Mutual Information and Conditional Mean Estimation in Poisson Channels

Dongning Guo
Electrical & Computer Engineering
Northwestern University
Evanston, IL 60208, USA
Email: dGuo@ece.northwestern.edu

Sergio Verdú
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: verdu@princeton.edu

Shlomo Shamai
Dept. of Electrical Engineering
Technion-Israel Institute of Technology
32000 Haifa, Israel
Email: sshlomo@ee.technion.ac.il

Abstract — **Following the recent discovery of new connections between information and estimation in Gaussian channels, this paper reports parallel results in the Poisson regime. Both scalar and continuous-time Poisson channels are considered. It is found that, regardless of the statistics of the input, the derivative of the input-output mutual information with respect to the dark current can be expressed in the expected difference between the logarithm of the input and the logarithm of its conditional mean estimate (noncausal in case of continuous-time). The same is true for the derivative with respect to input scaling, but with the logarithmic function replaced by $x \log x$.**

I. Introduction

Some fundamental relationships between input-output mutual information and conditional mean estimation have recently been discovered for additive Gaussian noise channels with arbitrary input [1, 2]. This paper develops parallel results for Poisson channels, the output of which is a Poisson random variable or process conditioned on the input. Such a channel occurs in direct-detection optical communication systems, in which incident radiation is intercepted by means of photon-sensitive devices to result a doubly Poisson point process, the rate of which is typically the intensity of the incident radiation plus a "dark current".

Reference [3] presents a review of major developments of communications theory in the Poisson regime. Signal detection in Poisson channels has been studied since 1960s and the general "Poisson matched filter" which yields optimal detection was found in 1969 [4]. Stochastic integration with Poisson point process observations was developed for various filtering problems in 1970s (e.g., [5]). In particular, the likelihood ratio for signal detection has been found as such an stochastic integral. Using martingale theory, the likelihood ratio for detection based solely on the observation is shown to be an "estimator-correlator" type of formula (e.g., [6], cf. [7]). Furthermore, the mutual information, which is the average of the log-likelihood ratio, can be expressed using the Liptser-Shiryaev formula as an integral of the expectation of the difference between the logarithm of the input and the logarithm of the causal conditional mean estimate [8, 5]. The capacity of Poisson channels under peak- and average-power limit was found [9, 10], where infinite bandwidth signaling is required to achieve the capacity. The reliability function at all rates below capacity is also known [11]. Bounds on the capacity are found under bandwidth-like constraints [12]. More recently, the high signal-to-noise ratio (SNR) asymptotic capacity of a peak and average power limited discrete-time Poisson channel is derived in [13] making use of an observation that the entropy of the output is lower bounded by the differential entropy of the input. Poisson multiple-access channels, Poisson broadcast channels, and Poisson arbitrarily varying channels are studied in references [14], [15] and [16] respectively.

Equipped with known likelihood ratios and stochastic integration techniques, this paper studies the input-output mutual information of Poisson channels in either scalar or continuous-time settings. A key result in this paper is that, regardless of the statistics of the input, the derivative of the input-output mutual information of a Poisson channel with respect to (the intensity of) the dark current is equal to the average error between the logarithm of the actual input and the logarithm of its conditional mean estimate (noncausal in case of continuous-time). In other words, the mutual information can be expressed as an integral of such an error as a function of the dark current. Unlike in the Gaussian channel case, scaling of the input cannot be absorbed into the dark current. The derivative of the mutual information with respect to the scaling can also be expressed in some error associated with the conditional mean estimate. In the continuous-time setting, together with the Liptser-Shiryaev formula [8], our results complete the triangle relationship of the mutual information and causal and noncausal conditional mean estimation.

The problem of Poisson channels studied in this paper is more involved than its counterpart in the Gaussian regime. Some of the difficulties are inherent to Poisson channels: 1) The dark current and scaling cannot be consolidated into one parameter; and 2) The channel conditioned on a degraded version of the output is no longer Poisson. Other difficulties are due to the fact that less is known about Poisson channels. For example, the hybrid continuous-discrete nature of the input-output pair ap-

pears harder to deal with; simple closed form expressions for conditional mean estimate and mutual information are known for few special cases; and little is known about "natural" metrics for measuring estimation errors.

In a wider context, this work reveals new connections between information theory and estimation theory as long as the input and output are related through conditional Poisson probability laws. The likelihood ratio plays a central role in linking mutual information and conditional mean estimation. The results allow certain information measures to be expressed using solely estimation errors and vice versa.

