Error Bounds for Joint Detection and Estimation of a Single Object with Random Finite Set Observation

Mohammad Rezaeian and Ba-Ngu Vo

Abstract

This paper considers the performance limits for joint detection and estimation from a finite set-valued observation that is stochastically related to the state or parameter of interest. Detection refers to inference about the existence of the state, whereas estimation refers to inference about its value, when detected. Since we need to determine the existence/non-existence of the state as well as its value, the usual notion of Euclidean distance error does not jointly capture detection and estimation error in a meaningful manner. Treating the state as set, which can be either empty or singleton, admits a meaningful distance error for joint detection and estimation. We derive bounds on this distance error for a widely used class of observation models. When existence of the state is a certainty, our bounds coincide with recent results on Cramér-Rao bounds for estimation only problems.

I. INTRODUCTION

In a practical state or parameter estimation problem the measuring device or sensor does not always receive the signal of interest. In addition the sensor may receive a spurious set of false measurements. The number of measurements (at each sampling instant) is random and there is no information on which is the measurement of interest or which are the false measurements [1], [2]. In the Bayesian estimation framework, the collection of measurement is treated as a realization of a random finite set or point process [2].

In this paper, we derive error bounds for the joint detection and estimation of a state observed as a realization of a random finite set. Our result is a generalization of the Cramér-Rao lower bound to real-world estimation problems where the measurement is not a vector but a finite set

of vectors that may not contain the signal of interest, and the state of system is not a vector, but a finite set that is either empty or singleton. This problem finds a host of applications in defence and surveillance [1], [2], where it is not known whether the target exists or not and the aim is to determine, from the measurement, the existence of the target and its state. In our model an observer tries to detect and estimate the position of a target using prior information about the target as well as an observation that consist of a random number of points in the signal space that are statistically dependent on the existence of the target and its kinematic state. Finding the limit of the accuracy (or error) that one can achieve for such a system is the purpose of this paper.

Since we need to determine the existence/non-existence of the state as well as its value, the usual notion of Euclidean distance error does not apply. Nonetheless, a mathematically consistent and physically meaningful distance, namely the Optimal Subpattern Assignment (OSPA) distance, is available [3] (in fact this distance was proposed for general finite-set-valued state). Our error bound for the joint detection and estimation problem is a lower bound on the expected OSPA error between the true state and the estimated state. The derivation hinges on the application of the information inequality which in turn requires a valid joint probability density for the state and the finite-set-valued measurement.

As opposed to the generic detection problem in the presence of noise that is posed as selection of two hypotheses [4] through the observation z_k,

$$
\begin{aligned}
H_0: & \quad z_k = n_k, \\
H_1: & \quad z_k = s_k + n_k,
\end{aligned}
\tag{1}
$$

where s_k is a point in the signal space and n_k is additive noise, our joint detection-estimation problem involves selecting the following hypotheses based on a set valued observation $Z = \{z_1, z_2, ..., z_n\}$.

$$
\begin{aligned}
H_0: & \quad Z = W \\
H_1: & \quad Z = \Theta(x) \cup W,
\end{aligned}
\tag{2}
$$

where Θ represents a random set statistically dependent on the state x, and W is an independent random set of spurious measurements called clutter. Moreover, if H_1 is selected, then the value of x needs to be estimated. Note that in this paper we use the notation that x is a random variable, $\mathbf{x} = (x_1, ..., x_n)$ is a random vector, and X a random finite set. In contrast to the generic detection problem where noise $n(t)$ is added to the signal $s(t)$ and $s(t) + n(t)$ is observed as

one point in the measurement space, here the clutter W is set-unioned with the signal $\Theta(x)$ in the measurement space. Moreover, it is not known which points belong to which category (signal or clutter).

Cramér-Rao bounds on the mean square error (MSE) is widely used in state estimation (but not detection) problems for systems in which the measurement vector $\mathbf{z}_k$ at time k statistically relates to the state vector $\mathbf{x}_k$ at that time through a conditional density function, $\mathbf{z}_k \sim f(\cdot|\mathbf{x}_k)$. The bounds are given by the inverse of the Fisher information matrix [5], [6]. For dynamic state estimation or filtering problems, the state estimate at time k makes use of the entire measurement history $\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_k$. A recursive formulation of the Fisher information matrix and error bounds over different time indices has been recently obtained in [7]. More recently this iterative bound has been further tighten in [8], [9].

In this paper we extend the Cramér-Rao bound to problem (2) which involves joint detection and estimation from a random number of measurements. Modeling the state as a random set X, the statistical relation of the observation and state in problem (2) can be written as

$$Z = \Theta(X) \cup W. \tag{3}$$

This measurement model is widely used in target tracking [2], [10]. In a dynamic estimation or filtering problem, the measurement, at any time instance k will be $Z_k = \Theta(X_k) \cup W_k$, and the state X_k is a Markov process. However, in the current set up, we restrict our attention to the static estimation problem (3), estimating the random set X. We assume that X and $\Theta(X)$ have at most cardinality one.

Our main results are presented in Section III. In two Theorems we derive bounds on the MSE of joint detection-estimation for two generic measurement models. We then discuss special cases of the bounds, namely, measurement without clutter, and estimation only problems. Sections IV is dedicated to proofs and detailed discussions of the bounds. While for simplicity the state space and the measurement space in Sections III and IV are considered to be one dimensional, we have extended the formulation to vector spaces in Section V.

II. BACKGROUND AND RELATED WORKS

A. Random finite set

Random finite set, or point process theory is a mathematical discipline that has long been used by statisticians in many diverse applications including agriculture, geology, and epidemiology

[11]. Finite set statistics (FISST) is a set of practical mathematical tools from point process theory, developed by Mahler for the study of multi-object systems [2]. Innovative multi-object filtering solutions derived from FISST such as the Probability Hypothesis Density (PHD) filters [12]- [15] have attracted substantial interest from academia and the commercial sector.

A random finite set (RFS) is simply a random variable that takes value as (unordered) finite set, i.e. a *finite-set-valued random variable*. It can be described by a probability density function over the space of finite sets. In this paper we only deal with two types of RFS, namely, Bernoulli and Poisson RFS. A Bernoulli RFS on a space $\mathcal{Z}$ is defined by two parameters r and $p(\cdot)$, where $0 \le r \le 1$ and $p(\cdot)$ is a density on $\mathcal{Z}$. The density f of this RFS on the space of finite sets is defined by [1]

$$
f(Z) = \begin{cases} 1 - r & Z = \emptyset \\ rp(z) & Z = \{z\} \\ 0 & \text{otherwise,} \end{cases} \tag{4}
$$

We refer to a Bernoulli RFS Z with parameters $r, p(\cdot)$ by $\mathcal{B}_Z(r, p(\cdot))$. A Poisson RFS on a space $\mathcal{Z}$ is uniquely characterized by its intensity function $v(\cdot)$ since the density of this RFS is given by

$$
f(\{z_1, z_2, ..., z_n\}) = e^{-\lambda} \prod_{i=1}^{n} v(z_i), \tag{5}
$$

where $\lambda = \int_{\mathcal{Z}} v(z) dz$. For a Poisson RFS, the cardinality distribution is a Poisson with rate λ, and given the number of points, the points themselves are independently distributed.

The set integral of a function g taking the finite subsets of $\mathcal{Z}$ to the reals is [2]

$$
\int g(Z) \delta Z \triangleq g(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{Z}^n} g_n(\mathbf{z}) d\mathbf{z} \tag{6}
$$

where $g_n((z_1, z_2, ..., z_n)) = g(\{z_1, z_2, ..., z_n\})$ for $z_1, z_2, ..., z_n$ being distinct, otherwise $g_n(\mathbf{z}) = 0$. The function g must have unit of $u^{-|Z|}$, where u is the unit of measurement in $\mathcal{Z}$. The mean of a function h of a RFS with density f is

$$
E[h] = \int h(Z) f(Z) \delta Z. \tag{7}
$$

---

[1]Strictly speaking, this density is the set derivative of the belief mass function. However, it has been shown in [16] that the set derivative of the belief function is equivalent to a probability density.

B. Information Inequality

Given a joint probability density $f(\cdot, \cdot)$ on $\mathcal{R}^2$, under regularity conditions and the existence of $\frac{\partial^2 \log f(x,y)}{\partial x^2}$, the information inequality states that [17], [6]

$$\int_{\mathcal{R}^2} f(x,y)(x - \hat{x}(y))^2 dxdy \geq -\left[\int_{\mathcal{R}^2} f(x,y)\frac{\partial^2 \log f(x,y)}{\partial x^2}dxdy\right]^{-1}, \qquad (8)$$

where $\hat{x}(y)$ is an unbiased estimate of x based on observation y. $\hat{x}(\cdot)$ is unbiased if $E_{|\dot{x}}\hat{x} = \dot{x}$, where $E_{|\dot{x}}$ is the expectation conditional on a specific realization $x = \dot{x}$. This inequality is written as

$$\sigma_e^2 \geq J^{-1},$$

where σ_e^2 is the means square of the unbiased estimation error in the left hand side of (8) and

$$J = -E_f\left[\frac{\partial^2 \log f(x,y)}{\partial x^2}\right]$$

is the *Fisher information* (the notation E_f means expectation with respect to density f).

For random vectors $\mathbf{x}$ and $\mathbf{y}$, the information inequality is stated in terms of the Fisher information matrix J,

$$[J]_{i,j} = -E_f\left(\frac{\partial^2 \log f(\mathbf{x}, \mathbf{y})}{\partial x_i \partial x_j}\right).$$

The information inequality states that the mean square error for estimation of i-th component of vector $\mathbf{x}$ upon observation of vector $\mathbf{y}$ is bounded by, [6],

$$\sigma_{ei}^2 = \int\int f(\mathbf{x}, \mathbf{y})(x_i - \hat{x}_i(\mathbf{y}))^2 d\mathbf{x}d\mathbf{y} \geq [J^{-1}]_{i,i}.$$

Given a set to point transformation T, a corresponding Cramér-Rao bound on the covariance of $T(X)$, where X is an RFS has been established in [18, Corollary 20].

In this paper we bound the error of the set-valued estimate in terms of the OSPA distance. Moreover, we apply the original information inequality directly on every summation term of set integrals, which we believe gives a tighter, and computationally tractable bounds compared to applying the information inequality to the probability space of random sets. The following illustrates the difference between the two approaches in a simple case, that will be further extended in Section III-E.

By extending the domain of random variable y to the space of finite sets on $\mathcal{R}$, the inequality (8) extends to

$$\int\int_{\mathcal{R}} f(x,Y)(x - \hat{x}(Y))^2 dx\delta Y \geq -\left[\int\int_{\mathcal{R}} f(x,Y)\frac{\partial^2 \log f(x,Y)}{\partial x^2}dx\delta Y\right]^{-1}. \qquad (9)$$

While, by applying the information inequality to each summation term of the set integral expansion of LHS of (9),

$$\sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathcal{Y}^n} \int_{\mathcal{R}} f_n(x, \mathbf{y})(x - \hat{x}(\mathbf{y}))^2 dx d\mathbf{y} \geq -\sum_{n=0}^{\infty} \frac{1}{n!} \left[\int_{\mathcal{Y}^n} \int_{\mathcal{R}} f_n(x, \mathbf{y}) \frac{\partial^2 \log f(x, \mathbf{y})}{\partial x^2} dx d\mathbf{y} \right]^{-1}.$$
(10)

Using convexity arguments and Jenson's inequality, we can show that (10) is a tighter inequality than (9). Moreover, the series in (10) converges from below to its limit, while the one in (9) is in fact a series converging from above. Therefore each element of the sequence corresponding to (10) is a lower bound, which makes this expression computationally tractable.

