ADAPTIVE COMMUNICATIONS AND SIGNAL PROCESSING LABORATORY

CORNELL UNIVERSITY, ITHACA, NY 14853

Asymptotic Locally Optimal Detector for Large-Scale Sensor Networks under the Poisson Regime

Y. Sung, L. Tong, and A. Swami

Technical Report No. ACSP-TR-10-03-01

October 16 2003

Abstract

We consider distributed detection with a large number of identical binary sensors deployed over a region where the phenomenon of interest (POI) has spatially varying signal strength. Each sensor makes a decision based on its own measurement, and the local decision of each sensor is sent to a fusion center using a random access protocol. The fusion center decides whether the POI has occurred under a global size constraint in the Neyman-Pearson formulation. Assuming a homogeneous Poisson distributed sensors, we show that the distribution of 'alarmed' sensors satisfies the locally asymptotically normal (LAN) condition. We then derive an asymptotically locally most powerful (ALMP) detector optimized jointly over the fusion form and local sensor threshold. Sufficient conditions on the spatial signal shape for the existence of ALMP detector are established. We show that the ALMP test statistic is a weighted sum of local decisions with the optimal weight function as the shape of the spatial signal, and the exact value of the signal strength is not required. We also derive the optimal threshold for each sensor. For the case of independent, identical distributed (i.i.d.) sensor observation, we show that the counting-based detector is also ALMP under the Poisson regime. The performance of the proposed detector is evaluated through analytic results and Monte-Carlo simulations, and compared with that of the counting-based detector for spatially-varying signals. The effect of mismatched signal shapes is also investigated.

Index Terms— Distributed detection, Spatially-varying signal, Spatial Poisson process, Locally asymptotically normal (LAN), Asymptotically locally most powerful (ALMP), Neyman-Pearson criterion, Fusion rule.

EDICS: 1-DRUS (Data fusion from multiple sensor types), 2-DETC (Detection theory and applications).

I. Introduction

A. Detection in Large Scale Sensor Field

We consider the detection of deterministic phenomena in a field of a large number of densely deployed microsensors. The sensors measure the phenomenon of interest (POI) and transmit their local data (decision) via wireless channels to a central site for global processing. A specific implementation is Sensor Networks with Mobile Access (SENMA) [33] where a mobile access point collects local decisions from sensors using random access schemes such as ALOHA as shown in Fig. 1. We assume that the number of sensors in the field is large, which makes it necessary that each sensor is inexpensive and has limited computation and communication capability.

Detection in a large scale sensor network faces several challenges not encountered in the

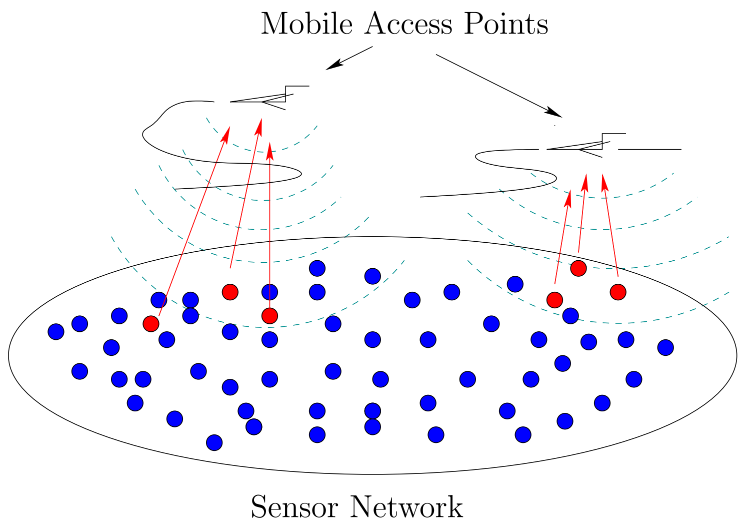

Fig. 1. Sensor Network with Mobile Access Points

classical distributed detection problem. First, inexpensive sensors are not reliable; they have low duty cycles and severe energy constraints. The communication link between a sensor and the central unit is specially weak due to a variety of implementation difficulties such as synchronization, fading, and interference from other sensors. The probability that the local decision at a particular sensor can be successfully delivered to the central unit can be very low. Second, POI in a wide geographic area generates spatially varying signals, which makes the observation at each sensor location dependent and not identically distributed. Furthermore, the strength of POI is unknown for many applications such as the detection of NBC activities. Third, the scale of the network makes it more practical to deploy sensors randomly without careful network layout. It is thus not possible to predict whether data from a particular sensor can be retrieved by the the central processing unit, especially when random access protocols are used. Consequently, the decision rule for each sensor should be optimized before deployment without knowing its exact location. In addition, because sensors may expire and the collection process random, the optimal decision should not critically depend on the number of available sensors and the collection process.

B. The Approach and Summary of Results

For large scale sensor networks, it is natural to consider asymptotic techniques, and one expects that central limit theorem will lead to the design under Gaussian statistics. For example, if the measurements at sensors are conditionally independent and identically distributed, it is well known that the global decision is made by counting the number of alarmed sensors collected from the sensor field, and the decision statistics will converge to a Gaussian random variable. When the measurements are not identically distributed across sensors, it is reasonable that the global detector should weight the decision of each sensor appropriately since sensors closer to the source produce more reliable decisions. But, what should be the optimal weighting when the local detection probability for each sensor is unknown beforehand? What are the factors affecting the weighting? Is there an assurance of asymptotic optimality?

Our approach is based on the Locally Asymptotically Normal (LAN) theory of Le Cam [1] [2]. (A brief summary of results of LAN related to our work is given in the Appendix.) Our goal is to find decision rules for sensors and the central unit that are asymptotically most powerful. Specifically, we find local and global decision rules that, for a given false alarm rate for the global decision, maximizes the the probability of detection as the number of sensors goes to infinity.

We model the POI over the region as a deterministic spatial signal with known shape but unknown signal strength. While the assumption of known signal shape is restrictive for it requires prior knowledge of POI, the model of unknown signal strength is almost necessary in practice because it is unreasonable to assume that POI can be calibrated. From a theoretical point of view, not knowing the signal strength makes the detection problem more difficult and also more interesting in the asymptotic regime. For example, the direct use of error exponent to characterize performance [30] [31] is no longer valid since the number of alternative hypothesis is uncountable. Indeed, had the signal strength is known, the error probability of any reasonable detector always decays to zero as the number of collected sensor detections increases.

We assume a Medium Access Control (MAC) protocol (such as ALOHA) is used to collect local decisions where each sensor has a probability p_m to transmit its decision successfully to the central unit. In order to exploit spatial correlation of the POI, it is necessary, as we will assume, that each sensor knows its location through the use of a certain geolocation device.

We make the assumption that randomly deployed sensors form a homogeneous spatial Poisson process with certain intensity. For the model of independent additive noise at each sensor, the marking by the local decision of each sensor is equivalent to a location-dependent thinning of the initial Poisson process; the alarmed sensors form a nonhomogeneous Poisson process. The process of retrieving sensor decisions from the sensor field is another thinning of the alarmed sensors. The Poisson assumption allows us to combine the two thinning processes at the MAC layer and the physical layer and model the alarmed sensors at the central unit as a nonhomogeneous Poisson field with an intensity as a function of the POI. Hence, the distributed detection problem is converted to the detection based on the intensity of the observed alarmed sensors.

To apply LAN theory of Le Cam, we derived (1) the sufficient conditions for the spatial signal shape that guarantees the existence of asymptotically locally most power (ALMP) detector; (2) the asymptotic local upper bound of power for any detector, and (3) an asymptotically locally jointly optimal rule over the fusion scheme and the single sensor threshold. For the special case that the power function of a single sensor is linear, the proposed detector is also asymptotically uniformly most powerful (AUMP).

Our numerical results are designed to answer a number of practical questions. Since the detector is based on asymptotic techniques, one questions what the size of the network is for which the asymptotic analysis is accurate. The simulations showed that the performance of the network of size 1000 matches well with the theoretic prediction. We showed that the ALMP detector proposed in this paper offers a significant gain over simplistic counting schemes. Since we assume the knowledge of signal waveform in the detector design, we considered the case of waveform mismatch in our simulation. The sensitivity of the mismatch of course depends on the specific shape of the signal waveform. For the class of symmetric exponentially decaying waveforms, we found that a simple step function approximation offers a graceful degradation. We also did not observe severe performance loss correspond to source location uncertainty.

C. Related Work

Distributed detection using multiple sensors and optimal fusion rules have been extensively investigated, see [14] [15] [16]. Many authors have derived optimal local detector and fusion rules under various situations based on different sets of assumptions, e.g., [17] [20] [19] [21]. For the fusion scheme, Chair and Varshney [18] showed that the optimal fusion rule is a likelihood ratio test on the decisions from sensors and becomes a threshold detector on the weighted sum of binary sensor decisions and the weight is obtained using the local detection and false alarm probability at each sensor under each hypothesis. However, the optimal criteria are obtained under the assumption that the hypotheses of the underlying phenomenon are simple, i.e., discrete and finite. These approaches require the knowledge of the detection probability as well as the false alarm probability at each sensor under each possible hypothesis, which is not proper for the detection of unknown signals that is the main interest in this paper.

The detection of unknown signal or signal of unknown amplitude was considered by several authors under the composite hypothesis formulation. The derivation of locally optimal detector for centralized scheme is known [23] [24]. Poor and Thomas considered the locally optimal detector for stochastic signals and compared the asymptotic relative performance between detectors using asymptotic relative efficiency (ARE) in the centralized detection scheme [25]. For the distributed or decentralized case, Aalo and Viswanathan considered the detection of an unknown signal via multilevel quantization and simple fusion rules [27]. However, no optimality for the fusion rule was considered. The works of Fedele, Izzo, and Paura [28] and Srinivasan [29] are perhaps the closest to our approach. In both cases, the authors considered the distributed scheme where multiple peripheral detectors or sensors are combined to a fusion center and the number of observations per each sensor goes to infinity. These assumptions are reasonable for the classical radar problem. For large scale microsensor networks, however, it is reasonable to assume that each sensor has only a few chances for observations and transmissions due to the limited power and to consider the asymptotic case where the number of sensors goes to infinity with a limited number of observations per each sensor as in this paper. Srinivasan derived the optimal local rule and fusion rule based on Bayes rule and summation over all realizations of sensor decisions

[29]. However, these approach is difficult to apply to get an explicit form of fusion rule and the local threshold beyond the case of several sensors with many observations per sensor.

The asymptotic case where the number of sensors goes to infinity was also considered by several authors from different perspectives. For example, the authors used the error exponent as the asymptotic performance measure to show the optimality of identical sensor for i.i.d. observations [30], [31]. In [32], the authors considered the optimality of identical binary sensors under the capacity limited reachback channel.

D. Notation

We denote a statistical model[1] by $(\Omega, \mathcal{X}, \mathcal{P})$ where Ω is the sample space and $\mathcal{X}$ is the σ-field defined on Ω. $\mathcal{P}$ is a class of parametric distributions $\{P_\theta, \theta \in \Theta\}$ defined on $\mathcal{X}$ for some parameter space Θ. The sequence of statistical models is denoted by $(\Omega^{(n)}, \mathcal{X}^{(n)}, \mathcal{P}^{(n)})$ where $\mathcal{P}^{(n)} = \{P_\theta^{(n)}, \theta \in \Theta\}$. Notice that the parameter space Θ doesn't change with the sequence index in our formulation and the superscript (n) does not denote the product space or measure in general. It can be an arbitrary sequence of measurable spaces and probability measures. For the product distribution of n i.i.d. P_θ, we use the notation $P_\theta^{\otimes n}$. For a sequence of random vectors $\mathbf{x}_n$ defined on $\Omega^{(n)}$, $\mathbb{E}_{n,\theta}\mathbf{x}_n$ is the statistical expectation of $\mathbf{x}_n$ under probability distribution $P_\theta^{(n)}$. The notation $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that $\mathbf{x}$ is Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The set of real numbers is denoted by $\mathbb{R}$. Vectors and matrices are written in boldface. Operation $(\cdot)^T$ indicates the matrix transpose.

II. System Model

We consider a large scale sensor network with identical binary sensors deployed over a wide area; we want to decide whether the POI has occurred in the area. We assume that each sensor makes a decision based on its own observation and that the local decisions are collected through a MAC at a central unit or fusion center where the global decision is made under a size (probability of false alarm) constraint. The POI is spatially varying, with a known shape function and an unknown magnitude. As an example, in the case of NBC activity, the signal strength is expected to be largest at the origin of the phenomenon,

[1] For more description, see Appendix I.

and decays away from the origin. We assume that the spatial signal is deterministic and denote the signal strength by

$$\gamma(\mathbf{x}) = \theta s(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ denotes the location, $\theta \in \Theta \triangleq [0, \infty)$ is an unknown amplitude, and $s(\mathbf{x})$ is a known[2] function which incorporates the information about the spatial variation of the underlying phenomenon.

A. Single Sensor

We assume that sensors make their local decisions independently without collaborating with other sensors. Since the exact value of the signal strength is unknown, we design each sensor to decide between the following hypotheses

$$\begin{aligned} H_0 :& \quad \gamma(\mathbf{x}) = 0 \quad \text{(null hypothesis)}, \\ H_1 :& \quad \gamma(\mathbf{x}) > 0 \quad \text{(alternative hypothesis)}, \end{aligned} \tag{2}$$

with local size constraint of α_0. Using the amplitude parameter θ, the hypotheses (2) is equivalently expressed by

$$\begin{aligned} H_0 :& \quad \theta = 0, \\ H_1 :& \quad \theta > 0. \end{aligned} \tag{3}$$

The local decision of sensor S_i located at $\mathbf{x}_i$ is denoted by

$$u_i = \begin{cases} 1 & \text{if } H_0 \text{ rejected}, \\ 0 & \text{otherwise}. \end{cases} \tag{4}$$

One possible sensor observation model is the additive Gaussian noise model as shown in Fig. 2, where the sensor observation Y_i is given by

$$Y_i = \gamma(\mathbf{x}_i) + N_i, \quad N_i \sim \mathcal{N}(0, \sigma_0), \tag{5}$$

where N_i is the sensor noise, assumed across sensors. For the additive Gaussian noise model, the local decision rule for (3) at each sensor is the uniformly most powerful (UMP) detector given by

$$Y_i \underset{H_0}{\overset{H_1}{\gtrless}} \tau_0, \tag{6}$$

[2]We do not assume that $s(\mathbf{x})$ is known *a priori* before sensor deployment or at the data retrieval period. See Section III-E for the estimation of $s(\mathbf{x})$ from collected binary sensor decisions. This includes the case where the shape function is parameterized, with unknown parameters that must be estimated.

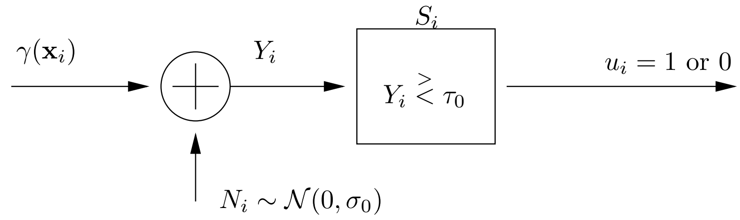

Fig. 2. Gaussian noise observation model at location $\mathbf{x}$

where $\tau_0 = \sigma_0 Q^{-1}(\alpha_0)$[3]. We define the following probability

$$p(\mathbf{x}_i) \triangleq \Pr\{u_i = 1\}. \tag{7}$$

That is, $p(\mathbf{x})$ is the probability that a sensor located at $\mathbf{x}$ declares its decision as H_1. The probability $p(\mathbf{x})$ is a function of the local threshold and the signal strength at $\mathbf{x}$ and is given by

$$p(\mathbf{x}) = \beta_{\tau_0}(\gamma(\mathbf{x})). \tag{8}$$

For the additive Gaussian observation model in (5), $\beta_{\tau_0}(\gamma(\mathbf{x}))$ is expressed as $Q\left(\frac{\tau_0 - \gamma(\mathbf{x})}{\sigma_0}\right)$.

