Capacity of the Poisson Channel with Random Noise Intensity

Michael R. Frey

Department of Statistics
University of North Carolina
Chapel Hill, NC 27599

LISS 33

August 1988 (revised March 1989)

Abstract - An upper bound B_f is given for the information capacity of the Poisson channel with marginally stationary random noise intensity and causal feedback. The capacity is shown to converge to B_f in the limit of longer and longer communication intervals for a class of random noise intensities including the case in which the noise intensity is nontime-varying. An upper bound B_{nf} on the capacity is also established for the Poisson channel with marginally stationary noise and no feedback. In this case also, for a class of random noise intensities including nontime-varying noise intensity, B_{nf} is found to be the capacity of the channel without feedback in the limit of longer and longer communication intervals. The fractional difference between B_f and B_{nf} is considered as a means to quantify the improvement afforded by feedback. Also, certain nonstandard encoder constraints are addressed and the importance of the encoder intensity peak constraint to the channel capacity problem is explored.

Supported by ARO Contract DAAl03-86-G-0022, ONR Contract N00014-86-K-0039, and ONR Contract N00014-89-5-1175.

These results were presented in part at the 23rd Annual Conference on Information Sciences and Systems, March 22-24, 1989, Johns Hopkins University.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for Public Release: Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Department of Statistics; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code): University of North Carolina, Chapel Hill, North Carolina 27514	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Office of Naval Research; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-86-K-0039 and N00014-89-J-1175
8c. ADDRESS (City, State and ZIP Code): Statistics & Probability Program, Arlington, VA 22217	10. SOURCE OF FUNDING NOS.: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.

11. TITLE (Include Security Classification): Capacity of the Poisson Channel with Random Noise Intensity

12. PERSONAL AUTHOR(S): M.R. Frey

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
TECHNICAL	FROM ____ TO ____	March 1989	17

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Poisson-type point processes; Information capacity, Direct-detection optical channel

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

An upper bound B_f is given for the information capacity of the Poisson channel with marginally stationary random noise intensity and causal feedback. The capacity is shown to converge to B_f in the limit of longer and longer communication intervals for a class of random noise intensities including the case in which the noise intensity is nontime-varying. An upper bound B_{nf} on the capacity is also established for the Poisson channel with marginally stationary noise and no feedback. In this case also, for a class of random noise intensities including nontime-varying noise intensity, B_{nf} is found to be the capacity of the channel without feedback in the limit of longer and longer communication intervals. The fractional difference between B_f and B_{nf} is considered as a means to quantify the improvement afforded by feedback. Also, certain nonstandard encoder constraints are addressed and the importance of the encoder intensity peak constraint to the channel capacity problem is explored.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☒ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL: C.R. Baker	22b. TELEPHONE NUMBER (Include Area Code): (919) 962-2189; 22c. OFFICE SYMBOL

I. Introduction

The Poisson channel is an additive noise channel with output $Y_t = X_t + N_t$ where $N = \{N_t\}_{0 \le t \le T}$ is the channel noise and $X = \{X_t\}_{0 \le t \le T}$ is the transmitted signal into which is encoded the message $\theta = \{\theta_t\}_{0 \le t \le T}$. X is sometimes viewed as the output of the channel encoder. All processes in the channel model are defined on a common probability space $(\Omega, \boldsymbol{F}, P)$. We write $\boldsymbol{F}^\theta$ for the natural history of θ, $\boldsymbol{F}^N$ for the natural history of N, etc.

Poisson-type point processes [6] are a class of simple point processes in which the compensator takes the form

$$\int_0^t a_s \, db(s)$$

where a_t is some nonnegative predictable process (the intensity of the point process) and $b(t)$ is a deterministic nonnegative right-continuous and nondecreasing function. In the Poisson channel (often called the Poisson-type point process channel) both X and N are Poisson-type point processes. We shall only consider Poisson-type point processes with $b(t) \equiv t$. The noise N in the Poisson channel is given to be directed by the H-predictable noise intensity Λ_t where H is a history such that $\boldsymbol{F}_t^N \subset H_t$ for all $t \in [0,T]$. X is given to be directed by the $\boldsymbol{F}^\theta \vee \boldsymbol{F}^Y$-predictable encoding intensity χ_t. Then the channel output $Y_t = X_t + N_t$ is also a Poisson-type point process directed by the intensity $\eta_t = \chi_t + \Lambda_t$ where η_t is predictable with respect to the global history $\boldsymbol{F}^\theta \vee H \vee \boldsymbol{F}^Y$. In the Poisson channel model, the message θ is encoded into the channel encoder output X_t indirectly via the encoding intensity $\chi_t = \chi_t(\theta, Y)$. $\boldsymbol{F}_t^\theta \vee \boldsymbol{F}_t^Y$-predictability of χ_t allows for nonanticipative message encoding and causal, noiseless, instantaneous feedback.

The augmented noise history H is introduced into the channel model so that the whole class of Poisson-type point processes with integrating function $b(t) = t$ can be considered as models for the channel noise. In particular, renewal processes, self-exciting processes and doubly stochastic Poisson processes fall into this class [7]. For instance, N is a doubly stochastic Poisson process for $H_t = \boldsymbol{F}_T^\Lambda \vee \boldsymbol{F}_t^N$ (suitably completed).

In the absence of constraints on the encoder output, the information capacity of the Poisson channel is infinite. A peak constraint $0 \le \chi_t \le c$ is generally imposed on the encoder intensity [1], [2], [3]. An average constraint

$$E\left[\int_0^T \chi_t \, dt\right] \le k_0 T$$

has also been considered [1], [3]. These and other encoder constraints are considered in Section V where some justification is given for emphasis on the peak constraint in this and earlier papers [1], [2], [3].

