A Generalization of Cramer-Rao Error Bound for Joint Detection and Estimation

COGIS 2009, Paris, 18-Nov-09

Mohammad Rezaeian and Ba-Ngu Vo

The University of Melbourne, Australia
Outline

- Review of information inequality and its variations
- Random set, and its application for modeling Joint detection-estimation from multiple measurements observations
- Two multi-object measurement models
- The error bounds for the two measurement models based on the information inequality
The simplest form of the Information Inequality

\[ f(x,y) \quad \hat{x}(\cdot) \quad \hat{\chi} \quad e = x - \hat{x}(y) \]

Assuming \( f(x, y) \) satisfies some regularity conditions and \( \hat{x}(y) \) is an unbiased estimator, \( \mathbb{E} \hat{x}_{|x=x} = x \)

\[
\int \int f(x, y)(x - \hat{x}(y))^2 \, dx \, dy \geq \left[ \int \int f(x, y) \frac{\partial^2 \log(f(x, y))}{\partial x^2} \, dx \, dy \right]^{-1}
\]

Variance of error \( \sigma_e^2 \geq J^{-1} \) Fisher information
The simplest form of the Information Inequality

\[ f(x,y) \xrightarrow{\hat{x}(\cdot)} \hat{x} \quad e = x - \hat{x}(y) \]

Assuming \( f(x,y) \) satisfies some regularity conditions and \( \hat{x}(y) \) is an unbiased estimator, \( \mathbb{E} \hat{x}_{|x=\hat{x}} = \hat{x} \)

\[
\int\int f(x,y)(x - \hat{x}(y))^2 \, dx \, dy \geq \left[ \int\int f(x,y) \frac{\partial^2 \log(f(x,y))}{\partial x^2} \, dx \, dy \right]^{-1}
\]

Variance of error

\[ \sigma_e^2 \geq J^{-1} \]

Fisher information
Cramer-Rao error bound

\[ f(x, y) \]

\[ \hat{x}(\cdot) \]

\[ \hat{x} \]

\[ e = x - \hat{x}(y) \]

\[ C = \text{covariance matrix of } e \]

\[ C_{i, j} = \int \int f(x, y) e_i e_j dxdy \]

\[ J_{i, j} = \int \int f(x, y) \frac{\partial^2 \log(f(x, y))}{\partial x_i \partial x_j} dxdy \]

Variance of error

\[ C \geq J^{-1} \]

Fisher information matrix

\[ C - J^{-1} \text{ is positive definite} \]
Posterior Cramer-Rao error bound

\[ C_k = \text{covariance matrix of } e_k \]

Given the dynamics of \( x_k \), a recursive formulation for \( J_k \) has been obtained.

\[ J_{k+1} = D_k^{33} - (D_k^{12})^T [J_k + D_k^{11}]^{-1} D_k^{12} + J_Y (k-1) \]
Our result: The Information Inequality for random sets

\[ F(X,Y) \]

\[ \hat{X}(\cdot) \]

\[ \hat{X} \]

\[ e(X, \hat{X}) \]

Need a clear definition of error

Finding a lower bound on the expected square error

\[ E[e^2] \geq ? \]

A random finite set (RFS) is simply a random variable that takes value as (unordered) finite set, 
\[ \text{RFS} = \text{finite-set-valued random variable} \]

Only two types of RFS are of interest here: 
* Bernoulli
* Poisson
Why random set

In multi-object systems, the hidden state to be estimated is a random number of objects that has random position (or kinetics).

The observation also consists of a random number of points in a signal space that their number and position are statistically dependent to the state.

Examples: Radar system, video tracking,....

In the conventional Bayesian analysis of these systems, state and measurements are treated as random vectors. FISST (Finite set statistics) has provided the tools to extend the Bayesian inference mechanisms to random sets.
Joint Detection-Estimation of a single target

Since the target may or may not exist, we cannot model it as a random variable. It is modeled by a random set $X$, $X = \emptyset$, $X = \{x\}$

The estimate $\hat{X}$ is also a set.

The estimate $|\hat{X}|$ corresponds to detection,

When $\hat{X} = \{\hat{x}\}$, then $\hat{x}$ will be the state estimate.

