Sensing Method for Two-Target Detection in Time-Constrained Vector Poisson Channel

Muhammad Fahad, *Member, IEEE*
Department of Applied Computing
Michigan Technological University
Houghton, MI 49931, USA
mfahad@mtu.edu

Daniel R. Fuhrmann, *Fellow, IEEE*
Department of Applied Computing
Michigan Technological University
Houghton, MI 49931, USA
fuhrmann@mtu.edu

Abstract

It is an experimental design problem in which there are two Poisson sources with two possible and known rates, and one counter. Through a switch, the counter can observe the sources individually or the counts can be combined so that the counter observes the sum of the two. The sensor scheduling problem is to determine an optimal proportion of the available time to be allocated toward individual and joint sensing, under a total time constraint. Two different metrics are used for optimization: mutual information between the sources and the observed counts, and probability of detection for the associated source detection problem. Our results, which are primarily computational, indicate similar but not identical results under the two cost functions.

Index Terms

sensor scheduling, vector Poisson channels.

1. INTRODUCTION

The importance of sensor scheduling emerges when it is not possible to collect unlimited amounts of data due to resource constraints. [1]. Generally speaking it is always advantageous to collect as much data as possible, provided such data can be properly managed and processed. However there are constraints imposed by available time, computational resources, memory requirements, communication bandwidth, available battery power, and sensor deployment patterns. This creates a need for an optimal or heuristic scheduling methodology to extract meaningful data from different available data sources in an efficient manner.

One of the major challenges in sensor scheduling problems involving switching among various sources arises from the combinatorial nature of the possible switching possibilities. This problem becomes further complicated if the sensor scheduling is done in continuous-time settings [2].

Information-theoretic approaches have long been used in sensor scheduling problems, especially for their invariance to invertible transformations of the data. Many of the early approaches towards sensor management were information-theoretic, for instance [3], [4]. Mutual information between observed data and unknown variables under test provides an intuitively appealing scalar metric for the performance of a given data acquisition methodology. That being said, it is also well-known that maximizing mutual information does not necessarily lead to optimal detection or estimation performance.

In this paper we address a very basic problem of *scheduling* a single sensor with multiple switch configurations to detect a binary 2-long random vector in a vector Poisson channel.
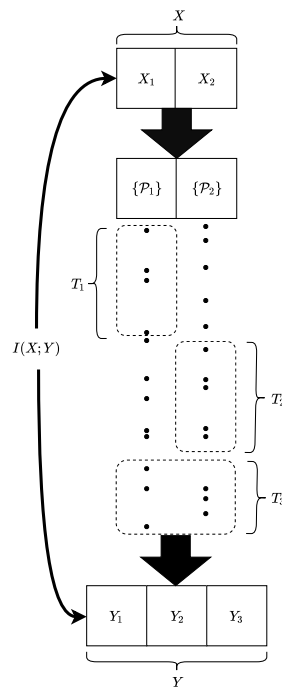
Fig. 1: Illustration of sensing method for Bayesian detection of $2-$long hidden random vector X from $3-$long observable random vector Y in vector Poisson channel under a total observation time constraint, where $\{\mathcal{P}_i\}$ is a conditional Poisson point process given Bernoulli distributed input X_i.

This might be of special interest in context of Poisson *compressive sensing*, where sparse signal conditions can be exploited for reduced computational cost and efficient sensor scheduling. We address this particular sensor scheduling problem from both *information theoretic* and *detection theoretic* perspectives. As will be seen, this problem is deceptively simple yet does not readily admit closed-form solutions.

The Poisson channel has been traditionally difficult to work with due to two inherent obstacles attached with it. First, added Poisson noise and scaling of input does not consolidate into a single parameter as SNR does in Gaussian channel. Second, scaling the support of a Poisson random variable (integers) does not result in another Poisson random variable. This is in contrast to Gaussian channels which are relatively well-studied from both information-theoretic and estimation-theoretic aspects [5], [6], [7]. The Poisson channel has the advantage in thought experiments such as ours of simple conceptual models for data acquisition based on counting (photons, say) and switches that either allow counts to either pass through or be discarded.

As will be seen when our problem is stated in more detail, fundamentally we are exploring a trade-off between SNR and ambiguity that arises when phenomena from multiple sources are combined in a single observation. Adding multiple Poisson processes results in a Poisson process at a higher rate; higher rates correspond to higher SNR, and generally higher SNR is seen as a good thing for detection and estimation performance. However, this SNR comes at a cost: with a single detector looking at counts from multiple processes, the observed Poisson counts cannot be disambiguated back to their original sources. In our proposed scheme, part of the time is spent looking at sources individually (to help with this disambiguation) and part of the time is spent observing the sources jointly (to increase SNR). Finding the right balance between these two approaches is the
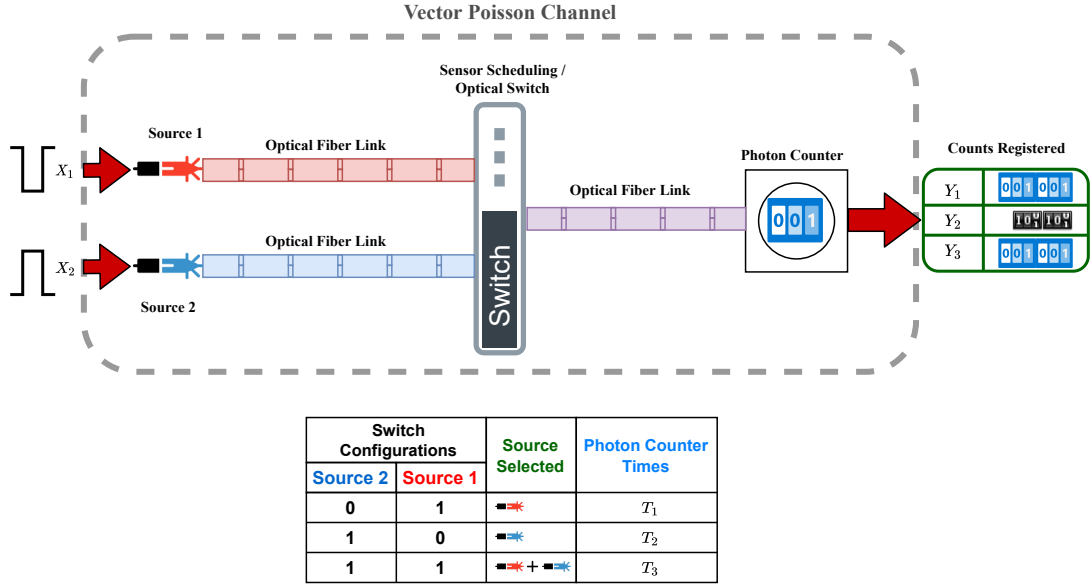
Switch Configurations		Source Selected	Photon Counter Times
Source 2	Source 1		
0	1		T_1
1	0		T_2
1	1		T_3

Fig. 2: An abstract example: Y_1, Y_2 and Y_3 are total counts of photons accumulated in times T_1, T_2 and T_3, respectively. LED source 1 (and source 2) initiates a homogeneous Poisson process of intensity λ_0 if Bernoulli distributed random input signal X_1 (and X_2) takes value of 0 with probability $(1-p)$; and λ_1 if input signal X_1 (and X_2) takes value of 1 with probability p. Once the input random vector $[X_0\ X_1]$ assumes any of the four possible states $[X_0\ X_1 : 00, 01, 10, 11]$, it doesn't change its state during the course of Photon counting for given time T.

heart of our problem.

In an unconstrained sensing-time setting, the vector Poisson channel has gained much research activity and interesting results are shown paralleling to that of vector Gaussian channel. One of the seminal works in linking the *information theory* and *estimation theory* in the context of scalar Poisson channel is done by Guo *et al.* in [8], where the derivative of mutual information with respect to signal intensity is equated with a function of conditional expectation; providing a ground for the possible Poisson counterpart of a Gaussian channel. This result for scalar Poisson channel was further refined by Atar and Weissman in [9] where exact relationship between derivative of mutual information and minimum mean loss error is given by providing the loss function $l(x, \hat{x}) = x \log(\frac{x}{\hat{x}}) - x + \hat{x}$ which shares a key property with squared error loss i-e optimality of conditional expectation with respect to mean loss criterion. This result together with $\frac{d}{d\alpha}I(X; \mathcal{P}(\alpha X))\big|_{\alpha=0} = E[X \cdot \log X] - E[X] \cdot \log(E[X])$ given in [8] provides us an exact answer to the question: for a given finite sensing time T and prior p, which of the two sensing mechanisms, individual or joint sensing, is better than the other? (however, hybrid sensing still remains elusive and this is we have investigated in this paper).

Wang *et al.* [10] unifies the vector Poisson and Gaussian channels by constructing a generalization of the classical Bregman divergence and extended the scalar result to vector case (unconstrained). They provide the gradient of mutual information with respect to their input scaling matrices for both Poisson and Gaussian channel. But, for a general vector Poisson channel existence of gradient of mutual information w.r.t scaling matrix is defined in terms of expected value of the Bregman divergence matrix with a strictly K-convex loss function and which requires the partial ordering interpretation [11] [10]
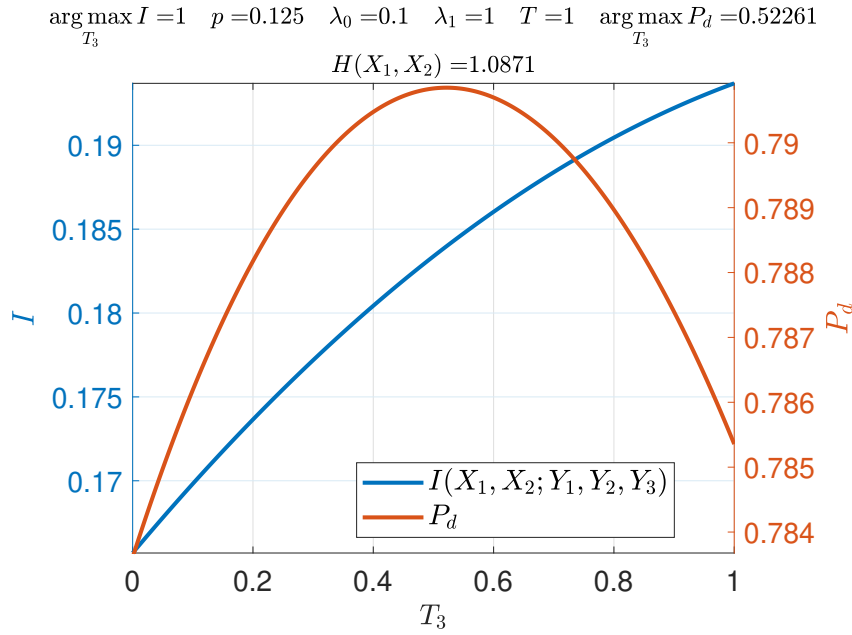
Fig. 3: Mutual information $I(X;Y)$ and Bayesian probability of total correct detections P_d vs. time T_3 where $0 \leq T_3 \leq T$ and $T_1 = T_2 = \frac{T-T_3}{2}$ such that time constraint $T = T_1 + T_2 + T_3$ is satisfied.

and which doesn't unify our problem. We are therefore also interested in exploring the general concave nature of our problem w.r.t scaling matrix (if it exists), which is discussed in coming sections.

