PREPRINT: Statistical Signal Processing Workshop, Gold Coast, Australia, June 2014.

THE CAUCHY-SCHWARZ DIVERGENCE FOR POISSON POINT PROCESSES

Hung Gia Hoang, Ba-Ngu Vo, Ba Tuong Vo[*]

Department of ECE
Curtin University
Bentley, WA 6102, Australia

Ronald Mahler

Lockheed Martin
1303 Corporate Center Dr., Suite 200
Eagan MN 55121, USA

ABSTRACT

Information theoretic divergences are fundamental tools used to measure the difference between the information conveyed by two random processes. In this paper, we show that the Cauchy-Schwarz divergence between two Poisson point processes is the half the squared L^2-distance between their respective intensity functions. Moreover, this can be evaluated in closed form when the intensities are Gaussian mixtures.

Index Terms— Poisson point process, information theoretic divergence, random finite sets

1. INTRODUCTION

The Poisson point process, which models "no interaction" or "complete spatial randomness" in spatial point patterns, is arguably one of the best known and most tractable of point processes [1]. Point process theory is the study of random counting measures with applications spanning numerous disciplines, see for example [2]. The Poisson point process itself arises in forestry, geology, biology, particle physics, communication networks and signal processing. The role of the Poisson point process in point process theory, in most respects, is analogous to that of the normal distribution in random vectors.

Information theoretic divergences measure the "difference" in information between two random variables, and are fundamental in information theory and statistical analysis [3]. The better known divergences include Kullback-Leibler, Rényi (or α-divergence) and their generalization Csiszár-Morimoto (or Ali-Silvey), as well as Jensen-Rényi and Cauchy-Schwarz. The Kullback-Leibler divergence can be evaluated analytically for Gaussians (random vectors), but for the more versatile class of Gaussian mixture [4], only Jensen-Rényi and Cauchy-Schwarz divergences can be evaluated in closed forms [5]. The Kullback-Leiber and Rényi divergences have also been studied for point processes or random finite sets [6]. However, so far except for trivial special cases, these divergences cannot be computed analytically and requires expensive approximations such as Monte Carlo.

In this paper, we show that for Poisson point processes, the Cauchy-Schwarz divergence between their probability densities is given by the square of the L^2-distance between their intensity functions. Geometrically, this result relates the angle subtended by the probability densities to the L^2-distance between the corresponding intensity functions. Moreover, for Gaussian mixture intensity functions, the L^2-distance, and hence the Cauchy-Schwarz divergence, can be evaluated in closed form. The Poisson point process enjoys numerous nice properties [1, 2], and our result is an interesting and useful addition.

The organization of the paper is as follows. Section 2 summarizes concepts in point process theory needed for the exposition of our result. In Section 3, the main results of the paper that establish the analytical formulation for the Cauchy-Schwarz divergence between two Poisson point processes is presented. Finally, Section 4 concludes the paper.

2. A BRIEF SUMMARY OF POINT PROCESSES THEORY

In this work we consider a state space $\mathscr{X} \subseteq \mathbb{R}^d$, and adopt the inner product notation $\langle f, g\rangle \triangleq \int f(x)g(x)dx$; the L^2-norm notation $\|f\| \triangleq \sqrt{\langle f, f\rangle}$; the multi-target exponential notation $h^X \triangleq \prod_{x\in X} h(x)$, where h is a real-valued function, with $h^{\emptyset} = 1$ by convention; and the indicator function notation

$$1_B(x) \triangleq \begin{cases} 1, & \text{if } x \in B \\ 0, & \text{otherwise} \end{cases}.$$

This section briefly summarizes concepts in point process theory needed for the exposition of our result. Point process theory, in general, is concerned with *random counting measures*. Our result is restricted to simple-finite point processes, which can be regarded as *random finite sets*. For simplicity, we omit the prefix "simple-finite" in the rest of the paper. For detailed treatments of the subject we refer the reader to the textbook [2].

