AD-A235 544

Information Capacity of the Poisson Channel
with Mean-Square-Constrained Encoder Intensity

Michael R. Frey

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

LISS 43

27 April 1990

DTIC
ELECTE
MAY1 3 1991
S E D

Abstract - A variety of results are known for the information capacity of the Poisson channel with a peak constraint $0 \leq \chi_t \leq c$ imposed on the encoder intensity χ_t. Certain of these results are shown to carry over in some form to the case of mean-square-constrained encoding intensity $E[\chi_t^2] \leq P^2$. "On-off keyed" encoder intensity is considered. All results are given for general finite channel base measure.

91

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	Approved for Public Release: Distribution Unlimited

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

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) Information Capacity of the Poisson Channel[*]

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

13a. TYPE OF REPORT TECHNICAL	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) 1990 April	15. PAGE COUNT 14

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Poisson channel, Shannon theory, channel capacity

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

A variety of results are known for the information capacity of the Poisson channel with a peak constraint $0 \le \chi_t \le c$ imposed on the encoder intensity χ_t. Certain of these results are shown to carry over in some form to the case of mean-square-constrained encoding intensity $E[\chi_t^2] \le P^2$. "On-off keyed" encoder intensity is considered. All results are given for general finite channel base measure.

TITLE CONT.: with Mean-Square-Constrained Encoder Intensity.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED XX SAME AS RPT. XX DTIC USERS ☐	

22a. NAME OF RESPONSIBLE INDIVIDUAL C.R. Baker	22b. TELEPHONE NUMBER (Include Area Code) (919) 962-2189	22c. OFFICE SYMBOL

DD FORM 1473, 83 APR   EDITION OF 1 JAN 73 IS OBSOLETE.

INTRODUCTION

The Poisson channel model, sometimes called the Poisson-type point process channel or the direct detection photon channel, models optical communications systems as described in [2], [6], [9], [15]. The Poisson channel is a continuous-time additive noise channel with output $Y_t = X_t + N_t$ where $N = \{N_t\}_{0 \leq t \leq T}$ is the channel noise and $X = \{X_t\}_{0 \leq t \leq T}$ is the transmitted signal into which is encoded the message $\theta = \{\theta_t\}_{0 \leq t \leq T}$. All processes in the channel model are defined on a common probability space (Ω, F, P). We write $I\!\!F^\theta$ for the natural history of θ, $I\!\!F^Y$ for the natural history of Y, etc. By history we mean a nondecreasing sequence of σ-algebras. The natural history of a process Z is $I\!\!F^Z$ where $I\!\!F_t^Z = \sigma[Z_s, 0 \leq s \leq t]$.

NOISE INTENSITY, $\lambda(t)$

MESSAGE, θ_t ⟶ ENCODER INTENSITY, $\chi_t(\theta, Y)$ ⟶ ⊕ ⟶ CHANNEL OUTPUT INTENSITY, $\eta_t = \chi_t + \lambda(t)$ $(Y_t = X_t + N_t)$
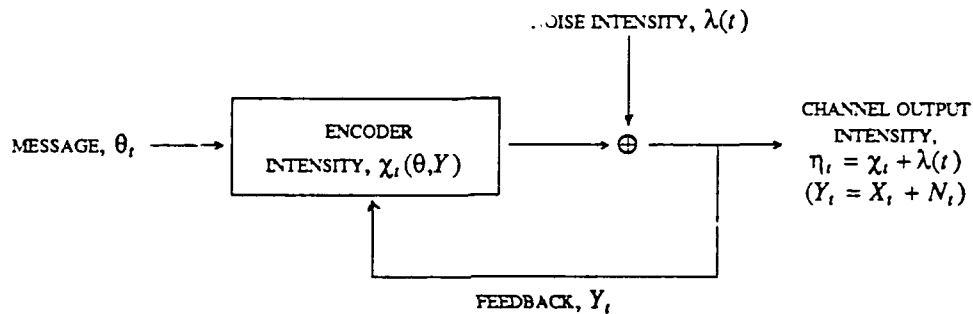
FEEDBACK, Y_t

Figure 1. Poisson channel model.

In the Poisson channel both X and N are Poisson-type point processes [11]. Thus, X and N have respective compensating measures

$$A(E) = \int_E \chi_t b(dt), \quad B(E) = \int_E \lambda(t) b(dt) \tag{1}$$

for all $E \in I\!\!B[0,T]$. b is called the channel base measure and is assumed to be finite; $b_T < \infty$ where $b_T \equiv b([0,T])/T$. The encoding intensity χ_t is required to be $I\!\!F^\theta \vee I\!\!F^Y$-predictable; this allows nonanticipative message encoding and causal, noiseless, instantaneous feedback from the channel output. The noise intensity $\lambda(t)$ is assumed to be nonrandom. Hence the channel noise N_t is a nonhomogeneous Poisson process. The channel output $Y_t = X_t + N_t$ is the sum of two Poisson-type point processes. Thus it is also a Poisson-type point process with intensity $\eta_t = \chi_t + \lambda(t)$ where η_t is predictable with respect to the global history $I\!\!F^\theta \vee I\!\!F^Y$. Poisson-type point process intensities are, by definition, nonnegative. Within the context of optical communications [2], [15], $\lambda(t)$ (see Figure 1) represents, nominally, noise due to background radiation as seen by the receiver. We make the usual assumption that the message and noise are independent, i.e., that the histories $I\!\!F^\theta$ and $I\!\!F^N$ are independent.

