A Strong Data Processing Inequality for Thinning Poisson Processes and Some Applications

Ligong Wang

A Strong Data Processing Inequality for Thinning Poisson Processes and Some Applications

Ligong Wang
ETIS – Université Paris Seine / Université de Cergy-Pontoise / ENSEA / CNRS
Cergy-Pontoise, France
ligong.wang@ensea.fr

***Abstract*—This paper derives a simple strong data processing inequality (DPI) for Poisson processes: after a Poisson process is passed through p-thinning—in which every arrival remains in the process with probability p and is erased otherwise, independently of the other points—the mutual information between the Poisson process and any other random variable is reduced to no more than p times its original value. This strong DPI is applied to prove tight converse bounds in several problems: a hypothesis test with communication constraints, a mutual information game, and a CEO problem.**

I. Introduction

A data processing inequality (DPI) states that, if the random variables $U \multimap X \multimap Y$ form a Markov chain in that order, then $I(U;Y) \leq I(U;X)$. For some channels (i.e., stochastic kernels) from X to Y, a stronger inequality holds: for any joint distributions on (U,X,Y) under which the Markov condition is satisfied and $P_{Y|X}$ is the given channel law, $I(U;Y) \leq \alpha I(U;X)$, where $\alpha < 1$ does not depend on the choice of the joint distribution. The latter inequality is usually called a "strong DPI." A simple example of strong DPI is the following [1, Exercise 3.19].

Example 1: If the channel $W(\cdot|\cdot)$ is such that, for some y_0, $W(y_0|x) \geq c$ for all x, then, for any joint distribution on (U,X,Y) satisfying the Markov condition $U \multimap X \multimap Y$ and with $P_{Y|X}$ given by $W(\cdot|\cdot)$,

$$I(U;Y) \leq (1-c)I(U;X). \tag{1}$$

Various strong DPIs have been derived for various channels. Some of them, like the one that is proven in the current work, are "input dependent," meaning that they hold for a specific distribution for X but not necessarily for all distributions.

Like normal DPIs, strong DPIs are useful tools in proving converse results in information theory and other areas. We refer to [2] for a survey of some known results.

The channel considered in this paper is *thinning* on point, in particular, Poisson, processes. Here, "p-thinning," $p \in (0,1)$, refers to the operation of independently erasing each arrival in a point process with probability $(1-p)$. Thinning occurs in many practical scenarios. For example, if the point process is a beam of photons, then p-thinning can describe passing the beam through a beamsplitter of transmissivity p, or detecting the photons with a photodetector of efficiency p.

Instead of a point process, thinning can also be defined on a nonnegative-integer-valued random variable: Y is the p-thinning of X if, conditional on $X = x \in \mathbb{Z}_0^+$, Y has the binomial distribution of parameters x and p; see [3]–[5] and references therein. Clearly, if the point process Y_0^T is the p-thinning of X_0^T, then the total number of arrivals in Y_0^T is the p-thinning of the total number of arrivals in X_0^T. In this work, we are only concerned with thinning of point processes.

The strong DPI we derive is the following: if Y_0^T is the p-thinning of a Poisson process X_0^T, then[1]

$$I(U;Y_0^T) \leq pI(U;X_0^T) \tag{2}$$

for all $U \multimap X_0^T \multimap Y_0^T$. This inequality is related to (1): heuristically, one can think of thinning as replacing every infinitesimal interval in X_0^T by an interval of the same length that contains no arrivals (corresponding to y_0 in Example 1) with probability $(1-p)$. Hence one can say, in a sense, that (2) follows from (1) and the memorylessness of the Poisson process. A formal derivation of (2) is given in Section II.

Inequality (2) is reminescent of [5, Lemma 1], which is a strong DPI concerning thinned Poisson random variables. The difference between the inequalities is that one concerns continuous-time random processes, while the other concerns random variables. Neither (2) nor [5, Lemma 1] appears to be a direct consequence of the other.