The rest of the paper is organized as follows: Section 2 introduces the problem description, explaining the Poisson vector channel in which the target is to be observed. Section 3 provides the *information theoretic* model, while Section 4 describes the *detection theoretic* model of the problem. Section 5 discusses the computed results from the previous two sections. Finally, Section 6 concludes the paper. *Notation:* Upper case letters denote random vectors. Realizations of the random vectors are denoted by lower case letters. A number in subscript is used to show the component number of the random vector. We use X_1 and Y_1 to represent scalar input and output random variables, respectively. The superscript $(\cdot)^\intercal$ denotes the matrix/vector transpose. Deterministic time variable T_i is used to denote the sensing time for i^{th} conditional Poisson process and T is a given finite time. α is an arbitrary positive scalar variable. Φ represents the scaling matrix. p is the prior probability of 1 for a Bernoulli random variable. $f_X(x)$ denotes the probability mass function of X, and the subscript will normally be dropped to simplify notation. $\text{Poiss}(y; \lambda) \equiv \frac{e^{-\lambda}\lambda^y}{y!}$.

2. PROBLEM DESCRIPTION

Let X_1 and X_2 be two independent and identical distributed (i.i.d) Bernoulli random variables with p being the probability of occurrence of 1. We may define probability mass function f of discrete random vector $X \equiv [X_1, X_2]^\intercal$ as

$$f_X(x) = \begin{cases} p^2 & x = [1 \quad 1]^\intercal \\ (1-p)^2 & x = [0 \quad 0]^\intercal \\ p(1-p) & x = [0 \quad 1]^\intercal \quad \text{or} \quad x = [1 \quad 0]^\intercal \end{cases} \tag{1}$$

Let $\{\mathcal{P}_1(t), t \geq 0\}$ and $\{\mathcal{P}_2(t), t \geq 0\}$ be two *conditional point Poisson processes* [12, pp. 88-89] such that their rate parameters, either λ_0 or λ_1 are determined by X_1 and X_2,

respectively. That is, $\lambda(x_i) = \lambda_0$ if $x_i = 0$ and λ_1 if $x_i = 1$, for $i = 1, 2$. λ_0 and λ_1 are assumed known. We may count the number of arrivals from the two *conditional Poisson arrival processes* in three possible configurations: two by individually counting the arrivals from the two processes $\mathcal{P}_1$, $\mathcal{P}_2$ and one by counting the arrivals from the sum of two processes $\{\mathcal{P}_1(t) + \mathcal{P}_2(t), t \geq 0\}$ with given rate parameter $\lambda(X_1) + \lambda(X_2)$ as illustrated in Figs.(1) and (2). The counting is performed such that at any given time, only one of the three possible configurations is active. Furthermore, because of the *independent increments* property of *conditional Poisson processes*, it is not necessary to switch back and forth among possible configurations; it is sufficient to be in configuration 1 for time T_1, followed by configuration 2 for time T_2, and then configuration 3 for time T_3. Additionally, counting is performed with a finite time constraint $T = T_1 + T_2 + T_3$ where T_1, T_2 and T_3 are the unknown time proportions in counting arrivals from processes $\{\mathcal{P}_1(t), t \geq 0\}$, $\{\mathcal{P}_2(t), t \geq 0\}$ and $\{\mathcal{P}_1(t) + \mathcal{P}_2(t), t \geq 0\}$, respectively and T is total time. After utilizing available time T, the above counting paradigm leads to a multivariate Poisson mixture model with four component in three dimensions; we write random vector $Y^\intercal \equiv [Y_1\ Y_2\ Y_3]$, so that $Y \in \{0, 1, 2, 3 \cdots \}^3$. Each Y_i is Poisson random variable given X; such that their conditional law is,

$$Y_1|X_1 \sim \text{Poiss}\left(y_1; \left(\lambda_0 \cdot (1 - X_1) + \lambda_1 \cdot X_1\right) \cdot T_1\right)$$

$$Y_2|X_2 \sim \text{Poiss}\left(y_2; \left(\lambda_0 \cdot (1 - X_2) + \lambda_1 \cdot X_2\right) \cdot T_2\right)$$

$$Y_3|(X_1 + X_2) \sim \text{Poiss}\left(y_3; \left((2 - (X_1 + X_2)) \cdot \lambda_0 + \right.\right.$$

$$\left.\left. (X_1 + X_2) \cdot \lambda_1\right) \cdot T_3\right). \tag{2}$$

The observed random vector Y carries *information* about hidden random vector $X^\intercal \equiv [X_1\ X_2]$. We are interested in strategies for selecting the times T_1, T_2, T_3 that optimize some cost related to the inference problem, either mutual information or probability of correct detection. Strategies that involve setting T_3 to 0 are referred to as *individual sensing*; any time spent looking at combined counts is termed *joint sensing*. The general case, which involves some individual and some joint sensing, will be referred to as *hybrid sensing*.

In matrix form we may relate the intensities of conditional Poisson distributed Y_i's as,

$$
\overbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}}^{\mathcal{M}:\,3\times 1} = \overbrace{\begin{bmatrix} T_1 & 0 & 0 \\ 0 & T_2 & 0 \\ 0 & 0 & T_3 \end{bmatrix}}^{\mathcal{D}:\,3\times 3} \overbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}}^{\mathcal{B}:\,3\times 2} \overbrace{\begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix}}^{\Lambda:\,2\times 1}
$$

$$
= \overbrace{\begin{bmatrix} T_1 & 0 \\ 0 & T_2 \\ T_3 & T_3 \end{bmatrix}}^{\Phi:\,3\times 2} \overbrace{\begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix}}^{\Lambda:\,2\times 1} \tag{3}
$$

$$
= \begin{bmatrix} \left(\lambda_0 \cdot (1 - X_1) + \lambda_1 \cdot X_1\right) \cdot T_1 \\ \left(\lambda_0 \cdot (1 - X_2) + \lambda_1 \cdot X_2\right) \cdot T_2 \\ \left((2 - (X_1 + X_2))\lambda_0 + (X_1 + X_2)\lambda_1\right)T_3 \end{bmatrix}
$$

where $\Lambda_1 = \lambda_0 \cdot (1 - X_1) + \lambda_1 \cdot X_1$ and $\Lambda_2 = \lambda_0 \cdot (1 - X_2) + \lambda_1 \cdot X_2$ are two scalar functions of random variables X_1 and X_2, respectively. From the optimization point of view and in terms of *sensor scheduling*, we are interested in finding the optimal time-allocation, (T_1, T_2, T_3), of total available time resource, T, that would maximize the *reward* i.e. either the mutual information or probability of total correct detections. We may say that we are
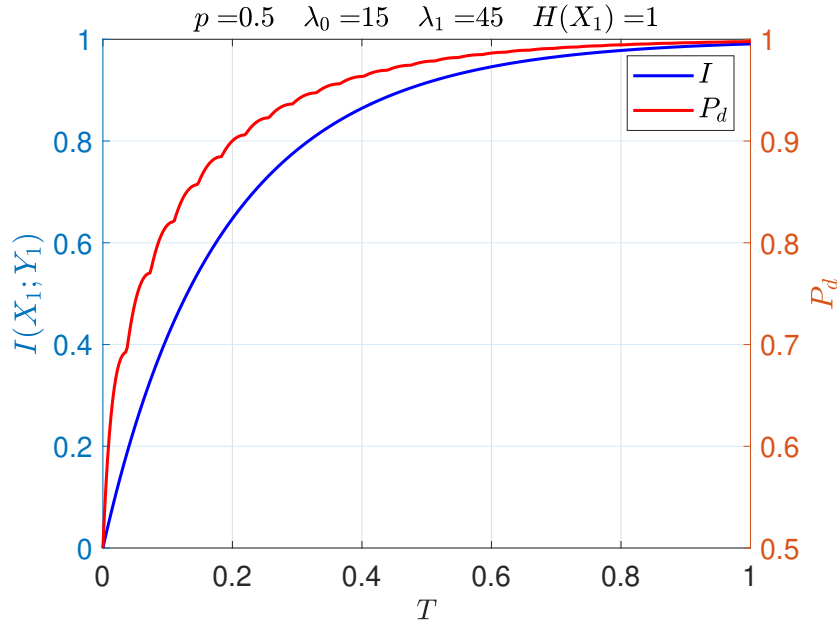
Fig. 4: Scalar Poisson channel: mutual information $I(X_1; Y_1)$ and probability of total correct detections P_d vs. time T.

interested in finding the diagonal matrix $\mathcal{D}$, such that the mutual information $I(X; Y)$ is maximized under time constraint $T = T_1 + T_2 + T_3$. Configuration matrix $\mathcal{B}$ represents all possible sensing combinations and Λ is the Poisson rate parameter vector. Vector $\mathcal{M}$ is the vector of intensities of the Poisson random variables Y_1, Y_2, and Y_3.

Mathematically we may write

$$\max_{T_1, T_2, T_3} I(X_1, X_2; Y_1, Y_2, Y_3) \text{ s.t. } T_1 + T_2 + T_3 = T. \tag{4}$$

We may rewrite the objective function

$$\max_{T_1, T_2, T_3} I(X_1, X_2; Y_1, Y_2, Y_3) \text{ s.t. } T_1 + T + T_3 = 1 \tag{5}$$

where $0 \leq T_1 \leq 1$, $0 \leq T_2 \leq 1$ and $0 \leq T_3 \leq 1$. Without loss of generality we can take $T = 1$ since the means of the observed Y_i variables are the product of rate and time, and any change in the total available time can be reflected in the rates, or equivalently changing the time units.

We extend our understanding by looking at the same problem from the *detection theoretic* aspect. For this, we maximize the Bayesian probability of total correct detection, P_d, of hidden random vector X from observable random vector Y, as

$$\max_{T_1, T_2, T_3} P_d \quad \text{s.t.} \quad T_1 + T_2 + T_3 = 1 \tag{6}$$

and compare the results to formerly computed *information theoretic* results. Simulations on empirical data are performed by varying different parameters involved in the model for further validation of computed results.

3. INFORMATION THEORETIC DESCRIPTION

A. Scalar Poisson channel

Firstly the scalar version of the Poisson channel is presented and then it is extended to the vector version of our problem. We start with mutual information between a scalar

Bernoulli random variable X_1 and Y_1 which is a scalar and two component Poisson mixture. The probability mass function of Y_1 is then given as

$$f(y_1) = (1 - p) \cdot \text{Poiss}(y_1; T\lambda_0) + p \cdot \text{Poiss}(y_1; T\lambda_1).$$

The mutual information I can be written as

$$I(X_1; Y_1) = H(Y_1) - H(Y_1|X_1)$$

where $H(\cdot)$ is the Shannon entropy, which is defined as a discrete functional $H : f \to -\sum_{Y \in \mathcal{Y}} f \cdot \log_2(f)$ and f is the probability mass function of random variate Y with $\mathcal{Y}$ as the corresponding support. We may write $H(Y_1)$ as

$$\begin{aligned}
H(Y_1) = -\sum_{y_1=0}^{\infty} \Big(&[(1-p) \cdot \text{Poiss}(y_1; T\lambda_0) + p \\
&\cdot \text{Poiss}(y_1; T\lambda_1)] \cdot \log_2[(1-p) \cdot \text{Poiss}(y_1; T\lambda_0) \\
&+ p \cdot \text{Poiss}(y_1; T\lambda_1)] \Big),
\end{aligned}$$

and

$$\begin{aligned}
H(Y_1|X_1) &= \sum_{x_1 \in \mathcal{X}_1} f(x_1) \sum_{y_1=0}^{\infty} -f(y_1|x_1) \cdot \log_2[f(y_1|x_1)] \\
&= -\Big((1-p) \sum_{y_1=0}^{\infty} \text{Poiss}(y_1; T\lambda_0) \\
&\qquad \cdot \log_2[\text{Poiss}(y_1; T\lambda_0)] + p \sum_{y_1=0}^{\infty} \text{Poiss}(y_1; T\lambda_1) \\
&\qquad \cdot \log_2[\text{Poiss}(y_1; T\lambda_1)] \Big).
\end{aligned}$$

In Fig. (4), for a scalar Poisson channel discussed above, $I(X_1; Y_1)$ and P_d illustrates a monotonic relationship w.r.t T, as expected.

