34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Computing the Trajectory Mutual Information Between a Point Process and an Analog Stochastic Process

Syed Ahmed Pasha[1] and Victor Solo[2], Fellow IEEE

Abstract— **In a number of application areas such as neural coding there is interest in computing, from real data, the information flows between stochastic processes one of which is a point process. Of particular interest is the calculation of the trajectory (as opposed to marginal) mutual information between an observed point process which is influenced by an underlying but unobserved analog stochastic process i.e. a state. Using particle filtering we develop a model based trajectory mutual information calculation for apparently the first time.**

I. INTRODUCTION

Recently there has been substantial practical investigation of mutual information and the related concept of entropy in neuroscience to measure information flows between various signals of interest for systems observed through point processes [1], [2]. Information transmission in the brain is partly by spike trains (axon potentials) [1], [2] and it is hoped that by studying these information flows some understanding of how the brain codes information can be achieved [3]. Also such studies could help in the design of neural prosthetic devices [4].

The traditional measure of information flows between stochastic processes computes instantaneous mutual information [5]. We refer to the traditional measure as *marginal* mutual information. The problem of interest in many applications including optical detection [6] is to measure the information flows between trajectories of the processes. We refer to this as *trajectory* mutual information to emphasize the fact that we are measuring mutual information between trajectories of random processes and not just instantaneous (i.e. marginal) mutual information.

In the recent work [7], new formulae for mutual information have been developed based on the state dependent stochastic intensity function of the point process. In related literature based on the stochastic intensity, the formula for mutual information for the particular case of an observed point process and an unobserved analog state has been developed in the important work of [8]. The only previous attempt at numerically calculating the mutual information between a point process and a state by means of a joint stochastic model is due to [3]. However that calculation is based on a Gaussian approximation of the predicted and posterior probability densities of the current state conditional on point process observations.

The calculation of trajectory mutual information between an observed point process and an unobserved analog process requires estimates of the (unconditional) probability density of the unobserved process as it evolves. In this paper, using iterated expectation we rewrite the formula in terms of probability density of the analog process conditional on point process observations. Using particle filtering [9], [10], [11] and Monte Carlo simulations we present for the first time a numerical calculation of the trajectory mutual information.

In the remainder of the paper we define the stochastic conditional intensity function in Section II. We give the formula for trajectory mutual information and computation details in Section III. This is followed by a review of particle filtering for point process observations in Section IV. In Section V we discuss the simulation setup and results. A discussion of the findings is offered in Section VI. The paper concludes with some final remarks in Section VII.

Notation. In the sequel subscript k denotes discrete time and subscript (t) denotes continuous time.

II. POINT PROCESSES

We define the point process notation and review some definitions from stochastic point process theory.

The point process is characterized by the counting process $N_{(t)} = \#$ events up to time t. Sampling the process at infinitesimal intervals δ gives the discrete-time equivalent $N_k = N_{(k\delta)}$. $\delta N_{(t)} = \#$ events in the interval $[t, t+\delta)$ or equivalently in discrete-time $\delta N_k = \delta N_{(k\delta)}$. The counting path up to time k is given by the sequence of incremental counts, $N_0^k = (\delta N_0 = \delta n_0, \dots, \delta N_k = \delta n_k)$ and the associated sequence of random variables by $N_{0,k} = (\delta N_0, \dots, \delta N_k)$. By $a(\delta) = o(\delta)$ we mean $a(\delta)/\delta \to 0$ as $\delta \to 0$. For the observation interval $[0, T]$, $T = n\delta$.

For the counting process N_k the stochastic conditional intensity λ_k at time k is given by,

$$P(\delta N_k = 1 | N_0^{k-1}) = \lambda_k \delta + o(\delta).$$

For a formal definition of the stochastic intensity refer to [12], [13].