Simple as it is, (2) can be used to prove tight converse bounds in various problems. We discuss three examples: a hypothesis test against independence, a mutual information game, and a CEO problem.

Inequality (2) is closely related to a problem of "covering point patterns" [6], [7]. The latter is a rate-distortion problem for point processes, where the reconstruction signal is a set that must contain all the arrivals, and the distortion is defined as the Lebesgue measure of the reconstruction set. Roughly speaking, equality in (2) can be approached by choosing U as an optimal reconstruction set for X_0^T. This relation is exploited in all three applications of (2) that we discuss. Our CEO problem is a multiple-user extension of the original rate-distortion problem of [6]. As for the other two problems, their formulation does not involve the concept of covering, but their optimal achievability schemes do.

The rest of this paper is arranged as follows: Section II derives the strong DPI; Section III demonstrates how to approach equality in it; Section IV studies some applications; and Section V concludes the paper with a few remarks.

[1]Here, mutual information can be defined via the Kullback-Leibler divergence between general probability measures.

II. THE INEQUALITY

A. Background

We provide some background on point processes. For more details, see [8], [9].

We describe point processes by their counting functions, thus a point process $X_0^T \triangleq \{X_t \colon t \in [0,T]\}$ on the interval $[0,T]$ is, with probability one, an integer-valued, non-decreasing, right-continuous random process satisfying $X_0 = 0$. The number of *arrivals* (i.e., points) in an interval $(t_1, t_2] \subseteq [0,T]$ contained in X_0^T is given by $X_{t_2} - X_{t_1}$. A point process is called *simple* if no two arrivals occur simultaneously, i.e., if with probability one all jumps in X_0^T are unit jumps. For a simple point process X_0^T one can define the *conditional intensity function*

$$\alpha_t(x_0^t) \triangleq \lim_{\Delta \downarrow 0} \frac{\mathsf{Pr}(X_{t+\Delta} - X_t \geq 1 | X_0^t = x_0^t)}{\Delta}, \quad t \in [0,T]. \tag{3}$$

The integrated conditional intensity function is the *compensator* of the point process.

Let P and Q be two probability distributions corresponding to simple point processes on $[0,T]$, with conditional intensity functions α and β, respectively. The Radon-Nikodym derivative between P and Q can be written as [8, (19.125)]

$$\frac{\mathrm{d}P}{\mathrm{d}Q}(x_0^T) = \exp\left\{ \int_0^T \log \frac{\alpha_t(x_0^t)}{\beta_t(x_0^t)} \,\mathrm{d}x_0^t - \int_0^T \left(\alpha_t(x_0^t) - \beta_t(x_0^t)\right) \mathrm{d}t \right\}. \tag{4}$$

It then follows that their Kullback-Leibler divergence is

$$D(P\|Q) = \mathsf{E}_P\left[\int_0^T \left(\alpha_t(X_0^t) \log \frac{\alpha_t(X_0^t)}{\beta_t(X_0^t)} - \alpha_t(X_0^t) + \beta_t(X_0^t) \right) \mathrm{d}t \right]. \tag{5}$$

A *Poisson process* is a point process whose conditional intensity function does not depend on the past realization of the process, and is hence a function of time alone, henceforth simply called the *intensity*. A Poisson process is called *homogeneous* if its intensity is constant over $t \in [0,T]$.

Given a point process X_0^T on the interval $[0,T]$ and a constant $p \in (0,1)$, the point process Y_0^T is called the *p-thinning* of $\mathcal{X}$ if

1) every arrival that occurs in Y_0^T also occurs in X_0^T with probability one; and
2) every arrival that occurs in X_0^T occurs in Y_0^T with probability p, independently both of X_0^T and of whether other arrivals in X_0^T occur in Y_0^T or not.

