Stochastic Counting in Sensor Networks, or: Noise is Good *

Y. M. Baryshnikov[1], E. G. Coffman[2], K. J. Kwak[2], and Bill Moran[3]

[1] Bell Labs
600 Mountain Ave.
Murray Hill, NJ 07974
ymb@research.bell-labs.com
[2] Electrical Engineering Dept.
Columbia University
New York, NY 10027
{egc,kjkwak}@ee.columbia.edu
[3] Electrical Engineering Dept.
University of Melbourne
Australia
b.moran@ee.unimelb.edu.au

Abstract. We propose a novel algorithm of counting indistinguishable objects by a collection of sensors. The information on multiple counting is recovered from the stochastic correlation patterns.

1 Introduction

One of the tasks fundamental to sensor applications is *counting*; that is, receiving and processing often noisy data, and returning the *number of objects* of interest located within a given sensor field. Such problems arise, for example, in the context of ecological or agricultural monitoring where, for example, a sensor might report the number of animals of a certain kind that it observes. In such situations, discriminating between different animals is almost certainly infeasible. Accordingly, we adopt the *minimality paradigm* (cf. [1]), and postulate that the sensors have the simplest functionality required to perform their task. In our case, while sensors can count the objects within their sensing ranges, they cannot identify them on the basis of location. in particular, locations relative to the domains of other sensors are unknown. One immediate hurdle to achieving the goal is then the over-counting of the objects: if one has no means to identify the objects, naïve summation of the counts reported by different sensors inflates the total count. Is it possible to correct this over-count without resorting to explicit identification of the counted objects? At face value, this goal seems to be completely unattainable, so it may come as something of a surprise that, by a process of *stochastic data fusion*, in this case fusing data with noise in the sensing process, we can design efficient counting procedures.

2 Problem Formulation

We assume that the ultimate task of the sensor network is to recover the total number of objects, henceforth referred to as *targets*. Targets τ potentially in the sensor field comprise a set $\mathcal{T}$. The set of targets actually in the sensor field is given by $\{\tau \in \mathcal{T} | X_\tau = 1\}$, where the indicator X_τ is 1 or 0 according as τ is or is not in the sensor field. Let $N := \sum_\tau X_\tau$ be the number of targets in the sensor field, which are to be observed by a collection $\mathcal{S}$ of sensors, σ. Typically, the sensing domain of a sensor is a disk of given radius. The sensor field is usually modeled as a rectangular subset of the union of the sensor domains, an assumption being that every point of the rectangular field is covered by at least one sensor. While perhaps useful to keep in mind as describing a canonical model, these assumptions are not needed in what follows.

We formalize target location relative to sensor domains by introducing an *incidence coefficient* $Z_{\sigma\tau}$ for each $\sigma \in \mathcal{S}$, $\tau \in \mathcal{T}$, which is 1 or 0 according as τ is or is not in the sensing area of σ. Thus, τ is in the intersection of $\sum_\sigma Z_{\sigma\tau}$ sensing domains. We can now define the target count C_σ by sensor σ as $C_\sigma = \sum_{\tau \in \mathcal{T}} Z_{\sigma\tau} X_\tau$, so the task of the sensor network we intend to design and analyze becomes the following:

Given the counts C_σ of the targets registered by each sensor σ, estimate the total number N of targets in the area covered by the sensors within the sensor field.

It is immediately clear that the problem as stated does not have a well-defined solution; indeed, the total count depends on the number of targets counted several times by different sensors. We have only the rough bounds $\max_{\sigma \in \mathcal{S}} C_\sigma \leq N \leq \sum_{\sigma \in \mathcal{S}} C_\sigma$. More precisely, as the well-known inclusion-exclusion formula indicates,

$$N = \sum_\sigma C_\sigma - \sum_{\sigma_1 < \sigma_2} C_{\sigma_1\sigma_2} + \ldots + (-1)^\ell \sum_{\sigma_1 < \cdots < \sigma_\ell} C_{\sigma_1\sigma_2\cdots\sigma_\ell} + \ldots \qquad (1)$$

where for an ordered subset $\sigma_1 < \sigma_2 < \ldots < \sigma_\ell$, $C_{\sigma_1\sigma_2\cdots\sigma_\ell}$ denotes the number of targets detected simultaneously by the sensors $\sigma_1, \sigma_2, \ldots, \sigma_\ell$. Recall that the summation terms can be interpreted as alternately compensating for the total under and over counts of the partial sums to their left.)