The rest of the paper is organized as follows. Section II studies scalar Poisson channels. Section III studies their continuous-time counterparts, and the paper is concluded in Section IV.

II. Scalar Poisson Channel

In a canonical Poisson random transformation, given a positive (≥ 0) input random variable X, the output Y is a Poisson random variable with its mean value equal to X:

$$P_{Y|X}(y|x) = \frac{1}{y!}x^y e^{-x}, \quad y = 0, 1, \dots \quad (1)$$

Such a transformation defines a canonical Poisson channel in the natural way. A general Poisson channel is similarly defined by a transformation whose output is a Poisson random variable conditioned on the input X with its mean equal to $(\alpha X + \lambda)$ with $\alpha, \lambda \geq 0$, i.e., X scaled by α plus a dark current λ. Given an arbitrary input random variable X, let $\mathcal{P}(\alpha X + \lambda)$ denote a conditionally Poisson random variable with its mean value equal to $(\alpha X + \lambda)$. This setting has a direct counterpart in the Gaussian regime where α is the scaling and λ corresponds to the Gaussian noise variance. Note that in the Gaussian case the scaling and the noise consolidate to a single degree of freedom—the SNR, whereas in the Poisson case they do not.

For future convenience, define an angle bracket operator for an arbitrary random variable X as the conditional mean estimate given $\mathcal{P}(X)$:

$$\langle X \rangle = \mathsf{E}\left\{ X \mid \mathcal{P}(X) \right\}$$

Note that $\langle X \rangle$ denotes in fact a class of random variables (depending on the choice of $Y = \mathcal{P}(X)$). For all purposes it suffices to consider an arbitrary version of it since we will always consider the expected value of some function of $\langle X \rangle$. It is important to note that $\langle \cdot \rangle$ is far from a linear operator: In general $\langle \alpha X + \lambda \rangle$ has different probability law from that of $\alpha \langle X \rangle + \lambda$ for all $\alpha \neq 0, 1$ and $\lambda > 0$.

Using the "incremental channel" device introduced in [2], one can study the change in mutual information due to an extra infinitesimal dark current and arrive at the following result.[1]

Lemma 1 *For every P_X with $\mathsf{E}\log X < \infty$, and $\lambda \to 0^+$,*

$$I\left(X; \mathcal{P}(X + \lambda)\right) - I\left(X; \mathcal{P}(X)\right)$$
$$= \lambda\,\mathsf{E}\left\{\log X - \log\langle X\rangle\right\} + o(\lambda). \quad (2)$$

Proof: See Appendix A. ∎

Lemma 1 essentially states that the decrease in mutual information due to an infinitesimal dark current is equal to the dark current times the mean error between the logarithm of the actual input and that of its conditional mean estimate (or, the expected value of the logarithm of the ratio of the actual input and its estimate). As a sanity check, the expectation in (2) is always negative due to Jensen's inequality; so is the derivative in (2) since adding dark current causes degradation. For a general Poisson channel, the following is true.

Theorem 1 *For every P_X with $\mathsf{E}\log X < \infty$ and every $\alpha, \lambda \geq 0$,*

$$\frac{\partial}{\partial\lambda}I\left(X; \mathcal{P}(\alpha X + \lambda)\right) = \mathsf{E}\left\{\log(\alpha X + \lambda) - \log\langle\alpha X + \lambda\rangle\right\}.$$

Proof: Treating $(\alpha X + \lambda)$ as a new random variable X' and applying Lemma 1 prove the theorem. ∎

Note that Theorem 1 resembles the central theorem in [1, 2] for Gaussian channels, which states that the derivative of the mutual information with respect to the SNR is equal to half the minimum mean-square error achieved by conditional mean estimation (the derivative with respect to the scaling or the noise level is immediate). The differential variable in the Poisson channel problem is the dark current, and the approximation quality is measured by the average error between the logarithm of the actual input and that of its conditional mean estimate.

Unlike in Gaussian channels, scaling of the input cannot be absorbed into the additive dark current in a Poisson channel. Taking both scaling and dark current into account, we have the following result.

Lemma 2 *For every P_X with $\mathsf{E}\left\{(X + 1)\log X\right\} < \infty$, and every $\delta \to 0$ and $\lambda \to 0^+$,*

$$I(X; \mathcal{P}((1+\delta)X + \lambda)) - I(X; \mathcal{P}(X))$$
$$= \lambda\,\mathsf{E}\left\{\log X - \log\langle X\rangle\right\} + o(\lambda)$$
$$+ \delta\,\mathsf{E}\left\{X\log X - \langle X\rangle\log\langle X\rangle\right\} + o(\delta).$$

Proof: See Appendix B. ∎

For a general Poisson channel, the derivative of the mutual information with respect to the scaling can also be expressed in some conditional mean estimation error.