Existence and uniqueness of the compensating measures A and B in the Poisson channel model can be established by means of the Doob-Meyer submartingale decomposition [2], [11], [14]. X and N are each submartingales, therefore each of them has associated with it a unique predictable increasing process, A_t and B_t, respectively, such that

$$X_t - A_t, \quad N_t - B_t$$

are each martingales. Making the identifications $A([0,t])=A_t$ and $B([0,t])=B_t$, the compensating measures are also seen to exist uniquely. Alternatively, uniqueness and existence of the measures A and B can be established using projection methods [5].

A notable feature of the Poisson channel model is the presence of two different sources of noise. Besides the channel noise represented by N_t, there is an encoding noise inherent to the channel. Encoding noise arises because the message θ is encoded indirectly into X via the intensity $\chi = \chi(\theta, Y)$. The path of X is influenced by χ_t and, also, by the innovation martingale $m_t = X_t - A_t$ deriving from the Doob-Meyer decomposition of X. Thus, a trajectory of X over a finite interval $[0,T]$ is insufficient to recover χ_t even when the noise intensity $\lambda(t)$ is identically zero. Hence, one speaks of both channel noise and encoder noise in the Poisson channel.

Information capacity is defined in terms of the average mutual information $I^T[\theta, Y]$ in the message and channel output processes, θ and Y over the interval $[0,T]$. 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 [13]

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

provided $\mu_{\theta Y} \ll \mu_{\theta \times Y}$; otherwise $I^T[\theta, Y] = \infty$. Expressions exist for the average mutual information over the interval $[0,T]$ in the Poisson channel with base measure b and channel output intensity η_t. Define

$$I_1 \equiv E\left[\int_0^T (\eta_t \ln \eta_t - \hat{\eta}_t \ln \hat{\eta}_t) b(dt)\right] \tag{2}$$

where $\hat{\eta}_t = E[\eta_t | F_t^Y]$. Note that, in the terminology of Boel, Varaiya, and Wong [1], $\hat{\eta}_t$ is the intrinsic local description of the channel output process Y whereas η_t is an extrinsic local description (with respect to the history $F^\theta \vee F^Y$.) According to Liptser and Shiryayev [11], $I^T[\theta, Y] = I_1$ provided $I_1 < \infty$ (and, as a consequence, $\mu_{\theta Y} \ll \mu_{\theta \times Y}$ for $I_1 < \infty$.) A useful equivalent expression for the channel information is

$$I_2 \equiv E\left[\int_0^T (\eta_t \ln \eta_t - \eta_t \ln \hat{\eta}_t) b(dt)\right]. \tag{3}$$

We have $I_2 = I_1$ so that $I^T[\theta, Y] = I_2$ if $I_2 < \infty$. The channel information capacity is

$$C_{\text{DPO}} = \sup_\theta \sup_\chi \frac{1}{T} I^T[\theta, Y]$$

where θ is any jointly measurable process defined over the interval $[0,T]$ and $\chi = \chi(\theta, Y)$ is any $F^\theta \vee F^Y$-predictable mapping.

The information capacity of the Poisson channel was first found by Kabanov [9] for the case of a peak-constrained encoder intensity $0 \le \chi_t \le c$ and constant noise intensity $\lambda(t) = \lambda$. Considering only Lebesgue channel base measure, he showed that

$$C = C(\lambda, c) \tag{4}$$

where

$$C(x, y) = \frac{x}{e}\left[1 + \frac{y}{x}\right]^{1+x/y} - x\left[1 + \frac{x}{y}\right]\ln\left[1 + \frac{y}{x}\right]. \tag{5}$$

The approach taken by Kabanov was adapted by Frey [6] to treat the case of time-varying noise intensity $\lambda(t)$ and time-varying peak constraint $0 \le \chi_t \le c(t)$. For time-varying channel parameters $\lambda(t)$ and $c(t)$ and finite channel base measure b, we have

$$C = \frac{1}{T} \int_0^T C(\lambda(t), c(t)) b(dt).$$

In other work along these lines, Davis [4] treated the case $\lambda(t) = \lambda$, $c(t) = c$, and Lebesgue b in which an additional average constraint

$$\frac{1}{T} \int_0^T E[\chi_t] dt \leq k$$

is imposed on the encoder intensity and showed how the capacity is modified. Recently, Wyner [16] showed that the coding capacity equals the information capacity for the case considered by Davis. Wyner also found an analytic expression for the channel error exponent in this case. An earlier contribution in this area is that of Massey [12].

In this paper we consider a mean-square constraint

$$E[\chi_t^2] \leq P^2 \tag{6}$$

$(P \geq 0)$ on the encoder intensity. This constraint is sufficient for finite information capacity. In fact, one has that

$$I_2 \leq E\left[\int_0^T (\chi_t \ln \chi_t - \chi_t \ln \hat{\chi}_t) b(dt)\right] = E\left[\int_0^T (\chi_t \ln \chi_t - \hat{\chi}_t \ln \hat{\chi}_t) b(dt)\right] \leq b([0,T]) \left[P^2 + \frac{1}{e}\right] < \infty$$

which, as already observed, is sufficient for $I^T[\theta, Y]$ to be expressible by either I_1 or I_2.

A mean-square constraint such as (6) or the similar constraints

$$\frac{1}{T} \int_0^T E[\chi_t^2] b(dt) \leq P^2, \quad E[\chi_t^2] \leq P^2(t)$$

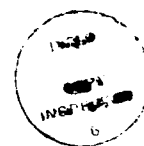have an intuitive power/energy interpretation. Also, they recall the mean-square constraint

$$\frac{1}{T} \int_0^T E[\phi_t^2] dt \leq P^2$$

appearing in Kadota, Zakai, and Ziv's treatment [10] of the additive white Gaussian noise channel

$$Y_t = \int_0^t \phi_t dt + W_t$$

where W_t is Wiener noise. These considerations and the fact that (6) is sufficient for finite capacity brought us to obtain the results we now present for the Poisson channel information capacity.