B. Vector Poisson channel

Mutual information between two random vectors can be defined as the difference between the total entropy in one random vector and the conditional entropy in the second random vector given the first vector. We write

$$I(X; Y) = H(Y) - H(Y|X) \tag{7}$$

The conditional entropy $H(Y|X)$ is calculated from the conditional probability mass functions $f(y|X = [x_1 \quad x_2]^\intercal)$ defined as

$$\begin{aligned}
&f(y|X = [0 \quad 0]^\intercal) \\
&= \text{Poiss}(y_1; \lambda_0 T_1) \cdot \text{Poiss}(y_2; \lambda_0 T_2) \cdot \text{Poiss}(y_3; 2\lambda_0 T_3), \\
&f(y|X = [0 \quad 1]^\intercal) \\
&= \text{Poiss}(y_1; \lambda_0 T_1) \cdot \text{Poiss}(y_2; \lambda_1 T_2) \cdot \text{Poiss}(y_3; (\lambda_0 + \lambda_1)T_3), \\
&f(y|X = [1 \quad 0]^\intercal) \\
&= \text{Poiss}(y_1; \lambda_1 T_1) \cdot \text{Poiss}(y_2; \lambda_0 T_2) \cdot \text{Poiss}(y_3; (\lambda_1 + \lambda_0)T_3), \\
&f(y|X = [1 \quad 1]^\intercal) \\
&= \text{Poiss}(y_1; \lambda_1 T_1) \cdot \text{Poiss}(y_2; \lambda_1 T_2) \cdot \text{Poiss}(y_3; 2\lambda_1 T_3).
\end{aligned}$$

The marginal probability mass function of Y is then given as

$$
\begin{aligned}
&f(y) \\
&= (1-p)^2 \cdot f(y|X = [0 \quad 0]^\intercal) + p(1-p)\cdot \\
&f(y|X = [0 \quad 1]^\intercal) + p(1-p) \cdot f(y|X = [1 \quad 0]^\intercal) + \\
&p^2 \cdot f(y|X = [1 \quad 1]^\intercal).
\end{aligned} \tag{8}
$$

Using the identity defined in (7) and the definition of entropy defined above, the mutual information $I(X;Y)$ becomes

$$
\begin{aligned}
&I(X;Y) \\
&= \sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} \sum_{y_3=0}^{\infty} -f(y_1,y_2,y_3) \cdot \log_2[f(y_1,y_2,y_3)]) \\
&\quad - \sum_{X \in \mathcal{X}} f(x) \Big[\sum_{y_1=0}^{\infty} \sum_{y_2=0}^{\infty} \sum_{y_3=0}^{\infty} -f(y_1,y_2,y_3|X) \\
&\quad \cdot \log_2[f(y_1,y_2,y_3|X)] \Big].
\end{aligned} \tag{9}
$$

A complete expression for mutual information is given in Appendix A.

Fig. (3) illustrates the concavity of mutual information as a function of T_3 (constrained problem), in a vector Poisson channel defined in (2) and illustrated in Fig. (1), when three sensing times are varied according to $(T_1, T_2, T_3) = (\frac{1-T_3}{2}, \frac{1-T_3}{2}, T_3)$ and $0 \leq T_3 \leq 1$. It can also be seen that the two metrics P_d and I are maximizing at two different input argument values.

Theorem 1: $I(X_1, X_2; Y_1, Y_2, Y_3)$ is concave in $T_3 = 0$ plane.

Proof:

$$I(X_1, X_2; Y_1, Y_2, Y_3)\Big|_{T_3=0} = I(X_1, X_2; Y_1, Y_2)$$

By chain rule of mutual information:

$$
\begin{aligned}
I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1) \\
&= I(X_1; Y_1) + I(X_2; Y_2).
\end{aligned} \tag{10}
$$

We note in (10) that each $I(X_i; Y_i)$ is solely a function of T_i and also concave in it [9, p. 1306][13]. We further know that sum of concave functions is a concave function. Therefore $I(X_1, X_2; Y_1, Y_2)$ is concave in T_1 and T_2 when $T_3 = 0$. This concludes the proof. ∎

Theorem 2: $I(X_1, X_2; Y_1, Y_2, Y_3)$ is symmetric in variables T_1 and T_2.

Proof: Mutual information $I(X_1, X_2; Y_1, Y_2, Y_3)$ given in appendix A is invariant under any permutation of variables T_1 and T_2. That means interchanging the two variables leaves the expression unchanged. ∎

If we further expand the expression for mutual information between X and Y using chain rule [14] [15] as,

$$
\begin{aligned}
&I(X_1, X_2; Y_1, Y_2, Y_3) \\
&= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1) + I(X_1, X_2; Y_3|Y_1, Y_2) \\
&= I(X_1; Y_1) + \overbrace{I(X_2; Y_1|X_1)}^{0} + \overbrace{I(X_1; Y_2|Y_1)}^{0} + \\
&\quad \overbrace{I(X_2; Y_2|Y_1, X_1)}^{I(X_2;Y_2)} + I(X_1, X_2; Y_3|Y_1, Y_2) \\
&= I(X_1; Y_1) + I(X_2; Y_2) + I(X_1, X_2; Y_3|Y_1, Y_2).
\end{aligned} \tag{11}
$$

$$I(X_1 X_2; Y_1 Y_2 Y_3) = I(X_1; Y_1) + I(X_2; Y_2) + I(X_1 X_2; Y_3 | Y_1 Y_2).$$

$$\arg\max I = 0.41414 \quad p = 0.1 \quad \lambda_0 = 5 \quad \lambda_1 = 20 \quad T = 1$$
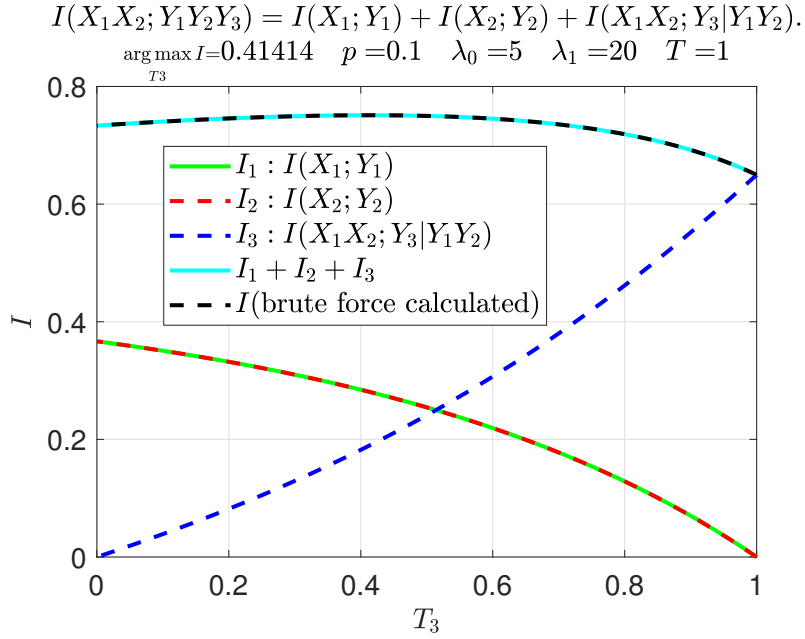

Fig. 5: Concavity and convexity of three terms in $I(X_1 X_2; Y_1 Y_2 Y_3) = I(X_1; Y_1) + I(X_2; Y_2) + I(X_1 X_2; Y_3 | Y_1 Y_2)$ when $(T_1, T_2, T_3) := (\frac{1-T_3}{2}, \frac{1-T_3}{2}, T_3)$ such that $0 \leq T_3 \leq 1$.

Note that in (11) that the first and second terms are indeed concave in (T_1, T_2, T_3). The third term, when computed, exhibits non-concavity in the respective domain. Fig. (5) illustrates this when mutual information is plotted along the line $(T_1, T_2, T_3) := (\frac{T-T_3}{2}, \frac{T-T_3}{2}, T_3)$ parametrized by $0 \leq T_3 \leq T$. It is interesting to note that the sum of three terms exhibits concavity, irrespective of the fact that the third term is non-concave. However, analytical investigation of the functional properties of this third term, $I(X_1 X_2; Y_3 | Y_1 Y_2)$, is not done in this work.

We computed mutual information between hidden random vector X and observable random vector Y for diverse set of intensities λ_0 and λ_1 along with different priors and total available time, $T = 1$, to observe the concavity of $I(X_1, X_2; Y_1, Y_2, Y_3)$ in T_1, T_2 and T_3. While we do not have a proof that the objective function is concave in a linear time constraint for the I metric, we have carried out extensive computational experiments and we have yet to observe a single case in which non-concavity is observed. Based on our observation we consistently noted the concave property of mutual information which led us to propose the following conjecture.

Conjecture 1: If $X \equiv [X_1 \ X_2]^\intercal$ be a non-negative random vector where X_1 and X_2 are mutually independent and identical distributed. Let random vector $Y \equiv [Y_1 \ Y_2 \ Y_3]^\intercal \in \mathbb{Z}_+^3$, jointly distributed with X such that conditional law is given as in (2) then $I(X_1, X_2; Y_1, Y_2, Y_3)$ is concave in (T_1, T_2, T_3) under time constraint $T = T_1 + T_2 + T_3$.

Corollary 1 (Feasible region for optimal solution): The two properties of mutual information *concavity* (if true) and *symmetry* (proved), guarantee that *maxima* must occur at the line of symmetry $(T_1, T_2, T_3) := (\frac{T-\alpha}{2}, \frac{T-\alpha}{2}, \alpha)$ where $0 \leq \alpha \leq T$.

One of the immediate consequences of exploiting *symmetry* and *concavity* of the problem is that it would reduce the search space for the optimal solution from two dimensions to one dimension.

Claim : Mutual information $I(X_1, X_2; Y_1, Y_2, Y_3)$ in general is not symmetric in p around $p = 0.5$ for fixed time proportions (T_1, T_2, T_3) for a conditionally multivariate Poisson (Y_1, Y_2, Y_3) given (X_1, X_2).