A *point process* or *random finite set* (RFS) X on $\mathscr{X}$ is random variable taking values in $\mathscr{F}(\mathscr{X})$, the space of finite

*Acknowledgment: This work is supported by the Australian Research Council under the Future Fellowship FT0991854 and Discovery Early Career Researcher Award DE120102388.

subsets of $\mathcal{X}$. Let $|X|$ denotes the number of elements in a set X. A point process X on $\mathcal{X}$ is said to be *Poisson* with a given *intensity function* u (defined on $\mathcal{X}$) if [2]:

1. for any $B \subseteq \mathcal{X}$ such that $\langle u, 1_B \rangle < \infty$, the random variable $|X \cap B|$ is Poisson distributed with mean $\langle u, 1_B \rangle$,
2. for any disjoint $B_1, ..., B_i \subseteq \mathcal{X}$, the random variables $|X \cap B_1|, ..., |X \cap B_i|$ are independent.

Since $\langle u, 1_B \rangle$ is the expected number of points of X in the region B, the intensity value $u(x)$ can be interpreted as the instantaneous expected number of points per unit hyper-volume at x. Consequently, $u(x)$ is not dimensionless in general. If hyper-volume (on $\mathcal{X}$) is measured in units of K (e.g. m^d, cm^d, in^d, etc.) then the intensity function u has unit K^{-1}.

The number of points of a Poisson point processs X is Poisson distributed with mean $\langle u, 1 \rangle$, and condition on the number of points the elements x of X are independently and identically distributed (i.i.d.) according to the probability density $u(\cdot)/\langle u, 1 \rangle$ [2]. It is implicit that $\langle u, 1 \rangle$ is finite since we only consider simple-finite point processes.

The probability distribution of a Poisson point process X with intensity function u is given by [2, pp. 15]

$$\Pr(X \in \mathcal{T}) = \sum_{i=0}^{\infty} \frac{e^{-\langle u,1 \rangle}}{i!} \int_{\mathcal{X}^i} 1_{\mathcal{T}}(\{x_1, ..., x_i\}) u^{\{x_1,...,x_i\}} d(x_1, ..., x_i), \quad (1)$$

for any (measurable) subset $\mathcal{T}$ of $\mathcal{F}(\mathcal{X})$, where $\mathcal{X}^i$ denotes an i-fold Cartesian product of $\mathcal{X}$, with the convention $\mathcal{X}^0 = \{\emptyset\}$, and the integral over $\mathcal{X}^0$ is $1_{\mathcal{T}}(\emptyset)$. A Poisson point process is completely characterized by its intensity function (or more generally the intensity measure).

Probability densities of point processes considered in this work are defined with respect to the reference measure μ given by

$$\mu(\mathcal{T}) = \sum_{i=0}^{\infty} \frac{1}{i! K^i} \int_{\mathcal{X}^i} 1_{\mathcal{T}}(\{x_1, ..., x_i\}) d(x_1, ..., x_i) \quad (2)$$

for any (measurable) subset $\mathcal{T}$ of $\mathcal{F}(\mathcal{X})$. The measure μ is analogous to the Lebesque measure on $\mathcal{X}$ (indeed it is the unnormalized distribution of a Poisson point process with unit intensity $u = 1/K$ when the state space $\mathcal{X}$ is bounded). Moreover, it was shown in [7] that for this choice of reference measure, the integral of a function $f : \mathcal{F}(\mathcal{X}) \to \mathbb{R}$, given by

$$\int f(X)\mu(dX) = \sum_{i=0}^{\infty} \frac{1}{i! K^i} \int_{\mathcal{X}^i} f(\{x_1, ..., x_i\}) d(x_1, ..., x_i), \quad (3)$$

is equivalent to Mahler's set integral [8]. Note that the reference measure μ, and the integrand f are all dimensionless.

Our main result involves Poisson point processes with probability densities of the form

$$f(X) = K^{|X|} e^{-\langle u,1 \rangle} u^X. \quad (4)$$

Note that for any (measurable) subset $\mathcal{T}$ of $\mathcal{F}(\mathcal{X})$

$$\int_{\mathcal{T}} f(X)\mu(dX) = \sum_{i=0}^{\infty} \frac{e^{-\langle u,1 \rangle}}{i!} \int_{\mathcal{X}^i} 1_{\mathcal{T}}(\{x_1, ..., x_i\}) u^{\{x_1,...,x_i\}} d(x_1, ..., x_i).$$

Thus, comparing with (1), f is indeed a probability density (with respect to μ) of a Poisson point process with intensity function u.