B. Parametric Poisson Model

Consider that a large number of identical sensors designed in II-A are deployed randomly and uniformly over a region A as shown in Fig. 3 (a). We assume that the initial distribution of sensors over the region is a homogeneous spatial Poisson process with local intensity λ_h. This is a reasonable model when the random location of each sensor is uniformly distributed over A. After deployment, each sensor makes a local decision about the underlying phenomenon. Specifically, sensor S_i located at $\mathbf{x}_i$ makes a binary decision u_i based on its observation, encodes its decision, and then sends its packet over MAC to the central unit.

Since we assume that each sensor makes the decision by itself and the sensor noise is independent, the local decision u_i is independent conditioned on the signal strength $\gamma(\mathbf{x})$. By the Poisson assumption on the initial sensor locations, the marking by the local decision of each sensor is equivalent to a location-dependent thinning of the original sensor distribution with thinning probability $p(\mathbf{x})$. Hence, the distribution of the *alarmed* sensors, i.e., sensors with $u_i = 1$, forms a nonhomogeneous spatial Poisson process [10].

[3] $Q(x)$ denotes the tail probability $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}t^2} dt$.

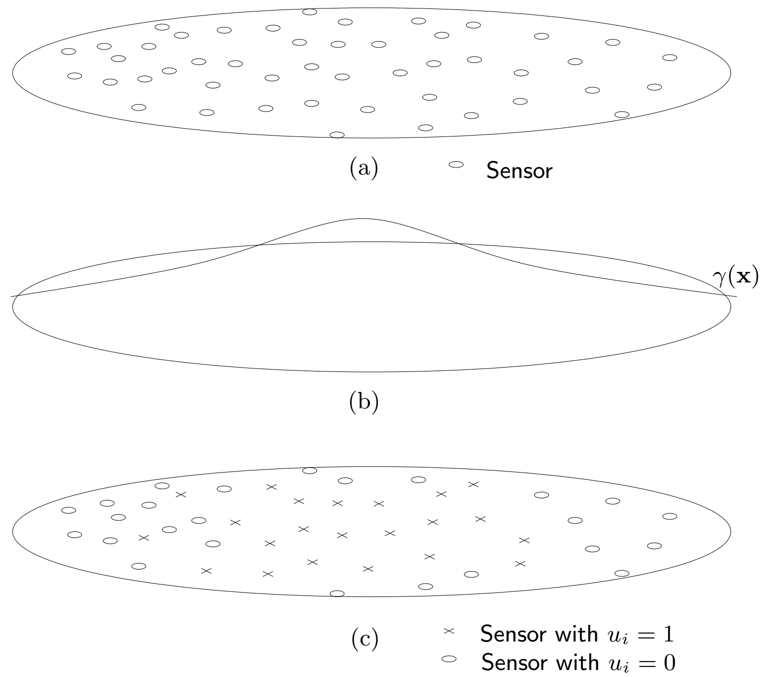

Fig. 3. (a) Initial sensor deployment over space (b) Signal strength of underlying phenomenon (c)Local decisions of sensors

During the data retrieval period, the local decisions of sensors are collected through a wireless channel. Sensor data can be lost during the transmission through MAC due to fading as well as collisions. We model this probabilistic loss as another thinning of the Poisson process of alarmed sensors. We assume that the probability that each sensor data is transmitted successfully through MAC is the same for all sensors, and denote it by p_m. Then, the second thinning is uniform over the region with probability p_m. So, the distribution of alarmed sensors at the fusion center or final data collector is a nonhomogeneous spatial Poisson process whose local intensity is given by

$$\begin{aligned} \lambda(\mathbf{x}) &= \lambda_h p_m p(\mathbf{x}) = \lambda_h p_m \beta_{\tau_0}(\theta s(\mathbf{x})), \\ &= \lambda_h p_m [\beta_{\tau_0}(0) + \beta'_{\tau_0}(0)\theta s(\mathbf{x}) + o(\theta)], \end{aligned} \tag{9}$$

where $\beta'_{\tau_0}(\gamma(\mathbf{x})) = \frac{\partial}{\partial \gamma(\mathbf{x})}\beta_{\tau_0}(\gamma(\mathbf{x}))$. When the function $\beta_{\tau_0}(\cdot)$ is linear or θ is in a small neighborhood of $\theta = 0$, the Poisson process of alarmed sensors is described by a nonhomogeneous intensity model parameterized by amplitude θ and is given by

$$\lambda(\theta, \mathbf{x}) = \theta \bar{f}(\mathbf{x}) + \bar{\lambda}, \quad \theta \in \Theta, \tag{10}$$

where

$$\bar{f}(\mathbf{x}) = \lambda_h p_m \beta'_{\tau_0}(0) s(\mathbf{x}), \quad \bar{\lambda} = \lambda_h p_m \beta_{\tau_0}(0), \tag{11}$$

for a given τ_0. Note that the intensity variation $\bar{f}(\mathbf{x})$ of alarmed sensors is a scaled version of the spatial signal shape $s(\mathbf{x})$.

The Poisson assumption on the initial sensor distribution converts the global detection of a spatially-varying signal to the detection of spatial Poisson processes based on the intensity of observed alarmed sensors; the spatial signal is mapped to the underlying intensity of the Poisson process of alarmed sensors. (This is what we mean by 'Poisson regime' in this paper.) As we will show later, the asymptotic detector requires knowledge of the signal $s(\mathbf{x})$, and hence of the location of the reporting sensors; however, the asymptotic detector will be shown to be surprisingly robust both with respect to the shape function $s(\mathbf{x})$ as well as its 'origin'. Also, the problem becomes simplified for incorporating data loss via the MAC.

C. Review of Poisson Process

The Poisson process X_A in a metric space A with a σ-field $\mathcal{B}$ is expressed in a simple manner by a counting measure notation as [11]

$$X_A(B) = \sum_{i:\ \mathbf{x}_i \in A} \epsilon_{\mathbf{x}_i}(B),\ \forall B \in \mathcal{B}, \tag{12}$$

where $\mathbf{x}_i$'s are random points in A and

$$\epsilon_{\mathbf{x}_i}(B) \triangleq \begin{cases} 1, & \mathbf{x}_i \in B \\ 0, & \mathbf{x}_i \notin B \end{cases}. \tag{13}$$

The Poisson process has the following properties.

(i) for every $B \in \mathcal{B}$, $X_A(B)$ is a Poisson random variable with mean $\Lambda(B) < \infty$,

(ii) for every finite disjoint sets $B_1, \ldots, B_k \in \mathcal{B}$, the random variables $X_A(B_1), \ldots, X_A(B_k)$ are independent.

Here, $\Lambda(\cdot)$ is called the intensity measure and its density[4] $\lambda(\mathbf{x})$, i.e., $\Lambda(d\mathbf{x}) = \lambda(\mathbf{x})d\mathbf{x}$, is called the (local) intensity. We define the stochastic integral for a given function g as [7]

$$I(g) \triangleq \int_A g(\mathbf{x}) X_A(dx) = \sum_{i:\ \mathbf{x}_i \in A} g(\mathbf{x}_i). \tag{14}$$

The probability distribution of Poisson process X_A is determined by the local intensity. For the case of parametric family of intensities like (10), the probability distribution is also parametrized with the same parameter and given by [12]

$$\begin{aligned} dP_\theta(X_A) &= \exp\left(\int_A \log \lambda(\theta, \mathbf{x}) X_A(d\mathbf{x}) - \int_A \lambda(\theta, \mathbf{x}) d\mathbf{x}\right), \\ &= \prod_{i:\mathbf{x}_i \in A} \lambda(\theta, \mathbf{x}_i) \exp\left(-\int_A \lambda(\theta, \mathbf{x}) d\mathbf{x}\right). \end{aligned} \tag{15}$$

The likelihood ratio between two distributions P_{θ_0} and P_{θ_1} is given by [7]

$$\begin{aligned} \frac{dP_{\theta_1}}{dP_{\theta_0}}(X_A) &= \exp(\int_A \log \frac{\lambda(\theta_1, \mathbf{x})}{\lambda(\theta_0, \mathbf{x})} X_A(d\mathbf{x}) \\ &\qquad - \int_A [\lambda(\theta_1, \mathbf{x}) - \lambda(\theta_0, \mathbf{x})] d\mathbf{x}). \end{aligned} \tag{16}$$

III. Detection of Spatially-varying Signal

In Section II-B, we assumed that the initial sensor distribution is Poisson, and showed that the original detection problem using identical binary sensors is converted to the problem (10, 11, 3) of detecting Poisson processes with different intensities.

Under the asymptotic local optimality, we focus on the detection of the alternative which converges to the null $\theta = 0$ where the distributions of the null and alternative hypothesis are asymptotically nonseparable. The existence of asymptotic locally optimal detection requires some conditions on the underlying statistical models. Le Cam's theory provides an analytic framework for such detection problems and gives an asymptotic optimal criterion. When a statistical model satisfies the locally asymptotically normal (LAN) condition, (See Appendix I.) we can construct an asymptotic local upper bound for the power for any sequence of detectors with a given asymptotic size, *and* we can construct a sequence of detectors that achieves this bound.

[4]For convenience, we assume that $\Lambda(\cdot)$ is differentiable.

In this section, we construct a sequence of statistical models for the Poisson process of the alarmed sensors and show that it satisfies the LAN condition. We also derive the asymptotic local upper bound for the power of any detector with a given size and the asymptotically locally most powerful (ALMP) detector for the model (10, 11, 3) under the Poisson regime.

A. Construction of Sequence of Statistical Models

We construct a sequence of statistical models $(\Omega^{(n)}, \mathcal{X}^{(n)}, \mathcal{P}^{(n)})$ where $(\Omega^{(n)}, \mathcal{X}^{(n)})$ is the measurable space of all possible realizations of the Poisson process $X_{A_n}^{(n)}$ of the alarmed sensors on space A_n and $\mathcal{P}^{(n)}$ is the corresponding family of probability distributions. Let $\lambda^{(n)}(\theta, \mathbf{x})$ be the local intensity of the Poisson process $X_{A_n}^{(n)}$. Then, the family of probability distributions $\mathcal{P}^{(n)} = \{P_\theta^{(n)}(X_{A_n}^{(n)}), \theta \in \Theta\}$ is given by (15). We are interested in the asymptotic scenario where the number of sensors deployed over a finite area goes to infinity. The model of increasing sensors in a finite area is described by increasing the initial intensity λ_h of sensor deployment.

Model 1 (Finite area and infinite sensor model) We initially distribute sensors over space A_n according to a homogeneous Poisson process with intensity $\lambda_h^{(n)}$ independently for each $n \geq 1$. We set

$$A_n = A, \text{ such that } |A| < \infty, \quad \text{for all } n = 1, 2, \ldots, \tag{17}$$

and choose the local intensity of initial sensor distribution over the space A as

$$\lambda_h^{(n)} = n\lambda_{h0}. \tag{18}$$

Then, we collect the local decision of each sensor and observe the realization $X_A^{(n)}$ of the alarmed sensor distribution. For each $n \geq 1$, the local intensity of $X_A^{(n)}$ is given, using (10, 11), by

$$\lambda^{(n)}(\theta, \mathbf{x}) = \theta n f(\mathbf{x}) + n\lambda_0, \quad \theta \in \Theta \tag{19}$$

where

$$f(\mathbf{x}) = \lambda_{h0} p_m \beta'_{\tau_0}(0) s(\mathbf{x}), \quad \lambda_0 = \lambda_{h0} p_m \beta_{\tau_0}(0), \tag{20}$$

and the sequence of probability distributions $\{P_\theta^{(n)}, \theta \in \Theta\}$ is given, using (15), by

$$dP_\theta^{(n)}(X_A^{(n)}) = \exp\left(\int_A \log \lambda^{(n)}(\theta, \mathbf{x}) X_A^{(n)}(d\mathbf{x}) - \int_A \lambda^{(n)}(\theta, \mathbf{x}) d\mathbf{x}\right). \tag{21}$$

The LAN property of distributions of Poisson processes has been investigated by several authors [6], [7], [8], who derived the conditions in terms of the local intensity for LAN. However, the authors considered a sequence of models where the observation area $A_n(\subset A_{n+1})$ goes to infinity for a fixed local intensity for all n, which is different from Model 1. We derive new conditions in terms of the spatial signal shape for the LAN property of the model with increasing sensors in a finite area.

Theorem 1: For Model 1, suppose that $f(\mathbf{x})$ satisfies the following conditions

(C.1) $f(\mathbf{x}) \geq 0, \quad \mathbf{x} \in A.$

(C.2) $\sup_{\mathbf{x}\in A} f(\mathbf{x}) < \infty.$

(C.3) $\int_A f(\mathbf{x}) d\mathbf{x} > 0.$

Then, the statistical model $\{P_\theta^{(n)}, \theta \in \Theta\}$ of the alarmed sensor distribution $X_A^{(n)}$ is LAN at $\theta = 0$; i.e., for every $h \geq 0$,

$$\log \frac{dP_{r_n(0)h}^{(n)}}{dP_0^{(n)}}(X_A^{(n)}) = h\Delta_{n,0} - \frac{1}{2}h^2 + o_{P_0^{(n)}}(1), \tag{22}$$

where the central sequence $\Delta_{n,0}$ and normalizing sequence $r_n(0)$ are given by

$$\Delta_{n,0} = \int_{A_n} r_n(0) \left(\frac{\dot{\lambda}^{(n)}(0,\mathbf{x})}{\lambda^{(n)}(0,\mathbf{x})} \right) \left[X_A^{(n)}(d\mathbf{x}) - \Lambda_0^{(n)}(d\mathbf{x}) \right], \tag{23}$$

$$r_n(0) = J_n(0)^{-1/2}, \quad J_n(0) = \int_A \left(\frac{\dot{\lambda}^{(n)}(0,\mathbf{x})}{\lambda^{(n)}(0,\mathbf{x})} \right)^2 \Lambda_0^{(n)}(d\mathbf{x}), \tag{24}$$

$$\dot{\lambda}^{(n)}(\theta,\mathbf{x}) = \frac{\partial}{\partial\theta}\lambda^{(n)}(\theta,\mathbf{x}), \quad \Lambda_0^{(n)}(d\mathbf{x}) = \lambda^{(n)}(0,\mathbf{x})d\mathbf{x} = n\lambda_{h0}d\mathbf{x},$$

and $\mathcal{L}(\Delta_{n,0}|P_0^{(n)}) \Rightarrow \mathcal{N}(0,1)$.

Proof: See Appendix II.

Here, the integration with random point measure in (23) is defined in (14). Condition *(C.1)* requires that the single sensor power function $\beta_{\tau_0}(\cdot)$ must be a nondecreasing function at the origin $\theta = 0$ for a given τ_0 and $s(\mathbf{x})$ is nonnegative; *(C.2)* is satisfied by any bounded $s(\mathbf{x})$, and *(C.3)* says that $s(\mathbf{x})$ is not identically zero over the sensor field. The conditions *(C.1)*-*(C.3)* are general enough to include most interesting cases. Examples of allowed

2-D signal shapes for any region with a finite area include: constant $f(\mathbf{x})$ or $s(\mathbf{x})$, a step function, Gaussian, or exponentially decaying as

$$s(x, y) = e^{-\eta r}, \quad r = \sqrt{x^2 + y^2}, \tag{25}$$

with some $\eta > 0$; indeed, any bounded non-negative function.

