Kullback's inequality and Cramer Rao bounds for point process models

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

28-01-2022 / 01-02-2022

CITATION

Clark, Daniel (2022): Kullback's inequality and Cramer Rao bounds for point process models. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.19085531.v1>

DOI

[10.36227/techrxiv.19085531.v1](https://doi.org/10.36227/techrxiv.19085531.v1)

Kullback's inequality and Cramér Rao bounds for point process models

Daniel E. Clark, *Senior Member, IEEE*

Abstract—A lower bound on the Kullback-Leibler divergence known as Kullback's inequality can be determined with the Legendre transform of the cumulant generating function. The Cramér Rao bound can be derived from Kullback's inequality as the inverse of the second-order term in a Taylor expansion. Analogous forms for Kullback's inequality and the Cramér Rao bound for point processes were recently derived using functional methods from quantum field theory. This article develops Kullback's inequality and the Cramér Rao bound for point process parametrisations as performance bounds for models used in multi-object filtering.

Index Terms—Cramér's rate function, Kullback-Leibler divergence, Kullback's inequality, Legendre-Fenchel transform, cumulant generating functional, point processes, multi-target tracking.

I. INTRODUCTION

A. Cramér Rao bounds for multi-target tracking applications

The Cramér Rao bound, defined as the inverse of the Fisher information, provides a lower bound on the variance of an unbiased estimator. The need for Cramér Rao type bounds for performance analysis of multi-target tracking systems has been recognised for over 30 years [1]–[3]. The posterior Cramér Rao lower bound applies this concept recursively to a filtering model to determine the best possible achievable variance for an unbiased estimator, i.e. the theoretical optimum. This is now well established in the domain of target tracking for analysing the performance of different tracking algorithms. This has been extended to multi-target tracking, eg. [4]–[7]. The Fisher information was recently studied in the context of multi-target tracking parameters by Houssineau et al [8]. The current article develops a general form of Cramér Rao bound for multi-target performance analysis.

Prior developments of the Cramér Rao bound for multi-target tracking have focussed on determining a covariance matrix for joint distribution describing a set of known targets. While this is potentially useful, it lacks flexibility in some respects: Firstly, it does not provide any inherent localisation with respect to the targets under surveillance. Secondly, it does not take into account the uncertainty in the number of targets in the surveillance region. Thirdly, the trace of the matrix is used in practice, which only provides a single number for the whole multi-target scenario.

The concept of covariance for point processes is a significantly departure from the usual covariance matrix of a vector.

The covariance for a point process describes the covariance between the number of objects in two operator-specified regions. This is a statistic that can be extracted from the multi-target posterior distribution describing the population of targets. A posterior Cramér Rao bound for a general multi-target system that has would then describe the minimum achievable such covariance for an online surveillance application. It will also enable direct comparison of different multi-target trackers.

The bound on the minimum achievable variance for the estimator of a parameter, equal to the inverse of the Fisher information [9], was derived independently by Rao [10] and Cramér [11] is fundamental for statistical analysis. The Cramér Rao bound provides a minimum achievable variance or covariance for a parameter for a univariate or vector-valued parameter. However, multi-target systems often have parameters that are described by functions and the variance and covariance for point processes are themselves functions with spatial variates. Consequently, the usual formulation of the Cramér Rao bound in these contexts is not applicable for providing a bound for the covariance of the process. Hence, a more general approach is required.

The mean and covariance have been used to describe the posterior distribution of targets since the development of the Kalman filter in the 1960s. An analogous concept for populations of targets came much later [12], which describes the variance in the number of targets localised to a user-specified subset of the surveillance region. This has been used to compare different filters, and has been proposed to analyse the performance of filters [13].

The covariance of a point process is defined as the covariance between the number of objects in two subsets of the state space. Hence, the usual vector-valued formulation is not directly applicable to point processes. Indeed, the covariance in the number of objects between different regions, as proposed for multi-target tracking analysis [12], is more applicable to large-scale tracking problems, e.g. [14] and sensor management since it provides an aggregate statistic for populations of objects. The Cramér Rao bound for point processes, or multi-target processes, can then be viewed as a minimum achievable covariance for populations of objects in operator-defined surveillance regions.

B. Kullback's inequality and the Cramér Rao bound

The Kullback-Leibler divergence [15], or relative entropy, plays a fundamental role in information statistics for determining the difference between two different probability distributions. Applications of entropy and relative entropy divergence have been applied to point processes [16], [17]

* Télécom SudParis, Évry, 91001 France. Email: daniel.clark@telecom-sudparis.eu. This work was supported by the Joint AFRL-Dstl Basic-Research Grant in Autonomous Signal Processing (AFOSR grant FA9550-19-1-7008 and Dstl Task No. 1000133068).

since the 1960s [18], though calculation of these concepts in practice can be challenging and there have been a number of different approaches for their determination, eg. [19], [20]–[22]. The results in the paper make use of a result from large deviations theory [23]–[25], that a lower bound for relative entropy can be found by taking the Legendre-Fenchel transform of its cumulant generating function, or Cramér’s rate function [23], which is known as Kullback’s inequality [26]–[29]. A generalisation of this result was derived by Donsker and Varadhan [26].

The Cramér Rao bound can be derived [29] by taking the second-order derivative of Kullback’s inequality [27] which relates the Kullback-Leibler divergence [15] to the moment generating function of a random variable. Following this derivation, a form of Cramér Rao bound for point processes and random measures was recently proposed [30]. The approach makes use of a result for the effective action in quantum field theory [31, p289], and makes an analogy between the propagator and inverse propagator to the Fisher information [9] and Cramér Rao bound respectively. The main contribution in the paper is the generalisation of the concept of the Cramér Rao bound [10], [11] to point processes and random measures. The proof follows that of Fuchs and Letta [29] that starts with Kullback’s inequality [27] between the Kullback-Leibler divergence and Cramér’s rate function [32], and finds the Fisher information [9] as a limit that retains Kullback’s inequality. The approach exploits the connection between Cramér’s rate functional [30] and the effective action, or generating functional of proper vertices [33, p164]. In particular, it is shown that the second-order term of the generating functional expansion provides a lower bound for a quantity similar to the Fisher information. This is inverted, using the inverse relation of Jona-Lasinio [34] between the second-order functional derivatives of the effective action and the cumulant generating functional, to give the Cramér Rao bound. This provides a bound for the covariance of the process.

C. Paper outline

The paper is structured as follows. Appendix A draws on key results in convex analysis for calculating convex conjugates of functions with the Legendre-Fenchel transform [35]–[37]. These are used to present a new result for the Legendre transform of a mixture function. Appendix B reviews Cramér’s rate function for random variables and reviews how to determine this for different parametrised random variables. The next section characterises point processes and random measures with the Laplace functional and cumulant generating functional. The recently derived Cramér rate functional, Kullback’s inequality and the Cramér Rao bound for point processes [30] are summarised in Section III. These results are then applied to determine Cramér’s rate functional for superpositions and mixtures of point processes in Section IV. Section V determines Cramér’s rate functional for parametrised point process and random measure models using functional derivatives to determine the Legendre-Fenchel transform. Section VI presents a detailed examination of Kullback’s inequality and the Cramér Rao bound for a particular parametrisation used in multi-target

tracking [38]. Appendix C presents tables of Cramer’s rate function and functional for different parametrisations.

II. POINT PROCESSES

In this section we describe the fundamental descriptors for point processes and the notion of a covariance. Following Daley and Vere-Jones [39, p.7], define point processes as follows.

Definition II.1 (Point process). *A point process ξ_λ is a measurable mapping:*

$$\xi_\lambda : (\Omega, \mathcal{F}, \lambda) \rightarrow (E_X, \mathbf{B}(E_X))$$

from probability space $(\Omega, \mathcal{F}, \lambda)$ to $(E_X, \mathbf{B}(E_X))$, where E_X is the state space of all boundedly finite integer-valued measures, and $\mathbf{B}(E_X)$ is the related Borel σ -algebra.

We shall use the Laplace functional to characterise point processes [39, p57] defined as follows.

Definition II.2 (Laplace functional). *The Laplace functional of a random measure ξ_λ with probability measure λ is defined with*

$$\mathcal{L}_\lambda(f) = \int e^{-\int_X f(x) \xi_\lambda(dx)} \lambda(d\xi). \quad (1)$$

where $f : X \rightarrow \mathbb{R}^+$ is on the space of non-negative bounded measurable functions of bounded support.

The probability generating functional is related to the Laplace functional with $\mathcal{L}_{\mu^\psi}(f) = \mathcal{G}(e^{-f})$.

The intensity measure [40, p133] on $A \subset X$ is defined with $\phi(A) = \mathbb{E}[\xi(A)]$, and the intensity function $\phi(x)$ is defined with $\phi(A) = \int_A \phi(x) dx$.

A generating functional for the cumulants is defined through the relation $\mathcal{W}_\lambda(f) = \log \mathcal{L}_\lambda(f)$. Cumulant densities $w_\lambda^{(n)}$ are defined through the generating functional

$$\mathcal{W}_\lambda(f) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int_{X^n} \prod_{i=1}^n f(x_i) dx_i w_\lambda^{(n)}(x_1, \dots, x_n). \quad (2)$$

The second cumulant density of the point process is determined with

$$w_\lambda^{(2)}(x, y) = \left. \frac{\delta^2 \mathcal{W}_\lambda(f)}{\delta f(x) \delta f(y)} \right|_{f=0}. \quad (3)$$

The covariance of the point process ξ on $A \times B$ is given by [39, p69].

$$\text{cov}(\xi_\lambda(A), \xi_\lambda(B)) = \int_{A \times B} w_\lambda^{(2)}(x, y) dx dy. \quad (4)$$

The variance is found by setting $A = B$ in the covariance, i.e.

$$\text{var}(\xi_\lambda(A)) = \text{cov}(\xi_\lambda(A), \xi_\lambda(A)). \quad (5)$$

In the following section the Cramér Rao bound for point processes is described in relation to Kullback’s inequality for point processes recently derived [30].

