# Local Entropy Statistics for Point Processes

Daniel E. Clark

***Abstract*— Point processes are often described with functionals, such as the probability generating functional, the Laplace functional, and the factorial cumulant generating functional. These are used to facilitate modelling of different processes and to determine important statistics via functional differentiation. In information theory, generating functions have also been defined for probability densities to determine information quantities such as the Shannon information and Kullback-Leibler divergence, though as yet there are no such analogues for point processes. The purpose of this article is to exploit the advantages of both types of generating function to facilitate the derivation of information statistics for point processes. In particular, a generating functional for point processes is introduced for determining statistics related to entropy and relative entropy based on Golomb's information function and Moyal's probability generating functional. It is shown that the information generating functional permits the derivation of a suite of statistics, including localised Shannon entropy and Kullback-Leibler divergence calculations.**

***Index Terms*— Information entropy.**

## I. Introduction

Generating function and functional expansions are used as a means of describing functions or stochastic processes to enable modelling and determination of specific statistical quantities of interest. For instance, the characteristic function can be viewed as the Fourier transform of a probability distribution and it inherits convenient properties for convolution from the Fourier transform. The Laplace functional [1], enables similar results to be determined for a more general class of stochastic processes. Generating functions for information have been developed, notably Golomb's information generating function that enables the determination of entropy moments [2], though are less widely used and developed. Methods for determination of entropy statistics for point processes are not new, dating back to McFadden's definition in 1965 [3]. However, it is still an active area of interest, with recent works on the development and application of the concepts [4], [5].

In this article, a generating functional for information statistics is developed for the purpose of determining statistics for point processes [6], [7] based on a unification of Golomb's information generating function and the Laplace functional [1]. The result is then placed in the context of point processes with the Laplace functional and probability generating functional [7] to determine an information functional version. This enables the determination of a suite of moment statistics related to entropy and relative entropy. In particular, by taking derivatives of the Laplace information functional, we can determine Shannon entropy [8], the Kullback-Leibler divergence [9], the Rényi divergence [10] and moment statistics of each of these.

The motivation for this work was to enable decisions to be made based on local assessments of information. There have been approaches in the point process literature for addressing a similar problem [4], [5], though usually restricted to providing the Shannon entropy.

The paper is structured as follows: In the following section, the basic concepts are introduced in the context of discrete random variables. Generating functions for information are discussed and defined in relation to Shannon entropy and Kullback-Leibler divergence. It is shown how to derive particular statistical quantities, from the generating functions. A summary of the key definitions and tools from point process theory are presented in Section III as an introduction to the new information generating functionals designed for point processes described in the context of Shannon entropy in Section IV, and for the Kullback-Leibler divergence in Section V. Worked examples illustrating these concepts are given in Section VI. The paper concludes in Section VII.

## II. Generating Functions and Entropy Statistics

In this Section we discuss generating functions for entropy statistics. In the following, we describe prior generating functions for entropy and relative entropy before introducing extended forms in the context of non-negative valued integer random variables and their moments.

### A. Golomb's Information Generating Function

In this section, we are concerned with ensembles of discrete probabilities. $p_1, p_2, \ldots,$ The entropy for an ensemble of probabiities, $p_1, \ldots, p_n$, discrete and finite was defined by Shannon [8] with the summation

$$H(p) = -\sum_{k=1}^{n} p_k \log p_k. \tag{1}$$

The relative entropy between two distributions $p_k$ and $q_k$ was defined by Kullback and Leibler [9] with

$$K(p,q) = \sum_{k=1}^{n} p_k \log \frac{p_k}{q_k}. \tag{2}$$

Manuscript received October 10, 2018; revised September 9, 2019; accepted September 9, 2019. Date of publication September 13, 2019; date of current version January 20, 2020. This work was supported in part by the Joint AFRL-DSTL Basic-Research in Autonomous Signal Processing AFSOR under Grant FA9550-19-1-7008 and in part by the DSTL Task under Grant 1000133068.

The author is with Télécom SudParis, 91001 Évry, France (e-mail: daniel.clark@telecom-sudparis.eu).

Communicated by P. Harremoës, Associate Editor for Probability and Statistics.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2019.2941213

Following this, Rényi [10] defined the entropy of order $\alpha$ for a discrete and finite distribution, and between two finite distributions $p_k, q_k$, $k = (0, 1, 2, \ldots)$ with

$$R^\alpha(p) = \frac{1}{\alpha - 1} \log \sum_{k=1}^{n} p_k^\alpha, \tag{3}$$

$$R^\alpha(p, q) = \frac{1}{\alpha - 1} \log \sum_{k=1}^{n} p_k^\alpha q_k^{1-\alpha}. \tag{4}$$

**Definition** II.1 (Golomb's information generating function for entropy). *Golomb defined an information generating function for entropy with the mathematical relation [2],*

$$T^\alpha(p) = \sum_{k=1}^{n} p_k^\alpha, \tag{5}$$

**Definition** II.2 (Guiasu and Reischer's information generating function for relative entropy). *Guiasu and Reischer [11] developed Golomb's idea to define an information generating function for relative entropy as follows.*

$$T^\alpha(p, q) = \sum_{k=1}^{n} \frac{p_k^\alpha}{q_k^{\alpha-1}}. \tag{6}$$

We can see the following relations Rényi entropy and divergence and the information generating functions with

$$R^\alpha(p_k) = \frac{1}{\alpha - 1} \log T^\alpha(p), \tag{7}$$

$$R^\alpha(p, q) = \frac{1}{\alpha - 1} \log T^\alpha(p, q). \tag{8}$$

Golomb noted that moments of entropy can be determined by taking derivatives of the information generating function with respect to $\alpha$ as follows,

$$\frac{\mathrm{d}^k}{\mathrm{d}\alpha^k} T^\alpha(p)\bigg|_{\alpha=1} = \sum_{r=1}^{n} (\log p_r)^k \, p_r. \tag{9}$$

There is a similar relation for the moments of information for the Kullback-Leibler divergence.

### *B. A New Information Generating Function for Non-Negative Integer-Valued Random Variables*

In this section, we are exclusively concerned with non-negative integer-valued random variables. Let $\xi$ be a random variable with $P(\xi = k) = p_k$, $k = (0, 1, \ldots)$.