B. ALMP detector under the Poisson regime

In the previous section, we showed that the Poisson process of alarmed sensors, with increasing initial sensor density, admits the LAN property for general spatial signal shapes. Once the LAN property is verified, the construction of an ALMP detector is straightforward using the central sequence $\Delta_{n,0}$.(See Appendix I.)

Theorem 2: For model 1, let the conditions (C.1)-(C.3) be satisfied. Then, an asymptotic upper bound of power for any sequence of detectors ϕ_n with size α, i.e., $\limsup_{n\to\infty} \mathbb{E}_{n,0}\phi_n \leq \alpha$, is given by

$$\limsup_{n\to\infty} \sup_{0<r_n(0)^{-1}\theta\leq M} \left[\mathbb{E}_{n,\theta}\phi_n - Q(Q^{-1}(\alpha) - r_n^{-1}(0)\theta)\right] \leq 0, \tag{26}$$

where

$$r_n(0) = \left(n\lambda_{h0}p_m \frac{(\beta'_{\tau_0}(0))^2}{\beta_{\tau_0}(0)} \int_A s^2(\mathbf{x})d\mathbf{x}\right)^{-1/2}. \tag{27}$$

Furthermore, the following sequence of (nonrandomized) detectors is asymptotically locally most powerful (ALMP) with size α for (10, 11, 3).

$$\phi_{n,opt} = \begin{cases} \text{Decide } H_0 & \text{if} \quad \Delta_{n,0} \leq Q^{-1}(\alpha), \\ \text{Decide } H_1 & \text{if} \quad \Delta_{n,0} > Q^{-1}(\alpha), \end{cases} \tag{28}$$

where

$$\begin{aligned} \Delta_{n,0} \quad &= \quad n^{-1/2}\lambda_0^{-1/2}\left(\int_A s^2(\mathbf{x})d\mathbf{x}\right)^{-1/2} \\ &\qquad \left(\sum_{\mathbf{x}_i\in A} s(\mathbf{x}_i) - n\lambda_0\int_A s(\mathbf{x})d\mathbf{x}\right), \end{aligned} \tag{29}$$

where $\mathbf{x}_i$'s are the random locations of alarmed sensors in A.

Proof: See Appendix II.

(26, 27) reveals how the various factors such as sensor density, probability of successful transmission through MAC, the spatial signal shape, and the single sensor threshold affect the asymptotic global power. Note that as expected, the power of the detector increases monotonically with sensor density, signal strength, and MAC transmission success rate. We observe that if the signal strength is halved, sensor density must be quadrupled in order to maintain the asymptotic performance. This is consistent with the notion of fusing independent signal decisions. Since the ALMP test statistic, the central sequence $\Delta_{n,0}$, has a limit distribution $\mathcal{N}(0,1)$ under the null distribution $P_0^{(n)}$ by the LAN condition, it is easy to see that the detector (28) has an asymptotic size of α. Notice that the ALMP test statistic consists of a weighted sum of binary sensor decisions. Theorem 2 describes how to optimally use the knowledge of the signal shape and the locations of the sensors. Each sensor decision is to be weighted and summed for global decision and the optimal weight is $s(\mathbf{x})$, the shape of underlying spatial signal $\gamma(\mathbf{x})$. In other words, the confidence for each sensor decision is proportional to the strength of the signal at the sensor location. This can be considered as a matched filtering in the spatial domain even though it is different from the conventional matched filtering since the received signal is 'sampled' at random points, with an input signal as the intensity function rather than the distorted version of the transmitted signal. The use of intensity function in the detection of Poisson processes has been investigated. In [13], the author considered a binary on-off detection problem in optical transmissions. The author assumed that the photon generation epochs were Poisson process and showed that the optimal weight is the intensity of input light under a Bayesian formulation of two simple hypotheses. However, the exact knowledge of the intensity of light is required rather than just the relative shape. For the proposed method, the optimal test can be implemented without obtaining the exact value of $\gamma(\mathbf{x})$ since the optimal weight requires only the local intensity variation $s(\mathbf{x})$ of alarmed sensors and any scaling of $s(\mathbf{x})$ is irrelevant in forming the statistic (29).

An intuitive interpretation is given by a step function which is given by

$$s(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in A_1 (\in A), \\ 0 & \mathbf{x} \in A \backslash A_1, \end{cases} \tag{30}$$

where $A\backslash A_1$ is the difference set. In this case, only the local decisions from the sensors located at the region A_1 where the phenomenon would occur are counted discarding the false alarms from the regions of no event. For more complicated signal shapes such as (25), the local decisions from sensors are weighted according to the relative strength of the underlying signal.

C. Optimization of Threshold for a Single Sensor

In Section III-B, we derived the ALMP detector for the global detection of a spatially-varying signal with an unknown amplitude under the Poisson regime assuming that identical binary sensors are deployed and that the signal shape, but not strength, is known. The conversion into the Poisson regime gives another simplification; The optimal local threshold for a single sensor described in Section II-A is obtained through the asymptotic (local) upper bound (26, 27). Since the bound is a function of the local threshold and achieved by the proposed ALMP detector with a reasonable number of sensors as shown in Section IV, the optimal threshold for a single sensor is the one that maximizes the asymptotic upper bound.

Theorem 3: Suppose that the power function $\beta_{\tau_0}(\gamma(\mathbf{x}))$ for a single sensor is continuous and piecewise differentiable in the second variable. Under the Poisson regime, the following threshold for a single sensor maximizes the global power for a fixed and sufficiently large number of sensors in the region.

$$\tau_{opt} = \arg\max_{\tau_0} \frac{\beta'_{\tau_0}(0)^2}{\beta_{\tau_0}(0)}, \tag{31}$$

where $\beta'_{\tau_0}(\gamma(\mathbf{x})) = \frac{\partial}{\partial\gamma(\mathbf{x})}\beta_{\tau_0}(\gamma(\mathbf{x}))$.

Proof: See Appendix II.

For the Gaussian noise model, we have $\beta_{\tau_0}(\gamma(\mathbf{x})) = Q(\frac{\tau_0-\gamma(\mathbf{x})}{\sigma_0})$, $\beta_{\tau_0}(0) = Q(\tau_0/\sigma_0)$ and $\beta'_{\tau_0}(0) = \frac{1}{\sqrt{2\pi}\sigma_0}\exp(-\frac{1}{2}\left(\frac{\tau_0}{\sigma_0}\right)^2)$. The optimal local size is $\alpha_0 = 0.27$, and the corresponding local threshold $\tau_o = 0.612\sigma_o$ (we verify this via simulations in IV). This threshold surprisingly coincides with the one that the authors obtained for nonparametric detection of symmetric distribution using *i.i.d.* observations [26]. This implies that we should design the single sensor assuming the signal shape is uniform over the area if the information of the signal shape is not available before sensor deployment and identical sensors are to be deployed over the field. (This is the case that we consider in Section II-B.) Notice that under the assumption of binary decisions and Poisson distributed sensors, the individual sensors need not be very good; a design with a PFA of 0.27 is optimal! We also note that for the AWGN model, $\frac{\beta'_{\tau_0}(0)^2}{\beta_{\tau_0}(0)}$ is fairly flat around its maximum so that it is not critical to use the optimal τ_o and α_o values.

Notice that the optimal fusion rule (29) and the local threshold (31) do not depend on the parameter θ. Hence, the optimal rule is actually an asymptotically uniformly most powerful detector when the model (10) is true, for example, when the power function for a single sensor is linear, or the signal is weak. However, in general, our conversion to the Poisson regime is only valid in the local neighborhood of $\theta = 0$ for a typical power function $\beta_{\tau_0}(\cdot)$.

D. Independent and Identically Distributed Observations

If the signal is constant,

$$s(\mathbf{x}) \equiv 1, \quad \mathbf{x} \in A, \tag{32}$$

then the sensor observations, for the model described in Section II, will be i.i.d. Optimality of the counting-based detector is given by the following corollary of Theorem 2.

Corollary 1: For i.i.d. sensor observations over A, the counting-based detector is ALMP with size α under the Poisson regime.

Proof: In this case, the central sequence is given by (see eq. 29)

$$\Delta_{n,0} = (n\lambda_0|A|)^{-1/2}(N^{(n)}(A) - n\lambda_0|A|), \tag{33}$$

where $N^{(n)}(A)$ is the number of alarmed sensors in space A and $|A|$ is the area of A. ∎

Note that $N^{(n)}(A)$ is a Poisson random variable with mean $n\lambda_0|A|$ under the null hypothesis. Since the mean and variance are equal for Poisson random variables, $\Delta_{n,0}$ is centered and normalized to have variance one. The Gaussian limit distribution of $\Delta_{n,0}$ under the null hypothesis is explained as follows. Suppose that we partition A into n subregions with an equal area for the nth experiment. Under the null hypothesis, we have $\lambda^{(n)}(0, \mathbf{x}) = n\lambda_0$ from (19). So, the number of alarmed sensors in each subregion is

Poisson random variable with mean $\lambda_0|A|$ regardless of n and is i.i.d. over subregions by the Poisson assumption. Since $N^{(n)}(A)$ is the sum of number of points in each subregion, it is a sum of n i.i.d. random variables and $\Delta_{n,0}$ converges in distribution to $\mathcal{N}(0,1)$ as n goes to infinity by the classical central limit theorem (CLT).

A different counting-based detector can be constructed for the i.i.d. case based on the Binomial distribution [22]. Let the number of sensors in A be K. Assuming all sensor observations are i.i.d., a detector of size α based on counting is given by

$$\text{Reject } H_0 \text{ if } T_K \triangleq u_1 + u_2 + \ldots + u_K > \tau_\alpha. \tag{34}$$

Under H_0, $\gamma(\mathbf{x}) = 0$ for all $\mathbf{x}$ and

$$u_i \overset{i.i.d.}{\sim} B(1, p_0), \quad p_0 = Pr\{u_i = 1 | \gamma(\mathbf{x}) = 0\} = \beta_{\tau_0}(0). \tag{35}$$

Recall that the u_i are the binary sensor decisions. Using the CLT, the asymptotic distribution of T_K is given by

$$\frac{1}{\sqrt{Kp_0(1-p_0)}} \left(\sum_{i=1}^{K} u_i - Kp_0 \right) \overset{D}{\Rightarrow} \mathcal{N}(0,1), \tag{36}$$

as the number of sensors K goes to infinity. Hence, the detector (34) is explicitly given by

$$\text{Reject } H_0 \text{ if } \sum_{i=1}^{K} u_i > Kp_0 + Q^{-1}(\alpha)\sqrt{Kp_0(1-p_0)}. \tag{37}$$

The distributions for the number of alarmed sensors in (33) and (36) have different variances for the same mean, i.e., $n\lambda_0|A| = Kp_0$, under two different models. It is well known that the binomial distribution converges to Poisson distribution with mean Λ when K goes to infinity with constraint $Kp_0 = \Lambda$. So, when p_0 is small, two distributions are almost equivalent.

E. Discussion

The construction of ALMP test statistic $\Delta_{n,0}$ in (29) requires the knowledge of several values such as the null intensity λ_0, the sensor locations, and the shape of the underlying spatial signal. But the null intensity can be obtained by the known or controllable parameters such as the density of the initial sensor distribution, the local false alarm probability

α_0, and the probability of successful transmission through MAC, p_m. In this section, we briefly discuss the estimation of $s(\mathbf{x})$.

As shown in Theorem 2, the ALMP weight is the spatial signal shape $s(\mathbf{x})$ under the Poisson regime. We consider an estimation method based on the collected sensor data. One simple way is to utilize the Poisson assumption itself. (10) reveals that the $f(\mathbf{x})$ is the local intensity variation of alarmed sensor distribution over space. Hence, the weight can be estimated from the alarmed sensors and their location directly. For example, we can use a nonparametric intensity model

$$f(\mathbf{x}) = \sum_j f_j I_{A_j}, \quad \bigcup_j A_j = A. \tag{38}$$

Assuming that λ_0 is known, the maximum likelihood estimator of $\theta f(\mathbf{x})$ for the model (38) is given by

$$\widehat{\theta f(\mathbf{x})} = \sum_j \frac{N(A_j)}{|A_j|} I_{A_j} - \lambda_0. \tag{39}$$

Since any scaling of $s(\mathbf{x})$ doesn't matter for obtaining $\Delta_{n,0}$, $\widehat{\theta f(\mathbf{x})}$ can be used as an estimate for the optimal weight function. However, for this estimation method, several independent measurements by sensors are required for the purposes of estimating weight and providing test statistics and also this method is useful only when θ is fairly large. We are currently investigating more efficient methods to estimate the signal variation from the sensor decisions. The performance degradation due to the mismatched signal shape is investigated in Section IV-B.

IV. Numerical Results

In this section, we present some simulation results. We used the receiver operating characteristics (ROC) as the performance criterion. The power of the proposed ALMP detector was evaluated by the analytic bound (26, 27) and also using Monte Carlo runs. The false alarm probability was also simulated to check the validity of detector design in the Neyman-Pearson context. The power of the proposed detector (Theorem 2) was compared with that of the counting-based detector (37) which has also an asymptotic size of α and neglects the spatial information. The performance degradation by mismatched signal shapes was also investigated.

A. Setup

We considered a two dimensional space A which is circular with radius one. The spatial signal shape was the symmetric exponential in (25) with different decay rates. The average number of sensors in A was chosen to be 1,000. For the local sensor function, we used the additive Gaussian noise model and the UMP detector with several local sizes described in Section II-A.

For the simulation of power and false alarm probability, 10,000 Monte Carlos runs were executed. For each run, the following procedures were performed. The locations of sensors were randomly generated according to a homogeneous Poisson process with the given mean intensity [9]. The local threshold for a single sensor was calculated from the local size constraint and set to be the same for all sensors, $\sigma_0 Q^{-1}(\alpha_0)$. Zero-mean Gaussian noise with variance $\sigma_0^2 = 1$ was generated independently for each sensor and added to the signal strength calculated from the sensor location and the amplitude parameter θ. Threshold detection was made based on the sum of the signal and noise for the local decision. Finally, the global decision was made based on the test statistic $\Delta_{n,0}$ and the number of alarmed sensors for the ALMP detector and the counting-based detector, respectively. The global thresholds for both detectors were determined via the Gaussian limit distribution. Throughout the simulations, the probability of successful data collection from each sensor was set to $p_m = 0.9$. The initial homogeneous density λ_h and the local false alarm probability were assumed to be known and the true values were used for the simulation.

B. Receiver Operating Characteristics

Fig. 4, 5 show the analytic results. For both plots, the decay rate η for the exponential signal was 3. Fig. 4 shows the analytic upperbound for power for global size 0.1 with respect to the number of sensors for fixed signal amplitudes. It also depicts the expected behavior, that if the signal strength θ decreases, we need more sensors to achieve the same performance. Fig. 5 shows the analytic bound and simulated power with respect to the false alarm probability. For the simulation curve, the actual false alarm probability was used rather than the designed size. As shown in the figure, the power of ALMP detector achieves the asymptotic upper bound with the network size of 1000.