III. KULLBACK'S INEQUALITY AND THE CRAMER RAO BOUND FOR POINT PROCESSES

This section summarises recently presented results by the author in [30] that are required to derive Kullback's inequality and the Cramer Rao bound for different point process models.

A. Kullback's inequality for point processes

We now define the rate functional for a point process recently introduced [30] as the following form of Legendre transform.

Definition III.1 (Cramér's rate functional for a point process [30]). *Define the Cramér rate functional $\Gamma_\lambda(\phi)$ to be Legendre transform of the cumulant generating functional $\mathcal{W}(f) = \log \mathcal{L}(f)$ of a point process.*

$$\Gamma_\lambda(\phi) = \sup_f \left[\int_X -f(x)\phi(x)dx - \mathcal{W}_\lambda(f) \right], \quad (6)$$

where f is a non-negative bounded measurable function of bounded support.

Definition III.2 (Kullback Leibler divergence for point processes). *Suppose that λ and μ are probability distributions of point processes. Then the Kullback-Leibler divergence [15] from λ and μ is defined as follows*

$$\mathcal{D}(\mu\|\lambda) = \int_{\mathcal{X}} \log \left(\frac{d\mu}{d\lambda}(\xi) \right) \mu(d\xi),$$

when μ is absolutely continuous with respect to λ , and infinite otherwise. Note that $0 \leq \mathcal{D}(\mu'\|\mu) \leq +\infty$, and $\mathcal{D}(\mu'\|\mu) = 0$ if and only if $\mu = \mu'$.

Theorem III.1 (Kullback's inequality for point processes [30]). *Let λ and μ^ϕ be probability distributions of point processes, where Γ_λ is the rate functional for point process with distribution λ , and μ^ϕ is absolutely continuous with respect to λ . Let ϕ be the intensity function of a point process with probability distribution μ^ϕ . Then the following inequality holds, cf. [41, p38] [29]*

$$\mathcal{D}(\mu^\phi\|\lambda) \geq \Gamma_\lambda(\phi). \quad (7)$$

Example III.1 (Poisson point process). *The Laplace functional of a Poisson point process with intensity function $\mu(x)$ is defined with [17]*

$$\mathcal{L}(f) = \exp \left(\int (e^{-f(x)} - 1) \mu(x) dx \right), \quad (8)$$

where $\mu(x)$ is the intensity function. Cramér's rate functional [30] is given by

$$\Gamma(\phi) = \int \phi(x) \log \frac{\phi(x)}{\mu(x)} dx - \int (\phi(x) - \mu(x)) dx. \quad (9)$$

We note that this is the Kullback-Leibler divergence between two Poisson distributions with intensity functions $\phi(x)$ and $\mu(x)$, eg. [19].

B. The Cramér Rao bound for point processes

Fuchs and Letta [29] derived the Cramér Rao bound as a second-order relation from Kullback's inequality. This proof

was recently developed for point processes [30]. Consider the following terms

$$I_{\mu^\psi}^{(2)}(x, y) = \frac{\delta^2}{\delta\phi(x)\delta\phi(y)} \mathcal{D}(\mu^\phi\|\mu^\psi) \Big|_{\phi=\psi} \quad (10)$$

$$\gamma_{\mu^\psi}^{(2)}(x, y) = \frac{\delta^2}{\delta\phi(x)\delta\phi(y)} \Gamma_\lambda(\phi) \Big|_{\phi=\psi}. \quad (11)$$

Then it was shown that following inequality holds

$$I_{\mu^\psi}^{(2)}(x, y) \geq \gamma_{\mu^\psi}^{(2)}(x, y). \quad (12)$$

The inverse of this relation gives a Cramér Rao bound [30]. However, the determination of the inverse is more challenging in a functional context. Jona-Lasinio [34] derived a relation between the second-order derivatives of the cumulant generating functional and the effective action as follows which we use to determine the inverse,

$$\delta(y - x) = \int_X \frac{\delta^2 \Gamma_{\mu^\psi}(\phi)}{\delta\phi(x)\delta\phi(z)} \frac{\delta^2 \mathcal{W}_{\mu^\psi}(f)}{\delta f(y)\delta f(z)} dz. \quad (13)$$

To find the inverse, we consider the following $\mathcal{I}_{\mu^\psi}(\phi) = \mathcal{D}(\mu^\phi\|\mu^\psi)$ and its Legendre transform, $\mathcal{V}_{\mu^\psi}(f)$, where

$$\mathcal{V}_{\mu^\psi}(f) = \sup_\phi \left[\int -f(x)\phi(x)dx - \mathcal{I}_{\mu^\psi}(\phi) \right]. \quad (14)$$

The relation becomes

$$\delta(y - x) = \int_X \frac{\delta^2 \mathcal{I}_{\mu^\psi}(\phi)}{\delta\phi(x)\delta\phi(z)} \frac{\delta^2 \mathcal{V}_{\mu^\psi}(f)}{\delta f(y)\delta f(z)} dz. \quad (15)$$

Now define the second-order terms,

$$g_{\mu^\psi}^{(2)}(x, y) = \frac{\delta^2}{\delta f(x)\delta f(y)} \mathcal{W}_{\mu^\psi}(f) \Big|_{f=0}, \quad (16)$$

$$v_{\mu^\psi}^{(2)}(x, y) = \frac{\delta^2}{\delta f(x)\delta f(y)} \mathcal{V}_{\mu^\psi}(f) \Big|_{f=0}. \quad (17)$$

Then the following inequality holds [30]

$$g_{\mu^\psi}^{(2)}(x, y) \geq v_{\mu^\psi}^{(2)}(x, y). \quad (18)$$

The covariance of the point process ξ_{μ^ψ} parameterised with intensity function ψ is then bounded with the following inequality [30],

$$\begin{aligned} \text{cov}(\xi_{\mu^\psi}(A), \xi_{\mu^\psi}(B)) &= \int_{A \times B} g_{\mu^\psi}^{(2)}(x, y) dx dy \\ &\geq \int_{A \times B} v_{\mu^\psi}^{(2)}(x, y) dx dy. \end{aligned} \quad (19)$$

This result shall be used in Section VI to analyse a particular point process model used in multi-target tracking.

IV. CRAMÉR'S RATE FUNCTIONAL FOR SUPERPOSITIONS AND MIXTURE POINT PROCESSES

A. Superpositions

The following Lemma show how to determine Cramér's rate functional for superpositions of point processes. A similar result was recently presented by Zajkowski [42] in the context of sums of random variables.

Lemma IV.1 (Legendre-Fenchel transform of a superposition of point processes). *Suppose that the distribution is described with a superposition of independent distributions described with characteristic functionals $\mathcal{L}_{\lambda_1}(f), \dots, \mathcal{L}_{\lambda_n}(f)$, i.e.*

$$\mathcal{L}_\lambda(f) = \prod_{i=1}^n \mathcal{L}_{\lambda_i}(f). \quad (20)$$

Then the Legendre-Fenchel transform of the cumulant generating functional $\mathcal{W}_\lambda(f) = \log \mathcal{L}_\lambda(f)$ is given by

$$(\mathcal{W}_{\lambda_1} + \dots + \mathcal{W}_{\lambda_n})^*(\phi) = \min_{\phi = \phi_1 + \dots + \phi_n} (\Gamma_{\lambda_1}(\phi_1) + \dots + \Gamma_{\lambda_n}(\phi_n)), \quad (21)$$

where $\Gamma_{\lambda_i}(\phi_i) = \mathcal{W}_{\lambda_i}^(\phi_i)$ is the Legendre-Fenchel transform of the i^{th} cumulant generating function.*

Proof. Note that the cumulant generating function is the sum of the component cumulant generating functions $\mathcal{W}_{\lambda_i}(f) = \log \mathcal{L}_{\lambda_i}(f)$, i.e.

$$\mathcal{W}_\lambda(f) = \sum_{i=1}^n \mathcal{W}_{\lambda_i}(f). \quad (22)$$

The proof follows from Theorem A.1. $\square$

Corollary IV.1 (Minimum Kullback-Leibler divergence for superposition processes). *The minimum Kullback-Leibler divergence between the superposition λ of n processes, $\lambda_1, \dots, \lambda_n$, and a process with mean ϕ is given by*

$$\begin{aligned} \Gamma_\lambda(\phi) &= \inf_{\mu: \mathbb{E}[\mu] = \phi} \mathcal{D}(P_\mu \| P_\lambda) \\ &= \min_{\phi = \phi_1 + \dots + \phi_n} \sum_{i=1}^n \inf_{\nu_i: \mathbb{E}[\nu_i] = \phi_i} \mathcal{D}(P_{\nu_i} \| P_{\lambda_i}), \end{aligned} \quad (23)$$

where $\bar{\nu}_i = \mathbb{E}[\nu_i]$ the process μ is of the form of a superposition of n processes, i.e.

$$\mathcal{L}_\mu(f) = \prod_{i=1}^n \mathcal{L}_{\mu_i}(f), \quad (24)$$

and the mean is of the form

$$\mathbb{E}[\mu] = \sum_{i=1}^n \mathbb{E}[\mu_i]. \quad (25)$$

This corollary gives an expression for the minimum Kullback-Leibler divergence for superpositions of n point processes or random measures. In the context of point processes, the rate function takes the first-order moment measure, or intensity measure, as its argument to return a lower bound on the relative entropy.