**Definition** II.3 (New information generating function for entropy). *Define the new information generating function for entropy $G_\xi^\alpha(z)$ of random variable $\xi$ by the series*

$$G_\xi^\alpha(z) = \sum_{k=0}^{\infty} p_k^{1-\alpha} z^k, \tag{10}$$

*where $0 \leq \alpha \leq 1$, and $z$ is a complex number.*

Note that if we set $\alpha = 0$, we have the usual probability generating function for non-negative integer-valued random variables. Thus we can determine the $k^{th}$ term $p_k$ via differentiation and setting the argument $z$ to be equal to zero, i.e.

$$p_k = \frac{1}{k!} \frac{\partial^k}{\partial z^k} G_\xi^\alpha(z)\bigg|_{z=0,\alpha=0} \quad (k = 0, 1, 2, \ldots). \tag{11}$$

The introduction of the generating function for entropy permits the treatment of theoretic problems for entropy with the methods of generating functions. For instance, if we take the first-order derivative,

$$\frac{\partial}{\partial \alpha} G_\xi^\alpha(1)\bigg|_{\alpha=0} = -\sum_{k=0}^{\infty} p_k \log p_k, \tag{12}$$

we find the Shannon entropy. In a similar manner to determining the terms in the sequence $p_k$, we can determine $k^{th}$-order contribution to the Shannon entropy with

$$-p_k \log p_k = \frac{1}{k!} \frac{\partial^k}{\partial z^k} \frac{\partial}{\partial \alpha} G_\xi^\alpha(z)\bigg|_{z=0,\alpha=0} \quad (k = 0, 1, 2, \ldots). \tag{13}$$

By analogy with the probability generating function, we can similarly define factorial moments of entropy by considering derivatives of both $z$ and $\alpha$. For instance, we can determine a moment statistic for entropy with

$$\frac{\partial}{\partial z} \frac{\partial}{\partial \alpha} G_\xi^\alpha(z)\bigg|_{z=1,\alpha=0} = -\sum_{k=1}^{\infty} k p_k \log p_k. \tag{14}$$

We now consider calculating some generating functions for well-known distributions.

*Example* 1 (Bernoulli distribution). *Let $\xi$ be a Bernoulli random variable. Then the new information generating function becomes*

$$G_\xi^\alpha(z) = (1 - p)^{1-\alpha} + p^{1-\alpha} z, \tag{15}$$

*where we recover the Shannon entropy with*

$$\frac{\partial}{\partial \alpha} G_\xi^\alpha(1)\bigg|_{\alpha=0} = -(1 - p)\log(1 - p) - p \log p. \tag{16}$$

*A first-order entropy statistic related to the cardinality is found by taking the derivative with respect to z, as well as $\alpha$ i.e.*

$$\frac{\partial}{\partial \alpha} \frac{\partial}{\partial z} G_\xi^\alpha(z)\bigg|_{\alpha=0,z=1} = -p \log p. \tag{17}$$

*Example* 2 (The Poisson distribution ). *Let $\xi$ be a random variable for a Poisson distribution with rate $\lambda$. Then the generating function becomes*

$$G_\xi^\alpha(z) = \sum_{k=0}^{\infty} \left( \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \right)^\alpha z^k. \tag{18}$$

*Using the same argument as earlier, we have the Shannon entropy for the Poisson distribution with*

$$\frac{\partial}{\partial \alpha} G_\xi^\alpha(1)\bigg|_{\alpha=0} = -\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \log\left( \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \right). \tag{19}$$

### C. Generating Functions for Relative Entropy

**Definition** II.4 (New information generating function for relative entropy). *Consider $\xi_1$ and $\xi_2$ to be the random variables of two distributions, where $P_0(\xi_0 = k) = p_{0,k}$, $P_1(\xi_1 = k) = p_{1,k}$ $k = (0, 1, \ldots)$. We define the generating function for relative entropy, $G^{\alpha}_{\xi_0,\xi_1}(z)$, of the random variables $\xi_0, \xi_1$ with the series*

$$G^{\alpha}_{\xi_0,\xi_1}(z) = \sum_{k=0}^{\infty} p_{0,k}^{1-\alpha} p_{1,k}^{\alpha} z^k, \tag{20}$$

*where $0 \leq \alpha \leq 1$, and $z$ is a complex number.*

*Example* 3 (Kullback-Leibler divergence). *The Kullback-Leibler divergence, or relative entropy, between two distributions can be determined with*

$$-\frac{\partial}{\partial \alpha} G^{\alpha}_{\xi_0,\xi_1}(1)\bigg|_{\alpha=0} = \sum_{k=0}^{\infty} p_{0,k} \log \frac{p_{0,k}}{p_{1,k}}. \tag{21}$$

We define this as the generating function for relative entropy between two distributions of non-negative integer-valued random variables. As above, we can determine the relative entropy by differentiating with respect to $\alpha$, i.e.

$$\sum_{k=0}^{\infty} p_{0,0} \log \frac{p_{0,0}}{p_{1,0}} = -\frac{\partial}{\partial \alpha} G^{\alpha}_{\xi}(0)\bigg|_{\alpha=0}; \tag{22}$$

$$p_{0,k} \log \frac{p_{0,k}}{p_{1,k}} = -\frac{1}{k!} \frac{\partial^k}{d\xi^k} \frac{\partial}{\partial \alpha} G^{\alpha}_{\xi}(x)\bigg|_{z=0,\alpha=0}. \tag{23}$$

These statistics will be extended to their point process counterparts in Section IV.

*Example* 4 (New generating function for relative entropy between two Poisson distributions). *Consider two Poisson distributions with respective parameters $\lambda_0$ et $\lambda_1$. The new generating function for relative entropy is written*

$$\begin{aligned} G^{\alpha}_{\xi_0,\xi_1}(z) &= \sum_{k=0}^{\infty} \left(\frac{\lambda_0^k}{k!} e^{-\lambda_0}\right)^{1-\alpha} \left(\frac{\lambda_1^k}{k!} e^{-\lambda_1}\right)^{\alpha} z^k \\ &= \exp\left(\lambda_0^{1-\alpha} \lambda_1^{\alpha} z - \lambda_0(1-\alpha) - \lambda_1 \alpha\right), \end{aligned} \tag{24}$$

*where the last line follows from the Taylor expansion of the exponential function.*

The Kullback-Leibler divergence is calculated with

$$-\frac{\partial}{\partial \alpha} G^{\alpha}_{\xi_0,\xi_1}(1)\bigg|_{\alpha=0} = \lambda_0 \log\left(\frac{\lambda_0}{\lambda_1}\right) - \lambda_0 - \lambda_1. \tag{25}$$

Note also that the Rényi divergence is simple to compute this case as

$$\begin{aligned} &\frac{1}{1-\alpha} \log G^{\alpha}_{\xi_0,\xi_1}(1) = \\ &\quad \frac{1}{\alpha - 1} \left((\lambda_0(1-\alpha) - \lambda_1 \alpha) \; -\lambda_0^{1-\alpha} \lambda_1^{\alpha}\right). \end{aligned} \tag{26}$$

In the following Sections we shall extend these ideas in the context of point processes.

## III. Point Process Fundamentals

In this section we provide the basic definitions and operations required for describing point processes, their moments, and functional descriptions. This summarises review material on point processes presented in [12]–[14].