III. Joint Estimation and Detection

We consider the problem of estimating the state (existence and kinematics) of a single target through an observation Z and prior knowledge about the target. The target is modeled as a Bernoulli random finite set $X \sim \mathcal{B}_X(r, f(\cdot))$ with $f(\cdot)$ defined on a subset of real numbers, $\mathcal{X} \subseteq \mathcal{R}$. The Bernoulli density $\mathcal{B}_X(r, f(\cdot))$ encapsulates all the prior information about the target, $r \in [0, 1]$ representing the probability of existence of target, and the density function $f(\cdot)$ representing the prior knowledge about position of target, if it exists. We refer to the special case of $r = 1$ as the asserted target system, where the problem is an estimation only problem. In this case the RFS X reduces to a random variable x on $\mathcal{X}$ with density $f(\cdot)$.

Since the target is modeled by a random set X, the estimate that we denote by $\hat{X}$ is a set. The estimation of $|X|$, the cardinality of X, corresponds to detection, which here can be either zero or one. When $|\hat{X}| = 1$, i.e. $\hat{X} = \{\hat{x}\}$, the estimate $\hat{X}$ also provides $\hat{x}$ as the state estimate. Therefore estimation of the finite set X is referred to as joint detection-estimation. The joint detection-estimation is based on measurement set Z, as a finite subset of a region $\mathcal{Z} \subset \mathcal{R}$, that is statistically dependent on the state X. The estimate $\hat{X}$ is a deterministic function of Z, and is denoted as $\hat{X}(Z)$, or to indicate its not-emptiness as $\hat{x}(Z)$.

We aim to find the performance limits of estimators measured by the average error between X and $\hat{X}(Z)$. Since X and $\hat{X}(Z)$ are random sets, we need a generalized concept of error between two sets. The conventional concept of error as the difference of two real values is not applicable. The error between two sets X and $\hat{X}$ should be defined by a metric $e(X, \hat{X})$, and

the mean square error as

$$\sigma^2 = E[e^2(X, \hat{X}(Z)] = \int \int f(X, Z)e^2(X, \hat{X}(Z))\delta X \delta Z, \qquad (11)$$

where $f(\cdot, \cdot)$ is the joint density of two random finite sets X and Z. The triangular inequality of the metric e is an important property for a principled performance measure.

Since X and $\hat{X}(Z)$ can have only cardinality zero or one, there are only four possibilities that need to be considered for defining error e. For both X and $\hat{X}(Z)$ having the cardinality one, the definition of error is as usual $e(X = \{x\}, \hat{X} = \{x'\}) = x - x'$, which is position error. The error will be zero for both having zero cardinality. We need only to define error for the two possible cardinality mismatch. Here we assign e_0 and e_1 to the error in the following two possible cardinality mismatch,

$$e_0 \triangleq e(X = \emptyset, \hat{X}(Z) = \{x'\}), \text{ for any } x',$$

$$e_1 \triangleq e(X = \{x\}, \hat{X}(Z) = \emptyset), \text{ for any } x.$$

Moreover, e is the optimal subpattern assignment (OSPA) metric [3] if $e_0 = e_1$ and
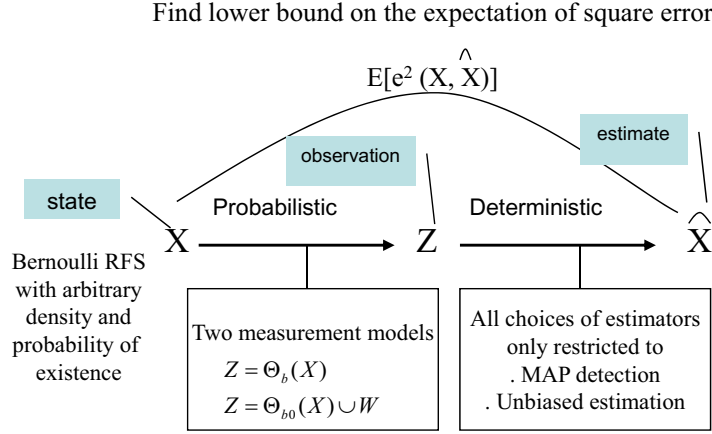
$$e_1 \geq |x - x'| \ \& \ \forall x, x' \in \mathcal{X}.$$

The results establish in the following hold for more general values of e_0 and e_1 (non-symmetric distance), so one can penalize the two cardinality errors differently.

The bounds on the MSE (11) are established for two specific measurement models that we define in Sections III-A and III-B and for estimators satisfying the following two conditions. The schematic representation of system variables, measurement models and constraints is shown in Figure 1.

1-*MAP detection:* The cardinality estimate for X, when Z is non-empty is based on the Maximum A-posterior Probability (MAP) criteria. In other words, for $Z \neq \emptyset$, the decision about whether the target exists or not ($|\hat{X}| = 1$ or $|\hat{X}| = 0$) is based on the maximum of the probabilities, $Pr(|X| = 1|Z), Pr(|X| = 0|Z)$. Although this criterion is Bayes optimal, it may not achieve the minimum mean square error defined according to (11). We need, however, the MAP assumption in deriving the bounds on MSE. We see later that without this restriction, characterizing the joint detector-estimator that achieves minimum MSE is impossible. For $Z = \emptyset$ we do not need the MAP detection assumption.

Fig. 1. Illustration of the joint detection-estimation problem and its generality. The observation Z is probabilistically related to the state X, where we have considered two probabilistic models for such relation, but the estimate $\hat{X}$ is a deterministic function of Z, where we have imposed only two constraints on such functions, and with such generality we find how far we can reduce $E[e^2(X, \hat{X})]$.

Find lower bound on the expectation of square error

$$E[e^2 (X, \hat{X})]$$

state

Bernoulli RFS with arbitrary density and probability of existence

observation

Probabilistic

$X \longrightarrow Z$

estimate

Deterministic

$\longrightarrow \hat{X}$

Two measurement models

$Z = \Theta_b(X)$

$Z = \Theta_{b0}(X) \cup W$

All choices of estimators only restricted to
. MAP detection
. Unbiased estimation

2- *Unbiased estimation:* We call an estimator $\hat{X}(\cdot)$ unbiased, if $E_{|\dot{x},|Z|=n}\hat{x} = \dot{x}$ for any $n > 0$ such that $Pr(|Z| = n|X = \{\dot{x}\}) > 0$, where the real function $\hat{x}(\cdot)$ is defined as $\hat{x}((z_1, \cdots, z_n)) = \hat{X}(\{z_1, \cdots, z_n\})$ when the RHS is nonempty, otherwise it is zero. This notation means that assuming the state is $X = \{\dot{x}\}$, the expected estimation over all possible n-points observations Z which are not mapped to empty set estimate is $\dot{x}$, i.e:

$$\int_{\mathcal{Z}^n} \hat{x}(\mathbf{z})g_n(\mathbf{z}|\dot{x})d\mathbf{z} = \dot{x}, \text{ for any } \dot{x} \text{ and any } n > 0 \text{ such that } Pr(|Z| = n|X = \{\dot{x}\}) > 0, \quad (12)$$

where $g_n(\mathbf{z}|\dot{x})$ is the density of observing at $\mathbf{z}$ which is not mapped to $\hat{X} = \emptyset$ given $X = \{\dot{x}\}$.

We now describe the two measurement models and derive the corresponding error bounds.

A. First measurement model

In this model there is no clutter in the observation[2], and the conditional density of the measurement Z given the state X is Bernoulli, i.e:

Measurement model I: $\qquad\qquad\qquad Z = \Theta_b(X),$

[2]We can fit this measurement model to more general systems that has independent clutter with any cardinality by considering the measurement space $\mathcal{Z}$ to be the space of random set observations, then any set observation is a point in this space. However this analysis is very complex. Here we consider a point in this model as one measurement (not a set of measurement).

where

$$f_{\Theta_b}(Z|X) = \mathcal{B}_Z(P_o(X), g(\cdot|X)),$$

and $g(\cdot|X)$ is the conditional probability density of observing a point z when the state is X. Hence, for any X, $\int_{\mathcal{Z}} g(z|X)dz = 1$. Also, $0 \leq P_o(X) \leq 1$ is the probability of observing a point in $\mathcal{Z}$ conditioned on state being at X. We remark that $P_o(\cdot)$ is not a density over $\mathcal{X}$. For clarity of exposition of results, we use the following notations for $X = \emptyset$,

$$q \triangleq P_o(\emptyset), \quad \psi(z) \triangleq g(z|\emptyset).$$

So equivalently we express f_{Θ_b} as,

$$f_{\Theta_b}(Z|X = \{x\}) = \begin{cases} 1 - P_o(x), & Z = \emptyset; \\ P_o(x)g(z|x), & Z = \{z\}; \\ 0, & \text{otherwise.} \end{cases} \qquad f_{\Theta_b}(Z|X = \emptyset) = \begin{cases} 1 - q, & Z = \emptyset; \\ q\psi(z), & Z = \{z\}; \\ 0, & \text{otherwise.} \end{cases}$$
(13)

The first expression models uncertainty in detection, and the second models false observations. In this representation $P_o(x)$ is the probability of observation when the state is at x, and $g(z|x)$ is the measurement likelihood when the target is detected. On the other hand, $\psi(\cdot)$ is the density of false alarm and q is the probability of false alarm. Although there is no independent clutter in this model, there can still be spurious measurement in the absence of target as false observation. In the special case of $q = 0$ that there is no such spurious measurement, we denote the measurement model by $\Theta_{b0}(X)$. Hence $f_{\Theta_{b0}}(\cdot|\emptyset)$ has all its probability mass at $Z = \emptyset$.

As it turns out, the bound on the MSE for this measurement model (and also for measurement model II that uses $\Theta_{b0}(X)$) are expressed in terms of the average probability of observation, defined as

$$\bar{P}_o = \int_{\mathcal{X}} f(x)P_o(x)dx, \tag{14}$$

and some other parameters defined based on the following densities (see Appendix A for the description of these densities),

$$\begin{aligned} f_d(x) &= f(x)P_o(x)/\bar{P}_o, \\ f_{d'}(x) &= f(x)(1 - P_o(x))/(1 - \bar{P}_o), \\ p_1(x, z) &= f_d(x)g(z|x). \end{aligned} \tag{15}$$

In particular all the bounds involve

$$\sigma_e^2 \triangleq \int_{\mathcal{X}} f_{d'}(x)(x - x^*)^2 dx, \tag{16}$$

where x^* and σ_e^2 are the mean and the variance of a random variable corresponding to the density $f_{d'}(\cdot)$. We obtain the following bound on the MSE of the first measurement model that we denote by σ_1^2.

Theorem 1: The mean square error of joint MAP detection and unbiased estimation of a single object with Bernoulli prior $\mathcal{B}_X(r, f(\cdot))$ and measurement model I is bounded by

$$\sigma_1^2 \geq r\bar{P}_o J_1'^{-1} + r(1 - \bar{P}_o)(1 - \mu_0)\sigma_e^2 + r(\mu_0(1 - \bar{P}_o) + \bar{P}_o - \bar{P}_o w_1)e_1^2 + (1 - r)(1 - \mu_1)e_0^2, \tag{17}$$

where

$$\mu_0 = \begin{cases} 0, & r \geq r_1^*; \\ 1, & r < r_1^*. \end{cases} \quad \mu_1 = \begin{cases} \eta_1, & r \geq r_1^*; \\ \eta_1 + 1 - q, & r < r_1^*. \end{cases} \quad r_1^* = \frac{(1-q)e_0^2}{(1-q)e_0^2 + (1-\bar{P}_o)(e_1^2 - \sigma_e^2)}, \tag{18}$$

$$\eta_1 = q - q \int_{\mathcal{Z}_0'} \psi(z)dz,$$

$$\mathcal{Z}_0' = \{z \in \mathcal{Z} : \beta_b(z) \geq \frac{1-r}{r\bar{P}_o}\}, \quad \beta_b(z) = \frac{\int p_1(x, z)dx}{q\psi(z)}, \tag{19}$$

and J_1' and w_1 are (assuming $J_1' = \infty$ for $\mathcal{Z}_0' = \emptyset$),

$$J_1' = -\frac{1}{w_1^2} \int_{\mathcal{X}} \int_{\mathcal{Z}_0'} p_1(x, z)[\frac{\partial^2 \log p_1(x, z)}{\partial x^2}]dzdx, \qquad w_1 = \int_{\mathcal{X}} \int_{\mathcal{Z}_0'} p_1(x, z)dzdx. \tag{20}$$

The bound in the above Theorem can also be written as

$$\sigma_1^2 \geq \begin{cases} r\bar{P}_o J_1'^{-1} + r(1 - \bar{P}_o)\sigma_e^2 + r(\bar{P}_o - \bar{P}_o w_1)e_1^2 + (1 - r)(1 - \eta_1)e_0^2, & r \geq \frac{(1-q)e_0^2}{(1-q)e_0^2 + (1-\bar{P}_o)(e_1^2 - \sigma_e^2)} \\ r\bar{P}_o J_1'^{-1} + r(1 - \bar{P}_o w_1)e_1^2 + (1 - r)(q - \eta_1)e_0^2, & r < \frac{(1-q)e_0^2}{(1-q)e_0^2 + (1-\bar{P}_o)(e_1^2 - \sigma_e^2)}. \end{cases} \tag{21}$$

The second expression in (21) can be obtained by replacing σ_e^2 with e_1^2 and adding to η_1 value of $1 - q$ in the first expression. We see in Lemma 1 in Section IV $\mathcal{Z}_0' \subset \mathcal{Z}$ is the region of observations that a MAP detector associates to the existence of target, hence η_1 is the probability of identifying false alarms in the system with MAP detector.