INFORMATION CAPACITY

We will assume without further mention that $b_T = 1$ as in the particular case of Lebesgue base measure. This entails no real loss of generality and clarifies the presentation. Four theorems are given. Theorem 1 considers the special case of the Poisson channel with zero noise intensity and states that the channel information capacity is

$$C = \frac{2}{e} P$$

for a mean-square constrained encoder intensity as in (6). Theorem 2 takes up the case of general $\lambda(t)$ and gives an expression for the capacity for the case in which the encoder intensity is restricted to only two values - the "on-off keying" case. Theorem 3 gives upper and lower bounds on the capacity for the general case in which "on-off keying" is not necessarily stipulated. Theorem 4 also treats $\lambda(t) \geq 0$ but with a formulation of the mean-square constraint on the encoder intensity which allows one to give an exact expression for the capacity. We conclude with a conjecture and some comments regarding jamming and coding capacity.

The following lemmas set the stage for Theorem 1.

Lemma 1: Let D be a Borel subset of $\mathbb{R}$ and define A_D to be the class of random variables with range in D. Let f be a real function and suppose f^{-1}, the inverse of f, exists on D. For $P \in D$, define $A = \{X \in A_D : E[f(X)] \leq f(P)\}$. Let g be a real function such that $g \circ f^{-1}$ is concave and nondecreasing. Then

$$\max_{X \in A} E[g(X)] = g(P).$$

Proof: Define $h = g \circ f^{-1}$. Then

$$E[g(X)] = E[h(f(X))] \leq h(E[f(X)]) \leq h(f(P)) = g(P).$$

Let $X = P$. Then $X \in A$ and $E[g(X)] = g(P)$. The result follows.

Lemma 2: Let $A = \{X \geq 0 : E[X^2] \leq P^2\}$ be the class of nonnegative random variables with constrained second moment. Then

$$\max_{X \in A} E[X \ln X] = \gamma(P) \tag{7}$$

where

$$\gamma(P) = \begin{cases} P \ln P, & P > e \\ \dfrac{P^2}{e}, & P \leq e \end{cases}.$$

Proof: $h(x) = \gamma(\sqrt{x})$ is concave and increasing for $x \geq 0$ so, by the previous lemma,

$$E[X \ln X] \leq E[\gamma(X)] \leq \gamma(P).$$

Therefore

$$\sup_{X \in A} E[X \ln X] \leq \gamma(P).$$

Suppose $P > e$. Choose $X = P$. Then $E[X \ln X] = \gamma(P)$. Suppose $P \le e$. Choose

$$X = \begin{cases} 0, & \text{w.p. } 1 - \dfrac{P^2}{e^2} \\ e, & \text{w.p. } \dfrac{P^2}{e^2} \end{cases}.$$

Then, again, $E[X \ln X] = (P^2/e^2) e \ln e = \gamma(P)$. Hence, for all $P \ge 0$, a random variable $X \in A$ exists such that $E[X \ln X] = \gamma(P)$. (7) is proved.

Lemma 3: For the Poisson channel with finite base measure b and mean-square-constrained encoding intensity as in (6),

$$C \le \left[\gamma(P) + \frac{1}{e} \right].$$

When there is no noise intensity present ($\lambda(t) = 0$),

$$C \ge \frac{2}{e} P.$$

Proof: The upper bound follows from Lemma 2 and an application of Jensen's inequality,

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

To establish the lower bound, one uses Brémaud's averaging principle [3] with a sequence of stationary random telegraph signal [8] message processes $\{\theta_t^{(m)}, m = 1,2,\ldots\}$ having common state space $\{0, eP\}$ and generator mA where A is the matrix

$$A = \begin{bmatrix} -1 & 1 \\ \dfrac{1-p}{p} & -\dfrac{1-p}{p} \end{bmatrix} \tag{8}$$

and $p = e^{-2}$. In the channel with these message processes, $E[w(\chi_t^{(m)})] = E[w(\theta_t^{(m)})] = (P/e) \ln(Pe)$ and $E[\chi_t^{(m)}] = E[\theta_t^{(m)}] = P/e$. Thus $C \ge (P/e) \ln(Pe) - (P/e) \ln(P/e) = 2P/e$.

Theorem 1: Consider the Poisson channel with finite base measure b and zero noise intensity. Suppose the encoding intensity χ_t satisfies the mean-square constraint $E[\chi_t^2] \le P^2$. Then the channel information capacity is $C = 2P/e$.

Proof: Let $I[X] = E[X \ln X] - E[X] \ln E[X]$ and define $B(P)$ for each $P \ge 0$ to be the class of non-negative random variables X such that $E[X^2] \le P^2$. For each $X \in B(P)$, we can write $X = PZ$ for some $Z \in B(1)$. For $X = PZ$, $I[X] = PI[Z]$. Thus

$$C = \sup_{X \in B(P)} I[X] = \sup_{PZ \in B(P)} I[PZ] = P \sup_{Z \in B(1)} I[Z].$$

We observe that the upper and lower bounds on capacity given in Lemma 3 coincide for $P = 1$. Therefore

$$\sup_{Z \in B(1)} I[Z] = \frac{2}{e}.$$

Thus $C = 2P/e$ and the proof is complete.