**Definition** III.1. *A point process $\Phi$ on $\mathcal{X}$ is a random variable on the process space $\mathfrak{X} = \bigcup_{n \geq 0} \mathcal{X}^n$, i.e., the space of finite sequences of points in $\mathcal{X}$. A realisation of $\Phi$ is a sequence $\varphi = (x_1, \ldots, x_n) \in \mathcal{X}^n$, representing a population of $n$ points with states $x_i \in \mathcal{X}$. Point processes can be described using their probability distribution $P_\Phi$ on the measurable space $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$, where $\mathcal{B}(\mathfrak{X})$ denotes the Borel $\sigma$-algebra of the process space $\mathfrak{X}$ [14]. The projection measure $P_\Phi^{(n)}$ of the probability distribution $P_\Phi$ on $\mathcal{X}^n$, $n \geq 0$, describes the realisations of $\Phi$ with $n$ elements; the projection measures of a point process are always defined as symmetric functions.*

We shall use the following two descriptions of point processes, the Laplace functional and the probability generating functional.

**Definition** III.2 (Laplace functional). *The* Laplace functional *$\mathcal{L}_\Phi$ of a point process $\Phi$ is defined by*

$$\mathcal{L}_\Phi(f) = \mathbb{E}_\Phi\left[\exp\left(-\sum_{x \in \Phi} f(x)\right)\right] \tag{27}$$

$$= \sum_{n \geq 0} \int \exp\left(-\sum_{i=1}^{n} f(x_i)\right) P_\Phi^{(n)}(\mathrm{d}x_1, \ldots, \mathrm{d}x_n), \tag{28}$$

*for test function $f : \mathcal{X} \to \mathbb{R}^+$.*

**Definition** III.3 (Probability generating functional). *The* probability generating functional *(p.g.fl.) $\mathcal{G}_\Phi$ of a point process $\Phi$ is defined by*

$$\begin{aligned} \mathcal{G}_\Phi(h) &= \mathbb{E}_\Phi\left[\left(\prod_{x \in \Phi} h(x)\right)\right] \\ &= \sum_{n \geq 0} \int \left(\prod_{i=1}^{n} h(x_i)\right) P_\Phi^{(n)}(\mathrm{d}x_1, \ldots, \mathrm{d}x_n), \end{aligned} \tag{29}$$

*for test function $h : \mathcal{X} \to [0, 1]$.*

The probability generating functional is analogous to the the probability generating functions for non-negative integer-valued random variables, replacing the complex number argument $z$ for the probability generating function with the function $h$, since we can determine the related probability statistics in a similar way via differentiation. The key difference is the replacement of the probability distribution $p_k$ with projection measures $P_\Phi^{(k)}$, $k \geq 0$.

*Example* 5 (Poisson point process on real line). *The probability generating functional of a Poisson point process on the real line with intensity measure $\mu_\Phi(A)$, is given by*

$$\mathcal{G}_\Phi(h) = \exp\left(\int (h(x) - 1) \mu_\Phi(\mathrm{d}x)\right), \tag{30}$$

*where $A \in \mathbb{R}$.*

Related functionals for computing the cumulants and factorial cumulants can be described as the logarithms of these quantities, eg.

$$\mathcal{W}_\Phi(f) = \log \mathcal{L}_\Phi(f), \tag{31}$$

Similarly to real-valued random variables, statistical moments can be defined for a point process $\Phi$ in order to provide an alternative description to its probability distribution $P_\Phi$ (or, equivalently, to its projection measures $P_\Phi^{(n)}$ for any $n \in \mathbb{N}$). The *n-th order moment measure* $\mu_\Phi^{(n)}$ of a point process $\Phi$ is the measure on $\mathcal{X}^n$ can be defined with [14]

$$\mu_\Phi^{(n)}(B_1 \times \cdots \times B_n) = \mathbb{E}_\Phi\left[\sum_{x_1,\ldots,x_n \in \Phi} \mathbb{1}_{B_1}(x_1)\ldots\mathbb{1}_{B_n}(x_n)\right], \tag{32}$$

for any regions $B_i \in \mathcal{B}(\mathcal{X})$, $1 \leq i \leq n$. The notation $\mathbb{1}_B$ denotes the indicator function, i.e., $\mathbb{1}_B(x) = 1$ if $x \in B$, and zero otherwise. The scalar $\mu_\Phi^{(n)}(B_1 \times \cdots \times B_n)$ estimates the joint localisation of sequence points within the regions $B_i$. For any Borel set $B \in_{\mathcal{X}}$, where $_{\mathcal{X}}$ is the Borel $\sigma$-algebra on $\mathcal{X}$, the integer-valued random variable

$$N_\Phi(B) = \sum_{x\in\Phi} 1_B(x) \tag{33}$$

counts the number of points falling inside $B$ according to the point process. We can determine the covariance of a point process $\Phi$ as [14], [15]

$$\mathrm{cov}_\Phi(N_\Phi(B), N_\Phi(B')) = \mu_\Phi^{(2)}(B \times B') - \mu_\Phi^{(1)}(B)\mu_\Phi^{(1)}(B'), \tag{34}$$

for any regions $B, B' \in \mathcal{B}(\mathcal{X})$.

Statistical quantities can be determined from the Laplace functional and p.g.fl. via differentiation. For instance, we can use the chain differential/derivative [16].

**Definition** III.4 (Chain differential). *given a functional $G$ and two functions $h, \eta : \mathcal{X} \to \mathbb{R}^+$, the (chain) differential of $G$ with respect to $h$ in the direction of $\eta$ is defined as [16]*

$$\delta G(h; \eta) = \lim_{n\to\infty} \frac{G(h + \varepsilon_n \eta_n) - G(h)}{\varepsilon_n}, \tag{35}$$

*when the limit exists and is identical for any sequence of real numbers $(\varepsilon_n)_{n\in\mathbb{N}}$ converging to $0$ and any sequence of functions $(\eta_n : \mathcal{X} \to \mathbb{R}^+)_{n\in\mathbb{N}}$ converging pointwise to $\eta$.*

The statistical quantities described in can then be extracted through the following differentiations:

$$P_\Phi^{(n)}(B_1 \times \cdots \times B_n) = \frac{1}{n!}\delta^n \mathcal{G}_\Phi(h; \mathbb{1}_{B_1}, \ldots, \mathbb{1}_{B_n})|_{h=0}, \tag{36}$$

$$\mu_\Phi^{(n)}(B_1 \times \cdots \times B_n) = (-1)^n \delta^n \mathcal{L}_\Phi(f; \mathbb{1}_{B_1}, \ldots, \mathbb{1}_{B_n})|_{f=0}, \tag{37}$$

for any regions $B_i \in \mathcal{B}(\mathcal{X})$, $1 \leq i \leq n$ [14].

