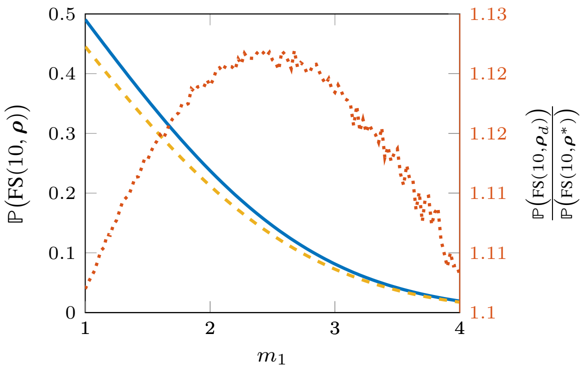

Figure 2. False selection probabilities achieved under equal allocation $\boldsymbol{\rho}_d$ (solid line) and under asymptotically optimal allocation $\boldsymbol{\rho}^*$ (dashed line) as a function of $m_1 := \mathbb{E}X_i(t)$ for a network with 4 processes under equal allocation. Their ratio is also depicted (dotted line, values on the right vertical axis). Other parameters are set as $m_j = 0$ for $j = 2, 3, 4$, $s = 2$, $N = 10$.

we provided: For example, the assumption of i.i.d. observations that we imposed in some places can certainly be relaxed. Also, our work can be extended to the setting where an unknown number of anomalous processes is present, in which case a process is declared anomalous when the LLR test statistic exceeds an appropriately specified threshold.

Acknowledgments

Julia Kuhn is supported by Australian Research Council (ARC) grant DP130100156. Michel Mandjes' research is partly funded by the NWO Gravitation Project NETWORKS grant 024002003.

References

[1] K. Cohen and Q. Zhao, "Asymptotically optimal anomaly detection via sequential testing," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2929–2941, 2015.

[2] D. Siegmund, *Sequential Analysis*. Springer-Verlag New York, 1985.

[3] L. Lai, H. V. Poor, Y. Xin, and G. Georgiadis, "Quickest search over multiple sequences," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5375–5386, 2011.

[4] A. G. Tartakovsky, X. R. Li, and G. Yaralov, "Sequential detection of targets in multichannel systems," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 425–445, 2003.

[5] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest outlying sequence detection in sensor networks," *Submitted, arXiv preprint arXiv:1411.0183*, 2014.

[6] Q. Zhao and J. Ye, "Quickest detection in multiple on–off processes," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 5994–6006, 2010.

[7] P. Glynn and S. Juneja, "A large deviations perspective on ordinal optimization," in *Proceedings of the 2004 Winter Simulation Conference*, vol. 1. IEEE, 2004.

[8] ——, "Ordinal optimization-empirical large deviations rate estimators, and stochastic multi-armed bandits," *arXiv preprint arXiv:1507.04564*, 2015.

[9] I. G. Shevtsova, "An improvement of convergence rate estimates in the Lyapunov theorem," *Doklady Mathematics*, vol. 82, no. 3, pp. 862–864, 2010. [Online]. Available: http://dx.doi.org/10.1134/S1064562410060062

[10] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Springer-Verlag, New York, 1998.

[11] M. Avriel, *Nonlinear programming: analysis and methods*. Courier Corporation, 2003.