B. Mixture processes

The Cramér rate function for mixture distributions is a novel result presented in Appendix B, Lemma B.2 based on Theorem A.4, Appendix A. The following Lemma shows the analogous result for a mixture of point processes.

Lemma IV.2 (Legendre-Fenchel transform of a mixture of point processes). *Suppose that the process λ is described with a mixture of Laplace functionals $\mathcal{L}_{\lambda_1}(f), \dots, \mathcal{L}_{\lambda_n}(f)$, i.e.*

$$\mathcal{L}_\lambda(f) = \sum_{i=1}^n w_{\lambda_i} \mathcal{L}_{\lambda_i}(f), \quad (26)$$

where $w_{\lambda_1} + \dots + w_{\lambda_n} = 1$. Then the Legendre-Fenchel transform of the cumulant generating function is given by

$$\begin{aligned} \Gamma_\lambda(\phi) &= \left(\log \sum_{i=1}^n w_{\lambda_i} \mathcal{L}_{\lambda_i} \right)^*(\phi) \\ &= \min_{\substack{w_{\mu_1} + \dots + w_{\mu_n} = 1, \\ \phi_1 + \dots + \phi_n = \phi}} \left(\sum_{i=1}^n w_{\mu_i} \log \left(\frac{w_{\mu_i}}{w_{\lambda_i}} \right) + w_{\mu_i} \Gamma_{\lambda_i} \left(\frac{\phi_i}{w_{\mu_i}} \right) \right), \end{aligned} \quad (27)$$

where Γ_{λ_i} is the Legendre-Fenchel transform of cumulant generating function $\mathcal{W}_{\lambda_i}(f)$

Proof. Note that

$$\Gamma_\lambda(\phi) = \left(\log \sum_{i=1}^n w_{\lambda_i} \exp(\mathcal{W}_{\lambda_i}) \right)^*(\phi)$$

and $w_{\lambda_i} \mathcal{L}_{\lambda_i}(f) = \exp(\mathcal{W}_{\lambda_i}(f) + \log w_{\lambda_i})$. The proof follows from Theorem A.4. $\square$

Corollary IV.2 (Minimum Kullback-Leibler divergence for mixture processes). *The minimum Kullback-Leibler divergence between a mixture λ of n processes, $\lambda_1, \dots, \lambda_n$, and a process with mean ϕ is given by*

$$\begin{aligned} \Gamma_\lambda(\phi) &= \min_{\substack{w_{\mu_1} + \dots + w_{\mu_n} = 1, \\ \phi_1 + \dots + \phi_n = \phi}} (\mathcal{D}(w_\mu \| w_\lambda) + \\ &\quad \sum_{i=1}^n w_{\mu_i} \left(\inf_{\nu_i: \mathbb{E}[\nu_i] = \phi_i} \mathcal{D}(P_{\nu_i} \| P_{\lambda_i}) \right)), \end{aligned} \quad (28)$$

where $\bar{\nu}_i = \mathbb{E}[\nu_i]$, the term on the left is the Kullback-Leibler divergence between the weights of the mixtures, i.e.

$$\mathcal{D}(w_\mu \| w_\lambda) = \sum_{i=1}^n w_{\mu_i} \log \left(\frac{w_{\mu_i}}{w_{\lambda_i}} \right), \quad (29)$$

and the process ϕ is described with a mixture of Laplace functionals, i.e. of the form

$$\mathcal{L}_\mu(f) = \sum_{i=1}^n w_{\mu_i} \mathcal{L}_{\mu_i}(f), \quad (30)$$

where $w_{\mu_1} + \dots + w_{\mu_n} = 1$, and the mean of the process is given by

$$\mathbb{E}[\mu] = \sum_{i=1}^n w_{\mu_i} \mathbb{E}[\mu_i]. \quad (31)$$

This corollary gives an expression for the minimum Kullback-Leibler divergence for mixtures of n point processes or random measures for a given intensity measure.

V. CRAMÉR'S RATE FUNCTIONAL FOR DIFFERENT PARAMETRISATIONS

This section derives the rate functional for some common parametrisations.

Example V.1 (Bernoulli point process). *The Bernoulli point process extends the Bernoulli distribution to the case with intensity $\psi(x)$, eg. [39, p530]. The Laplace functional of a Bernoulli process is given by, eg. [19],*

$$\mathcal{L}(f) = \int 1 - \psi(x) + e^{-f(x)}\psi(x)dx. \quad (32)$$

Cramér's rate function is given by

$$\begin{aligned} \Gamma(\phi) = & \int \phi(x) \log \left(\frac{\phi(x)}{\psi(x)} \right) dx \\ & + \left(1 - \int \phi(x)dx \right) \log \left(\frac{1 - \int \phi(x)dx}{1 - \int \psi(x)dx} \right). \end{aligned} \quad (33)$$

Proof. The cumulant generating functional is then

$$\mathcal{W}(f) = \log \left(\int 1 - \psi(x) + e^{-f(x)}\psi(x)dx \right). \quad (34)$$

The Legendre-Fenchel transform then becomes

$$\begin{aligned} \mathcal{W}^*(\phi) = & \sup_f \left(\langle \phi, -f \rangle - \log \left(\int 1 - \psi(x) + \int e^{-f(x)}\psi(x)dx \right) \right). \end{aligned} \quad (35)$$

As above, we take the functional derivative to give

$$\begin{aligned} \phi(y) = & \frac{-\delta}{\delta f(y)} \mathcal{W}(f) \\ = & \frac{e^{-f(y)}\psi(y)}{\mathcal{L}(f)}, \end{aligned} \quad (36)$$

which gives

$$e^{-f(y)}\psi(y) = \phi(y)\mathcal{L}(f). \quad (37)$$

Rearranging for $-f(y)$, we have

$$-f(y) = \log \left(\frac{\phi(y)}{\psi(y)} \right) + \mathcal{W}(f). \quad (38)$$

Substituting (37) into the definition of the Laplace functional, we get

$$\mathcal{L}(f) = \int 1 - \psi(y) + \phi(y)\mathcal{L}(f)dy, \quad (39)$$

and thus

$$\mathcal{L}(f) = \frac{1 - \int \psi(y)dy}{1 - \int \phi(y)dy}, \quad (40)$$

which, substituting into (38), we get

$$-f(y) = \log \left(\frac{\phi(y)}{\psi(y)} \right) + \log \left(\frac{1 - \int \psi(y)dy}{1 - \int \phi(y)dy} \right). \quad (41)$$

Finally, substitution of $-f$ into (35), we get the result. $\square$

We note that this is the Kullback-Leibler divergence between two Bernoulli processes with rates $\phi(x)$ and $\psi(x)$, eg. [19], [43].

Example V.2 (Panjer point process). *The Laplace functional of a Panjer point process [44] is given by*

$$\mathcal{L}(f) = \left(1 - (1 - \int e^{-f(x)}s(x)dx)/\beta \right)^{-\alpha}. \quad (42)$$

Cramér's rate function is given by

$$\begin{aligned} \Gamma(\phi) = & \int \phi(x) \log \left(\frac{\phi(x)(\beta+1)}{\alpha s(x) + \phi(x)} \right) + \alpha \log \left(\frac{\alpha s(x)(\beta+1)}{\beta(\alpha s(x) + \phi(x))} \right) dx. \end{aligned} \quad (43)$$

Proof. The cumulant generating functional becomes

$$\mathcal{W}(f) = -\alpha \log \left(1 - (1 - \int e^{-f(x)}s(x)dx)/\beta \right). \quad (44)$$

Taking the functional derivative leads to

$$\phi(y) = \frac{-\delta}{\delta f(y)} \mathcal{W}(f) = \frac{\alpha e^{-f(y)} \frac{1}{\beta} s(y)}{(1 - (1 - \int e^{-f(x)}s(x)dx)/\beta)} \quad (45)$$

Then $-f(y) = \log \left(\frac{\phi(y)(\beta+1)}{\alpha s(y) + \phi(y)} \right)$ is substituted into the Legendre-Fenchel transform. $\square$

Example V.3 (Binomial point process). *Following the extension of a Bernoulli distribution to a binomial distribution, the Laplace functional of a binomial point process is given by*

$$\mathcal{L}(f) = \left(\int 1 - \psi(x) + \psi(x)e^{-f(x)}dx \right)^n. \quad (46)$$

Cramér's rate functional is given by

$$\begin{aligned} \Gamma(\phi) = & \int \phi(x) \log \left(\frac{\phi(x)}{n\psi(x)} \right) dx \\ & + \left(n - \int \phi(x)dx \right) \log \left(\frac{1 - \int \frac{\phi(x)}{n}dx}{1 - \int \psi(x)dx} \right). \end{aligned} \quad (47)$$

Proof. The cumulant generating functional becomes

$$\mathcal{W}(f) = n \log \left(\int 1 - \psi(x) + \psi(x)e^{-f(x)}dx \right). \quad (48)$$

The result follows by using and the result for the Bernoulli point process and the scaling property for Legendre transforms, i.e. $(n\Lambda)^*(\phi) = n\Lambda^*\left(\frac{\phi}{n}\right)$. $\square$

Example V.4 (Poisson-binomial process). *Consider the Laplace functional of the Poisson-binomial process, which is composed of n Bernoulli processes [45], i.e.*

$$\mathcal{L}(f) = \prod_{i=1}^n \left(\int 1 - \psi_i(x) + e^{-f(x)}\psi_i(x)dx \right). \quad (49)$$

Cramér's rate function is given by

$$\begin{aligned} \Gamma(\phi) = & \min_{\phi = \phi_1 + \dots + \phi_n} \left(\sum_{i=1}^n \int \phi_i(x) \log \left(\frac{\phi_i(x)}{\psi_i(x)} \right) dx \right. \\ & \left. + \left(1 - \int \phi_i(x)dx \right) \log \left(\frac{1 - \int \phi_i(x)dx}{1 - \int \psi_i(x)dx} \right) \right) \end{aligned} \quad (50)$$