3. THE CAUCHY-SCHWARZ DIVERGENCE FOR POISSON POINT PROCESSES

The Cauchy-Schwarz divergence is based on the Cauchy-Schwarz inequality for inner products, and is defined for two random vectors with probability densities f and g by [9]

$$D_{CS}(f, g) = -\ln \frac{\langle f, g \rangle}{\|f\| \, \|g\|}. \quad (5)$$

The argument of the logarithm in (5) is non-negative (since probability densities are non-negative) and does not exceed one (by the Cauchy-Schwarz inequality). Moreover, it can be interpreted as the cosine of the angle subtended by f and g in $L^2(\mathcal{X}, \mathbb{R})$, the space of square integrable functions taking $\mathcal{X}$ to $\mathbb{R}$. Note that $D_{CS}(f, g)$ is symmetric and positive unless $f = g$, in which case $D_{CS}(f, g) = 0$.

Geometrically, the Cauchy-Schwarz divergence determines the information "difference"between random vectors from the angle between their probability densities. The Cauchy-Schwarz divergence can also be interpreted as an approximation to the Kullback-Leibler divergence [5]. Hence, the Cauchy-Schwarz divergence between two densities of random variables has been employed in many information theoretic applications, especially in machine learning and pattern recognition [5].

For point processes, the Csiszár-Morimoto divergence, which includes Kullback-Leibler and Rényi, were formulated in [10] by replacing standard (Lebesque) integrals with set integrals [8]. However, the Cauchy-Schwarz divergence cannot be extended to point processes by simply replacing the standard integral with the set integral since the latter is not unit compatible for defining inner products and norms of set derivatives. If we "blindly" define $\langle f, g \rangle$ via set integral as

$$\int f(X)g(X)\delta X = \sum_{i=0}^{\infty} \frac{1}{i!} \int f(\{x_1, ..., x_i\}) g(\{x_1, ..., x_i\}) d(x_1, ..., x_i), \quad (6)$$

then the inner product is not well-defined. The reason is that the i-th term in the above sum has units of K^i, and hence these terms cannot be added together due to unit mismatch. For example, if $K = m^3$, then the first term is unitless, the second term is in m^3, the third term is in m^6, etc., and it is meaningless to add these together.

Using the standard notion of density and integration summarized in Section 2, we can define the inner product $\langle f,g\rangle_\mu \triangleq \int f(X)g(X)\mu(dX)$, and corresponding norm $\|f\|_\mu \triangleq \sqrt{\langle f,f\rangle_\mu}$ on $L^2(\mathscr{F}(\mathscr{X}),\mathbb{R})$, the space of square integrable functions taking $\mathscr{F}(\mathscr{X})$ to $\mathbb{R}$. Interestingly, the inner product between multi-target exponentials is given by the following result.

Lemma 1. *Let $f(X) = K^{|X|}u^X$ and $g(X) = K^{|X|}v^X$ with u, $v \in L^2(\mathscr{X},\mathbb{R})$ (measured in units of K^{-1}). Then $\langle f,g\rangle_\mu = e^{K\langle u,v\rangle}$.*

Proof:

$$\begin{aligned}\langle f,g\rangle_\mu &= \int K^{2|X|}\,[uv]^X\,\mu(dX)\\ &= \sum_{i=0}^{\infty}\frac{K^i}{i!}\left[\int_{\mathscr{X}} u(x)v(x)dx\right]^i \quad \text{(using (3))}\\ &= \sum_{i=0}^{\infty}\frac{K^i}{i!}\langle u,v\rangle^i = e^{K\langle u,v\rangle}.\square\end{aligned}$$

In the spirit of using the angle between probability densities to determine the information "difference", the Cauchy-Schwarz divergence can be extended to point processes as follows.

Definition 2. *The Cauchy-Schwarz divergence between the probability densities f and g of two point processes with respect to the reference measure μ is defined by*

$$D_{CS}(f;g) = -\ln\frac{\langle f,g\rangle_\mu}{\|f\|_\mu\,\|g\|_\mu}. \tag{7}$$

The following Proposition asserts that *the Cauchy-Schwarz divergence between two Poisson point process is half the squared distance between their intensity functions*.