B. Second measurement model

The second measurement model that we consider here is more general where observation is affected by independent Poisson distributed clutter. In this model the spurious measurements are considered completely independent of target generated observation, hence we use $\Theta_{b0}(X)$ instead of $\Theta_b(X)$ for the target generated signal where in the absence of target $\Theta_{b0}(\emptyset) = \emptyset$ with probability one.

Measurement model II : $$Z = \Theta_{b0}(X) \cup W,$$

where W is a Poisson RFS with a given intensity function $v(\cdot)$ and $\lambda = \int v(z) dz$.

The error bound for the second measurement model is based on a sequence of joint probability densities $p_n(\cdot, \cdot)$ over $\mathcal{X} \times \mathcal{Z}^n$ that we define as (note $p_1(\cdot, \cdot)$ is defined in (15))

$$p_n(x, \mathbf{z}) \triangleq \frac{1}{n\lambda^{n-1}} \sum_{i=1}^{n} p_1(x, z_i) \prod_{j \neq i} v(z_j), \quad n = 1, 2, \cdots \tag{22}$$

In general, $p_n(x, \mathbf{z})$ is the joint probability density of $(x, \mathbf{z})$ conditioned on $|Z| = n$, and that one of the points of Z is target generated (see Appendix A). We find the following bound on the MSE for the second measurement model that we denote by σ_2^2.

Theorem 2: The mean square error of joint MAP detection and unbiased estimation of a single object with Bernoulli prior $\mathcal{B}_X(r, f(\cdot))$ and measurement model II is bounded by

$$\sigma_2^2 \geq r\bar{P}_o \sum_{n=0}^{\infty} C_n J_{1+n}'^{-1} + r(1-\bar{P}_o)(1-\mu_2)\sigma_e^2 + r(\mu_2(1-\bar{P}_o) + \bar{P}_o - \bar{P}_o \sum_{n=0}^{\infty} C_n w_{n+1})e_1^2 + (1-r)(1-\mu_2)e_0^2, \tag{23}$$

where $C_n = \frac{e^{-\lambda}\lambda^n}{n!}$ is the discrete Poisson distribution, and

$$\mu_2 = \begin{cases} \eta_2, & r \geq \frac{e_0^2}{e_0^2 + (1-\bar{P}_o)(e_1^2 - \sigma_e^2)}; \\ \eta_2 + e^{-\lambda}, & r < \frac{e_0^2}{e_0^2 + (1-\bar{P}_o)(e_1^2 - \sigma_e^2)}. \end{cases} \tag{24}$$

$$\eta_2 = e^{-\lambda_0} - e^{-\lambda}, \quad \lambda_0 = \int_{\mathcal{Z}_0'^1} v(z) dz.$$

$$\mathcal{Z}_0'^n = \{\mathbf{z} \in \mathcal{Z}^n : \beta(\mathbf{z}) \geq 1 + \frac{1-2r}{r\bar{P}_o}\}, \quad \beta(\mathbf{z}) = \sum_{i=1}^{n} \frac{\int p_1(x, z_i) dx}{v(z_i)},$$

and J_n' and w_n are (assuming $J_n' = \infty$ for $\mathcal{Z}_0'^n = \emptyset$)

$$J_n' = -\frac{1}{w_n^2} \int_{\mathcal{X}} \int_{\mathcal{Z}_0'^n} p_n(x, \mathbf{z}) \left[\frac{\partial^2 \log p_n(x, \mathbf{z})}{\partial x^2}\right] d\mathbf{z} dx, \quad w_n = \int_{\mathcal{X}} \int_{\mathcal{Z}_0'^n} p_n(x, \mathbf{z}) d\mathbf{z} dx. \tag{25}$$

Here also η_2 is the probability of identifying false alarms in the system with MAP detector, and $\mathcal{Z}_0'^n$ is the region in $\mathcal{Z}^n$ associated to existence of target when an n-point observation happens.

For the special case that $\mathcal{Z}_0'^n = \mathcal{Z}^n$, we denote J_n' by J_n which is the Fisher information corresponding to $p_n(x, \mathbf{z})$

$$J_n = -E_{p_n} \frac{\partial^2 \log p_n(x, \mathbf{z})}{\partial x^2} = -\int_{\mathcal{X}} \int_{\mathcal{Z}^n} p_n(x, \mathbf{z}) \left[\frac{\partial^2 \log p_n(x, \mathbf{z})}{\partial x^2} \right] d\mathbf{z} dx. \tag{26}$$

C. Specialization and common results for the two measurement models

First we note that the integral defining J_1' in Theorem 1 and J_n' for $n = 1$ in Theorem 2 have different regions of integration $\mathcal{Z}_0'$ and $\mathcal{Z}_0'^1$, but when the regions become $\mathcal{Z}$ they are both equal to J_1. Both bounds are continuous functions of r (despite the discontinuity in μ_0, μ_1 and μ_2 with respect to r). In both Theorems η_1 and η_2 are integrals of the conditional density $\gamma(Z|X = \emptyset)$ over all non-empty sets in the detector null region (associated to the non-existence of target). The jump in μ_1 and μ_2 in the two bounds is equal to $Pr(Z = \emptyset | X = \emptyset)$.

The expressions for the two bounds in Theorems 1 and 2 (as functions of r) change when r exceeds some $r^* < 1$. We refer to the case $r \geq r^*$ as High Probability of Existence (HPE) phase and $r < r^*$ as Low Probability of Existence (LPE) phase. Here we discuss how the expressions for the bounds in the two theorems reduce significantly in three special cases.

1) asserted target: Since $r^* < 1$, the asserted target system, $r = 1$ is always in HPE phase. Moreover for $r = 1$, we have $\mathcal{Z}_0' = \mathcal{Z}$ in the first Theorem, and $\mathcal{Z}_0'^n = \mathcal{Z}^n$ in the second one. Therefore for asserted target systems $J_n' = J_n$, $w_n = 1$ for all n, and $\eta_1 = \eta_2 = 0$. Denoting MSE by $\underline{\sigma}^2$ for the asserted target systems we have therefore,

Corollary 1: The MSE of unbiased estimation of an asserted target system with measurement model I is bounded by

$$\underline{\sigma}_1^2 \geq \bar{P}_o J_1^{-1} + (1 - \bar{P}_o)\sigma_e^2. \tag{27}$$

Corollary 2: The MSE of the unbiased estimation of an asserted target system with measurement model II is bounded by

$$\underline{\sigma}_2^2 \geq \bar{P}_o \sum_{n=0}^{\infty} C_n J_{1+n}^{-1} + (1 - \bar{P}_o)\sigma_e^2. \tag{28}$$

We also prove these two corollaries directly from the definition of MSE for the asserted target system after the proof of relevant Theorems in Section IV. Since Fisher information is non-negative, the bound in (28) still holds if the infinite sum is truncated to a finite number of terms. Of course, the more terms in the sum the tighter the bound will be. The same is true for the first summation term in Theorem 2. In fact J'_n is $\frac{1}{w_n}$ times the Fisher information of density $p'_n(x, \mathbf{z}) = \frac{1}{w_n} p_n(x, \mathbf{z})$ over $\mathcal{X} \times \mathcal{Z}'^n_0$.

2) No clutter system: For the measurement model $Z = \Theta_{b0}(X)$, substituting $q = 0$ in the lower bound in Theorem 1, we have $\mathcal{Z}'_0 = \mathcal{Z}$, therefore $J'_1 = J_1$, $w_1 = 1$, and $\eta_1 = 0$. Moreover, for $q = 0$ there is no ambiguity of detection when $Z \neq \emptyset$ (c.f (13)). Therefore the MAP restriction can be removed from Theorem 1 for this measurement model. Henceforth, denoting MSE for this measurement model by σ_0^2, we have

Corollary 3: The MSE of joint detection and unbiased estimation of a single object with measurement model $Z = \Theta_{b0}(X)$ is bounded by

$$\sigma_0^2 \geq \begin{cases} r(\bar{P}_o J_1^{-1} + (1 - \bar{P}_o)\sigma_e^2) + (1 - r)e_0^2, & r \geq \frac{e_0^2}{e_0^2 + (1 - \bar{P}_o)(e_1^2 - \sigma_e^2)}; \\ r(\bar{P}_o J_1^{-1} + (1 - \bar{P}_o)e_1^2), & r < \frac{e_0^2}{e_0^2 + (1 - \bar{P}_o)(e_1^2 - \sigma_e^2)}. \end{cases} \quad (29)$$

As we expect, the error bound for the measurement model $Z = \Theta_{b0}(X)$ can also be obtained from that of measurement Model II with $\lambda = 0$. In fact, for $\lambda = 0$, $v(\cdot) \equiv 0$, hence $\mathcal{Z}'^m_0 = \mathcal{Z}^n$, $\lambda_0 = \lambda$, $\eta_2 = 0$, and therefore (23) reduces to (29).

3) Constant probability of observation: For uniform sensor field of view, i.e: $P_o(\cdot)$ is a constant $\bar{P}_o$, the bounds in Corollaries 1 and 2 and also each LPE and HPE segment of the bound in Corollary 3 will be linear functions of $\bar{P}_o$. This is because in this case we have $f_d(\cdot) = f_{d'}(\cdot) = f(\cdot)$, and $p_1(x, z)$ reduces to $p(x, z) = f(x)g(z|x)$, hence J_n is independent of $\bar{P}_o$. Note that in this case $\sigma_e^2 = \sigma_f^2$ is the variance with respect to the prior density $f(\cdot)$.