Proof. The proof follows from Lemma V.1 and Theorem A.1. $\square$

VI. APPLICATION TO A MULTI-TARGET TRACKING MODEL

A. Analysis of the Poisson multi-Bernoulli mixture process

In this section we analyse a particular model used for multi-target tracking applications, known as the Poisson multi-Bernoulli mixture process introduced by Williams [38] which is a mixture process where each term in the mixture is the superposition of a Poisson process and a Poisson-binomial process. This is chosen since it involves both superpositions and a mixture, hence we can demonstrate the application of the results introduced in the previous sections. The probability generating functional is of the following form,

$$\mathcal{G}(h) = \sum_{j=1}^N w^{(j)} \mathcal{G}^{(j)}(h), \quad (51)$$

$$\mathcal{G}^{(j)}(h) = \mathcal{G}_p^{(j)}(h) \prod_{i=1}^{n(j)} \mathcal{G}_b^{(i,j)}(h) \quad (52)$$

where each $\mathcal{G}_p^{(j)}(h)$ is Poisson and each $\mathcal{G}_b^{(i,j)}(h)$ is Bernoulli with intensities $\mu_p^{(j)}$ and $\mu_b^{(i,j)}$ respectively and the following probability generating functionals.

$$\mathcal{G}_p^{(j)}(h) = \exp \left(\int (h(x) - 1) \mu_p^{(j)}(x) dx \right), \quad (53)$$

$$\mathcal{G}_b^{(i,j)}(h) = \int 1 - \mu_b^{(i,j)}(x) + h(x) \mu_b^{(i,j)}(x) dx. \quad (54)$$

The related Laplace functionals have the same structure, i.e.

$$\mathcal{L}(f) = \sum_{j=1}^N w^{(j)} \mathcal{L}^{(j)}(f), \quad (55)$$

$$\mathcal{L}^{(j)}(f) = \mathcal{L}_p^{(j)}(f) \prod_{i=1}^{n(j)} \mathcal{L}_b^{(i,j)}(f). \quad (56)$$

The cumulant generating functional is then given by

$$\mathcal{W}(f) = \log \left(\sum_{j=1}^N w^{(j)} \mathcal{L}^{(j)}(f) \right). \quad (57)$$

It is straightforward to show that the intensity of the whole process is given by

$$\mu(x) = \sum_{j=1}^N w^{(j)} \left(\mu_p^{(j)}(x) + \sum_{i=1}^{n(j)} \mu_b^{(i,j)}(x) \right). \quad (58)$$

The following lemma computes the Cramér rate functional for this model.

Lemma VI.1 (Cramér rate functional for Poisson multi-Bernoulli mixture process). *The rate functional for the process described above is of the form*

$$\Gamma(\phi) = \min_{\substack{\alpha^{(1)} + \dots + \alpha^{(N)} = 1, \\ \phi^{(1)} + \dots + \phi^{(N)} = \phi}} \left(\sum_{i=1}^N \alpha^{(i)} \log \left(\frac{\alpha^{(i)}}{w^{(i)}} \right) + \alpha^{(i)} \Gamma^{(i)} \left(\frac{\phi^{(i)}}{\alpha^{(i)}} \right) \right), \quad (59)$$

where the term on the right is given by

$$\Gamma^{(j)}(\eta) = \min_{\eta = \phi_p^{(j)} + \phi_b^{(1,j)} + \dots + \phi_b^{(n(j),j)}} \left(\Gamma_p^{(j)}(\phi_p^{(j)}) + \sum_{i=1}^{n(j)} \Gamma_b^{(i,j)}(\phi_b^{(i,j)}) \right), \quad (60)$$

whose terms are calculated with the rate functionals for the Poisson and Bernoulli processes, i.e.

$$\Gamma_p^{(j)}(\phi_p^{(j)}) = \int \phi_p^{(j)}(x) \log \left(\frac{\phi_p^{(j)}(x)}{\mu_p^{(j)}(x)} \right) dx - \int (\phi_p^{(j)}(x) - \mu_p^{(j)}(x)) dx, \quad (61)$$

$$\Gamma_b^{(i,j)}(\phi_b^{(i,j)}) = \int \phi_b^{(i,j)}(x) \log \left(\frac{\phi_b^{(i,j)}(x)}{\mu_b^{(i,j)}(x)} \right) dx + \left(1 - \int \phi_b^{(i,j)}(x) dx \right) \log \left(\frac{1 - \int \phi_b^{(i,j)}(x) dx}{1 - \int \mu_b^{(i,j)}(x) dx} \right). \quad (62)$$

Proof. The proof follows from first applying Lemma IV.2 to determine the rate functional for the mixture process, then Lemma IV.1 for the superposition in each mixture component, and finally we use the rate functionals for the Poisson and Bernoulli processes in Examples III.1 and V.1 respectively. $\square$

The following Lemma shows that the Kullback-Leibler divergence of two Poisson multi-Bernoulli processes with the same structure is the same form as the rate functional.

Lemma VI.2 (Kullback-Leibler divergence between two Poisson multi-Bernoulli mixture processes of the same structure). *The probability generating functional is of the following form,*

$$\mathcal{G}'(h) = \sum_{j=1}^N \alpha^{(j)} \mathcal{G}'^{(j)}(h), \quad (64)$$

$$\mathcal{G}'^{(j)}(h) = \mathcal{G}'_p^{(j)}(h) \prod_{i=1}^{n(j)} \mathcal{G}'_b^{(i,j)}(h), \quad (65)$$

where

$$\mathcal{G}'_p^{(j)}(h) = \exp \left(\int (h(x) - 1) \phi_p^{(j)}(x) dx \right), \quad (66)$$

$$\mathcal{G}'_b^{(i,j)}(h) = \int 1 - \phi_b^{(i,j)}(x) + h(x) \phi_b^{(i,j)}(x) dx. \quad (67)$$

The Kullback-Leibler divergence between this process and the previously defined Poisson multi-Bernoulli process is given by

$$\mathcal{D}(P_\phi \| P_\mu) = \sum_{i=1}^N \alpha^{(i)} \log \left(\frac{\alpha^{(i)}}{w^{(i)}} \right) + \alpha^{(i)} \Gamma^{(i)} \left(\frac{\phi^{(i)}}{\alpha^{(i)}} \right), \quad (68)$$

where $\Gamma^{(i)}$ is defined in Lemma VI.1, i.e.

$$\Gamma^{(j)}(\phi^{(j)}) = \Gamma_p^{(j)}(\phi_p^{(j)}) + \sum_{i=1}^{n(j)} \Gamma_b^{(i,j)}(\phi_b^{(i,j)}), \quad (69)$$

and $\Gamma_p^{(j)}, \Gamma_b^{(i,j)}$ are the terms defined in Lemma VI.1.

Proof. Note that this takes the same form as in Corollary IV.2 since the Poisson and Bernoulli processes are uniquely characterised by their intensity functions hence their rate functionals are equal to the Kullback-Leibler divergence. However, we can show this directly as follows. We first apply the chain rule for Kullback-Leibler divergence, see e.g. [46, Theorem 2.5.3]. More specifically, the Kullback-Leibler divergence between mixtures of n processes, $\lambda_1, \dots, \lambda_n$, and $\nu_1, \dots, \nu_n$, with respective weights $w_{\lambda_1}, \dots, w_{\lambda_n}$, and $w_{\nu_1}, \dots, w_{\nu_n}$, i.e.

$$\mathcal{D}(P_\nu \| P_\lambda) = \sum_{i=1}^n w_{\nu_i} \log \left(\frac{w_{\nu_i}}{w_{\lambda_i}} \right) + \sum_{i=1}^n w_{\nu_i} \mathcal{D}(P_{\nu_i} \| P_{\lambda_i}), \quad (70)$$

where the term on the left is the Kullback-Leibler divergence between the weights of the mixtures. Each component j is composed of the superposition of an independent Poisson process and independent Bernoulli processes, hence the divergence is additive. Finally note that the divergence between two Poisson processes or Bernoulli processes are equal to their respective rate functionals. $\square$

The following lemma determines the covariance of the process.

Lemma VI.3 (The covariance of a Poisson multi-Bernoulli mixture process). *The covariance of the Poisson multi-Bernoulli mixture process defined above is of the form*

$$\begin{aligned} \text{cov}(\xi(A), \xi(B)) &= \sum_{j=1}^N w^{(j)} \text{cov}(\xi^{(j)}(A), \xi^{(j)}(B)) \\ &+ \sum_{j=1}^N w^{(j)} \mu^{(j)}(A) \mu^{(j)}(B) \\ &- \left(\sum_{j=1}^N w^{(j)} \mu^{(j)}(A) \right) \left(\sum_{j=1}^N w^{(j)} \mu^{(j)}(B) \right), \end{aligned} \quad (71)$$

where the j^{th} covariance and intensities are given by

$$\text{cov}(\xi^{(j)}(A), \xi^{(j)}(B)) = \quad (72)$$

$$\text{cov}(\xi_p^{(j)}(A), \xi_p^{(j)}(B)) + \sum_{i=1}^{n(j)} \text{cov}(\xi_b^{(i,j)}(A), \xi_b^{(i,j)}(B)),$$

$$\mu^{(j)}(A) = \mu_p^{(j)}(A) + \sum_{i=1}^{n(j)} \mu_n^{(i,j)}(A), \quad (73)$$

where the covariance and intensities of the individual components are found with

$$\text{cov}(\xi_p^{(j)}(A), \xi_p^{(j)}(B)) = \mu_p^{(j)}(A \cap B), \quad (74)$$

$$\text{cov}(\xi_b^{(i,j)}(A), \xi_b^{(i,j)}(B)) = \mu_b^{(i,j)}(A \cap B) - \mu_b^{(i,j)}(A) \mu_b^{(i,j)}(B), \quad (75)$$

$$\mu_p^{(j)}(A) = \int_A \mu_p^{(j)}(x) dx, \quad (76)$$

$$\mu_b^{(i,j)}(A) = \int_A \mu_b^{(i,j)}(x) dx. \quad (77)$$

Proof. The first equality for the covariance follows from the law of total covariance. The calculation of $\text{cov}^{(j)}(\xi(A), \xi(B))$ and $\mu^{(j)}(A), \mu^{(j)}(B)$ follow from the independence (due to superposition) of the terms for each j . The terms $\mu_p^{(j)}(A)$ and $\text{cov}(\xi_p^{(j)}(A), \xi_p^{(j)}(B))$ are the mean and covariance for a Poisson process. The terms $\mu_b^{(i,j)}(A)$ and $\text{cov}(\xi_b^{(i,j)}(A), \xi_b^{(i,j)}(B))$ are the mean and covariance of a Bernoulli process [13]. $\square$

In the following Lemma we shall consider the case in Lemma VI.2 where $\alpha^{(j)} = w^{(j)}$, since it relates to the particular model in the Gaussian prediction model in the following section. When this is not the case, we have to invert (67), which will involve a minimisation over $w^{(j)}$.