Theorem 2 *For every P_X with $\mathsf{E}\left\{X\log X\right\} < \infty$, and every $\alpha > 0$ and $\lambda \geq 0$,*

$$\frac{\partial}{\partial\alpha}I(X; \mathcal{P}(\alpha X + \lambda)) = \mathsf{E}\Big\{X\log(\alpha X + \lambda)$$
$$- \frac{1}{\alpha}\left(\langle\alpha X + \lambda\rangle - \lambda\right)\log\langle\alpha X + \lambda\rangle\Big\}. \quad (3)$$

Proof: Note that for $\alpha > 0$,

$$I(X; \mathcal{P}((\alpha + \epsilon)X + \lambda)) - I(X; \mathcal{P}(\alpha X + \lambda))$$
$$= I(X; \mathcal{P}((1 + \epsilon/\alpha)(\alpha X + \lambda) - \epsilon\lambda/\alpha)) \quad (4)$$
$$- I(X; \mathcal{P}(\alpha X + \lambda))$$

For $\epsilon \to 0^-$, applying Lemma 2 to the right hand side of (4) establishes (3). The same is true for $\epsilon \to 0^+$ by the smoothness of the mutual informations with respect to scaling. ∎

The expectation in (3) is positive due to Jensen's inequality since $(x - \lambda) \log x$ is a convex function. Note that (3) does not apply directly to the special case of $\alpha = 0$, which describes the mutual information that corresponds to a very small input. In particular, at $\alpha = 0$ but $\lambda > 0$, the derivative of $I(X; \mathcal{P}(\alpha X + \lambda))$ with respect to α is found to be 0 by taking the limit $\alpha \to 0^+$. At the point $\alpha = \lambda = 0$, the derivative can be obtained from Theorem 2 by noting that

$$\lim_{\alpha \to 0} \frac{1}{\alpha} \langle \alpha X \rangle = \mathsf{E} X.$$

Therefore,

Corollary 1 *For every P_X with $\mathsf{E}\{X \log X\} < \infty$ and every $\alpha \to 0$,*

$$I(X; \mathcal{P}(\alpha X)) = \alpha \mathsf{E}\{X \log X - \mathsf{E} X \log \mathsf{E} X\} + o(\alpha).$$

It is interesting to note the similarity of the above results to those under Gaussian channels [2].

Theorems 1–2 and Corollary 1 give the two directional derivatives of the mutual information for all $\alpha, \lambda \geq 0$. Thus the Taylor expansion of the mutual information to the first order in the scaling and dark current is known at the vicinity of every (α, λ).

Clearly, the mutual information can be regarded as an integral of the estimation errors as a function of the dark current. For example,

Corollary 2 *For every P_X with $\mathsf{E} \log X < \infty$,*

$$I(X; \mathcal{P}(X)) = \int_0^\infty \mathsf{E}\{\log \langle X + \xi \rangle - \log(X + \xi)\} \, d\xi.$$

Alternatively, the mutual information can be expressed as an integral of the estimation errors as a function of the scaling.

Corollary 3 *For every P_X with $\mathsf{E}\{X \log X\} < \infty$,*

$$I(X; \mathcal{P}(X)) = \int_0^1 \mathsf{E}\left\{X \log(\alpha X) - \frac{\langle \alpha X \rangle}{\alpha} \log \langle \alpha X \rangle\right\} d\alpha.$$

One can regard the mutual information $I(X; \mathcal{P}(\alpha X + \lambda))$ as a potential field on the 2-dimensional plain $\{(\alpha, \lambda) \,|\, \alpha, \lambda \geq 0\}$. It is clear that the mutual information vanishes for $\alpha \to 0$ or $\lambda \to \infty$. Thus the mutual information at any (α, λ) pair can be regarded as a path integral of some estimation errors from any point $(\alpha_0, \lambda_0) \in \{(a, b) \,|\, a = 0 \text{ or } b = \infty\}$ to the point (α, λ).

Entropy also admits a similar expression [17].