Corollary: Consider the Poisson channel with finite base measure b, zero noise intensity, and encoder intensity satisfying

$$E[\chi_t^2] \le P^2(t), \quad 0 \le t \le T$$

where $P(t)$ is b-integrable. Then the channel capacity is

$$\frac{1}{T}\frac{2}{e}\int_0^T P(t)b(dt).$$

Proof: If $P(t)$ is b-integrable then $I_1 < \infty$ and the usual expressions (2) and (3) for channel information can be used. Then the result follows from approximating $P(t)$ by a simple function $P_n(t)$ and passing to the limit as $P_n(t) \to P(t)$ pointwise.

Corollary: The results of Theorem 1 are unchanged by substituting

$$\frac{1}{T}\int_0^T E[\chi_t^2]\,dt \le P^2$$

for the stronger constraint $E[\chi_t^2] \le P^2$ used in Theorem 1.

Proof: Write $E[\chi_t^2] = m(t)$. From the previous corollary

$$C = \sup_{m(t) \in \Gamma} \frac{1}{T}\frac{2}{e}\int_0^T \sqrt{m(t)}b(dt) \tag{9}$$

where Γ is the class of nonnegative functions

$$\Gamma = \{m(t) : m(t) \ge 0, \ \frac{1}{T}\int_0^T m(t)b(dt) \le P^2\}.$$

The square root function is nonengative, increasing, and concave so by a "waterpouring" argument [6] $m(t)$ is optimally chosen in (9) to be the constant $m(t) = P^2$. Thus $C = 2P/e$ and the corollary is proved.

We now turn to the case of nonzero noise intensity. This case is not as tractable as the case of zero noise intensity treated in Theorem 1 and its corollaries and at present, with one exception, only bounds and asymptotic results can be given. The exception referred to is the "on-off keying" case - the case in which the encoder intensity switches between only two values (neither of which are necessarily zero). For "on-off keyed" encoder intensities, we can and do (Theorem 2) give an expression for the capacity. It is clear that when the encoder intensity is restricted to only two values, then one of these values should be chosen to be zero (to maximize the channel information rate.) Transitions of the encoder intensity between its zero value and its second (positive) value might typically be accomplished by turning on and off a power source. Hence the nomenclature "on-off keying."

For $x \ge 0$, define

$$k(x) = \frac{x}{e}\left[1 + \frac{1}{x}\right]^{x+1} - x. \tag{10}$$

Also, for channel parameters λ and P, let

$$A = \{a \ge 0: a^2 k(\lambda/a) \ge P^2\}.$$

The function $f(a) = a^2 k(\lambda/a)$ is increasing. Therefore A is a semiinfinte interval of the form $[a_o, \infty)$ where $a_o^2 k(\lambda/a_o) = P^2$. Theorem 2 follows readily from the following lemma.

Lemma 1: Let $J[X] = E[(X + \lambda)\ln(X + \lambda)] - (E[X] + \lambda)\ln(E[X] + \lambda)$ and define $\mathcal{B}_P$ for each $P > 0$ to be the class of nonnegative random variables X having two possible values and such that $E[X^2] \le P^2$.

Then

$$\sup_{X \in B_P} I[X] = \max_{a \in A} \frac{P^2}{a} \ln \frac{ak(\lambda/a)+\lambda}{P^2 k(\lambda a/P^2)/a +\lambda} . \qquad (11)$$

Proof: For $P = 0$, the RHS of (11) is zero. Therefore, in this case (11) is true. Thus we only address $P > 0$. Let B_{P0} be the class of nonnegative random variables X of the form

$$X = \begin{cases} a, & p \\ 0, & 1-p \end{cases} . \qquad (12)$$

Then as was observed above

$$\sup_{X \in B_P} I[X] = \sup_{X \in B_{P0}} I[X].$$

Hence we need to show that

$$\sup_{X \in B_{P0}} I[X] = \max_{a \in A} \frac{P^2}{a} \ln \frac{ak(\lambda/a)+\lambda}{P^2 k(\lambda a/P^2)/a +\lambda} . \qquad (13)$$

The proof (13) is conducted in 3 steps.

Step 1: For $X \in B_{P0}$ as in (12) with $0 \le p \le 1$, suppose $a < P$ and write

$$Y = \begin{cases} a+\delta, & p \\ 0, & 1-p \end{cases} .$$

$Y \in B_{P0}$ for $\delta \in [0, P-a]$. Define $I_\delta = I[Y]$. We have

$$\frac{\partial I_\delta}{\partial \delta} = p \ln \frac{a+\delta+\lambda}{p(a+\delta)+\lambda} .$$

$\frac{\partial I_\delta}{\partial \delta} \ge 0$ for $\delta \in [0, P-a]$ so $I[X]=I_0 \le I_{P-a}$ and

$$\sup_{X \in B_{P0}} I[X] = \sup_{X \in B_{PS}} I[X] \qquad (14)$$

where B_{PS} is the class of random variables $X \in B_{P0}$ as in (12) such that $a \ge P$.