The covariance can be determined via the second-order derivative of the cumulant generating functional, since the second-order cumulant is equal to the covariance, i.e.

$$\mathrm{cov}_\Phi(N_\Phi(A), N_\Phi(B)) = \delta^2 \mathcal{W}_\Phi(f; 1_A, 1_B)\Big|_{f=0}. \tag{38}$$

When a functional $G$ is defined as an integral with respect to a measure $\mu$ on $\mathcal{X}$ which is absolutely continuous with respect to the reference measure $\lambda$, the term $\delta G(f, \delta_x)$ will be understood as the Radon-Nikodym derivative of the measure $\mu' : B \mapsto \delta G(f, \mathbb{1}_B)$ evaluated at point $x$, i.e.

$$\delta G(f, \delta_x) := \frac{\mathrm{d}\mu'}{\mathrm{d}\lambda}(x), \tag{39}$$

for any appropriate function $f$ on $\mathcal{X}$ and any point $x \in \mathcal{X}$. In the context of this paper, this property holds for the pgfl $\mathcal{G}_\Phi$ of any point process $\Phi$ since its probability distribution $P_\Phi$ admits a density wrt the reference measure $\lambda$. In particular,

$$p_\Phi^{(n)}(x_1, \ldots, x_n) = \frac{1}{n!}\delta^n \mathcal{G}_\Phi(h; \delta_{x_1}, \ldots, \delta_{x_n})|_{h=0}, \tag{40}$$

for any points $x_i \in \mathcal{X}$, $1 \leq i \leq n$. Similarly, the first-order moment density, or *intensity function*, can be determined from the Laplace functional or p.g.fl. as follows

$$\begin{aligned}\mu_\Phi^{(1)}(x) &= \delta\mathcal{G}_\Phi(h; \delta_x)|_{h=1} \\ &= -\delta\mathcal{L}_\Phi(f; \delta_x)|_{f=0}.\end{aligned} \tag{41}$$

## IV. Entropy Statistics for Point Processes

In this section we generalise the statistics introduced in Section II to point processes. In the next subsection, we discuss the concepts in relation to Shannon entropy, and the following subsection their related Kullback-Leibler relative entropy counterparts.

Combining the information function of Golomb [2], and the probability generating functional of Moyal [7], we introduce an information generating functional as follows.

**Definition** IV.1 (Information generating functional for entropy). *Define the information generating functional for entropy for point processes with*

$$\mathcal{G}_\Phi^\alpha(h) = \mathbb{E}_\Phi\left[\left(\prod_{x\in\Phi} h(x)\right) p_\Phi^{-\alpha}\right] \tag{42}$$

$$= \sum_{n\geq 0} \int \prod_{i=1}^n h(x_i)\ p_\Phi^{(n)}(x_1, \ldots, x_n)^{-\alpha}\ P_\Phi^{(n)}(\mathrm{d}x_1, \ldots, \mathrm{d}x_n),$$

*where the densities and measures of a point process are defined in Section III.*

First note that if we set $h = 1$, then we can determine the Shannon entropy with

$$\frac{\partial}{\partial\alpha}\mathcal{G}_\Phi^\alpha(1)\bigg|_{\alpha=0} = -\mathbb{E}_\Phi\left[\log p_\Phi\right]. \tag{43}$$

More generally, we can determine entropy moments [2] with higher-order derivatives i.e.

$$\frac{\partial^m}{\partial\alpha^m}\mathcal{G}_\Phi^\alpha(1)\bigg|_{\alpha=0} = (-1)^m \mathbb{E}_\Phi\left[(\log p_\Phi)^m\right]. \tag{44}$$

Now consider using the functional in the same way as the probability generating functional to determine local entropy statistics. To determine the Shannon entropy related to $n$ points in the regions $B_1, \ldots, B_n$, we can use the analogous relations for the p.g.fl. [7]. Using the differentials defined in Section III. this is calculated with

$$\frac{\partial}{\partial\alpha}\,\delta^n\mathcal{G}^{\alpha}_{\Phi}(h;1_{B_1},\ldots,1_{B_n})\big|_{h=0,\alpha=0} = -\int_{B_1\times\ldots\times B_n}\log p^{(n)}_{\Phi}(x_1,\ldots,x_n)P^{(n)}_{\Phi}(\mathrm{d}x_1,\ldots,\mathrm{d}x_n). \tag{45}$$

This result is interpreted as the local contribution to the total Shannon entropy restricted to $n$ points in regions $B_1,\ldots,B_n$. Note that due to the linearity of the sum and integrals this is additive. Usage of this result may be too cumbersome to use in practice since it may require consideration of all possible numbers of points in specific regions. An alternative description of local entropy statistics can be considered by analogy of point process population moments.

**Definition** IV.2 (Entropy intensity). *We can determine an entropy statistic for a particular region that also considers the population number. We shall refer to this concept as the entropy intensity, which shall be defined with*

$$-\mathbb{E}_{\Phi}\left[N_{\Phi}(B)\log p_{\Phi}\right] = \frac{\partial}{\partial\alpha}\delta\mathcal{G}^{\alpha}_{\Phi}(h;1_B)|_{\alpha=0,h=1} = -\frac{\partial}{\partial\alpha}\delta\mathcal{L}^{\alpha}_{\Phi}(f;1_B)|_{\alpha=0,f=0}, \tag{46}$$

*where the definition of the counting measure $N_{\Phi}(B)$ is given in Section III.*

By analogy with the means of determining the intensity measure with the probability generating functional via differentiation, the entropy intensity can be found with the first-order functional derivative. Note that this the counting measure restricts the analysis of the entropy to a particular region. For instance, if the result is zero, then either the entropy is zero or there are no points. If the result is high, then either there are many points or the entropy is high.

Higher-order moments can be found by defining a functional in relation to the Laplace functional as follows.

**Definition** IV.3 (Laplace information functional for entropy). *For point processes, the Laplace information functional for Shannon entropy is determined with*

$$\mathcal{L}^{\alpha}_{\Phi}(f) = \mathbb{E}_{\Phi}\left[\exp\left(-\sum_{x\in\Phi}f(x)\right)p_{\Phi}^{-\alpha}\right]. \tag{47}$$

Setting $f = 0$ gives the equivalent to the information generating function, so that the Shannon entropy moments are found with

$$\mathbb{E}_{\Phi}\left[(\log p_{\Phi})^m\right] = (-1)^m\,\frac{\partial^m}{\partial\alpha^m}\mathcal{L}^{\alpha}_{\Phi}(f)\bigg|_{\alpha=0}. \tag{48}$$

Equivalently, setting $\alpha = 0$ gives the Laplace functional, so that the population moments become

$$\mathbb{E}_{\Phi}\left[N_{\Phi}(B_1)\ldots N_{\Phi}(B_n)\right] = (-1)^n\delta^n\mathcal{L}^{\alpha}_{\Phi}(f;1_{B_1},\ldots,1_{B_n})|_{f=0}. \tag{49}$$