To define channel capacity let μ_θ, μ_Y, and $\mu_{\theta Y}$ be the marginal and joint measures induced by the message and output processes, θ and Y, on the spaces S_θ, S_Y, and $S_\theta \times S_Y$ where S_θ and S_Y are the spaces of trajectories of θ and Y over the interval $[0,T]$. Write the induced product measure as $\mu_{\theta \times Y}$. Then, the average mutual information in θ and Y over the interval $[0,T]$ is [5]

$$I^T[\theta, Y] = E\left[\ln \frac{d\mu_{\theta Y}}{d\mu_{\theta \times Y}}\right]$$

provided $\mu_{\theta Y} \ll \mu_{\theta \times Y}$; otherwise $I^T[\theta, Y] = \infty$. Using an obvious notation, the conditional mutual information [5] in θ and Y given a trajectory of the noise intensity is

$$I^T[\theta, Y \mid \Lambda] = E\left[\ln \frac{d\mu_{\theta Y \mid \Lambda}}{d\mu_{\theta \times Y \mid \Lambda}} \mid \Lambda\right].$$

The channel information capacity is

$$C = \sup_{\theta} \sup_{X} \frac{1}{T} I^T[\theta, Y]$$

where the suprema are taken over all admissible message processes, θ, and all admissible channel encodings, X. The class of admissible message processes is generally taken to be all jointly measurable real-valued processes with index set $[0,T]$. A further comment on admissibility of messages and encodings for the Poisson channel can be found in [2].

We shall assume that the message and noise are independent. Specifically, it is assumed that the histories H and $\boldsymbol{F}^\theta$ are independent. This assumption has the essential consequence that the average mutual information in the channel is upper-bounded by the expected conditional information given the path of the noise intensity. Independence of the histories H and $\boldsymbol{F}^\theta$ also implies that, in the no-feedback case, χ_t and Λ_t are independent processes.

The capacity of the Poisson channel has been previously treated for cases in which the noise intensity is a real constant $\Lambda_t = \lambda$ [2], [3] and for cases in which the noise intensity is a deterministic function of time $\Lambda_t = \lambda(t)$ [1]. Applications of and further references to the Poisson channel model are given in [4]. We refer to [1] for motivation for treating the channel noise intensity as a stochastic process.

It will be convenient to have notations for certain conditional expectations. For any random variable z define

$$\begin{aligned} \hat{z} &= E[z \mid \boldsymbol{F}_{t-}^Y], \\ \hat{\hat{z}} &= E[z \mid \boldsymbol{F}_{t-}^Y \vee \boldsymbol{F}_T^\Lambda], \\ \tilde{z} &= E[z \mid \boldsymbol{F}_{t-}^Y \vee \boldsymbol{F}_{t-}^\theta]. \end{aligned}$$

II. Average Mutual Information for Random Noise Intensity

The average mutual information in the Poisson channel with random noise intensity is given by Liptser and Shiryayev [6] to be

$$I^T[\theta,Y] = E\left[\int_0^T (w(\chi_t + \tilde{\Lambda}_t) - w(\hat{\chi}_t + \hat{\Lambda}_t))dt\right] \tag{1}$$

where $w(x) = x \ln x$. Note that $\tilde{\chi}_t = \chi_t$. For the case of nonrandom noise intensity, $\hat{\Lambda}_t = \tilde{\Lambda}_t = \Lambda_t$. In this case the usually intractable calculation of $\hat{\Lambda}_t$ and $\tilde{\Lambda}_t$ is avoided and (1) is greatly simplified. In the case of stochastic noise intensity, the presence of $\hat{\Lambda}_t$ and $\tilde{\Lambda}_t$ limits the direct usefulness of the expression for $I^T[\theta,Y]$. The conditional information given the path of the noise intensity

$$I^T[\theta,Y \mid \Lambda] = E\left[\int_0^T (w(\chi_t + \Lambda_t) - w(\overset{*}{\chi}_t + \Lambda_t))dt \mid F_T^\Lambda\right]$$

sidesteps this difficulty. We have the following inequality:

Proposition 1: Suppose the message and noise intensity processes θ and Λ are independent. Then

$$I^T[\theta,Y] \le E\left[\int_0^T (w(\chi_t + \Lambda_t) - w(\overset{*}{\chi}_t + \Lambda_t))dt\right].$$

Proof: This is a special case of a general inequality which follows from Kolmogorov's formula [8]:

$$I[X,Z] + E[I[X,Y \mid Z]] = I[X,(Y,Z)]$$

for random elements X, Y, Z provided $\mu_{XYZ} \ll \mu_X \times \mu_{YZ}$. Independence of X and Z implies $I[X,Z] = 0$. Also, $I[X,(Y,Z)] \ge I[X,Y]$ so, in the context of the Poisson channel,

$$I^T[\theta,Y] \le E[I^T[\theta,Y \mid \Lambda]].$$

Proposition 2:

$$I^T[\theta,Y] \ge E\left[\int_0^T (w(\chi_t + \tilde{\Lambda}_t) - w(\overset{*}{\chi}_t + \Lambda_t))dt\right].$$

Proof: Using Jensen's inequality, we have $E[w(\overset{*}{\chi}_t + \Lambda_t)] \ge E[w(\hat{\chi}_t + \hat{\Lambda}_t)]$. Use this inequality in (1) and the result follows.

Let Φ be the class of marginally stationary noise intensities Λ_t such that $|\tilde{\Lambda}_t - \Lambda_t| \to 0$ a.s. as $t \to \infty$. Define

$$\Phi_M = \{\Lambda_t \in \Phi : E[w(\Lambda_t)] < \infty\},$$

$$\Phi_B = \{\Lambda_t \in \Phi : \Lambda_t \text{ is bounded}\}.$$

Let $L \ge 0$ be a r.v. such that L is bounded or $E[L \ln L]$ is finite. Then $\Lambda_t = L$ belongs to Φ_M

or Φ_B. More generally, let $p(t)$ be a bounded nonnegative periodic function with period t_0 and suppose Π is uniformly distributed over $[0,t_0]$ and independent of L. Then $\Lambda_t = Lp(t-\Pi)$ belongs to Φ_M or Φ_B.