Step 2: Fix $P > 0$. Suppose $X \in B_{PS}$ with a fixed ($P \le a$). Then, as a consequence of the inequality $E[X^2] \le P^2$, $p = P\{X = a\}$ is restricted to the range

$$0 \le p \le \frac{P^2}{a^2} . \qquad (15)$$

Let us identify the value of p which maximizes $I[X]$. Define

$$m = \frac{\phi(a,\lambda)}{a}, \quad b = \lambda \ln \lambda$$

where $\phi(x,y) = (x+y)\ln(x+y) - y \ln y$. Then, for $X \in B_{PS}$,

$$I[X] = E[(X+\lambda)\ln(X+\lambda)] - (E[X]+\lambda)\ln(E[X]+\lambda)$$

$$= E[(X+\lambda)\ln(X+\lambda) - mX - b] + mE[X] + b$$

$$- (E[X]+\lambda)\ln(E[X]+\lambda)$$

$$= mE[X] + b - (E[X]+\lambda)\ln(E[X]+\lambda).$$

Define $I(\lambda) = mx + b - (x+\lambda)\ln(x+\lambda)$. For all $x \ge 0$,

$$\frac{dI}{dx} = m - 1 - \ln(x+\lambda), \quad \frac{d^2I}{dx^2} = -\frac{1}{(x+\lambda)} \le 0.$$

Thus $I(x)$ has a unique maximum at $x = e^{m-1} - \lambda$. Using $k(x)$ defined in (10), $p = E[X]/a$, and the identity

$$\phi(a,\lambda) = a \ln(ak(\lambda/a) + \lambda) + a ,$$

we have that, subject to the constraint in (15), the choice of p which maximizes $I[X]$ is

$$p = k(\lambda/a) \wedge \frac{P^2}{a^2} . \tag{16}$$

Let $\mathbb{B}_Z$ be the random variables $X \in \mathbb{B}_{PS}$ with p given as in (16). Then

$$\sup_{X \in \mathbb{B}_{PS}} I[X] = \sup_{X \in \mathbb{B}_Z} I[X] . \tag{17}$$

Step 3: For $X \in \mathbb{B}_Z$, $I[X]$ assumes one of two forms depending on p in (16). After some algebra, one has

$$I[X] = \begin{cases} ak(\lambda/a) \ln \dfrac{ak(\lambda/a) + \lambda}{ak(\lambda/a)k(\lambda/(ak(\lambda/a))) + \lambda} , & a^2 k(\lambda/a) \leq P^2 \\[3mm] \dfrac{P^2}{a} \ln \dfrac{ak(\lambda/a) + \lambda}{P^2 k(\lambda/(P^2/a))/a + \lambda} , & a^2 k(\lambda/a) > P^2 \end{cases} \tag{18}$$

Consider the first case in (18) in which $a^2 k(\lambda/a) \leq P^2$. Letting $\alpha = a/\lambda$, we have $I[X]/\lambda = G(\alpha k(1/\alpha))$ where

$$G(s) = s \ln \frac{s+1}{sk(1/s)+1} .$$

Now

$$\frac{d}{d\alpha} \alpha k(1/\alpha) = \frac{(1+\alpha)^{1+1/\alpha}}{e\,\alpha^2} \left[\alpha - \ln(1+\alpha) \right] \geq 0$$

for all $\alpha > 0$. Thus $\alpha k(1/\alpha)$ is a nondecreasing function of α. Then, also, $G(s)$ is a nondecreasing function of s. Therefore, for $a^2 k(\lambda/a) \leq P^2$ the maximum value of $I[X]$ is found at a satisfying $a^2 k(\lambda/a) = P^2$. Hence

$$\sup_{X \in \mathbb{B}_{PS}} I[X] = \max_{a \in A} \frac{P^2}{a} \ln \frac{ak(\lambda/a) + \lambda}{P^2 k(\lambda a/P^2)/a + \lambda} . \tag{19}$$

Therefore, by (14), (17), and (19), our desired result (13) is obtained.

Theorem 2: For the Poisson channel with noise intensity λ, mean square constraint parameter P, finite b, and "on-off keyed" encoder intensity, the information capacity is $C = D_o(\lambda, P)$ where

$$D_o(\lambda,P) = \max_{a \in A} \frac{P^2}{a} \ln \frac{ak(\lambda/a) + \lambda}{P^2 k(\lambda a/P^2)/a + \lambda} .$$

Also, as $P \to \infty$,

$$D_o(\lambda,P) - \frac{2}{e} P \tag{20}$$

for any fixed λ.

Proof: By Jensen's inequality,

$$C \leq \sup_{X \in \mathbb{B}_P} I[X] .$$

By the usual choice of sequence of random telegraph signals $\{\theta^{(m)}\}$ and using Brémaud's averaging principle in taking the limit as $m \to \infty$, we find that

$$\mathbf{C} \geq \sup_{X \in \mathbb{B}_P} I[X].$$

By Lemma 4, then, $\mathbf{C} = D_o(\lambda, P)$.

To prove (20), first observe that

$$D_o(0, P) = \max_{a \in A} \frac{P^2}{a} \ln \frac{a/e}{P^2/e} = \frac{2}{e} P.$$

Thus $D_o(\lambda, P) \leq 2P/e$. Consider $a = eP$. We have $a \in A$ so

$$D_o(\lambda, P) \geq \frac{P^2}{eP} \ln \frac{ePk(\lambda/(eP)) + \lambda}{P^2 k(\lambda eP/P^2)/(eP) + \lambda}$$

$$= \lambda \frac{2}{e} Q + \lambda \frac{Q}{e} \ln \frac{k(1/(eQ)) + 1/(eQ)}{k(e/Q) + e/Q}$$

where for convenience we have used $Q = P/\lambda$. Expanding $k(x)$ as in [4, Lemma 2] and using the logarithm expansion $\ln(1 + x) = x + o(x)$ for $x \to 0$, we obtain

$$D_o(\lambda, P) \geq \lambda \frac{2}{e} Q + \lambda \frac{Q}{e} \ln \frac{1 + \dfrac{23}{12} \dfrac{1}{eQ} + o(1/Q)}{1 + \dfrac{23}{12} \dfrac{e}{Q} + o(1/Q)}$$

$$= \lambda \frac{2}{e} Q + \lambda \frac{Q}{e} \left[\frac{23}{12} \frac{1}{eQ} - \frac{23}{12} \frac{e}{Q} + o(1/Q) \right]$$

$$= \lambda \frac{2}{e} Q + \frac{23}{12} \lambda \frac{e - 1}{e^2} + o(1).$$

(20) follows and the proof is complete.