III. CONTINUOUS-TIME POISSON CHANNELS

A canonical continuous-time Poisson channel is described as the following. Given a positive input process $\{X_t\}$ during a given time interval $[0, T]$, the output is a realization of a Poisson point process $\{Y_t\}$ whose time-varying expectation at any time t is equal to the integral of the "rate function" X_t. Precisely, for all $0 \leq t \leq s \leq T$,

$$\mathsf{P}\{Y_s - Y_t = k \,|\, \{X_t\}\} = \frac{1}{k!}\Lambda^k e^{-\Lambda}$$

where

$$\Lambda = \int_t^s X_\xi \, d\xi.$$

A more general Poisson channel can be regarded as the canonical channel with its input replaced by the rate function $\alpha X_t + \lambda$. Let its output at time t be denoted by $\mathcal{P}_t(\alpha X_t + \lambda)$, and the output process in the interval $[r, s]$ by $\mathcal{P}_r^s(\alpha X_t + \lambda)$. Also, define the conditional mean operator for an arbitrary process $\{X_t\}$ as

$$\langle X_t \rangle_s = \mathsf{E}\{X_t \,|\, \mathcal{P}_0^s(X_t)\}.$$

Note that the superscript s dictates the amount of observation available to the conditional mean estimator. In particular, $\langle X_t \rangle_t$ is referred to as causal (filtering) conditional mean estimate, and $\langle X_t \rangle_T$ is referred to as noncausal (smoothing) conditional mean estimate.

The input-output mutual information is related to the optimal causal estimate of the input via the following theorem, which takes a similar form as Duncan's result for Gaussian channels [18].

Theorem 3 (Liptser and Shiryaev [8]) *For every input with $\mathsf{E}\int_0^T X_t \log X_t \, dt < \infty$,*

$$I\left(X_0^T; \mathcal{P}_0^T(X_t)\right)$$
$$= \int_0^T \mathsf{E}\{X_t \log X_t - \langle X_t \rangle_t \log \langle X_t \rangle_t\} \, dt. \quad (5)$$

By virtue of the infinite divisibility of Poisson distributions, the "incremental channel" idea developed in [19] can be applied to study the decrease in mutual information due to an increase in dark current. The derivative of the mutual information with respect to the dark current is thus found.

Theorem 4 *For every input with $\mathsf{E}\int_0^T \log X_t \, dt < \infty$,*

$$\frac{\partial}{\partial\lambda} I\left(X_0^T; \mathcal{P}_0^T(\alpha X_t + \lambda)\right)$$
$$= \int_0^T \mathsf{E}\{\log(\alpha X_t + \lambda) - \log \langle \alpha X_t + \lambda \rangle_T\} \, dt.$$

Theorem 4 can also be proved by examining the likelihood ratio as in the proof of Lemma 1 (Appendix A). Note that since probability density functions no longer exist in the continuous-time setting, one has to resort to

the Radon-Nikodym derivative. Consider two hypotheses. Let P_1 denote the probability measure under input $\{X_t\}$, and P_0 denote the probability measure under zero input, where the dark current is λ_0 in both cases. The following Radon-Nikodym derivative [5, p. 180] between the two probability measures is the key:

$$\log \frac{dP_1^{XY}}{dP_0^{XY}} = \int_0^T \log\left(1 + \frac{X_t}{\lambda_0}\right) \, dY_t - \int_0^T X_t \, dt. \quad (6)$$

From (6) one can also derive the Radon-Nikodym derivative when only the observation $\{Y_t\}$ is accessible only. In fact the "estimator-correlator" principle found in Gaussian channels is also relevant here. That is, the resulting log-likelihood ratio $\log dP_1^Y / dP_0^Y$ is given by (6) only with X_t replaced by the causal estimate $\mathsf{E}\{X_t \mid Y_0^t\}$. Note that expectation in this case is in general a stochastic integral with respect to a counting process. Detailed proof of all continuous-time results is omitted. Interested readers are referred to [17].

Similar to that under scalar channels in Section II, the derivative with respect to scaling can be obtained from known likelihood ratios.

Theorem 5 *For every input with* $\mathsf{E}\int_0^T X_t \log X_t \, dt < \infty$,

$$\frac{\partial}{\partial \alpha} I\left(X_0^T; \mathcal{P}_0^T(\alpha X_t + \lambda)\right) = \int_0^T \mathsf{E}\Bigg\{ X_t \log(\alpha X_t + \lambda)$$
$$- \frac{1}{\alpha}\left(\langle \alpha X_t + \lambda\rangle_T - \lambda\right) \log \langle \alpha X_t + \lambda\rangle_T \Bigg\} \, dt.$$

Proof: See [17]. ∎

As a counterpart to Corollary 1, the mutual information under a small scaling is immediate.