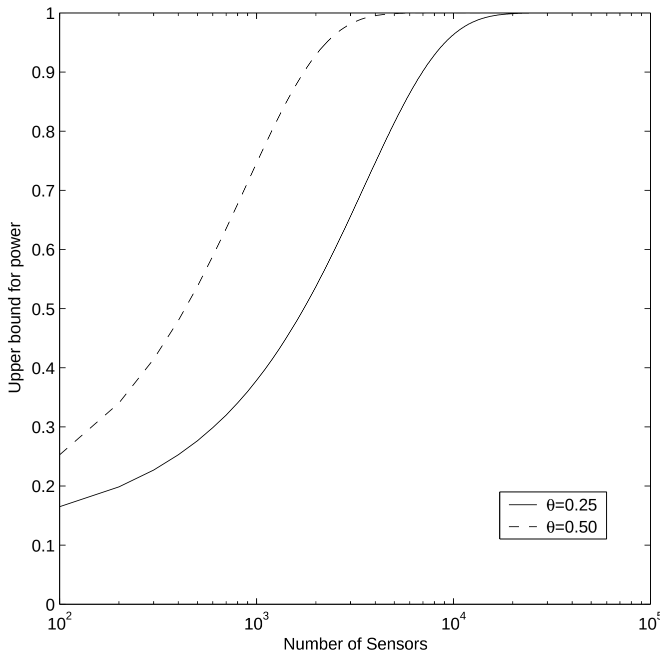

Fig. 4. Upperbound for power ($\alpha_0 = 0.1$(single sensor)$, P_F = 0.1$)

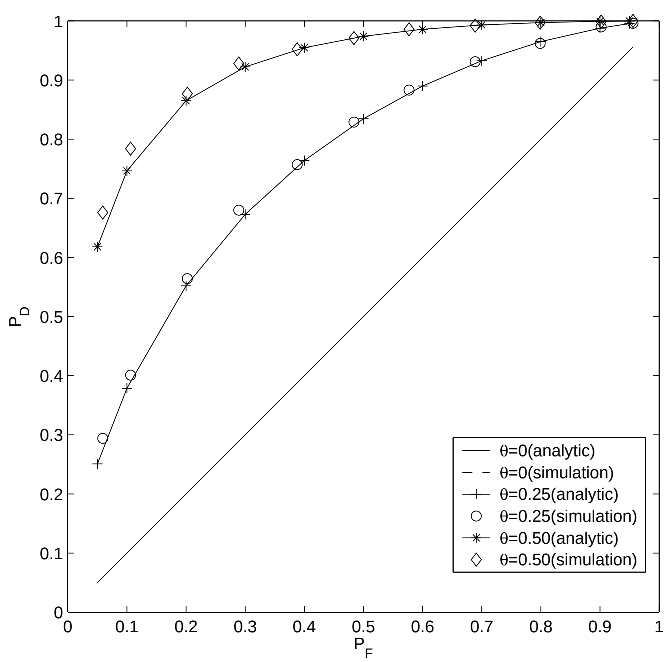

Fig. 5. ROC - analytic versus simulation ($\alpha_0 = 0.1$)

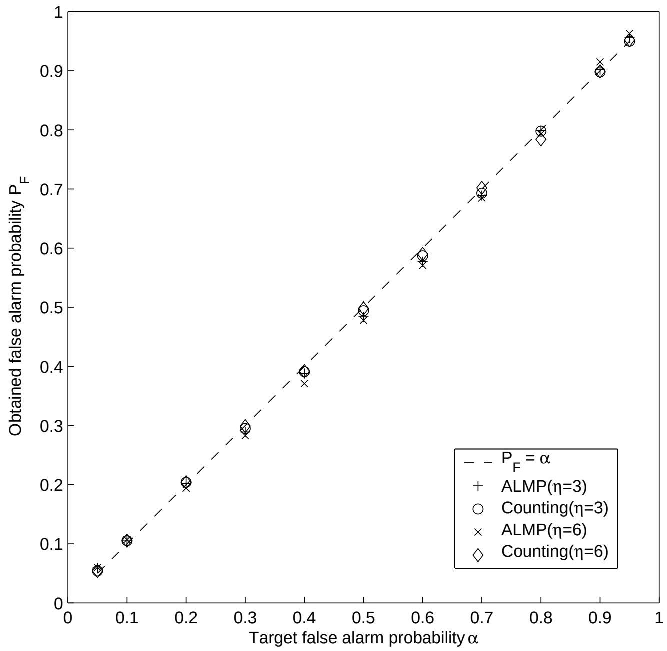

Fig. 6. Actual false alarm probability with respect to the designed size.

One important feature of the proposed detector is that $\Delta_{n,0}$ has a Gaussian limit distribution and the global threshold is based on it. Fig. 6 shows the actual false alarm probability obtained by simulation versus the designed size for the additive Gaussian noise model with local size 0.1 and verifies the convergence in distribution of the test statistic with a network size of 1000 sensors. As shown, the actual false alarm probabilities closely follow the designed size. Notice that the actual false alarm probability of the ALMP detector for the decay rate $\eta = 6$ deviates more than that for the case of $\eta = 3$ whereas the deviations are almost the same for the counting-based detector. This is because the test statistic $\Delta_{n,0}$ for the ALMP detector is the sum of local decisions weighted by the spatial signal shape while the weight is uniform over the space for the counting-based detector. For the exponential with a large decay rate, the sensor decisions around the origin dominate the overall distribution. Hence, the distribution is concentrated around the mean and deviates more from the Gaussian, which is clearly seen by the maximum deviation around $\alpha = 0.5$. This effect is more severe with a larger decay rate such as $\eta = 6$.

Fig. 7 shows the ROC of the ALMP detector using different local thresholds for a single sensor. The additive Gaussian input model was used and the average number of sensors in A was kept the same at 1000 for all four cases in the figure. As shown, the global power changes with the local size of each sensor and the maximum is achieved between the local sizes of 0.2 and 0.3 as predicted in Section III-C

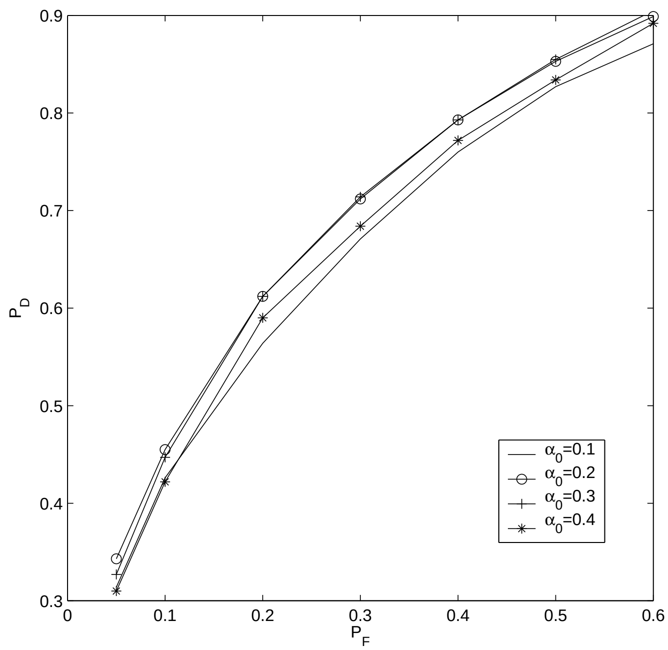

Fig. 7. ROC with different local sizes($\theta = 0.5$)

Fig. 8, 9 show the ROC's of the proposed ALMP and counting-based detector for the additive Gaussian sensor model with local size 0.1. Fig. 8 shows the case where the spatial signal changes quickly within the region A. The ALMP detector utilizing the spatial information drastically improves the detector performance over the counting-based approach. Fig. 9 show the ROC's for a different decay rate with the same setup as in Fig. 8. As the decay rate becomes small or the signal becomes more uniform over the space, the performance gained by utilizing the spatial information becomes less significant. The meaning of asymptotic local optimality is clearly evident from Fig. 9. In this case, we have a larger overall power in the space than the cases in Fig.8, since the signal decays slowly over the space with the same peak at the origin. Since the peak value of $s(\mathbf{x})$ in (25) is one and the variance for sensor input noise is chosen to be one, the maximum signal-to-noise ratio (SNR) for a sensor located at the origin is 0 dB when the amplitude parameter θ is one in the figure. Even though the SNR of 0 dB is very small for a single sensor, we

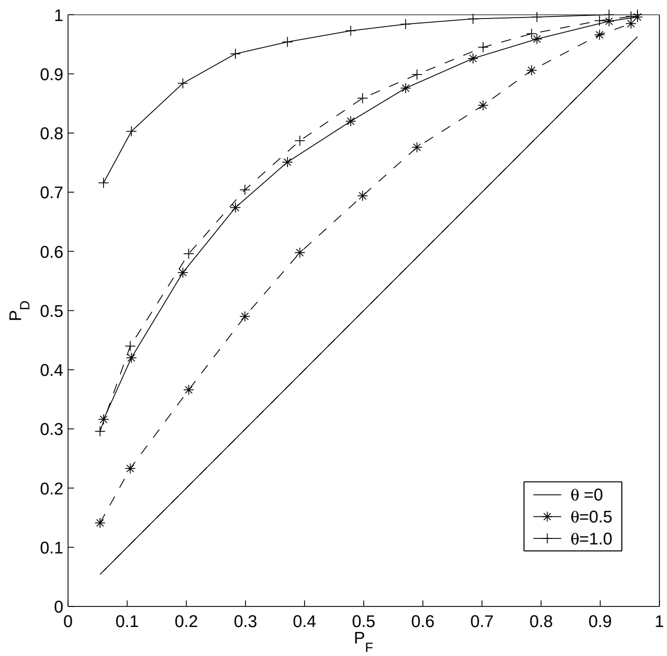

Fig. 8. ROC - additive Gaussian sensor model and $\eta = 6$ (solid line: ALMP detector, dashed line: counting-based detector).

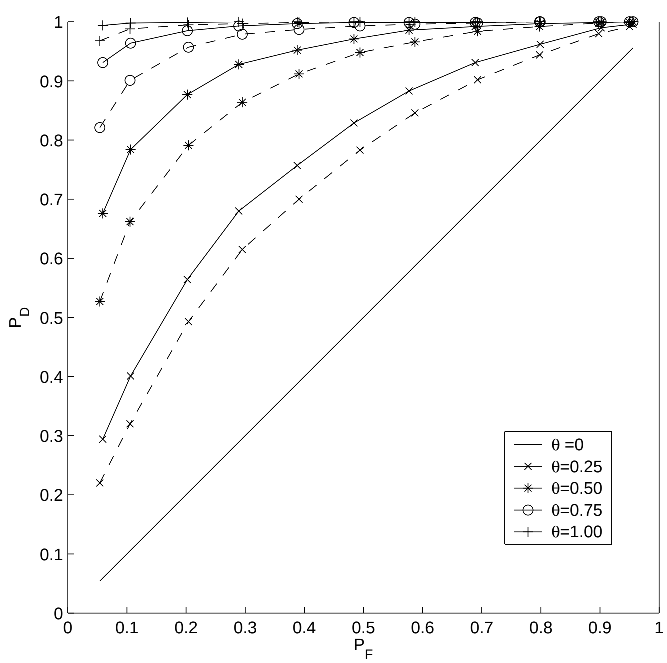

Fig. 9. ROC - additive Gaussian sensor model and $\eta = 3$ (solid line: ALMP detector, dashed line: counting-based detector)

have a large number of observations for the entire sensor field (the average number is 1000 sensors). Hence, the global power for $\theta = 1$ already reaches almost unity for both the ALMP and counting-based detector and the comparison above $\theta = 1$ is less meaningful in this case. However, the performance within the local neighborhood of the null hypothesis is clearly distinguishable in all the figures.

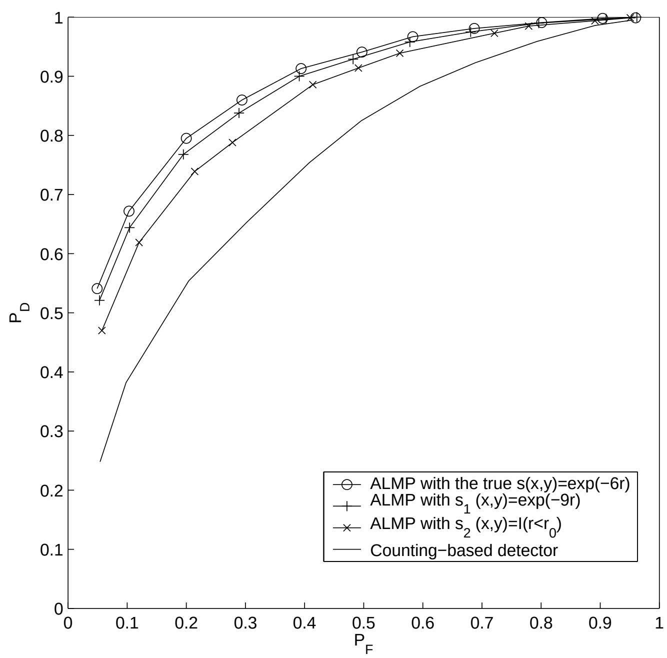

Fig. 10. ROC with mismatched signal shapes ($\alpha_0 = 0.27,\ \theta = 0.75$)

Up to now, the true signal shape was used to obtain the ROC of ALMP detectors. Fig. 10 shows the ROC of the proposed detector with mismatched signal shapes. The true signal shape of POI was the symmetric exponential with $\eta = 6$. We used two mismatched shapes as the weighting function to construct the central sequence. First, we considered the symmetric exponential $s_1(x, y)$ with a different decay rate $\eta = 9$. As expected, the proposed detector with the mismatched shape performs worse than the true ALMP detector. However, for the case of $s_1(x, y)$, the performance almost approaches the ALMP detector since $s_1(x, y)$ is quite similar to the original shape. So, we further approximated the signal shape by a step function

$$s_2(x,y) = \begin{cases} 1 & \sqrt{x^2+y^2} \le r_0 \\ 0 & \text{otherwise} \end{cases}, \tag{40}$$

where r_0 was determined such that the spatial 'power' of $s_2(x,y)$ covers 90 % of that of the original signal, i.e., $\int_A s_2^2(x,y)dxdy = 0.9\int_A s^2(x,y)dxdy$. In this case, even though the degradation from the true ALMP becomes larger, it still shows good performance compared with the true ALMP detector. It is shown that the rough knowledge of the signal shape is enough to get most of the advantage of the ALMP detector for the spatially-varying signal. Fig. 11 shows the ROC of the ALMP detector using the same signal shape

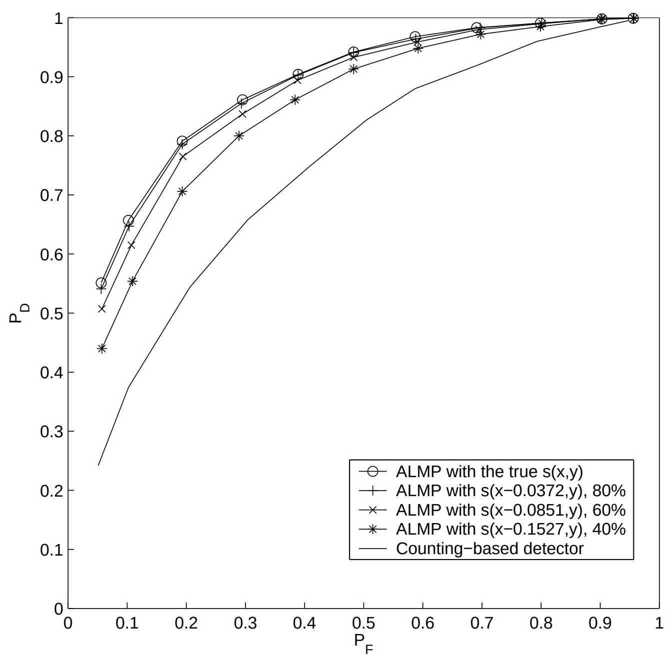

Fig. 11. ROC with mismatched centers ($\alpha_0 = 0.27,\ \theta = 0.75$)

with mismatched centers. The same parameters were used as Fig.10 for the true signal. For the mismatched signal, we used the symmetric exponential with the same decay rate but different centers. The displacements of center correspond to the positions of $80, 60, 40\%$ from the amplitude of the true center. As shown, even with a rough estimation of the center, the performance degradation is not severe compared with the perfectly matched case.