In terms of the dimensionless quantities

$$Q = \frac{P}{\lambda}, \quad \alpha = \frac{a}{\lambda},$$

we have $D_o(\lambda, P) = \lambda D_o(1, Q)$ where

$$D_o(1, Q) = \sup_{\alpha^2 k(1/\alpha) \leq Q^2} \frac{Q^2}{\alpha} \ln \frac{\alpha k(1/\alpha) + 1}{Q^2 k(\alpha/Q^2)/\alpha + 1}.$$

The shape of the curve defined by $D_o(1, Q)$ is remarkably similar to that defined by $C(1, Q)$. Both are increasing concave functions. Kabonov found in [8] that $C(1, Q) \sim Q/e$. Thus the two curves each have linear asymptotes. These comments are illustrated in Figure 2.
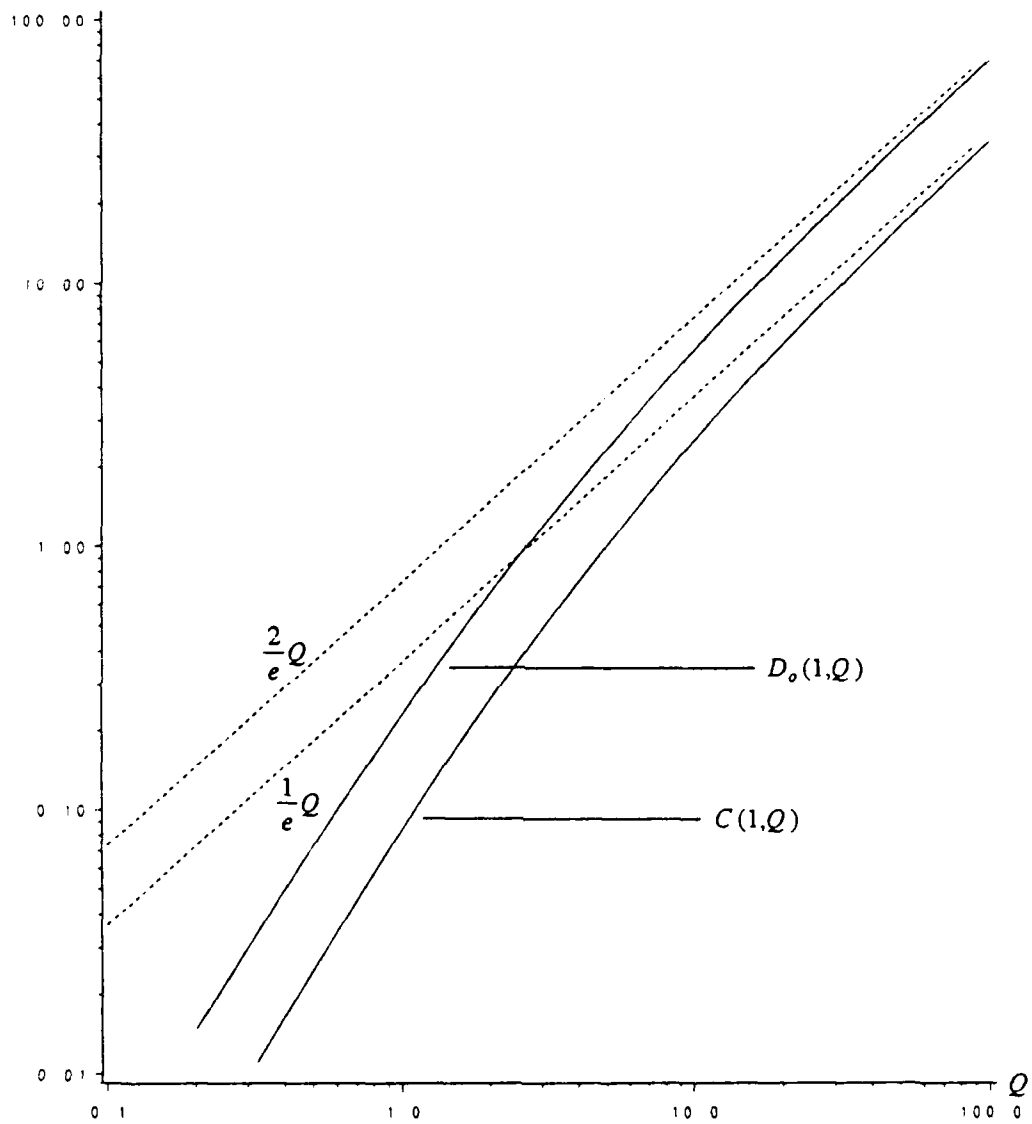
Figure 2. Log-log plots of $D_o(1,Q)$ and $C(1\ Q)$ and their large-Q linear
asymptotes in terms of the dimensionless quantity $Q = P/\lambda$.