Therefore estimation of the finite set $\hat{X}$ is referred to as joint detection-estimation. The joint detection-estimation is a deterministic function of $Y$

All previous Error bounds assumes that the target exist.
The word detection doesn’t mean inferring the existence of target, but measuring an observation from the target.
Models

\[ F(X) \]

\[ \Gamma(Y \mid X) \]

- 1-Bernoulli random set
  - \(|X| = 0\), or \(|X| = 1\)
  - \(X = \emptyset\), \(X = \{x\}\)

- 2- Union of Bernoulli and Poisson random sets

$X$ and $Y$ connected by $\Gamma(Y \mid X)$.
Bernoulli and Poisson random sets

**Bernoulli**: defined by \((r, f(.))\),

\[
F(X) = \begin{cases} 
1 - r & X = \emptyset \\
rf(x) & X = \{x\}
\end{cases}
\]

\[
\int f(x)dx = 1
\]

**Poisson**: defined by an intensity function \(v(.)\),

\[
F(\{x_1, x_2, ..., x_n\}) = e^{-\lambda} \prod_{i=1}^{n} v(x_i)
\]

where \(\lambda = \int v(x)dx\)
Target prior probability density

\[ F(X) = \begin{cases} 
1 - r & X = \emptyset \\
r f(x) & X = \{x\} 
\end{cases} \]

\[ \int f(x) dx = 1 \]

\( F(X) \) contains all prior information about the target
\( r \) is the probability of existence of target
\( f(\cdot) \) is the density of target position if it exists

Bernoulli random set
Measurement models:

\[ \Gamma(Y \mid X) \]

First measurement model: \( Y = S \)

\( S \) is stochastically depended to \( X \)
\[ |S| = 0, \text{ or, } |S| = 1 \]

\[
F_S(S \mid X = \{x\}) = \begin{cases} 
1 - p_o(x) & S = \emptyset \\
p_o(x)g(s \mid x) & S = \{s\} 
\end{cases}
\]

\[
F_S(S \mid X = \emptyset) = \begin{cases} 
1 - p_{fa} & S = \emptyset \\
p_{fa}\varphi(s) & S = \{s\} 
\end{cases}
\]

 \( p_o(x) \) is the probability of observation (conventionally called detection) when target is at \( x \)

\( g(s \mid x) \) is the probability density for observation at \( s \), given that the target is at \( x \)

\( p_{fa} \) is the probability of false alarm

\( \varphi(.) \) is the density of false alarm
Measurement models:

First measurement model: \( Y = S \)

Second measurement model: \( Y = S_{nf} \cup W \)

\[ F_S(S \mid X = \{x\}) = \begin{cases} 
1-p_o(x) & S = \emptyset \\
p_o(x)g(s \mid x) & S = \{s\} 
\end{cases} \]

\[ F_S(S \mid X = \emptyset) = \begin{cases} 
1-p_{fa} & S = \emptyset \\
p_{fa}\varphi(s) & S = \{s\} 
\end{cases} \]

\( W \) is a Poisson random set independent of \( X \) (clutter)

\( S_{nf} \) is \( S \) when \( p_{fa} = 0 \) (no false alarm)
Performance measure: Definition of error

The difference of two random sets $X$ and $\hat{X}$ is meaningless. A metric distance should be used for quantizing the dissimilarity between two sets. $e(X, \hat{X})$ can be considered as the OSPA distance of the two sets.

For $X, \hat{X}$ having maximum cardinality 1, the distance is defined as

$$e(X, \hat{X}) = \begin{cases} 
|x - \hat{x}|, & X = \{x\}, \hat{X} = \{\hat{x}\} \\
0, & X = \hat{X} = \emptyset \\
E_0, & X = \emptyset, \hat{X} = \{\hat{x}\} \\
E_1, & X = \{x\}, \hat{X} = \emptyset 
\end{cases}$$

OSPA = OPTIMAL SUBPATTERN ASSIGNMENT

For two given $E_0, E_1$
Performance measure for joint detection-estimation:
Mean Square Error (MSE)

\[ \sigma^2 = \mathbb{E}e^2(X, \hat{X}) \]