Corollary 4 *For every input with* $\mathsf{E}\int_0^T X_t \log X_t \, dt < \infty$,

$$I\left(X_0^T; \mathcal{P}_0^T(\alpha X_t)\right)$$
$$= \alpha \int_0^T \mathsf{E}\{X_t \log X_t - \mathsf{E}X_t \log \mathsf{E}X_t\} \, dt + o(\alpha).$$

In fact, the Liptser-Shiryaev formula (5) admits a new intuitive proof using Corollary 4 and the incremental channel idea. The mutual information due to an infinitesimal extra observation time interval $(t, t + \alpha)$ is equal to the conditional mutual information of the same Poisson channel during the extra time interval given the past observation. By expanding the small interval to unit length, this conditional mutual information can be regarded as the mutual information of a channel with diluted input with a scaling of α, which, by Corollary 4, is obtained essentially as

$$\alpha \int_0^1 \mathsf{E}\{X_{t+\alpha s} \log X_{t+\alpha s} - \langle X_{t+\alpha s}\rangle_t \log \langle X_{t+\alpha s}\rangle_t\} \, ds.$$

Theorem 3 is then established by continuity assuming that $\{X_t\}$ is continuous:

$$\lim_{\alpha \to 0} \langle X_{t+\alpha s}\rangle_t = \langle X_t\rangle_t.$$

The mutual information as a function of the scaling and dark current also allows a Taylor expansion at the vicinity of every (α, λ). Informed with the two derivatives with respect to the scaling and dark current respectively, one can regard the mutual information as a potential field and calculate $I\left(X_0^T; \mathcal{P}_0^T(\alpha X_t + \lambda)\right)$ as a path integral. The continuous-time counterparts of Corollaries 2 and 3 are immediate from Theorems 4 and 5, and are omitted here (see [17]).

Combining the Liptser-Shiryaev formula and the new results that link the mutual information and noncausal estimate, one establishes the following curious relationships between causal and noncausal estimation errors:

$$I\left(X_0^T; \mathcal{P}_0^T(X_t)\right)$$
$$= \int_0^T \mathsf{E}\{X_t \log X_t - \langle X_t\rangle_t \log \langle X_t\rangle_t\} \, dt$$
$$= \int_0^\infty \int_0^T \mathsf{E}\{\log(X_t + \lambda) - \log \langle X_t + \lambda\rangle_T\} \, dt \, d\lambda$$
$$= \int_0^1 \int_0^T \mathsf{E}\Bigg\{ X_t \log(\alpha X_t)$$
$$- \frac{\langle \alpha X_t\rangle_T}{\alpha} \log \langle \alpha X_t\rangle_T \Bigg\} \, dt \, d\alpha.$$

Note that the results under scalar Poisson channels in Section II can be obtained by considering the special case of time-invariant input in the continuous-time setting. For example, the increase of the mutual information due to the scaling $(1 + \delta)$ is an outcome of the Liptser-Shiryaev formula (Theorem 3). Let $X_t \equiv X$. By (5),

$$\frac{d}{dt} I(X; Y_0^t) = \mathsf{E}\{X \log X - \langle X\rangle_t \log \langle X\rangle_t\}$$

where $\langle X\rangle_t = \mathsf{E}\{X \mid Y_0^t\}$. Clearly,

$$I\left(X; \mathcal{P}_0^1((1+\delta)X)\right) - I\left(X; \mathcal{P}_0^1(X)\right)$$
$$= \delta \mathsf{E}\{X \log X - \langle X\rangle \log \langle X\rangle\} + o(\delta).$$

The desired result is obtained once we note that increasing the observation time from 1 to $1 + \delta$ is equivalent to keeping the observation time to $[0, 1]$ but scaling the intensity of the input by $1 + \delta$.

IV. Conclusion

New relationships between the input-output mutual information and conditional mean estimation in Poisson channels have been identified in this paper. By studying the change in mutual information due to an infinitesimal increase in dark current and/or input scaling, the derivatives of the mutual information with respect to dark current and scaling are expressed in the expected difference

in a function ($\log x$ or $x \log x$) evaluated at the actual input and the same function evaluated at its conditional mean estimate. The general relationships hold for both scalar and continuous-time Poisson channels. In particular, the results for scalar channels can be obtained from their respective continuous-time counterparts by considering time-invariant inputs. The results extend connections between estimation and information theory to the Poisson regime. It would be very interesting to see if these results lead to new applications.