Lemma VI.4 (The Cramér Rao bound for a particular Poisson multi-Bernoulli mixture process). *Consider the case in VI.2 where $\alpha^{(j)} = w^{(j)}$, so that the mixture weights are equal. The Cramér Rao bound is given by*

$$\sum_{j=1}^N w^{(j)} \text{cov}(\xi^{(j)}(A), \xi^{(j)}(B)). \quad (78)$$

Proof. Since $\alpha^{(j)} = w^{(j)}$, we have

$$\mathcal{I}(\phi) = \sum_{i=1}^N w^{(i)} \Gamma^{(i)} \left(\frac{\phi^{(i)}}{w^{(i)}} \right), \quad (79)$$

Using the scaling property of Legendre transforms $(a\Lambda)^*(\phi) = a\Lambda^*\left(\frac{\phi}{a}\right)$, from Theorem A.1, we have

$$\mathcal{V}(f) = \min_{f=f_1+\dots+f_N} \left(\sum_{j=1}^N w^{(j)} \mathcal{W}^{(j)}(f_j) \right). \quad (80)$$

From the law of total differentiation, we have

$$v^{(2)}(x, y) = \frac{\delta^2}{\delta f(x) \delta f(y)} \mathcal{V}(f) \Big|_{f=0} = \quad (81)$$

$$\begin{aligned} &\sum_{j=1}^N w^{(j)} \frac{\delta^2}{\delta f_j(x) \delta f_j(y)} \mathcal{W}^{(j)}(f_j) \Big|_{f_j=0} \\ &= \sum_{j=1}^N w^{(j)} g^{(2)}(x, y). \end{aligned}$$

Hence the Cramér Rao bound is given by

$$\begin{aligned} &\sum_{j=1}^N w^{(j)} \int_{A \times B} g^{(2)}(x, y) dx dy \\ &= \sum_{j=1}^N w^{(j)} \text{cov}(\xi^{(j)}(A), \xi^{(j)}(B)). \end{aligned} \quad (82)$$

$\square$

In the previous Lemma, it is interesting to note that the Cramér Rao bound is lower than the covariance of the process calculated in Lemma VI.1. This is due to the fact that the hypothesis weights do not change in the prediction model.

B. Gaussian prediction model

Let us consider a scenario where the posterior probability generating functional is of the form of a Poisson multi-Bernoulli mixture as described in subsection A. Suppose that each target evolves according to the branching model

$$\mathcal{G}_{k|k-1}(h) = \mathcal{G}_{k-1}(\mathcal{G}_s(h|\cdot)), \quad (83)$$

where each target evolves according to Bernoulli survival model

$$\mathcal{G}_s(h|y) = 1 - p_s + p_s \int f_{k|k-1}(x|y), \quad (84)$$

where p_s is the probability of survival and $f_{k|k-1}(x|y) = \mathcal{N}(x; \mathbf{F}y, \mathbf{Q})$ is a Gaussian Markov transition.

Suppose that the intensities are Gaussian mixtures of the forms (c.f. [38]).

$$\mu_b^{(i,j)}(x) = p_b^{(i,j)} \mathcal{N}(x; m_{\mu,b}^{(i,j)}, \Sigma_{\mu,b}^{(i,j)}), \quad (85)$$

$$\mu_p^{(j)}(x) = \mu_p^{(j)}(X) \sum_{\ell=1}^{m(j)} \omega_{p,\mu}^{(j,\ell)} \mathcal{N}(x; m_{\mu,p}^{(j,\ell)}, \Sigma_{\mu,p}^{(j,\ell)}). \quad (86)$$

$$\phi_b^{(i,j)}(x) = q_b^{(i,j)} \mathcal{N}(x; m_{\phi,b}^{(i,j)}, \Sigma_{\phi,b}^{(i,j)}), \quad (87)$$

$$\phi_p^{(j)}(x) = \phi_p^{(j)}(X) \sum_{\ell=1}^{m(j)} \omega_{\mu,p}^{(j,\ell)} \mathcal{N}(x; m_{\phi,p}^{(j,\ell)}, \Sigma_{\phi,p}^{(j,\ell)}), \quad (88)$$

where $\omega_{p,\mu}^{(j,\ell)} \geq 0$, $\sum_{\ell=1}^{m(j)} \omega_{p,\mu}^{(j,\ell)} = 1$, and

$$q_b^{(i,j)} = p_s p_b^{(i,j)}, \quad (89)$$

$$\phi_p^{(j)}(X) = p_s \mu_p^{(j)}(X), \quad (90)$$

$$m_{\phi,b}^{(i,j)} = \mathbf{F} m_{\mu,b}^{(i,j)}, \quad (91)$$

$$\Sigma_{\phi,b}^{(i,j)} = \mathbf{F} \Sigma_{\mu,b}^{(i,j)} \mathbf{F}^T + \mathbf{Q}, \quad (92)$$

$$m_{\phi,p}^{(j,\ell)} = \mathbf{F} m_{\mu,p}^{(j,\ell)}, \quad (93)$$

$$\Sigma_{\phi,p}^{(j,\ell)} = \mathbf{F} \Sigma_{\mu,p}^{(j,\ell)} \mathbf{F}^T + \mathbf{Q}. \quad (94)$$

Since $\alpha^{(j)} = w^{(j)}$, we have

$$\mathcal{I}(\phi) = \sum_{i=1}^N w^{(j)} \Gamma^{(j)} \left(\frac{\phi^{(j)}}{w^{(j)}} \right), \quad (95)$$

Lemma VI.5 (Calculation of the Kullback-Leibler divergence and rate functionals for linear Gaussian prediction model). *Then the Bernoulli and Poisson rate functionals $\Gamma_b^{(j)}(\phi_b^{(j)})$ and $\Gamma_p^{(j)}(\phi_p^{(j)})$ are of the forms*

$$\begin{aligned} \Gamma_b^{(j)}(\phi_b^{(j)}) = & q_b^{(i,j)} \log \left(\frac{q_b^{(i,j)}}{p_b^{(i,j)}} \right) + \left(1 - q_b^{(i,j)} \right) \log \left(\frac{1 - q_b^{(i,j)}}{1 - p_b^{(i,j)}} \right) \\ & + \frac{q_b^{(i,j)}}{2} \left(m_{\phi,b}^{(i,j)} - m_{\mu,b}^{(i,j)} \right)^T (\Sigma_{\mu,p}^{(j,\ell)})^{-1} \left(m_{\phi,b}^{(i,j)} - m_{\mu,b}^{(i,j)} \right) \\ & + \frac{q_b^{(i,j)}}{2} \left(\text{tr} \left((\Sigma_{\mu,b}^{(j,\ell)})^{-1} \Sigma_{\phi,b}^{(j,\ell)} \right) - d + \log \frac{|\Sigma_{\phi,b}^{(j,\ell)}|}{|\Sigma_{\mu,b}^{(j,\ell)}|} \right), \end{aligned} \quad (96)$$

and

$$\begin{aligned} \Gamma_p^{(j)}(\phi_p^{(j)}) = & \phi_p^{(j)}(X) \log \left(\frac{\phi_p^{(j)}(X)}{\mu_p^{(j)}(X)} \right) - \left(\phi_p^{(j)}(X) - \mu_p^{(j)}(X) \right) + \\ & \phi_p^{(j)}(X) \left(\sum_{j=1}^N \alpha^{(j)} \log \left(\frac{\alpha_i}{w^{(j)}} \right) + \right. \\ & \left. \frac{1}{2} \sum_{j=1}^N \alpha^{(j)} \left(m_{\phi,b}^{(i,j)} - m_{\mu,b}^{(i,j)} \right)^T (\Sigma_{\mu,p}^{(j,\ell)})^{-1} \left(m_{\phi,b}^{(i,j)} - m_{\mu,b}^{(i,j)} \right) \right. \\ & \left. + \frac{1}{2} \sum_{j=1}^N \alpha^{(j)} \left(\text{tr} \left((\Sigma_{\mu,p}^{(j,\ell)})^{-1} \Sigma_{\mu,\phi}^{(j,\ell)} \right) - d + \log \frac{|\Sigma_{\mu,\phi}^{(j,\ell)}|}{|\Sigma_{\mu,p}^{(j,\ell)}|} \right) \right). \end{aligned} \quad (97)$$