2.1 Model specifications

In the model above, we abstracted to the extreme the properties of sensor counting systems. It is instructive to map the model to some concrete realizations, to indicate the situations in which the techniques developed in this paper might be applied.

– *Domain coverage* Many areas of application of sensor networks involve a domain, planar or spatial, and a finite collection of subdomains (sensing

areas), labeled by the sensors, and consisting of the points where a target can be registered by the labeling sensor. An example of such applications would be an environmental monitoring system (counting the number of animals), or area surveillance systems, aiming at keeping track of the total number of persons or items of equipment in a facility. In this case each count (reading) returns instantaneous data on the number of targets in the sensed domain. Realistically, the domains are bound to overlap (to ensure complete coverage) making the estimation of the total count highly nontrivial.

– *Linear coverage* In applications to transportation, the sensors can be placed along a transportation link, and register the number of targets passing by. In this case, the number of targets registered by a sensor would encompass the routes crossing the location of the sensor. Notice that in this setting, the readings need not, and in fact cannot be instantaneous, as the targets require time to move between the locations of the sensors.

– *Communication networks* Application to communication networks requires an extension of the previous case to more general graphs. In this context targets represent the unique packets (or communication sessions) between some source-destination pairs, and a sensor corresponds to a counter of such packets or sessions traversing a router.

In each of these contexts, the problem is nontrivial, because of the overlap issue described earlier, together with the anonymous nature of the targets.

2.2 Stochastic overlap recovery

The total lack of information describing the overlap of the sensing areas precludes any rational hypothesis building when it comes to eliminating target over-counts. On the other hand, one can reason that *if the target visibility is not deterministic, but stochastic, then the contributions to the counts by different sensors coming from overlapping sensing areas might be correlated in some useful way.* One is led to speculate on whether the extra information given by correlations can in turn lead to a solution to our problem.

This paper shows indeed that exploitation of this simple idea parlays into a solution to our counting problem. In a sense, we acquire the ability to detect overlaps of the sensing areas at a sacrifice in the deterministic nature of the observations. "Noise is good" in that, by measuring randomly perturbed signals, we gather additional information, as compared to the situation when the measurements are noiseless; in the latter case, all measurements by a sensor would be identical and contribute nothing to our knowledge of the target counts.

The basic quantitative measure of correlation between two random variables is the covariance. We will see that the covariance between the counts by different sensors reflects the nature of the intersection between the sensing areas. To capture intersections of higher orders, however, one needs correlations of higher order. There are many statistical tools suitable for exploring the correlation structure of the data. This paper applies the notion of *cumulants*, which generalizes that of covariance coefficients, and which carries all the information of stochastic interdependency of the random variables being observed.

3   Related Work

The counting problem has been discussed in many papers in the system engineering context, see e.g. [2]. Most of the papers, however, assume that there is either explicit or implicit identification of the targets, or incremental counting, where the targets are counted one-by-one.

Earlier work by Marzullo [3] proposed an interval-based algorithm to detect faults and recover from errors. Assume at most f out of n sensors can be faulty and each sensor records the interval in which an object is detected. Marzullo calculates the smallest interval that contains all of the intersections of the $n - f$ intervals guaranteed to contain the true value that the event occurred. Based on Marzullo's work, Iyengar et al [4] propose another interval-based algorithm that produces a smaller interval under the assumption that the intervals of false sensing are sufficiently close.

In [5], the author addresses the question of counting anonymous targets in the context of robotic exploration. This poses challenges quite different from those arising in the context of this paper.

Brooks pointed out in [6] that integration (or fusion) of sensor readings is crucial to automating sensor networks, and that such distributed computing reinforces integrity of the system. To increase the precision of data, Brooks proposes sensor fusion and Byzantine agreement. To overcome faults in sensor readings, Clouqueur el al [7] propose two collaborative signal processing algorithms: value fusion and decision fusion. Sensors exchange a parameter (this can be consumed energy, required communication bandwidth, etc) to reach agreement as to whether a decision was correct or not, so recovery from the fault will be possible. The concept of *spatial sensor mining* (derivation of conclusions from distributed information collected over time), is introduced in [8]. They propose a change of focus from many sensors on a single target to many sensors on many targets.

Natheta [9] addresses the interesting but rather narrow problem of counting the number of nodes in the sensor network itself.

