NATO Undersea Research Centre
Centre de Recherche Sous-Marine de l'OTAN

Reprint Series NURC-PR-2006-005

A Bayesian approach to predicting an unknown number of targets based on sensor performance

Bryan, Karna; Carthel, Craig

August 2006

Originally published in:

FUSION 2006, The 9th International Conference on Information Fusion 2006, Florence (Italy), 10-13 July 2006, Conference Proceedings.

NATO Undersea Research Centre (NURC)

NURC conducts world class maritime research in support of NATO's operational and transformational requirements. Reporting to the Supreme Allied Commander Transformation, the Centre maintains extensive partnering to expand its research output, promote maritime innovation and foster more rapid implementation of research products.

The Scientific Programme of Work (SPOW) is the core of the Centre's activities and is organized into four Research Thrust Areas:

- Expeditionary Mine Countermeasures (MCM) and Port Protection (EMP)
- Reconnaissance, Surveillance and Undersea Networks (RSN)
- Expeditionary Operations Support (EOS)
- Command and Operational Support (COS)

NURC also provides services to other sponsors through the Supplementary Work Program (SWP). These activities are undertaken to accelerate implementation of new military capabilities for NATO and the Nations, to provide assistance to the Nations, and to ensure that the Centre's maritime capabilities are sustained in a fully productive and economic manner. Examples of supplementary work include ship chartering, military experimentation, collaborative work with or services to Nations and industry.

NURC's plans and operations are extensively and regularly reviewed by outside bodies including peer review of the research, independent national expert oversight, review of proposed deliverables by military user authorities, and independent business process certification. The Scientific Committee of National Representatives, membership of which is open to all NATO nations, provides scientific guidance to the Centre and the Supreme Allied Commander Transformation.

NOTE: The NURC Reprint series reprints papers and articles published by NURC authors in the open literature as an effort to widely disseminate NURC products. Users are encouraged to cite the original article where possible.

A Bayesian approach to predicting an unknown number of targets based on sensor performance

K. Bryan
Command and Operational Support Department
NATO Undersea Research Centre
La Spezia, Italy
bryan@nurc.nato.int

C. Carthel
Command and Operational Support Department
NATO Undersea Research Centre
La Spezia, Italy
carthel@nurc.nato.int

Abstract - ***Estimating remaining targets after some attempt has been made to detect an overall, unknown number of targets is critical to determining the potential threat associated with these remaining targets. This paper presents a Bayesian approach to calculate the distribution on the number of remaining targets given the sensor performance and the number of targets detected. For a single sensor, a closed form posterior distribution on remaining targets is derived. For multiple sensors, the corresponding posterior distribution is developed. A naive implementation of this calculation is shown to be computationally prohibitive, and an efficient means for performing the calculation is presented.***

Keywords: sensor management, Dirichlet-Multinomial hierarchical model.

1 Introduction

In Mine Countermeasure operations, sensors with an associated probability of detection attempt to find an unknown number of mines on the seafloor. After these operations, some estimation must be made of the number of remaining mines in order to predict the remaining threat in the area. The combined efforts of multiple sensors allocated to sub-areas, each having an associated probability of detection, must also be considered.

An estimation of the remaining targets after a certain level of search is required to evaluate the potential threat associated with these remaining targets. We consider the problem of estimating an unknown number of stationary targets based on the number of targets detected by a given sensor and the probability of detection associated with this sensor. The problem of determining an unknown number of targets is then considered for a larger area where many sensors are working independently in several sub-areas of the overall area. The quantity of interest is the total number of unknown targets in the entire area. Given uniform sensor performance, the algorithm provides the same results for a single sensor working in an area, or multiple sensors working in several sub-areas or this original area when the total number of targets detected is the same.

A standard application of Bayes' Theory is to estimate the unknown success probability p of the Binomial distribution for a population of a fixed size given a certain number of observed successes. Looking at the Binomial distribution from another perspective, an estimation can also be made of the population size based on a given success probability and number of observed successes. In general, the problem of estimating the number of trials n for the Binomial distribution has received little attention [2].

The distribution on remaining targets is developed in Section 2. For the single sensor case, assuming a infinite uniform distribution on the total number of targets, the posterior distribution is shown to follow a Negative Binomial distribution. The posterior distribution on remaining targets is then developed for multiple sensors. The prior distribution should be chosen in a way to give the same results as the single sensor case under certain circumstances. This requires the introduction of a hierarchical model to provide a flexible prior distribution giving intuitive results. Section 3 sets up the calculations of interest as expected values. The separable form of these expectation calculations is then exploited in Section 4 giving an efficient way of calculating the expected values for the quantities of interest.