Under the *No Simultaneity* [7] condition (more commonly known by the less informative term of orderliness which just means no more than one event can occur in a small interval of time), N_k is conditionally a Bernoulli process, i.e.,

$$P(\delta N_k = 0 | N_0^{k-1}) = 1 - \lambda_k \delta + o(\delta).$$

The point processes of interest here are those which evolve under the influence of another stochastic process. Let $x_{(t)}$ denote such a process with sampled version $x_k = x_{(k\delta)}$. The sample path of the process up to time k is denoted by $X_1^k = (X_1 = x_1, \dots, X_k = x_k)$ and $X_{1,k} = (X_1, \dots, X_k)$.

[1]S.A. Pasha is with School of Electrical and Information Engineering, University of Sydney, NSW 2006, Australia ahmed.pasha at sydney.edu.au

[2]V. Solo is with the School of Electrical Engineering and Telecommunications, University of New South Wales, NSW 2052, Australia v.solo at unsw.edu.au

Suppose X_k is the state of an analog stochastic process at time k and the stochastic intensity depends on the underlying unobserved state. Then, the state dependent stochastic intensity is given by

$$P(\delta N_k = 1|X_k = x_k) = \lambda_{k,x_k}\delta + o(\delta),$$

and the No Simultaneity condition gives,

$$P(\delta N_k = 0|X_k = x_k) = 1 - \lambda_{k,x_k}\delta + o(\delta).$$

Furthermore,

$$\begin{aligned} P(\delta N_k = 1) &= \int P(\delta N_k = 1|x_k)p(x_k)d(x_k) \\ &= \beta_k\delta + o(\delta), \end{aligned}$$

where $\beta_k = E(\lambda_{k,x_k})$ defines the marginal rate function. Moreover, $P(\delta N_k = 0) = 1 - \beta_k\delta + o(\delta)$.

III. MUTUAL INFORMATION

We review some information-theoretic calculations pertaining to the concepts of marginal and trajectory mutual information.

In the discrete case, the entropy of a vector-valued random variable X with probability distribution $p(x) = P(X = x)$ is defined as [14], $H(X) = -\Sigma_x p(x) lnp(x)$. If y is a realization of a vector-valued random variable Y, then the stochastic conditional entropy is defined as [7],

$$H(X|Y = y) = -\Sigma_x p(x|y) lnp(x|y).$$

Taking expectation gives the average conditional entropy [14], $H(X|Y) = \Sigma_y P(Y = y)H(X|Y = y)$. For the bivariate random variable (X, Y), mutual information is defined as [14],

$$\mathcal{I}(X; Y) = H(X) + H(Y) - H(X, Y).$$

A. Marginal Mutual Information

Turning to the process case, traditional data analysis using mutual information measure association for $(X_{(t+u)}, Y_{(t)})$ at lag u. This involves brute force calculation based on the definition [1], [2], [5]. Following the approach in [7], we express the marginal mutual information at time t in terms of the underlying stochastic conditional intensity functions.

Let $(X_k, \delta N_k)$ denote the bivariate random variable at time k comprising of the underlying unobserved state and the incremental count. We follow the notation in [7] and subscript H and $\mathcal{I}$ by δ. The marginal mutual information $\mathcal{I}_{\delta,k}$ i.e., mutual information at time k is given by,

$$\mathcal{I}_{\delta,k} = H_\delta(X_k) + H_\delta(\delta N_k) - H_\delta(X_k, \delta N_k).$$

The chain rule for entropy gives

$$\mathcal{I}_{\delta,k} = H_\delta(\delta N_k) - H_\delta(\delta N_k|X_k).$$

Consider first

$$\begin{aligned} H_\delta(\delta N_k) &= -P(\delta N_k = 1) lnP(\delta N_k = 1) \\ &\quad - P(\delta N_k = 0) lnP(\delta N_k = 0) \\ &= -\beta_k\delta ln\beta_k + \beta_k\delta - \beta_k\delta ln\delta + o(\delta). \end{aligned}$$

Next consider

$$\begin{aligned} H_\delta(\delta N_k|X_k = x_k) &= -P(\delta N_k = 1|x_k) lnP(\delta N_k = 1|x_k) \\ &\quad - P(\delta N_k = 0|x_k) lnP(\delta N_k = 0|x_k) \\ &= \lambda_{k,x_k}\delta ln\lambda_{k,x_k} + \lambda_{k,x_k}\delta - \lambda_{k,x_k}\delta ln\delta + o(\delta). \end{aligned}$$