**Definition** IV.4 (Joint entropy-population moments). *The joint entropy-population moments can be determined by differentiating with respect to $f$ and $\alpha$, i.e.*

$$\mathbb{E}_{\Phi}\left[\left(\prod_{i=1}^{n}N_{\Phi}(B_i)\right)(\log p_{\Phi})^m\right] = (-1)^{m+n}\frac{\partial^m}{\partial\alpha^m}\delta^n\mathcal{L}^{\alpha}_{\Phi}(f;1_{B_1},\ldots,1_{B_n})|_{f=0,\alpha=0}. \tag{50}$$

Following the definition of the Laplace information functional, an analogous cumulant information functional can be defined with

$$\mathcal{W}^{\alpha}_{\Phi}(f) = \log\mathcal{L}^{\alpha}_{\Phi}(f). \tag{51}$$

The second derivative of the cumulant generating function gives the covariance. Considering the Laplace functional ($\alpha = 0$), the population covariance between the number of points in $A$ and the number of points in $B$ can be determined with

$$\mathrm{cov}(N_{\Phi}(A),N_{\Phi}(B)) = \delta^2\mathcal{W}^{\alpha}_{\Phi}(f;1_A,1_B)|_{f=0,\alpha=0}. \tag{52}$$

Similarly setting $f = 0$ and differentiating with respect to $\alpha$ gives the variance in the entropy, i.e.

$$\mathrm{var}(\log p_{\Phi}) = \frac{\partial^2}{\partial\alpha^2}\mathcal{W}^{\alpha}_{\Phi}(0)\bigg|_{\alpha=0} = \mathbb{E}_{\Phi}\left[(\log p_{\Phi})^2\right] - \mathbb{E}_{\Phi}\left[\log p_{\Phi}\right]^2. \tag{53}$$

Another covariance statistic can be found between the number of points in a particular region $A$ and the entropy, which we define as follows.

**Definition** IV.5 (Entropy-population covariance). *The covariance between the entropy and the population of points in region $A$ is determined with*

$$\mathrm{cov}(N_{\Phi}(A),-\log p_{\Phi}) = \frac{\partial}{\partial\alpha}\delta\mathcal{W}^{\alpha}_{\Phi}(f;1_A)|_{f=0,\alpha=0} = -\mathbb{E}_{\Phi}\left[N_{\Phi}(A)\log p_{\Phi}\right] + \mathbb{E}_{\Phi}\left[N_{\Phi}(A)\right]\mathbb{E}_{\Phi}\left[\log p_{\Phi}\right]. \tag{54}$$

This statistic is interesting since it quantifies how much the number of points in a particular region (local) correlates with the entropy (global). Hence it can be computed in different regions and compared to determine which region correlates more strongly with the entropy. We give a simple example for exactly one point in the case of a Gaussian distribution.

*Example* 6 (Gaussian distribution). *Suppose that the population contains exactly one point, and that it is described with a Gaussian distribution $\mathcal{N}(x;m,P)$ with mean $m$ and variance $P = \sigma^2$. Then the Laplace information functional is given by*

$$\mathcal{L}^{\alpha}_{\Phi}(f) = \int\mathcal{N}(x;m,P)^{1-\alpha}\mathrm{e}^{-f(x)}\mathrm{d}x. \tag{55}$$

*Then the entropy-population covariance density is*

$$-\frac{\partial}{\partial\alpha}\delta\mathcal{W}^{\alpha}_{\Phi}(f;\delta_y)|_{f=0,\alpha=0} = -\mathcal{N}(y;m,P)\left(\log\mathcal{N}(y;m,P)+\frac{1}{2}\log(2\pi\mathrm{e}\sigma^2)\right), \tag{56}$$

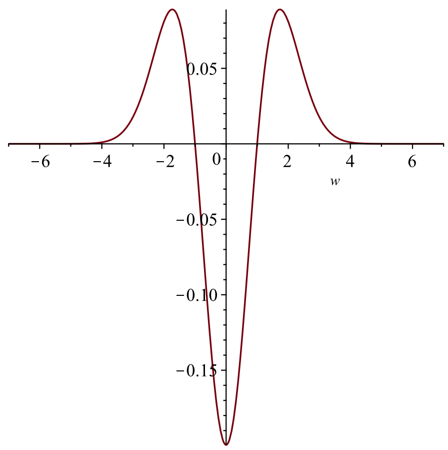



Fig. 1. Entropy-population covariance for Gaussian distribution.

*where we note that the Shannon entropy for a Gaussian distribution is equal to*

$$-\int \mathcal{N}(x; m, P) \log \mathcal{N}(x; m, P) \mathrm{d}x = \frac{1}{2}\log(2\pi \mathrm{e}\sigma^2). \quad (57)$$

*Figure 1 considers an example with mean $m = 0$, $\sigma = 1$. Note that the covariance is negative around the mean of the distribution, indicating the correlation least with the entropy, and the lobes show a positive correlation indicating a stronger correlation in the tails of the distribution.*

## V. Relative Entropy Statistics for Point Processes

In this section we extend the approach described in the previous section for relative entropy statistics. This will involve replacing the methods defined for Shannon entropy for Kullback-Leibler entropy using the same approach outlined in Section II. Hence, we begin by defining a information generating functional as follows.

**Definition** V.1 (Information generating functional for relative entropy). *Define the information generating functional for relative entropy between point processes $\Phi$ and $\Psi$ with*

$$\mathcal{G}^{\alpha}_{\Phi,\Psi}(h) = \mathbb{E}_{\Phi}\left[\left(\prod_{x\in\Phi} h(x)\right)\left(\frac{p_\Phi}{p_\Psi}\right)^{-\alpha}\right]. \quad (58)$$

We see that the Kullback-Leibler divergence is found with

$$-\frac{\partial}{\partial\alpha}\mathcal{G}^{\alpha}_{\Phi,\Psi}(1)\bigg|_{\alpha=0} = \mathbb{E}_{\Phi}\left[\log\frac{p_\Phi}{p_\Psi}\right]. \quad (59)$$

To determine the local Kullback-Leibler entropy statistic related to $n$ points in the regions $B_1, \ldots, B_n$, we can use the analogous relations for the p.g.fl. [7], eg.

$$-\frac{\partial}{\partial\alpha}\delta^n\mathcal{G}^{\alpha}_{\Phi,\Psi}(h; 1_{B_1}, \ldots, 1_{B_n})|_{h=0,\alpha=0} = \quad (60)$$
$$\int_{B_1\times\ldots\times B_n} \log\frac{p^{(n)}_\Phi(x_1,\ldots,x_n)}{p^{(n)}_\Psi(x_1,\ldots,x_n)} P^{(n)}_\Phi(\mathrm{d}x_1,\ldots,\mathrm{d}x_n).$$