2 Distribution on targets remaining

A Bayesian approach is used to determine a distribution on targets remaining in the area. Before developing this distribution, we review the basic definitions in

Bayesian inference [4, 5] . We would like to make statements about an unknown parameter θ given data y. The *joint probability distribution* $P(\theta, y)$ follows from the definition of conditional probability:

$$P(\theta, y) = P(\theta)P(y|\theta), \tag{1}$$

where $P(\theta)$ is referred to as the *prior distribution* and $P(y|\theta)$ as the *sampling distribution*, or the *likelihood function*. Bayes' rule follows from another application of the definition of conditional probability:

$$P(\theta|y) = \frac{P(\theta, y)}{P(y)} = \frac{P(\theta)P(y|\theta)}{P(y)}, \tag{2}$$

where $P(y) = \sum_\theta P(\theta)P(y|\theta)$ sums over all possible values of θ. The symbol "$\propto$," proportional to, will be used for the *unnormalized posterior density*

$$P(\theta|y) \propto P(\theta)P(y|\theta), \tag{3}$$

as considering terms up to a constant of normalization can be beneficial. When the unnormalized posterior density is not a recognizable form (and therefore the constant is not known), the constant of normalization can be computed by summing (or integrating in the case θ is continuous) the unnormalized posterior density over the entire sample space $P(y) = \sum_\theta P(\theta)P(y|\theta)$.

2.1 Single Sensor: the Posterior Distribution

Consider the game where a fair coin is flipped an unknown number of times and the resulting number of heads is five. Intuition would suggest that the coin was flipped about ten times, but some variation on this would be expected. Observing only the number of heads from an unknown number of coin flips is the same problem as having observed a certain number of targets from an unknown number of total targets. Assuming a probability of detection p for a given sensor which has detected m targets, the distribution of the total number of targets is Binomial(n, p) where $n = m + r$ and r represents the number of remaining targets.

In order to derive a closed form posterior distribution, an improper (infinite), uniform prior distribution on the number of targets in the area is chosen, $n \sim \texttt{DiscreteUniform}(0, \infty)$. This can be thought of as $\lim_{N\to\infty} \frac{1}{N+1}$ `for` $n = 0, 1, 2, \ldots$. Computationally, $\texttt{DiscreteUniform}(0, N)$ for some large N would give similar results as the improper prior. In either case, as N does not depend on n, this term can be considered as the constant of normalization. Using (3) to estimate the unknown n based on observed m, the posterior distribution for n given m is:

$$\begin{aligned} P(n|m) &\propto P(n)P(m|n) && (4)\\ &\propto P(m|n) && (5)\\ &\propto \binom{n}{m} p^m (1-p)^{(n-m)} && (6)\\ &\propto \binom{n}{m} p^{m+1} (1-p)^{(n-m)}, && (7) \end{aligned}$$

where (7) follows from the fact that p is a constant with respect to n. This is the kernel of the $\texttt{NegativeBinomial}(m+1, p)$ for $n = m, m+1, \ldots$, and putting this into the context of remaining targets $r = n - m$, $r \sim \texttt{NegativeBinomial}(m+1, p)$ for $r = 0, 1, 2, \ldots$.

Thus, assuming the improper uniform prior, given the observation of five heads in an unknown number of coin flips of a fair coin, the number of tails follows a $\texttt{NegativeBinomial}(6, 0.5)$ distribution.

2.2 Multiple Sensors

Assume now that for each sensor $i \in 1, \ldots, T$ working in some sub-area, there is an associated number of detected targets m_i and probability of detection p_i. To give some intuition to the problem, suppose we have two weighted coins, coin one with associated probability p_1 of landing on heads, and coin two with probability p_2. The same game can be constructed as in 2.1. For example, given $p_1 = 0.25$ and $p_2 = 0.75$ where we observe five heads from coin one and none from coin two, we can ask the number of total flips from both coins, as well as the number of likely flips from coin one and coin two.

In the context of this example, the combinatorial nature of the sample space is apparent. In the case of a single coin, we were concerned only with the number of flips. In the case of two coins, we have several ways that ten flips may have occurred: ten flips of coin one, nine of coin one and one of coin two, and so on. In fact, for k coins and n flips, there are $\binom{n+k-1}{k-1}$ ways this may have been observed [1].

