Pointwise Relations between Information and Estimation in the Poisson Channel

Jiantao Jiao
Dept. of Electrical Engineering
Stanford University
Email: jiantao@stanford.edu

Kartik Venkat
Dept. of Electrical Engineering
Stanford University
Email: kvenkat@stanford.edu

Tsachy Weissman
Dept. of Electrical Engineering
Stanford University
Email: tsachy@stanford.edu

***Abstract*—Identities yielding optimal estimation interpretations for mutual information and relative entropy - paralleling those known for minimum mean squared estimation under additive Gaussian noise - were recently discovered for the Poisson channel by Atar and Weissman. We express these identities as equalities between expectations of the associated estimation and information theoretic random variables such as the actual estimation loss and the information density. By explicitly characterizing the relations between these random variables we show that they are related in much stronger pointwise senses that directly imply the known expectation identities while deepening our understanding of them. As an example for the nature of our results, consider the equality between the mutual information and the mean cumulative filtering loss of the optimal filter in continuous-time estimation. We show that the difference between the information density and the cumulative filtering loss is a martingale expressible as a stochastic integral. This explicit characterization not only directly recovers the previously known expectation relation, but allows to characterize other distributional properties of the random variables involved where some of the original objects of interest emerge in new and surprising roles. For example, we find that the increasing predictable part of the Doob-Meyer decomposition of the information density (which is a sub-martinagle) is nothing but the cumulative loss of the optimal filter.**

I. Introduction

Fundamental relations between measures of information and estimation-loss are providing crucial insights in both fields.

Relations between information and estimation for the Gaussian channel, date at least as far back as "de Bruijn's" identity by Stam [1] in 1959, which presented the derivative of a differential entropy as a Fisher information, the mean square score; and Duncan's theorem [2] in 1970 on the relation between filtering squared error and the input-output mutual information. A more recent line of work has reignited interest and research on the intimate relationship between mutual information and minimum mean squared error (MMSE) for signals corrupted by Gaussian noise. Barron [3] presented an integral version of the "de Bruijn's" identity, while imposing only a finite second moment constraint on the channel input. In [4], Guo et al. discovered the I-MMSE result and established its equivalence to "de Bruijn's" identity. In continuous-time [4], [5], this translates to a relation between the derivative of the mutual information w.r.t. signal-to-noise ratio (SNR), and the squared smoothing error. Together with Duncan's result, the authors of [4] established a simple expression for the causal MMSE at SNR level γ - as the average smoothing MMSE with SNR uniformly distributed between 0 and γ. These relations have further been extended to incorporate mismatch at the decoder, for both the scalar Gaussian channel by Verdú in [6], and for the continuous-time setting by Weissman in [7]. Interestingly, such precise "optimal estimation interpretations" for information-theoretic quantities are not unique to the Gaussian channel and mean squared loss.

We now turn our attention to the Poisson channel, which is central to our discussion in the current work. The Poisson channel is canonically used to describe direct detection optical communication. The channel input is a non-negative stochastic process $X^T = \{X_t, 0 \le t \le T\}$, and conditionally on X^T, the output Y^T is a non-homogeneous Poisson process with intensity function $\gamma \cdot X^T$. Here γ plays the role of the channel SNR. We refer to [8] for a discussion on the Poisson channel in communication theory, and to [9] for its applications in a broader context.

In [10], the authors proposed a natural estimation loss function under this channel model. $\ell : [0,\infty) \times [0,\infty) \to [0,\infty]$, is defined as

$$\ell(x,\hat{x}) \stackrel{\triangle}{=} x \log \frac{x}{\hat{x}} - x + \hat{x} \tag{1}$$

The loss function in (1) satisfies several properties that make it a very natural one for the Poisson channel [10, Lemma 2.2]. In particular, the minimizer of the average loss turns out to be the conditional expectation, as in the case of MMSE. Further, ℓ satisfies the scaling property, $\ell(\gamma x, \gamma \hat{x}) = \gamma\, \ell(x,\hat{x})$.