Proposition 9: Suppose $\Lambda_t \in \Phi_M \cup \Phi_B$. Then, for the Poisson channel

$$\lim_{T \to \infty} \frac{1}{T} I^T[\theta, Y] = \lim_{T \to \infty} \frac{1}{T} E\left[\int_0^T (w(\chi_t + \Lambda_t) - w(\hat{\chi}_t + \Lambda_t))dt\right].$$

Proof: Use the upper and lower bounds for $I^T[\theta, Y]$ given in Propositions 1 and 2 and the result follows.

According to Proposition 3,

$$\lim_{T \to \infty} \frac{1}{T} I^T[\theta, Y] = \lim_{T \to \infty} \frac{1}{T} E[I^T[\theta, Y \mid \Lambda]]$$

for noise intensities $\Lambda_t \in \Phi_M \cup \Phi_B$.

III. CHANNEL CAPACITY

In this section we give capacity results for the Poisson channel with and without feedback. For clarity, we content ourselves with treating marginally stationary noise intensities; that is, noise intensities having a common marginal distribution $F(x) = P\{\Lambda_t \le x\}$ for all $t \in [0,T]$. However, the results and their proofs given below extend in a natural way to nonstationary noise intensities. Also, we only impose a peak constraint on the encoder intensity. Using the method of Davis [3], the extension to encoder intensities both peak- and average-constrained is obvious. We also note that the results given here for a constant peak constraint on the encoder intensity can be adjusted to apply to a time-varying peak constraint by simple function approximation of the peak constraint as in [1].

Theorem 4: Let the noise process N_t in a Poisson channel have marginally stationary noise intensity Λ_t with marginal distribution $F(t) = P\{\Lambda_t \le x\}$ and suppose Λ_t is independent of the message process θ_t. Suppose the encoder output intensity χ_t is peak-constrained, $0 \le \chi_t \le c$ and require χ_t to be $\mathbf{F}^\theta \vee \mathbf{F}^Y$-adapted so that causal feedback is possible. Then

$$\mathcal{C} \le B_f \equiv E[C(\Lambda_t)]$$

where

$$C(x) = \frac{x}{c}\left(1 + \frac{c}{x}\right)^{1+x/c} - x\left(1 + \frac{x}{c}\right)\ln\left(1 + \frac{c}{x}\right) .$$

Proof: By Proposition 1

$$I^T[\theta,Y] \le E\left[E\left[\int_0^T(\phi(\chi_t,\Lambda_t) - \phi(\hat{\chi}_t,\Lambda_t))dt \mid \Lambda_t\right]\right] .$$

From [1], for nonrandom noise intensity $\lambda(t)$, we have

$$E\left[\int_0^T(\phi(\chi_t,\lambda(t)) - \phi(\hat{\chi}_t,\lambda(t)))dt\right] \le \int_0^T C(\lambda(t))dt$$

so, for marginally stationary Λ_t,

$$\frac{1}{T}I^T[\theta,Y] \le E\left[\frac{1}{T}\int_0^T C(\Lambda_t)dt\right] = E[C(\Lambda_t)] .$$

Theorem 5: Consider the Poisson channel presented in Theorem 4 but now restrict the encoding intensity to be $\mathbf{F}^\theta$-adapted so that no feedback is allowed. Then

$$\mathcal{C} \le B_{nf} \equiv \max_{\frac{1}{e} \le p \le \frac{1}{2}} E[h(p,\Lambda_t)]$$

where

$$h(p,\Lambda_t) = p\,\phi(c,\Lambda_t) - \phi(pc,\Lambda_t)$$

with

$$\phi(x,y) = (x + y)\ln(x + y) - y\ln y .$$

Proof: χ_t is $\boldsymbol{F}^{\theta}$-adapted so χ_t and Λ_t are independent. Thus $E[\hat{\chi}_t \mid \Lambda_t] = E[\chi_t]$. Then, by Jensen's inequality we have $E[w(\hat{\chi}_t + \Lambda_t)] \geq E[w(E[\chi_t] + \Lambda_t)]$. So, by Proposition 1,

$$I^T[\theta, Y] \leq \int_0^T \left(E[\phi(\chi_t, \Lambda_t)] - E[\phi(E[\chi_t], \Lambda_t)] \right) dt \ .$$

Let Λ_0 be Λ_t for $t = 0$. The independence of Λ_t and χ_t together with the marginal stationarity of Λ_t gives $E[\phi(\chi_t, \Lambda_t)] = E[\phi(\chi_t, \Lambda_0)]$. Also $E[\phi(E[\chi_t], \Lambda_t)] = E[\phi(E[\chi_t], \Lambda_0)]$. Then

$$I^T[\theta, Y] \leq \int_0^T \left(E[\phi(\chi_t, \Lambda_0)] - E[\phi(E[\chi_t], \Lambda_0)] \right) dt \ .$$

Thus

$$\frac{1}{T} I^T[\theta, Y] \leq \sup_{t \in [0,T]} \left(E[\phi(\chi_t, \Lambda_0)] - E[\phi(E[\chi_t], \Lambda_0)] \right) \ .$$