From our literature survey, it appears that no previous research has considered how to eliminate redundancy in data, to recover from errors, and hence to retrieve correct total target counts.

4   Preliminaries

This section introduces the necessary probabilistic tools (see e.g. [10] for greater detail). For a random variable X define the *cumulant generating function* as

$$\kappa(t) = \sum_n c_n(X)\frac{t^n}{n!} = \ln \mathbb{E}e^{tX}, \tag{2}$$

where we make the assumption that the Laplace transform of X is defined in some neighborhood of 0. In particular, the first two *factorial coefficients* are the mean and variance

$$c_1(X) = \mathbb{E}X \qquad c_2(X) = \mathbb{V}X \tag{3}$$

Given a collection of random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, not necessarily independent, define the cumulants $\kappa_n(X_1, \ldots, X_n)$ as

$$\kappa_n(X_1, \ldots, X_n) = \frac{\partial^n}{\partial t_1 \cdots \partial t_n} g(\mathbf{t})|_{\mathbf{t}=0},$$

where $\mathbf{t} = (t_1, \ldots, t_n)$ and $g(\mathbf{t}) = ln(\mathbb{E}(e^{\mathbf{t} \cdot \mathbf{X}}))$ is the logarithm of the Laplace transform of the vector-valued random variable $\mathbf{X}$. For example, the cumulant of the first order, $\kappa_1(X_1)$ is just the expected value of X_1; the joint cumulant of two random variables equals their covariance,

$$\kappa_2(X_1, X_2) = \mathbb{E}X_1 X_2 - \mathbb{E}X_1 \mathbb{E}X_2, \tag{4}$$

and the cumulant of three random variables X_1, X_2, and X_3 is given by

$$\begin{aligned} \kappa_3(X_1, X_2, X_3) &= \mathbb{E}(X_1 X_2 X_3) - \mathbb{E}(X_1 X_2)\mathbb{E}(X_3) \\ &- \mathbb{E}(X_1 X_3)\mathbb{E}(X_2) - \mathbb{E}(X_2 X_3)\mathbb{E}(X_1) + 2\mathbb{E}(X_1)\mathbb{E}(X_2)\mathbb{E}(X_3) \end{aligned} \tag{5}$$

In general, the cumulant $\kappa_n(X_1, \ldots, X_n)$ of order n is a polynomial in joint moments of the random variables and can be recovered recursively from the general formula

$$\mathbb{E}\prod_{i=1}^{n} X_i = \sum_{\Pi = P_1 \amalg P_2 \ldots \amalg P_k} \prod_{j=1}^{k} \kappa(P_j), \tag{6}$$

where Π is a partition of the index set $\{1, 2, \ldots, n\}$ into blocks P_j, and $c(P_j)$ is the cumulant of order $|P_j|$ of the random variables with indices in P_j. Thus, $\mathbb{E}X_1 X_2 = \kappa_2(X_1, X_2) + \kappa_1(X_1)\kappa_1(X_2)$ and

$$\begin{aligned} \mathbb{E}X_1 X_2 X_3 &= \kappa_3(X_1, X_2, X_3) + \kappa_2(X_1, X_2)\kappa_1(X_3) + \kappa_2(X_2, X_3)\kappa_1(X_3) \\ &+ \kappa_2(X_1, X_3)\kappa_1(X_2) + \kappa_1(X_1)\kappa_1(X_2)\kappa_1(X_3). \end{aligned}$$

We will be using the following important facts concerning cumulant functions.

PROPERTY 1: The cumulants are symmetric functions of their arguments;

PROPERTY 2: The cumulants are multilinear:

$$\kappa_n(X_1' + X_1'', X_2, \ldots, X_n) = \kappa_n(X_1', X_2, \ldots, X_n) + \kappa_n(X_1'', X_2, \ldots, X_n);$$

PROPERTY 3: If the random variables $X_1, \ldots, X_n$ can be split into two groups, say $\mathbf{X}' = (X_1, \ldots, X_k)$ and $\mathbf{X}'' = (X_{k+1}, \ldots, X_n)$, so that $\mathbf{X}'$ and $\mathbf{X}''$ are independent, then

$$\kappa_n(X_1, \ldots, X_n) = 0.$$

PROPERTY 4: If the X_i are different notations for the same random variable X, then $\kappa_n(X, \ldots, X) = c_n(X)$, the n-th factorial coefficient of the cumulant generating function of X.