It will be useful to have a notation for the information capacity of the Poisson channel with mean-square constraint. Define $D(\lambda, P)$ to be the capacity of the channel with constant noise intensity λ and mean-square-constrained encoder intensity $E[\chi_t^2] \le P^2$. In this notation Theorem 1 states that $D(0, P) = 2P/e$. Analogous to the fact that $C(\lambda z, cz) = zC(\lambda, c)$ for Kabanov's capacity function stated in (5), we have

$$D(\lambda z, Pz) = zD(\lambda, P).\tag{21}$$

To see this we use $w(x) \equiv x \ln x$ and write

$$D(\lambda z, Pz) = \sup_\theta \sup_{E[\chi_t^2] \le (zP)^2} E\left[\int_0^T \left[w(\chi_t + z\lambda) - w(\hat{\chi}_t + z\lambda)\right] b(dt)\right]$$

$$= \sup_\theta \sup_{E[(\chi_t/z)^2] \le P^2} z\left\{E\left[\int_0^T \left[w(\chi_t/z + \lambda) - w(\hat{\chi}_t/z + \lambda)\right] b(dt)\right]\right.$$

$$\left. + E\left[\int_0^T \left[(\chi_t/z + \lambda)\ln\lambda - (\hat{\chi}_t/z + \lambda)\ln\lambda\right] b(dt)\right]\right\}.$$

The second term within the braces is zero so (21) follows.

From (21) we have $D(\lambda, P) = \lambda D(1, P/\lambda)$ for $P > 0$. Thus, determining $D(\lambda, P)$ is reduced to the problem of finding the one-parameter function $D(1, \cdot)$.

Theorem 3: For the Poisson channel with noise intensity λ, mean-square constraint parameter P, and finite b, the information capacity $C = D(\lambda, P)$ admits the bounds

$$D_o(\lambda, P) \le D(\lambda, P) \le \frac{2}{e}P.\tag{22}$$

Also, as $P \to \infty$,

$$D(\lambda, P) \sim \frac{2}{e}P\tag{23}$$

for any fixed λ.

Proof: The first inequality in (22) follows from Theorem 2 and the second from Theorem 1. (23) follows directly from (21) in Theorem 2.

The next theorem is related to and motivated by results obtained by Davis for polarization modulation and by Frey [6] for peak-constrained encoder intensity. Davis [4] showed that, when operating two orthogonally polarized, separately modulated Poisson channels, channel capacity is maximized when encoder intensity is not distributed over both channels but, instead, is concentrated solely in one channel. This was because of the convexity of the channel capacity function $C(x, y)$ in y. Frey [6] showed that for this same reason it is also better not to distribute encoder intensity over time but, rather, to concentrate it into as short a time interval as possible. This result, obtained for peak-constrained encoder intensity, is equally valid for mean-square-constrained encoder intensity. Thus consider the Poisson channel with continuous finite base measure b, nonrandom noise intensity $\lambda(t)$, and encoding intensity χ_t. Suppose the encoder intensity is mean-square-constrained $0 \le E[\chi_t^2] \le P^2(t)$ but that $P(t)$ is not some given function. Suppose, instead, that $P(t)$ may be chosen freely subject only to the constraint

$$\frac{1}{T}\int_0^T P(t)dt \le Q\tag{24}$$

for some given $Q > 0$. Then, in Theorem 4, the channel capacity is found to be

$$C = \frac{2Q}{e}.\tag{25}$$

Notice that $\lambda(t)$ is missing from this expression. In the proof of (25) the power $P(t)$ available to the encoder is concentrated into as short a time interval as possible to obtain a rate of average mutual information in the channel closer and closer to $2Q/e$. Concentration of the encoder intensity into a short time interval permits it to be very large without violating the constraint (24) on $P(t)$. By concentrating the encoder intensity into a short time interval, it can be made so large that it completely overshadows whatever noise intensity is present in the interval in which the encoding intensity is applied. In passing to the limit, the magnitude of the noise intensity becomes irrelevant.

Theorem 4: Suppose the noise and encoded message processes, N and X, in a Poisson channel have intensities $\lambda(t)$ and χ_t with respect to a finite continuous base measure b. Also, suppose the encoder intensity χ_t is $\mathbb{F}_t^\theta \sim \mathbb{F}_t^\gamma$-adapted and mean-square-constrained $0 \le E[\chi_t^2] \le P^2(t)$ and allow $P(t)$ to be chosen freely provided only that $P(t) \in \Gamma$ where Γ is the class of nonnegative functions

$$\Gamma = \{P(t) \ge 0: \frac{1}{T}\int_0^T P(t)b(dt) \le Q\}.$$

Then the channel capacity is $\mathbb{C} = \dfrac{2Q}{e}$.

Proof: By Corollary 1 of Theorem 1,

$$\mathbb{C} = \sup_{P \in \Gamma} \frac{1}{T}\int_0^T D(\lambda(t), P(t))b(dt).$$

Define

$$\Gamma_1 = \{P(t) \ge 0: \frac{1}{T}\int_0^T P(t)b(dt) = Q\}.$$

$D(x,y)$ is nondecreasing in y so

$$\mathbb{C} = \sup_{P \in \Gamma_1} \frac{1}{T}\int_0^T D(\lambda(t), P(t))b(dt).$$

$D(x,y)$ is nonincreasing in x and $D(0,y)=2y/e$ so

$$\mathbb{C} \le \sup_{P \in \Gamma_1} \frac{1}{T}\int_0^T D(0, P(t))b(dt) = \sup_{P \in \Gamma_1} \frac{1}{T}\int_0^T \frac{2P(t)}{e}b(dt) = \frac{2}{e}Q.$$