Finally, the effect of inaccurate sensor locations was investigated. Fig. 12 shows the performance of the perturbed sensor locations. The same signal shape with $\eta = 6$ was used for all ALMP detectors but the perturbed location of sensor were used at the fusion, i.e., $\hat{s}(\mathbf{x}) = s(\mathbf{x} + \Delta\mathbf{x})$ where the perturbation $\Delta\mathbf{x}$ was generated independently for each sensor with Gaussian distribution of standard deviations 1 %, 5 %, 10 % with respect to the radius of the total space. As shown, the ALMP detector is robust with respect to the sensor location error and a rough estimation of sensor location after deployment is enough.

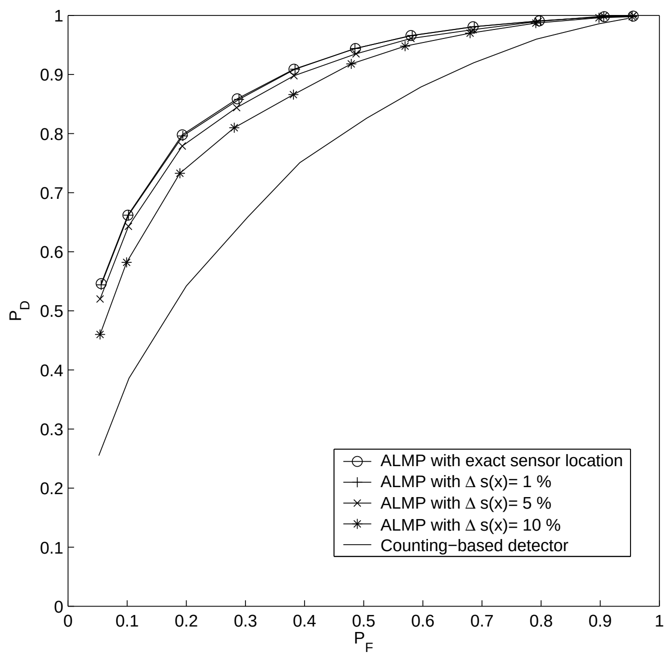

Fig. 12. ROC with inaccurate sensor locations ($\alpha_0 = 0.27,\ \theta = 0.75$)

V. Conclusion

We considered a global detection problem based on (inaccurate) binary decisions from local threshold sensors. The local 'alarm' probability is described by a generic function $\beta_{\tau_0}(\theta s(\mathbf{x}))$, where τ_0 is the local threshold, $s(\mathbf{x})$ is the known underlying signal shape, and θ is the unknown amplitude. By assuming a homogeneous Poisson distribution of the sensors, we mapped the global detection problem to one of detecting Poisson process with different intensity. Under a small signal strength assumption, asymptotically (in the number of sensors in a fixed area) locally most powerful (AMLP) detector was derived using Locally Asymptotic Normal (LAN) theory. It was shown that the conditions for applying LAN theory lead to reasonable assumptions on the underlying spatial signal. The AMLP fusion rule is a threshold detector for the weighted sum of local decisions, where the weight is proportional to the signal strength. The AMLP detector requires knowledge of the sensor locations, the signal shape, and a parameter which is the product of the local detector function $\beta_{\tau_0}(0)$, the MAC success probability, and the average density of the sensor locations. We have also shown (Theorem 3) how to optimize the local threshold so as to maximize the global power function. The asymptotic local optimality of the counting-based detector was established for the case of constant spatial signals. Numerical examples were provided to verify the theoretical results. Several of these examples also demonstrate that the proposed detector is robust under conditions of signal mismatch, including wrong signal shape, translated signal shape, and location calibration errors. Efficient estimators of $s(\mathbf{x})$, or its parameters, based on the binary sensor decisions, is currently under investigation.

Appendix I: Locally Asymptotically Normal (LAN) Theory

The theory of LAN was first proposed by Le Cam [1] [2]. In this section, we briefly introduce the LAN theory. The LAN theory provides simplifications of asymptotic statistics via the existence of a randomized statistic in a limit Gaussian model using convergence in distribution of loglikelihood ratios. It makes it possible to establish a minimax bound for the risk of arbitrary estimators and to construct the asymptotically locally most powerful (ALMP) detector for composite hypothesis test which we use in this paper [1] [3] [4].

A. Sequence of Statistical Experiments and LAN

The statistical experiment or model $(\Omega, \mathcal{X}, \mathcal{P})$ is described as follows. An event $X \in \mathcal{X}$ is observed such that the probability distribution of X is from a parametric family of probability measures $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, all defined on the same measurable space $(\Omega, \mathcal{X})$; the true parameter θ is unknown. The statistical experiment or model is simply denoted by $\{P_\theta, \theta \in \Theta\}$. We consider a sequence of statistical experiments $(\Omega^{(n)}, \mathcal{X}^{(n)}, \mathcal{P}^{(n)})$ where $\mathcal{P}^{(n)} = \{P_\theta^{(n)}, \theta \in \Theta\}$. For example, we can construct a sequence of statistical experiments with a location parameter.

Example 1 (Gaussian location model) The random variables $X_1, \ldots, X_n$ are i.i.d. with

$X_1 \sim P_\theta \triangleq \mathcal{N}(\theta, 1),\ \theta \in \mathbb{R}$. We set the sequence of probability spaces as the product space

$$(\Omega^{(n)}, \mathcal{X}^{(n)}) = (\Omega^{\otimes n}, \mathcal{X}^{\otimes n}), \quad P_\theta^{(n)} = P_\theta^{\otimes n}, n = 1, 2, \ldots.$$

Here, we consider $X^{(n)} \triangleq (X_1, \ldots, X_n)$ as a single observation in $\mathcal{X}^{(n)}$.

Definition 1 (LAN) [4] The sequence of statistical models $\{P_\theta^{(n)}, \theta \in \Theta\}$ is LAN at $\theta_0 \in \Theta$ if there exist matrices $\mathbf{r}_n(\theta_0)$ and $\mathbf{I}_{\theta_0}$ and random vectors Δ_{n,θ_0} such that $\mathcal{L}(\Delta_{n,\theta_0}|\theta_0) \Rightarrow \mathcal{N}(0, \mathbf{I}_{\theta_0})$ and for every $\mathbf{h}$

$$\log \frac{dP^{(n)}_{\theta_0+\mathbf{r}_n(\theta_0)\mathbf{h}}}{dP^{(n)}_{\theta_0}}(X^{(n)}) = \mathbf{h}^T \Delta_{n,\theta_0}(X^{(n)}) - \frac{1}{2}\mathbf{h}^T \mathbf{I}_{\theta_0}\mathbf{h} + o_{P^{(n)}_{\theta_0}}(1), \tag{41}$$

where $\mathcal{L}(\Delta_{n,\theta_0}|\theta_0) \Rightarrow \mathcal{N}(0, \mathbf{I}_{\theta_0})$ denotes that Δ_{n,θ_0} converges in distribution to $\mathcal{N}(0, \mathbf{I}_{\theta_0})$ under $P^{(n)}_{\theta_0}$ probability and $o_{P^{(n)}_{\theta_0}}(1)$ represents a term that converges to zero in $P^{(n)}_{\theta_0}$ probability.

Here, θ_0 is called the base parameter, Δ_{n,θ_0} the central sequence, and $\mathbf{I}_{\theta_0}$ the Fisher information matrix which actually coincides with the conventional definition of Fisher information for smooth parametric families. The experiment $\{P^{(n)}_{\theta_0+\mathbf{r}_n(\theta_0)\mathbf{h}}\}$ is called the local experiment with local parameter $\mathbf{h}$. LAN is a property of a (parametric) statistical model itself and implies that the loglikelihood ratio between the experiment based on θ_0 and local experiment admits a pointwise quadratic approximation in the neighborhood of θ_0 with a Gaussian linear term and a deterministic quadratic term (Not all statistical models satisfy the LAN condition). The statistical model in Example 1 is a good example of LAN family. It is LAN for all $\theta \in \mathbb{R}$ with $r_n(\theta) = \frac{1}{\sqrt{n}}$, $\Delta_{n,\theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \theta)$, and $I_\theta = 1$.

Now consider two sequences of probability distributions $\{P^{(n)}\}$ and $\{\tilde{P}^{(n)}\}$ defined on the same measurable space $(\Omega^{(n)}, \mathcal{X}^{(n)})$. As n goes to infinity, the relation between two sequences of distributions can be classified in several cases.

Definition 2 (Entire Asymptotic Separation) [5] The sequences $\{P^{(n)}\}$ and $\{\tilde{P}^{(n)}\}$ are entirely asymptotically separated if there is a subsequence $n_k \to \infty$, $k \to \infty$ and sets $A_{n_k} \in \mathcal{X}^{(n)}$ such that

$$P^{(n_k)}(A_{n_k}) \to 1 \text{ and } \tilde{P}^{(n_k)}(A_{n_k}) \to 0 \text{ as } k \to \infty. \tag{42}$$

For the i.i.d. Gaussian model in Example 1, if we choose $P^{(n)} = P_{\theta=1}^{\otimes n}$ and $\tilde{P}^{(n)} = P_{\theta=-1}^{\otimes n}$, two sequences are entirely asymptotically separated. In this case, we can set $A_n = \{(X_1, \dots, X_n) : \sum_{i=1}^n X_n > 0\}$. The distributions of $\sum_{i=1}^n X_n$ under two sequences are Gaussian with mean n and $-n$, respectively with an equal standard deviation $\sqrt{n}$. Hence, the distance between two distributions becomes unbounded as n goes to infinity and (42) is satisfied. Another extreme case in asymptotic relation is given by the following definition.

Definition 3 (Contiguity) [5] The sequence $\{\tilde{P}^{(n)}\}$ is contiguous to the sequence $\{P^{(n)}\}$ if for all $A_n \in \mathcal{X}^{(n)}$,

$$P^{(n)}(A_n) \to 0 \text{ implies } \tilde{P}^{(n)}(A_n) \to 0. \tag{43}$$

The concept of contiguity and entire asymptotic separation can be considered as asymptotic versions of absolute continuity and singularity between two probability measures, respectively. Note that the entire asymptotic separation is symmetric whereas the contiguity is not. If the contiguity condition also holds in the other direction, two sequences of probability measures are called mutually contiguous. It is also possible that the asymptotic relation between two sequences of probability distributions does not fall into either case in the above.

The simplification of asymptotics by the LAN theory comes from the relations between two experiments $\{P_{\theta_0}^{(n)}\}$ and $\{P_{\theta_0+\mathbf{r}_n(\theta_0)\mathbf{h}}^{(n)}\}$. When a statistical model satisfies the LAN condition, two sequences of distributions $\{P_{\theta_0}^{(n)}\}$ and $\{P_{\theta_0+\mathbf{r}_n(\theta_0)\mathbf{h}}^{(n)}\}$ are mutually contiguous and in limit these two experiments are represented by Gaussian location models $\{\mathcal{N}(0, \mathbf{I}_{\theta_0}^{-1})\}$ and $\{\mathcal{N}(\mathbf{h}, \mathbf{I}_{\theta_0}^{-1})\}$ as n goes to infinity. Hence, the asymptotic analysis of detection between the null hypothesis and contiguous alternative can be based on these limit Gaussian experiments that we can handle easily.

B. Asymptotically Locally Most Powerful Detector

The detection in the asymptotic situation, i.e., with infinite number of samples or observations has been well investigated for M-ary hypothesis testing or channel capacity problems. In the formulation of M-ary hypothesis testing, the number of hypotheses is fixed regardless of the sample size whereas the number of hypotheses increases with a certain ratio to the sample size or the codeblock length for channel capacity. The asymptotic analysis in these formulations is based on the entire asymptotic separation between discrete hypotheses. It is well known that the error probability converges to zero exponentially for M-ary hypothesis testing or for the case that the data rate is less than the channel capacity as the sample size goes to infinity. In this case, the rate of convergence or error exponent usually serves as a criterion for the asymptotic performance. However, the detection problem of signal with a continuous and unknown strength such as the model in (1, 3) does not lend itself to the error exponent approach easily due to the uncountable nature of the alternative hypothesis.

Moreover, for the test of the hypothesis (1, 3) with a good separation of parameters between the null and alternative hypothesis, the error exponent may not be a proper criterion in asymptotic situations where a sufficiently large number of observation samples are already available. Specifically, suppose that the alternative hypothesis is bounded away from the null. That is,

$$\begin{aligned} H_0 :&\quad \theta = 0 \\ H_1 :&\quad \theta > \theta_1, \end{aligned} \tag{44}$$

where $\theta_1 > 0$ is some constant. As the sample size goes to infinity, two sequences of distributions $\{P_0^{(n)}\}$ and $\{P_{\theta_1}^{(n)}\}$ becomes entirely asymptotically separated for most interesting cases and the power (probability of detection) of any reasonable detector approaches unity as the sample size becomes large (with possibly different convergence rate). Suppose that we have two such detectors and a sufficiently large number of samples. Then, the powers of two detectors are already very close to unity, and the convergence rate is no longer a proper measure for assessing the performance of a detector in the asymptotic regime[5]. The asymptotic local optimality is another choice for such an asymptotic scheme. The detection is focused on the alternative which is close to the null $\theta = 0$ where the distributions of the null and alternative hypothesis are still nonseparable or mutually contiguous. In our case, the asymptotic criteria focus on the low signal-to-noise (SNR) range as in [25] [28] [29] since the parameter θ represents the amplitude of the spatial signal.

The LAN theory provides an analytic framework for such detection problems and gives

[5]The error exponent or convergence rate can be considered as an approach to the case that each observation sample has a fixed SNR. It can be used to determine a detector requiring the minimum number of samples for a given error probability.

an asymptotic optimality criterion. When a statistical model satisfies the LAN condition, we can construct an asymptotic local upper bound of the power for any sequence of detectors with a given asymptotic size and a sequence of detectors that achieves this bound.

Theorem 4 (Asymptotic local upper bound) [1], [4] Let $\{\phi_n\}$ be any sequence of asymptotic α-tests for hypothesis (3). That is,

$$\limsup_{n\to\infty} \mathbb{E}_{n,0}\phi_n \le \alpha.$$

Suppose that the sequence of statistical models $\{P_\theta^{(n)}, \theta \in \Theta = [0,\infty)\}$ is LAN at $\theta = 0$ with normalizing sequence $r_n(0)$ ($\to 0$), central sequence $\Delta_{n,0}$, and FIM I_0. Then, for any $M > 0$,

$$\limsup_{n\to\infty} \sup_{0<r_n(0)^{-1}\theta\le M} \left[\mathbb{E}_{n,\theta}\phi_n - Q(Q^{-1}(\alpha) - r_n^{-1}(0)\theta I_0^{1/2})\right] \le 0. \tag{45}$$

Furthermore, the following procedure is an asymptotic α-test for the hypotheses (3) that achieves the bound

$$\begin{aligned} &\text{Take } H_0 \quad \text{if} \quad I_0^{-1/2}\Delta_{n,0} \le Q^{-1}(\alpha), \\ &\text{Take } H_1 \quad \text{if} \quad I_0^{-1/2}\Delta_{n,0} > Q^{-1}(\alpha), \end{aligned} \tag{46}$$

where $\Delta_{n,0}$ is the central sequence.

Here, $\theta \in (0, r_n(0)M)$ is the local neighborhood of the null hypothesis depending on the sample size where two hypotheses are still not separable. A detector that achieves the asymptotic local upper bound with asymptotic size α is called the asymptotically locally most powerful (ALMP) detector with size α. The detector (46) is known as the *score test* and is an ALMP detector. Furthermore, for the one-sided detection problem, the ALMP detector is most powerful not only in the local neighborhood of null parameter but also in the entire parameter space $\Theta = (0,\infty)$ asymptotically [3].