Proposition 3. *The Cauchy-Schwarz divergence between the probability densities f and g of two Poisson point processes with respective intensity functions u and $v \in L^2(\mathscr{X},\mathbb{R})$ (measured in units of K^{-1}), is*

$$D_{CS}(f;g) = \frac{K}{2}\|u-v\|^2. \tag{8}$$

Proof: Applying Lemma 1 to the numerator and denominator of (7) and cancelling out the constants $e^{-\langle u,1\rangle}$, $e^{-\langle v,1\rangle}$ we have

$$\begin{aligned}D_{CS}(f;g) &= -\ln\left(e^{K\langle u,v\rangle-\frac{K}{2}\langle u,u\rangle-\frac{K}{2}\langle v,v\rangle}\right)\\ &= \frac{1}{2}\left[K\langle u,u\rangle - 2K\langle u,v\rangle + K\langle v,v\rangle\right]\\ &= \frac{K}{2}\|u-v\|^2.\square\end{aligned}$$

Note that the LHS of (8) is defined with respect to the reference measure μ given in (2), thus the equality might not hold if a different reference measure is used. However, the RHS of (8) is invariant of the unit K of the space of intensity functions. Indeed, suppose that the unit of the hyper-volume in the intensity function space has been changed from K to ρK, the value of the intensity function is also changed accordingly as illustrated in Fig. 1. For example, the intensity measured in m^3 is 10^3 times larger than that measured in dm^3. As a result, the change of the squared distance $\|u_\rho - v_\rho\|^2$ cancels out the change of unit ρK in the RHS of (8), making it unit-invariant.

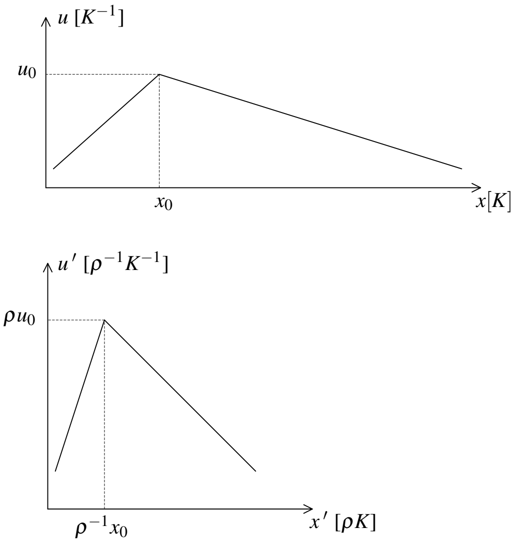

Fig. 1: Change of unit in the intensity function space

The above result has a nice geometric interpretation that relates the angle subtended by the probability densities in $L^2(\mathscr{F}(\mathscr{X}),\mathbb{R})$ to the distance between the corresponding intensity functions in $L^2(\mathscr{X},\mathbb{R})$ as depicted in Fig. 2. More specifically, for Poisson point processes: *the secant of the angle between by their probability densities equals the exponential of half the squared distance between their intensity functions.*

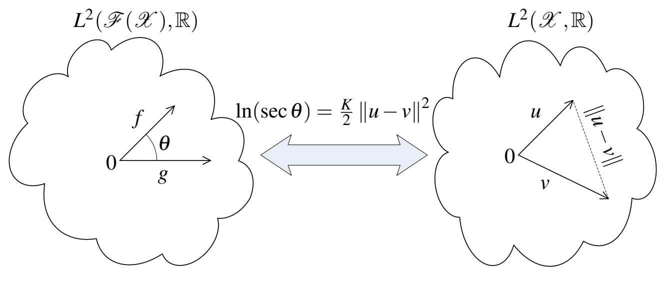

Fig. 2: Geometric interpretation of Proposition 3

Proposition 3 has important implications in the approximation of Poisson point processes through their intensity functions. It is intuitive that the "difference" between the Poisson distributions vanishes as the distance between their intensity functions tends to zero. However, it was not clear that a reduction in the error between the intensity functions necessarily implies a reduction in the "difference" between the associated Poisson distributions. Our result not only verifies that the "difference" between the distributions is reduced, it also quantifies the reduction.