Proof. Consider the first term in the rate functional for the Bernoulli process $\phi_b^{(j)}(x)$. Then we can separate the parts related to the constant and to the Gaussian due to the additivity of the logarithm, i.e.

$$\begin{aligned} \int \phi_p^{(j)}(x) \log \left(\frac{\phi_p^{(j)}(x)}{\mu_p^{(j)}(x)} \right) dx = & q_p^{(j)} \log \left(\frac{q_p^{(j)}}{p_b^{(j)}} \right) \\ & + q_p^{(j)} \int \mathcal{N}(x; m_{\phi,b}^{(i,j)}, \Sigma_{\phi,b}^{(i,j)}) \log \left(\frac{\mathcal{N}(x; m_{\phi,b}^{(i,j)}, \Sigma_{\phi,b}^{(i,j)})}{\mathcal{N}(x; m_{\mu,b}^{(i,j)}, \Sigma_{\mu,b}^{(i,j)})} \right) dx \end{aligned} \quad (98)$$

The result comes from the Kullback-Leibler divergence for Gaussians [47, p189]. The result for the Poisson case uses the chain rule for Kullback-Leibler divergence. $\square$

Corollary VI.1 (The variance and the Cramér Rao bound). *Following Lemmas VI.3 and VI.4, the variance for the terms in this model over the state space X become*

$$\text{var}(\xi_p^{(j)}(X)) = \text{cov}(\xi_p^{(j)}(X), \xi_p^{(j)}(X)) = \mu_p^{(j)}(X), \quad (99)$$

$$\text{var}(\xi_b^{(i,j)}(X)) = \text{cov}(\xi_b^{(i,j)}(X), \xi_b^{(i,j)}(X)) = q_b^{(i,j)}(1 - q_b^{(i,j)}). \quad (100)$$

The calculations above relate to the Kullback-Leibler divergence and the Cramér Rao a particular model for the multi-target prediction where it is possible to calculate the expressions analytically. This approach is also applicable to multi-target scenarios where there is control in the population of targets [48]. The possibility of target births was omitted, though the variance of the birth model could be added to the expression for Cramér Rao bound to deal with this scenario.

To consider the update model, we could either consider calculating the expressions, e.g. the Kullback-Leibler divergence, for a particular measurement set or average over all possible measurement sets. The former case brings complications since the structure of the posterior is not generally the same as the prior since there are additional terms introduced due to the different data association configurations. In the latter case, we can determine the mutual information which is the expectation of the Kullback-Leibler divergence between the prior and the posterior. See [21]. for a systematic approach for calculating the mutual information in multi-sensor scenarios.

APPENDIX A
LEGENDRE-FENCHEL TRANSFORMS

A. Fundamentals

The Legendre transform [49] is an involutive transformation on real-valued convex functions. It is used in physics to relate the Hamiltonian to the Lagrangian in thermodynamics. Fenchel proposed a generalisation of the Legendre transform to non-convex functions, known as the Legendre-Fenchel transform or convex conjugate [50], which is a fundamental result in convex analysis. The Legendre-Fenchel transform is used in large deviations theory [51]–[53] based on Crmer's rate function [23]–[25], [32], to calculate the probability of rare events.

Definition A.1 (Legendre-Fenchel transform). *The Legendre-Fenchel transform of function $\Lambda : X \rightarrow \mathbb{R} \cup \{+\infty\}$ is a function defined on the topological dual space X^* of X determined with*

$$\Lambda^*(\phi) = \sup_{x \in X} (\langle \phi, x \rangle - \Lambda(x)), \quad (101)$$

where $\phi \in X^*$.

Theorem A.1 (Legendre-Fenchel transform of a summation (Theorem 1 in [35])). *Let $\Lambda_1, \dots, \Lambda_k$ belong to the set of functions $\{\Lambda : X \rightarrow \mathbb{R} \cup \{+\infty\}\}$ which are convex, lower-continuous and not identically equal to $\{+\infty\}$ in X . Suppose that there is a point in the intersection of the domains of $\Lambda_1, \dots, \Lambda_k$, where all but one Λ_i are continuous. Then, for all $\phi \in X^*$,*

$$(\Lambda_1 + \dots + \Lambda_n)^*(\phi) = \min_{\phi = \phi_1 + \dots + \phi_n} (\Lambda_1^*(\phi_1) + \dots + \Lambda_n^*(\phi_n)). \quad (102)$$

Theorem A.2 (Log-sum-exp function (p482, [37])). *Consider the following function, known as the log-sum-exponential function. Given $\epsilon > 0$, let $\theta : \mathbb{R}^n \rightarrow \mathbb{R}$ be defined as follows.*

$$\theta(y_1, \dots, y_n) = \log(e^{y_1} + \dots + e^{y_n}). \quad (103)$$

The Legendre-Fenchel transform of this is calculated as [37, p482],

$$\theta^*(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i \log \alpha_i, \quad (104)$$

if $\alpha_1 + \dots + \alpha_n = 1$ and $\alpha_1 \geq 0, \dots, \alpha_n \geq 0$, or is equal to $+\infty$ if not.

This function is useful when considering composite functions, as shown in the note by Hiriart-Urruty [35]. In particular, we shall make use of the following theorem.

Theorem A.3 (Corollary 4 [35]). *Let $\Lambda_1, \dots, \Lambda_k$ belong to the set of functions $\{\Lambda : X \rightarrow \mathbb{R} \cup \{+\infty\}\}$ which are convex, lower-continuous and not identically equal to $\{+\infty\}$ in X . Suppose that there is a point in the intersection of the domains of $\Lambda_1, \dots, \Lambda_k$ where all but one Λ_i are continuous. Then, for*

all $\phi \in X^*$,

$$(\theta \circ (\Lambda_1, \dots, \Lambda_n))^*(\phi) = \min_{\substack{\alpha_1 + \dots + \alpha_n = 1, \\ \alpha_1 \geq 0, \dots, \alpha_n \geq 0, \\ \phi_1 + \dots + \phi_n = \phi.}} \sum_{i=1}^n \alpha_i \log \alpha_i + \sum_{i=1}^n \alpha_i \Lambda_i^*\left(\frac{\phi_i}{\alpha_i}\right). \quad (105)$$

B. The Legendre-Fenchel transform for the logarithm of mixtures of functions

Random variables are often defined in terms of weighted mixtures of moment generating functions. Hence, their cumulant generating function is defined by the logarithm of this weighted mixture. We can use the operations discussed above to determine the Legendre transform with the following theorem.

Theorem A.4 (Legendre transform of the logarithm of a mixture function). *Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be defined as follows.*

$$g(y_1, \dots, y_n) = \sum_{i=1}^n w_i e^{\Lambda_i(y_i)}, \quad (106)$$

where $w_1 + \dots + w_n = 1$ and $w_i \geq 0$ for $i = 1, \dots, n$. Then the Legendre-Fenchel transform of

$$\log g(y_1, \dots, y_n) = \log \left(\sum_{i=1}^n w_i e^{\Lambda_i(y_i)} \right) \quad (107)$$

is calculated with

$$\min_{\substack{\alpha_1 + \dots + \alpha_n = 1, \\ \alpha_1 \geq 0, \dots, \alpha_n \geq 0, \\ \phi_1 + \dots + \phi_n = \phi.}} \sum_{i=1}^n \alpha_i \log \left(\frac{\alpha_i}{w_i} \right) + \sum_{i=1}^n \alpha_i \Lambda_i^*\left(\frac{\phi_i}{\alpha_i}\right). \quad (108)$$

Proof. The proof follows by noting that

$$\log g(y_1, \dots, y_n) = (\theta \circ (\Lambda_1 + \log w_1, \dots, \Lambda_n + \log w_n))(y_1, \dots, y_n). \quad (109)$$

Hence, using the Legendre transform of the log-sum-exp function in Theorem A.2 (from [37, p482]) and Theorem A.3 (Corollary 4 from [35]), and the Legendre-Fenchel transform of the sum of a function and a constant, the result follows. Specifically, suppose that c is a constant. Then the Legendre-Fenchel transform of the sum of a function f with the constant c is given by

$$(\Lambda + c)^*(\phi) = \Lambda^*(\phi) - c. \quad (110)$$

Then we have

$$(\log g)^*(\phi) = (\theta \circ (\Lambda_1 + \log w_1, \dots, \Lambda_n + \log w_n))^*(\phi). \quad (111)$$

□

Note that the constituent mixture components Λ_i are typically described on the same space. This can be determined by setting all of the arguments to be the same, i.e. $y_1 = \dots = y_n$.

Though this is a direct application of Corollary 4 from [35], the author believes that this is the first statement of this particular result for mixture distributions.

APPENDIX B

CRÁMER'S RATE FUNCTION FOR RANDOM VARIABLES

In this section, we review Cramér's rate function for random variables and how to determine the Legendre transform with some simple parametrisations.

A. The cumulant generating functions and the Legendre-Fenchel transform

Definition B.1 (Moment generating function). *The moment generating function of random variable ξ is given by*

$$M(z) = \mathbb{E} [e^{z\xi}]. \quad (112)$$

When ξ is a non-negative integer-valued random variable, the moment generating function becomes

$$M(z) = \sum_{k=0}^{\infty} p_k e^{zk}, \quad (113)$$

and when ξ is a continuous-valued random variable, the moment generating function becomes

$$M(z) = \int e^{z\xi} p(\xi) d\xi. \quad (114)$$

Definition B.2 (Cumulant generating function). *The cumulant generating function is found by taking the logarithm of $M(z)$, i.e.*

$$\Lambda(z) = \log M(z). \quad (115)$$

Definition B.3 (Cramér's rate function for random variables). *The rate function in large deviations theory [51] is found by taking the Legendre-Fenchel transform of the cumulant generating function, i.e.*

$$\Lambda^*(\phi) = \sup_z (z\phi - \Lambda(z)). \quad (116)$$

B. Cramér's rate function for sums of random variables

The following Lemma was recently presented by Zająkowski [42] for determining Cramér's rate function for sums of random variables. This uses Theorem A.1, the result of Hiriart-Urruty for calculating the Legendre-Fenchel transform of a summation [35]. We first present the Lemma, and then illustrate it for a Poisson-binomial distribution [45], which is formed with the sum of independent Bernoulli distributions.