In the supand above, t enters only through χ_t. Let $\mu(A) \equiv P \circ \chi_t^{-1}(A)$. The support of μ is contained in $[0,c]$. Define

$$Q_\mu(g) \equiv \int_0^c g(x) \mu(dx) \ .$$

Then $Q_\mu(i) = E[\chi_t]$ where i is the identity function. Also $Q_\mu(E[\phi_0]) = E[\phi(\chi_t, \Lambda_0)]$ using the independence of χ_t and Λ_0. We can now write

$$C \leq \sup_{\mu \in M} \left[Q_\mu(E[\phi_0]) - E[\phi_0(Q_\mu(i))] \right]$$

where M is the set of all probability measures on $([0,c], \boldsymbol{B}[[0,c]])$. For any $\mu \in M$ there is a $k \in [0,c]$ such that $Q_\mu(i) = k$ so

$$C \leq \sup_{0 \leq k \leq c} \sup_{\substack{\mu \in M \\ Q_\mu(i) = k}} \left[Q_\mu(E[\phi_0]) - E[\phi_0(Q_\mu(i))] \right]$$

$$= \sup_{0 \leq k \leq c} \left[\sup_{\substack{\mu \in M \\ Q_\mu(i) = k}} Q_\mu(E[\phi_0]) - E[\phi_0(k)] \right] .$$

$Q_\mu \circ E$ is linear so

$$\sup_{\substack{\mu \in M \\ Q_\mu(i) = k}} Q_\mu(E[\phi_0]) = \sup_{\substack{\mu \in M \\ Q_\mu(i) = k}} Q_\mu \left(E[\phi_0 - \frac{\phi_0(c)}{c} i] \right) + Q_\mu \left(E[\frac{\phi_0(c)}{c} k] \right) .$$

The function $\phi_0(x) - \frac{x}{c}\phi_0(c)$ is maximum at $x = 0$ and $x = c$ (for any value of Λ_0.) Thus

$$E[Q_\mu(\phi_0 - \frac{\phi_0(c)}{c} i)]$$

is maximized for $Q_\mu(i) = k$ by the probability measure

$$\mu(\{0\}) = \frac{c-k}{c} , \qquad \mu(\{c\}) = \frac{k}{c}$$

in which case $E[Q_\mu(\phi_0 - \frac{\phi_0(c)}{c} i)]$. So

$$C \leq \sup_{0 \leq k \leq c} E\left[\frac{\phi_0(c)}{c} k - \phi_0(k) \right]$$

$$= \sup_{0 \leq p \leq 1} \left[pE[\phi(c, \Lambda_0)] - E[\phi(pc, \Lambda_0)] \right]$$

$$= \sup_{0 \le p \le 1} E\left[h(p, \Lambda_t)\right].$$

For $0 < p < 1$, $H(p) = E[h(p, \Lambda_t)]$ has the first and second derivatives

$$\frac{dH}{dp} = E[\phi(c, \Lambda_t)] - c - cE[\ln(pc + \Lambda_t)]$$

$$= cE[\ln(K_t + \Lambda_t)] - cE[\ln(pc + \Lambda_t)] \tag{2}$$

where $K_t = k(\Lambda_t)$ with

$$k(x) = \frac{x}{c}\left(1 + \frac{c}{x}\right)^{1+x/c} - x$$

and

$$\frac{d^2H}{dp^2} = -c^2 E\left[\frac{1}{pc + \Lambda_t}\right].$$

$H(p)$ is concave and so has a unique maximum over $p \in [0,1]$. To see that $H(p)$ has its maximum in the subinterval $p \in [1/e, 1/2]$ we just note that according to (2), $p_{\max}$ maximizing $H(p)$ must satisfy

$$E\left[\ln \frac{p_{\max} c + \Lambda_t}{K_t + \Lambda_t}\right] = 0$$

and the range of k is $[c/e, c/2]$. Thus

$$\sup_{0 \le p \le 1} E[h(p, \Lambda_t)] = \max_{\frac{1}{e} \le p \le \frac{1}{2}} E[h(p, \Lambda_t)].$$

The proof is complete.

Of course, $B_{nf} \le B_f$ since

$$B_{nf} = \max_{\frac{1}{e} \le p \le \frac{1}{2}} E[h(p, \Lambda_t)]$$

and

$$B_f = E[C(\Lambda_t)] = E\left[\max_{\frac{1}{e} \le p \le \frac{1}{2}} h(p, \Lambda_t)\right].$$

Theorem 6: Let the noise intensity Λ_t in a Poisson channel belong to Φ_M or Φ_B. Suppose the encoder intensity is peak-constrained and no feedback is allowed. Then $\mathcal{C} \to B_{nf}$ as $T \to \infty$.

Proof: There exists a sequence of message processes $\{\theta^{(m)}\}$ [2] such that [1]

$$\lim_{m \to \infty} I^T[\theta^{(m)}, Y \mid \Lambda] = \int_0^T h(p, \Lambda_t)\, dt$$

where $p \in [0,1]$ is a parameter of the process $\theta^{(m)}$. By Proposition 3

$$\lim_{m \to \infty} \lim_{T \to \infty} \frac{1}{T} I^T[\theta^{(m)}, Y] = \lim_{T \to \infty} \frac{1}{T} \int_0^T E[h(p, \Lambda_t)]\, dt$$

Λ_t is marginally stationary so

$$\lim_{m \to \infty} \lim_{T \to \infty} \frac{1}{T} I^T [\theta^{(m)}, Y] = E[h(p, \Lambda_t)] .$$

This is true for all $p \in [0,1]$ so

$$\lim_{T \to \infty} C \geq \sup_{p \in [0,1]} E[h(p, \Lambda_t)] .$$

This together with Theorem 5 proves the result.

Efficient communication through the Poisson channel requires that the encoder be "tuned" to the channel; i.e., that the encoder be adjusted in accordance with the amount of noise present. This requires knowledge of the channel noise intensity. In cases where the noise intensity is random the state of the noise intensity process is unknown to the sender. An encoding scheme incorporating feedback is then needed whereby the sender makes a feedback-based estimate of the noise intensity and then adjusts the encoding process to agree with that estimate. Without feedback the capacity cannot exceed $B_{nf} \leq B_f$. The next theorem states that $C \to B_f$ as $T \to \infty$ if causal feedback is allowed.