Since \( X \), \( \hat{X} \) are random sets, and \( \hat{X} \) is a deterministic function of \( Y \)

\[ \sigma^2 = \int \int F(X, Y)e^2(X, \hat{X}(Y))\delta X\delta Y \]

\[ \sigma^2 = \int \int F(X)\Gamma(Y \mid X)e^2(X, \hat{X}(Y))\delta X\delta Y \]

\( \sigma^2 \geq \)

We find a lower bound on \( \sigma^2 \) over all possible choices of joint detector-estimators \( \hat{X}(\cdot) \)

obeying some restrictions
Restrictions on the joint detector-estimator

1 - Unbiased estimation
For any cardinality n of observation Y,
If \( X = \{ \hat{x} \} \) \( \Rightarrow \) \( E\hat{X}(\tilde{Y}_n) = \hat{x} \)

\( \tilde{Y}_n \) is the set of Y not mapped to \( \Theta \), \( \hat{X}(\tilde{Y}_n) \neq \Theta \)

2 - MAP Detection
For \( Y \neq \Theta \),
\[ |\hat{X}(Y)| = 1 \iff P(|X| = 1|Y) > P(|X| = 0|Y) \]
Error bound for the first measurement model

\[ F(X) = \begin{cases} 1-r & X = \Theta \\ rf(x) & X = \{x\} \end{cases} \]

\[ F_S(S \mid X = \{x\}) = \begin{cases} 1-p_o(x) & S = \Theta \\ p_o(x)g(s \mid x) & S = \{s\} \end{cases} \]

\[ F_S(S \mid X = \Theta) = \begin{cases} 1-p_{fa} & S = \Theta \\ p_{fa}\phi(s) & S = \{s\} \end{cases} \]

\[ \bar{p}_o = \int f(x)p_o(x)dx \]

\[ f_+(x) = \frac{f(x)p_o(x)}{\bar{p}_o} \]

\[ f_-(x) = \frac{f(x)(1-p_o(x))}{1-\bar{p}_o} \]

\[ p_1(x,s) = f_+(x)g(s \mid x) \]

Conditional density of target position knowing that an observation has happened

Conditional density of target position and signal knowing that an observation has happened
Error bound for the first measurement model

\[
F(X) = \begin{cases} 
1 - r & X = \emptyset \\
rf(x) & X = \{x\}
\end{cases}
\]

\[
F_s(S | X = \{x\}) = \begin{cases} 
1 - p_o(x) & S = \emptyset \\
p_o(x)g(s | x) & S = \{s\}
\end{cases}
\]

\[
F_s(S | X = \emptyset) = \begin{cases} 
1 - p_{fa} & S = \emptyset \\
p_{fa}\varphi(s) & S = \{s\}
\end{cases}
\]

\[
\bar{p}_o = \int f(x)p_o(x)dx \quad f_+(x) = \frac{f(x)p_o(x)}{\bar{p}_o} \quad f_-(x) = \frac{f(x)(1 - p_o(x))}{1 - \bar{p}_o}
\]

\[
p_1(x, s) = f_+(x)g(s | x)
\]

\[
\Delta = \left\{ s : \frac{\bar{p}_o\int p_1(x, s)dx}{q\varphi(s)} \geq \frac{1 - r}{r} \right\}
\]

**Estimator’s operating region:**
The set of signals \( s \) that is associated to the existence of target when using MAP detector
Error bound for the first measurement model

\[ p_1(x, s) = f_+(x) g(s \mid x) \]

\[ p'_1(x, s) = p_1(x, s) / \int_{\Delta} p_1(x, s) dsdx \]

Fisher information for Joint detection Estimation

\[ J_1 = -\int_{\Delta} p'_1(x, s) \frac{\partial^2 \log(p'_1(x, s))}{\partial x^2} dsdx \]

\[ \Delta = \left\{ s : \frac{\overline{p}_o \int p_1(x, s) dx}{q \varphi(s)} \geq 1 - r \right\} \]