The main result of [10] is to establish, for the Poisson channel, a set of relations paralleling those from the Gaussian setting. The results state that Duncan's theorem [2], the I-MMSE relationship [4], [5], the relationship between relative entropy and cost of mismatch [6], [7], and the relation between causal and non-causal errors - hold verbatim, for the Poisson channel, upon replacing the squared error loss by the loss function in (1).

Recently, in [11], information-estimation relations for the Gaussian channel were unified, and presented as special instances of general pointwise identities involving more fundamental measures of information, and estimation loss. The current work can be viewed, in the same vein, as a step towards gaining a similar understanding of "pointwise" information-estimation relations, for the Poisson channel. Along with strengthening and unifying known identities, we establish connections between information theoretic quantities, and change of measure formulae for point processes - which have natural estimation theoretic interpretations of their own.

As an illustration of our results, let us first consider the relation between input-output mutual information and the causal minimum mean loss in estimation (CMLE) presented in [10], for the loss function defined in (1):

$$I(X^T; Y^T) = \gamma \cdot \mathsf{cmle}_{P,P}(\gamma), \tag{2}$$

where the processes X^T and Y^T are related via the Poisson channel, and $\mathsf{cmle}_{P,P}(\gamma)$ denotes the expected cumulative loss in *causal* filtering of X^T based on Y_γ^T when $X^T \sim P$ i.e.,

$$\mathsf{cmle}_{P,P}(\gamma) = \mathbb{E}_P \left[\int_0^T \ell(X_t, \mathbb{E}_P[X_t | Y_\gamma^t]) dt \right]. \tag{3}$$

We denote the input-output information density as

$$\imath(X^T; Y^T) \triangleq \log \frac{dP_{X^T, Y^T}}{dP_{X^T} dP_{Y^T}}, \tag{4}$$

and note that its mean is simply the mutual information between X^T and Y^T, i.e.,

$$\mathbb{E}[\, \imath(X^T; Y^T)\,] = I(X^T; Y^T). \tag{5}$$

The input-output information density $\imath(X^T; Y^T)$, plays an important role, among other applications, in characterizing the fundamental limits of communication in the finite blocklength regime [12].

Note that (2) can be equivalently stated as

$$\mathbb{E}\left[\mathcal{E}_T\right] = 0, \tag{6}$$

where $\mathcal{E}_T$ is defined as

$$\mathcal{E}_T \triangleq \imath(X^T; Y^T) - \gamma \int_0^T \ell(X_t, \mathbb{E}[X_t | Y_\gamma^t])\, dt. \tag{7}$$

What can we say about $\mathcal{E}_T$, beyond that fact that it has zero mean? One of our main contributions in this work is to present an explicit characterization of $\mathcal{E}_T$. We establish that the random process $\mathcal{E}_T$ (indexed by T) is in fact a stochastic integral, which by virtue of being a martingale and having zero mean directly recovers the identity in (2). In addition to providing us with a pointwise information-estimation relation, $\mathcal{E}_T$ emerges in the classical Doob-Meyer decomposition of the sub-martingale $\imath(X^T; Y^T)$. Thus, along with unifying existing identities between information and estimation, our results also shed light on the fundamental structure of the underlying processes defining these quantities.

The rest of this paper is organized as follows. In Section II, we describe the problem setting and terminology. In Section III, we present our main results for causal estimation. We also consider the presence of noiseless feedback, as well as the classical Doob-Meyer decomposition theorem for sub-martingales. In Section IV, we present pointwise relations for non-causal estimation. We conclude with a summary in Section V.

II. Preliminaries and Notation

Throughout, we use capital letters X, Y, Θ to denote random variables, and small letters x, y, θ to denote their specific realizations. X^T, Y^T, Θ^T denote random processes in the interval $[0, T]$. Our treatment of point processes invokes concepts that are standard in the literature on continuous-time stochastic processes, such as filtrations and martingales. We refer the reader to a standard reference (such as [13]) for definitions and properties.

A. The Poisson Channel

Throughout our discussion, the input to the Poisson channel will be denoted by the predictable stochastic process X^T. We assume the input process X^T is bounded. The channel output Y^T, conditioned on X^T, is a non-homogeneous Poisson process with intensity $\gamma \cdot X^T$. In a communication setting, the channel input can be a coded version of an underlying time-varying message Θ^T.