Example: For a scalar random variable X_1 and conditional Poisson variable Y_1 with scaling factor T, we have equation (115) on page (10) of [8], given below

$$\frac{d}{dT}I(X_1;Y_1)\Big|_{T=0} = E[\Lambda_1 \cdot \log \Lambda_1] - E[\Lambda_1] \cdot \log(E[\Lambda_1]).$$
(12)

where $\Lambda_1 = \lambda_0 \cdot (1 - X_1) + \lambda_1 \cdot X_1$. When we consider a random vector X (instead of scalar random variable as above), and take distribution on X as given in (1). Dividing variable total time T equally in counting arrivals from X_1 and X_2 separately: $T_1 = T_2 = \frac{T}{2}$ and $T_3 = 0$, we have

$$I(X_1, X_2; Y_1, Y_2, Y_3)\Big|_{\left(T_1=\frac{T}{2}, T_2=\frac{T}{2}, T_3=0\right)} = $$
$$2 \cdot I(X_1; Y_1)\Big|_{\left(T_1=\frac{T}{2}\right)}$$
(13)

Let T be a free variable in (13), we have

$$\frac{d}{dT}\left(I(X_1, X_2; Y_1, Y_2, Y_3)\Big|_{\left(T_1=\frac{T}{2}, T_2=\frac{T}{2}, T_3=0\right)}\right)\Bigg|_{T=0} = $$
$$(1-p) \cdot \lambda_0 \cdot \log_2(\lambda_0) + p \cdot \lambda_1 \cdot \log_2(\lambda_1) -$$
$$\left((1-p) \cdot \lambda_0 + p \cdot \lambda_1\right) \cdot \log_2\left((1-p) \cdot \lambda_0 + p \cdot \lambda_1\right)$$
(14)

Equation (14) is concave in p

$$\frac{d^2}{dp^2}\left(\frac{d}{dT}\left(I(X_1, X_2; Y_1, Y_2, Y_3)\Big|_{\left(T_1=\frac{T}{2}, T_2=\frac{T}{2}, T_3=0\right)}\right)\Bigg|_{T=0}\right)$$
$$= -\frac{(\lambda_0 - \lambda_1)^2}{\mathrm{Ln}(2) \cdot \left((1-p)\lambda_0 + p \cdot \lambda_1\right)} < 0$$
(15)

Maxima of equation (14) occurs at $p = \frac{4-e}{e}$ when $\lambda_0 = 2$ and $\lambda_1 = 4$. Since (14) is concave in p, then if it was symmetric, maxima must had occurred at $p = 0.5$. ∎

4. DETECTION THEORETIC DESCRIPTION

A. *Maximization of probability of total correct detections in sensing time intervals*

In previous section, we discussed and presented the metric of mutual information I between hidden random vector X, and observable vector Y. Here we approach the original sensing problem as a multi-hypothesis detection problem and find the optimal solution by minimizing the Bayesian risk [16, pp.220] in sensing times i-e (T_1, T_2, T_3). We define Bayes risk r as

$$r = (1-p)^2 \Big[P_{00|00} C_{00|00} + P_{01|00} C_{01|00}$$
$$+ P_{10|00} C_{10|00} + P_{11|00} C_{11|00} \Big] + p(1-p)$$
$$\Big[P_{00|01} C_{00|01} + P_{01|01} C_{01|01} + P_{10|01} C_{10|01}$$
$$+ P_{11|01} C_{11|01} \Big] + p(1-p) \Big[P_{00|10} C_{00|10}$$
$$+ P_{01|10} C_{01|10} + P_{10|10} C_{10|10} + P_{11|10} C_{11|10} \Big]$$
$$+ p^2 \Big[P_{00|11} C_{00|11} + P_{01|11} C_{01|11} + P_{10|11} C_{10|11}$$
$$+ P_{11|11} C_{11|11} \Big],$$

where $P_{kl \mid ij}$ is the probability that $X = [x_1^i \; x_2^j]^\intercal$ is true while decision $X = [x_1^k \; x_2^l]^\intercal$ is made; similarly for $C_{kl \mid ij}$. Setting all costs for which $[x_1^i \; x_2^j]^\intercal \neq [x_1^k \; x_2^l]^\intercal$ to one and $[x_1^i \; x_2^j]^\intercal = [x_1^k \; x_2^l]^\intercal$ to zero, we have

$$r = (1-p)^2 \Big[P_{01 \mid 00} + P_{10 \mid 00} + P_{11 \mid 00} \Big] + p(1-p)$$
$$\Big[P_{00 \mid 01} + P_{10 \mid 01} + P_{11 \mid 01} \Big] + p(1-p)$$
$$\Big[P_{00 \mid 10} + P_{01 \mid 10} + P_{11 \mid 10} \Big] + p^2$$
$$\Big[P_{00 \mid 11} + P_{01 \mid 11} + P_{10 \mid 11} \Big]. \tag{16}$$

We are interested in minimizing this Bayes risk r in (T_1, T_2, T_3) i-e

$$\min_{T_1,T_2,T_3} r \quad \text{s.t. } T_1 + T_2 + T_3 = 1. \tag{17}$$

Note that while minimizing r in (T_1, T_2, T_3), the decisions boundaries would be changing accordingly and become function of (T_1, T_2, T_3). Equivalently, we may say that

$$\max_{T_1,T_2,T_3} P_d \quad \text{s.t. } T_1 + T_2 + T_3 = 1 \tag{18}$$

where P_d is probability of total correct detections, $P_d = 1 - r$. In the next section we compute I ans P_d, for different channel parameters, given in Appendices A and B, respectively.

5. COMPUTATIONAL AND SIMULATION RESULTS

The primary purpose of performing the computations and simulations is twofold: the first is to see under what circumstances which of the sensing method: individual, joint or hybrid sensing is optimal since analytical solutions for the two objective functions do not exist; and the second is to investigate whether the two metrics I and P_d leads to the same or different maximizing input arguments. The objective function, defined in (4) is observed to be a convex optimization problem when a pre-defined channel scaling matrix structure is imposed. For all computational purposes, we always approximate the Poisson mass function by truncating it; truncation is done by a rectangular window that extends from lower limit 0 to the upper limit where Poisson pmf drops in value below double precision machine epsilon ($\epsilon = 2^{-53}$) as given in Appendix C.

Fig. (6) illustrates how the optimal input argument T_3 moves from 0 to 1 as prior $0 < p < 1$ is varied while input intensities are held fixed. Ninety linearly spaced discrete points in the range $(0, 0.5]$ are first taken for prior p and then another ten linearly spaced points are taken in the interval $(0.5, 1)$, constituting a total of hundred points of p for computations. For every value of prior p; another hundred linearly spaced points for time proportion T_3 are taken in the closed interval $[0, 1]$ such that $T_1 = T_2 = \frac{1-T_3}{2}$ and $T_1 + T_2 + T_3 = 1$. At every instant of p, hundred values of I are computed corresponding to hundred time proportion. Maximum value of I is then sought and the corresponding value of T_3 is plotted against prior p. The plot signifies that if the prior p is closer to 0 then all time T be invested in joint sensing whereas as prior increases from 0 towards 1, individual sensing progressively starts to take more and more proportion of total available time T. This phenomena is further explored in scatter plots where input intensities are varied too.

In Fig. (11), ternary plots of I are shown as prior p is varied. The vertices of the equilateral triangle are at $(T_1, 0, 0)$, $(0, T_2, 0)$ and $(0, 0, T_3)$ with $T_1 + T_2 + T_3 = 1$. It can be seen that the distribution of time resource $T = 1$ changes as p is varied. As p tends to 0, the optimal value of T_3 tends to consume all the available time resource T
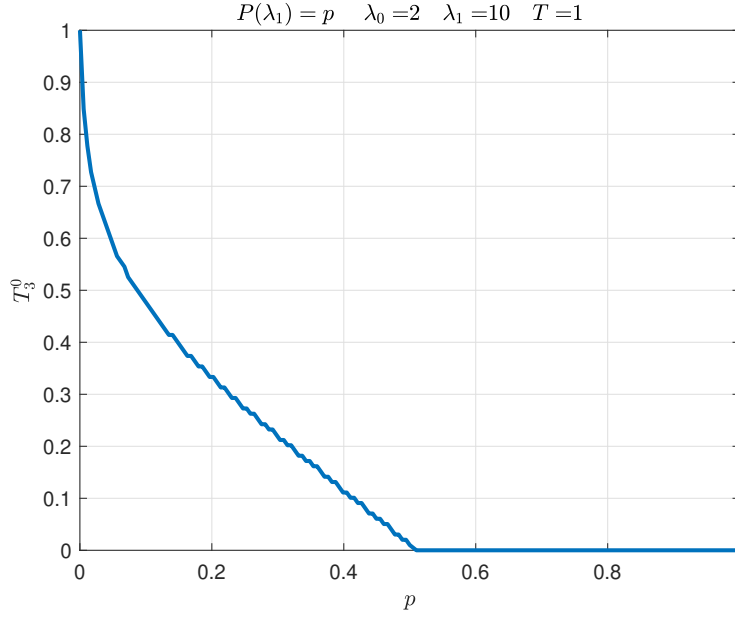
Fig. 6: Optimal time T_3^0 vs. prior probability p with $T_1 = T_2$ and $T_1 + T_2 + T_3 = T$, considering I as a metric.

leaving optimal values of T_1 and T_2 approaching to 0. Mirror symmetry in arguments T_1 and T_2 can be observed. It can be seen further that the optimal argument always lies on the line that bisects the equilateral triangle into two right angle triangles i.e. $(T_1, T_2, T_3) := (\frac{1-T_3}{2}, \frac{1-T_3}{2}, T_3)$ where $0 \leq T_3 \leq 1$.

A relationship between mutual information and MAP (maximum a posteriori) detection for our problem is investigated in Fig. (12), in which Bayesian probability of total correct detections, P_d, given in Appendix B is plotted against respective mutual information I given in Appendix A. The I and P_d curves are not optimizing on the same input arguments. We further counter-checked our MAP detector performance with empirical means as shown in Fig. (13) by generating 10^5 samples from a multivariate Poisson mixture random vector Y, at each time instant (T_1, T_2, T_3) for fixed given parameters: p, λ_0, λ_1. Total of 200 linearly spaced time instants of T_3 are taken in the range $[0, 1]$ and assuming $T_1 = T_2 = \frac{1-T_3}{2}$. The samples are taken in the same proportion (of 10^5) from each of the four components of mixture Y as their respective prior probabilities suggest. Each of the samples is then passed through a MAP detector to detect which of the four mixture component is the source of that sample. These 10^5 detections are then compared with their true sources to compute C_d at each of time instant T_3. We found that this empirical total correct detections C_d is consistent with analytic P_d. The computing method for I is described step-by-step in the flowcharts given in Fig.(9) and Fig.(10). The bounds on the entropy of Poisson variable are available in [17], we use them to validate that the computed Poisson entropies are within bounds.