Lemma B.1 (Cramér's rate function for sums of random variables). *Suppose that the distribution is described with a superposition of independent distributions described with moment generating functions $M_1(z), \dots, M_n(z)$, i.e.*

$$M(z) = \prod_{i=1}^n M_i(z). \quad (117)$$

Then the Legendre-Fenchel transform of the cumulant generating function $\Lambda(z) = \log M(z)$ is given by

$$(\Lambda_1 + \dots + \Lambda_n)^*(\phi) = \min_{\phi = \phi_1 + \dots + \phi_n} (\Lambda_1^*(\phi_1) + \dots + \Lambda_n^*(\phi_n)), \quad (118)$$

where $\Lambda_i^*(\phi_i)$ is the Legendre-Fenchel transform of the i^{th} cumulant generating function.

C. Cramér's rate function for mixture distributions

The following Lemma determines Cramér's rate function for arbitrary mixtures of random variables. This provides a lower bound on the Kullback-Leibler divergence from a mixture distribution to a distribution of a given mean.

Lemma B.2 (Legendre-Fenchel transform of a mixture of distributions). *Suppose that the distribution is described with a mixture of moment generating functions $M_1(z), \dots, M_n(z)$, i.e.*

$$M(z) = \sum_{i=1}^n w_i M_i(z), \quad (119)$$

where $w_1 + \dots + w_n = 1$. *Then the Legendre-Fenchel transform of the cumulant generating function is given by*

$$\begin{aligned} \Lambda^*(\phi) &= \left(\log \sum_{i=1}^n w_i M_i \right)^* (\phi) \\ &= \min_{\substack{\alpha_1 + \dots + \alpha_n = 1, \\ \alpha_1 \geq 0, \dots, \alpha_n \geq 0, \\ \phi_1 + \dots + \phi_n = \phi}} \left(\sum_{i=1}^n \alpha_i \log \left(\frac{\alpha_i}{w_i} \right) + \sum_{i=1}^n \alpha_i \Lambda_i^* \left(\frac{\phi_i}{\alpha_i} \right) \right), \end{aligned} \quad (120)$$

where Λ_i^* is the Legendre-Fenchel transform of cumulant generating function $\Lambda_i(z) = \log M_i(z)$.

Proof. Note that $\Lambda(z) = \log \left(\sum_{i=1}^n w_i e^{\Lambda_i(z)} \right)$ and that $w_i M_i(z) = \exp(\Lambda_i(z) + \log w_i)$. The proof follows from Theorem A.4. $\square$

Example B.1 (Cramér's rate function for a mixture of Gaussians). *Consider a mixture of Gaussians with weights w_i , means m_i and covariances Σ_i for $i = 1, \dots, n$ with moment generating function*

$$M(z) = \sum_{i=1}^n w_i \exp \left(m_i^T z + \frac{1}{2} z^T \Sigma_i z \right). \quad (121)$$

Cramér's rate function is given by

$$\begin{aligned} \Gamma(\phi) &= \min_{\substack{\alpha_1 + \dots + \alpha_n = 1, \\ \alpha_1 \geq 0, \dots, \alpha_n \geq 0, \\ \phi_1 + \dots + \phi_n = \phi}} \left(\sum_{i=1}^n \alpha_i \log \left(\frac{\alpha_i}{w_i} \right) + \right. \\ &\quad \left. \frac{1}{2} \sum_{i=1}^n \alpha_i \left((\alpha_i^{-1} \phi_i - m_i)^T \Sigma_i^{-1} (\alpha_i^{-1} \phi_i - m_i) \right) \right). \end{aligned} \quad (122)$$

Proof. The moment generating function for a multivariate Gaussian with mean m and covariance Σ is given by

$$M(z) = \exp \left(m^T z + \frac{1}{2} z^T \Sigma z \right). \quad (123)$$

Cramér's rate function is given by [51, p13]

$$\Lambda^*(\phi) = \frac{1}{2} (\phi - m)^T \Sigma^{-1} (\phi - m). \quad (124)$$

The result then follows from Lemma B.2. $\square$

APPENDIX C

TABLES OF CRAMÉR'S RATE FUNCTIONS FOR DISTRIBUTIONS, POINT PROCESSES, AND RANDOM MEASURES

Distribution	m.g.f., $M(z)$	Cramér's rate function, $\Lambda^*(\phi) = \sup_z (z\phi - \Lambda(z))$
Poisson	$\exp(\lambda(e^z - 1))$	$\phi \log(\phi/\lambda) + \lambda - \phi$
Bernoulli	$1 - p + pe^z$	$\phi \log\left(\frac{\phi}{p}\right) + (1 - \phi) \log\left(\frac{1-\phi}{1-p}\right)$
Binomial	$(1 - p + pe^z)^n$	$\phi \log\left(\frac{\phi}{np}\right) + (n - \phi) \log\left(\frac{1-\phi/n}{1-p}\right)$
Panjer	$(1 - (1 - e^z)/\beta)^{-\alpha}$	$\phi \log\left(\frac{\phi(\beta+1)}{\alpha+\phi}\right) + \alpha \log\left(\frac{\alpha(\beta+1)}{\beta(\alpha+\phi)}\right)$
Poisson-binomial	$\prod_{i=1}^n (1 - p_i + p_i e^z)$	$\min_{\phi_1+\dots+\phi_n} \left(\sum_{i=1}^n \phi_i \log\left(\frac{\phi_i}{p_i}\right) + (1 - \phi_i) \log\left(\frac{1-\phi_i}{1-p_i}\right) \right)$
Gamma	$(1 - z/\beta)^{-\alpha}$	$\phi\beta - \alpha + \alpha \log(\frac{\alpha}{\phi\beta})$
Gaussian	$\exp(m^T z + \frac{1}{2} z^T \Sigma z)$	$\frac{1}{2} (m - \phi)^T \Sigma^{-1} (m - \phi)$
Gaussian mixture	$\sum_{i=1}^n w_i e^{m_i^T z + \frac{1}{2} z^T \Sigma_i z}$	$\min_{\substack{\alpha_1+\dots+\alpha_n=1, \\ \alpha_1 \geq 0, \dots, \alpha_n \geq 0, \\ \phi_1+\dots+\phi_n=\phi}} \left(\sum_{i=1}^n \alpha_i \log\left(\frac{\alpha_i}{w_i}\right) + \sum_{i=1}^n \frac{\alpha_i}{2} (m_i - \frac{\phi_i}{\alpha_i})^T \Sigma_i^{-1} (m_i - \frac{\phi_i}{\alpha_i}) \right)$

Table I

TABLE OF CRAMÉR'S RATE FUNCTION FOR PROBABILITY DISTRIBUTIONS. SEE GRIMMETT AND STIRZAKER [54, p186-188] FOR DESCRIPTIONS OF THE CHARACTERISTIC FUNCTIONS $\Phi(z) = M(iz)$ FOR THE BERNOULLI, BINOMIAL, MULTIVARIATE GAUSSIAN, AND GAMMA DISTRIBUTIONS. THE MOMENT GENERATING FUNCTION FOR THE PANJER DISTRIBUTION (GENERALISED NEGATIVE-BINOMIAL DISTRIBUTION) IS GIVEN IN FACKLER [55]. CRAMÉR'S RATE FUNCTION FOR A MULTIVARIATE GAUSSIAN IS GIVEN IN VARADHAN [51, p13]. THE POISSON-BINOMIAL DISTRIBUTION IS DESCRIBED IN LE CAM [45].

Laplace functional, $\mathcal{L}(f)$	Cramér's rate functional, $\Gamma(\phi) = \sup_f (\langle -f, \phi \rangle - \mathcal{W}(f))$
$\exp\left(\int \lambda(x)(e^{-f(x)} - 1)dx\right)$	$\int \phi(x) \log \frac{\phi(x)}{\lambda(x)} dx - \int (\phi(x) - \lambda(x)) dx$
$\int 1 - \psi(x) + e^{-f(x)} \psi(x) dx$	$\int \phi(x) \log\left(\frac{\phi(x)}{\psi(x)}\right) dx + (1 - \int \phi(x) dx) \log\left(\frac{1 - \int \phi(x) dx}{1 - \int \psi(x) dx}\right)$
$(\int 1 - \psi(x) + \int e^{-f(x)} \psi(x) dx)^n$	$\int \phi(x) \log\left(\frac{\phi(x)}{n\psi(x)}\right) dx + (n - \int \phi(x) dx) \log\left(\frac{1 - \frac{1}{n} \int \phi(x) dx}{1 - \int \psi(x) dx}\right)$
$(1 - (1 - \int e^{-f(x)} s(x) dx)/\beta)^{-\alpha}$	$\int \phi(x) \log\left(\frac{\phi(x)(\beta+1)}{\alpha s(x) + \phi(x)}\right) + \alpha \log\left(\frac{\alpha s(x)(\beta+1)}{\beta(\alpha s(x) + \phi(x))}\right) dx$
$\prod_{i=1}^n (\int 1 - \psi_i(x) + \psi_i(x) e^{-f(x)} dx)$	$\min_{\phi_1+\dots+\phi_n} \left(\sum_{i=1}^n \int \phi_i(x) \log\left(\frac{\phi_i(x)}{\psi_i(x)}\right) dx + (1 - \int \phi_i(x) dx) \log\left(\frac{1 - \int \phi_i(x) dx}{1 - \int \psi_i(x) dx_i}\right) \right)$
$\exp\left(-\int \log(1 + \beta^{-1} f(x)) \alpha(x) dx\right)$	$\int (\beta \phi(x) - \alpha(x)) dx + \int \log\left(\frac{\alpha(x)}{\beta \phi(x)}\right) \alpha(x) dx$

Table II

TABLE OF CRAMÉR'S RATE FUNCTION FOR POINT PROCESSES AND RANDOM MEASURES. (GENERALISATIONS OF THE MODELS IN ORDER AS ABOVE): (I) POISSON POINT PROCESS [17]; (II) BERNOULLI POINT PROCESS [56]; (III) BINOMIAL POINT PROCESS; (IV) PANJER POINT PROCESS [44]; (V) POISSON-BINOMIAL POINT PROCESS [57]; (VI) GAMMA RANDOM MEASURE [40, p167].