We now specify further conditions on the input process, so that it does not anticipate the message or the channel output. To this effect, we introduce the corresponding filtrations generated by the random processes.

Let $\mathcal{F}_t^\Theta$ be the filtration generated by Θ_s up to t, i.e.,

$$\mathcal{F}_t^\Theta = \sigma\{\Theta_s, 0 \le s \le t\}. \tag{8}$$

Analogously, $\mathcal{F}_t^Y$ is defined as the filtration generated by Y. up to t, i.e.,

$$\mathcal{F}_t^Y = \sigma\{Y_s, 0 \le s \le t\}. \tag{9}$$

With respect to the history

$$\mathcal{G}_t = \mathcal{F}_T^\Theta \vee \mathcal{F}_t^Y, \tag{10}$$

the point process Y_t has intensity X_t. Further, the channel input X_t is predictable according to the filtration $\mathcal{F}_t^\Theta \vee \mathcal{F}_t^Y$. This condition says two things. First, coding does not anticipate the message [since at each time t, X_t depends on the process Θ_t only through $(\Theta_s, 0 \le s \le t)$]. Second, non-anticipative feedback is allowed, i.e., at time t, X_t may depend on $(Y_s, 0 \le s < t)$.

The full probability law governing (Θ^T, X^T, Y^T) is denoted by P. Under probability law μ, Y^T is a Poisson process with constant intensity 1, and (Θ^T, X^T) are independent of the output Y^T. Further, under μ, we assume the distribution of Θ^T is identical to that under law P, i.e.,

$$P_{\Theta^t} \equiv \mu_{\Theta^t}. \tag{11}$$

The measure μ will be useful later, as a reference measure for the application of Girsanov-type change of measure theorems.

Having defined the probability basis, we now specify the law governing the processes in each of the underlying filtrations. Let us denote by $P_t, P_{\Theta^t}, P_{X^t}$, and P_{Y^t} the restrictions of P to $\mathcal{G}_t, \mathcal{F}_t^\Theta, \mathcal{F}_t^X$ and $\mathcal{F}_t^Y$, respectively. From [13], we have the following characterizations for the corresponding Radon-Nikodym derivatives.

Theorem 1 (Theorem T3 and R8, Chapter 6, [13]): The Radon-Nikodym derivatives $\frac{dP_t}{d\mu_t}$ and $\frac{dP_{Y^t}}{d\mu_{Y^t}}$ are well defined and given by:

$$\frac{dP_t}{d\mu_t} = \exp\left\{\int_0^t \log\gamma X_s dY_s + \int_0^t (1-\gamma X_s)ds\right\}, \quad (12)$$

$$\frac{dP_{Y^t}}{d\mu_{Y^t}} = \exp\left\{\int_0^t \log\gamma \hat{X}_s dY_s + \int_0^t (1-\gamma \hat{X}_s)ds\right\}, \quad (13)$$

where $\hat{X}_s$ is the predictable projection of X_s on $\mathcal{F}_s^Y$, i.e. the conditional mean estimate given the causal history of Y^s,

$$\hat{X}_s = \mathbb{E}[X_s|Y^s]. \quad (14)$$

Having introduced the probabilistic framework and tools, we now study the estimation loss.

B. Minimum Mean Loss Estimation in the Poisson Channel

Throughout this paper, we use $\ell(x,\hat{x})$ to denote the loss function defined in (1). As shown in [10], the conditional expectation $\mathbb{E}[X|\mathcal{H}]$ uniquely minimizes $\mathbb{E}[\ell(X,\hat{X})]$ among all random variables $\hat{X}$ measurable with respect to $\mathcal{H} \subset \sigma(X)$. The result carries over directly to stochastic processes.