5 Sensing Area Intersections and Cumulants

Let us return to the original problem of estimating the number of targets detected in the sensing domains of a collection of sensors.

5.1 Bernoulli visibility variables

We introduce the often justifiable assumption that the targets are visible to the sensors not continuously, but intermittently, in some time-dependent fashion. *Further, we assume that the visibility of the targets is random and independent across the targets.* Situations where such assumptions are valid range from environmental deployment of sensor networks, where the targets might be animals not covered by foliage, to perimeter protection applications, where moving targets are detected and counted. We do not go into further details here, for our interest is restricted to the algorithms underlying the solutions to the counting problem complicated by overlapping sensing areas and imperfect visibility.

Up to this point, X_τ has indicated (deterministically) whether τ was located somewhere in the sensor field. It is convenient now to consider that all targets are in the sensor field and let the X_τ model the random *visibility* of the targets. Formally,

$$X_\tau = \begin{cases} 1 \text{ if the target is } \textit{visible}; \\ 0 \text{ otherwise} \end{cases}$$

are *independent $\{0, 1\}$-valued random variables.* The sensor counts $C_\sigma = \sum_\tau Z_{\sigma\tau} X_\tau$, then become random variables. Sensing noise is defined precisely by the following visibility assumption.

> *The i.i.d. random variables X_τ are* `Bernoulli`(p) *distributed, that is, they independently take the value 1 with probability p, and 0 with probability $1 - p$.*

We infer that the count C_σ has the `Binomial`(p, N_σ) distribution, where $N_\sigma = \sum_\tau Z_{\sigma\tau}$ is the number of targets in σ's sensing area.

To appreciate the basic ideas behind our counting algorithm, we need only consider a system of two sensors. For this case, if A_1 and A_2 denote the two sensing areas, then the sensor-field partition of interest consists of the disjoint subsets $A_1 \setminus A_2$, $A_2 \setminus A_1$, and $A_1 \bigcap A_2$, which we denote by $A_{1\setminus2}, A_{2\setminus1}, A_{12}$. A similar notation applies to the number of targets in the respective areas, e.g., $N_{1\setminus2}, N_{12}, \ldots$ and to the counts $C_{1\setminus2}, C_{12}, \ldots$. To verify the fact that the cumulant for the two sensor counts C_1 and C_2 is simply their covariance, write

$$\kappa_2(C_1, C_2) = \kappa_2(C_{1\setminus2} + C_{12}, C_2) = \kappa_2(C_{12}, C_2)$$
$$= \kappa_2(C_{12}, C_{2\setminus1} + C_{12}) = \kappa_2(C_{12}, C_{12})$$
$$= N_{12} c_2(\textit{Bernoulli}(p)) = p(1 - p) N_{12}$$

where Properties 2 and 3 have been applied in the first two lines and Property 4 in the last line.
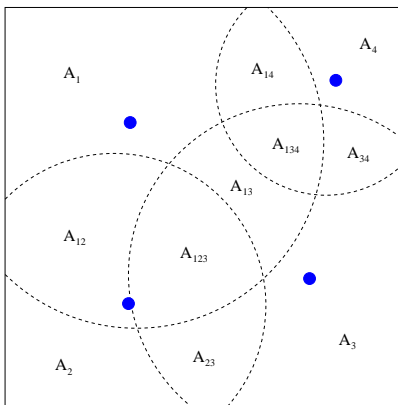
Fig. 1. Sensor Coverage Area

This simple formula carries two useful messages. Firstly, assuming p is known, knowledge of the covariance gives information about the size of the intersections of the sensing areas A_1 and A_2. Secondly, this information can be recovered only when target visibility is *truly stochastic*; that is, when the probability of visibility is neither 0, in which case the sensor system would not function, nor 1, in which case we would be left with our original intractable problem.

To recover the counts in intersections of more than two sensing areas (see Figure 1, for example), one takes higher cumulants. More precisely, using again the standard properties of cumulants, and the simpler notation $c_n(p) \equiv c_n(\texttt{Bernoulli(p)})$, we have

Lemma 1. *For sensors $\sigma_1, \sigma_2, \ldots, \sigma_n$ (repetitions allowed), let $C_k = \sum_\tau Z_{\sigma_k \tau} X_\tau,\quad 1 \leq k \leq n$, be the corresponding counts. To account for repetitions, suppose that the sensors can be divided into two groups, say $\{\sigma_1, \ldots, \sigma_i\}$ and $\{\sigma_{i+1}, \ldots, \sigma_n\}$, such that each sensor in the latter group is also a sensor in the former. Then the n-th order cumulant is given by*

$$\kappa_n(C_1, C_2, \ldots, C_n) = N_{12\ldots i} \cdot c_n(p).$$

where, extending earlier notation, $N_{12\ldots i}$ denotes the number of targets in the intersection of the sensing areas of the first i (i.e., the distinct) sensors.