For multiple sensors, we require that the estimation of remaining targets be the same as in 2.1 when the probability of detection and total targets detected does not change (i.e. the resulting overall distribution for one sensor with $p = 0.8$ and $m = 10$ is the same as for 10 sensors, each with $p_i = 0.8$ and $m_i = 1$). This assumption means that the total number of targets needs to be viewed as a single entity. For example, a mine field can be viewed not as a collection of individual mines, but as a single deterrent with a common goal of blocking transit. This assumption plays an important role in the derivation of the posterior distribution.

The derivation of the posterior distribution is divided into the calculation of the likelihood function (Section 2.3) and of the prior distribution (Section 2.4). The resulting posterior distribution is then given in 2.5.

2.3 The Likelihood Function

The likelihood function is a straightforward generalization of the single sensor case. Given the number of targets detected in each of the T sensors's sub-areas $\mathbf{m} = (m_1, \ldots, m_T)$ and the probabilities of detection for each sensor $\mathbf{p} = (p_1, \ldots, p_T)$, we determine the likelihood function for the total targets in the areas of $\mathbf{n} = (n_1, \ldots, n_T)$. Each p_i is the probability of detecting any given target in the sub-area. Then for each sub-area, we determine the likelihood of having observed m_i targets given that n_i targets are actually in

the sub-area. As the probability depends only on the sub area, information in each sub-area is independent. Thus, the likelihood of the data $\mathbf{m}$ given $\mathbf{n}$ and $\mathbf{p}$ is given by:

$$P(\mathbf{m}|\mathbf{n},\mathbf{p}) = \prod_{i=1}^{T} \binom{n_i}{m_i} p_i^{m_i}(1-p_i)^{(n_i-m_i)}. \quad (8)$$

2.4 The Prior Distribution

In order to obtain the same result as in 2.1, the same infinite uniform prior on the total number of targets $n = \sum_{i=1}^{T} n_i$ must be assumed. This is not the same as assuming a uniform prior distribution on the entire sample space. For example, consider the idea of flipping two fair coins a number of times and observing only information about the number of heads observed from each coin. Our prior distribution must be defined for $n = 0, 1, 2, \ldots$, i.e. for zero total flips, one total flip, two total flips, etc. From two coins, there is one way to have no flips, two ways to have one flip, three ways to have two flips, and $n+1$ ways to have n flips. If all of these outcomes have the same prior weight, more weight is assigned to three total flips than to two total flips because there are more ways that more flips could be realized. For k coins the situation would be even more noticeable as there are $\binom{n+k-1}{k-1}$ ways of realizing n flips.

From this discussion it becomes apparent that the prior distribution must be defined on $\mathbf{n} = (n_1, \ldots, n_T)$ and $n = \sum_{i=1}^{T} n_i$ must follow a uniform distribution to ensure the same results as 2.1 in the case where sensor performance and total targets detected remains constant. The prior distribution is now expressed as the product of the distribution on total targets and the distribution of the targets between sub-areas given the total number of targets. As $P(n)$ does not depend on n and is therefore part of the constant of normalization, we then see that the choice of prior is really a choice of the distribution of targets among sub-areas $\mathbf{n} = (n_1, \ldots, n_T)$ for a fixed number of total targets $n = \sum_{i=1}^{T} n_i$. I.e.,

$$P(\mathbf{n}) = P(n)P(\mathbf{n}|n) \quad (9)$$
$$\propto P(\mathbf{n}|n). \quad (10)$$

The first choice of prior on the $\mathbf{n} = (n_1, \ldots, n_T)$ for a fixed $n = \sum_{i=1}^{T} n_i$ is the Multinomial Distribution (2.4.1). The posterior distribution resulting from this choice is shown to depend on the observed m_i only via their sum $M = \sum_{i=1}^{T} m_i$. A hierarchical model is then considered for $P(\mathbf{n}|n)$ (2.4.3).