For stochastic processes X^T and Y^T related via the Poisson channel with SNR level γ, we define the expected cumulative loss in *non-causal* filtering of X^T based on Y_γ^T when $X^T \sim P$, but the *non-causal* filter used is optimized for $X^T \sim Q$, as

$$\mathsf{mle}_{P,Q}(\gamma) \triangleq \mathbb{E}_P\left[\int_0^T \ell(X_t, \mathbb{E}_Q[X_t|Y_\gamma^T])dt\right]. \quad (15)$$

Similarly, we extend the definition in (3) to denote the mismatched expected cumulative loss in *causal* filtering as follows

$$\mathsf{cmle}_{P,Q}(\gamma) \triangleq \mathbb{E}_P\left[\int_0^T \ell(X_t, \mathbb{E}_Q[X_t|Y_\gamma^t])dt\right]. \quad (16)$$

Throughout this paper Y^T is shorthand for Y_γ^T when the SNR level is fixed and there is no ambiguity.

III. Main Results

In this section, we present our main results for causal estimation in the Poisson channel. We first discuss pointwise extensions of the I-CMLE theorem [10] (with and without feedback), and in the context of mismatched estimation. We then relate these characterizations with the canonical Doob–Meyer decomposition of sub-martingales.

A. Continuous-time Causal Estimation in the absence of Feedback

In [14, Chapter 19.5], [15], and [10], the following I-CMLE identity was developed:

$$I(X^T;Y^T) = \gamma \cdot \mathsf{cmle}_{P,P}(\gamma), \quad (17)$$

which is equivalently stated as

$$\mathbb{E}\left[\mathcal{E}_T\right] = 0, \quad (18)$$

for $\mathcal{E}_T$ defined in (7). The random process $\mathcal{E}_T$, indexed by T, can be interpreted as the "tracking error between the information density and the cumulative filtering loss".

Theorem 2: Let $\mathcal{E}_T$ denote the stochastic process defined in (7). Then

$$\mathcal{E}_T = \int_0^T \log\frac{X_s}{\hat{X}_s}(dY_s - \gamma X_s ds) \quad a.s. \quad (19)$$

Theorem 2 gives us a strikingly simple expression for the "tracking error" process $\mathcal{E}_T$. It is simply a stochastic integral with respect to the compensated Poisson process $Y_t - \int_0^t \gamma X_s ds$, which is a uniformly integrable martingale by Theorem 18.7 of [14] since we assume the input process is bounded, hence has mean zero. The stochastic integrals we obtain in the sequel are all uniformly integrable martingales by similar reasons. We do not prove Theorem 2 here, since it follows as a corollary of Theorem 3 discussed below.

B. Continuous Time Causal Estimation where Feedback may be Present

In [14, Chapter 19.5] and [10], it was shown that even when feedback is allowed, there exist similar relationships expressing the mutual information between the time-varying message Θ^T and the output process Y^T, as an integral of the minimum mean loss in causal estimation. Concretely, the relationship is given by:

$$I(\Theta^T;Y^T) = \gamma \cdot \mathsf{cmle}_{P,P}(\gamma), \quad (20)$$

which is equivalent to

$$\mathbb{E}\left[\mathcal{E}_T^{(F)}\right] = 0, \quad (21)$$

where

$$\mathcal{E}_T^{(F)} \triangleq \imath(\Theta^T;Y^T) - \gamma\int_0^T \ell(X_t, \mathbb{E}[X_t|Y_\gamma^t])dt. \quad (22)$$

A natural question is, how to explicitly characterize $\mathcal{E}_T^{(F)}$ when feedback may be present? Our next theorem provides the answer.

Theorem 3: Let $\mathcal{E}_T^{(F)}$ denote the stochastic process defined in (22). Then

$$\mathcal{E}_T^{(F)} = \int_0^T \log\frac{X_s}{\hat{X}_s}(dY_s - \gamma X_s ds) \quad a.s. \quad (23)$$

Proof: The main idea is to find an explicit expression for $\imath(\Theta^T;Y^T)$, via an appropriate change of measure.

$$\begin{aligned}
\imath(\Theta^T;Y^T) &= \log\frac{dP_{\Theta^T,Y^T}}{dP_{\Theta^T}dP_{Y^T}} && (24)\\
&= \log\frac{dP_{\Theta^T,Y^T}}{d\mu_{\Theta^T,Y^T}} - \log\frac{dP_{\Theta^T}dP_{Y^T}}{d\mu_{\Theta^T,Y^T}} && (25)\\
&= \log\frac{dP_{\Theta^T,Y^T}}{d\mu_{\Theta^T,Y^T}} - \log\frac{d\mu_{\Theta^T}dP_{Y^T}}{d\mu_{\Theta^T,Y^T}} && (26)\\
&= \log\frac{dP_{\Theta^T,Y^T}}{d\mu_{\Theta^T,Y^T}} - \log\frac{dP_{Y^T}}{d\mu_{Y^T}}, && (27)
\end{aligned}$$

where (26) follows from (11), and (27) is due to the fact that under law μ_{Θ^T,Y^T}, Θ^T and Y^T are independent.