For an unconstrained problem of choosing individual sensing over jointing sensing for a given prior and fixed intensities for an infinitesimal or finite T; the first derivative of I at $T = 0$ given in [8] (along with knowing the fact that I is concave in T [9]) might be helpful. It is also helpful in verifying some of the numerical results of I with analytical ones. In Fig.(7) we have $\frac{d}{dT} I(X_1, X_2; Y_3)\big|_{T=0}$ and $\frac{d}{dT} I(X_1, X_2; Y_1, Y_2)\big|_{T=0}$ vs. p for $\lambda_0 = 2$ and $\lambda_1 = 100$. We can see that at $p = 0.5$ the first derivatives of both the
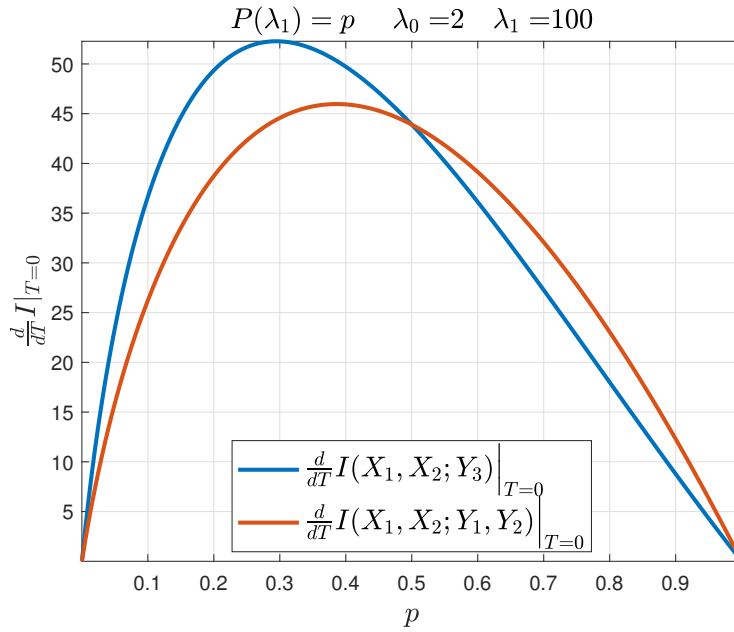
Fig. 7: Individual 2-target detection problem: $\frac{d}{dT}I(X_1, X_2; Y_3)\Big|_{T=0}$ and $\frac{d}{dT}I(X_1, X_2; Y_1, Y_2)\Big|_{T=0}$ with $(T_1 = \frac{T}{2}, T_2 = \frac{T}{2})$ vs. p.
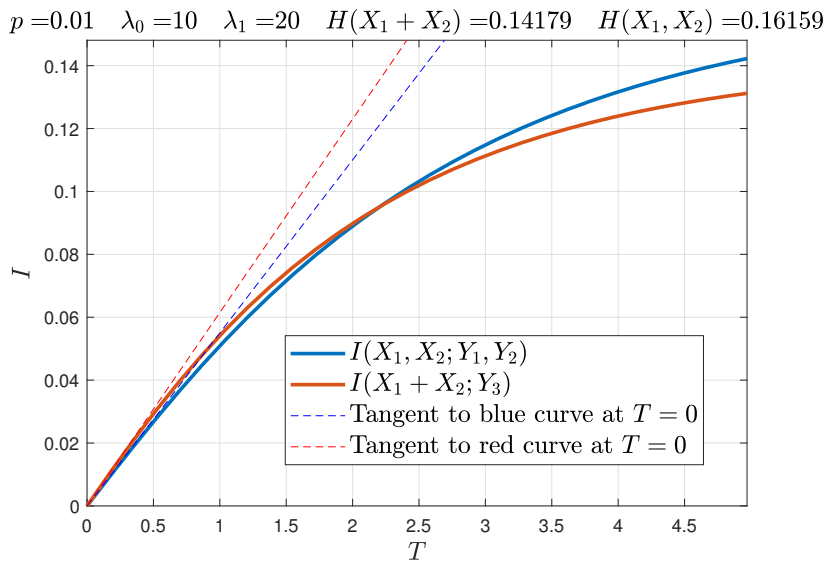

Fig. 8: Joint sensing and individual sensing vs. T. Note that joint sensing is better sensing method when $T < 2.25$. For $T > 2.25$ individual sensing is better.

joint and individual sensing w.r.t time variable T are same; this means that no matter what the given time is, individual sensing is never worse than the joint sensing at $p = 0.5$. Whereas, for $p < 0.5$ we need to further know what the given finite time T is, to decide which of the two methods is better over the other for that given time T. As can be seen in Fig.(8) that the two I curves cross each other, making one better over the other in two different time segments. Further note that in Fig. (8) the numerical derivative values of I at $T = 0$ are 0.055011541474291 and 0.061490803759598 for individual and

joint sensing, respectively when I is computed according to the algorithm defined in flowcharts. Now compare these numerical derivative values with analytically calculated ones 0.055011540931595 and 0.061490807291534; the two are accurate to eight decimal places, this validates that computations are accurate enough.

Fig. (14) and Fig. (15) illustrates the scatter plots, for a range of varying input intensities in constrained problem w.r.t I metric. The scatter plots on the left hand side illustrates the optimal value of mutual information at any point $(\lambda_0 \cdot T, \lambda_1 \cdot T)$ where $\lambda_0 \cdot T < \lambda_1 \cdot T$ and $0 < \lambda_1 \cdot T \leq 5$ for a given prior p and $T = 1$. The corresponding maximizing argument is shown on the left hand side of the scatter plots. It is noticed that for prior $p < 0.5$, farther the four components of multivariate mixture Poisson are from each other in terms of their intensities; more the optimal solution moves towards the individual sensing. In other words if the four pmfs (under hypothesis testing) gets closer, more the optimal sensing relies on joint sensing for prior $p < 0.5$. If the prior $p \geq 0.5$ then irrespective of the other parameter values in the model, individual sensing is the optimal strategy from I metric.

6. Conclusion

In this work we have addressed a simple abstract sensor scheduling problem in which two Poisson sources are observed either individually or jointly with a switching mechanism and a single counter. The scheduling problem is to choose the times associated with the different switch configurations given a total time constraint.

For our specific problem we were able to solve the optimization problem using computational methods, primarily because of the relatively small number of free parameters. For the case in which mutual information is the cost function, we observed but were unable to prove that the cost function is concave in the switch times; if true this reduces the optimization problem to a one-dimensional search. We observed that the two different metrics of mutual information and probability of correct detection did not always lead to the same scheduling solution. A second interesting result was that, under the mutual information metric, when the prior p on the sources was greater than 0.5, individual sensing was always the best that could be done.

In a certain sense our computational results are not particularly interesting. In most cases the cost functions are quite flat and there were no dramatic variation in mutual information or probability of correct detection across the space of possible switch configuration times. A communications engineer faced with the problem exactly as we have posed it would be hard-pressed to justify doing anything other than the simplest and most obvious solution, namely taking the available time and dividing it equally between individual observation of the two sources.

However, the fact that there is *some* variation in performance based in intelligent choice of the switch times leaves open the possibility of the value of investigations like ours for more complex problems. Several extensions immediately come to mind: more sources (perhaps hundreds or thousands, not two), unknown Poisson rates, different unknown or nonexistent priors, different source distributions. Unfortunately, the jump to a larger number of sources brings with it a combinatorial explosion which would rule out the brute-force optimization techniques used in this paper in very short order. Alternate switching strategies might include an adaptive one in which the switching time and configurations are determined online or in real time depending on the observations up to the current time. We recommend and are interested in continuing investigations along these lines.

APPENDIX A
EXPRESSION OF MUTUAL INFORMATION $I(X_1, X_2; Y_1, Y_2, Y_3)$

$$I(X;Y) = -\sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}\sum_{y_3=0}^{\infty}\left[\left((1-p)^2 \cdot \text{Poiss}(y_1; \lambda_0 T_1) \cdot \text{Poiss}(y_2; \lambda_0 T_2)\cdot\right.\right.$$

$$\text{Poiss}(y_3; 2\lambda_0 T_3) + p(1-p) \cdot \text{Poiss}(y_1; \lambda_0 T_1) \cdot \text{Poiss}(y_2; \lambda_1 T_2) \cdot \text{Poiss}(y_3; (\lambda_0 + \lambda_1)T_3)+$$

$$p(1-p) \cdot \text{Poiss}(y_1; \lambda_1 T_1) \cdot \text{Poiss}(y_2; \lambda_0 T_2) \cdot \text{Poiss}(y_3; (\lambda_1 + \lambda_0)T_3) + p^2 \cdot \text{Poiss}(y_1; \lambda_1 T_1)\cdot$$

$$\left.\text{Poiss}(y_2; \lambda_1 T_2) \cdot \text{Poiss}(y_3; 2\lambda_1 T_3)\right) \cdot \left(\log_2[(1-p)^2 \cdot \text{Poiss}(y_1; \lambda_0 T_1) \cdot \text{Poiss}(y_2; \lambda_0 T_2)\cdot\right.$$

$$\text{Poiss}(y_3; 2\lambda_0 T_3) + p(1-p) \cdot \text{Poiss}(y_1; \lambda_0 T_1) \cdot \text{Poiss}(y_2; \lambda_1 T_2) \cdot \text{Poiss}(y_3; (\lambda_0 + \lambda_1)T_3)+$$

$$p(1-p) \cdot \text{Poiss}(y_1; \lambda_1 T_1) \cdot \text{Poiss}(y_2; \lambda_0 T_2) \cdot \text{Poiss}(y_3; (\lambda_1 + \lambda_0)T_3)+$$

$$\left.\left.p^2 \cdot \text{Poiss}(y_1; \lambda_1 T_1) \cdot \text{Poiss}(y_2; \lambda_1 T_2) \cdot \text{Poiss}(y_3; 2\lambda_1 T_3)]\right)\right] +$$

$$\left[\left((1-p)^2 \cdot \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}\sum_{y_3=0}^{\infty}(\text{Poiss}(y_1; \lambda_0 T_1) \cdot \text{Poiss}(y_2; \lambda_0 T_2) \cdot \text{Poiss}(y_3; 2\lambda_0 T_3))\cdot\right.\right.$$

$$\left.\log_2[\text{Poiss}(y_1; \lambda_0 T_1) \cdot \text{Poiss}(y_2; \lambda_0 T_2) \cdot \text{Poiss}(y_3; 2\lambda_0 T_3)]\right)+$$

$$\left(p(1-p) \cdot \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}\sum_{y_3=0}^{\infty}(\text{Poiss}(y_1; \lambda_0 T_1) \cdot \text{Poiss}(y_2; \lambda_1 T_2) \cdot \text{Poiss}(y_3; (\lambda_0 + \lambda_1)T_3)\cdot\right.$$

$$\left.\log_2[\text{Poiss}(y_1; \lambda_0 T_1) \cdot \text{Poiss}(y_2; \lambda_1 T_2) \cdot \text{Poiss}(y_3; (\lambda_0 + \lambda_1)T_3)]\right) + \left(p(1-p)\cdot\right.$$

$$\sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}\sum_{y_3=0}^{\infty}(\text{Poiss}(y_1; \lambda_1 T_1) \cdot \text{Poiss}(y_2; \lambda_0 T_2) \cdot \text{Poiss}(y_3; (\lambda_1 + \lambda_0)T_3)\cdot$$

$$\left.\log_2[\text{Poiss}(y_1; \lambda_1 T_1) \cdot \text{Poiss}(y_2; \lambda_0 T_2) \cdot \text{Poiss}(y_3; (\lambda_1 + \lambda_0)T_3)]\right)+$$

$$\left(p^2 \cdot \sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}\sum_{y_3=0}^{\infty}(\text{Poiss}(y_1; \lambda_1 T_1) \cdot \text{Poiss}(y_2; \lambda_1 T_2) \cdot \text{Poiss}(y_3; 2\lambda_1 T_3)\cdot\right.$$

$$\left.\left.\log_2[(\text{Poiss}(y_1; \lambda_1 T_1) \cdot \text{Poiss}(y_2; \lambda_1 T_2) \cdot \text{Poiss}(y_3; 2\lambda_1 T_3)]\right)\right].$$