**Definition** V.2 (Relative entropy intensity). *Following the result above for Shannon entropy, we can determine the first-order population statistic related to the relative entropy with*

$$\mathbb{E}_{\Phi}\left[N_\Phi(B)\log\frac{p_\Phi}{p_\Psi}\right]. \quad (61)$$

**Definition** V.3 (Laplace information functional for relative entropy). *For point processes, the Laplace information functional for relative entropy is determined with*

$$\mathcal{L}^{\alpha}_{\Phi,\Psi}(f) = \mathbb{E}_{\Phi}\left[\exp\left(-\sum_{x\in\Phi} f(x)\right)\left(\frac{p_\Phi}{p_\Psi}\right)^{-\alpha}\right]. \quad (62)$$

Similarly, a cumulant informational functional for relative entropy can be determined with

$$\mathcal{W}^{\alpha}_{\Phi,\Psi}(f) = \log\mathcal{L}^{\alpha}_{\Phi,\Psi}(f), \quad (63)$$

which enables the determination of the Rényi divergence with

$$\frac{1}{\alpha-1}\mathcal{W}^{\alpha}_{\Phi,\Psi}(0). \quad (64)$$

**Definition** V.4 (Relative entropy-population covariance). *As in the Shannon entropy, case, we can consider the covariance between the number of points in a particular region A and the entropy, i.e.*

$$\mathrm{cov}(N_\Phi(A), \log\frac{p_\Phi}{p_\Psi}) = \frac{\partial}{\partial\alpha}\delta\mathcal{W}^{\alpha}_{\Phi,\Psi}(f; 1_A)|_{f=0,\alpha=0} \quad (65)$$
$$= \mathbb{E}_{\Phi}\left[N_\Phi(A)\log\frac{p_\Phi}{p_\Psi}\right] - \mathbb{E}_{\Phi}\left[N_\Phi(A)\right]\mathbb{E}_{\Phi}\left[\log\frac{p_\Phi}{p_\Psi}\right].$$

This statistic is interesting since it quantifies how much the number of points in a particular region correlates with the global relative entropy between the two processes. An illustration of the use of this concept is given in the following section.

*Example* 7 (Poisson point process). *To illustrate the concept for point processes, we consider two Poisson processes $\Phi$ and $\Psi$, described with probability generating functionals given in Section III, where $\mu_\Phi$ and $\mu_\Phi$ are the intensity measures of processes $\Phi$ and $\Psi$ respectively. We shall assume that these admit densities with respect to the Lebesgue measure $\lambda$, which we shall write $\mu_\Phi(x)$ and $\mu_\Psi(x)$ respectively. Following the relative entropy between two Poisson distributions in Section II, we determine that the relative entropy Laplace information functional becomes*

$$\mathcal{L}^{\alpha}_{\Phi,\Psi}(f) = \exp\left(\int e^{-f(x)}\mu_\Phi(x)^{1-\alpha}\mu_\Psi(x)^{\alpha}(x)\lambda(\mathrm{d}x)\right) \quad (66)$$
$$\times\exp\left(-\int((1-\alpha)\mu_\Phi(x)+\alpha\mu_\Psi(x))\,\lambda(\mathrm{d}x)\right).$$

*The Kullback-Leibler divergence is found by setting $f = 0$ and differentiating with respect to $\alpha$, i.e.*

$$-\frac{\partial}{\partial\alpha}\mathcal{L}^{\alpha}_{\Phi,\Psi}(0)|_{\alpha=0} = \quad (67)$$
$$\int\left(\mu_\Phi(x)\log\left(\frac{\mu_\Phi(x)}{\mu_\Psi(x)}\right) - \mu_\Phi(x) + \mu_\Psi(x)\right)\lambda(\mathrm{d}x).$$

*To determine the information statistics defined above, we shall use the cumulant information functional, i.e. the logarithm of the Laplace information functional, which takes the simple form*

$$\mathcal{W}^{\alpha}_{\Phi,\Psi}(f) = \int e^{-f(x)}\mu_{\Phi}(x)^{1-\alpha}\mu_{\Psi}(x)^{\alpha}(x)\lambda(\mathrm{d}x) - \int ((1-\alpha)\mu_{\Phi}(x) + \alpha\mu_{\Psi}(x))\,\lambda(\mathrm{d}x). \tag{68}$$

*The Rényi divergence is then*

$$\frac{1}{1-\alpha}\mathcal{W}^{\alpha}_{\Phi,\Psi}(0) = \frac{1}{1-\alpha}\left(\int \mu_{\Phi}(x)^{1-\alpha}\mu_{\Psi}(x)^{\alpha}(x)\lambda(\mathrm{d}x) - \int ((1-\alpha)\mu_{\Phi}(x) + \alpha\mu_{\Psi}(x))\,\lambda(\mathrm{d}x)\right). \tag{69}$$

*Using Definition III.9, the relative entropy-population covariance becomes*

$$\frac{\partial}{\partial\alpha}\delta\mathcal{W}^{\alpha}_{\Phi,\Psi}(f;1_A)\Big|_{f=0,\alpha=0} = \int_A \log\left(\frac{\mu_{\Phi}(x)}{\mu_{\Psi}(x)}\right)\mu_{\Phi}(x)\lambda(\mathrm{d}x). \tag{70}$$

*Note that this closely resembles the Kullback-Leibler divergence, though the probabilities are replaced with point process intensities. We could also use the density form of this result, i.e.*

$$\frac{\partial}{\partial\alpha}\delta\mathcal{W}^{\alpha}_{\Phi,\Psi}(f;\delta_y)\Big|_{f=0,\alpha=0} = \mu_{\Phi}(y)\log\left(\frac{\mu_{\Phi}(y)}{\mu_{\Psi}(y)}\right). \tag{71}$$

*This example is illustrated in the following Section.*

## VI. Worked Examples

In this section we consider the application of the results for particular case studies to demonstrate the potential utility.