Plugging (12) into (27), we obtain

$$\imath(\Theta^T;Y^T) = \int_0^T \log \frac{X_s}{\hat{X}_s} dY_s - \gamma \int_0^T (X_s - \hat{X}_s) ds. \quad (28)$$

The cumulative minimum mean loss for causal estimation $\gamma \cdot \mathsf{cmle}_{P,P}(\gamma)$ can be expressed as

$$\gamma \cdot \mathsf{cmle}_{P,P}(\gamma) = \int_0^T \mathbb{E}\left[\gamma \left(X_s \log \frac{X_s}{\hat{X}_s} - X_s + \hat{X}_s \right)\right] ds. \quad (29)$$

Since $\mathcal{E}_T^{(F)}$ is defined as the difference of the left-hand sides of (28) and (29), we have

$$\mathcal{E}_T^{(F)} = \int_0^T \log \frac{X_s}{\hat{X}_s} (dY_s - \gamma X_s ds). \quad (30)$$

■

C. Continuous Time Causal Mismatched Estimation

In [10], Atar and Weissman showed that if P and Q are two admissible probability measures, then

$$D(P_{Y^T} \| Q_{Y^T}) = \gamma \cdot [\mathsf{cmle}_{P,Q}(\gamma) - \mathsf{cmle}_{P,P}(\gamma)], \quad (31)$$

where the left-hand side is the Kullback–Leibler divergence between the two output laws P_{Y^T} and Q_{Y^T}.

Here we present the pointwise version of (31). Define the relative information between two probability measures P, Q, where $P \ll Q$, as

$$\imath(P\|Q) \triangleq \log \frac{dP}{dQ}, \quad (32)$$

and note that

$$\mathbb{E}_P[\imath(P_{Y^T} \| Q_{Y^T})] = D(P_{Y^T} \| Q_{Y^T}). \quad (33)$$

Let $\mathcal{M}_T$ denote the mismatched tracking error:

$$\imath(P_{Y^T}\|Q_{Y^T}) - \gamma \left(\int_0^T \ell(X_s, \hat{X}_s^Q) ds - \int_0^T \ell(X_s, \hat{X}_s^P) ds \right), \quad (34)$$

where $\hat{X}_s^P$ is the predictable projection of X_s on $\mathcal{F}_s^Y$ under law P, and $\hat{X}_s^Q$ is the predictable projection of X_s on $\mathcal{F}_s^Y$ under law Q.

Theorem 4: If P and Q are two probability measures, then

$$\mathcal{M}_T = \int_0^T \log \frac{\hat{X}_s^P}{\hat{X}_s^Q} (dY_s - \gamma X_s ds) \quad P - a.s. \quad (35)$$

Proof: The proof of Theorem 4 follows by applying the change of measure formulae (12) to $\frac{dP_{Y^T}}{d\mu_{Y^T}}$ and $\frac{dQ_{Y^T}}{d\mu_{Y^T}}$, followed by standard algebraic manipulations. ■

Thus far, we have obtained pointwise extensions to identities relating mutual information and causal mean loss for the Poisson channel. We observed that the corresponding tracking error, is always expressible as a stochastic integral, and is in fact a martingale. We now show that such a pointwise result allows us to prove a stronger conditional result, relating the information density and cumulative filtering loss.