Appendix B
Expression of Probability of total correct detections P_d

$$P_d =$$
$$\sum_{(y_1, y_2, y_3)} f_{00} : \{(f_{00} >= f_{01}) \& (f_{00} >= f_{10}) \& (f_{00} >= f_{11})\}+$$
$$\sum_{(y_1, y_2, y_3)} f_{01} : \{(f_{01} > f_{00}) \& (f_{01} >= f_{10}) \& (f_{01} >= f_{11}))\}+$$
$$\sum_{(y_1, y_2, y_3)} f_{10} : \{(f_{10} > f_{00}) \& (f_{10} > f_{01}) \& (f_{10} >= f_{11})\}+$$
$$\sum_{(y_1, y_2, y_3)} f_{11} : \{(f_{11} > f_{00}) \& (f_{11} > f_{01}) \& (f_{11} > f_{10})\}.$$

Where

$$f_{00} \equiv (1-p)^2 \cdot \mathrm{Poiss}(y_1; \lambda_0 T_1) \cdot \mathrm{Poiss}(y_2; \lambda_0 T_2) \cdot \mathrm{Poiss}(y_3; 2\lambda_0 T_3),$$
$$f_{01} \equiv p(1-p) \cdot \mathrm{Poiss}(y_1; \lambda_0 T_1) \cdot \mathrm{Poiss}(y_2; \lambda_1 T_2) \cdot \mathrm{Poiss}(y_3; (\lambda_0 + \lambda_1)T_3),$$
$$f_{10} \equiv p(1-p) \cdot \mathrm{Poiss}(y_1; \lambda_1 T_1) \cdot \mathrm{Poiss}(y_2; \lambda_0 T_2) \cdot \mathrm{Poiss}(y_3; (\lambda_1 + \lambda_0)T_3),$$
$$f_{11} \equiv p^2 \cdot \mathrm{Poiss}(y_1; \lambda_1 T_1) \cdot \mathrm{Poiss}(y_2; \lambda_1 T_2) \cdot \mathrm{Poiss}(y_3; 2\lambda_1 T_3),$$
$$\& : \text{logical AND operation}$$

Appendix C
Numerical approximation of the Poisson pmf and mutual information

$$\mathrm{Poiss}(x; \lambda) \approx \begin{cases} \frac{e^\lambda \lambda^x}{x!} & x \le x_c \\ 0 & x > x_c \end{cases} \tag{19}$$

where x_c is such that cummulative distribution function of Poisson random variable at x_c is approximately equal to 1. i.e. $\mathrm{CDFPoiss}(x_c; \lambda) \approx 1$. When represented in double precision floating-point arithmetic in IEEE 754 standard; this means that numerical values of $\mathrm{CDFPoiss}(x_c; \lambda)$ and $\mathrm{CDFPoiss}(x_c + 1; \lambda)$ are identical and equal to numerical one.

The infinite summations in mutual information expression given in appendix A are all truncated from 0 to $x_c = \mathrm{CDFPoiss}^{-1}(1 - \epsilon; (\lambda_1 + \lambda_1)T)$ for a given value of λ_1 and T. $\mathrm{CDFPoiss}^{-1}$ is inverse Poisson cdf and ϵ is, the machine precision number, 2^{-53} [18].

Appendix D
Code description

Since computations for I can be done in parallel, we may exploit this fact by vectorization of the algorithm and then evaualting I at grid points in parallel as defined in flowchart in Fig.(9) and Fig.(10) for fast computational throughput. However, it must be noted that this way of vectorized computations by first performing the truncation of Poisson distribution and then forming the grid of points, for further calculation of I, is not effective for large values of Poisson intensities involved in the problem. This is because the grid size increases cubely w.r.t truncated support of involved pmfs. So, for large values of Poisson intensities, say $\lambda >> 100$, we need to resort to Monte Carlo method.

For calculation of P_d, the same computed conditional Poisson pmfs can be used for all the comparisons that are required in conditional expression given in Appendix B.
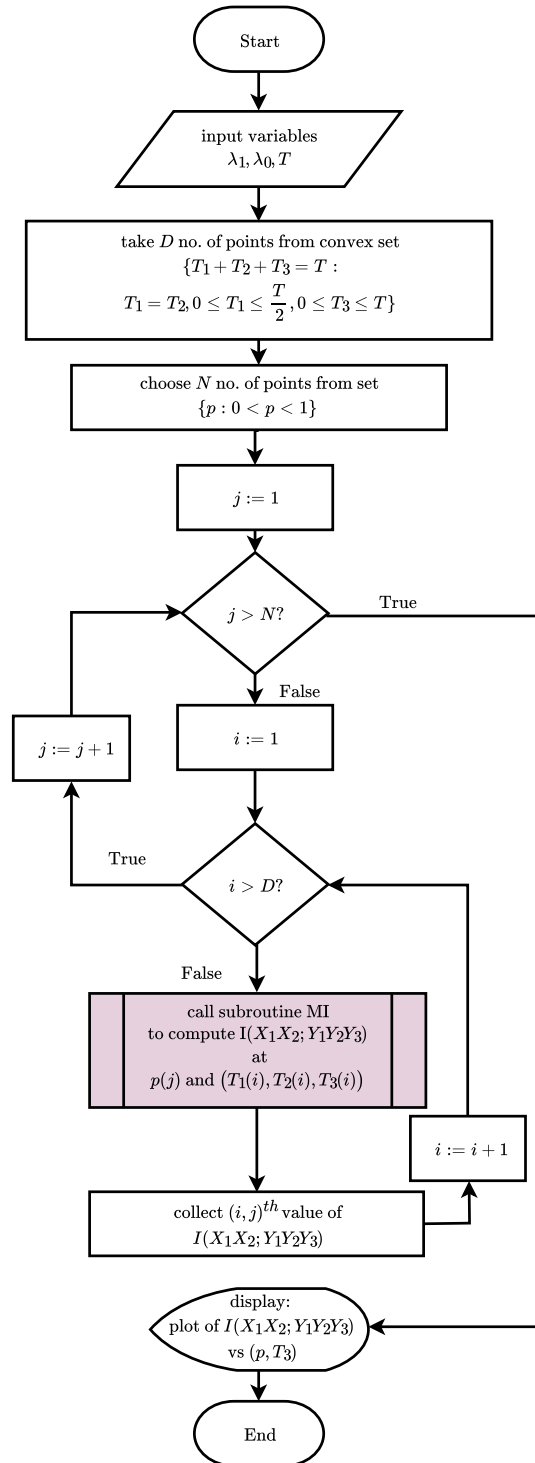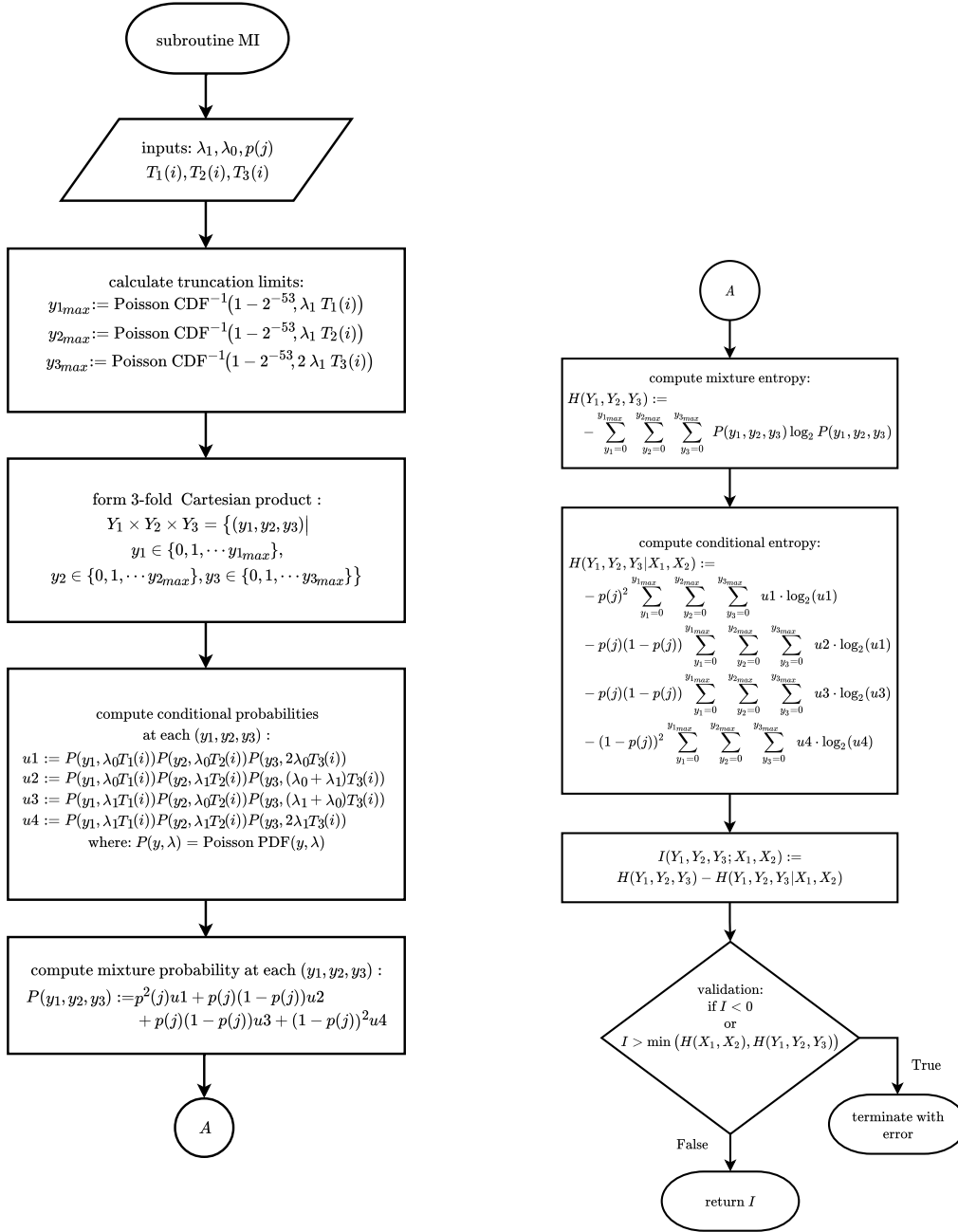
Fig. 9: Flowchart 1: Algorithm description for computation of I.