D. Examples

Example 1: Gaussian measurement models: Here we consider a no-clutter measurement model I with constant probability of observation where observation $z = ax + v$ happens with probability of $\bar{P}_o$. In this linear Gaussian measurement model, v is a zero mean Gaussian random variable with variance of σ_v^2, hence $g(z|x) = \mathcal{N}_z(ax, \sigma_v^2)$. Let the prior target distribution be a Bernoulli

$\mathcal{B}(r, \mathcal{N}(x_0, \sigma_x^2))$ for a given r, x_0, σ_x^2, i.e.: we know that a target may exist with probability of r and if it exists it is Gaussian distributed around x_0 with a variance of σ_x^2. The Fisher information for the joint density $p(x, z) = f(x)g(z|x)$ with respect to x will be $J_1 = \frac{1}{\sigma_x^2} + \frac{a^2}{2\sigma_v^2}$. The expected error for a joint detector- estimator is then bounded by

$$\sigma^2 \geq \begin{cases} r(\sigma_x^2 - \bar{P}_o(\frac{a^2\sigma_x^4}{2\sigma_v^2 + a^2\sigma_x^2})) + (1-r)e_0^2, & r \geq \frac{e_0^2}{e_0^2 + (1-\bar{P}_o)(e_1^2 - \sigma_x^2)}; \\ r(e_1^2 - \bar{P}_o(e_1^2 - \frac{2\sigma_x^2\sigma_v^2}{2\sigma_v^2 + a^2\sigma_x^2})), & r < \frac{e_0^2}{e_0^2 + (1-\bar{P}_o)(e_1^2 - \sigma_x^2)}. \end{cases} \tag{30}$$

In general, if the priors $f(\cdot)$ is Gaussian and the conditional density $g(z|x)$ is also Gaussian (e.g: additive Gaussian measurement noise, $z = \phi(x) + v$), with the assumption of constant $P_o(\cdot)$, letting $p(x, z) = f(x)g(z|x) = f_z(z)h(x|z)$ the bound in Corollary 3 reduces to

$$\sigma_0^2 \geq \begin{cases} r(\sigma_f^2 - \bar{P}_o(\sigma_f^2 - \underline{\sigma}_h^2)) + (1-r)e_0^2, & r \geq \frac{e_0^2}{e_0^2 + (1-\bar{P}_o)(e_1^2 - \sigma_f^2)}; \\ r(e_1^2 - \bar{P}_o(e_1^2 - \underline{\sigma}_h^2)), & r < \frac{e_0^2}{e_0^2 + (1-\bar{P}_o)(e_1^2 - \sigma_f^2)}. \end{cases} \tag{31}$$

where $\int_z f_z(z) \frac{1}{\sigma_h^2(z)} dz \triangleq \frac{1}{\underline{\sigma}_h^2}$ and $\sigma_h^2(z)$ is the variance of $h(\cdot|z)$. From Jensen's inequality and $\sigma_f^2 - E_z \sigma_h^2 = \frac{2}{\ln 2\pi e} I_p(x; z) \geq 0$, (see [19]), we have $\sigma_f^2 \geq \underline{\sigma}_h^2$, hence the bound (31) (also (30)) is decreasing with the probability of observation $\bar{P}_o$.
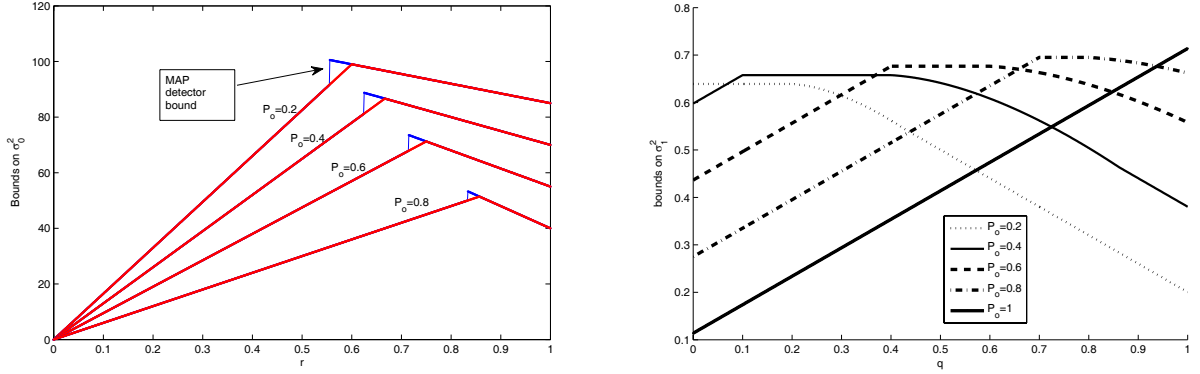
Figure 2 (left) shows the bounds on the MSE for the no-clutter measurement model $Z = \Theta_{b0}(X)$ with a constant probability of observation $\bar{P}_o$, and Gaussian prior and conditional densities $f(\cdot)$ and $g(\cdot|\cdot)$. The figure also shows the increase in the MSE bound in a small region of r when the MAP detector is used for $Z = \emptyset$. For $r = 1$, the linear decrease of the bound as $\bar{P}_o$ increases, according to the first expression in (31) is apparent in Figure 2.

Example 2: calculation of bound in Theorem 1: Let $x, z \in [0, 1]$, and over this region, $f(x) = 2x(e^x - 1)$, $g(z|x) = \frac{xe^{xz}}{e^x - 1}$. The false alarm density is assumed to be uniform, $\psi(z) = 1$. The error bound for this system over variation of q, as the probability of false alarm, for $r = 0.6$ and various P_o, and $e_0^2 = 1.2, e_1^2 = e_0^2 + \sigma_e^2$, is depicted in Figure 2 (right). For this system we have $p_1(x, z) = 2x^2 e^{xz}$, and $\beta_b(z) = 2(\frac{z^2 - 2z + 2}{qz^3})e^z - \frac{4}{qz^3}$, therefore,

$$\begin{aligned} \mathcal{Z}_0' &= \{z \in [0, 1] : \beta_b(z) > \frac{(1-r)}{rP_o}\}, & w_1 &= \int_{\mathcal{Z}_0'} \int_0^1 p_1(x, z) dx dz, \\ J_1' &= \int_{\mathcal{Z}_0'} \frac{4}{z}(e^z - 1) dz, & \eta_1 &= q(1 - \int_{\mathcal{Z}_0'} dz). \end{aligned} \tag{32}$$

We see from Figure 2 that the bounds on the MSE are not always increasing with probability of false alarm q. For low probability of observation P_o, higher probability of false alarm can be helpful in reducing the total MSE. This is because the detection error has a big contribution in

Fig. 2. Left- Lower bounds on MSE for joint detection-estimation in the general case and also with the restriction of MAP detector. The bounds are for Gaussian systems with $\sigma_f^2 = 100, \underline{\sigma}_h^2 = 25$, and cardinality error quantities $e_0^2 = 150, e_1^2 = 200.$. Right- Bounds on MSE as functions of false alarm probability and probability of observation.

the total MSE, and high probability of false alarm together with low probability of observation results in less possibility for detection error. In this case with the observation of some z, the target most likely doesn't exists, and with no observation, the target most likely exists. In the extreme case of $P_o = 0, q = 1$, the detector works perfectly, and the only source for MSE is σ_e^2, which happens with probability r, otherwise MSE is zero. The non-linearity of bound over $\bar{P}_o$ in this case is apparent from Figure 2.

E. Alternative approaches and comparison with previous results

Here we obtain alternative bounds for the systems in Corollaries 1 and 2, and compare them with previous results. First we note that the bound (27) could also be written as

$$\underline{\sigma}_1^2 \geq \bar{P}_o J_1^{-1} + (1 - \bar{P}_o) J_0^{-1}, \tag{33}$$

where $J_0 = -E(\frac{\partial^2 \log f_{d'}(x)}{\partial x^2})$. This is because for the random variable with density $f_{d'}(\cdot)$, the variance σ_e^2 and J_0 obey the inequality $\sigma_e^2 \geq J_0^{-1}$ with equality only if $f_{d'}(\cdot)$ is Gaussian (hence the bound (27) is generally tighter than (33)). This bound is derived by applying the information inequality on each summation term of the set integral expansion of the MSE (see Equation (50) in Section IV). On the other hand, as it was discussed in Section II-B, we can derive the following bound on $\underline{\sigma}_1^2$ by applying the information inequality directly on the joint density $f(x, Z)$,

$$\underline{\sigma}_1^2 \geq (\bar{P}_o J_1 + (1 - \bar{P}_o) J_0)^{-1}. \tag{34}$$

However, by the the convexity of inverse function and using Jenson's inequality, we can show that the bound (33) always dominates (34).

Here we show that (33) corresponds to the Enumeration bound $P_k(\text{Enum})$ and (34) corresponds to the Information Reduction Factor bound $P_k(\text{IRF})$ discussed in [8], specialized to one time instant. In [8] these two Posterior (filtering) Cramér-Rao bounds have been compared for a system that in which the state evolves according to $x_{k+1} = f_k(x_k) + w_k$ and the observation $z_k = h_k(x_k) + v_k$ happens with probability $\bar{P}_o$, where w_k and v_k are independent zero mean Gaussian processes with covariance matrix Σ_k, R_k, respectively. By specializing to

$$x_{k+1} = x_k, \qquad z_k = h_k(x_k) + v_k,$$

the error in the estimation of x_k in this system will be the same as the error in estimating a single random variable x from k conditionally independent measurements, e.g: from k different sensors, with the measurement model I in the asserted target case. Here, we assume that the estimate $\hat{x}$ will be improved by consecutive measurements, which may or may not contain target generated observations. With the above specialization, the following terms defining $P_k(\text{Enum})$ and $P_k(\text{IRF})$ bounds in [8] reduces to (assuming constant $P_o(x)$)

$$[F_k^{-1}]^T J_k F_k^{-1} = J_k, \qquad H_{k+1}^T R_{k+1}^{-1} H_{k+1} = (J_1 - J_0),$$

where F_k and H_k were the Jacobian of the f_k and h_k, respectively. With this substitutions, the recursive formulation of the Fisher information $\underline{J}$ in the IRF bound will be $\underline{J}_{k+1} = \underline{J}_k + \bar{P}_o(J_1 - J_0)$, starting with $\underline{J}_0 = J_0$. Therefore the IRF bound with k measurements will be

$$P_k(IRF) = [J_0 + k\bar{P}_o(J_1 - J_0)]^{-1}. \tag{35}$$

For the Enumeration bound, assuming $P_k = \sum_{i=1}^{2^k} w_{k,i} J_{k,i}^{-1}$, starting from $P_0 = J_0^{-1}$, we get from [8, Eq. 21],

$$P_{k+1} = (1 - \bar{P}_o)P_k + \bar{P}_o \sum_{i=1}^{2^k} w_{k,i}(J_{k,i} + J_1 - J_0)^{-1}. \tag{36}$$
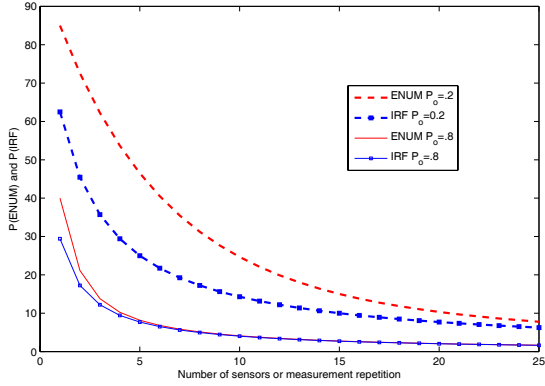
This results in

$$P_k(\text{Enum}) = \sum_{m=0}^{k} B_m (J_0 + m(J_1 - J_0))^{-1}, \tag{37}$$

where $B_m = \binom{k}{m} \bar{P}_o^m (1 - \bar{P}_o)^{k-m}$ is the Binomial distribution. We note that the bounds in (33) and (34) are special case of $P_k(\text{Enum})$ and $P_k(\text{IRF})$, respectively, for $k = 1$ measurement. We

also note that due to $J_1 \geq J_0$ both bounds $P_k(IRF)$ and $P_k(\text{Enum})$ are decreasing with k. Figure 3 shows that for the system parameters shown in Figure 2 with $r = 1$, the two bounds are asymptotically the same. Generally, $P_k(\text{Enum})$ is a tighter bound, but it is harder to calculate.

There is also a similarity between the bound in Corollary 2 and the improved posterior Cramér Rao lower bound for asserted target system derived in [9]. Specializing the bound in [9] to one sensor and one time instance, the bound is expressed as $\underline{\sigma_2^2} \geq \mathbb{E}_m J^{-1}(m)$, where in the notation [9], $J(m)$ is the Fisher information conditional on the number of measurement to be m. In (28), with the condition that a target generated point exists, C_n is also the probability of observing $m = n + 1$ points, hence the summation term is $\sum_m P(m)J^{-1}(m) = \mathbb{E}_m J^{-1}(m)$, which is the conditional expectation of J^{-1} on the condition that one point is target generated. Similarly, we can write (37) as $P_k(\text{Enum}) = \mathbb{E}_m J^{-1}(m)$, where $J(m) = J_0 + m(J_1 - J_0)$. Therefore, the Enumeration bound (37) is also the same as the bound for asserted target system derived in [9] by conditioning on the number of target generated observations in k measurements.