### A. Shannon Entropy Examples

Here we consider the Laplace functional of a Bernoulli point process, where the spatial density is described with a Gaussian distribution, i.e.

$$\mathcal{L}_{\Phi}(f) = (1-p) + p\int \mathcal{N}(x;m,P)\mathrm{e}^{-f(x)}\mathrm{d}x, \tag{72}$$

where $p$ is the probability of point existence, and $\mathcal{N}(x;m,P)$ is a Gaussian density function with argument $x$, mean $m$ and covariance $P$. We consider a scenario with potentially two points, described with independent Bernoulli point processes, i.e. their Laplace functional of the composite process is given by the superposition,

$$\mathcal{L}_{\Phi}(f) = \mathcal{L}_{\Phi_0}(f)\mathcal{L}_{\Phi_1}(f), \tag{73}$$

of independent Bernoulli components

$$\mathcal{L}_{\Phi_0}(f) = (1-p_0) + p_0\int \mathcal{N}(x;m_0,P_0)\exp(-f(x))\mathrm{d}x, \tag{74}$$

$$\mathcal{L}_{\Phi_1}(f) = (1-p_1) + p_1\int \mathcal{N}(x;m_1,P_1)\exp(-f(x))\mathrm{d}x. \tag{75}$$

We note that if each Bernoulli component were taken separately, their information functionals become

$$\mathcal{L}_{\Phi_0}(f) = (1-p_0)^{1-\alpha} + p_1^{1-\alpha}\int \mathcal{N}(x;m_0,P_0)^{1-\alpha}\mathrm{e}^{-f(x)}\mathrm{d}x, \tag{76}$$

$$\mathcal{L}_{\Phi_1}(f) = (1-p_1)^{1-\alpha} + p_1^{1-\alpha}\int \mathcal{N}(x;m_1,P_1)^{1-\alpha}\mathrm{e}^{-f(x)}\mathrm{d}x. \tag{77}$$

Unfortunately, it is important to note that the functional of the composite process no longer factorises as the product of individual independent components. In this case, the information functional becomes

$$\begin{aligned}\mathcal{L}^{\alpha}_{\Phi}(f) &= (p_0\ p_1)^{1-\alpha} + \\ &\int (p_0\mathcal{N}(x;m_0,P_0)(1-p_1) + \\ &p_1\mathcal{N}(x;m_1,P_1)(1-p_0))^{1-\alpha}\,\mathrm{e}^{-f(x)}\mathrm{d}x \\ &+ \int (p_0\mathcal{N}(x;m_0,P_0)p_1\mathcal{N}(x;m_1,P_1))^{1-\alpha}\,\mathrm{e}^{-f(x)-f(y)}\mathrm{d}x\mathrm{d}y.\end{aligned} \tag{78}$$

The first-moment, or intensity, is found in the usual way by considering the first-order derivative of the Laplace functional, i.e. by considering the derivative with respect to $f$ when $\alpha = 0$. Rather than use the measure, we use the density, so that the intensity function is calculated with

$$-\delta\mathcal{L}^{\alpha}_{\Phi}(f;\delta_y)\big|_{f=0} = p_0\mathcal{N}(y;m_0,P_0) + p_1\mathcal{N}(y;m_1,P_1). \tag{79}$$

Conversely, to find the Shannon entropy, we set $f = 0$ and differentiate with respect to $\alpha$. These operations are rather tedious in practice and lead to many terms and the results of these operations do not always lead to greater insight into their nature. Hence, we do not reproduce all of these derivatives. However, due to the mechanical nature, these operations can be automated by using computer algebra packages. In particular, the derivatives here were produced using the Physics package of Maple [17]. The entropy intensity is calculated with

$$-\frac{\partial}{\partial\alpha}\delta\mathcal{L}^{\alpha}_{\Phi}(f;\delta_y)\Big|_{\alpha=0,f=0}, \tag{80}$$

and the covariance between the entropy and the population is given with

$$\frac{\partial}{\partial\alpha}\delta\mathcal{W}^{\alpha}_{\Phi}(f;\delta_y)\Big|_{\alpha=0,f=0}. \tag{81}$$

To illustrate these results, we consider a one-dimensional scenario with parameters: $p_0 = 0.5, p_2 = 0.5, m_0 = -10, m_1 = 10, P_0 = 1, P_1 = 4$. This has been chosen to demonstrate the case with well-separated points, where the probability of existence is the same but the variances of the Gaussians are different. In this scenario, the intensity related to each point is the same, i.e. 0.5. Thus, if we compute the expectations localised around each point, it is not possible to distinguish between them. The entropy intensities are computed to be 1.93 and 2.10 respectively. Since the existence probabilities are equal, we can attribute the higher value to be

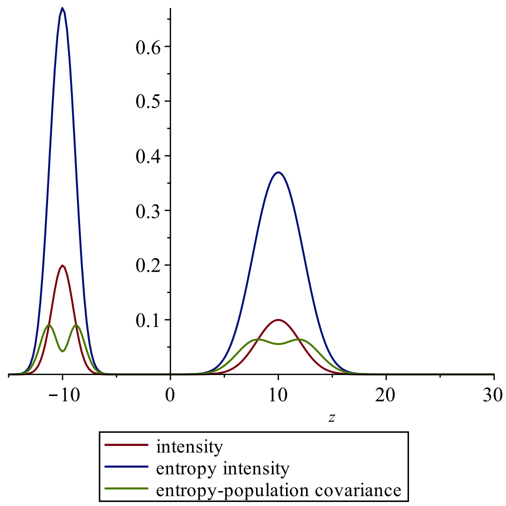



Fig. 2. Entropy statistics.

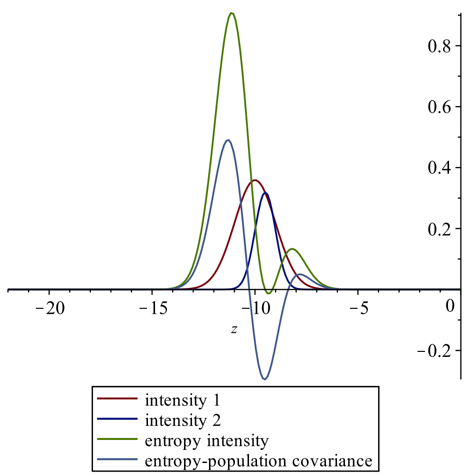
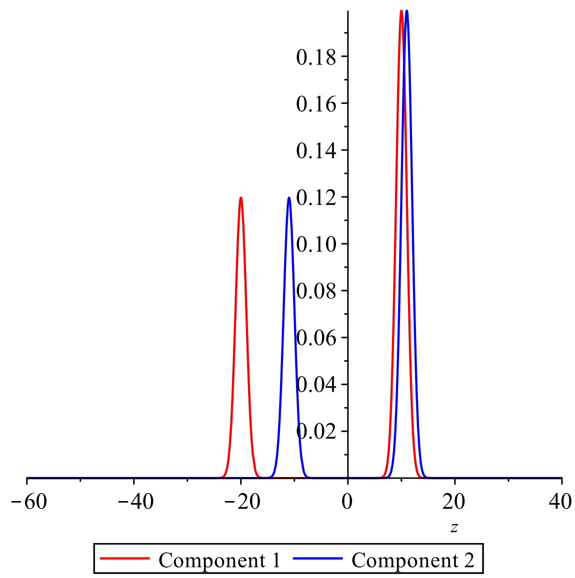
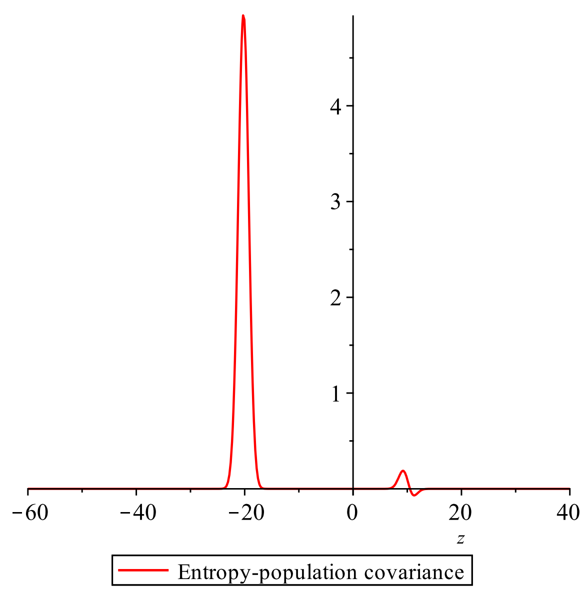



Fig. 3. Relative entropy statistics.