Thus, to recover the count in the intersection of n sensing areas, $A_1, A_2, \ldots, A_n$ (again assuming that the visibility probability p is known), it is enough to estimate the cumulant $\kappa_n(C_1, C_2, \ldots, C_n)$ given that the n-th factorial coefficient of the Bernoulli random variable satisfies $c_n(p) \neq 0$.

But this last condition presents us with a potential problem. The cumulant $c_n(p)$ as a function of p is a polynomial of degree n with zeros at 0 and 1 and a further $(n-2)$ *interior* zeros in the interval $(0,1)$; a fact that can be easily
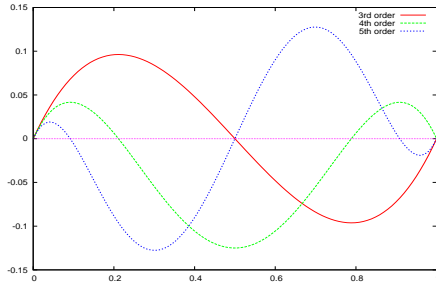
Fig. 2. Cumulants for Bernoulli variables as functions of p: $c_3(1/2) = 0$; in fact, more generally, cumulants of odd orders vanish at $1/2$.

derived from the well-known formula

$$c_{n+1} = p \cdot (1 - p) \cdot \frac{dc_n}{dp} \qquad (7)$$

which in turn follows from the explicit expression for the cumulant generating function of Bernoulli random variables,

$$\ln \mathbb{E} e^{s\text{Bernoulli}(p)} = \ln[pe^s + (1 - p)].$$

For example, plots of the third and fourth cumulant polynomials in p are shown in Figure 2. However, it is easy to verify that the interior zeros of consecutive cumulant polynomials interlace, and hence, if $c_n(p) = 0$, then $c_{n+1}(p) \neq 0$. Therefore, if the cumulant of order n vanishes for a given probability of visibility p, the next cumulant $c_{n+1}(p)$ will be non-vanishing at p, and the size of the intersection of the sensing areas $A_1, A_2, \ldots, A_n$ can be reconstructed, by Lemma 1, from the cumulant

$$\kappa_{n+1}(C_1, C_2, \ldots, C_n, C_n) = N_{12\ldots n} c_{n+1}(p). \qquad (8)$$

5.2 Poisson Counts

The preceding theory addressed the situation where the number of targets in each sensing domain is fixed and finite and only the visibility patterns are intermittent. Another realistic scenario would be that of *infinitely* many potential targets in each sensing area, with the *measure of a sensing area* being quite general. For simplicity, this measure will be taken here as the classical Lebesgue measure, in which case the measure is just the area, a_σ. Correspondingly, in lieu of the binomial random variables describing the observed counts visible by each sensor, one is driven to the assumption that, for each sensor σ, the count of visible targets is *Poisson distributed*, with the parameter of the Poisson law being proportional to a_σ. Of course, by extension, the number of targets visible

in the intersections of the sensing areas is Poisson distributed as well, again with a parameter proportional to the areas of the intersections.

More formally, we assume that, as before, a finite number of sensors $\sigma \in \mathcal{S}$ are given, and that corresponding sensing domains A_σ have areas A_σ. For any collection of sensors $\mathcal{S}' \subseteq \mathcal{S}$, we denote the intersection of their sensing areas by $A_{\mathcal{S}'}$, and the area of this intersection by $A_{\mathcal{S}'}$. The visible objects in A_σ are counted by C_σ and are assumed to be a sample of a Poisson point process with intensity λa_σ; equivalently, we have an intensity-λ Poisson pattern of points in the plane, with those points of the pattern falling in A_σ being stochastically the same as a sample of a rate λa_σ Poisson process. Note particularly that, according to this definition, the counts of the visible targets in non-intersecting areas are independent. The overall goal is to recover the total number, N, of objects, which is Poisson distributed with mean $\mathbb{E}N = \lambda| \cup_{\sigma \in \mathcal{S}} A_\sigma|$ from the individual counts C_σ, which requires estimates of the counts for all intersections, as before.