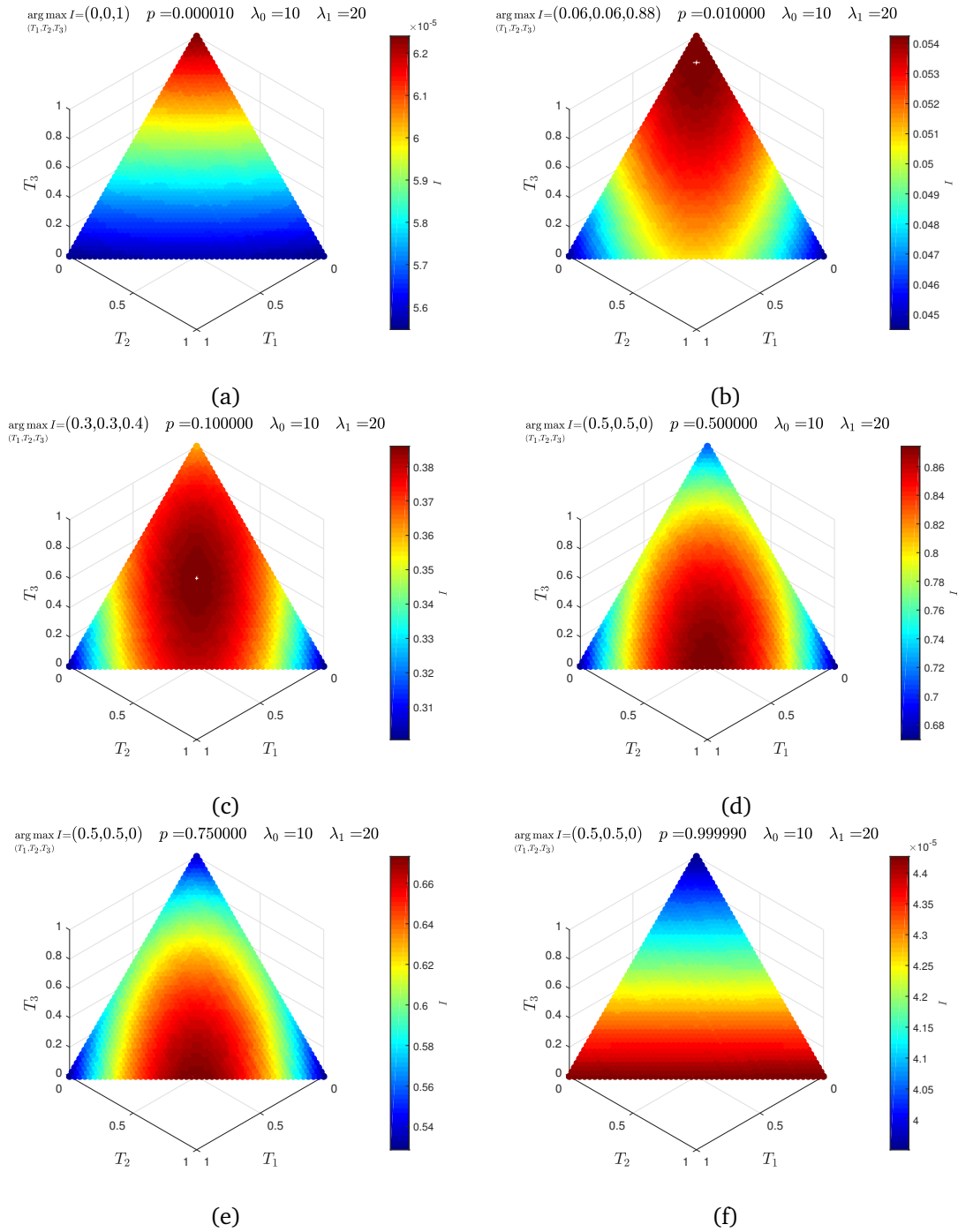Fig. 10: Flowchart 2: subroutine for computation of I.

Fig. 11: $I(X;Y)$ vs. (T_1, T_2, T_3) under time constraint $T_1 + T_2 + T_3 = 1$ for $\lambda_0 = 10$, $\lambda_1 = 20$, $T = 1$ and varying *prior* probability p. It can be seen from (a)-(f) that as prior p varies from 0.00001 to 0.5, the optimal solution drifts from $(0, 0, 1)$ to $(0.5, 0.5, 0)$ and stays there as p is varied further from 0.5 to 0.99999 along the line of symmetry $(T_1, T_2, T_3) := (\frac{1-\alpha}{2}, \frac{1-\alpha}{2}, \alpha)$ where $0 \leq \alpha \leq 1$.
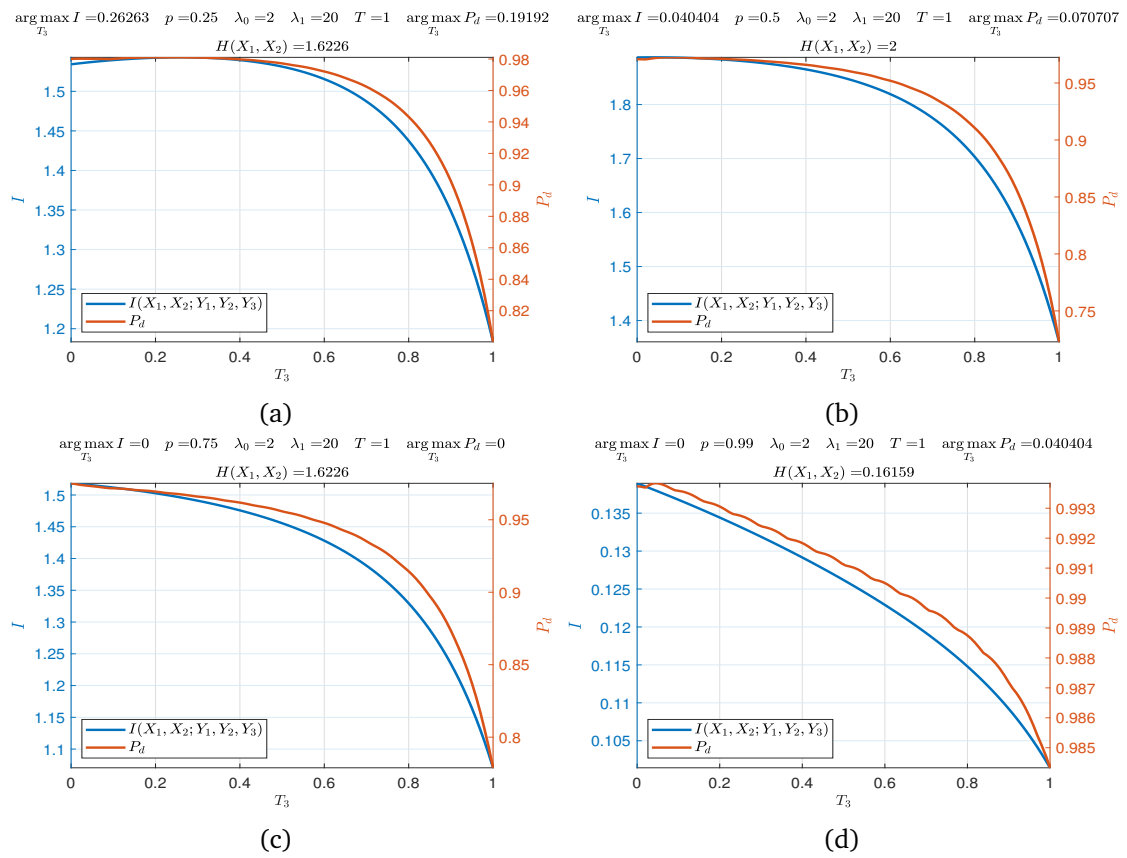
Fig. 12: Mutual information $I(X;Y)$ and probability of total correct detections P_d vs. T_3 for *prior* probabilities of $0.25, 0.5, 0.75$ and 0.99.
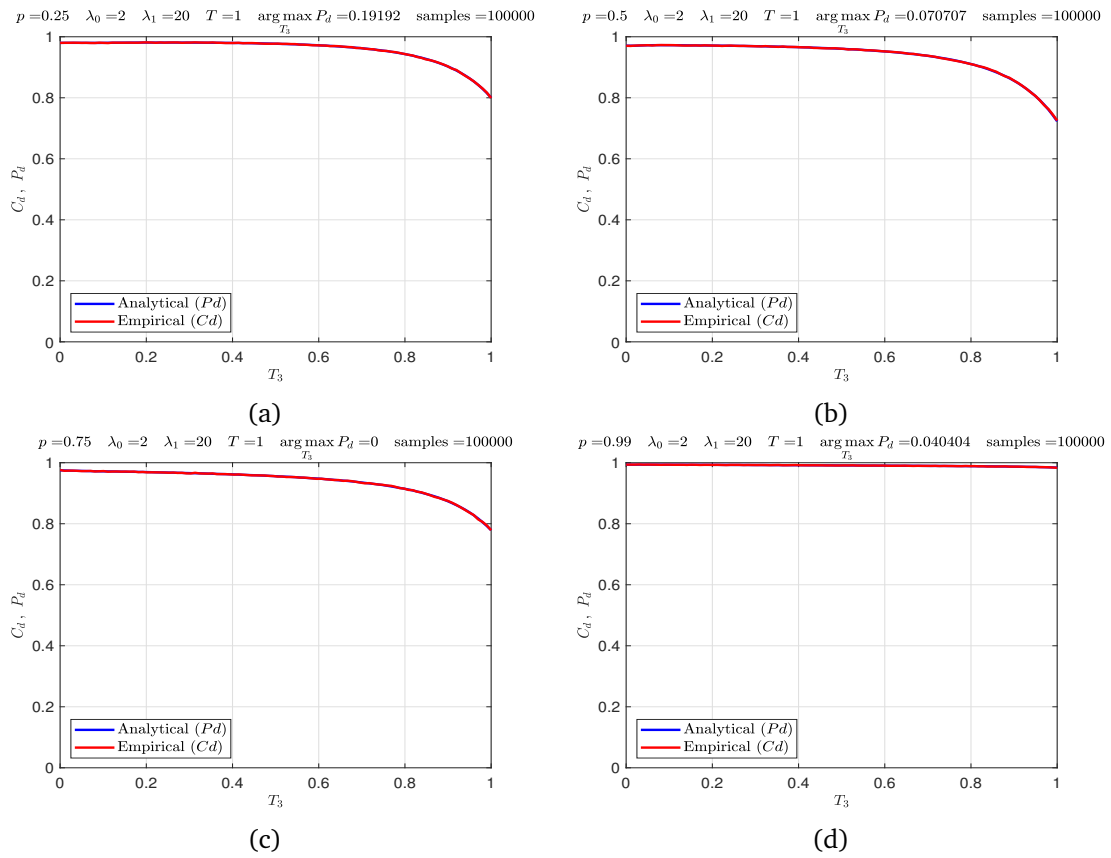
Fig. 13: Empirical Correct-decision rate C_d and analytical probability of total correct detections P_d vs. T_3 for *prior* probabilities of $0.25, 0.5, 0.75$ and 0.99.
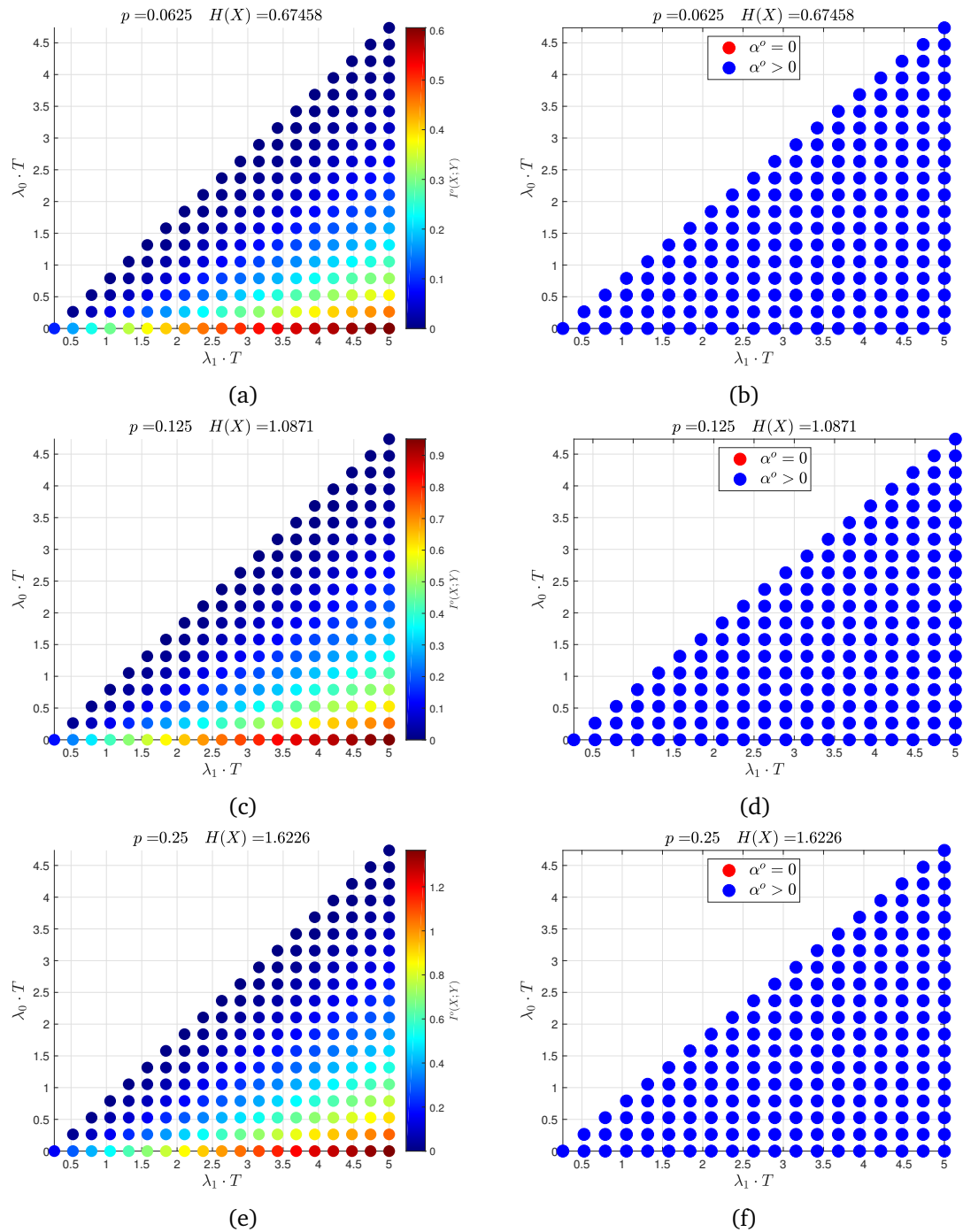
Fig. 14: Left: $I^O(X;Y)$ vs. $(\lambda_0 T, \lambda_1 T)$ in the region $\lambda_1 T > \lambda_0 T$, right: corresponding optimal argument parameter α^O vs. $(\lambda_0 T, \lambda_1 T)$ for varying *prior* probabilities p. The search for each optimal argument α^O for any fixed: $(\lambda_0 T, \lambda_1 T)$ and p is performed over the line $(T_1, T_2, T_3) := (\frac{1-\alpha}{2}, \frac{1-\alpha}{2}, \alpha)$ where $0 \le \alpha \le 1$ and $T_1 + T_2 + T_3 = 1$.
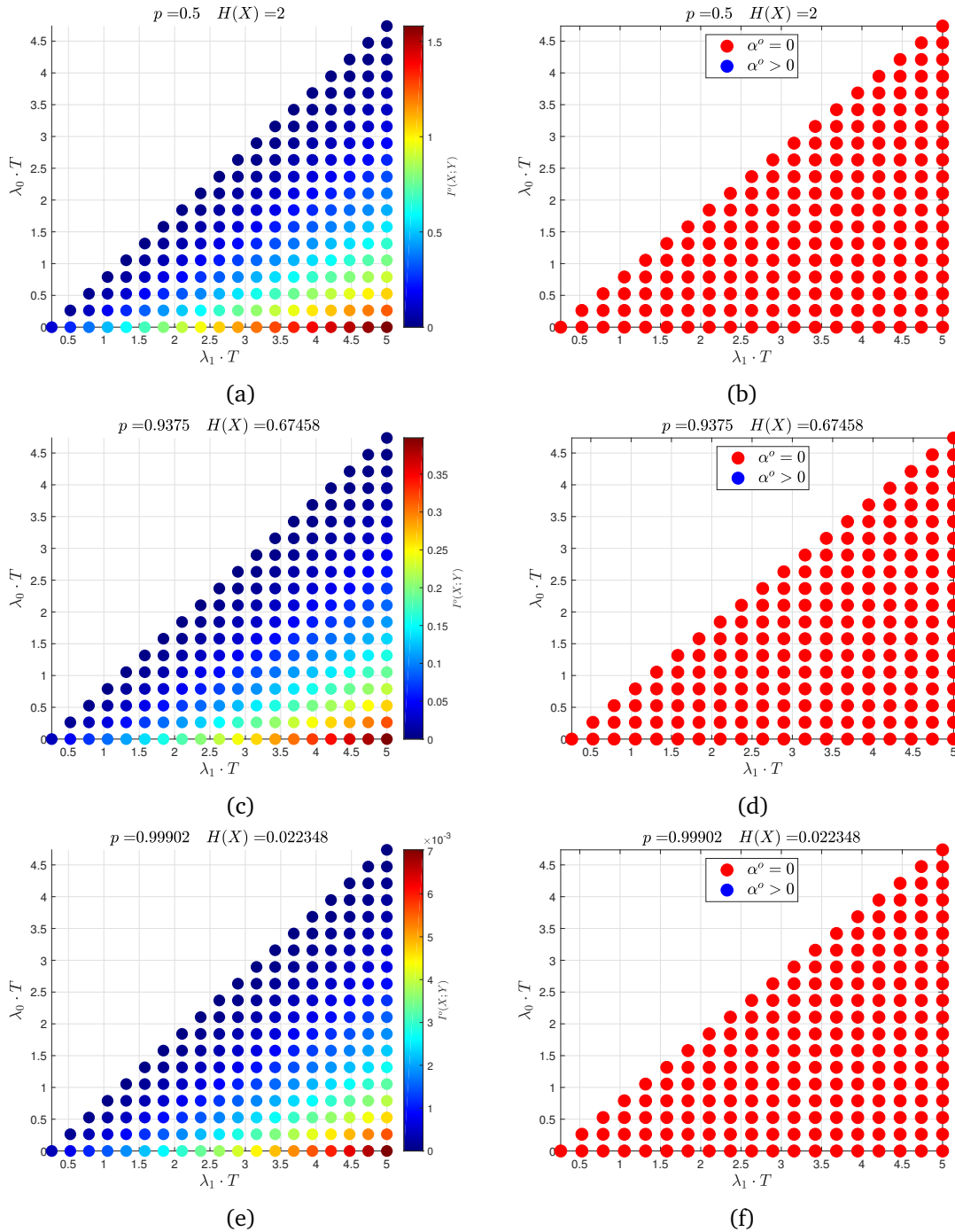
Fig. 15: Left: $I^O(X;Y)$ vs. $(\lambda_0 T, \lambda_1 T)$ in the region $\lambda_1 T > \lambda_0 T$, right: corresponding optimal argument parameter α^O vs. $(\lambda_0 T, \lambda_1 T)$ for varying *prior* probabilities p. The search for each optimal argument α^O for any fixed: $(\lambda_0 T, \lambda_1 T)$ and p is performed over the line $(T_1, T_2, T_3) := (\frac{1-\alpha}{2}, \frac{1-\alpha}{2}, \alpha)$ where $0 \le \alpha \le 1$ and $T_1 + T_2 + T_3 = 1$.