(a) Poisson intensity of two processes.




(b) Entropy-population covariance for relative entropy

Fig. 4. Relative entropy statistics between two Poisson components.

related to the higher variance in the second Gaussian component. The entropy-population covariance calculations restricted to the two components are 0.35 and 0.53, demonstrating that the second Gaussian is more strongly correlated with the entropy. Figure 2 shows the value of these functions. We can see a dip in the entropy-population covariance near the peak of the entropy, which relates to better localisation. Note that due to the linearity of the integral, the global covariance is the sum of these components, i.e. 0.88, giving an additive expression for the statistic.

When considering large populations with small global variations in the probability of existence, the entropy intensity may provide a useful statistic. However, this should be used with caution, since it is unable to disambiguate between high population number and high entropy. It is likely that the entropy-population covariance is a more useful local statistic, since it is able to quantify the correlation with the local intensity of the population with the global entropy. Higher values can be interpreted as regions with higher entropy.

### B. Relative Entropy Examples

*1) Relative Entropy Statistics Between Two Bernoulli Components:* In this section we consider the relative entropy between two Bernoulli components. The parameters for the components are $p_0 = 0.9, p_2 = 0.6, m_0 = -10, m_1 = -9.5, P_0 = 1, P_1 = 4$, and the quantities are described in Figure 3. The Kullback-Leibler divergence is the same as the entropy intensity for this scenario, 1.71, and the entropy-population covariance is equal to 0.43. We see that the entropy intensity is greatest when there is the least overlap between the intensities, and lowest when there is most overlap, as desired. The entropy-population covariance is also positive, showing a strong correlation between the population statistic and the entropy statistic. Importantly, the entropy-population covariance is negative when there is a strong overlap between the entropies, showing that there is a negative correlation between the intensities.

*2) Relative Entropy Statistics Between Two Poisson components:* In this section we consider the relative entropy entropy-covariance between two Poisson components, equation (71), where each of which is a mixture of two Gaussians with intensity functions

$$\mu_\Phi(y) = \omega_0 \mathcal{N}(y; m_0, P_0) + \omega_1 \mathcal{N}(y; m_1, P_1), \tag{82}$$

$$\mu_\Psi(y) = \omega_2 \mathcal{N}(y; m_2, P_2), +\omega_3 \mathcal{N}(y; m_3, P_3), \tag{83}$$

respectively. The parameters for the first Poisson component are $\omega_0 = 0.3, m_0 = -20, P_0 = 1, \omega_1 = 0.5, m_1 = 10, P_1 = 1$, and the second component parameters are $\omega_2 = 0.9, m_2 = 10, P_2 = 1, \omega_3 = 0.5, m_3 = -11, P_3 = 1$. The intensities of these processes on the is plotted in Figure 4(a). The entropy-population covariance is plotted in Figure 4(b). In the section $y < 0$, we see that there is a peak around $y = -20$, since Component 2 is different where Component 1 is non-zero. Note also that around $y = -11$, the entropy-population covariance is zero since Component 1 is zero. In the Section where $z > 0$, we see that there is a smaller contribution around $z = 10$ since the two Poisson components are similar but not equal. The integration on these two sections ($y < 0$ and $y > 0$) is 16.1 and 0.25 respectively which shows that a higher correlation to the relative entropy in the section where $y < 0$.

## VII. Discussion

By considering moment statistic calculations based on a new generating functional formulation, it is shown that it is possible to calculate localised information statistics. A new covariance statistic between the population of points and the entropy has been proposed that can help determine local regions where the entropy correlates most strongly with the population of points.

## References


[1] A. N. Kolmogorov, "La transformation de Laplace dans les espaces linéaires," *CR Acad. Sci. Paris*, vol. 200, pp. 1717–1718, Jan. 1935.

[2] S. Golomb, "The information generating function of a probability distribution (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 12, no. 1, pp. 75–77, Jan. 1966.

[3] J. A. McFadden, "The entropy of a point process," *J. Soc. Ind. Appl. Math.*, vol. 13, no. 4, pp. 988–994, 1965.

[4] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: General Theory and Structure*, vol. 2. New York, NY, USA: Springer, 2007.

[5] F. Baccelli and J. O. Woo, "On the entropy and mutual information of point processes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 695–699.

[6] M. S. Bartlett, "Processus stochastiques ponctuels," *Annales De L'Institut Henri Poincaré*, vol. 14, no. 1, pp. 35–60, 1954.

[7] J. E. Moyal, "The general theory of stochastic population processes," *Acta Mathematica*, vol. 108, no. 1, pp. 1–31, 1962.

[8] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[9] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[10] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Probab., Contrib. Theory Statist.*, vol. 1. Oakland, CA, USA: The Regents of the Univ. of California, 1961.

[11] S. Guiasu and C. Reischer, "The relative information generating function," *Inf. Sci.*, vol. 35, no. 3, pp. 235–241, 1985.

[12] E. D. Delande, M. Üney, J. Houssineau, and D. E. Clark, "Regional variance for multi-object filtering," *IEEE Trans. Signal Process.*, vol. 62, no. 13, pp. 3415–3428, Jul. 2014.

[13] I. Schlangen, E. D. Delande, J. Houssineau, and D. E. Clark, "A second-order PHD filter with mean and variance in target number," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 48–63, Jan. 2018.

[14] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*. Hoboken, NJ, USA: Wiley, 1997.

[15] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical Analysis and Modelling of Spatial Point Patterns*, vol. 70. Hoboken, NJ, USA: Wiley, 2008.

[16] P. Bernhard, "Chain differentials with an application to the mathematical fear operator," *Nonlinear Anal., Theory, Methods Appl.*, vol. 62, no. 7, pp. 1225–1233, Sep. 2005.

[17] E. S. Cheb-Terrab, "Maple procedures for partial and functional derivatives," *Comput. Phys. Commun.*, vol. 79, no. 3, pp. 409–424, 1994.