Appendix II: Proofs

Lemma 1: Let $X_{A_n}^{(n)}$ be the sequence of Poisson processes (or corresponding statistical model) with probability distribution $\{P_\theta^{(n)}, \theta \in \Theta = [0,\infty)\}$ induced by intensity measure $\Lambda_\theta^{(n)}(d\mathbf{x}) \triangleq \lambda^{(n)}(\theta,\mathbf{x})d\mathbf{x},\ \mathbf{x} \in A_n$. Let the conditions *(B.1)-(B.3)* below be satisfied. Then, the statistical model $\{P_\theta^{(n)}, \theta \in \Theta\}$ is LAN at $\theta_0 \in \Theta$ with central sequence Δ_{n,θ_0} and normalizing sequence $r_n(\theta_0)$ defined as (23) and (24), respectively.

(B.1) All intensity measures $\{\Lambda_\theta^{(n)}(d\mathbf{x}), \theta \in \Theta\}, n = 1, 2, \ldots$ are equivalent or mutually absolutely continuous for all $\theta \in \Theta$ and $\Lambda_\theta^{(n)}(A_n) < \infty$ for all n. We define

$$S_n(\theta, \theta_0; \mathbf{x}) \triangleq \frac{\Lambda_\theta^{(n)}(d\mathbf{x})}{\Lambda_{\theta_0}^{(n)}(d\mathbf{x})} = \frac{\lambda^{(n)}(\theta, \mathbf{x})}{\lambda^{(n)}(\theta_0, \mathbf{x})}.$$

(B.2) The function $S_n(\theta, \theta_0; \mathbf{x}), \theta, \theta_0 \in \Theta, x \in A_n$ is continuously differentiable with respect to θ at θ_0. We define

$$\Psi_n(\theta, \theta_0; \mathbf{x}) \triangleq 2\sqrt{S_n(\theta, \theta_0; \mathbf{x})}, \quad \dot{\Psi}_n(\theta, \theta_0; \mathbf{x}) \triangleq \frac{\partial}{\partial \theta}\Psi_n(\theta, \theta_0; \mathbf{x})$$

$$\dot{\Psi}_n(\theta_0, \theta_0; x) = \frac{\dot{\lambda}^{(n)}(\theta_0, \mathbf{x})}{\lambda^{(n)}(\theta_0, \mathbf{x})}$$

Then, the quantity

$$J_n(\theta_0) \triangleq \int_{A_n} |\dot{\Psi}_n(\theta_0, \theta_0; \mathbf{x})|^2 \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \tag{47}$$

is positive (> 0) at $\theta_0 \in \Theta$ and

$$r_n(\theta_0) \triangleq J_n(\theta_0)^{-1/2} \to 0 \text{ as } n \to \infty. \tag{48}$$

(B.3)

$$\lim_{n\to\infty} \int_{A_n} |r_n(\theta_0)\dot{\Psi}_n(\theta_0, \theta_0; \mathbf{x})|^3 \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \to 0, \tag{49}$$

and for every $C > 0$,

$$\lim_{n\to\infty} r_n^2(\theta_0) \sup_{|\theta - y| + |\theta - z| < r_n(\theta_0)C} \int_{A_n} \Big[\frac{\dot{\lambda}^{(n)}(z, \mathbf{x})}{\lambda^{(n)}(y, \mathbf{x})} - \frac{\dot{\lambda}^{(n)}(\theta_0, \mathbf{x})}{\lambda^{(n)}(\theta_0, \mathbf{x})}\Big]^2 \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) = 0. \tag{50}$$

Proof: In Appendix III.

Proof of Theorem 1

Since

$$\lambda^{(n)}(\theta, \mathbf{x}) = \theta n f(x) + n\lambda_0, \ \mathbf{x} \in A, \ \lambda_0 > 0$$

for Model 1 and $f(\mathbf{x}) \geq 0,\ \mathbf{x} \in A$ by the condition *(C.1)*, the family of intensity measures $\{\Lambda_\theta^{(n)}(d\mathbf{x})\}$ are equivalent for $\theta \geq 0$ and all n. Hence, *(B.1)* is satisfied and $\Psi_n(\theta, \theta_0; \mathbf{x})$ is well defined and its derivative is given by

$$\dot{\Psi}_n(\theta, \theta_0; \mathbf{x}) = \frac{nf(\mathbf{x})}{\theta n f(\mathbf{x}) + n\lambda_0} / \sqrt{\frac{\theta n f(\mathbf{x}) + n\lambda_0}{\theta_0 n f(\mathbf{x}) + n\lambda_0}}$$

and at $\theta_0 = 0$,

$$\dot{\Psi}_n(0, 0; \mathbf{x}) = \lambda_0^{-1} f(\mathbf{x}).$$

$$\begin{aligned} J_n(0) &= \int_{A_n} |\dot{\Psi}_n(0, 0; \mathbf{x})|^2 \Lambda_0^{(n)}(d\mathbf{x}), \\ &= \int_A \lambda_0^{-2} f^2(\mathbf{x}) n \lambda_0 d\mathbf{x}, \\ &= n\lambda_0^{-1} \int_A f^2(\mathbf{x}) d\mathbf{x}, \end{aligned} \tag{51}$$

since

$$A_n = A \text{ for all } n,\ |A| < \infty.$$

By *(C.2)* and *(C.3)*,

$$0 < C_0 := \int_A f^2(\mathbf{x}) d\mathbf{x} < \infty \Rightarrow 0 < J_n(0) < \infty \text{ for all } n,$$

and

$$\lim_{n\to\infty} J_n(0) = \infty.$$

Hence, *(B.2)* is satisfied.

Define

$$\begin{aligned} r_n(0) &\triangleq J_n(0)^{-1/2} \\ &= n^{-1/2} \lambda_0^{1/2} \left[\int_A f^2(\mathbf{x}) d\mathbf{x} \right]^{-1/2} = n^{-1/2} \lambda_0^{1/2} C_0^{-1/2}. \end{aligned}$$

Now check (49).

$$\begin{aligned} & \int_{A_n} |r_n(0) \dot{\Psi}_n(0, 0; \mathbf{x})|^3 \Lambda_0^{(n)}(d\mathbf{x}) \\ = & \int_A n^{-3/2} \lambda_0^{3/2} [\int_A f^2(\mathbf{x}) d\mathbf{x}]^{-3/2} \lambda_0^{-3} f^3(\mathbf{x}) n \lambda_0 d\mathbf{x} \\ = & \ n^{-1/2} \lambda_0^{-1/2} [\int_A f^2(\mathbf{x}) d\mathbf{x}]^{-3/2} \int_A f^3(\mathbf{x}) d\mathbf{x}\ . \end{aligned}$$

Since $f(\mathbf{x}) \geq 0$, $\int_A f(\mathbf{x}) > 0$, and $\sup_{\mathbf{x}\in A} f(\mathbf{x}) = M < \infty$,

$$0 < \int_A f^3(\mathbf{x}) d\mathbf{x} < \infty.$$

We have

$$\int_{A_n} |r_n(0)\dot{\Psi}_n(0,0;\mathbf{x})|^3 \Lambda_0(d\mathbf{x}) \to 0 \text{ as } n \to \infty.$$

Hence, the Lindeberg condition (49) is satisfied.

For given $C > 0$,

$$\begin{aligned}
& \sup_{|y|+|z|<r_n(0)C} \int_{A_n} r_n^2(0) \left[\frac{\dot{\lambda}^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})} - \frac{\dot{\lambda}^{(n)}(0,\mathbf{x})}{\lambda^{(n)}(0,\mathbf{x})} \right]^2 \Lambda_0^{(n)}(d\mathbf{x}) \\
= & \sup_{0\leq y<n^{-1/2}C'} \int_A n^{-1}\lambda_0^2 C_0^{-1} \left[\frac{f(x)}{yf(x)+\lambda_0} - \frac{f(x)}{\lambda_0} \right]^2 n\lambda_0 dx \\
= & \sup_{0\leq y<n^{-1/2}C'} \lambda_0^3 C_0^{-1} \int_A \left[\frac{f(x)}{yf(x)+\lambda_0} - \frac{f(x)}{\lambda_0} \right]^2 dx \\
= & \sup_{0\leq y<n^{-1/2}C'} \lambda_0^3 C_0^{-1} \int_A \left[\frac{f(x)}{\lambda_0} \right]^2 \left[\frac{1}{yf(x)/\lambda_0+1} - 1 \right]^2 dx \\
\leq & \sup_{0\leq y<n^{-1/2}C'} \lambda_0^3 C_0^{-1} \left[\frac{1}{yM/\lambda_0+1} - 1 \right]^2 \int_A \left[\frac{f(x)}{\lambda_0} \right]^2 dx \\
\leq & \lambda_0^3 C_0^{-1} \left[\frac{1}{n^{-1/2}C'M/\lambda_0+1} - 1 \right]^2 \int_A \left[\frac{f(x)}{\lambda_0} \right]^2 dx \\
\to & 0 \text{ as } n \to \infty.
\end{aligned}$$

where $C' = \lambda_0^{1/2} C_0^{-1/2} C$. Here, we used the fact that $h(x)$ defined in (52) is monotone increasing for $x \geq 0$.

$$h(x) = \left(\frac{1}{ax+1} - 1 \right)^2, \text{ for any } a > 0. \tag{52}$$

Hence, *(B.3)* is satisfied. ■

Proof of Theorem 2

$$\frac{\dot{\lambda}^{(n)}(0,\mathbf{x})}{\lambda^{(n)}(0,\mathbf{x})} = \frac{nf(\mathbf{x})}{\theta n f(\mathbf{x}) + n\lambda_0}|_{\theta=0} = \lambda_0^{-1} f(\mathbf{x}),$$

$$r_n(0) = \left[\int_A \left(\frac{\dot{\lambda}^{(n)}(0,\mathbf{x})}{\lambda^{(n)}(0,\mathbf{x})} \right)^2 \lambda^{(n)}(0,\mathbf{x}) d\mathbf{x} \right]^{-1/2},$$

$$= n^{-1/2}\lambda_0^{1/2}\left(\int_A f^2(\mathbf{x})d\mathbf{x}\right)^{-1/2},$$

$$\begin{aligned}
\Delta_{n,0} &= r_n(0)\int_A f(\mathbf{x})[X^{(n)}(d\mathbf{x}) - \Lambda_0^{(n)}(d\mathbf{x})], \\
&= n^{-1/2}\lambda_0^{-1/2}\left(\int_A c^2 s^2(\mathbf{x})d\mathbf{x}\right)^{-1/2} \\
&\qquad \left(\int_A cs(\mathbf{x})X^{(n)}(d\mathbf{x}) - n\lambda_0 \int_A cs(\mathbf{x})d\mathbf{x}\right), \\
&= n^{-1/2}\lambda_0^{-1/2}\left(\int_A s^2(\mathbf{x})d\mathbf{x}\right)^{-1/2} \\
&\qquad \left(\sum_{i:\ \mathbf{x}_i \in A} s(\mathbf{x}_i) - n\lambda_0 \int_A s(\mathbf{x})d\mathbf{x}\right).
\end{aligned} \tag{53}$$

Here, we used the fact that $f(\mathbf{x})$ is a scaled version of $s(\mathbf{x})$ ($f(\mathbf{x}) = cs(\mathbf{x}),\ c > 0$). The last step is by the definition of stochastic integral. The ALMP detector is obtained by Theorem 4.

■

Proof of Theorem 3

By Theorem 4, the asymptotic local upper bound for the global power is given by

$$Q(Q^{-1}(\alpha) - r_n^{-1}(0)\theta).$$

Since $Q(\cdot)$ is a monotone decreasing function, the maximum upper bound for a fixed θ is achieved by maximizing $r_n^{-1}(0)$ for a given n. Since $r_n(0)$ is given, using (20), by

$$\begin{aligned}
r_n(0) &= n^{-1/2}\lambda_0^{1/2}\left[\int_A f^2(\mathbf{x})d\mathbf{x}\right]^{-1/2} \\
&= \left(n\lambda_{h0}p_m \frac{(\beta'_{\tau_0}(0))^2}{\beta_{\tau_0}(0)} \int_A s^2(\mathbf{x})d\mathbf{x}\right)^{-1/2},
\end{aligned}$$

the theorem follows. ■

Appendix III: Proof of Lemma 1

This proof is mostly cited from [6], [7] and more detailed steps are involved.

Let $\Psi_n(\theta,\theta_0;\mathbf{x})$ be defined as follows.

$$\Psi_n(\theta,\theta_0;\mathbf{x}) = 2\sqrt{S_n(\theta,\theta_0;\mathbf{x})}, \quad \mathbf{x}\in A_n, \quad \theta,\theta_0\in\Theta,$$

where

$$S_n(\theta,\theta_0;\mathbf{x}) = \frac{\Lambda_\theta^{(n)}(d\mathbf{x})}{\Lambda_{\theta_0}^{(n)}(d\mathbf{x})} = \frac{\lambda^{(n)}(\theta,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}, \quad \mathbf{x}\in A_n, \qquad \Lambda_\theta^{(n)}(d\mathbf{x}) = \lambda^{(n)}(\theta,\mathbf{x})d\mathbf{x}.$$

Then, we have

$$\dot\Psi_n(\theta,\theta_0;\mathbf{x}) \triangleq \frac{\partial}{\partial\theta}\Psi_n(\theta,\theta_0;\mathbf{x}) = \frac{\partial}{\partial\theta}2\sqrt{\frac{\lambda^{(n)}(\theta,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}} = 2\frac{1}{2}\left(\frac{\dot\lambda^{(n)}(\theta,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}\right)/\left(\sqrt{\frac{\lambda^{(n)}(\theta,\mathbf{x})}{\lambda^{(n)}(\theta_0,x)}}\right).$$

Let's define the following quantities.

$$q_n(\theta_0,\mathbf{x}) \triangleq \dot\Psi_n(\theta_0,\theta_0;\mathbf{x}) = \dot\Psi_n(\theta,\theta_0;\mathbf{x})|_{\theta=\theta_0} = \frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})},$$

$$J_n(\theta_0) \triangleq \int_{A_n}\dot\Psi_n(\theta_0,\theta_0;\mathbf{x})\dot\Psi_n^T(\theta_0,\theta_0;\mathbf{x})\Lambda_{\theta_0}^{(n)}(d\mathbf{x}) = \int_{A_n}\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})}d\mathbf{x},$$

$$r_n(\theta_0) \triangleq J_n^{-1/2}(\theta_0),$$

$$\Delta_{n,\theta_0} \triangleq \int_{A_n} r_n(\theta_0)\dot\Psi_n(\theta_0,\mathbf{x})[X^{(n)}(d\mathbf{x})-\Lambda_{\theta_0}^{(n)}(d\mathbf{x})] = \int_{A_n} r_n(\theta_0)\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}[X^{(n)}(d\mathbf{x})-\Lambda_{\theta_0}^{(n)}(d\mathbf{x})].$$

Since the probability measures $\{P_\theta^{(n)},\theta\in\Theta\}$ are equivalent if the corresponding intensity measure $\{\Lambda_\theta^{(n)},\theta\in\Theta\}$ [7], by condition *(B.1)*, all measures $\{P_\theta^{(n)},\theta\in\Theta\}$ are equivalent and for every h, the likelihood ratio is well defined and denoted by

$$Z_n(h) = \frac{dP_{\theta_0+r_n(\theta_0)h}^{(n)}}{dP_{\theta_0}^{(n)}}(X_{A_n}^{(n)}). \tag{54}$$