IV. PROOF OF MAIN THEOREMS AND DISCUSSIONS

Here we prove the bounds on the MSE for the two fundamental measurement models described in Section III in the general case of $r \in [0, 1]$, followed by some special attentions to $r = 1$. The proof of Theorem 1 provides some backgrounds for the proof of Theorem 2. We also discuss the effect of the MAP detection criterion on the MSE, and the necessity of imposing this criteria for finding the error bounds, in particular the complexity that arises in characterizing the minimum

MSE detector for systems with false alarms.

First we note that for the first measurement model the observation likelihood function is $\gamma(Z|X) = f_{\Theta_b}(Z|X)$ defined in (13), and for the second measurement model the likelihood function is

$$\gamma(Z|X = \emptyset) = \ e^{-\lambda}v^Z, \quad Z = \{z_1, z_2, ..., z_n\}, \quad \text{(we note that } \gamma(\emptyset|X = \emptyset) = e^{-\lambda})$$

$$\gamma(Z|X = \{x\}) = \ \tfrac{1-P_o(x)}{e^{\lambda}}v^Z + \tfrac{P_o(x)}{e^{\lambda}}\sum_{z\in Z} g(z|x)(v)^{Z-\{z\}}, \quad \text{(note } \gamma(\emptyset|x) = e^{-\lambda}(1 - P_o(x))).$$
$$(38)$$

Hereafter, for simplicity we use the notation $v^Z \triangleq \prod_{i=1}^{n} v(z_i)$ for a function $v(\cdot)$ and a set $Z = \{z_1, z_2, ..., z_n\}$. The first expression is due to the fact that $f_{\Theta_0}(\emptyset) = \emptyset$ with probability one and W has Poisson distribution. The second expression is from [10].

We also need the following results.

Lemma 1: With observation $Z \neq \emptyset$, the MAP detector assigns $\hat{X} = \emptyset$ if and only if

- for the first measurement model $z \in \mathcal{Z}_0$, where $\mathcal{Z}_0 = \mathcal{Z} - \mathcal{Z}_0'$

- for the second measurement model $\mathbf{z} \in \mathcal{Z}_0^n$ for some n, where $\mathcal{Z}_0^n = \mathcal{Z}^n - \mathcal{Z}_0'^n$.

Proof: For the first measurement model, using Bayes rule on the posterior probabilities $p_{01}(z) \triangleq P(X = \emptyset|Z = \{z\})$ and $p_{11}(z) = P(X \neq \emptyset|Z = \{z\})$,

$$
\begin{aligned}
p_{01}(z) = \ & P(X = \emptyset)P(Z = \{z\}|X = \emptyset)/K = \ \tfrac{(1-r)q\psi(z)}{K} \\
p_{11}(z) = \ & P(X \neq \emptyset)P(Z = \{z\}|X \neq \emptyset)/K' = \ \tfrac{r\bar{P}_o \int p_1(x,z)dx}{K},
\end{aligned}
\tag{39}
$$

where K is a normalizing factor. Note that $P(Z = \{z\}|X \neq \emptyset) = \int_{\mathcal{X}} f(x)\gamma(z|X = \{x\})dx$. Hence, the MAP detector upon observing $Z = \{z\}$ will assign $\hat{X} = \emptyset$ if $p_{01}(z) > p_{11}(z)$, i.e.:

$$\beta_b(z) < \frac{1-r}{r\bar{P}_o} \Leftrightarrow z \in \mathcal{Z}_0.$$

For the second model, the posterior probabilities are functions of random finite set Z,

$$
\begin{aligned}
p_{01}(Z) = \ & P(X = \emptyset)P(Z|X = \emptyset)/K = \ \tfrac{(1-r)e^{-\lambda}v^Z}{K} \\
p_{11}(Z) = \ & P(X \neq \emptyset)P(Z|X \neq \emptyset)/K' = \ \tfrac{r\int f(x)\gamma(Z|X=\{x\})dx}{K}.
\end{aligned}
\tag{40}
$$

Considering the relation

$$\frac{e^{\lambda}}{v^Z} \int f(x)\gamma(Z|X = \{x\})dx = 1 - \bar{P}_o + \bar{P}_o\beta(Z),$$

we have $p_{11}(Z) < p_{01}(Z)$ iff $1 - \bar{P}_o + \bar{P}_o\beta(Z) < \frac{1-r}{r}$. Hence, if the estimator receives a non-empty set Z with cardinality n, it will assign $\hat{X} = \emptyset$ if $Z \in \mathcal{Z}_0^n$, otherwise it will consider the existence of target. ∎

Lemma 2: With observation $Z = \emptyset$, in both measurement models, the best nonempty state estimate that has minimum effect on the total MSE is $\hat{x}(\emptyset) = x^*$, where x^* is the mean corresponding to the density $f_{d'}(\cdot)$.

Proof: Extending the set integral definition of MSE, we see that the total MSE is the summation of a number of terms, and the effect of estimate $\hat{X}(\emptyset)$ only appears on the following two terms,

$$P(X = \emptyset, Z = \emptyset)e^2(\emptyset, \hat{X}(\emptyset)) + \int_{\mathcal{X}} f(x)\gamma(Z = \emptyset|x)e^2(x, \hat{X}(\emptyset))dx.$$

Assuming $\hat{X}(\emptyset) = \hat{x}$ be non-empty, the first term does not change with $\hat{x}$. Letting $\hat{X}(\emptyset) = x^* + d$, the second term, for the first measurement model is equal to

$$\int_{\mathcal{X}} f(x)(1 - P_o(x))(x - \hat{X}(\emptyset))^2 dx = (1 - \bar{P}_o) \int f_{d'}(x)(x - x^* - d)^2 dx = (1 - \bar{P}_o)(\sigma_e^2 + d^2),$$

which will be minimized if $d = 0$, i.e.: $\hat{X}(\emptyset) = x^*$. For the second measurement model this term is

$$\int_{\mathcal{X}} f(x)\gamma(Z = \emptyset|x)e^2(x, \hat{X}(\emptyset))dx = \int_{\mathcal{X}} f(x)e^{-\lambda}(1 - P_o(x))(x - \hat{X}(\emptyset))^2 dx,$$

which again will be minimized for $\hat{X}(\emptyset) = x^*$. ∎

Note however that x^* is not the MAP estimate for nonempty $\hat{x}(\emptyset)$. In fact, (see Appendix A) $f_{d'}(x)$ is the posterior density taking into account the event $Z = \emptyset$, and therefore the MAP estimate for $Z = \emptyset$ is $\hat{x}(\emptyset) = \arg\max f_{d'}(x)$.

Proof: [Theorem 1] Extending the set integral definition of MSE in (11) by $f(X, Z) = \mathcal{B}_X(r, f(\cdot))f_{\Theta_b}(Z|X)$,

$$\sigma_1^2 = (1 - r)(1 - q)e^2(\emptyset, \hat{X}(\emptyset)) + (1 - r) \int_{\mathcal{Z}} q\psi(z)e^2(\emptyset, \hat{X}(z))dz$$

$$+ r \int_{\mathcal{X}} f(x)(1 - P_o(x))e^2(x, \hat{X}(\emptyset))dx + r \int_{\mathcal{Z}} \int_{\mathcal{X}} f(x)P_o(x)g(z|x)e^2(x, \hat{X}(z))dxdz. \quad (41)$$

According to Lemma 1,

$$z \in \mathcal{Z}_0' \Rightarrow \quad e^2(x, \hat{X}(z)) = (x - \hat{x}(z))^2, \quad e^2(\emptyset, \hat{X}(z)) = e_0^2$$

$$z \in \mathcal{Z}_0 \Rightarrow \quad e^2(x, \hat{X}(z)) = e_1^2. \quad e^2(\emptyset, \hat{X}(z)) = 0,$$

Hence by breaking the two integrations over $\mathcal{Z}$ in (41) into regions of $\mathcal{Z}_0$ and $\mathcal{Z}_0'$, we have,

$$\sigma_1^2 = (1-r)(1-q)e^2(\emptyset, \hat{X}(\emptyset)) + (1-r)qe_0^2 \int_{\mathcal{Z}_0'} \psi(z)dz + r \int_{\mathcal{X}} f(x)(1-P_o(x))e^2(x, \hat{X}(\emptyset))dx$$

$$+ re_1^2 \int_{\mathcal{Z}_0} \int_{\mathcal{X}} f(x)P_o(x)g(z|x)dxdz + r \int_{\mathcal{Z}_0'} \int_{\mathcal{X}} f(x)P_o(x)g(z|x)(x-\hat{x}(z))^2 dxdz, \quad (42)$$

which according to (15) and definition of η_1 it reduces to

$$\sigma_1^2 = (1-r)(1-q)e^2(\emptyset, \hat{X}(\emptyset)) + (1-r)(q-\eta_1)e_0^2 + r(1-\bar{P}_o) \int_{\mathcal{X}} f_{d'}(x)e^2(x, \hat{X}(\emptyset))dx$$

$$+ r\bar{P}_o e_1^2 \int_{\mathcal{Z}_0} \int_{\mathcal{X}} p_1(x,z)dxdz + r\bar{P}_o \int_{\mathcal{Z}_0'} \int_{\mathcal{X}} p_1(x,z)(x-\hat{x}(z))^2 dxdz. \quad (43)$$

Defining the density $p_1'(x,z) = p_1(x,z)/w_1$ over $\mathcal{X} \times \mathcal{Z}_0'$ and applying the information inequality on this density with the assumption that (12) for $n=1$, and $g_1(z|x) = g(z|x)/\int_{\mathcal{Z}_0'} g(z|x)dz$ holds, we obtain

$$\sigma_{e1}'^2 \triangleq \int_{\mathcal{Z}_0'} \int_{\mathcal{X}} p_1'(x,z)(x-\hat{x}(z))^2 dxdz \geq \left[-\int_{\mathcal{X}} \int_{\mathcal{Z}_0'} p_1'(x,z) \left[\frac{\partial^2 \log p_1'(x,z)}{\partial x^2} \right] dzdx \right]^{-1}$$

$$= \left[-\int_{\mathcal{X}} \int_{\mathcal{Z}_0'} p_1'(x,z) \left[\frac{\partial^2 \log p_1(x,z)}{\partial x^2} \right] dzdx \right]^{-1} = (w_1 J_1')^{-1}. \quad (44)$$

Therefore the last integral in (43) is bounded by $J_1'^{-1}$, i.e.:

$$\sigma_1^2 \geq (1-r)(1-q)e^2(\emptyset, \hat{X}(\emptyset)) + r(1-\bar{P}_o) \int_{\mathcal{X}} f_{d'}(x)e^2(x, \hat{X}(\emptyset))dx$$

$$+ (1-r)(q-\eta_1)e_0^2 + r\bar{P}_o e_1^2 (1-w_1) + r\bar{P}_o J_1'^{-1}. \quad (45)$$

The estimator can have two possible assignments for $\hat{X}(\emptyset)$, referred to LPE and HPE estimates, which based on Lemma 2 are as follows,

$$\begin{aligned} \text{LPE:} \quad & Z = \emptyset \Rightarrow \hat{X} = \emptyset \\ \text{HPE:} \quad & Z = \emptyset \Rightarrow \hat{X} = \{x^*\}. \end{aligned} \quad (46)$$

If it chooses the LPE rule, the bound in (45) will be

$$\sigma_{1|LPE}^2 \geq r(1-\bar{P}_o)e_1 + (1-r)(q-\eta_1)e_0^2 + r\bar{P}_o(1-w_1)e_1^2 + r\bar{P}_o J_1'^{-1}, \quad (47)$$

If it chooses the HPE rule, the bound will be

$$\sigma_{1|HPE}^2 \geq (1-r)(1-q)e_0^2 + r(1-\bar{P}_o)\sigma_e^2 + (1-r)(q-\eta_1)e_0^2 + r\bar{P}_o(1-w_1)e_1^2 + r\bar{P}_o J_1'^{-1}, \quad (48)$$