Underlying the analysis and results in both [2] and this paper are common properties of Gaussian and Poisson distributions, namely, 1) infinite divisibility of Gaussian and Poisson distributions; and 2) independent increments of Gaussian and Poisson processes. In fact, the entire class of processes with independent increments is known as Lévy processes and can be characterized by not much more than a mixture of Wiener and Poisson processes [20]. It is even speculated in [2] that information and estimation satisfy similar relationships as long as the output is a Lévy process conditioned on the input.

A. Proof of Lemma 1

With some abuse of notation, we use $\mathcal{P}(\alpha X + \lambda)$ and $Y^{\alpha,\lambda}$ interchangeably, i.e., $Y^{\alpha,\lambda}$ stands for a Poisson random variable conditioned on X with its mean value equal to $\alpha X + \lambda$.

Proof: [Lemma 1] Since unit scaling is assumed, let the notation be shortened: $Y^0 = Y^{1,0}$ and $Y^\lambda = Y^{1,\lambda}$. Then,

$$I(X;Y^0) - I(X;Y^\lambda) = \mathsf{E}\left\{\log L(X,Y^0,Y^\lambda)\right\} \quad (7)$$

where the expectation is over the joint probability measure of X, Y^0 and Y^λ, and the likelihood ratio is

$$L(x,y^0,y^\lambda) = \frac{P_{Y^0|X}(y^0|x) P_{Y^\lambda}(y^\lambda)}{P_{Y^0}(y^0) P_{Y^\lambda|X}(y^\lambda|x)}.$$

Here the conditional Poisson distribution $P_{Y^0|X}$ is given by (1) and its marginal is

$$P_{Y^0}(y) = \frac{1}{y!}\mathsf{E}\left\{X^y e^{-X}\right\}.$$

Also, $P_{Y^\lambda|X}(y|x)$ and $P_{Y^\lambda}(y)$ are similarly defined with x and X replaced by $x + \lambda$ and $X + \lambda$ respectively. Easily, the log-likelihood ratio is

$$\log L(X,Y^0,Y^\lambda) = Y^0 \log X - Y^\lambda \log(X + \lambda)$$
$$+ \log \frac{\mathsf{E}\left\{(X' + \lambda)^{Y^\lambda} e^{-X'} \mid Y^\lambda\right\}}{\mathsf{E}\left\{(X')^{Y^0} e^{-X'} \mid Y^0\right\}} \quad (8)$$

where X' is identically distributed as X but independent of all other random variables. Let $N^\lambda = Y^\lambda - Y^0$. Then $N^\lambda \sim \mathcal{P}(\lambda)$ is independent of X. The expectation of the log-likelihood ratio (8) can be further evaluated as

$$-\lambda \mathsf{E} \log X - \mathsf{E}\left\{(X + \lambda)\log(X + \lambda)\right\} + \mathsf{E}U_1 + \mathsf{E}U_2 + o(\lambda) \quad (9)$$

where

$$U_1 = \log \frac{\mathsf{E}\left\{(X' + \lambda)^{N^\lambda}(X')^{Y^0} e^{-X'} \mid Y^0, N^\lambda\right\}}{\mathsf{E}\left\{(X')^{Y^0} e^{-X'} \mid Y^0\right\}}$$

$$U_2 = \lambda \frac{\mathsf{E}\left\{Y^0(X' + \lambda)^{N^\lambda}(X')^{Y^0} e^{-X'} \mid Y^0, N^\lambda\right\}}{\mathsf{E}\left\{(X' + \lambda)^{N^\lambda}(X')^{Y^0} e^{-X'} \mid Y^0, N^\lambda\right\}}.$$

Note that N^λ takes the value of 1 with probability $\lambda + o(\lambda)$ and zero value with probability $1 - \lambda + o(\lambda)$. Then

$$\mathsf{E}U_1 = \lambda \mathsf{E}\left\{\log \frac{\mathsf{E}\left\{(X' + \lambda)(X')^{Y^0} e^{-X'} \mid Y^0\right\}}{\mathsf{E}\left\{(X')^{Y^0} e^{-X'} \mid Y^0\right\}}\right\} + o(\lambda)$$

$$= \lambda \mathsf{E} \log \langle X \rangle + o(\lambda). \quad (10)$$

Meanwhile,

$$\mathsf{E}U_2 = \lambda \mathsf{E}\left\{\frac{\mathsf{E}\left\{Y^0(X')^{Y^0-1} e^{-X'} \mid Y^0\right\}}{\mathsf{E}\left\{(X')^{Y^0} e^{-X'} \mid Y^0\right\}}\right\} + o(\lambda)$$

$$= \lambda + o(\lambda). \quad (11)$$

Putting (9), (10) and (11) together yields

$$\mathsf{E}\left\{\log L(X,Y^0,Y^\lambda)\right\} = \lambda \mathsf{E}\left\{\log \langle X \rangle - \log X\right\} + o(\lambda),$$

and hence (2) by (7). ∎

B. Proof of Lemma 2

We first prove the following result.