If Y_0^T is the p-thinning of X_0^T, then (see [9, Corollary 6.2.6])

$$\lim_{\Delta \downarrow 0} \frac{\mathsf{Pr}(\{Y_{t+\Delta} - Y_t \geq 1\} | X_0^t = x_0^t, Y_0^t = y_0^t)}{\Delta} = p\alpha_t(x_0^t). \tag{6}$$

Hence, if β is the conditional intensity function for Y_0^T, then

$$\beta(y_0^t) = p \cdot \mathsf{E}_{X_0^t | Y_0^t = y_0^t}\left[\alpha_t(X_0^t)\right]. \tag{7}$$

B. Proof of the Inequality

We formally state the strong DPI in terms of both mutual information and Kullback-Leibler divergence.

Theorem 1 (Mutual Information): Let X_0^T be a Poisson process on $[0,T]$, and let Y_0^T be the p-thinning of X_0^T for some $p \in (0,1)$. Let U be any random variable such that $U \multimap X_0^T \multimap Y_0^T$ form a Markov chain. Then

$$I(U; Y_0^T) \leq pI(U; X_0^T). \tag{8}$$

Theorem 2 (Kullback-Leibler Divergence): Let Q denote the distribution of a Poisson process on $[0,T]$. Let P be another distribution that is absolutely continuous with respect to Q. Further, let P' and Q' be the distributions resulting from the p-thinning of P and Q, respectively, for some $p \in (0,1)$. Then

$$D(P'\|Q') \leq pD(P\|Q). \tag{9}$$

The two versions of the strong DPI are equivalent; see [2, Theorem 4]. We provide a proof for Theorem 2.

Proof of Theorem 2: For notational convenience, we use X_0^T and Y_0^T to denote point processes before and after p-thinning, respectively. Let α and β be the conditional intensity functions corresponding to P and P', respectively. Further, let λ_t, $t \in [0,T]$, be the intensity corresponding to Q, then the intensity corresponding to Q' is $p\lambda_t$, $t \in [0,T]$. It then follows from (5) that

$$D(P\|Q) = \mathsf{E}_P\left[\int_0^T \left(\alpha(X_0^t) \log \frac{\alpha(X_0^t)}{\lambda_t} - \alpha(X_0^t) + \lambda_t \right) \mathrm{d}t \right] \tag{10}$$

$$D(P'\|Q') = \mathsf{E}_{P'}\left[\int_0^T \left(\beta(Y_0^t) \log \frac{\beta(Y_0^t)}{p\lambda_t} - \beta(Y_0^t) + p\lambda_t \right) \mathrm{d}t \right]. \tag{11}$$

The claim follows immediately by recalling (7), convexity of the function $a \mapsto a \log a$, and Jensen's inequality. ■

Note that (8) need *not* hold if X_0^T is not Poisson; similarly, (9) need *not* hold if Q does not represent a Poisson process. The following simple example demonstrates the former. Let U be uniform on $\{0,1\}$. Conditional on $U = u$, X_0^T is a homogeneous Poisson process of intensity λu. Hence X_0^T is a doubly-stochastic Poisson process (Cox process). One can see that, if $p\lambda T$ is large, then with high probability $U = 1$ will result in at least one arrival in Y_0^T, whereas with probability one $U = 0$ will result in no arrival in Y_0^T, hence $I(U; Y_0^T)$ is close to one bit. Similarly, $I(U; X_0^T)$ is also close to one bit.

III. APPROACHING EQUALITY IN (8) BY COVERING THE ARRIVALS

Both (8) and (9) are tight. For (9) one can easily verify that equality is achieved when P is also Poisson. To see how to approach equality in (8), we first briefly review a related problem that is studied in [6], [7].