As in the case of Bernoulli random variables, the cumulants are ideally suited to our problem:

Lemma 2. *The n-th order cumulant is given by*

$$\kappa_n(C_1, C_2, \ldots, C_n) = \lambda a_{12\ldots n}.$$

The proof of this lemma again follows immediately from the standard properties of cumulants and the fact that the cumulant generating function for the Poisson random variable of parameter λa is $e^{\lambda a} - 1$. In other words, for all cumulants $c_n(\texttt{Poisson}(\lambda a)) = \lambda a$.

6 Target Count Recovery

The results presented above indicate how a procedure might be designed for target-count recovery using statistical estimates of the cumulants. We concentrate on the more challenging case of Bernoulli visibility variables and start by giving the algorithm in broad outline; we then flesh it out with details that depend on specific circumstances.

Target Count Recovery Algorithm
Assume that the counts are sampled at unit rate from the same population of targets over a time period of duration, T, a given parameter.

1. For our Bernoulli case, the moments in (3) can be expressed as $\mathbb{E}C_\sigma = pN_\sigma$, $\mathbb{V}C_\sigma = p(1-p)N_\sigma$, which imply

$$p := 1 - \mathbb{V}C_\sigma/\mathbb{E}C_\sigma.$$

Computing the first two empirical moments and then substituting into the above gives an estimate for the visibility probability, p.
2. From (7) compute the numbers $c_n(p)$ to one greater, say $s+1$, than the maximum depth (order) of intersections in the sensor field.
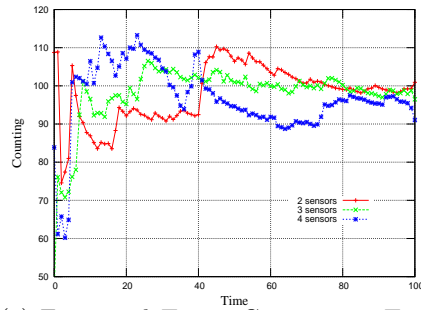
(a) Estimated Target Count : 100 Target, p=0.7
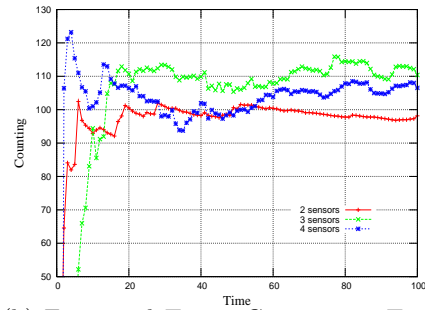
(b) Estimated Target Count : 100 Target, p=0.85
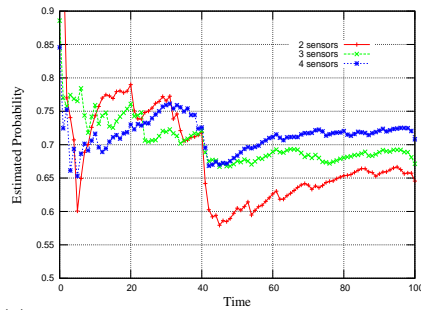
(c) Estimated Visibility Probability : 100 Target, p=0.7
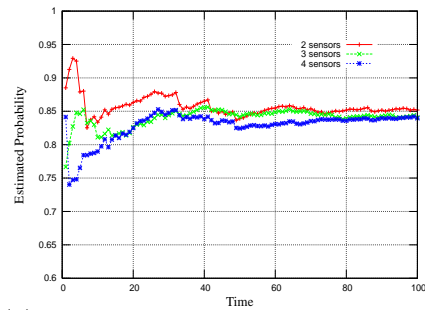
(d) Estimated Visibility Probability : 100 Target, p=0.85

Fig. 3. Time trace of estimated target count and estimated visibility probability with 2,3, and 4 sensors.

3. Compute the remaining cumulants $\kappa_3, \ldots, \kappa_s$ using formulas derived from (6) (recall that the first two are given by the coefficients computed in step 1).
4. Compute the estimates of the $N_{12\ldots n}$ from (2) unless $c(p)$ is too close to 0, in which case use (8) instead.
5. Compute the estimate for N from the data in step 4 substituted into the inclusion-exclusion formula (1).

Remarks: The estimate for p in step 1 can be averaged over several sensors, to improve its quality.