**Estimator’s operating region:**
The set of signals \( s \) that is associated to the existence of target when using MAP detector
Error bound for the first measurement model

\[
\sigma^2 \geq \begin{cases} 
 p_o J_1^{-1} + a(r) & r \geq r^* \\
 p_o J_1^{-1} + b(r) & r < r^* 
\end{cases} \\
 r^* = \frac{(1 - p_{fa}) E_0^2}{(1 - p_{fa}) E_0^2 + (1 - \bar{p}_o)(E_1^2 - \sigma_-^2)}
\]

\[
a(r) = (1 - \bar{p}_o) \sigma_-^2 + r \bar{p}_o (1 - w) E_1^2 + (1 - r)(1 - \eta) E_0^2
\]

\[
b(r) = r(1 - \bar{p}_o w) E_1^2 + (1 - r)(p_{fa} - \eta) E_0^2
\]

\[
\eta = p_{fa}(1 - \int_\Delta \varphi(y) dy)
\]

The probability of recognizing false alarms

\[\sigma_-^2\] is the variance corresponding to density \( f_- (\cdot) \)

A two phase formula, \( r \geq r^* \) (HPE), \( r < r^* \) (LPE) appears due to two possible assignments for \( X \) when \( Y = \emptyset \).
Extension to vector state space

\[ F(X) = \begin{cases} 1 - r & X = \Theta \\ rf(x) & X = \{x\} \end{cases} \]

\[ F_S(S \mid X = \{x\}) = \begin{cases} 1-p_o(x) & S = \Theta \\ p_o(x)g(s \mid x) & S = \{s\} \end{cases} \]

\[ F_S(S \mid X = \Theta) = \begin{cases} 1-p_{fa} & S = \Theta \\ p_{fa}\varphi(s) & S = \{s\} \end{cases} \]

\[ C > ? \]

\[ C_{i,j} = \int \int f(X,Y)e_i e_j \delta X \delta Y \]

\[ C_{i,i} > ? \text{ Bounds for diagonal elements of covariance matrix (component MSE)} \]

\[ \sigma_i^2 \geq \begin{cases} p_o[J_1^{-1}]_{i,i} + a_i(r) & r \geq r_i^* \\ p_o[J_1^{-1}]_{i,i} + b(r) & r < r_i^* \end{cases} \]

\[ r_i^* = \frac{(1-p_{fa})E_0^2}{(1-p_{fa})E_0^2 + (1-p_o)(E_1^2 - \sigma_{-i}^2)} \]

\[ a_i(r) = r(1-\bar{p}_o)\sigma_{-i}^2 + r\bar{p}_o(1-w)E_1^2/n + (1-r)(1-\eta)E_0^2/n \]

\[ b(r) = r(1-\bar{p}_o w)E_1^2/n + (1-r)(p_{fa} - \eta)E_0^2/n \]

\[ \eta = p_{fa}(1-\int_\Delta \varphi(s)ds) \]

\[ \sigma_{-i}^2 \text{ is the variance corresponding to density } f_{-i}(\cdot), \text{ the marginalization of } f_{-i}(\cdot) \text{ to } i-\text{th component} \]
Special case of $r=1$

$$F(X) = \begin{cases} 
1 - r & X = \Theta \\
rf(x) & X = \{x\}
\end{cases}$$

$x \rightarrow x$, $s \rightarrow s$

$C_{i,i} > \ ?$ Bounds for diagonal elements of covariance matrix (component MSE)

$$\sigma_i^2 \geq \begin{cases} 
p_o[J_{1_i}^{-1}]_{i,i} + a_i(r) & r \geq r_i^* \\
p_o[J_{1_i}^{-1}]_{i,i} + b(r) & r < r_i^*
\end{cases}$$

$$a_i(r) = r(1 - \bar{p}_o)\sigma_i^{-2} + r\bar{p}_o(1 - w)E_1^2 / n + (1 - r)(1 - \eta)E_0^2 / n$$

$$r_i^* = \frac{(1 - p_{fa})E_0^2}{(1 - p_{fa})E_0^2 + (1 - \bar{p}_o)(E_i^2 - \sigma_i^2)}$$

$$\Delta = \left\{ s : \frac{\bar{p}_o \int p_1(x,s)dx}{q\varphi(s)} \geq \frac{1 - r}{r} \right\}$$

$$p'_1(x,s) = p_1(x,s) / \int_{\Delta} \int p_1(x,s)dsdx$$

$w = 1$

$$J_1 = -\int_{\Delta} \int p_1(x,s) \frac{\partial^2 \log(p_1(x,s))}{\partial x^2} dsdx$$
Second measurement model:

First measurement model: \( Y = S \)

Second measurement model: \( Y = S_{nf} \cup W \)

\[ \Gamma(Y \mid X) = \begin{cases} 
1-p_o(x) & Y = \emptyset \\
p_o(x)g(y \mid x) & Y = \{y\} 
\end{cases} \]

\[ \Gamma(Y \mid X = \emptyset, W = \emptyset) = \begin{cases} 
1 & Y = \emptyset \\
0 & Y \neq \emptyset 
\end{cases} \]

\[ \Gamma(Y \mid X = \{x\}) = \frac{1-p_o(x)}{e^{\lambda}} \prod_{y \in Y} v(y) + \frac{p_o(x)}{e^{\lambda}} \sum_{y \in Y} g(y \mid x) \prod_{y' \in Y, y' \neq y} v(y') \]

\[ \Gamma(Y \mid X = \emptyset) = e^{-\lambda} \prod_{y \in Y} v(y) \]
Fisher information for the second measurement model

For every cardinality $n$ as cardinality of $Y$, define

$$p_n(x, y) = \frac{1}{n\lambda^{n-1}} \sum_{i=1}^{n} p_1(x, y_i) \prod_{j \neq i} v(y_j)$$

$$p'_n(x, y) = p'_n(x, y) / w_n \Rightarrow J'_n$$

\[ \Delta^n = \left\{ y : \sum_{i=1}^{n} \int p_1(x, y_i) dx \geq \frac{1-2r}{rP_o} \right\} \]

The set of $n$-vector observations $y$ that is associated to the existence of target when using MAP detector
Error bound for the second measurement model

\[ \sigma^2 \geq \begin{cases} 
\bar{p}_o \sum \alpha_n (J_{n+1}')^{-1} + a(r) & r \geq r^* \\
\bar{p}_o \sum \alpha_n (J_{n+1}')^{-1} + b(r) & r < r^* 
\end{cases} \\
r^* = \frac{(1 - p_{fa})E_0^2}{(1 - p_{fa})E_0^2 + (1 - \bar{p}_o)(E_i^2 - \sigma^2)}
\]

where \( \alpha_n = \frac{e^{-\lambda} \lambda^n}{n!} \) is Poisson distribution

A two phase formula, \( r \geq r^* \) (HPE), \( r < r^* \) (LPE)
Appears due to two possible assignments for \( X \) when \( Y = \emptyset \)
For $r=1$, the bound match with the Enumeration bound

$$\sigma^2 \geq \bar{p}_o \sum \alpha_n (J'_{n+1})^{-1} + (1 - \bar{p}_o)\sigma_-^2$$

$$\alpha_n = \Pr\{\text{clutter # points is } n\} = \frac{e^{-\lambda} \lambda^n}{n!}$$

$$\bar{p}_o = \Pr\{a \text{ target generated point exists in the observation}\}$$

$$(J'_{n+1})^{-1} = \text{error bound for one target generated point plus n clutter points}$$

$$\sigma_-^2 = \text{error bound without a target generated point}$$

Enumeration bounds

$$\sum_m \Pr\{m \text{ measurments}\} J^{-1}(m)$$

Information reduction factor bound

$$(\sum_m \Pr\{m \text{ measurments}\} J(m))^{-1}$$
Conclusion

- We extended the Cramer-Rao bounds for estimation only problems to joint detection and estimation problems.
- The state and observation for such problems are best modeled by random sets.
- The error is best defined by OSPA distance of the true set and the estimated set.
- Two widely used class of measurement models were considered for the joint detection-estimation.
- The error bounds consist of inverse Fisher information restricted to the estimator’s operating region.
Thank you