Let x_0^T be the realization of a point process on $[0,T]$. A *covering function* $\hat{x}_0^T$ is a $\{0,1\}$-valued Lebesgue-measurable function on $[0,T]$. Define

$$d(x_0^T, \hat{x}_0^T) \triangleq \begin{cases} \frac{\mu_{\mathrm{L}}(\hat{x}^{-1}(1))}{T}, & \text{if all arrivals in } x_0^T \text{ are in } \hat{x}^{-1}(1), \\ \infty, & \text{otherwise,} \end{cases} \tag{12}$$

where $\mu_{\mathrm{L}}(\cdot)$ denotes the Lebesgue measure. We call $\hat{x}^{-1}(1)$ the *covering set*. A rate-distortion pair (R,D) is said to be achievable on a point process if, for every large enough T, there exists an encoder mapping a point pattern to a covering function taken from a codebook of size $\lfloor e^{TR} \rfloor$ such that the expected distortion between the point pattern and the covering function does not exceed D. The rate-distortion function $R(D)$ is the infimum over R such that (R,D) is achievable.

For a homogeneous Poisson process of intensity λ, [6] shows

$$R(D) = -\lambda \log D, \quad D \in (0,1). \tag{13}$$

The following lemma is a slight variation of (13): instead of the size of the codebook, it concerns the mutual information between the source and the reconstruction. The proof of this lemma is similar to that of (13), and is omitted due to space limitations.

Lemma 1: Let X_0^T be the homogeneous Poisson process of intensity λ on $[0,T]$. For all $D \in (0,1)$,

$$\lim_{T\to\infty} \inf_{\mathsf{E}\left[d(X_0^T, \hat{X}_0^T)\right] \le D} \frac{I(X_0^T; \hat{X}_0^T)}{T} = -\lambda \log D. \tag{14}$$

We can now use Lemma 1 to demonstrate how equality in (8) can be approached. Fix any $\epsilon > 0$ and $D \in (0,1)$. Let X_0^T be homogeneous Poisson of intensity λ,[2] and let Y_0^T be the p-thinning of X_0^T, which is homogeneous Poisson of intensity $p\lambda$. By Lemma 1, for large enough T, one can choose a covering function $\hat{X}_0^T$ such that $\hat{X}_0^T \multimap X_0^T \multimap Y_0^T$, $\mathsf{E}\left[d(X_0^T, \hat{X}_0^T)\right] \le D$, and $I(X_0^T; \hat{X}_0^T) \le -\lambda T \log D + T\epsilon$. Since Y_0^T is obtained by thinning X_0^T, we know that any $\hat{X}_0^T$ that covers X_0^T must also cover Y_0^T. This implies that $\mathsf{E}\left[d(Y_0^T, \hat{X}_0^T)\right] = \mathsf{E}\left[d(X_0^T, \hat{X}_0^T)\right] \le D$. Again by Lemma 1 it follows that $I(Y_0^T; \hat{X}_0^T) \ge -p\lambda T \log D - T\epsilon$. Hence choosing $U = \hat{X}_0^T$ as above for arbitrarily small ϵ will asymptotically achieve equality in (8) when T tends to infinity.

Below we recall two more relevant results from [6], [7] that will be used in Section IV.

Lemma 2 (General Processes [6]): On any point process $\{X_t, t \in \mathbb{R}_0^+\}$ satisfying

$$\lim_{t\to\infty} \mathsf{Pr}(X_t > (\lambda+\delta)t) = 0 \quad \text{for all } \delta > 0, \tag{15}$$

the rate-distortion pair $(R(D)+\epsilon, D)$, with $R(D)$ given in (13), is achievable for all $\epsilon > 0$.

[2] A similar construction also works for inhomogeneous Poisson processes.

Lemma 3 (Arbitrary Point Patterns [7]): Fix any $\epsilon > 0$ and let $R(D)$ be given in (13). For every large enough T, there exists a codebook containing $e^{T(R(D)+\epsilon)}$ covering sets each having Lebesgue measure not exceeding DT, such that every subset of $[0,T]$ of cardinality not exceeding λT is contained in at least one of the covering sets in this codebook.