Theorem 7: For the Poisson channel with marginally stationary noise intensity Λ_t belonging to Φ_M or Φ_B, peak-constrained encoder intensity, and causal feedback, $C \to B_f$ as $T \to \infty$.

Proof: As in the proof of Theorem 6 we have

$$\lim_{m \to \infty} \lim_{T \to \infty} \frac{1}{T} I^T [\theta^{(m)}, Y] = \lim_{T \to \infty} \frac{1}{T} \int_0^T E[h(p, \Lambda_t)] dt .$$

Choose $p = k(\tilde{\Lambda}_t)/c$ and check that $C(x) = h(k(x)/c, x)$. For $\Lambda_t \in \Phi_B \cup \Phi_M$

$$\lim_{t \to \infty} E[h(k(\tilde{\Lambda}_t)/c, \Lambda_t)] = \lim_{t \to \infty} E[C(\Lambda_t)] = E[C(\Lambda_t)] .$$

So

$$\lim_{T \to \infty} C \geq \lim_{T \to \infty} \lim_{m \to \infty} \frac{1}{T} I^T [\theta^{(m)}, Y] = E[C(\Lambda_t)] .$$

This completes the proof.

IV. Improvement Afforded by Feedback

For Poisson channels with nonrandom noise intensity $B_f = B_{nf}$ consistent with the fact that causal feedback does not increase channel capacity in this case [1]. For stochastic noise intensity, however, $B_f \neq B_{nf}$ in general. Thus for Poisson channels with random noise intensity Λ_t the fractional difference

$$\frac{B_f - B_{nf}}{B_f}$$

is, in the sense of Theorems 4, 5, 6, and 7, a measure of the improvement afforded by the use of causal feedback. Let Λ_t be marginally stationary and let L be a random variable with distribution $F(x) = P\{\Lambda_t \leq x\}$. Define D to be the set of all such L, i.e., D is the set of all nonnegative random variables. We shall show that

$$\sup_{L \in D} \frac{B_f - B_{nf}}{B_f} \doteq 0.0163 .$$

This means that, while causal feedback offers some improvement in the case of stochastic noise intensity, the improvement may never be very large. In particular, in the cases where B_f and B_{nf} are the channel capacities, the improvement afforded by feedback never exceeds 1.63% of that achievable without feedback. An analytic expression is obtained for the above supremum by way of the following series of lemmas. W.l.o.g. we take $c = 1$.

Lemma 8: Let D_b be the set of all bounded nonnegative random variables. Then

$$\sup_{L \in D} \frac{B_f - B_{nf}}{B_f} = \sup_{L \in D_b} \frac{B_f - B_{nf}}{B_f} .$$

Proof: For any nonnegative r.v. L and positive real number β let $L_\beta = L\,1_{\{L \leq \beta\}} + \beta 1_{\{L > \beta\}}$. Then $L_\beta \in D_b$. Write $B_f(L)$ and $B_{nf}(L)$ for B_f and B_{nf} evaluated, respectively, for a marginally stationary noise intensity whose marginal distribution is that of the r.v. L. We have

$$\begin{aligned}
&\left| \frac{B_f(L) - B_{nf}(L)}{B_f(L)} - \frac{B_f(L_\beta) - B_{nf}(L_\beta)}{B_f(L_\beta)} \right| \\
&\qquad = \left| \frac{B_{nf}(L)}{B_f(L)} - \frac{B_{nf}(L)}{B_f(L_\beta)} + \frac{B_{nf}(L)}{B_f(L_\beta)} - \frac{B_{nf}(L_\beta)}{B_f(L_\beta)} \right| \\
&\qquad \leq \frac{B_{nf}(L)}{B_f(L) B_f(L_\beta)} \left| B_f(L) - B_f(L_\beta) \right| + \frac{\left| B_{nf}(L) - B_{nf}(L_\beta) \right|}{B_f(L_\beta)} .
\end{aligned}$$

Now $C(\cdot)$ is a decreasing function so

$$E[C(L)] \leq E[C(L_\beta)] \leq E[C(L)] + C(\beta) P\{L > \beta\} .$$

Then, as $\beta \to \infty$,

$$\begin{aligned}
\left| B_f(L) - B_f(L_\beta) \right| &= \left| E[C(L)] - E[C(L_\beta)] \right| \\
&\leq C(\beta) P\{L > \beta\} \\
&\to 0 .
\end{aligned}$$

Also

$$\frac{\partial}{\partial\lambda}h(p,\lambda) = p\ln(1+1/\lambda) - \ln(1+p/\lambda)$$

and, for $|x| \to \infty$,

$$\ln(1+1/x) = \frac{1}{x} - \frac{1}{2x^2} + o(x^{-2})$$

so, for $\lambda \to \infty$,

$$\lambda^2 \frac{\partial}{\partial\lambda}h(p,\lambda) = -\frac{p(1-p)}{2} + o(1)\,.$$

Thus, for all λ large enough, $h(p,\lambda)$ is decreasing in λ. Then, for β large enough,

$$E[h(p,L)] \le E[h(p,L_\beta)] \le E[h(p,L)] + h(p,\beta)P\{L>\beta\}$$

so that, as $\beta \to \infty$,

$$\begin{aligned}
|B_{nf}(L) - B_{nf}(L_\beta)| &= |\max_p E[h(p,L)] - \max_p E[h(p,L_\beta)]| \\
&\le \max_p |E[h(p,L)] - E[h(p,L_\beta)]|\ . \\
&\le \max_p h(p,\beta)P\{L>\beta\} \\
&= C(\beta)P\{L>\beta\} \\
&\to 0\,.
\end{aligned}$$

This completes the proof.