Lemma 3 *For every P_X with $\mathsf{E}\{X \log X\} < \infty$, and $\delta \to 0$,*

$$I(X;\mathcal{P}((1+\delta)X) - I(X;\mathcal{P}(X))$$
$$= \delta \mathsf{E}\left\{X \log X - \langle X \rangle \log \langle X \rangle\right\} + o(\delta). \quad (12)$$

Proof: Consider first the case $\delta \to 0^+$. Let $Y = \mathcal{P}(X)$ and $\Delta = \mathcal{P}(\delta X)$ be independent conditioned on X. Let also $Y^\delta = Y + \Delta$. Then, the left hand side of (12) is

$$I\left(X;Y^\delta\right) - I(X;Y)$$

$$= \mathsf{E}\left\{\log \frac{p_{Y^\delta|X}\left(Y^\delta|X\right)}{p_{Y^\delta}\left(Y^\delta\right)} - \log \frac{p_{Y|X}(Y|X)}{p_Y(Y)}\right\}$$

$$= \mathsf{E}\left\{\Delta \log X - \delta X + \log \mathsf{E}\left\{(X')^Y e^{-X'} \mid Y\right\}\right.$$
$$\left. - \log \mathsf{E}\left\{(X')^{Y^\delta} e^{-(1+\delta)X'} \mid Y^\delta\right\}\right\}$$

$$= \mathsf{E}\left\{\delta X \log X - \delta X + \log \mathsf{E}\left\{(X')^Y e^{-X'} \mid Y\right\}\right.$$
$$\left. - \log \mathsf{E}\left\{(X')^Y (X')^\Delta e^{-X'} e^{-\delta X'} \mid Y, \Delta\right\}\right\}, (13)$$

where X' takes the same distribution as X but independent of X, Y and Δ. Since we are interested in the expansion of (13) in the first order of δ, it is harmless to

replace Δ by a random variable that takes the value of 1 with probability δX conditioned on X and 0 otherwise. Hence

$$
\mathsf{E}\left\{\log \mathsf{E}\left\{(X')^Y (X')^\Delta e^{-X'} e^{-\delta X'} \,\Big|\, Y, \Delta\right\}\right\}
$$
$$
= \mathsf{E}\left\{(1-\delta X)\log \mathsf{E}\left\{(X')^Y e^{-X'} e^{-\delta X'} \,\Big|\, Y\right\}\right.
$$
$$
\left. +\delta X \log \mathsf{E}\left\{(X')^{Y+1} e^{-X'} e^{-\delta X'} \,\Big|\, Y\right\}\right\} + o(\delta)
$$
$$
= \mathsf{E}\left\{(1-\delta X)\log \mathsf{E}\left\{(X')^Y e^{-X'} \,\Big|\, Y\right\}\right.
$$
$$
+\delta \frac{\mathsf{E}\left\{(X')^{Y+1} e^{-X'} \,\Big|\, Y\right\}}{\mathsf{E}\left\{(X')^Y e^{-X'} \mid Y\right\}}
$$
$$
\left. +\delta X \log \mathsf{E}\left\{(X')^{Y+1} e^{-X'} \,\Big|\, Y\right\}\right\} + o(\delta)
$$

Plugging into (13) yields (12).

The case of $\delta \to 0^-$ can be similarly proved by letting $Y^\delta = \mathcal{P}((1+\delta)X)$, $\Delta = \mathcal{P}(-\delta X)$, $Y = Y^\delta + \Delta$, and essentially repeating the above. ∎

Proof: [Lemma 2] Applying Lemmas 1 and 3, the change of the mutual information due to δ and λ can be written as

$$
I(X; \mathcal{P}((1+\delta)X + \lambda)) - I(X; \mathcal{P}(X))
$$
$$
= I(X; \mathcal{P}((1+\delta)X + \lambda)) - I(X; \mathcal{P}((1+\delta)X))
$$
$$
+ I(X; \mathcal{P}((1+\delta)X)) - I(X; \mathcal{P}(X))
$$
$$
= \lambda \mathsf{E}\left\{\log((1+\delta)X) - \log\langle(1+\delta)X\rangle\right\}
$$
$$
+ \delta \mathsf{E}\left\{X\log X - \langle X\rangle \log\langle X\rangle\right\} + o(\lambda) + o(\delta).
$$