IV. Applications

A. Test Against Independence

Hypothesis testing with communication constraints is a classic problem in which physically separated observers make decisions based on their own observations together with the messages sent to them by other observers [10], [11]. Here we consider a Poisson version of this problem. Let X_0^T be a homogeneous Poisson process of intensity λ on $[0,T]$, and fix $p \in (0,1)$. Given $H=0$, Y_0^T is the p-thinning of X_0^T; given $H=1$, Y_0^T is a homogeneous Poisson process of intensity $p\lambda$ that is independent of X_0^T. The transmitter observes X_0^T and describes it to the receiver using TR nats, where $R>0$ is the available communication rate (in nats per second). The receiver observes Y_0^T and guesses the value of H based on Y_0^T and the message from the transmitter. Let $P_\epsilon(T,R)$, $\epsilon>0$, denote the smallest achievable error probability of the receiver guessing $H=0$ when $H=1$, provided that the error probability of it guessing $H=1$ when $H=0$ is not more than ϵ. Further let

$$\theta(R) \triangleq -\lim_{\epsilon\downarrow 0} \lim_{T\to\infty} \frac{\log P_\epsilon(T,R)}{T}. \tag{16}$$

Theorem 3: For all $R>0$,

$$\theta(R) = pR. \tag{17}$$

Proof: Converse. Following standard arguments [10] we have,

$$\theta(R) \le (1+\delta_T)\frac{I(M; Y_0^T)}{T}, \tag{18}$$

where M denotes the message sent by the transmitter to the receiver, δ_T tends to zero as T tends to infinity, and the mutual information is computed under $H=0$. Clearly, $M \multimap X_0^T \multimap Y_0^T$ form a Markov chain, hence Theorem 1 implies

$$I(M; Y_0^T) \le pI(M; X_0^T) \le p \cdot TR. \tag{19}$$

Combining (18) and (19) and letting T go to infinity yields $\theta(R) \le pR$.

Achievability. By Lemma 3, the transmitter can construct a codebook containing $(e^{RT}-1)$ covering sets, each having Lebesgue measure DT, such that every point pattern containing no more than $(1+\delta)\lambda T$ arrivals is covered by at least one covering set, where

$$D = (1+\epsilon)e^{-\frac{R}{(1+\delta)\lambda}}. \tag{20}$$

Here δ and ϵ are arbitrarily small positive numbers. Label the covering sets by 1 to $(e^{RT}-1)$. The transmitter's strategy is the following: if $X_T \le (1+\delta)\lambda T$, it looks for a covering set in the codebook to cover all arrivals in X_0^T, and sends its index to the receiver; if $X_T > (1+\delta)\lambda T$, it sends zero. Upon receiving a nonzero index, the receiver reconstructs the corresponding covering set and compares it with Y_0^T. If the covering set covers all arrivals in Y_0^T, and if $Y_T \geq (1-\delta)p\lambda T$, then it guesses $H = 0$; otherwise it guesses $H = 1$. If the receiver receives index zero, then it also guesses $H = 1$.

Next we analyze the error probabilities. When $H = 0$, if $X_T \leq (1+\delta)\lambda T$, then the covering set chosen by the transmitter will cover Y_0^T with probability one. Hence the receiver will guess $H = 1$ only if either $X_T > (1+\delta)\lambda T$ or $Y_T < (1-\delta)p\lambda T$. The probabilities of both cases tend to zero as T tends to infinity, by the Law of Large Numbers.

When $H = 1$, Y_0^T is a Poisson process of intensity $p\lambda$ generated independently of X_0^T and, hence, also of the covering set chosen by the transmitter. The receiver will guess $H = 0$ if, and only if, Y_0^T contains more than $(1-\delta)p\lambda T$ arrivals, and all these arrivals lie in the chosen covering set. Note that, given that the realization of a homogeneous Poisson process contains k arrivals on $[0,T]$, these arrivals, when randomly labeled, are independent and identically distributed, uniformly on $[0,T]$ [12, Theorem 2.4.6]. Thus the probability of any of the arrivals lying within a certain subset of $[0,T]$ of Lebesgue measure DT is D. We then have

$$\begin{aligned}
\mathsf{Pr}(\text{error}|H=1) &= \sum_{k=(1-\delta)p\lambda T}^{\infty} \mathsf{Pr}(Y_T = k)\cdot D^k \\
&\leq \sum_{k=(1-\delta)p\lambda T}^{\infty} \mathsf{Pr}(Y_T = k)\cdot D^{(1-\delta)p\lambda T} \\
&\leq D^{(1-\delta)p\lambda T} \\
&= (1+\epsilon)^{(1-\delta)}\cdot e^{-\frac{(1-\delta)pRT}{(1+\delta)}}. \qquad (21)
\end{aligned}$$