Lemma 9: Let D_2 be the family of all random variables L of the form

$$L = \begin{cases} 0 & \text{w.p. } \epsilon \\ \beta & \text{w.p. } 1-\epsilon \end{cases}$$

for $\beta > 0$ and $\epsilon \in (0,1)$. Then

$$\sup_{L \in D_b} \frac{B_f - B_{nf}}{B_f} = \sup_{L \in D_2} \frac{B_f - B_{nf}}{B_f}\,.$$

Proof: Suppose $L \in D_b$ and let $\beta = \operatorname{esssup}_\Omega L$. There is a unique $L_2 \in D_2$ with

$$L_2 = \begin{cases} 0 & \text{w.p. } \epsilon \\ \beta & \text{w.p. } 1-\epsilon \end{cases}$$

for some $\epsilon \in (0,1)$ such that $E[C(L)] = E[C(L_2)]$. Moreover $B_{nf}(L_2) \le B_{nf}(L)$ so $B_f(L) - B_{nf}(L) \le B_f(L_2) - B_{nf}(L_2)$. Thus

$$\left.\frac{B_f - B_{nf}}{B_f}\right|_L \le \left.\frac{B_f - B_{nf}}{B_f}\right|_{L_2}$$

Noting that $D_2 \subset D_b$, we get the desired result.

B_{nf} may be written $B_{nf} = E[h(p_{max},L)]$ where, by (2), p_{max} is obtained from $E[\ln(p_{max}+L)] = E[\ln(k(L)+L)]$. For $L \in D_2$, p_{max} is a function of ϵ and β. Its partial derivatives with respect to ϵ and β can also be obtained from $E[\ln(p_{\max} + L)] = E[\ln(k(L) + L)]$ by implicit differentiation. Then, by the usual methods of differential calculus, the function

$$f(\epsilon,\beta) \equiv \left.\frac{B_f - B_{nf}}{B_f}\right|_{L_2}$$

can be shown to approach its supremum over $\{(\epsilon,\beta) : 0 \leq \epsilon \leq 1, \beta > 0\}$ as $\epsilon \to 0$ with $\beta = (e\,\epsilon)^{-1}$. Thus,

$$\sup_{L \in D_2} \frac{B_f - B_{nf}}{B_f} = \lim_{\epsilon \to 0} \left.\frac{B_f - B_{nf}}{B_f}\right|_{L_\epsilon}$$

where

$$L_\epsilon = \begin{cases} 0 & \text{w.p. } \epsilon \\ \frac{1}{e\,\epsilon} & \text{w.p. } 1-\epsilon \end{cases}.$$

Lemma 10: For $\epsilon > 0$ and $L = L_\epsilon$,

$$B_f = \frac{8+e^2}{8e}\epsilon + o(\epsilon).$$

Proof: For $L = L_\epsilon$,

$$B_f = \epsilon C(0) + (1-\epsilon) C\left(\frac{1}{e\,\epsilon}\right).$$

By Th. 2 in [3],

$$C(x) = \frac{1}{8x} + o\left(\frac{1}{x}\right).$$

Using $C(0) = 1/e$, this leads to the desired result.

Lemma 11: For $\epsilon > 0$ and $L = L_\epsilon$,

$$B_{nf} = \epsilon p_\epsilon (1 + e p_\epsilon / 2) + o(\epsilon)$$

where p_ϵ is the value of p in $[1/e, 1/2]$ which maximizes $H(p, L) = H(p, L_\epsilon)$.

Proof: Define $K_\epsilon = k(L_\epsilon)$. Then

$$\begin{aligned} B_{nf} &= p_\epsilon E[\phi(1, L_\epsilon)] - E[\phi(p_\epsilon, L_\epsilon)] \\ &= p_\epsilon + p_\epsilon E[\ln(K_\epsilon + L_\epsilon)] - E[(p_\epsilon + L_\epsilon)\ln(p_\epsilon + L_\epsilon)] + E[L_\epsilon \ln L_\epsilon] \\ &= p_\epsilon + p_\epsilon \{E[\ln(K_\epsilon + L_\epsilon)] - E[\ln(p_\epsilon + L_\epsilon)]\} - E[L_\epsilon \ln(p_\epsilon + L_\epsilon)] + E[L_\epsilon \ln L_\epsilon]. \end{aligned}$$

Recall from (2) that $E[\ln(K_\epsilon + L_\epsilon)] = E[\ln(p_\epsilon + L_\epsilon)]$. Thus

$$\begin{aligned} B_{nf} &= p_\epsilon - E[L_\epsilon \ln(p_\epsilon + L_\epsilon)] + E[L_\epsilon \ln L_\epsilon] \\ &= p_\epsilon - (1-\epsilon)\frac{1}{e\,\epsilon}\ln\left(p_\epsilon + \frac{1}{e\,\epsilon}\right) + (1-\epsilon)\frac{1}{e\,\epsilon}\ln\frac{1}{e\,\epsilon} \\ &= p_\epsilon - (1-\epsilon)\frac{1}{e\,\epsilon}\ln(1 + e\,\epsilon p_\epsilon). \end{aligned}$$

Using $\ln(1+x) = x - x^2/2 + o(x^2)$ the desired result follows.

Theorem 12:

$$\sup_{L \in D} \frac{B_f - B_{nf}}{B_f} = \frac{\frac{8+e^2}{8e} - p_0\left(1+\frac{ep_0}{2}\right)}{\frac{8+e^2}{8e}}$$

where

$$p_0 \equiv \lim_{\epsilon \to 0} p_\epsilon \doteq 0.436815$$

satisfies $\ln p_0 + ep_0 = e/2 - 1$.