Therefore a lower bound on MSE will be the minimum of the bounds on the right hand sides of (47) and (48). Consequently, a joint detector- estimator can have a lower MSE if it chooses the LPE rule when,

$$r(1 - \bar{P}_o)e_1 + (1 - r)(q - \eta_1)e_0^2 < (1 - r)(1 - \eta_1)e_0^2 + r(1 - \bar{P}_o)\sigma_e^2, \qquad (49)$$

or equivalently when $r < r_1^*$, otherwise it chooses the HPE rule. This gives the bound in (21), which is equivalent to (17). ∎

For the asserted target system, $r = 1$, where $X = \emptyset$ is impossible, a more direct proof can be followed. In this case the MSE $\underline{\sigma}_1^2$ will only have the last two terms in (41) with $r = 1$, and since in Lemma 1, $\hat{X} = \emptyset$ is not an option for estimator, $\mathcal{Z}_0'$ covers all $\mathcal{Z}$, hence

$$\underline{\sigma}_1^2 = \int_{\mathcal{X}} f(x)(1 - P_o(x))(x - \hat{x}(\emptyset))^2 dx + \int_{\mathcal{Z}} \int_{\mathcal{X}} f(x)P_o(x)g(z|x)(x - \hat{x}(z))^2 dxdz$$

$$= (1 - \bar{P}_o) \int_{\mathcal{X}} f_{d'}(x)(x - \hat{x}(\emptyset))^2 dx + \bar{P}_o \int_{\mathcal{Z}} \int_{\mathcal{X}} p_1(x, z)(x - \hat{x}(z))^2 dxdz, \quad (50)$$

which from Lemma 2 and the information inequality on the density $p_1(\cdot, \cdot)$ gives the bound in Corollary 1.

Considering the characterizations of $f_{d'}(\cdot)$ and $p_1(\cdot, \cdot)$ in Appendix A, the bound in Corollary 1 is easily explained as follows. The probability of observing $Z = \emptyset$ is $1 - \bar{P}_o$. But by $Z = \emptyset$, the estimator has to make decision based on priors conditioned on not having any observation, i.e.: the density $f_{d'}(x)$. This estimation makes MSE no less than σ_e^2. On the other hand, with the remaining probability $\bar{P}_o$ it observes $Z = \{z\}$ for some z, and makes its estimation based on joint density p_1 with MSE $\sigma_{e1}^2 = \int \int p_1(x, z)(x - \hat{x}(z))^2 dxdz \geq J_1^{-1}$. The combination of these conditional MSE gives total MSE, hence $\underline{\sigma}_1^2 \geq \bar{P}_o \sigma_{e1}^2 + (1 - \bar{P}_o)\sigma_e^2$.

The above intuitive explanation extends to general $r \in [0, 1]$ in Corollary 3, where for $Z \neq \emptyset$ there is no ambiguity about the existence of target. The main difference is that in $r \neq 1$ when the observer do not receive anything, $Z = \emptyset$, uncertainty about target existence arises. For High Probability of Existence, high r, with $Z = \emptyset$ the estimator assumes target exists, so if target actually exists it will make the same MSE as the asserted target system, but if target doesn't exist, it makes error e_0. On the contrary, knowing that the target has a LPE, with $Z = \emptyset$ the estimator assumes target doesn't exists, so it makes no error if the target doesn't exists, but if the target exists, it makes error e_1 when it doesn't receive anything. As (29) shows, on deciding

whether the target exists or not when $Z = \emptyset$, the detector naturally chooses the decision that makes the least contribution to the MSE.

For the proof of Theorem 2 we note that the expression for η_2 in this theorem can be written as

$$\eta_2 = e^{-\lambda_0} - e^{-\lambda} = \sum_{n=1}^{\infty} \frac{e^{-\lambda}(\lambda - \lambda_0)^n}{n!} = e^{-\lambda} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{Z}_0^n} v^{\mathbf{z}} d\mathbf{z},$$

where we also use $v^{\mathbf{z}} = \prod_i v(z_i)$ for a vector $\mathbf{z}$. Therefore we can use the equality

$$e^{-\lambda} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{Z}_0'^n} v^{\mathbf{z}} d\mathbf{z} = 1 - e^{-\lambda} - \eta_2. \tag{51}$$

Proof: [Theorem 2] Using the observation likelihood function (38) for the second measurement model in the definition of MSE in (11),

$$\sigma_2^2 = (1-r)e^{-\lambda}e^2(\emptyset, \hat{X}(\emptyset)) + (1-r)e^{-\lambda} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{Z}^n} v^{\mathbf{z}} e^2(\emptyset, \hat{X}(\mathbf{z})) d\mathbf{z}$$

$$+ r(1 - \bar{P}_o)e^{-\lambda} \int_{\mathcal{X}} f_{d'}(x) e^2(x, \hat{X}(\emptyset)) dx + r \int_{\mathcal{X}} f(x) \frac{1 - P_o(x)}{e^\lambda} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{Z}^n} v^{\mathbf{z}} e^2(x, \hat{X}(\mathbf{z})) d\mathbf{z} dx$$

$$+ r \int_{\mathcal{X}} f(x) \frac{P_o(x)}{e^\lambda} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{Z}^n} \sum_{i=1}^{n} g(z_i|x) \prod_{j \neq i} v(z_j) e^2(x, \hat{X}(\mathbf{z})) d\mathbf{z} dx. \tag{52}$$

We show that expression (23) for $\mu_2 = \eta_2$ and $\mu_2 = \eta_2 + e^{-\lambda}$ correspond to the lower bounds on MSE when HPE and LPE rules are chosen for $Z = \emptyset$, respectively. Since the intersection of these two bounds is (by finding the r that makes the two bounds equal)

$$r^* = \frac{e_0^2}{e_0^2 + (1 - \bar{P}_o)(e_1^2 - \sigma_e^2)},$$

and since for any r one can select the HPE or LPE rules that achieve less MSE, the lower bound (23) for μ_2 defined as (24) will be established.

HPE: $r \geq r^*, Z = \emptyset \Rightarrow \hat{X} = \{x^*\}$

Applying $e(\emptyset, \hat{X}(\emptyset)) = e_0$ and $e(X = \{x\}, \hat{X}(\emptyset)) = x - x^*$ under HPE rule in (52),

$$\sigma_2^2 = (1-r)e^{-\lambda}e_0^2 + r(1 - \bar{P}_o)e^{-\lambda}\sigma_e^2 + (1-r)e^{-\lambda} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{Z}^n} v^{\mathbf{z}} e^2(\emptyset, \hat{X}(\mathbf{z})) d\mathbf{z}]$$

$$+ r(1 - \bar{P}_o) \sum_{n=1}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} \int_{\mathcal{Z}^n} f_z(\mathbf{z}) \int_{\mathcal{X}} f_{d'}(x) e^2(x, \hat{X}(\mathbf{z})) d\mathbf{z} dx$$

$$+ r\bar{P}_o \sum_{n=1}^{\infty} \frac{n\lambda^{n-1}e^{-\lambda}}{n!} \int_{\mathcal{X}} \int_{\mathcal{Z}^n} p_n(x, \mathbf{z}) e^2(x, \hat{X}(\mathbf{z})) d\mathbf{z} dx, \tag{53}$$

where f_z is a density over $\mathcal{Z}$. According to Lemma 1 for the MAP detector,

$$\mathbf{z} \in \mathcal{Z}_0'^n \Rightarrow e^2(x, \hat{X}(\mathbf{z})) = (x - \hat{x}(\mathbf{z}))^2, e^2(\emptyset, \hat{X}(\mathbf{z})) = e_0^2$$

$$\mathbf{z} \in \mathcal{Z}_0^n \Rightarrow e^2(x, \hat{X}(\mathbf{z})) = e_1^2, e^2(\emptyset, \hat{X}(\mathbf{z})) = 0,$$

Hence breaking the integrations over $\mathcal{Z}^n$ into integration over $\mathcal{Z}_0^n$ and $\mathcal{Z}_0'^n$ gives,

$$\sigma_2^2 = (1-r)e^{-\lambda}e_0^2 + r(1-\bar{P}_o)e^{-\lambda}\sigma_e^2 + (1-r)(1-e^{-\lambda}-\eta_2)e_0^2$$

$$+ r(1-\bar{P}_o) \times \sum_{n=1}^{\infty} \frac{e^{-\lambda}}{n!} \left[\int_{\mathcal{Z}_0'^n} v^{\mathbf{z}'} d\mathbf{z}' \right] \int_{\mathcal{Z}_0'^n} f_z'(\mathbf{z}) \int_{\mathcal{X}} f_{d'}(x)(x-\hat{x}(\mathbf{z}))^2 dx d\mathbf{z}$$

$$+ e_1^2 r(1-\bar{P}_o) \sum_{n=1}^{\infty} \frac{e^{-\lambda}}{n!} \int_{\mathcal{Z}_0^n} v^{\mathbf{z}} d\mathbf{z} + r\bar{P}_o \sum_{n=1}^{\infty} \frac{n\lambda^{n-1}e^{-\lambda}}{n!} \int_{\mathcal{X}} \int_{Z_0'^n} p_n(x,\mathbf{z})(x-\hat{x}(\mathbf{z}))^2 d\mathbf{z} dx$$

$$+ r\bar{P}_o e_1^2 \sum_{n=1}^{\infty} \frac{n\lambda^{n-1}e^{-\lambda}}{n!}[1-w_n], \quad (54)$$

where we have used (51) and $f_z'(\cdot)$ is a density over $\mathcal{Z}_0'^n$. Similar to the proof of Lemma 2, we infer

$$\int_{\mathcal{Z}_0'^n} f_z'(\mathbf{z}) \int_{\mathcal{X}} f_{d'}(x)(x-\hat{x}(\mathbf{z}))^2 dx d\mathbf{z} \geq \sigma_e^2. \quad (55)$$

Moreover, by defining the density $p_n'(x, \mathbf{z}) = p_n(x, \mathbf{z})/w_n$ on $\mathcal{X} \times \mathcal{Z}_0'^n$ and applying the information inequality on this density (similar to derivation of (44), replacing w_1 with w_n and $\mathcal{Z}_0'$ with $\mathcal{Z}_0'^n$) with the assumption that (12) holds for $g_n(\mathbf{z}|x) = p_n(x, \mathbf{z})/\int_{\mathcal{Z}_0'^n} p_n(x, \mathbf{z})d\mathbf{z}$ we get

$$\int_{\mathcal{X}} \int_{Z_0'^n} p_n(x, \mathbf{z})(x-\hat{x}(\mathbf{z}))^2 d\mathbf{z} dx \geq J_n'^{-1}. \quad (56)$$

Applying the inequalities (55) and (56) and equality (51) to (54) we obtain the bound

$$\sigma_2^2 \geq (1-r)e^{-\lambda}e_0^2 + r(1-\bar{P}_o)e^{-\lambda}\sigma_e^2 + (1-r)(1-e^{-\lambda}-\eta_2)e_0^2$$

$$+ r(1-\bar{P}_o)(1-e^{-\lambda}-\eta_2)\sigma_e^2 + e_1^2 r(1-\bar{P}_o)\eta_2 + r\bar{P}_o \sum_{n=0}^{\infty} C_n J_{1+n}'^{-1} + r\bar{P}_o e_1^2 (1 - \sum_{n=0}^{\infty} C_n w_{n+1}),$$

$$(57)$$

which is the same as (23) when $\mu_2 = \eta_2$.