Denote $\tilde q_n = r_n(\theta_0)\dot\Psi_n(\theta_0,\theta_0;\mathbf{x})$, $\theta_h = \theta_0 + r_n(\theta_0)h$, and $S_n = S_n(\theta_h,\theta_0;\mathbf{x})$. Using (16), the log-likelihood ratio is explicitly given by

$$\begin{aligned}\log Z_n(h) &= \int_{A_n}\log S_n(\theta_h,\theta_0;\mathbf{x})X^{(n)}(d\mathbf{x}) - \Lambda_{\theta_h}^{(n)}(A_n) + \Lambda_{\theta_0}^{(n)}(A_n),\\ &= \int_{A_n}\log S_n X^{(n)}(d\mathbf{x}) - \int_{A_n}[S_n-1]\Lambda_{\theta_0}^{(n)}(d\mathbf{x}),\\ &= \int_{A_n}\log S_n[X^{(n)}(d\mathbf{x}) - \Lambda_{\theta_0}^{(n)}(d\mathbf{x})] - \int_{A_n}[S_n - 1 - \log S_n]\Lambda_{\theta_0}^{(n)}(d\mathbf{x}),\end{aligned}$$

$$
\begin{aligned}
&= \int_{A_n} (\log S_n - h\tilde{q}_n)[X^{(n)}(d\mathbf{x}) - \Lambda_{\theta_0}^{(n)}(d\mathbf{x})] + \int_{A_n} h\tilde{q}_n[X^{(n)}(d\mathbf{x}) - \Lambda_{\theta_0}^{(n)}(d\mathbf{x})] \\
&\quad -\frac{1}{2}\int_{A_n} (h\tilde{q}_n)^2\Lambda_{\theta_0}^{(n)}(d\mathbf{x}) + \frac{1}{2}\int_{A_n} (h\tilde{q}_n)^2\Lambda_{\theta_0}^{(n)}(d\mathbf{x}) - \int_{A_n} [S_n - 1 - \log S_n]\Lambda_{\theta_0}^{(n)}(d\mathbf{x}), \\
&= h\int_{A_n} \tilde{q}_n[X^{(n)}(d\mathbf{x}) - \Lambda_{\theta_0}^{(n)}(d\mathbf{x})] - \frac{1}{2}\int_{A_n} (h\tilde{q}_n)^2\Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \\
&\quad + \int_{A_n} (\log S_n - h\tilde{q}_n)[X^{(n)}(d\mathbf{x}) - \Lambda_{\theta_0}^{(n)}(d\mathbf{x})] \\
&\quad - \int_{A_n} [S_n - 1 - \log S_n - \frac{1}{2}(h\tilde{q}_n)^2]\Lambda_{\theta_0}^{(n)}(d\mathbf{x}), \\
&= h\int_{A_n} \tilde{q}_n[X^{(n)}(d\mathbf{x}) - \Lambda_{\theta_0}^{(n)}(d\mathbf{x})] - \frac{1}{2}\int_{A_n} (h\tilde{q}_n)^2\Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \\
&\qquad + R_1^{(n)}(h, \theta_0, X^{(n)}) - R_2^{(n)}(h, \theta_0, X^{(n)}). \qquad (55)
\end{aligned}
$$

Hence, for the LAN property, we need to show that for every h, the first term converges in distribution to a Gaussian random variable uder $P_{\theta_0}^{(n)}$ probability , the sencond term to a deterministic quantity, and the remaining terms go to zero in $P_{\theta_0}^{(n)}$ probability.

Consider the following equality. For any fixed h,

$$
\begin{aligned}
\frac{\partial}{\partial s}\log S_n(\theta_0 + r_n hs, \theta_0; \mathbf{x}) &= hr_n \frac{\frac{\partial}{\partial s}S_n(\theta_0 + r_n hs, \theta_0; \mathbf{x})}{S_n(\theta_0 + r_n hs, \theta_0; \mathbf{x})} \\
\int_0^1 \frac{\partial}{\partial s}\log S_n(\theta_0 + r_n hs, \theta_0; \mathbf{x})ds &= \int_0^1 hr_n \frac{\frac{\partial}{\partial s}S_n(\theta_0 + r_n hs, \theta_0; \mathbf{x})}{S_n(\theta_0 + r_n hs, \theta_0; \mathbf{x})}ds \\
\log S_n(\theta_0 + r_n h, \theta_0; \mathbf{x}) - \log S_n(\theta_0, \theta_0; \mathbf{x}) &= \int_0^1 hr_n \frac{\frac{\partial}{\partial s}S_n(\theta_0 + r_n hs, \theta_0; \mathbf{x})}{S_n(\theta_0 + r_n hs, \theta_0; \mathbf{x})}ds \\
\log S_n(\theta_0 + r_n h, \theta_0; \mathbf{x}) &= \int_0^1 hr_n \frac{\frac{\partial}{\partial s}S_n(\theta_0 + r_n hs, \theta_0; \mathbf{x})}{S_n(\theta_0 + r_n hs, \theta_0; \mathbf{x})}ds \\
&= \int_0^1 hr_n \frac{\frac{\partial}{\partial s}\frac{\lambda^{(n)}(\theta_0 + r_n hs, \mathbf{x})}{\lambda^{(n)}(\theta_0, \mathbf{x})}}{\frac{\lambda^{(n)}(\theta_0 + r_n hs, \mathbf{x})}{\lambda^{(n)}(\theta_0, \mathbf{x})}}ds \\
&= \int_0^1 hr_n \frac{\frac{\partial}{\partial s}\lambda^{(n)}(\theta_0 + r_n hs, \mathbf{x})}{\lambda^{(n)}(\theta_0 + r_n hs, \mathbf{x})}ds \\
&= \int_0^1 hr_n \dot{\Psi}_n(\theta_0 + r_n hs, \theta_0 + r_n hs; \mathbf{x})ds
\end{aligned}
\qquad (56)
$$

where, $r_n(\theta_0)$ is simply denoted by r_n (and hereafter). We can also denote $\tilde{q}_n$ as follows.

$$
\tilde{q}_n = r_n\dot{\Psi}_n(\theta_0, \theta_0; \mathbf{x}) = r_n\dot{\Psi}_n(\theta_0, \theta_0; \mathbf{x})\int_0^1 ds = \int_0^1 r_n\dot{\Psi}_n(\theta_0, \theta_0; \mathbf{x})ds \qquad (57)
$$

Now, consider the first remaining term.

$$R_1^{(n)}(h,\theta_0,\mathbf{x}_{A_n}) \;=\; \int_{A_n} (\log S_n - h\tilde{q}_n)[X^{(n)}(d\mathbf{x}) - \Lambda_{\theta_0}^{(n)}(d\mathbf{x})].$$

By the property of stochastic integral and Poisson process, $\mathbb{E}\ R_1^{(n)}(h,\theta_0,\mathbf{x}_{A_n}) = 0$ under $P_{\theta_0}^{(n)}$ probability and the variance under $P_{\theta_0}^{(n)}$ probability is given by [7]

$$\begin{aligned}
\mathbb{E}|R_1^{(n)}(h,\theta_0,\mathbf{x}_{A_n})|^2 &= \int_{A_n} [\log S_n - h\tilde{q}_n]^2\, \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \\
&= \|\log S_n - h\tilde{q}_n\|_{\theta_0}^2 \\
&\qquad \text{by Eq.(56) and (57),} \\
&= \|\int_0^1 hr_n\dot{\Psi}_n(\theta_0+r_nhs,\theta_0+r_nhs;\mathbf{x})ds - \int_0^1 hr_n\dot{\Psi}_n(\theta_0,\theta_0;\mathbf{x})ds\|_{\theta_0}^2 \\
&= \|\int_0^1 \Big(hr_n\dot{\Psi}_n(\theta_0+r_nhs,\theta_0+r_nhs;\mathbf{x}) - hr_n\dot{\Psi}_n(\theta_0,\theta_0;\mathbf{x})\Big)\, ds\|_{\theta_0}^2 \\
&= \|\int_0^1 hr_n\Big(\dot{\Psi}_n(\theta_0+r_nhs,\theta_0+r_nhs;\mathbf{x}) - \dot{\Psi}_n(\theta_0,\theta_0;\mathbf{x})\Big)\, ds\|_{\theta_0}^2 \\
&\le \int_0^1 \|hr_n\Big(\dot{\Psi}_n(\theta_0+r_nhs,\theta_0+r_nhs;\mathbf{x}) - \dot{\Psi}_n(\theta_0,\theta_0;\mathbf{x})\Big)\|_{\theta_0}^2 ds \\
&= |\mathbf{h}|^2\int_0^1 \|r_n\Big(\dot{\Psi}_n(\theta_0+r_nhs,\theta_0+r_nhs;\mathbf{x}) - \dot{\Psi}_n(\theta_0,\theta_0;\mathbf{x})\Big)\|_{\theta_0}^2 ds \\
&= |\mathbf{h}|^2\sup_{0\le s\le 1} \|r_n\Big(\dot{\Psi}_n(\theta_0+r_nhs,\theta_0+r_nhs;\mathbf{x}) - \dot{\Psi}_n(\theta_0,\theta_0;\mathbf{x})\Big)\|_{\theta_0}^2
\end{aligned}$$

Since $r_n \to 0$ and $\dot{\Psi}_n(\theta_0,\theta_0,\mathbf{x})$ is continous by the condition *(B.2)*, the righthand side term goes to zero. Since the convergence in L_2 implies the convergence in probability, we have

$$R_1^{(n)}(h,\theta_0,X^{(n)}) \to 0 \text{ in } P_{\theta_0}^{(n)} \text{ probability.}$$

Next, consider the second term $R_2^{(n)}(h,\theta_0,X^{(n)})$ which is deterministic.

$$R_2^{(n)}(h,\theta_0,X^{(n)}) \;=\; \int_{A_n} [S_n - 1 - \log S_n - \frac{1}{2}(h\tilde{q}_n)^2]\Lambda_{\theta_0}^{(n)}(d\mathbf{x})$$

Consider the following equalities. The first part of $R_2^{(n)}(h,\theta_0,X^{(n)})$ is expressed by

$$\begin{aligned}
&\int_{A_n} [S_n(\theta_0+r_nh,\theta_0;\mathbf{x}) - 1 - \log S_n(\theta_0+r_nh,\theta_0;\mathbf{x})]\Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \\
=\; &\int_{A_n} [S_n(\theta_0+r_nh,\theta_0;\mathbf{x}) - 1 - \log S_n(\theta_0+r_nh,\theta_0;\mathbf{x})]\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}
\end{aligned}$$

$$
\begin{aligned}
&= \int_{A_n} \left[\lambda^{(n)}(\theta_0 + r_n h, \mathbf{x}) - \lambda^{(n)}(\theta_0, \mathbf{x}) - \lambda^{(n)}(\theta_0, \mathbf{x}) \log \frac{\lambda^{(n)}(\theta_0 + r_n h, \mathbf{x})}{\lambda^{(n)}(\theta_0, \mathbf{x})} \right] d\mathbf{x} \\
&= \int_{A_n} \int_{\theta_0}^{\theta_0 + r_n h} \int_{\theta_0}^{y} \frac{\dot{\lambda}^{(n)}(z, \mathbf{x})}{\lambda^{(n)}(y, \mathbf{x})} \frac{\dot{\lambda}^{(n)}(y, \mathbf{x})}{\lambda^{(n)}(\theta_0, \mathbf{x})} \, dz dy \lambda^{(n)}(\theta_0, \mathbf{x}) d\mathbf{x}
\end{aligned}
$$

For the last step, we used the following.

$$
\begin{aligned}
& \int_{A_n} \int_{\theta_0}^{\theta_0 + r_n h} \int_{\theta_0}^{y} \frac{\dot{\lambda}^{(n)}(z, \mathbf{x})}{\lambda^{(n)}(y, \mathbf{x})} \frac{\dot{\lambda}^{(n)}(y, \mathbf{x})}{\lambda^{(n)}(\theta_0, \mathbf{x})} \, dz dy \lambda^{(n)}(\theta_0, \mathbf{x}) d\mathbf{x} \\
&= \int_{A_n} \int_{\theta_0}^{\theta_0 + r_n h} \int_{\theta_0}^{y} \frac{\dot{\lambda}^{(n)}(z, \mathbf{x})}{\lambda^{(n)}(y, \mathbf{x})} \dot{\lambda}^{(n)}(y, \mathbf{x}) \, dz dy d\mathbf{x} \\
&= \int_{A_n} \int_{\theta_0}^{\theta_0 + r_n h} \frac{\dot{\lambda}^{(n)}(y, \mathbf{x})}{\lambda^{(n)}(y, \mathbf{x})} \int_{\theta_0}^{y} \dot{\lambda}^{(n)}(z, \mathbf{x}) \, dz dy d\mathbf{x} \\
&= \int_{A_n} \int_{\theta_0}^{\theta_0 + r_n h} \frac{\dot{\lambda}^{(n)}(y, \mathbf{x})}{\lambda^{(n)}(y, \mathbf{x})} [\lambda^{(n)}(y, \mathbf{x}) - \lambda^{(n)}(\theta_0, \mathbf{x})] \, dy d\mathbf{x} \\
&= \int_{A_n} \int_{\theta_0}^{\theta_0 + r_n h} [\dot{\lambda}^{(n)}(y, \mathbf{x}) - \lambda^{(n)}(\theta_0, \mathbf{x}) \frac{\dot{\lambda}^{(n)}(y, \mathbf{x})}{\lambda^{(n)}(y, \mathbf{x})}] \, dy d\mathbf{x} \\
&= \int_{A_n} [\lambda^{(n)}(y, \mathbf{x}) - \lambda^{(n)}(\theta_0, \mathbf{x}) \log \lambda^{(n)}(y, \mathbf{x})]_{\theta_0}^{\theta_0 + r_n h} d\mathbf{x} \\
&= \int_{A_n} \left[\lambda^{(n)}(\theta_0 + r_n h, \mathbf{x}) - \lambda^{(n)}(\theta_0, \mathbf{x}) - \lambda^{(n)}(\theta_0, \mathbf{x}) \log \frac{\lambda^{(n)}(\theta_0 + r_n h, \mathbf{x})}{\lambda^{(n)}(\theta_0, \mathbf{x})} \right] d\mathbf{x}.
\end{aligned}
$$

The second part of $R_2^{(n)}(h, \theta_0, \mathbf{x}_{A_n})$ is now expressed by

$$
\begin{aligned}
\frac{1}{2} h^2 &= \frac{1}{2} h^2 r_n^2(\theta_0) J_n(\theta_0), \quad \text{by def. of } r_n \\
&= \frac{1}{2} h^2 r_n^2(\theta_0) \int_{A_n} \frac{\dot{\lambda}^{(n)}(\theta_0, \mathbf{x})^2}{\lambda^{(n)}(\theta_0, \mathbf{x})^2} \lambda^{(n)}(\theta_0, \mathbf{x}) d\mathbf{x} \\
&= \int_{A_n} \int_{\theta_0}^{\theta_0 + r_n h} \int_{\theta_0}^{y} \frac{\dot{\lambda}^{(n)}(\theta_0, \mathbf{x})^2}{\lambda^{(n)}(\theta_0, \mathbf{x})^2} \, dz dy \lambda^{(n)}(\theta_0, \mathbf{x}) d\mathbf{x},
\end{aligned}
$$

where we used

$$
\begin{aligned}
\int_{\theta_0}^{y} dz &= y - \theta_0 \\
\int_{\theta_0}^{\theta_0 + \varphi h} [y - \theta_0] dy &= \frac{1}{2} h^2 r_n^2.
\end{aligned}
$$

So, we express $R_2^{(n)}(h, \theta_0, X^{(n)})$ as

$$
R_2^{(n)}(h, \theta_0, , X^{(n)}) = \int_{A_n} [S_n - 1 - \log S_n - \frac{1}{2}(h \tilde{q}_n)^2] \Lambda_{\theta_0}^{(n)}(d\mathbf{x})
$$

$$
\begin{aligned}
&= \int_{A_n}\int_{\theta_0}^{\theta_0+r_nh}\int_{\theta_0}^{y}\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}\; dzdy\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\\
&\quad -\int_{A_n}\int_{\theta_0}^{\theta_0+r_nh}\int_{\theta_0}^{y}\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}\; dzdy\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\\
&= \int_{A_n}\int_{\theta_0}^{\theta_0+r_nh}\int_{\theta_0}^{y}\left[\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}\right]\; dzdy\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\\
&= \int_{\theta_0}^{\theta_0+r_nh}\int_{\theta_0}^{y}\int_{A_n}\left[\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}\right]\;\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}dzdy.
\end{aligned}
$$