Taking expectations,

$$H_\delta(\delta N_k|X_k) = -\delta E(\lambda_{k,x_k} ln\lambda_{k,x_k}) + \beta_k\delta - \beta_k\delta ln\delta + o(\delta).$$

Taking the difference of the entropies, the mutual information at time t between an observed incremental count and an unobserved analog state is given by

$$\mathcal{I}(X_{(t)}, \delta N_{(t)}) = \delta E(\lambda_{(t,x)} ln\lambda_{(t,x)}) - \beta_{(t)}\delta ln\beta_{(t)} + o(\delta).$$

B. Trajectory Mutual Information

The formula for calculating the mutual information between the trajectories of an unobserved analog state and an observed point process is [8],[7, Section IV],

$$\begin{aligned} &\mathcal{I}(X_{(0,T)}, N_{(0,T)}) \\ &\quad = \int_0^T E(\lambda_{(t,x)} \ln \lambda_{(t,x)})dt - \int_0^T E(\hat{\lambda}_{(t)} \ln \hat{\lambda}_{(t)})dt, \quad (1) \end{aligned}$$

Note that the expectation in the first term is with respect to the probability measure of the unobserved process which is not known. We use iterated expectation to rewrite the formula as $\mathcal{I}(X_{(0,T)}, N_{(0,T)}) = E(J_T)$,

$$J_T = \int_0^T E(\lambda_{(t,x)} \ln \lambda_{(t,x)}|N_{(0,t)})dt - \int_0^T \hat{\lambda}_{(t)} \ln \hat{\lambda}_{(t)}dt,$$

with $\hat{\lambda}_{(t)} = E(\lambda_{(t,x)}|N_{(0,t)})$.

Then the expectation in the first term of J_T and $\hat{\lambda}_{(t)}$ can be evaluated using estimates of the filtering density conditional on point process observations. We use the auxiliary particle filter [10] to obtain estimates of the filtering density at each time step. $E(J_T)$ is approximated by a Monte Carlo average.

IV. PARTICLE FILTERING

The computation of trajectory mutual information in (1) requires estimates of the conditional density $p(x_k|N_0^k)$. The conditioning on the point process history implies that even under linear dynamical models for the state and intensity processes a closed form solution does not exist. We employ the particle filtering approach to solve the estimation problem. In the following we briefly explain the auxiliary particle filter [10] in the context of point process observations, referring to the substantial theoretical and practical development of the technique for the case of discrete time observations [11].

A. Auxiliary Particle Filter

In [10] the auxiliary particle filter was proposed as an extension of the particle filter [9] that was shown to improve the statistical efficiency of the sampling method by simulating particles with high predictive likelihoods. Sampling from the joint density of the state x_{k+1} and index i of the particles, the index which is an auxiliary variable can then be dropped to produce a discrete support of the filtering density. The unnormalized weights are defined based on the information of the incremental count δN_k as follows

$$\tilde{w_k^{(j)}} = \frac{p(\delta N_k|x_k^{(j)})p(x_k^{(j)}|x_{k-1}^{(i)^{(j)}})}{g(x_k^{(j)}, i^{(j)}|N_0^k)}, \; j = 1, \ldots, M$$

where

$$g(x_k, i|N_0^k) \propto p(\delta N_k|\xi_k^{(i)})p(x_k|x_{k-1}^{(i)})w_k^{(i)}$$

with $\xi_k^{(i)}$ as some charaterization of $x_k|x_{k-1}^{(i)}$. $w_k^{(i)}$ are the normalized weights. Marginalizing the state x_k,

$$g(i|N_0^k) \propto \int p(\delta N_k|\xi_k^{(i)})p(x_k|x_{k-1}^{(i)})w_k^{(i)}dx_k$$
$$= p(\delta N_k|\xi_k^{(i)})w_k^{(i)}.$$

This suggests an algorithm that samples from $g(x_k, i|N_0^k)$ by simulating index $i \sim g(i|N_0^k)$ and then sampling from $p(x_k|x_{k-1}^{(i)})$. The weights are redefined as $\tilde{w}_k^{(j)} = \frac{p(\delta N_k|x_k^{(j)})}{p(\delta N_k|\xi_k^{(j)})}, j = 1, \ldots, M$.