LPE: $r < r^*, Z = \emptyset \Rightarrow \hat{X} = \emptyset$

Applying $e(\emptyset, \hat{X}(\emptyset)) = 0$ and $e(X \neq \emptyset, \hat{X}(\emptyset)) = e_1$ for LPE in (52), we obtain an expression that its only difference with (53) is that the term $(1 - r)e^{-\lambda}e_0^2 + r(1 - \bar{P}_o)e^{-\lambda}\sigma_e^2$ in (53) is replaced with $r(1 - \bar{P}_o)e^{-\lambda}e_1^2$, therefore instead of (57) we have

$$\sigma_2^2 \geq r(1 - \bar{P}_o)e^{-\lambda}e_1^2 + (1 - r)(1 - e^{-\lambda} - \eta_2)e_0^2 + r(1 - \bar{P}_o)(1 - e^{-\lambda} - \eta_2)\sigma_e^2 + e_1^2 r(1 - \bar{P}_o)\eta_2$$
$$+ r\bar{P}_o \sum_{n=0}^{\infty} C_n J_{1+n}'^{-1} + r\bar{P}_o e_1^2 (1 - \sum_{n=0}^{\infty} C_n w_{n+1}). \quad (58)$$

This inequality reduces to (23) when we apply $\mu_2 = \eta_2 + e^{-\lambda}$. ∎

For the asserted target system $r = 1$ with the second measurement model, the expression for MSE only has the following two terms

$$\underline{\sigma}_2^2 = \int_{\mathcal{X}} f(x)\frac{1 - P_o(x)}{e^\lambda} \sum_{n=0}^{\infty} \frac{1}{n!} \int_{Z^n} v^{\mathbf{z}}(x - \hat{x}(\mathbf{z}))^2 d\mathbf{z}dx +$$
$$\int_{\mathcal{X}} f(x)\frac{P_o(x)}{e^\lambda} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{Z^n} \sum_{i=1}^{n} g(z_i|x) \prod_{j \neq i} v(z_j)(x - \hat{x}(\mathbf{z}))^2 d\mathbf{z}dx. \quad (59)$$

From lemma 2, the first term is bounded by

$$e^{-\lambda}(1 - \bar{P}_o) \sum_{n=0}^{\infty} \frac{1}{n!} \int_{Z^n} v^{\mathbf{z}} \int_{\mathcal{X}} f_{d'}(x)(x - \hat{x}(\mathbf{z}))^2 dxd\mathbf{z}$$
$$\geq (1 - \bar{P}_o) \sum_{n=0}^{\infty} \frac{e^{-\lambda}}{n!} \int_{Z^n} v^{\mathbf{z}}\sigma_e^2 d\mathbf{z} = (1 - \bar{P}_o)\sigma_e^2, \quad (60)$$

and from the information inequality the second term is bounded by

$$\bar{P}_o e^{-\lambda} \sum_{n=1}^{\infty} \frac{n\lambda^{n-1}}{n!} \int_{\mathcal{X}} \int_{Z^n} p_n(x, \mathbf{z})(x - \hat{x}(\mathbf{z}))^2 d\mathbf{z}dx$$
$$= \bar{P}_o \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} \int_{\mathcal{X}} \int_{Z^{n+1}} p_{n+1}(x, \mathbf{z})(x - \hat{x}(\mathbf{z}))^2 d\mathbf{z}dx \geq \bar{P}_o \sum_{n=0}^{\infty} C_n J_{1+n}^{-1}, \quad (61)$$

which directly proves Corollary 2.

Equation (61) indicates that the estimation of x, based on observing n points in $\mathcal{Z}$, involves the joint density $p_n(x, \mathbf{z})$. Each term will be minimized if $\hat{X}_n(\mathbf{z}) = E_{p_n}[x|\mathbf{z}]$. This gives an operational characteristics to the sequence of densities $p_n(x, \mathbf{z})$, which is the joint density capturing all information on hand for estimating x based on the observation Z with cardinality n.

A. *MAP detection restriction*

The bound in Theorem 1 is restricted to MAP detection criterion for $Z \neq \emptyset$, but no restriction for $Z = \emptyset$. We can show that using a MAP detector for $Z = \emptyset$ will shift the HPE-LPE transition point from r_1^* to

$$r_M^* = \frac{1 - q}{2 - \bar{P}_o - q}.$$

In other words, MAP detection for $Z = \emptyset$ moves the switching point between the two linear bounds (21) to an r, which is different from the intersection of the bounds, hence increasing the bound for some r. Figure 2 illustrates this effect. However, if we choose e_0 and e_1 such that $\sigma_e^2 = e_1^2 - e_0^2$, then $r_M^* = r_1^*$ and using the MAP detector for $Z = \emptyset$ can achieve minimum MSE.

Although the MAP criterion is lifted for $Z = \emptyset$, here we show that it cannot be lifted for $Z \neq \emptyset$ (except when $q = 0$, as shown before). This is due to the complexity and interrelation of the detector and estimator that we discuss next. To find a bound on MSE without the MAP detection condition we need to find the best choice of $\mathcal{Z}_0'$ (and its complement $\mathcal{Z}_0$) that minimizes the MSE expression in Equation (43). The part of this equation that dependents on $\mathcal{Z}_0'$ is (due to the dependency of w_1, η_1, and $\sigma_{e1}'^2$ to the region $\mathcal{Z}_0'$)

$$L = r\bar{P}_o\sigma_{e1}'^2 + r\bar{P}_o(1 - w_1)(e_1^2 - \sigma_{e1}'^2) - \eta_1(1 - r)e_0^2.$$

For any infinitesimal region Δz around a point z to be added to $\mathcal{Z}_0'$, equivalently removed from $\mathcal{Z}_0$, the net change in the L quantity will be

$$\Delta L = [r\bar{P}_o d(z) - r\bar{P}_o q\psi(z)\beta_b(z)(e_1^2 + d(z)) + q\psi(z)(1 - r)e_0^2]\Delta z,$$

where

$$d(z) = \int_{\mathcal{X}} p_1(x, z)(x - \hat{x}(z))^2 dx \Big/ \int_{\mathcal{X}} p_1(x, z)dx.$$

We note that $d(z)$ depends on the estimator $\hat{x}(z)$ at z. The MSE minimizer will add a point z to the region $\mathcal{Z}_0'$ if $\frac{\Delta L}{\Delta z}$ is negative at point z, i.e: adding this point will decrease L, but ΔL depends on the estimator $\hat{x}(\cdot)$. This shows the extreme complexity in defining $\mathcal{Z}_0'$ for the detector that minimizes the MSE and its intricate interconnection with the position estimator function that jointly may achieve a lower MSE than with the MAP detector assumption. For the same reason the MAP detector restriction for $Z \neq \emptyset$ must be imposed for Theorem 2.

V. Vector Extensions

Using the vector version of the information inequality mentioned in Section II, we can extend the bounds on all MSE $\sigma_1^2, \sigma_2^2, \underline{\sigma}_1^2, \underline{\sigma}_2^2$, and σ_0^2 where $\mathcal{X}$ and $\mathcal{Z}$ instead of being subsets of $\mathcal{R}$, are subsets of $\mathcal{R}^l, \mathcal{R}^m$, respectively. The extension for $\mathcal{Z}$ from subset of $\mathcal{R}$ to $\mathcal{R}^m$ is straightforward. None of the formulations in the previous sections will change as long as we interpret $\int_{\mathcal{Z}} dz$ as vector integration over $\mathcal{R}^m$ and $\int_{\mathcal{Z}^n} d\mathbf{z}$ as vector integration over $\mathcal{R}^m \times \mathcal{R}^m \times \cdots \times \mathcal{R}^m$ for n products. The definition of RFS densities and the information inequality will not change as long as the relevant densities in z are considered as densities in $\mathcal{R}^m$. With this consideration, the definition of $\mathcal{Z}_0^n$ and $\mathcal{Z}_0'^n$ for both Theorems and Lemma 1 will be unchanged. Hence, we leave the representation of z in $\mathcal{R}^m$ same as $\mathcal{R}$, and focus only on the multidimensional $\mathcal{X}$, where for clarity, we denote the random variable representing the state in $\mathcal{X} \subset \mathcal{R}^l$ by $\mathbf{x}$, and its components by $x_i, i = 1, \cdots, l$, replacing all xs with $\mathbf{x}$ in all the previous densities.

Here we find error bounds for the joint detection and estimation of x_i. To this end, we define component error metrics $\mathbf{e}_i$, as $\mathbf{e}_i(X, \hat{X}) = x_i - \hat{x}_i$, when $|X| = |\hat{X}| = 1$, and $\mathbf{e}_i(\emptyset, \emptyset) = 0, \mathbf{e}_i(\emptyset, \mathbf{x}') = e_0/\sqrt{l}, \mathbf{e}_i(\mathbf{x}, \emptyset) = e_1/\sqrt{l}$ for any $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$. The mean square error for component i is defined as

$$\sigma_i^2 = E[\mathbf{e}_i^2(X, \hat{X}(Z)] = \int \int f(X, Z) \mathbf{e}_i^2(X, \hat{X}(Z)) \delta X \delta Z, \tag{62}$$

where $f(\cdot, \cdot)$ is the joint density of two random sets X and Z. Here for vector extensions, index i for σ indicates generic component wise mean square error. We also denote $\mathbf{x}^*$ as the mean of $\mathbf{x}$ under the density $f_{d'}(\cdot)$, and $\sigma_{e,i}^2$ as the variance of its i-th component under this density. We note that if the metric e in Section III is extended to $\mathcal{R}^l$ with the metric being equal to the the Euclidian distance in $\mathcal{R}^l$ when $|X| = |\hat{X}| = 1$, we have $e^2(X, \hat{X}) = \sum_i \mathbf{e}_i^2(X, \hat{X})$ and σ^2 and σ_i^2 in (11) and (62) are related by $\sigma^2 = \sum_i \sigma_i^2$.

Using the vector version of the information inequality, the component error bound for the measurement model I in Theorem 1 is

$$\sigma_{1,i}^2 \geq r\bar{P}_o[\mathbf{J}_1'^{-1}]_{i,i} + r(1-\bar{P}_o)(1-\mu_0)\sigma_{e,i}^2 + r(\mu_0(1-\bar{P}_o) + \bar{P}_o - \bar{P}_o w_1)e_1^2/l + (1-r)(1-\mu_1)e_0^2/l, \tag{63}$$

and for the measurement model II in Theorem 2 is

$$\sigma_{2,i}^2 \geq$$
$$r\bar{P}_o \sum_{n=0}^{\infty} C_n [\mathbf{J}_{1+n}'^{-1}]_{i,i} + r(1-\bar{P}_o)(1-\mu_2)\sigma_{e,i}^2 + r(\mu_2(1-\bar{P}_o) + \bar{P}_o - \bar{P}_o \sum_{n=0}^{\infty} C_n w_{n+1})e_1^2/l + (1-r)(1-\mu_2)e_0^2/l,$$

$$(64)$$

where (for (63), replacing $\mathcal{Z}_0'^1$ below with $\mathcal{Z}_0'$ and $\mathbf{z}$ with z)

$$[\mathbf{J}_n']_{i,j} = -\frac{1}{w_n^2} \int_{\mathcal{X}} \int_{\mathcal{Z}_0'^n} p_n(\mathbf{x}, \mathbf{z}) \left[\frac{\partial^2 \log p_n(\mathbf{x}, \mathbf{z})}{\partial x_i \partial x_j} \right] d\mathbf{z} d\mathbf{x}, \qquad w_n = \int_{\mathcal{X}} \int_{\mathcal{Z}_0'^n} p_n(\mathbf{x}, \mathbf{z}) d\mathbf{z} d\mathbf{x}, \quad (65)$$

and the rest of parameters are defined similar to Theorems 1 and 2, except that the discontinuity point for μ_0, μ_1 and μ_2 will be at $e_0^2/(e_0^2 + (1-\bar{P}_o)(e_1^2 - l\sigma_{e,i}^2))$. Here we give an outline of proof for (63) and (64) by extending the proof of Theorems 1 and 2, but mentioning only the main differences and avoiding many parallel discussions. We note that Lemma 1 will not change and in Lemma 2 under $Z = \emptyset$ the best non-empty estimate for x_i that has the least contribution in the component MSE σ_i will be x_i^*. With these considerations, parallel to (43) we get from (62),

$$\sigma_{1,i}^2 = (1-r)(1-q)\mathbf{e}_i^2(\emptyset, \hat{X}(\emptyset)) + (1-r)(q-\eta_1)e_0^2/l + r(1-\bar{P}_o) \int_{\mathcal{X}} f_{d'}(\mathbf{x})\mathbf{e}_i^2(\mathbf{x}, \hat{X}(\emptyset))d\mathbf{x}$$

$$+ r\bar{P}_o e_1^2/l \int_{\mathcal{Z}_0} \int_{\mathcal{X}} p_1(\mathbf{x}, z)d\mathbf{x}dz + r\bar{P}_o \int_{\mathcal{Z}_0'} \int_{\mathcal{X}} p_1(\mathbf{x}, z)(x_i - \hat{x}(z)_i)^2 d\mathbf{x}dz, \quad (66)$$