It remains to express $\mathsf{E}\log\langle(1+\delta)X\rangle$ in terms of $\mathsf{E}\log\langle X\rangle$ and the conditional mean estimates. Let $Y = \mathcal{P}(X)$ and $\Delta = \mathcal{P}(\delta X)$ be independent conditioned on X. Then

$$
\langle(1+\delta)X\rangle = (1+\delta)\frac{\mathsf{E}\left\{(X')^{Y+\Delta+1} e^{(1+\delta)X'} \,\Big|\, Y, \Delta\right\}}{\mathsf{E}\left\{(X')^{Y+\Delta} e^{(1+\delta)X'} \,\Big|\, Y, \Delta\right\}}. \tag{14}
$$

Again, it is harmless to replace Δ by a random variable that takes the value of 1 with probability δX conditioned on X and 0 otherwise. Some algebra leads to

$$
\mathsf{E}\left\{\log\langle(1+\delta)X\rangle\right\} = \mathsf{E}\left\{\log((1+\delta)X)\right\}
$$
$$
+ \delta \mathsf{E}\left\{X\log\frac{\langle X^2\rangle}{\langle X\rangle^2} + X - \frac{\langle X^2\rangle}{\langle X\rangle}\right\} + o(\delta). \tag{15}
$$

Plugging (15) into (14) proves the lemma since also

$$
|\delta\lambda| \leq (\max(|\delta|, |\lambda|))^2 \leq \max\left(|\delta|^2, |\lambda|^2\right) = o(\delta) + o(\lambda).
$$
∎

References

[1] D. Guo, S. Shamai, and S. Verdú, "Mutual information and MMSE in Gaussian channels," in *Proceedings 2004 IEEE International Symposium on Information Theory*, p. 347, Chicago, IL, USA, June 2004.

[2] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *submitted to IEEE Trans. on Inform. Theory*, 2004.

[3] S. Verdú, "Poisson communication theory," *Invited talk*, March 25 1999. The International Technion Communication Day in honor of Israel Bar-David. Available at http://www.princeton.edu/∼verdu/.

[4] I. Bar-David, "Communication under the Poisson regime," *IEEE Trans. Inform. Theory*, vol. 15, pp. 31–37, Jan. 1969.

[5] P. Brémaud, *Point processes and queues*. New York: Springer-Verlag, 1981.

[6] D. L. Snyder, "Filtering and detection for doubly stochastic Poisson processes," *IEEE Trans. Inform. Theory*, vol. 18, pp. 91–102, Jan. 1972.

[7] T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 15, pp. 350–361, May 1969.

[8] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes*, vol. II. New York: Springer-Verlag, 1978.

[9] Y. M. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probability and Its Applications*, vol. 23, pp. 143–147, 1978.

[10] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inform. Theory*, vol. 26, pp. 710–715, Nov. 1980.

[11] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel—Part I and Part II," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1449–1471, Nov. 1988.

[12] S. Shamai and A. Lapidoth, "Bounds on the capacity of a spectrally constrained Poisson channel," *IEEE Trans. Inform. Theory*, vol. 39, pp. 19–29, Jan. 1993.

[13] A. Lapidoth and S. M. Moser, "Bounds on the capacity of the discrete-time Poisson channel," in *Proceedings 41st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Oct. 2003.

[14] A. Lapidoth and S. Shamai, "The Poisson multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 488–501, March 1998.

[15] A. Lapidoth, I. E. Telatar, and R. Urbanke, "On wideband broadcast channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 3250–3258, Dec. 2003.

[16] S. I. Bross and S. Shamai, "Capacity and decoding rules for the Poisson arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 49, pp. 3076–3093, Nov. 2003.

[17] D. Guo, S. Shamai, and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *in preparation for IEEE Trans. on Inform. Theory*, 2004.

[18] T. E. Duncan, "On the calculation of mutual information," *SIAM Journal of Applied Mathematics*, vol. 19, pp. 215–220, July 1970.

[19] D. Guo, *Gaussian Channels: Information, Estimation and Multiuser Detection*. PhD thesis, Princeton University, 2004.

[20] J. Bertoin, *Lévy Processes*. Cambridge University Press, 1996.