Since both δ and ϵ can be arbitrarily small, the exponent on the right-hand side of (21) can be arbitrarily close to pR. ■

B. A Poisson Mutual Information Game

Various types of games have been studied in the literature where the quantity that the players wish to maximize or minimize is a mutual information. For example, [4] studies a mutual information game on a single Poisson random variable, and [13] studies a game on Gaussian vectors. In the following we study a mutual information game on a Poisson process.

Let X_0^T be a homogeneous Poisson process on $[0,T]$ with intensity λ. Player 1 chooses a stochastic kernel $P_{Y_0^T|X_0^T}$ that (randomly) adds at most μ arrivals per second to X_0^T to form the point process Y_0^T. Precisely, $P_{Y_0^T|X_0^T}$ must satisfy

$$\mathsf{Pr}\left(Y_T - X_T > (\mu+\delta_T)T\right) \leq \epsilon_T \qquad (22)$$

for some ϵ_T and δ_T that tend to zero as T tends to infinity. Player 1 is *not allowed to remove* any points from X_0^T. Player 2 then chooses a stochastic kernel $P_{U|Y_0^T}$, where U may take value in an arbitrary measurable set, such that

$$I(U;Y_0^T) \geq \alpha T \qquad (23)$$

for a given positive constant α. The quantity that Player 1 wishes to maximize and that Player 2 wishes to minimize is

$$\frac{1}{T}\left\{I(U;Y_0^T) - I(U;X_0^T)\right\}. \qquad (24)$$

Theorem 4: The asymptotic value of the above game is

$$\lim_{T\to\infty} \sup_{P_{Y_0^T|X_0^T}} \inf_{P_{U|Y_0^T}} \frac{1}{T}\left\{I(U;Y_0^T) - I(U;X_0^T)\right\} = \frac{\mu}{\lambda+\mu}\alpha. \qquad (25)$$

Proof: We first propose a strategy for Player 1 to guarantee that, for all $T > 0$ and for all choices by Player 2,

$$I(U;Y_0^T) - I(U;X_0^T) \geq \frac{\mu}{\lambda+\mu}\alpha T. \qquad (26)$$

This strategy is simply to add to X_0^T a homogeneous Poisson process of intensity μ that is independent of X_0^T. Then Y_0^T is homogeneous Poisson of intensity $(\lambda+\mu)$, and X_0^T is the $\lambda/(\lambda+\mu)$-thinning of Y_0^T. It follows from Theorem 1 that

$$I(U;X_0^T) \leq \frac{\lambda}{\lambda+\mu} I(U;Y_0^T), \qquad (27)$$

which, combined with (23), yields (26).

Next, for any strategy chosen by Player 1, we propose a corresponding strategy for Player 2. By Player 1's constraint (22), Y_0^T satisfies

$$\lim_{T\to\infty} \mathsf{Pr}\left(Y_T > (\lambda+\mu+\delta)T\right) = 0 \qquad (28)$$

for all positive δ. By Lemma 2, the rate-distortion pair $(\alpha, e^{-\alpha/(\lambda+\mu)})$ is achievable on $\{Y_t, t \in \mathbb{R}_0^+\}$, i.e., for any $\epsilon > 0$, for large enough T, there exists a codebook consisting of $e^{(\alpha+\epsilon)T}$ covering functions $\hat{y}_0^T$ using which one can achieve expected distortion

$$\mathsf{E}\left[d(Y_0^T, \hat{Y}_0^T)\right] \leq e^{-\alpha/(\lambda+\mu)}. \qquad (29)$$