In general, the L^2-norm and hence the Cauchy-Schwarz divergence above cannot be numerically evaluated in closed form. However, for Poisson point processes with Gaussian mixture intensity function, applying the Gaussian identity [11]

$$\langle \mathcal{N}(\cdot;\mu_0,\Sigma_0), \mathcal{N}(\cdot;\mu_1,\Sigma_1)\rangle = \mathcal{N}(\mu_0;\mu_1,\Sigma_0+\Sigma_1),$$

to (8) yields an analytic expression for the Cauchy-Schwarz divergence. This is stated more concisely in the following result.

Corollary 4. *The Cauchy-Schwarz divergence between two Poisson point processes with Gaussian mixture intensities:*

$$u(x) = \sum_{i=0}^{N_u} w_u^{(i)} \mathcal{N}(x; m_u^{(i)}, P_u^{(i)}), \tag{9a}$$

$$v(x) = \sum_{i=0}^{N_1} w_v^{(i)} \mathcal{N}(x; m_v^{(i)}, P_v^{(i)}) \tag{9b}$$

(measured in units of K^{-1}) is given by

$$\begin{aligned} D_{CS} = {} & \frac{1}{2}\sum_{i=1}^{N_u}\sum_{j=1}^{N_u} w_u^{(i)} w_u^{(j)} \mathcal{N}\left(m_u^{(i)}; m_u^{(j)}, P_u^{(i)} + P_u^{(j)}\right) \\ & + \frac{1}{2}\sum_{i=1}^{N_v}\sum_{j=1}^{N_v} w_v^{(i)} w_v^{(j)} \mathcal{N}\left(m_v^{(i)}; m_v^{(j)}, P_v^{(i)} + P_v^{(j)}\right) \\ & - \sum_{i=1}^{N_u}\sum_{j=1}^{N_v} w_u^{(i)} w_v^{(j)} \mathcal{N}\left(m_u^{(i)}; m_v^{(j)}, P_u^{(i)} + P_v^{(j)}\right) \end{aligned} \tag{10}$$

This Corollary has important implications in Gaussian mixture reduction for intensity functions. The result provides mathematical justification for Gaussian mixture intensity reduction based on L^2-error. Furthermore, since Gaussian mixtures can approximate any density to any desired accuracy [12], Corollary 4 enables the Cauchy-Schwarz divergence between two Poisson point processes to be approximated to any desired accuracy.

4. CONCLUSIONS

In this paper, we have extended the Cauchy-Schwarz divergence to point processes by using the inner product of their probability densities and showed that for Poisson point processes, this divergence is half the squared distance between the intensity functions. Moreover, for Gaussian mixture intensity functions this divergence can be evaluated analytically. Our result is an addition to the list of interesting properties of Poisson point processes and has important implications in numerical approximations. Future works will expand the results to other information-theoretic measures that are invariant under the change of reference measure, such as the Bhattacharyya distance.

5. REFERENCES

[1] J.F.C. Kingman, *Poisson Processes*, Oxford studies in probability. Oxford University Press, 1993.

[2] J. Moller and R. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*, Chapman & Hall CRC, 2004.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, New York, NY, USA, 1991.

[4] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 145–151, 1991.

[5] F. Nielsen, "Closed-form information-theoretic divergences for statistical mixtures," in *2012 21st International Conference on Pattern Recognition (ICPR)*, 2012, pp. 1723–1726.

[6] B. Ristic and B-N. Vo, "Sensor control for multi-object state-space estimation using random finite sets," *Automatica*, vol. 46, no. 11, pp. 1812–1818, 2010.

[7] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multi-target filtering with random finite sets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 4, pp. 1224–1245, 2005.

[8] R. Mahler, "Multi-target Bayes filtering via first-order multi-target moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, 2003.

[9] K. Kampa, E. Hasanbelliu, and J. C. Principe, "Closed-form Cauchy-Schwarz pdf divergence for mixture of gaussians," in *Proc. The 2011 International Joint Conference on Neural Networks*. 2011, pp. 2578–2585, IEEE.

[10] R. Mahler, "Global posterior densities for sensor management," in *Proc. SPIE*, 1998, vol. 3365, pp. 252–263.

[11] Y. C. Ho and R. C. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Trans. Autom. Control*, vol. 9, no. 4, pp. 333–339, 1964.

[12] J.T.-H. Lo, "Finite-dimensional sensor orbits and optimal nonlinear filtering," *IEEE Trans. Inf. Theory*, vol. 18, no. 5, pp. 583–588, 1972.