2.4.1 Multinomial Distribution

To determine the prior distribution we require a distribution on the on the $\mathbf{n} = (n_1, \ldots, n_T)$ for a fixed $n = \sum_{i=1}^{T} n_i$. Given $\mathbf{x} = (x_1, \ldots, x_T)$ where each x_i represents an assumed a priori probability that any given target is in the sub-area of sensor i, a first approach is to consider a $\texttt{Multinomial}(\mathbf{x}, \mathbf{n})$ distribution with pdf

$$f(\mathbf{n}|\mathbf{x}) = \frac{n!}{\prod_{i=1}^{T} n_i!} \prod_{i=1}^{T} x_i^{n_i}. \quad (11)$$

In order to understand the implications of this choice, the resulting posterior distribution on remaining mines must be examined. Equation 3 is applied to the unknown $\mathbf{n}$ for observed data $\mathbf{m}$. Combining equations 8, 10 and 11 the posterior distribution is:

$$P(\mathbf{n}|\mathbf{m},\mathbf{p})$$
$$\propto P(\mathbf{m}|\mathbf{n})P(\mathbf{n}) \quad (12)$$
$$\propto \prod_{i=1}^{T} \binom{n_i}{m_i} p_i^{m_i}(1-p_i)^{(n_i-m_i)} \cdot \cdot \frac{n!}{\prod_{i=1}^{T} n_i!} \prod_{i=1}^{T} x_i^{n_i} \quad (13)$$
$$\propto n! \prod_{i=1}^{T} \left[\frac{1}{(n_i-m_i)!} (1-p_i)^{(n_i-m_i)} x_i^{n_i} \right], \quad (14)$$

by cancelling $\prod_{i=1}^{T} n_i!$ and removing all terms which are constant with respect to $\mathbf{n}$, i.e depend only on $\mathbf{m}$ and $\mathbf{p}$. In terms of remaining targets $\mathbf{r}$ for $r_i = n_i - m_i$ we have:

$$P(\mathbf{r}|\mathbf{m},\mathbf{p}) \propto n! \prod_{i=1}^{T} \left[\frac{1}{r_i!} (1-p_i)^{r_i} x_i^{r_i} \right], \quad (15)$$

since $x_i^{m_i}$ is also constant. What is interesting about the posterior distribution resulting from the multinomial prior on $\mathbf{n}|n$ is that the observed m_i's have completely disappeared in all but the $n! = (\sum_{i=1}^{T}(m_i+r_i))!$ term. This happened because the term $\prod_{i=1}^{T}(m_i+r_i)!$ was cancelled from the numerator of the likelihood function and the denominator of the prior distribution. This cancellation of terms left the posterior distribution dependent on the observed m_i's only by their sum $M = \sum_{i=1}^{T} m_i$. This means that there is no distinction between the case where we flip two coins and observe five heads from coin one and zero from coin two, and the case where we observe zero from coin one and five from coin two, as well as the case of two from one and three from the other. In this case, the probabilities x_i are the only term influencing the estimation of remaining targets by type. As the x_i are unknown and only an estimate, we would like to find an approach which allows the data to influence the final answers. That is, we would like the prediction of coin flips from coin one and coin two to depend on the number of observed heads from coin one and coin two.

In order to do this, a mixture model is proposed. The concept of a mixture model is first introduced using the beta-binomial mixture distribution for a two sensor case in 2.4.2 and this is then developed for the multiple sensor case using the Dirichlet-Multinomial mixture model in 2.4.3.

2.4.2 The beta-binomial mixture distribution

A random variable Y is *mixture distribution* if the distribution of Y depends on a quantity which also has

a distribution [3, 6]. Mixture distributions are also referred to as hierarchical models. An example of a mixture distribution is the Beta-Binomial distribution where

$$Y\big|P \sim \texttt{Binomial}(n, P)$$
$$P \sim \texttt{Beta}(\alpha, \beta)$$

and the Beta distribution is defined by:

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)\Gamma(\alpha)} x^{\alpha-1}(1-x)^{\beta-1} \tag{16}$$

for $\alpha > 0, \beta > 0, 0 \le x \le 1$.

The marginal distribution of Y over the joint distribution of (Y, P) can then be computed. Here, the binomial distribution is being used to allocate the targets to the sub-areas of sensor one and sensor two. Total targets are represented by n, y is the number of targets in the sub-area of sensor one, and p is the probability that a target is in sub-area one. As p is unknown, we consider it to be a random variable P and give P a $\texttt{Beta}(\alpha, \beta)$ distribution for some chosen α, β, i.e. the $f(x|\alpha,\beta)$ from equation 16 is substituted for $f(p)$ in equation 19. The marginal distribution of Y is then:

$$\begin{aligned}
&P(Y=y)\\
&= P(Y=y, 0\le P\le 1) && (17)\\
&= \int_0^1 f(y,p)\mathrm{dp} && (18)\\
&= \int_0^1 f(y|p)f(p)\mathrm{dp} && (19)\\
&= \int_0^1 [\binom{n}{y} p^y (1-p)^{n-y}]\cdot\\
&\quad [\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{(\alpha-1)}(1-p)^{(\beta-1)}]\mathrm{dp} && (20)\\
&= \binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot\\
&\quad \int_0^1 p^{(y+\alpha-1)}(1-p)^{(n-y+\beta-1)}\mathrm{dp} && (21)\\
&= \binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}, && (22)
\end{aligned}$$

since the integrand is the kernel of the $\texttt{Beta}(y+\alpha, n-y+\beta)$ pdf and therefore the integral must integrate to the reciprocal of the normalizing constant.

The Beta-Binomial mixture was shown using parameters more familiar to the Binomial, n,m and p, and Beta, α and β, distributions. Putting this back into the notation of multiple sensors with an index i, for the prior distribution on $P(n_1, n_2|n)$ we substitute $n_1 = y$ and $n_2 = n - y$ the beta-binomial mixture above, i.e. $n_1, n_2\big|n \sim \texttt{Beta}(n_1+\alpha, n_2+\beta)$. Although the two parameters α and β must still be chosen (as the x_i in 2.4.1), they provide much more flexibility than the x_i. For example, $\alpha = \beta = 1$ simplifies to $\frac{1}{n+1}$. This means that for a fixed n the prior distribution is uniform on the sample space as for each n we have $n+1$ combinations and each combination is given equal weight. For $\alpha = \beta = 1$, there is no assumption about the relative densities for the two sensors. The data determine the posterior distribution. If there is good prior information to determine the relative densities for the sub-areas, α and β can be chosen so that they will have more of an impact on the final answers. Large α, β will have a larger impact on the final answers and smaller values will have less influence on the final answers.

2.4.3 A Dirichlet-Multinomial Mixture

The same idea can be applied to a Dirichlet-Multinomial mixture for T sensors. We start with the Multinomial distribution described in 2.4.1. Then, we consider the x_i's not as fixed parameters, but as random variables. The $\texttt{Dirichlet}(\alpha_1, \ldots, \alpha_T)$ distribution for $\mathbf{X} = (X_1, \ldots, X_T)$ is defined as:

$$f(\mathbf{x}|\boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^T \alpha_i)}{\prod_{i=1}^T \Gamma(\alpha_i)} \prod_{i=1}^T x_i^{\alpha_i - 1}, \tag{23}$$

for $\alpha_i > 0$, $0 < x_i < 1$ and $\sum_{i=1}^T x_i = 1$ (see [5, 7]).

Using this in the same way as the Beta Distribution was used with the Binomial Distribution, the Dirichlet-Multinomial hierarchical model is of the form:

$$\begin{aligned}
&N_1, N_2, \ldots, N_T\big|X_1, X_2, \ldots, X_T \sim\\
&\qquad \texttt{Multinomial}(n_1, \ldots, n_T, x_1, \ldots, x_T)\\
&X_1, X_2, \ldots, X_T \sim\\
&\qquad \texttt{Dirichlet}(\alpha_1, \ldots, \alpha_T).
\end{aligned}$$

As in 2.4.2, we compute the marginal distribution of $(N_1, \ldots, N_T)$ by conditioning on $(X_1, \ldots, X_T)$:

$$\begin{aligned}
&P(\mathbf{n}|n)\\
&= \int_{x_1=0}^1 \cdots \int_{x_{T-1}=0}^{1-\sum_{i=1}^{T-2} x_i} f(\mathbf{n}, \mathbf{x}) dx_1 \cdots dx_{T-1} && (24)\\
&= \int_{x_1=0}^1 \cdots \int_{x_{T-1}=0}^{1-\sum_{i=1}^{T-2} x_i} f(\mathbf{n}|\mathbf{x}) f(\mathbf{x}) dx_1 \cdots dx_{T-1} && (25)\\
&= \int_{x_1=0}^1 \cdots \int_{x_{T-1}=0}^{1-\sum_{i=1}^{T-2} x_i} \frac{n!}{n_1! \cdots n_T!} x_1^{n_1} \cdots x_T^{n_T}\\
&\quad \frac{\Gamma(\sum_{i=1}^T \alpha_i)}{\prod_{i=1}^T \Gamma(\alpha_i)} \prod_{i=1}^T x_i^{\alpha_i - 1} dx_1 \cdots dx_{T-1} && (26)\\
&= \frac{n!}{n_1! \cdots n_T!} \frac{\Gamma(\sum_{i=1}^T \alpha_i)}{\prod_{i=1}^T \Gamma(\alpha_i)}\\
&\quad \int_{x_1=0}^1 \cdots \int_{x_{T-1}=0}^{1-\sum_{i=1}^{T-2} x_i} x_i^{n_i+\alpha_i-1} dx_1 \cdots dx_{T-1} && (27)\\
&= \frac{n!}{n_1! \cdots n_T!} \frac{\Gamma(\sum_{i=1}^T \alpha_i)}{\prod_{i=1}^T \Gamma(\alpha_i)} \frac{\prod_{i=1}^T \Gamma(n_i+\alpha_i)}{\Gamma(\sum_{i=1}^T (n_i+\alpha_i))} && (28)
\end{aligned}$$

again using the fact that $\prod_{i=1}^T x_i^{n_i+\alpha_i-1}$ the kernel of a $\texttt{Dirichlet}(\alpha_1+n_1, \ldots, \alpha_T+n_T)$ pdf, and therefore the integral must be equal to the reciprocal of the normalization constant. Note that $n = \sum_{i=1}^T n_i$ has also been used to simplify the notation.

As were α and β in the Beta-Binomial mixture, the the α_i are fixed parameters. Choosing all $\alpha_i = 1$ has the same effect of giving uniform weight to all the combinations of the n_i for a fixed n. Large α_i will have more impact on the posterior distribution than small values of α_i.

2.5 The Posterior Distribution

Combining equations 8 and 28 using 3, the posterior distribution on $n_{1,1}, \ldots, n_T$ given the m_i,p_i is then defined by:

$$\begin{aligned} P(\mathbf{n}\big|\mathbf{m},\mathbf{p}) & \\ &\propto P(\mathbf{m}\big|\mathbf{n})P(\mathbf{n}) && (29)\\ &\propto \prod_{i=1}^{T}\binom{n_i}{m_i}p_i^{m_i}(1-p_i)^{(n_i-m_i)}\cdot \\ &\quad \cdot\frac{n!}{\prod_{i=1}^{T} n_i!}\frac{\Gamma(\sum_{i=1}^{T}\alpha_i)}{\prod_{i=1}^{T}\Gamma(\alpha_i)}\cdot\frac{\prod_{i=1}^{T}\Gamma(n_i+\alpha_i)}{\Gamma(n+\sum_{i=1}^{T}\alpha_i)} && (30)\end{aligned}$$

As with the single sensor case we consider the posterior distribution on the targets remaining $r_i = n_i - m_i$ for $i = 1, \ldots, T$:

$$\begin{aligned} P(\mathbf{r}\big|\mathbf{m},\mathbf{p}) & \\ &\propto \prod_{i=1}^{T}\frac{\Gamma(r_i +_m i+1)}{\Gamma(m_i+1)\Gamma(r_i+1)}p_i^{m_i}(1-p_i)^{r_i}\cdot \\ &\quad \cdot\frac{\Gamma((\sum_{i=1}^{T} r_i+m_i)+1)}{\prod_{i=1}^{T}\Gamma(r_i+m_i+1)}\frac{\Gamma(\sum_{i=1}^{T}\alpha_i)}{\prod_{i=1}^{T}\Gamma(\alpha_i)}\cdot \\ &\quad \frac{\prod_{i=1}^{T}\Gamma(m_i+r_i+\alpha_i)}{\Gamma(\sum_{i=1}^{T} r_i+m_i+\sum_{i=1}^{T}\alpha_i)}, && (31)\end{aligned}$$

using the fact that $x! = \Gamma(x+1)$.

Cancelling terms and removing constant terms with respect to $\mathbf{r}$,

$$\begin{aligned} P(\mathbf{r}\big|\mathbf{m},\mathbf{p}) & \\ &\propto \frac{\Gamma(1+M+\sum_{i=1}^{T} r_i)}{\Gamma(\Lambda+M+\sum_{i=1}^{T} r_i)}\cdot \\ &\quad \prod_{i=1}^{T}\left({q_i}^{r_i}\frac{\Gamma(m_i+r_i+\alpha_i)}{\Gamma(r_i+1)}\right) && (32)\end{aligned}$$

where $\Lambda = \sum_{i=1}^{T}\alpha_i$, $M = \sum_{i=1}^{T} m_i$, and $q_i = 1-p_i$. As the individual m_i's are still part of the resulting posterior, rather than just the M as in 15, the distribution on targets remaining in the overall area and sub-areas will depend on the observed m_i. This means that five heads resulting from coin one and zero from coin two will give a different prediction on total flips from coin one than would zero heads from coin one and five from coin two, even if the overall distribution of total flips would be the same for fair coins.