The size of the codebook guarantees

$$I(Y_0^T;\hat{Y}_0^T) \leq H(\hat{Y}_0^T) \leq (\alpha+\epsilon)T. \qquad (30)$$

If $I(Y_0^T;\hat{Y}_0^T) \geq \alpha T$, then Player 2 chooses $U = \hat{Y}_0^T$; otherwise it chooses $U = (\hat{Y}_0^T, K)$, where K is an arbitrary bit string containing further information about Y_0^T (for example, K may be chosen as part of the bit string describing the exact position of the first arrival in Y_0^T) such that

$$\alpha T \leq I(U;Y_0^T) \leq (\alpha+\epsilon)T. \qquad (31)$$

Now note that

$$\mathsf{E}\left[d(X_0^T, \hat{Y}_0^T)\right] = \mathsf{E}\left[d(Y_0^T, \hat{Y}_0^T)\right] \leq e^{-\alpha/(\lambda+\mu)}. \qquad (32)$$

By Lemma 1, (32) implies that, for large enough T,

$$I(U;X_0^T) \geq I(\hat{Y}_0^T;X_0^T) \geq \left(\frac{\lambda}{\lambda+\mu} - \epsilon\right)\alpha T. \qquad (33)$$

Combining this with (31), we obtain

$$I(U;Y_0^T) - I(U;X_0^T) \leq \left(\frac{\mu}{\lambda+\mu} + 2\epsilon\right)\alpha T. \qquad (34)$$

where ϵ can be arbitrarily close to zero. ■

C. A Poisson CEO Problem

The "CEO (Chief Executive Officer) problem" in information theory usually refers to distributed source coding where the objective is to reconstruct a source from coded noisy observations of the source. Optimal solution to the CEO problem in the general setting is still unknown, but several special cases have been solved [14]–[16]. Here we consider a CEO problem where the source is a Poisson process.

Let X_0^T, W_0^T, and V_0^T be three mutually independent homogeneous Poisson processes on $[0,T]$ of intensities λ, μ, and ν, respectively. The source is X_0^T. Encoder 1 observes Y_0^T, which is the union of X_0^T and W_0^T: with probability one $Y_t = X_t + W_t$ for all $t \in [0,T]$. Encoder 2 observes Z_0^T, which is the union of X_0^T and V_0^T. Encoder 1 maps Y_0^T to a message M_1 containing TR_1 nats, and Encoder 2 maps Z_0^T to M_2 containing TR_2 nats. The CEO then maps the pair (M_1, M_2) to a covering function $\hat{X}_0^T$, which, as in Section III, is a $\{0,1\}$-valued Lebesgue-measurable function on $[0,T]$. Define the distortion function $d(\cdot,\cdot)$ as in (12). A triple (R_1, R_2, D) is said to be achievable if, for large enough T, there exists a triple of Encoder 1, Encoder 2, and CEO as above that achieves $\mathsf{E}\left[d(X_0^T, \hat{X}_0^T)\right] \le D$. The closure of the set of all achievable triples is called the rate-distortion region of the CEO problem.

First consider some simple cases. When $R_2 = 0$, Encoder 1 can use a codebook of covering sets for its own observation Y_0^T. By producing the covering set chosen by Encoder 1, the CEO is guaranteed to cover X_0^T, all arrivals in which are also in Y_0^T. Hence we can achieve all rate-distortion triples (R_1, R_2, D) satisfying

$$R_1 > -(\lambda+\mu)\log D \tag{35a}$$
$$R_2 = 0. \tag{35b}$$

Similarly, we can also achieve all triples such that

$$R_1 = 0 \tag{36a}$$
$$R_2 > -(\lambda+\nu)\log D. \tag{36b}$$

By time-sharing between the two strategies above we know that the rate-distortion region contains all triples (R_1, R_2, D) satisfying

$$\frac{R_1}{\lambda+\mu} + \frac{R_2}{\lambda+\nu} \ge -\log D. \tag{37}$$

As we next show, this simple time-sharing strategy is optimal.