Next we show $\mathbb{C} \ge 2Q/e$ to complete the proof. Let $G = \{t \in [0,T]: \lambda(t) < L\}$ and choose L so that $b(G) > 0$. Define $\lambda_L(t) = L$ on G and $\lambda_L(t) = \infty$ elsewhere on $[0,T]$. Then $D(\lambda(t), P(t)) \ge D(\lambda_L(t), P(t))$. Therefore

$$\mathbb{C} \ge \sup_{P \in \Gamma_1} \frac{1}{T}\int_0^T D(\lambda_L(t), P(t))b(dt)$$

$$= \sup_{P \in \Gamma_1} \frac{1}{T}\int_G D(L, P(t))b(dt).$$

Let S be the set of all nonnegative b-measurable simple functions on $[0,T]$. Then

$$\mathbb{C} \ge \sup_{\bar{P} \in \Gamma_1 \cap S} \frac{1}{T}\int_G D(L, \bar{P}(t))b(dt)$$

$$= \sup_{M > P} \sup_{\substack{\bar{P} \in \Gamma_1 \cap S \\ \bar{P} \le M}} \frac{1}{T}\int_G D(L, \bar{P}(t))b(dt)$$

$$= \sup_{M > Q_1} \sup_{\substack{\bar{P} \in \Gamma_1 \cap S_1 \\ \bar{P} \le M}} \frac{1}{T}\int_G D(L, \bar{P}(t))b(dt)$$

where $Q_1 = QT'b(G)$ and $S_1 \subset S$ is the set of all functions in S which vanish outside G. For each $M > Q_1$, let $\tilde{P}_M(t) = M 1_A(t)$ where $A \subset G$ with $b(A) = QT/M$. $\tilde{P}_M \in \Gamma_1 \cap S_1$ so

$$C \geq \sup_{M > Q_1} \frac{1}{T} \int_G D(L, \tilde{P}_M(t))b(dt)$$

$$\geq \sup_{M > Q_1} \frac{1}{T} b(A) D(L, M)$$

$$= P \sup_{M > Q_1} \frac{1}{M} D(L, M).$$

We have

$$\sup_{M > Q_1} \frac{D(L, M)}{M} = \sup_{M > Q_1} D(L/M, 1) = D(0, 1) = 2/e$$

so the proof is complete.

Final comments: Much more could be said about the Poisson channel with the mean-square constraint considered here. Following Wyner [16], one could derive the channel coding capacity by considering the Poisson channel as a discrete binary Z-channel [7] and taking the appropriate limits. One would find that the coding capacity and the information capacity were equal. Also following Wyner, one might try to calculate the channel error exponent. Whether or not an analytic expression exists for the error exponent in the present case and what its form might be are unknown. Even to obtain only the cut-off rate [4], [12], [16] would be interesting. All these various results could be extended to the case of time-varying channel parameters $\lambda(t)$ and $P(t)$. Then, as in [6] for the peak-constrained Poisson channel with noise intensity $\lambda(t)$, the optimal jamming solution could be pursued. For the mean-square-constrained "on-off keyed" Poisson channel, the information capacity $D_o(\lambda, P)$ is nonnegative, decreasing, and convex in λ. Therefore, the optimal jamming intensity will be nonrandom and "waterfilling" [6]. If $D(\lambda, P)$ proves to be convex in λ then in this case too, one would obtain a "waterfilling" solution to the jamming problem.

In the peak-constrained Poisson channel the capacity with and without an "on-off keying" restriction on the encoder intensity is the same. Based on this, some computer calculations, and the fact that

$$D(0, P) = D_o(0, P) \left[= \frac{2}{e} P \right],$$

we conjecture that the same holds true for the Poisson channel with mean square constraint; i.e., $D(\lambda, P) = D_o(\lambda, P)$. Note that this amounts to showing that $D(1, P) = D_o(1, P)$ for all $P \geq 0$.

Finally, it is worth noting that the encoder intensity χ_t is $F^\theta \vee F^Y$-predictable in the Poisson channel model considered here. Thus our capacity results are all results for the Poisson channel with causal feedback. However, the possible presence of channel feedback was exploited in none of our proofs; in fact, implicitly or explicitly only the trivial encoding $\chi_t(\theta) = \theta_t$ is used in the proofs. Thus the capacity results presented in this paper are equally valid for the no-feedback Poisson channel with F^θ-predictable encoder intensity.

REFERENCES

[1]   R. Boel, P. Varaiya, and E. Wong, "Martingales on jump processes. II: Applications," *SIAM Jour. Control*, Vol. 13, No. 5, pp. 1022-1061, 1975.

[2]   P. Brémaud, *Point Processes and Queues*, Springer-Verlag, New York, 1981.

[3]   P. Brémaud, "An Averaging Principle for Filtering a Jump Process with Point Process Observations," *IEEE Trans. Info. Theory*, Vol. 34, No. 3, pp. 582-585, 1988.

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

[5]   C. Dellacherie, *Capacités et Processus Stochastiques*, Springer-Verlag, Berlin, 1972.

[6]   M.R. Frey, "Capacity of the Poisson Channel with Time-Varying Noise Intensity and Jamming," *LISS Tech. Report Series*, No. 25, Department of Statistics, University of North Carolina, May 1988.

[7]   S. W. Golomb, "The Limiting Behavior of the Z-Channel," *IEEE Trans. on Info. Theory*, Vol. 26, No. 3, p. 372, May 1980.

[8]   P. Hoel, S. Port, and C. Stone, *Introduction to Stochastic Processes*, Houghton Mifflin, Co., Boston, 1972.

[9]   Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probability and Applications*, Vol. 23, 1978.

[10] T.T. Kadota, M. Zakai, and J. Ziv, "Mutual Information of the White Gaussian Channel With and Without Feedback," *IEEE Trans. Info. Theory*, Vol. 17, No. 4, July 1971.

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

[12] J.L. Massey, "Capacity, Cutoff Rate, and Coding for a Direct-Detection Optical Channel," *IEEE Trans. Commun.*, Vol. 29, No. 11, Nov. 1981.

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

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

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

[16] A.D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel - Part I, II," *IEEE Trans. on Info. Theory*, Vol. 34, No. 6, November 1988.