3 Expectation Calculations

The goal of the derivation above is to calculate various measures of effectiveness which can be represented as expected values with respect to distribution 32. Of particular interest are the two measures of effectiveness:

- The expected number of remaining targets,

and

- The threat, or probability of damage, to traffic transiting the area.

The first quantity can be expressed, using the linearity of the expectation operator, by

$$E\left(\sum_{i=1}^{T} r_i\right) = \sum_{i=1}^{T} E\left(r_i\right). \qquad (33)$$

For the second quantity, define a to be the probability that a transiting vessel will not be damaged by a single target. Then the probability of not being damaged by r targets is a^r. Thus, in a single area the probability of safe transit is $1 - a^r$. Assuming a transiting vessel must pass each of T areas safely given probability of safe transit a_i for a single target in each sub-area and r_i targets in each sub-area, the threat given $\mathbf{r} = (r_1, \ldots, r_T)$ remaining targets is $\prod_{i=1}^{T} {a_i}^{r_i}$. Defining D to be the event that a ship transiting the minefield is damaged by any target,

$$P(\mathrm{D}\big|\mathbf{R}=\mathbf{r}) = 1 - \prod_{i=1}^{T} a_i^{r_i} \qquad (34)$$

The expected value of this probability over $\mathbf{R} = (R_1, \ldots, R_T)$ is then:

$$\begin{aligned} P(\mathrm{D}) &= \sum_{r_1=0}^{\infty}\cdots\sum_{r_T=0}^{\infty} P(\mathrm{D}\big|\mathbf{R}=\mathbf{r})P(\mathbf{R}=\mathbf{r}) && (35)\\ &= \sum_{r_1=0}^{\infty}\cdots\sum_{r_T=0}^{\infty} P(\mathrm{D}\big|\mathbf{r})P(\mathbf{r}\big|\mathbf{m},\mathbf{p}) && (36)\\ &= \mathtt{E}(P(\mathrm{D}\big|\mathbf{R}=\mathbf{r})) && (37)\\ &= \mathtt{E}(1-\prod_{i=1}^{T} a_i^{r_i}) && (38)\\ &= 1-\mathtt{E}(\prod_{i=1}^{T} a_i^{r_i}). && (39)\end{aligned}$$

In both cases, the desired measure of effectiveness can be represented in terms of the expected values, with respect to distribution 32, of separable functions.

4 Implementation of Expectation Calculations

In order to simplify the presentation, define

$$K(\mathbf{r}) = \frac{\Gamma\left(1+M+\sum_{i=1}^{T} r_i\right)}{\Gamma\left(\Lambda+M+\sum_{i=1}^{T} r_i\right)}$$

and

$$h_i(r) = {q_i}^r \frac{\Gamma(r+m_i+\alpha_i)}{\Gamma(r+1)}.$$

Assume, here and below, that the function $g : N^T \to R$ is separable and can be written in the form

$$g(\mathbf{r}) = g_1(r_1)\, g_2(r_2) \cdots g_T(r_T) \tag{40}$$

Finally, define

$$\begin{aligned} S_N(g) &= \sum_{r_1=0}^{N} \cdots \sum_{r_T=0}^{N} K(\mathbf{r}) \prod_{i=1}^{T} h_i(r_i) g_i(r_i) && (41) \\ &= \sum_{r_1=0}^{N} h_1(r_1) g_1(r_1) \cdots \\ &\qquad \sum_{r_T=0}^{N} h_T(r_T) g_T(r_T) \cdot K(\mathbf{r}) && (42) \end{aligned}$$

and

$$S(g) = \lim_{N\to\infty} S_N(g). \tag{43}$$

By direct comparison of equation 32 and 41 we see that the normalization constant for distribution 32 is S(1), and that the expected value of separable function g with respect to distribution 32 is given by $E(g) = S(g)/S(1)$.