Theorem 5: The rate-distortion region for the above CEO problem is characterized by (37).

Proof: It remains only to prove the converse part. To this end, recall Lemma 1: to achieve $d(X_0^T, \hat{X}_0^T) \le D$, one needs

$$I(X_0^T; \hat{X}_0^T) \ge T(R(D) - \epsilon) \tag{38}$$

with $R(D)$ given by (13) and ϵ tending to zero as T tends to infinity. Since $X_0^T \multimap\!\!- (M_1, M_2) \multimap\!\!- \hat{X}_0^T$ form a Markov chain, we further need

$$I(X_0^T; M_1, M_2) \ge T(R(D) - \epsilon). \tag{39}$$

Further, $M_1 \multimap\!\!- X_0^T \multimap\!\!- M_2$ also form a Markov chain, so

$$I(X_0^T; M_1, M_2) \le I(X_0^T; M_1) + I(X_0^T; M_2). \tag{40}$$

Next note $X_0^T \multimap\!\!- Y_0^T \multimap\!\!- M_1$ and the fact that X_0^T is the $\lambda/(\lambda+\mu)$-thinning of Y_0^T. Theorem 1 implies

$$I(X_0^T; M_1) \le \frac{\lambda}{\lambda+\mu} I(Y_0^T; M_1) \le \frac{\lambda}{\lambda+\mu} TR_1. \tag{41}$$

Similarly,

$$I(X_0^T; M_2) \le \frac{\lambda}{\lambda+\nu} TR_2. \tag{42}$$

Combining (13) and (39)–(42) proves the converse. ∎

V. Concluding Remarks

We have derived a simple strong DPI concerning the thinning operation on Poisson processes. Its application should not be limited to the few examples discussed in this paper.

Additionally, this work shows that the rate-distortion problem of covering Poisson processes studied in [6] can be a useful mathematical tool in various scenarios, where "covering" need not have an operational meaning in itself. Indeed, generating such a covering set not only approaches equality in the above strong DPI, but also constitutes the optimal solution to the hypothesis testing problem and the mutual information game that we considered.

References

[1] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[2] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and Bayesian networks," 2015.

[3] P. Harremoës, O. Johnson, and I. Kontoyiannis, "Thinning, entropy, and the law of thin numbers," *IEEE Trans. Inform. Theory*, vol. 56, no. 9, pp. 4228–4244, Sept. 2010.

[4] ——, "Thinning and the law of small numbers," in *Proc. IEEE Int. Symp. Inform. Theory*, Nice, France, June 24–29, 2007.

[5] Y. Yu, "Monotonic convergence in an information-theoretic law of small numbers," *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5412–5422, Dec. 2009.

[6] A. Lapidoth, A. Malär, and L. Wang, "Covering point patterns," *IEEE Trans. Inform. Theory*, vol. 61, no. 9, pp. 4521–4533, Sept. 2015.

[7] A. Mazumdar and L. Wang, "Covering arbitrary point patterns," in *Proc. 50th Allerton Conf. Comm., Contr. and Comp.*, Monticello, IL, October 1–5, 2012.

[8] R. S. Lipster and A. N. Shiryaev, *Statistics of Random Processes II: Applications*, 2nd ed. Springer Verlag, 2001.

[9] G. Last and A. Brandt, *Marked Point Processes on the Real Line: The Dynamic Approach*, ser. Probability and Its Applications. Springer Verlag, 1995.

[10] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inform. Theory*, vol. 32, no. 4, pp. 533–542, July 1986.

[11] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inform. Theory*, vol. 33, no. 6, pp. 759–772, Nov. 1987.

[12] J. R. Norris, *Markov Chains*. Cambridge University Press, 1997.

[13] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639–1667, June 2002.

[14] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[15] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 43, no. 5, pp. 1549–1559, Sept. 1997.

[16] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inform. Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.