As the order of cumulants grows, the convergence of sampled values to the limit becomes slower. Thus, it makes sense to try to limit the order of cumulants needed for estimation by stopping the procedure when the count precision reaches a desired degree. To this end, the monotone character of the partial sums in the inclusion-exclusion formula allows one to halt the computations when the upper and lower bounds are close enough.

Indeed, as

$$\sum_{\sigma} N_{\sigma} \geq \sum_{\sigma} N_{\sigma} - \sum_{\sigma_1 < \sigma_2} N_{\sigma_1 \sigma_2} + \sum_{\sigma_1 < \sigma_2 < \sigma_3} N_{\sigma_1 \sigma_2 \sigma_3} \geq \ldots \geq N \geq$$

$$\geq \sum_{\sigma} N_{\sigma} - \sum_{\sigma_1 < \sigma_2} N_{\sigma_1 \sigma_2} + \sum_{\sigma_1 < \sigma_2 < \sigma_3} N_{\sigma_1 \sigma_2 \sigma_3} - \sum_{\sigma_1 < \sigma_2 < \sigma_3 < \sigma_4} N_{\sigma_1 \sigma_2 \sigma_3 \sigma_4} \geq \sum_{\sigma} N_{\sigma} - \sum_{\sigma_1 < \sigma_2} N_{\sigma_1 \sigma_2},$$

one can iteratively compute the upper bounds (with the deepest intersections involved of odd order) and lower bounds (with the deepest intersections of even order) until the required precision is achieved. ∎

7 Experimental Results

In this section, we simulate stochastic count recovery in the two-dimensional (planar) and one-dimensional sensoria.

7.1 Planar sensorium

Here is our setup:

- 2, 3, and 4 sensors are located in a 50 × 50 area of interest.
- The sensing radius is equal to 35 for all three cases, so that the sensing regions provide at most a 2-cover of the area of interest.
- 100 targets are distributed uniformly at random over the sensing area and sensors detect each target with probability p. Each sensor counts the number of detected targets within its sensing radius.

From the sampled data, we calculate the sample mean and variance to estimate the visibility probability and the sample cumulants to estimate and recover the total number of targets in sensing area. Experiments are performed with different p: 0.7 and 0.85 respectively. Over time, the estimated target counting and detection probability converge to the actual value as in Figure 3. This demonstrates the statements in the preceding sections. Note that as the visibility probability increases, the estimated target count and the estimated visibility probability converge faster with smaller errors.

Next we extend our experiment to a more general case. We assume 24 sensors are distributed uniformly at random over 100 × 100 sensing area. Each sensor has sensing radius of 15. (This ensure that any point in the area can be covered by at most 3 sensors.) We assume the visibility probability p is equal to 0.85.

Experiments are performed with 100, 200, 300, and 400 targets respectively. We have run extensive simulations to get a mean estimated target count and estimated visibility probability. The standard deviation is provided as a measure of error. As in the previous experiments, estimated target count, estimated visibility probability and cumulants are calculated from the collected data of 24 sensors. For comparison, to help understand how well the estimation scheme performs, we provide the simple summation of all counting from 24 sensors. As the

(a) Estimated Target Count (b) Estimated Visibility Probability
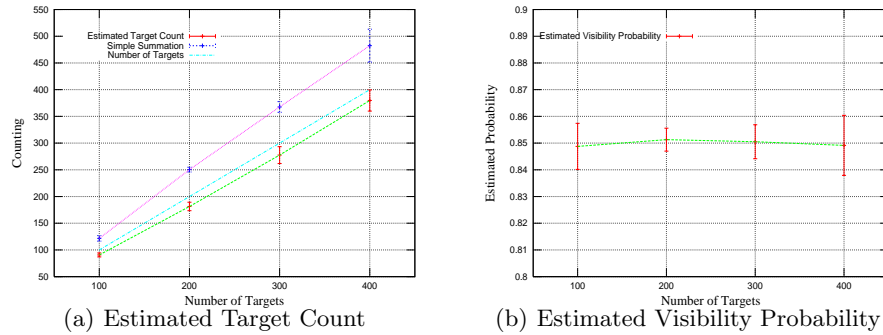
Fig. 4. Estimated target count and estimated visibility probability. 24sensors with 100,200,300, and 400 targets. (p=0.85)