Theorem 5: When there is *no* feedback, the conditional expectation of $\mathcal{E}_T$, conditioned on the message and input history $\mathcal{F}_T^{\Theta,X}$, is zero almost surely:

$$\mathbb{E}[\mathcal{E}_T | \mathcal{F}_T^{\Theta,X}] = 0 \quad P - a.s. \quad (36)$$

Proof: This statement follows by a direct application of the pointwise mismatch result in Theorem 4. In particular, we choose P to be the law under which the underlying message and channel input is deterministic, and let Q denote the actual law governing them. We skip the details for brevity. ■

Remark 1: Note that Theorem 5 implies that the "tracking error" $\mathcal{E}_T$ not only has mean zero, which is a natural consequence of being a stochastic integral– but also has conditional expectation zero, when it is conditioned on the entire history of the message and the input. This is a stronger statement, and in particular directly implies the zero-mean result in (2).

D. Doob–Meyer Decomposition Interpretations

Reviewing the results in Section III-A,B,C, we find an underlying commonality: we expressed a process as a sum of two parts - one being a martingale, and the other having a cumulative loss interpretation. In fact, what we have obtained is the Doob–Meyer decomposition for a special class of sub-martingales, namely the information density, and the relative information.

Recall the Doob–Meyer decomposition Theorem (cf., Chap. 3.3 of [16]), which states that under mild regularity conditions, a right-continuous sub-martingale Z_t has the unique decomposition

$$Z_t = A_t + M_t, \quad (37)$$

as the sum of an increasing, predictable process A_t, and a martingale M_t. In general, it is non-trivial to obtain the Doob–Meyer decomposition for continuous-time processes.

It is straightforward to check that $\imath(\Theta^T; Y^T)$ and $\imath(P_{Y^T}\|Q_{Y^T})$ are sub-martingales indexed by T, via the chain rule for mutual information and relative entropy, respectively. Note that results of Section III-B subsume their counterparts in Section III-A. Now, equations (23) and (35) can be re-written as follows:

$$\begin{aligned}\imath(\Theta^T;Y^T) &= \gamma \int_0^T \ell(X_t, \mathbb{E}[X_t|Y_\gamma^t]) dt \\ &\quad + \int_0^T \log \frac{X_s}{\hat{X}_s} (dY_s - \gamma X_s ds), \quad (38)\\ \imath(P_{Y^T}\|Q_{Y^T}) &= \gamma \int_0^T \ell(\hat{X}_s^P, \hat{X}_s^Q) ds \\ &\quad + \int_0^T \log \frac{\hat{X}_s^P}{\hat{X}_s^Q} (dY_s - \gamma \hat{X}_s^P ds). \quad (39)\end{aligned}$$

The first terms on the right hand sides of these equalities are obviously non-decreasing, while the second terms are martingales, and hence we've explicitly obtained the Doob-Meyer decomposition of the quantities on the left hand sides. Thus, it is quite satisfying to see the pointwise relations in (38)-(39) emerge as answers to a fundamental question pertaining to an interesting class of sub-martingales having strong connections with information theory. Further, the predictable components of these decomposition have optimal estimation loss interpretations. Thus, the pointwise "tracking error" emerges as a bridge between the underlying objects in information and estimation for the Poisson channel.

Remark 2: We note in passing that the Doob–Meyer decomposition interpretations of "relations between information and estimation" discussed above have natural analogues in the Gaussian channel setting of [11]. We defer the details to a more comprehensive version of this manuscript, currently under preparation.

IV. Non-Causal Estimation

Having studied the pointwise extensions for causal estimation in Section III, we now proceed to unravel the D-MLE theorem in [10], which relates the relative entropy of the output laws to the cost of mismatch in non-causal estimation:

$$D(P_{Y^T}\|Q_{Y^T}) = \int_0^\gamma [\mathsf{mle}_{P,Q}(\alpha) - \mathsf{mle}_{P,P}(\alpha)]. \tag{40}$$

We now present the pointwise analogue of (40). Define $\mathcal{M}_T^{(NC)}$ to be the difference between the relative information and the integral of the cumulative cost of mismatch, i.e.,

$$\begin{aligned} &\mathcal{M}_T^{(NC)} \\ &\triangleq \imath(P_{Y^T}\|Q_{Y^T}) \\ &\quad - \int_0^\gamma \int_0^T \ell(X_t, \mathbb{E}_Q[X_t|Y_\alpha^T]) - \ell(X_t, \mathbb{E}_P[X_t|Y_\alpha^T])\, dt\, d\alpha. \end{aligned} \tag{41}$$