REFERENCES

[1] A. O. Hero, D. Castañón, D. Cochran, and K. Kastella, *Foundations and Applications of Sensor Management*. Springer Science & Business Media, 2007.

[2] H. Lee, K. L. Teo, and A. E. Lim, "Sensor scheduling in continuous time," *Automatica*, vol. 37, no. 12, pp. 2017–2023, 2001.

[3] J. M. Manyika and H. F. Durrant-Whyte, "Information-theoretic approach to management in decentralized data fusion," in *Applications in Optical Science and Engineering*. International Society for Optics and Photonics, 1992, pp. 202–213.

[4] W. W. Schmaedeke, "Information-based sensor management," in *Optical Engineering and Photonics in Aerospace Sensing*. International Society for Optics and Photonics, 1993, pp. 156–164.

[5] Y. Yang and R. S. Blum, "MIMO radar waveform design based on mutual information and minimum mean-square error estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 1, 2007.

[6] M. Payaró and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3613–3628, 2009.

[7] S. Verdú, "Mismatched estimation and relative entropy," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3712–3720, 2010.

[8] D. Guo, S. Shamai, and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1837–1849, 2008.

[9] R. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the Poisson channel," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1302–1318, 2012.

[10] L. Wang, D. E. Carlson, M. R. Rodrigues, R. Calderbank, and L. Carin, "A Bregman matrix and the gradient of mutual information for vector Poisson and Gaussian channels," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2611–2629, 2014.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[12] S. M. Ross, *Stochastic Processes*. Wiley New York, 1996, vol. 2.

[13] M. Fahad, "Sensing Methods for Two-Target and Four-Target Detection in Time-Constrained Vector Poisson and Gaussian Channels," Ph.D. dissertation, Michigan Technological University, Houghton, MI, 49931-1295, May 2021. [Online]. Available: https://doi.org/10.37099/mtu.dc.etdr/1193

[14] R. W. Yeung, *Information Theory and Network Coding*. Springer Science & Business Media, 2008.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

[16] T. A. Schonhoff and A. A. Giordano, *Detection and Estimation Theory and its Applications*. Pearson College Division, 2006.

[17] J. A. Adell, A. Lekuona, and Y. Yu, "Sharp bounds on the entropy of the Poisson law and related quantities," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2299–2306, 2010.

[18] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*. SIAM, 1997, vol. 50.

Muhammad Fahad (M'17) received the Ph.D. and M.S degree in Electrical Engineering from Michigan Technological University, Houghton, MI in 2021 and 2016, respectively. He received M.S. degree in Electrical Engineering, from University of Engineering and Technology, Lahore, Pakistan and B.E. degree *(with distinction)* in Electronics Engineering from Mehran University of Engineering and Technology, Pakistan, in 2007 and 2005, respectively.

He was a Fulbright Ph.D. Scholar during 2011-2017 and currently serving as an Adjunct Assistant Professor in the Department of Applied Computing, College of Computing at Michigan Technological University, Houghton, MI. His research interests are in sensor management and Poisson processes.

Daniel R. Fuhrmann (S'78-M'84-SM'95-F'10) received the B.S.E.E. degree *(cum laude)* from Washington University, St. Louis, MO, USA, in 1979 and the M.A., M.S.E., and the Ph.D. degrees from Princeton University, Princeton, NJ, in 1982 and 1984, respectively.

From 1984 to 2008, he was with the Department of Electrical Engineering, now the Department of Electrical and Systems Engineering, Washington University. In the 2000-2001 academic year, he was a Fulbright Scholar visiting the Universidad Nacional de La Plata, Buenos Aires, Argentina. He is currently the Dave House Professor and Chair of the Department of Applied Computing, Michigan Technological University, Houghton, MI. His research interests lie in various areas of statistical signal and image processing, including sensor array signal processing, radar systems, and adaptive sensing.

Dr. Fuhrmann is a former Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was the Technical Program Chairman for the 1998 IEEE Signal Processing Workshop on Statistical Signal and Array Processing, and General Chairman of the 2003 IEEE Workshop on Statistical Signal Processing. He has been an ASEE Summer Faculty Research Fellow with the Naval Underwater Systems Center, New London, CT, a consultant to MIT Lincoln Laboratory, Lexington, MA, and an ASEE Summer Faculty Fellow with the Air Force Research Laboratory, Dayton, OH.