number of target increases, so do the errors, but the error of our proposed scheme is fairly small compared to that of simple summation. The estimated visibility probability is reasonably accurate and provides up to double digit precision as in Figure 4.
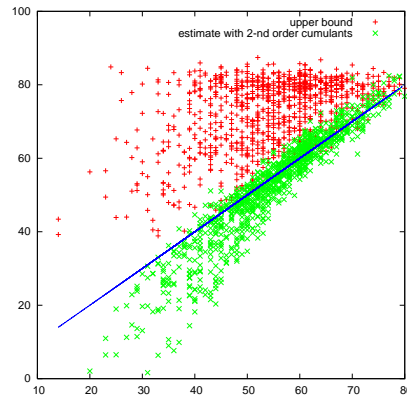
7.2 Linear sensorium

Here the substrate sarrying the targets is one-dimensional, with the point-wise targets. The sensors are indentified with certain intervals on the serorium, counting the number of visible targets falling within the corresponding interval.

Unlike the previous setup, in this set of experiments we randomized also over the sensor positions. We consider random placements of $S = 5$ sensors on an interval so that the total length of the sensors (not necessarily the total covered length) is equal to the interval length, see Figure 7.2. The intervals are chosen intependently, with uniformly distributed endpoints in the unit interval; the targets are uniformly distributed.
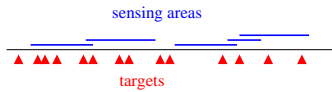
We again apply our stochastic count recovery algorithm taking into account only the cumulants of the second order. We can see that second order cumulants (via imputed second order overlaps) already give quite reasonabe approximation of the actual number of targets in the covered area.

The results are shown on Figure 5.

Fig. 5. The plot shows the scatter plots of the actual number of targets in the linear sensorium and upper bound (the sum total of the countes by each sensor, in red) and the estimates based on the inferred overlaps (in green).

8   Conclusion

We have described an algorithm for target count recovery from measurements by a sensor network, where each sensor is capable of counting the number of targets of interest in its sensing region. Overlaps of these regions preclude the possibility of using simple summation of counts as an effective estimate of the total number of targets.

We show that, by exploiting the stochastic nature of the measurements, and stochastic independence between sensors and between targets, it is possible over multiple measurements to arrive at an estimate of the total number of targets present. To do this we employ a method based on cumulants that quantifies non-independence between measurements and thereby permits estimation of the number of targets in the (multiple) overlaps of sensing regions of different sensors. Together with the inclusion-exclusion these numbers provide the estimated total number of targets. The method is described for both Bernoulli and Poisson detection models, and simulations are given to demonstrate the effectiveness of the method.

In a future work we extend the described methodology to the problem of Internet traffic measurements and monitoring.

References

1. Yu. Baryshnikov and R. Ghrist. Target enumeration via integration over planar sensor networks. Technical report, UIUC, 2007.
2. B. R. Son, S. C. Shin, and J. Kim. Implementation of the real-time people counting system using wireless sensor networks. *International Journal of Multimedia and Uniquitous Engineering*, 2(2):63–80, 2007.

3. K. Marzullo. Tolerating failures of continuous-valued sensors. *ACM Transactions on Computer Systems*, 8:284–304, 1990.

4. S.S. Iyengar and L. Prasad. A general computational framework for distributed sensing andfault-tolerant sensor integration. *IEEE Transactions on System, Man and Cybernetics*, 25(4):643–650, 1995.

5. B. Gfeller, M. Mihalak, S. Suri, E. Vicari, and P. Widmayer. Counting targets with mobile sensors in an unknown environment. preprint.

6. R. Brooks and S. S. Iyengar. Robust distributed computing and sensing algorithm. *Computer*, 29(6):53–60, 1996.

7. T. Clouqueur, P. Ramanathan, K. K. Saluja, and K. Wang. Value-fusion versus decision-fusion for fault-tolerance in collaborative target detection in sensor networks. In *Proceedings of the Conference on Information Fusion*, pages TuC2/25–TuC2/30, 2001.

8. J. Heidemann and N. Bulusu. Using geospatial information in sensor networks. In *Proceedings of the Workshop on Intersection of Geospatial Information and Information Technology*, 2001.

9. S. Nath, P. Gibbons, S. Seshan, and Z. Anderson. Synopsis diffusion for robust aggregation in sensor networks. In *Proceedings of SenSys' 04*, pages 250–262, 2004.

10. *Mathematics of Statistics*, pages 77–82. Princeton, NJ: Van Nostrand, 1951.