Our next result provides an explicit characterization of $\mathcal{M}_T^{(NC)}$:

Theorem 6: If P and Q are two probability measures, then P-a.s, we have

$$\mathcal{M}_T^{(NC)} = \int_0^\gamma \int_0^T \log \frac{\mathbb{E}_P[X_s|Y_\alpha^T]}{\mathbb{E}_Q[X_s|Y_\alpha^T]} (dY_{s,\alpha} - X_s\, ds\, d\alpha). \tag{42}$$

Here $Y_{s,\alpha}$ denotes the output of the Poisson channel at time s and SNR level α.

The techniques involved in proving Theorem 6 are similar to those presented in Section III, but applied to a new filtration defined with respect to the SNR level. For the setting and point process construction we refer to Section VII-A of [10].

Compared to $\mathcal{E}_T, \mathcal{E}_T^{(F)}, \mathcal{M}_T$ – even though $\mathcal{M}_T^{(NC)}$ has zero mean, thereby implying (40), it is not adapted to the filtration indexed by T. The reason behind this is quite simple: since the definition of $\mathcal{M}_T^{(NC)}$ is based on non-causal filtering loss, it clearly anticipates the output process.

Another observation is that, following similar techniques to proving Theorem 5, Theorem 6 implies the pointwise analogue of the I-MMLE relationship [10], which states that

$$I(X^T;Y^T) = \int_0^\gamma \mathsf{mle}_{P,P}(\alpha)\, d\alpha. \tag{43}$$

Note however, that in the non-causal framework we assume there is *no* feedback. Indeed, it is easy to construct examples involving the presence of feedback, where (43) no longer holds.

V. Conclusions

We present pointwise analogues of information-estimation identities in the continuous-time Poisson channel. We explicitly characterize the tracking error between the information density (relative information) and cumulative loss (cost of mismatch), and show that under causal estimation, this quantity is a stochastic integral with respect to the compensated output Poisson process. This is directly analogous to the pointwise information-estimation relations for the Gaussian channel, discovered recently in [11], where the tracking error emerged as an Itô integral. Our results unify, and develop new insights into the information density process, as well as improve the understanding of relations between information and estimation for a classical channel.

Acknowledgment

This research is partially supported by two Stanford Graduate Fellowships and NSF grant CCF-0939370. Jiantao Jiao would like to thank Xiaodong Li for helpful discussions on the Doob–Meyer decomposition.

References

[1] A. J. Stam, "Some inequalities satisfied by the quantities of information of fisher and shannon," *Information and Control*, vol. 2, pp. 101–112, 1959.

[2] T. Duncan, "On the calculation of mutual information," *SIAM J. Appl. Math.*, vol. 19, pp. 215–220, 1970.

[3] A. Barron, "Entropy and the central limit theorem," *The Annals of probability*, pp. 336–342, 1986.

[4] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1283, 2005.

[5] M. Zakai, "On mutual information, likelihood ratios, and estimation error for the additive gaussian channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3017–3024, 2005.

[6] S. Verdú, "Mismatched estimation and relative entropy," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3712–3720, 2010.

[7] T. Weissman, "The relationship between causal and noncausal mismatched estimation in continuous-time awgn channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4256–4273, 2010.

[8] S. Verdú, "Poisson communication theory," *International Technion Communication Day in Honor of Israel Bar-David*, 1999.

[9] J. Grandell, *Mixed Poisson Processes*. Chapman & Hall, 1997.

[10] R. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the poisson channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1302–1318, 2012.

[11] K. Venkat and T. Weissman, "Pointwise relations between information and estimation in gaussian noise," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6264–6281, 2012.

[12] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, 2010.

[13] P. Brémaud, *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, 1981.

[14] R. Liptser and A. N. Shiryaev, *Statistics of Random Processes II: Applications, 2nd ed.* Springer-Verlag, 2001.

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

[16] R. Liptser and A. N. Shiryaev, *Statistics of Random Processes I: General Theory, 2nd ed.* Springer-Verlag, 2000.