V. SIMULATION

A simulation study of the spike activity of the hippocampal place cells analyzed in [3] is presented. The spike activity is modeled as a doubly stochastic process [6] where the stochastic intensity depends on the unobserved spatial position of the animal. We illustrate the computation of information flows between the trajectory of the animal and the spiking activity of a place cell.

1) Simulation Development: A typical free foraging experiment is considered where the rat moves in a $[0,1] \times [0,1]\, m^2$ region. An electrode implanted in the hippocampus of the rat records neural activity at intervals of $0.8\, ms$. The model of the stochastic intensity to generate the spike train and the dynamics of the animal to generate the spatial positions are discussed below.

The state dependent stochastic intensity is modeled as a two-dimensional Gaussian surface,

$$\lambda_{(t,x)} = \exp(\alpha - \frac{1}{2}(x_t - \eta)^T \Sigma^{-1}(x_t - \eta))$$

with known parameter values $\alpha = 3.5, \eta = (0.5, 0.5)^T$ and $\Sigma = \sigma_\lambda^2 I$ where $\sigma_\lambda = 11.5\, cm$ and I is the identity matrix. This generates responses centered at $(0.5, 0.5)$; at a distance of $\sigma_\lambda = 11.5\, cm$ the firing rate is about $20\, Hz$ while at $2\sigma_\lambda = 23\, cm$ it is about $4.5\, Hz$.

The temporal evolution of the state is given by the continuous time Ornstein-Uhlenbeck process,

$$dx_{(t)} = \theta(\mu - x_{(t)})dt + \sigma_x dW_{(t)}$$

where the time constant $1/\theta = 3.3\, s$ and the standard deviation $\sigma_x = 10\, cm$, $W_{(t)}$ is a Brownian motion of variance t. Also $\mu = (0.5, 0.5)^T$.

The x and y components then each have standard errors $\approx \sqrt{\frac{\sigma_x^2}{2\theta}} = 13\, cm$. The magnitude of fluctuations in x,y relative to the center η of the intensity are $38\, cm$ and $33.6\, cm$ respectively.

The solution of the stochastic differential equation is simulated for $100\, s$ using the implicit Euler approximation (which preserves stability without additional constraints on the parameters),

$$x_{k+1} = \frac{2\theta\delta}{2+\theta\delta}\mu + \frac{2-\theta\delta}{2+\theta\delta}x_k + \frac{2}{2+\theta\delta}\sigma\sqrt{\delta}w_k$$

where $\delta = 25\, ms$ is the discretization step, and w_k is the normal random variable with mean $(0,0)^T$ and covariance I_2.

The point process is simulated for $100\, s$ under the No Simultaneity condition using the thinning procedure [15], [16].

Estimates of the conditional density $p(x_k|N_0^k)$ are obtained via the auxiliary particle filter using $M = 250$ particles which was found to perform better than the sampling importance resampling (SIR) filter [11].

The expectation in the second term in (1) is obtained by a Monte Carlo average from 50 runs involving simulation of the point process followed by particle filtering. The second integral in the formula is then evaluated numerically.

The expectation in the first term in (1) is evaluated by the iterated expectation using estimates of the conditional density $p(x_k|N_0^k)$ and taking a Monte Carlo average from 50 runs. Evaluating the integral and taking the difference with the second term gives the measure of information flow between the trajectories.

2) Simulation Results: The true trajectory of the rat from an initial position $(0.48, 0.49)\, m$ is shown in Fig. 1(a) where it is overlaid on a contour plot of the intensity. The particle filter estimates of the position and the error in the estimates along each axis are shown in Fig. 1(b),(c). The results indicate that the particle filter performs reasonably well considering the estimates are formed from observations of the neuronal firings only shown in Fig. 2.