REFERENCES

- [1] F. E. Daum, "Cramér—rao type bounds for random set problems," in *Random Sets*. Springer, 1997, pp. 165–183.
- [2] F. E. Daum, "Bounds on performance for multiple target tracking," *IEEE Transactions on Automatic Control*, vol. 35, no. 4, pp. 443–446, 1990.
- [3] Streit, Roy L., "Multiple target tracking," in *Poisson Point Processes*. Springer, 2010, pp. 147–178.
- [4] C. Hue, J. . Le Cadre, and P. Perez, "Posterior cramer-rao bounds for multi-target tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 37–49, 2006.
- [5] B. Ristic, A. Farina, and M. Hernandez, "Cramér-rao lower bound for tracking multiple targets," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 151, no. 3, pp. 129–134, 2004.
- [6] W. F. Leven, "Approximate cramer-rao bounds for multiple target tracking," Ph.D. dissertation, Georgia Institute of Technology, 2006.
- [7] B. Ristic, A. Farina, and M. Hernandez, "Cramér-rao lower bound for tracking multiple targets," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 151, no. 3, pp. 129–134, 2004.
- [8] Houssineau, Jeremie and Singh, Sumeetpal S and Jasra, Ajay, "Identification of MultiObject Dynamical Systems: Consistency and Fisher Information," *SIAM Journal on Control and Optimization*, vol. 57, no. 4, pp. 2603–2627, 2019.
- [9] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Phil. Trans. Royal Soc. A*, vol. 222, no. 594–604, pp. 309–368, 1922.
- [10] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, p. 81–89, 1945.
- [11] H. Cramer, "Mathematical methods of statistics princeton university press princeton," 1946.
- [12] E. D. Delande, M. Üney, J. Houssineau, and D. E. Clark, "Regional Variance for Multi-Object Filtering," *Signal Processing, IEEE Transactions on*, vol. 62, no. 13, pp. 3415–3428, Jul. 2014.
- [13] Delande, ED and Houssineau, J and Clark, DE, "Regional variance in target number: analysis and application for multi-Bernoulli point processes," *IET Seminar on Target Tracking*, 2014.
- [14] M. A. Campbell, D. E. Clark, and F. de Melo, "An algorithm for large-scale multitarget tracking and parameter estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 4, pp. 2053–2066, 2021.
- [15] Kullback, Solomon and Leibler, Richard A, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [16] Bartlett, MS, "Processus stochastiques ponctuels," *Ann. Inst. Henri Poincaré*, vol. 14, pp. 35–60, 1954.
- [17] Moyal, JE, "The general theory of stochastic population processes," *Acta mathematica*, vol. 108, no. 1, pp. 1–31, 1962.
- [18] McFadden, JA, "The entropy of a point process," *Journal of the Society for Industrial and Applied Mathematics*, vol. 13, no. 4, pp. 988–994, 1965.
- [19] Clark, Daniel E, "Local entropy statistics for point processes," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 1155–1163, 2020.
- [20] J. L. Williams, "An efficient, variational approximation of the best fitting multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 258–273, 2014.
- [21] D. E. Clark, "Multi-sensor network information for linear-gaussian multi-target tracking systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4312–4325, 2021.
- [22] B. Ristic, B.-N. Vo, and D. Clark, "A note on the reward function for phd filters with sensor control," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 1521–1529, 2011.
- [23] Cramér, Harald, "On a new limit theorem of the theory of probability," *Uspekhi Matematicheskikh Nauk*, no. 10, pp. 166–178, 1944.
- [24] Gärtner, Jürgen, "On large deviations from the invariant measure," *Theory of Probability & Its Applications*, vol. 22, no. 1, pp. 24–39, 1977.
- [25] Ellis, Richard S and others, "Large deviations for a general class of random vectors," *The Annals of Probability*, vol. 12, no. 1, pp. 1–12, 1984.
- [26] Donsker, MD and Varadhan, SRS, "Asymptotic evaluation of certain Markov process expectations for large time—III," *Communications on pure and applied Mathematics*, vol. 29, no. 4, pp. 389–461, 1976.
- [27] I. Sanov, "On the probability of large deviations of random quantities," *Mat. Sb*, vol. 42, pp. 14–44, 1957.
- [27] S. Kullback *et al.*, "Certain inequalities in information theory and the cramer-rao inequality," *The Annals of Mathematical Statistics*, vol. 25, no. 4, pp. 745–751, 1954.
- [29] A. Fuchs and G. Letta, "L'inégalité de Kullback. Application à la théorie de l'estimation," *Séminaire de probabilités de Strasbourg*, vol. 4, pp. 108–131, 1970.
- [30] D. E. Clark, "A Cramér Rao bound for point processes," *IEEE Transactions on Information Theory*, 2022.
- [31] Itzykson, Claude and Zuber, Jean-Bernard, *Quantum field theory*. Courier Corporation, 2012.
- [32] Cramér, Harald, "Sur un nouveau théorème-limite de la théorie des probabilités," *Actual. Sci. Ind.*, vol. 736, pp. 5–23, 1938.
- [33] Zinn-Justin, Jean, *Quantum field theory and critical phenomena*. Clarendon Press, 1996.
- [34] Jona-Lasinio, G, "Relativistic field theories with symmetry-breaking solutions," *Il Nuovo Cimento (1955-1965)*, vol. 34, no. 6, pp. 1790–1795, 1964.
- [35] Hiriart-Urruty, Jean-Baptiste, "A note on the Legendre-Fenchel transform of convex composite functions," in *Nonsmooth Mechanics and Analysis*. Springer, 2006, pp. 35–46.
- [36] Hiriart-Urruty, Jean-Baptiste and Lemaréchal, Claude, *Fundamentals of convex analysis*. Springer Science & Business Media, 2012.
- [37] Rockafellar, R Tyrrell and Wets, Roger J-B, *Variational analysis*. Springer Science & Business Media, 2009, vol. 317.
- [38] J. L. Williams, "Marginal multi-bernoulli filters: Rfs derivation of mht, jipda, and association-based member," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 1664–1687, 2015.
- [39] Daley, Daryl J and Vere-Jones, David, *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.
- [40] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes. vol. I. , Elementary theory and methods*, ser. Probability and its applications. New York, Berlin, Paris: Springer, 2003.
- [41] Kullback, S, "Information theory and statistics," Wiley, New York, 1959.
- [42] K. Zajkowski, "A variational formula on the Cramér function of series of independent random variables," *Positivity*, vol. 21, no. 1, pp. 273–282, 2017.
- [43] B. Ristic and S. Arulampalam, "Bernoulli particle filter with observer control for bearings-only tracking in clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2405–2415, 2012.
- [44] I. Schlangen and E. D. Delande and J. Houssineau and D. E. Clark, "A Second-Order PHD Filter With Mean and Variance in Target Number," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 48–63, Jan 2018.
- [45] Le Cam, Lucien, "An approximation theorem for the Poisson binomial distribution," *Pacific Journal of Mathematics*, vol. 10, no. 4, pp. 1181–1197, 1960.
- [46] Cover, Thomas M and Thomas, Joy A, *Elements of information theory (2nd ed.)*. John Wiley & Sons, 2012.
- [47] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.
- [48] D. E. Clark, "Stochastic multi-object guidance laws for interception and rendezvous problems," *IEEE Transactions on Automatic Control*, 2021.
- [49] Legendre, Adrien Marie, *Mémoire sur l'intégration de quelques équations aux différences partielles*. Imprimerie royale, 1787.
- [50] Fenchel, Werner, "On conjugate convex functions," *Canadian Journal of Mathematics*, vol. 1, no. 1, pp. 73–77, 1949.
- [51] Varadhan, SR Srinivasa, *Large deviations and applications*. Siam, 1984, vol. 46.
- [52] Ellis, Richard S, *Entropy, large deviations, and statistical mechanics*. Springer, 2007.
- [53] Touchette, Hugo, "The large deviation approach to statistical mechanics," *Physics Reports*, vol. 478, no. 1-3, pp. 1–69, 2009.
- [54] G. Grimmett, G. R. Grimmett, D. Stirzaker *et al.*, *Probability and random processes*. Oxford university press, 2001.
- [55] M. Fackler, "Panjer class united - one formula for the Poisson, Binomial, and Negative Binomial distribution," *ASTIN colloquium*, 2009.
- [56] J. A. Smith, "Statistical modeling of daily rainfall occurrences," *Water Resources Research*, vol. 23, no. 5, pp. 885–893, 1987.
- [57] P. Deheuvels and D. Pfeifer, "Poisson approximations of multinomial distributions and point processes," *Journal of multivariate analysis*, vol. 25, no. 1, pp. 65–89, 1988.