Now the integrand in the above equation is expressed by

$$
\begin{aligned}
&\int_{A_n}|\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}|\;\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x} \qquad (58)\\
=&\int_{A_n}|\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}-\frac{\dot\lambda^{(n)}(y,\mathbf{x})^2}{\lambda^{(n)}(y,\mathbf{x})^2}+\frac{\dot\lambda^{(n)}(y,\mathbf{x})^2}{\lambda^{(n)}(y,\mathbf{x})^2}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}|\;\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\\
\le&\int_{A_n}|\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}-\frac{\dot\lambda^{(n)}(y,\mathbf{x})^2}{\lambda^{(n)}(y,\mathbf{x})^2}|\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}+\int_{A_n}|\frac{\dot\lambda^{(n)}(y,\mathbf{x})^2}{\lambda^{(n)}(y,\mathbf{x})^2}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}|\;\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\\
=&\int_{A_n}|\frac{\dot\lambda^{(n)}(y,\mathbf{x})^2}{\lambda^{(n)}(y,\mathbf{x})^2}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}|\;\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}+\int_{A_n}|\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}-\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}|\frac{|\dot\lambda^{(n)}(y,\mathbf{x})|}{\lambda^{(n)}(y,\mathbf{x})}\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\\
&\quad\text{using } a^2-b^2=(a+b)(a-b) \text{ and Cauchy-Schwarz inequality,}\\
\le&\sqrt{\int_{A_n}|\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}|^2\;\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\int_{A_n}|\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}+\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}|^2\;\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}}\\
&+\sqrt{\int_{A_n}|\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}-\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}|^2\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\int_{A_n}\frac{\dot\lambda^{(n)}(y,\mathbf{x})^2}{\lambda^{(n)}(y,\mathbf{x})^2}\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}}. \qquad (59)
\end{aligned}
$$

Check the nondiminishing term in the above equation.

$$
\begin{aligned}
&\int_{A_n}\frac{\dot\lambda^{(n)}(y,\mathbf{x})^2}{\lambda^{(n)}(y,\mathbf{x})^2}\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\\
=&\int_{A_n}|\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}+\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}|^2\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\\
\le&\ 2\int_{A_n}|\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}|^2\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}+2\int_{A_n}|\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}|^2\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\\
=&\ r_n^{-2}(1+o(1)),\quad\text{by (50).} \qquad (60)
\end{aligned}
$$

Hence, (58) is bounded by

$$
\begin{aligned}
&\int_{A_n}|\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}|\;\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\\
\le&\sqrt{\int_{A_n}|\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}-\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}|^2\;\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\int_{A_n}|\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}+\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}|^2\;\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}}
\end{aligned}
$$

$$
\begin{aligned}
& +\sqrt{\int_{A_n} |\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})} - \frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}|^2 \lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x} \int_{A_n} \frac{\dot\lambda^{(n)}(y,\mathbf{x})^2}{\lambda^{(n)}(y,\mathbf{x})^2}\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}} \\
& \quad \text{using } |a+b|^2 \le 2a^2+2b^2 \\
\le & \sqrt{\int_{A_n} |\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})} - \frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})}|^2 \ \lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x} \int_{A_n} \left(2\frac{\dot\lambda^{(n)}(y,\mathbf{x})^2}{\lambda^{(n)}(y,\mathbf{x})^2} + 2\frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}\right) \ \lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}} \\
& +\sqrt{\int_{A_n} |\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})} - \frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}|^2 \lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x} \int_{A_n} \frac{\dot\lambda^{(n)}(y,\mathbf{x})^2}{\lambda^{(n)}(y,\mathbf{x})^2}\lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}} \\
& \quad \text{using (60),(50)} \\
= & \sqrt{r_n^{-2}o(1)r_n^{-2}(1+o(1))} + \sqrt{r_n^{-2}o(1)r_n^{-2}(1+o(1))} \\
= & r_n^{-2}o(1) \ \text{ since } \sqrt{o(1)} = o(1) \qquad (61)
\end{aligned}
$$

where $o(1)$ is uniform over $y, z : |\theta_0 - y| + |\theta_0 - z| < Cr_n(\theta_0)$ by (50) in the condition *(B.3)*. Now the convergence of $R_2^{(n)}(h, \theta_0, X^{(n)})$ to zero is established as follows.

$$
\begin{aligned}
R_2^{(n)}(h,\theta_0,X^{(n)}) &= \int_{\theta_0}^{\theta_0+r_nh}\int_{\theta_0}^{y}\left(\int_{A_n}\left[\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})} - \frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}\right] \ \lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\right)dzdy \\
&\le \int_{\theta_0}^{\theta_0+r_nh}(y-\theta_0) \sup_{z:|\theta_0-z|<Cr_n(\theta_0)}\Big(\int_{A_n}\Big[\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})} \\
&\qquad - \frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}\Big] \ \lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\Big)dy \\
&\le \sup_{y,z:|\theta_0-z|+|\theta_0-y|<Cr_n(\theta_0)}\Big(\int_{A_n}\Big[\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})} \\
&\qquad - \frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}\Big] \ \lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\Big)\int_{\theta_0}^{\theta_0+r_nh}(y-\theta_0)dy \\
&= [\frac{1}{2}y^2-\theta_0 y]_{\theta_0}^{\theta_0+r_nh} \sup_{y,z:|\theta_0-z|+|\theta_0-y|<Cr_n(\theta_0)}\Big(\int_{A_n}\Big[\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})} \\
&\qquad - \frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}\Big] \ \lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\Big) \\
&= \frac{1}{2}h^2r_n(\theta_0)^2 \sup_{y,z:|\theta_0-z|+|\theta_0-y|<Cr_n(\theta_0)}\Big(\int_{A_n}\Big[\frac{\dot\lambda^{(n)}(z,\mathbf{x})}{\lambda^{(n)}(y,\mathbf{x})}\frac{\dot\lambda^{(n)}(y,\mathbf{x})}{\lambda^{(n)}(\theta_0,\mathbf{x})} \\
&\qquad - \frac{\dot\lambda^{(n)}(\theta_0,\mathbf{x})^2}{\lambda^{(n)}(\theta_0,\mathbf{x})^2}\Big] \ \lambda^{(n)}(\theta_0,\mathbf{x})d\mathbf{x}\Big) \\
&\quad \text{by (61)} \\
&= \frac{1}{2}h^2r_n^2r_n^{-2}o(1) \\
&= h^2o(1).
\end{aligned}
$$

Hence, we have

$$R_2^{(n)}(h, \theta_0, \mathbf{x}_{A_n}) \to 0 \quad \text{as } n \to \infty. \tag{62}$$

The quadratic term in (55) is given by

$$\begin{aligned}
& \int_{A_n} (h\tilde{q}_n)^2 \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \\
= & \int_{A_n} h^2 |\tilde{q}_n|^2 \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \\
= & \; h^2 \int_{A_n} |\tilde{q}_n|^2 \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \\
= & \; h^2,
\end{aligned}$$

by the definition of $\tilde{q}_n$. Hence, the Fisher information I_{θ_0} is one. Finally, check the convergence of the central sequence.

$$\begin{aligned}
\Delta_{n,\theta_0} &= \int_{A_n} \tilde{q}_n [X^{(n)}(d\mathbf{x}) - \Lambda_{\theta_0}^{(n)}(d\mathbf{x})] \\
&= \int_{A_n} r_n(\theta_0) \dot{\Psi}(\theta_0, \theta_0; \mathbf{x}) [X^{(n)}(d\mathbf{x}) - \Lambda_{\theta_0}^{(n)}(d\mathbf{x})]
\end{aligned}$$

We need to check Lindeberg condition. For any $\epsilon > 0$,

$$\begin{aligned}
L_n(\epsilon) &= \int_{A_n} |r_n(\theta_0)\dot{\Psi}_n(\theta_0, \theta_0; \mathbf{x})|^2 I_{\{|r_n(\theta_0)\dot{\Psi}_n(\theta_0,\theta_0;\mathbf{x})|>\epsilon\}} \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \\
&\leq \int_{A_n} |r_n(\theta_0)\dot{\Psi}_n(\theta_0, \theta_0; \mathbf{x})|^2 \frac{|r_n(\theta_0)\dot{\Psi}_n(\theta_0, \theta_0; \mathbf{x})|}{\epsilon} I_{\{|r_n(\theta_0)\dot{\Psi}_n(\theta_0,\theta_0;\mathbf{x})|>\epsilon\}} \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \\
&= \epsilon^{-1} \int_{A_n} |r_n(\theta_0)\dot{\Psi}_n(\theta_0, \theta_0; \mathbf{x})|^3 I_{\{|r_n(\theta_0)\dot{\Psi}_n(\theta_0,\theta_0;\mathbf{x})|>\epsilon\}} \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \\
&\leq \epsilon^{-1} \int_{A_n} |r_n(\theta_0)\dot{\Psi}_n(\theta_0, \theta_0; \mathbf{x})|^3 \Lambda_{\theta_0}^{(n)}(d\mathbf{x}) \to 0,
\end{aligned}$$

where the last step is by (49) in the condition *(B.3)*. Hence, we have

$$\Delta_{n,\theta_0} \Rightarrow \mathcal{N}(0, 1),$$

under $P_{\theta_0}^{(n)}$ probability. This concludes the proof. ■

References

[1] Lucien Le Cam, "Locally Asymptotically Normal Families of Distributions," *University of California Publications in Statistics*, vol. 3, no. 2, pp. 37-98, 1960.

[2] Lucien Le Cam and Grace Lo Yang, *Asymptotics in Statistics: Some Basic Concepts*, 2nd Edition, Springer-Verlag, New York, 2000.

[3] George G. Roussas, *Contiguity of Probability Measures: Some Applications in Statistics*, Cambridge at the university press, 1972.

[4] A.W. van der Vaart, *Asymptotic Statistics*, Cambridge university press, 1998.

[5] A.N. Shiryaev, *Probability*, 2nd Edition, Springer-Verlag, New York, 1995.

[6] Yu.A. Kutoyants, "Locally Asymptotic Normality for Processes of Poisson Type," *Soviet Journal of Contemporary Mathematical Analysis,* vol.14, pp.1-18, 1979.

[7] Yu.A. Kutoyants, *Statistical Inference for Spatial Poisson Processes*, Springer-Verlag, New York, 1998.

[8] Friedrich Liese and Udo Lorz, "Contiguity and LAN-property of Sequences of Poisson processes," *Kybernetika*, vol. 35, no. 3, pp.281-308, 1999.

[9] P.A.W. Lewis and G.S.Shedler, "Simulation of Nonhomogeneous Poisson Processes by Thinning," *Naval Res. Logistics Quart.*, vol. 26, no. 3, pp. 403-413, 1979.

[10] Sidney I. Resnick, *Adventures in Stochastic Processes*, Birkhäuser, Boston, 1992.

[11] Alan F. Karr, *Point Processes and Their Statistical Inference*, Marcel Dekker, Inc. New York and Basel, 1986.

[12] D.R. Snyder and M.I. Miller, *Random Point Processes in Time and Space*, Springer-Verlag, New York, 1991.

[13] Israel Bar-David, "Communications under the Poisson Regime," *IEEE Trans. Information Theory,* vol.15, no.1, pp.31-37, Jan. 1969.

[14] J.N.Tsitsiklis,"Decentralized detection," in *Advances in Statistical Signal Processing, vol 2- Signal Detection,* H.V.Poor and J.B.Thomas, JAI Press, 1990.

[15] R. Viswanathan and P.K.Varshney, "Distributed Detection with Multiple Sensors: Part I-Fundamentals ," *Proceedings of IEEE*, vol.85, no.1 pp.54-63, Jan. 1997.

[16] R.S.Blum, S.A.Kassam, and H.V.Poor, "Distributed Detection with Multiple Sensors: Part II-Advanced Topics ," *Proceedings of IEEE*, vol.85, no.1 pp.64-79, Jan. 1997.

[17] Robert R. Tenney and Nils R. Sandell, "Detection with Distributed Sensors," *IEEE Trans on Aerospace and Electronic Systems,* vol. 17, no. 4, pp. 501- 509, July 1981.

[18] Z. Chair and P.K. Varshney, "Optimal Data Fusion in Multiple Sensor Detection Systems," *IEEE Trans. on Aerospace and Electronic Systems,* vol.22, no.1, pp.98-101, Jan. 1986.

[19] S.C.A. Thomopoulos, R. Viswanathan, and D. K. Bougoulias, "Optimal Distributed Detection Fusion," *IEEE Trans. Aerospace and Electronic Systems,* vol. 25, no. 5, pp. 761 -765, Sep. 1989.

[20] I.Y.Hoballarh and P.K.Varshney, "Distributed Bayesian Signal Detection," *IEEE Trans. Information Theory,* vol.35, no.5, pp.995-1000, Sep. 1989.

[21] Q. Yan and R. S. Blum, "Distributed signal detection under the Neyman-Pearson criterion," *IEEE Transactions on Information Theory,* vol.47, no.4, pp. 1368-1377, May 2001

[22] R. Viswanathan and V. Aalo, "On Counting Rules in Distributed Detection," *IEEE Trans. on Acoustics, Speech, and Signal Processing,* vol.37, no.5, pp.772-775, May 1989.

[23] J.Capon, "On the asymptotic efficiency of locally opitmum detectors," *IRE Trans. Information Theory*, vol.7, pp.67-71, Apr. 1961.

[24] H.V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd Edition, Springer, New York, 1994.

[25] H.V.Poor and J.B.Thomas, "Locally Optimum Detection of Discrete-Time Signals in Non-Gaussian Noise," *Journal of Acoustical Society of America*, vol.63, no.1, pp.75-80, Jan. 1978.

[26] S.A. Kassam and J.B. Thomas, "Dead-Zone Limiter: An Application of Conditional Tests in Nonparametric Detection," *Journal of Acoustical Socieity of America*, vol. 60, no. 4, pp.857-862, Oct. 1976.

[27] V. Aalo and R. Viswanathan, "Multilevel Quantisation and Fusion Scheme for the Decentralized Detection of an Unknown Signal," *IEE Proc. Radar, Sonar Navig.,* vol.141, no.1, pp.37-44, Feb. 1994.

[28] G.Fedele,L.Izzo, and L.Paura, "Optimum and suboptimum space-diversity detection of weak signals in non-Gaussian noise," *IEEE Trans. Communications*, vol.32, no.9, pp.990-997, Sep. 1984.

[29] R. Srinivasan, "The Detection of Weak Signals using Distributed Sensors," *Proc. of the Conference on Information Sicences and Systems*, pp.541-545, Princeton, NJ, March, 1990.

[30] John N. Tsitsiklis, "Decentralized Detection by a Large Number of Sensors," *Math. Control Signals System,* no.1 pp.167-182, 1988.

[31] Po-Ning Chen and Adrian Papamarcou, "New Asymptotic Results in Parallel Distributed Detection," *IEEE Trans. Information Theory,* vol. 39, no. 6, pp. 1847-1863, Nov. 1993.

[32] J.-F. Chamberland and V. Veeravalli, "Decentralized Detection in Sensor Networks," *IEEE Trans. Signal Processing,* vol. 51, no. 2, pp. 407-416, Feb. 2003.

[33] L.Tong, Q.Zhao, and S.Adireddy, "Sensor Networks with Mobile Agents," Proc. *2003 MILCOM*, Boston, MA, Oct. 2003.