As shown in Fig. 3(a), the trajectory mutual information rate drops initially before settling. The information rate is informative since it captures the variations in the mutual information flow between the trajectories of the rat and the neuronal activity of the hippocampal place cell. Also, variations in the trajectory mutual information rate are captured by the moving window mutual information rate as shown in Fig. 3(b) for a window size of $h = 25\, ms$.

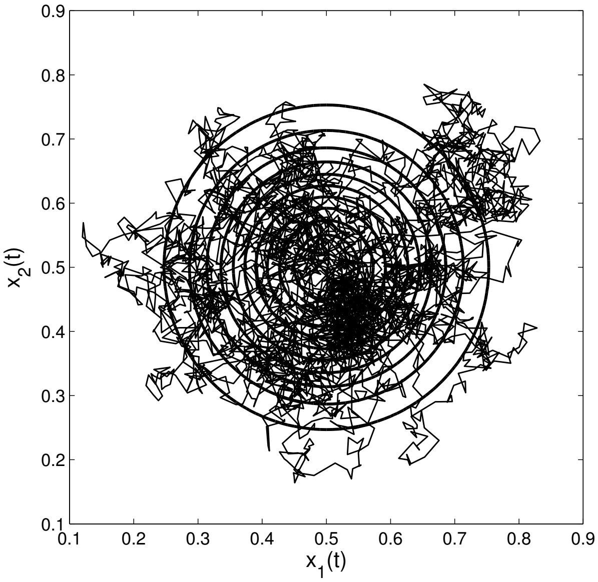
VI. DISCUSSION

The continuous time Ornstein-Uhlenbeck process is a suitable means to model a scenario where the rat freely forages in an open environment but returns to a location where some food or water may be present. Furthermore, by encoding the spatial information in the spike train through the intensity function, neuronal firings in bursts can be produced as the rat approaches the location $(0.5, 0.5)\, m$ (as shown in Fig. 2) using the 2-dimensional Gaussian surface to model the intensity. The particle filtering provides a reliable approach to decode the spatial information of the rat as it forages given observations of the neuronal firing times only.

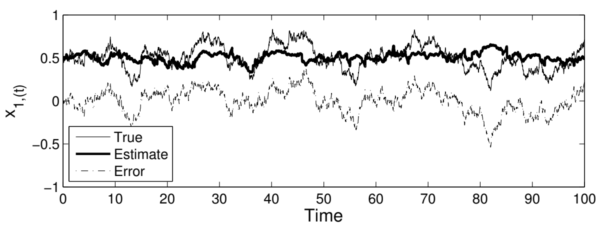
The mutual information between the trajectories of the rat's spatial position and neuronal firings in the place cell is expected to grow with time as the animal forages in the environment. This is also confirmed by Fig. 3(a) which suggests that information gain is approximately linear with time. The variations in the information gain due to the path taken by the rat are apparent in Fig. 3(b) where the peaks of the moving window mutual information rate coincide with those of the neuronal firings as the rat approaches $(0.5, 0.5)\, m$.

VII. CONCLUSIONS

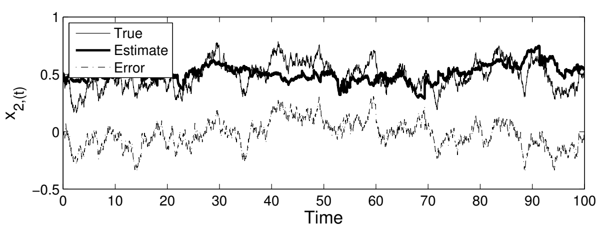
In this paper we have discussed for apparently the first time a proper model based computation of true trajectory



(a) True Trajectory



(b) Trajectory along x-axis



(c) Trajectory along y-axis

Fig. 1. True Trajectory and Particle Filter Estimates.

mutual information between a point process and an analog process when the intensity function of the point process depends on the unobserved analog process. Our implementation is based on the formula for trajectory mutual information due to [8] but developed further in [7] using the conditional Bernoulli heuristic and the stochastic conditional intensity function. Estimates of the conditional density of the state are obtained using particle filtering. The method has been illustrated with a simulation study. Future work will develop such a calculation for multi-unit neuronal recordings from the hippocampus place cells from real studies.