Defining the joint density of $p_1'(\mathbf{x}, z) = p_1(\mathbf{x}, z)/w_1$, on $\mathcal{X} \times \mathcal{Z}_0'$ and matrix

$$\mathbf{l}_1[i,j] = -E_{p_1'(\mathbf{x},z)} \frac{\partial^2 \log(p_1'(\mathbf{x}, z))}{\partial x_i \partial x_j} = -\frac{1}{w_1} \int_{\mathcal{X}} \int_{\mathcal{Z}_0'} p_1(\mathbf{x}, z)[\frac{\partial^2 \log p_1(\mathbf{x}, z)}{\partial x_i \partial x_j}]dzd\mathbf{x} = w_1 \mathbf{J}_1', \quad (67)$$

from the information inequality,

$$\sigma_{e1,i}'^2 = \int_{\mathcal{X}} \int_{\mathcal{Z}_0'} p_1'(\mathbf{x}, z)(x_i - \hat{x}_i(z))^2 dzd\mathbf{x} \geq [\mathbf{l}_1^{-1}]_{i,i} = \frac{1}{w}[\mathbf{J}_1'^{-1}]_{i,i}. \tag{68}$$

Therefore the last integral in (66) is bounded by $[\mathbf{J}_1'^{-1}]_{i,i}$. The rest of the arguments in the proof of Theorem 1 leads to (63).

In the proof of Theorem 2, by replacing e with $\mathbf{e}_i$ in (52) we obtain for $\sigma_{2,i}^2$ under HPE an equation similar to (54) in which σ_e^2, e_0^2 and e_1^2 are replaced with $\sigma_{e,i}^2, e_0^2/l$ and e_1^2/l, respectively and the variable x is replaced by vector $\mathbf{x}$. Similar to (68), by defining the density $p_n'(\mathbf{x}, \mathbf{z}) = p_n(\mathbf{x}, \mathbf{z})/w_n$ on $\mathcal{X} \times \mathcal{Z}_0'^n$ and applying the information inequality on this density, we get

$$\int_{\mathcal{X}} \int_{\mathcal{Z}_0'^n} p_n(\mathbf{x}, \mathbf{z})(x_i - \hat{x}(\mathbf{z})_i)^2 d\mathbf{z}d\mathbf{x} \geq [\mathbf{J}_n'^{-1}]_{i,i}, \tag{69}$$

which leads to (64) by the rest of the argument for HPE and LPE in the proof of Theorem 2.

By specialization of (63) and (64), the bounds in Corollaries 1, 2 and 3 extend to,

$$\underline{\sigma}^2_{1,i} \geq \bar{P}_o[\mathbf{J}_1^{-1}]_{i,i} + (1 - \bar{P}_o)\sigma^2_{e,i}, \tag{70}$$

$$\underline{\sigma}^2_{2,i} \geq \bar{P}_o \sum_{n=0}^{\infty} C_n[\mathbf{J}_{1+n}^{-1}]_{i,i} + (1 - \bar{P}_o)\sigma^2_{e,i}, \tag{71}$$

$$\sigma^2_{0,i} \geq \begin{cases} r(\bar{P}_o[\mathbf{J}_1^{-1}]_{i,i} + (1 - \bar{P}_o)\sigma^2_{e,i}) + (1 - r)e_0^2/l, & r \geq \frac{e_0^2}{e_0^2 + (1 - \bar{P}_o)(e_1^2 - l\sigma^2_{e,i})}; \\ r(\bar{P}_o[\mathbf{J}_1^{-1}]_{i,i} + (1 - \bar{P}_o)e_1^2/l), & r < \frac{e_0^2}{e_0^2 + (1 - \bar{P}_o)(e_1^2 - l\sigma^2_{e,i})}. \end{cases} \tag{72}$$

where $\mathbf{J}_n$ is the special case of $\mathbf{J}'_n$ in (65) for $\mathcal{Z}'^n_0 = \mathcal{Z}^n$.

VI. CONCLUSION

This paper considered error bounds for the joint detection and estimation of a state partially observed as a realization of a random finite set. In particular, we derived bounds for two generic observation models. In the first model the observation consists of at most one point, while in the second model the observation consists of multiple points, which include clutter and detection uncertainty. We also discussed particular cases of importance for example the estimation only problems where the target exists with certainty, but its state needs to be estimated. The part of the bounds related to estimation of the state conditioned on its existence is a generalization of the well known Cramér-Rao lower bound to finite-set-valued observation. The inverse of Fisher information in the Cramér-Rao lower bound has been replaced by the average of J'^{-1}_n over all n, where J'_n is the Fisher information over operating region of estimator in n dimensional space $\mathcal{Z}^n$, and the average is taken with respect to the cardinality distribution of clutter, assuming clutter W has cardinality $|W| = n - 1$.

The results are envisaged to be extendable to filtering problems where the state evolution is governed by a dynamic and estimation of its value is undertaken under the consideration of all past observations. Also extension is possible to multi-object systems where the state is a general RFS instead of a Bernoulli RFS.

Here we characterize densities $f_d(x).f_{d'}(\cdot), p_1(\cdot, \cdot)$ and $p_n(\cdot, \cdot)$ that are used in the paper. We can assign a binary random variable $S \in \{0, 1\}$ to the two events $Z = \emptyset, Z = \{z\}$ in (13), where $P_o(x) = Pr(S = 1|x)$, and $g(z|x) = f(Z|x, S = 1)$. We have

$$
\begin{aligned}
Pr(S = 1) &= \int Pr(S = 1|x)f(x)dx = \int P_o(x)f(x)dx &&= \bar{P}_o, \\
f(x|S = 1) &= \frac{f(x)Pr(S=1|x)}{Pr(S=1)} &&= f_d(x), \\
f(x|S = 0) &= \frac{f(x)Pr(S=0|x)}{Pr(S=0)} &&= f_{d'}(x), \\
f(x, Z|S = 1) &= f(x|S = 1)f(Z|x, S = 1) = f_d(x)g(z|x) &&= p_1(x, z).
\end{aligned}
\tag{73}
$$

These relations show that $f_d(x) = f(x|S = 1)$ is the posterior probability of x knowing that we have received an observation, and $f_{d'}(x) = f(x|S = 0)$ is the posterior probability of x otherwise. The density of $p_1(\cdot, \cdot)$ is the joint density of x, Z knowing that observation has happened.

For characterizing the densities $p_n(\cdot, \cdot)$ in (22), we note that for the measurement model II,

$$
f(Z|x, S = 0) = e^{-\lambda}v^Z,
\tag{74}
$$

because with $S = 0$ all observation is Poisson clutter, and

$$
f(Z|x, S = 1) = \sum_{z \in Z} g(z|x)e^{-\lambda}v^{Z-\{z\}},
\tag{75}
$$

which is the union of disjoint events that each of the points be state generated, while the rest are clutter. Further conditioning of (75) to $|Z| = n$, gives

$$
f(\mathbf{z}|x, S = 1, |Z| = n) = \sum_{i=1}^{n} g(z_i|x) \prod_{j \neq i} v^{z_j}/K,
\tag{76}
$$

where K is the normalizing factor $K = \int_{\mathcal{Z}^r} \sum_{i=1}^{n} g(z_i|x) \prod_{j \neq i} v^{z_j} = n\lambda^{n-1}$. Multiplying (76) by $f(x|S = 1, |Z| = n) = f_d(x)$ gives

$$
f(x, \mathbf{z}|S = 1, |Z| = n) = p_n(x, \mathbf{z}).
$$

Therefore the joint density $p_n(., .)$ is the joint density (x, z) conditioned on $|Z| = n$, and that one of the points of Z is state generated.

We also note that the likelihood function (38) can be obtained by (74) and (75),

$$
\gamma(Z|x) = f(Z|x, S = 0)Pr(S = 0|x) + f(Z|x, S = 1)Pr(S = 1|x).
$$

REFERENCES

[1] Y. Bar-Shalom and T. Fortmann, Tracking and Data Association. Academic Press, San Diego, 1988.

[2] R. P. S. Mahler, Statistical Multisource-Multitarget Information Fusion. Artech House, 2007.

[3] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," IEEE Trans. Signal Processing, vol. 56, no. 8, Aug. 2008, pp. 3447–3457.

[4] T. Kailath and H. V. Poor, "Detection of Stochastic Processes", IEEE Transactions on Information Theory, Vol. 44, NO. 6, October 1998, pp 2230-2259.

[5] H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part I. New York, Wiley, 1968.

[6] B. Bobrovsky, and M. Zakai, "A Lower Bound on the Estimation Error for Markov Processes", IEEE Transaction on Automatic control, Dec. 1975, pp 785-788.

[7] P. Tichavsky, C. Muravchik, A. Nehorai, "Posterior Cramér Rao Bounds for discrete time nonlinear filtering", IEEE Transaction on Signal Processing 46(5), May 1998, pp 1701-1722.

[8] M. Hernandez, B. Ristic, and A. Farina, L. Timmoneri, "A Comparison of Two CramérRao Bounds for Nonlinear Filtering with $Pd < 1$", IEEE Transactions on Signal Processing, Vol. 52, No.9 Sept 2004, pp 2361-2370.

[9] M. Hernandez, A. Farina, and B. Ristic "PCRLB for Tracking in Cluttered Environments: Measurement Sequence Conditioning Approach" IEEE Transactions on Aerospace and Electronic systems, Vol. 42, No. 2, April 2006, pp 680-704.

[10] B-T. Vo, B-N. Vo, and A. Cantoni, "Bayesian Filtering With Random Finite Set Observations" IEEE Transactions on Signal processing, Vol. 56, No. 4, April 2008, pp 1313-1326.

[11] D. Stoyan, W. Kendall, J. Mecke, Stochastic geometry and its applications. Chichester, New York, Wiley, 1995.

[12] R. P. S. Mahler, "Multi-target Bayes filtering via first-order multi-target moments", *IEEE Trans. Aerospace & Electronic Systems*, Vol. 39, No. 4, 2003, pp. 1152-1178.

[13] B.-N. Vo, W.-K. Ma, "The Gaussian mixture Probability Hypothesis Density filter," *IEEE Trans. Signal Processing, IEEE Trans. Signal Processing*, Vol. 54, No. 11, 2006, pp. 4091-4104.

[14] R. P. S. Mahler, "PHD filters of higher order in target number", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 43, No. 4, 2007, pp. 1523-1543.

[15] B.-T. Vo, B.-N. Vo, A. Cantoni, "Analytic implementations of the Cardinalized Probability Hypothesis Density Filter," IEEE Trans. Signal Processing, Vol. 55, No. 7, part 2, 2007, pp. 3553-3567.

[16] B.-N. Vo, S. Singh and A. Doucet, "Sequential Monte Carlo methods for Bayesian Multi-target filtering with Random Finite Sets," IEEE Trans. Aerospace and Electronic Systems, Vol. 41, No. 4, 2005, pp. 1224-1245.

[17] V. Poor, An Introduction to Signal Detection and Estimation. Springer-Verlag, 1994.

[18] I. Goodman, R. Mahler, H.T. Nguyen, Mathematics of Data Fusion. Kluwer Academic Ppublisher, 1997.

[19] T.M. Cover and J.A. Thomas, Elements of Information Theory. Wiley, New York, 1991.