Proof: By the preceding lemmas,

$$\sup_{L \in D} \frac{B_f - B_{nf}}{B_f} = \lim_{\epsilon \to 0} \frac{B_f - B_{nf}}{B_f}\Bigg|_{L_\epsilon} = \frac{\frac{8+e^2}{8e} - p_\epsilon\left(1+\frac{ep_\epsilon}{2}\right)}{\frac{8+e^2}{8e}} + o(1).$$

Thus all that remains to be shown is that $\ln p_0 + ep_0 = e/2 - 1$. Using $E[\ln(K_\epsilon + L_\epsilon)] = E[\ln(p_\epsilon + L_\epsilon)]$ from (4) we get

$$(1+\frac{1}{e\epsilon})(1-\epsilon)\ln(1+e\epsilon) - 1 = \epsilon \ln p_\epsilon + (1-\epsilon)\ln(1+e\epsilon p_\epsilon).$$

The LHS of this equation can be expressed

$$\text{LHS} = (\frac{e}{2} - 1)\epsilon + o(\epsilon)$$

while for the RHS, using $p_\epsilon = p_0 + o(1)$, we get

$$\text{RHS} = \epsilon \ln p_0 + e\epsilon p_0 + o(\epsilon).$$

The proof is complete.

V. ENCODER OUTPUT CONSTRAINTS

The usual type of constraint imposed on the encoder in determining the capacity of the Poisson channel is a peak constraint [1], [2], [3]:

$$0 \le \chi_t \le c \ , \quad 0 \in [0,T] \tag{3}$$

or an average constraint [1], [3]:

$$E\left[\int_0^T \chi_s \, ds\right] \le k_0 T \ . \tag{4}$$

In this paper, also, we have used primarily a peak constraint on encoder output intensity. This emphasis in favor of a peak constraint is now given some justification.

(3) and (4) are constraints on the intensity (and compensator) of the encoder output. One might rather consider constraints imposed directly upon the output X_t of the encoder. We write P_c to signify the class of conditionally Poisson processes. For $X \in P_c$ we have [4, p. 407]

$$Var\,[X_t] = Var\,[A_t] + E\,[A_t]$$

where A_t is the compensator of X_t. Thus for $X \in P_c$ the constraint $Var\,[X_T] \le P$ is equivalent to the pair of constraints on the encoder compensator:

$$E\,[A_T] \le k_0$$

$$Var\,[A_T] \le V$$

with $0 \le k_0 \le P$, $0 \le V \le P$, and $k_0 + V \le P$.

For clarity let us write $C(c, k_0, V, P)$ for the capacity given the following constraints on the encoder output:

$$0 \le \chi_t \le c \ ,$$

$$0 \le E\,[A_T] \le k_0 \, ,$$

$$0 \le Var\,[A_T] \le V \ ,$$

$$0 \le Var\,[X_T] \le P \ .$$

To exclude a constraint the corresponding argument of $C(\cdot,\cdot,\cdot,\cdot)$ is set to ∞; $C(\infty,\cdot,\cdot,\cdot)$ is the channel capacity with no peak constraint on the encoder intensity, etc. A subscript is used $C_{P_c}(\cdot,\cdot,\cdot,\cdot)$ to indicate that, for the encoder output X, we require $X \in P_c$. Using this notation, we have

$$C_{P_c}(\infty,\infty,\infty,P) = \sup_{(k_0,V)\in R} C_{P_c}(\infty,k_0,V,\infty)$$

where $R = \{(x,y) : x \ge 0, \ y \ge 0, \ x+y \le P\}$.

We now show, for $\Lambda_t = \lambda$, that

$$C(\infty,\infty,\infty,P) = \infty \ .$$

The proof uses several lemmas.

Lemma 13: Let θ_t, $0 \le t \le T$ be a random telegraph signal [2]; i.e., a left-continuous homogeneous Markov jump process with states $\{0,1\}$ and infinitesimal parameter matrix

$$\begin{bmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{bmatrix} = \begin{bmatrix} -m & m \\ \dfrac{m(1-p)}{p} & -\dfrac{m(1-p)}{p} \end{bmatrix} .$$

Set $P\{\theta_0=1\} = p$. Then θ_t is stationary, mean-square continuous, $E[\theta_t] = p$, and

$$R_\theta(s,t) \equiv E[\theta_s \theta_t] = p^2 + p(1-p)e^{-\frac{m}{p}|s-t|} .$$

Proof: Using some obvious notation

$$\begin{aligned} E[\theta_t] &= P\{\theta_t = 1\} \\ &= P_{01}(t)(1-p) + P_{11}(t)p \\ &= \left[p - pe^{-\frac{m}{p}t}\right](1-p) + \left[p + (1-p)e^{-\frac{m}{p}t}\right]p \\ &= p \ . \end{aligned}$$

Consider $s \leq t$. Then

$$\begin{aligned} R_\theta(s,t) &= P\{\theta_t=1, \theta_s=1\} \\ &= P\{\theta_{t-s}=1, \theta_0=1\} \\ &= P_{11}(t-s)p \\ &= \left[p + (1-p)e^{-\frac{m}{p}(t-s)}\right]p \ . \end{aligned}$$

Lemma 14: Suppose θ_t is the telegraph signal described above and let

$$A_t = \int_0^t c\, \theta_s\, ds \ .$$

Then $E[A_t] = pct$ and

$$R_A(s,t) = p^2c^2st + \frac{2c^2p^2(1-p)}{m} s \wedge t + \frac{c^2p^3(1-p)}{m^2}\left[e^{-\frac{m}{p}s} + e^{-\frac{m}{p}t} - e^{-\frac{m}{p}|s-t|} - 1\right].$$

Proof: We just note that θ_t is mean-square continuous so

$$R_A(s,t) = \int_0^s\int_0^t c^2 R_\theta(u,v)\,dv\,du \ .$$

Lemma 15: For A_t and θ_t as above and $t \in [0,T]$, $Var[A_t] \to 0$ as $m \to \infty$.

Proof:

$$Var[A_t] = 2\frac{c^2}{m}p^2(1-p)t + 2\frac{c^2p^3(1-p)}{m^2}\left[e^{-\frac{m}{p}t} - 1\right].$$

Lemma 16: $\mathcal{C}_{P_t}(c, k_0, V, \infty) = \mathcal{C}_{P_t}(c, k_0, \infty, \infty)$ for all $V > 0$.