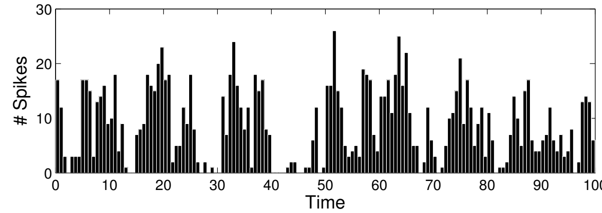
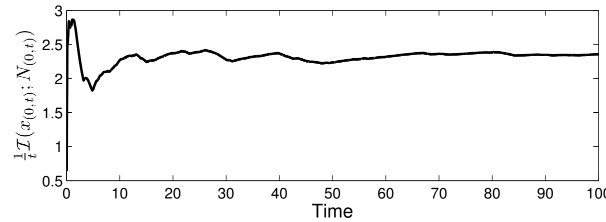
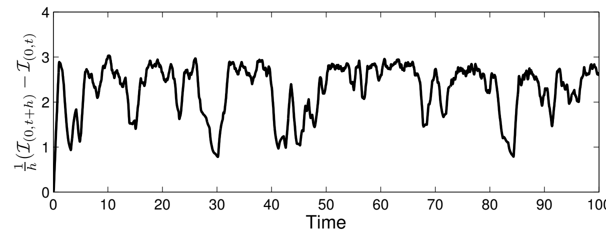


Fig. 2. Spike Activity in the Hippocampal Place Cell.



(a) Mutual Information Rate



(b) Moving Window Mutual Information Rate

Fig. 3. Information Flows in the Hippocampal Place Cell.

References

[1] F. Rieke, D. Warland, R. de Ruyter van Stvenink, and W. Bialek, *Spikes: Exploring the Neural Code*, MIT Press, Boston, MA, 1997.

[2] P Dayan and LF Abbott, *Theoretical Neuroscience*, MIT Press, Cambridge, MA, 2001.

[3] R Barbieri, LM Frank, DP Nguyen, MC Quirk, V Solo, MA Wilson, and EN Brown, "Dynamic analyses of information encoding in neural ensembles," *Neural Computation*, vol. 16, pp. 277–307, 2004.

[4] DM Taylor, SIH Tillery, and AB Schwartz, "Direct cortical control of 3D neuroprosthetic devices," *Science*, vol. 296, pp. 1829–1832, 2002.

[5] L. Paninski, "Estimation of entropy and mutual information," *Neural Computation*, vol. 15, pp. 1191–1253, 2003.

[6] DL Snyder, *Random Point Processes*, J. Wiley, New York, MA, 1975.

[7] V. Solo, "System identification with analog and counting process observations II: Mutual information," in *Proc. IEEE CDC*, 2005, pp. 3237–3242.

[8] P. Bremaud, "On the information carried by a stochastic point process," *Revue du Cethededc*, vol. 43, pp. 45–70, 1975.

[9] NJ Gordon, DJ Salmond, and AFM Smith, "Novel approach to nonlinear/nonGaussian Bayesian state estimation," *Proc. IEE-F*, vol. 140, pp. 104–117, 1993.

[10] MK Pitt and N Shepard, "Filtering via simulation: auxiliary particle filters," *J. American Stat. Assoc.*, vol. 94, pp. 590–599, 1999.

[11] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer, New York, MA, 2001.

[12] A. Karr, *Point Processes and Their Statistical Inference*, Marcel Dekker, New York, MA, 1991.

[13] DJ Daley and D Vere-Jones, *An Introduction to the Theory of Point Processes*, Springer-Verlag, New York, MA, 1988.

[14] T. Cover and J. Thomas, *Elements of Information Theory*, J Wiley, New York, MA, 1991.

[15] PAW Lewis and GS Shedler, "Simulation of nonhomogeneous Poisson processes by thinning," *Naval Res. Logistics Quart.*, vol. 26, pp. 403–413, 1979.

[16] Y. Ogata, "On Lewis' simulation method for point processes," *IEEE Trans. Information Theory*, vol. 27, pp. 23–31, 1981.