Proof: Make explicit the dependence of θ_t on the parameter m by writing $\theta_t = \theta_t^{(m)}$. Then with the right choice of p [3],

$$\lim_{m\to\infty} \frac{1}{T} I^T[\theta^{(m)}, Y] = \mathcal{C}_{P_t}(c, k_0, \infty, \infty) \ .$$

For any $V > 0$, there is an m_0 such that $Var[A_t] \le V$ for all $m \ge m_0$. Thus $\mathcal{C}_{P_\epsilon}(c, k_0, V, \infty) = \mathcal{C}_{P_\epsilon}(c, k_0, \infty, \infty)$.

Theorem 17: Let $\Lambda_t = \lambda$ be a real constant. Then $\mathcal{C}(\infty,\infty,\infty,P) = \infty$.

Proof: Choose $(k_0, V) \in R$. Then

$$\begin{aligned}
\mathcal{C}(\infty,\infty,\infty,P) &\ge \mathcal{C}_{P_\epsilon}(\infty,\infty,\infty,P) \\
&\ge \mathcal{C}_{P_\epsilon}(\infty,k_0,V,\infty) \\
&\ge \lim_{c \to \infty} \mathcal{C}_{P_\epsilon}(c, k_0, V, \infty) \\
&= \lim_{c \to \infty} \mathcal{C}_{P_\epsilon}(c, k_0, \infty, \infty) \\
&= \lim_{c \to \infty} \mathcal{C}(c, k_0, \infty, \infty) \\
&= \lim_{c \to \infty} k_0 \frac{\phi(c,\lambda)}{c} - \phi(k_0,\lambda) \\
&= \lim_{c \to \infty} k_0 [\ln(k(\lambda)+\lambda) + 1] - \phi(k_0,\lambda) \\
&\ge \lim_{c \to \infty} k_0 [\ln(c/e + \lambda) + 1] - \phi(k_0,\lambda) \\
&= \infty .
\end{aligned}$$

The above approach shows as well that $\mathcal{C}(\infty, k_0, V, P) = \infty$ for all $k_0 > 0$, $V > 0$, $P > 0$.

To further indicate the importance of the peak constraint, recall the following capacity result from [1]: let the noise intensity be nonrandom and suppose $0 \le \chi_t \le c(t)$ where $c(t)$ is chosen freely subject only to

$$\frac{1}{T}\int_0^T c(t)dt \le P$$

for some given $P > 0$. Then $\mathcal{C} = P/e$. If this encoder constraint is reformulated to eliminate the peak constraint; i.e. the similar, but weaker, constraint

$$\frac{1}{T}\int_0^T \chi_t \, dt \le P$$

is imposed, then the capacity is infinite.

Thus, it appears that a peak constraint on the encoder intensity or its equivalent for the encoder compensator is needed - at least for the group of constraints considered here - to give a well-posed capacity problem.

VI. SUMMARY

Significant differences exist between Poisson channels with nonrandom noise intensity and those with stochastic noise intensity. In the nonrandom case it is known [1], [2] that causal feedback does not increase channel capacity. By contrast, we have shown that causal feedback does increase capacity (though it seems by only a small amount) when the noise intensity is random. Also, expressions are available [1] for the channel capacity in the nonrandom case for $T < \infty$ while, for stochastic noise intensity and $T < \infty$, only least upper bounds have so far been obtained. These differences arise out of the simplification of the expression for average mutual information that is possible in the case of nonrandom noise intensity. In the case of nonrandom noise intensity $\Lambda_t = \lambda(t)$, the expression for the channel information

$$I^T[\theta,Y] = E\left[\int_0^T (\phi(\chi_t,\tilde{\Lambda}_t) - \phi(\hat{\chi}_t,\hat{\Lambda}_t))dt\right] \tag{5}$$

simplifies to

$$I^T[\theta,Y] = E\left[\int_0^T (\phi(\chi_t,\lambda(t)) - \phi(\hat{\chi}_t,\lambda(t)))dt\right] .$$

For stochastic noise intensity $\hat{\Lambda}_t$ and $\tilde{\Lambda}_t$ remain, making (5) relatively intractable. Only for a certain class of noise intensities can much headway be made. For $\Lambda_t \in \Phi_M \cup \Phi_B$, it was shown that for the Poisson channel with feedback

$$\lim_{T\to\infty} C = E[C(\Lambda_t)]$$

while, without feedback,

$$\lim_{T\to\infty} C = \max_{\frac{1}{e}\le p \le \frac{1}{2}} \left(pE[\phi(c,\Lambda_t)] - E[\phi(pc,\Lambda_t)]\right) .$$

ACKNOWLEDGEMENT

It is a pleasure to thank Prof. C.R. Baker for his patient encouragement and interest in this effort.

VII. References

[1] M.R. Frey, "Capacity of the Poisson Channel with Time-Varying Noise Intensity and Jamming," LISS Technical Report, No. 25, Dept. of Statistics, Univ. of N.C., May 1988.

[2] Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probabilty and Applications*, Vol. 23, 1974.

[3] M. Davis, "Capacity and cut-off rate for Poisson-type channels," *IEEE Transactions on Information Theory*, Vol. 26, No. 6, pp. 710-715, 1980.

[4] D.L. Snyder, *Random Point Processes*, John Wiley & Sons, Inc., 1975.

[5] M.S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco, 1964.

[6] R. Liptser and A. Shiryayev, *Statistics of Random Processes, Vol. II, Applications*, Springer-Verlag, New York, 1977.

[7] A. Segall and T. Kailath, "The Modeling of Randomly Modulated Jump Processes," *IEEE Transactions on Information Theory*, Vol. 21, No. 2, pp. 135-143, 1975.