Since a closed form expression for $S(g)$ is unlikely to exist, we assume that N has been chosen large enough so that $S_N(g)$ is a sufficient close to $S(g)$ and discuss the calculation of $S_N(g)$. From 41 we see that the obvious method of computing $S_N(g)$ has complexity $O\left(N^T\right)$ which severely limits the applicability of distribution 41. However, this limitation can be overcome.

The difficulty in calculating $S_N(g)$ is caused by inseparability of the kernel function K. However, K still has a particularly nice structure. Specifically, if we define

$$\mathbf{k}(x) = \frac{\Gamma(1+M+x)}{\Gamma(\Lambda+M+x)} \tag{44}$$

with $x \in \{0, 1, \ldots, NT\}$, then

$$K(\mathbf{r}) = \mathbf{k}\left(\sum_{i=1}^{T} r_i\right). \tag{45}$$

Representing $\mathbf{k}$ by its discrete Fourier expansion, we obtain

$$\mathbf{k}(x) = \sum_{j=0}^{NT} c_j b_j^x \tag{46}$$

where

$$\begin{aligned} b_j &= \frac{\exp(2\pi j\ \mathbf{i})}{\sqrt{NT+1}} && (47) \\ c_j &= \sum_{n=0}^{NT} \mathbf{k}(n)\ \overline{b}_j^n. && (48) \end{aligned}$$

Using this expansion to represent K, we have

$$\begin{aligned} K(\mathbf{r}) &= \mathbf{k}\left(\sum_{i=1}^{T} r_i\right) && (49) \\ &= \sum_{j=0}^{NT} c_j b_j^{(r_1+r_2+\cdots+r_T)} && (50) \\ &= \sum_{j=0}^{NT} \left(c_j \prod_{k=1}^{T} b_j^{r_k}\right). && (51) \end{aligned}$$

Thus, while K is inseparable, it is "nearly separable" in the sense that it can be represented as a finite sum of separable functions.

Inserting 51 into the equation 42 we get, after some simplification,

$$\begin{aligned} S_N(g) &= \sum_{j=0}^{NT} c_j \sum_{r_1=0}^{N} b_j^{r_1} h_1(r_1)\, g_1(r_1) \\ &\qquad \cdots \sum_{r_T=0}^{N} b_j^{r_T} h_T(r_T)\, g_T(r_T) && (52) \\ &= \sum_{j=0}^{NT} c_j \left(\prod_{k=1}^{T} \left(\sum_{r_k=0}^{N} b_j^{r_k}\ g_k(r_k)\ \ h_k(r_k)\right)\right). \end{aligned}$$

The computational complexity of calculating $S_N(g)$ using expression 52 is $O\left(N^2T^2\right)$. This provides remarkable speed-up in computation time and greatly increases the applicability of distribution 32.

5 Summary

This paper presents a method of estimating the number of targets remaining in an area, and the threat to transiting traffic, after an attempt has been made to detect an overall, unknown number of targets. The method is based on expectation calculations with respect to a distribution which was derived using a Dirichlet-Multinomial prior distribution and assumptions about the effectiveness of the original search attempt. The resulting calculations, in their canonical form, are computationally prohibitive. An alternative form of the calculations is presented which provides a mechanism for calculating the expectations in a negligible amount of time.

References

[1] R.C. Bose and B. Manvel, *Introduction To Combinatorial Theory*, John Wiley & Sons, 1984.

[2] I. Olkin, A.J. Petkau and J.V. Zidek, A Comparison of n Estimators for the Binomial Distribution, *Jornal of the American Statistical Association*, 76(375):637-642, Sept 1981.

[3] G. Casella and R.L. Berger, *Statistical Inference*, Duxbury Press, 1990.

[4] A. Gelman, J.B. Carlin, H.S. Stern and D.B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC, 1995.

[5] J.M. Bernardo and A.F.M. Smith, *Bayesian Theory*, John Wiley & Sons, Ltd, 1994.

[6] L. Wasserman, *All of Statistics: A Concise Course in Satistical Inference*, Springer, 2004.

[7] N.L. Johnson, S. Kotz and N. Balakrishnan, *Discrete Multivariate Distributions*, John Wiley & Sons, Inc, 1997.

Intentionally blank page

Document Data Sheet

Security Classification		Project No.
RELEASABLE TO THE PUBLIC		
Document Serial No.	**Date of Issue**	**Total Pages**
NURC-PR-2006-005	August 2006	11 pp.

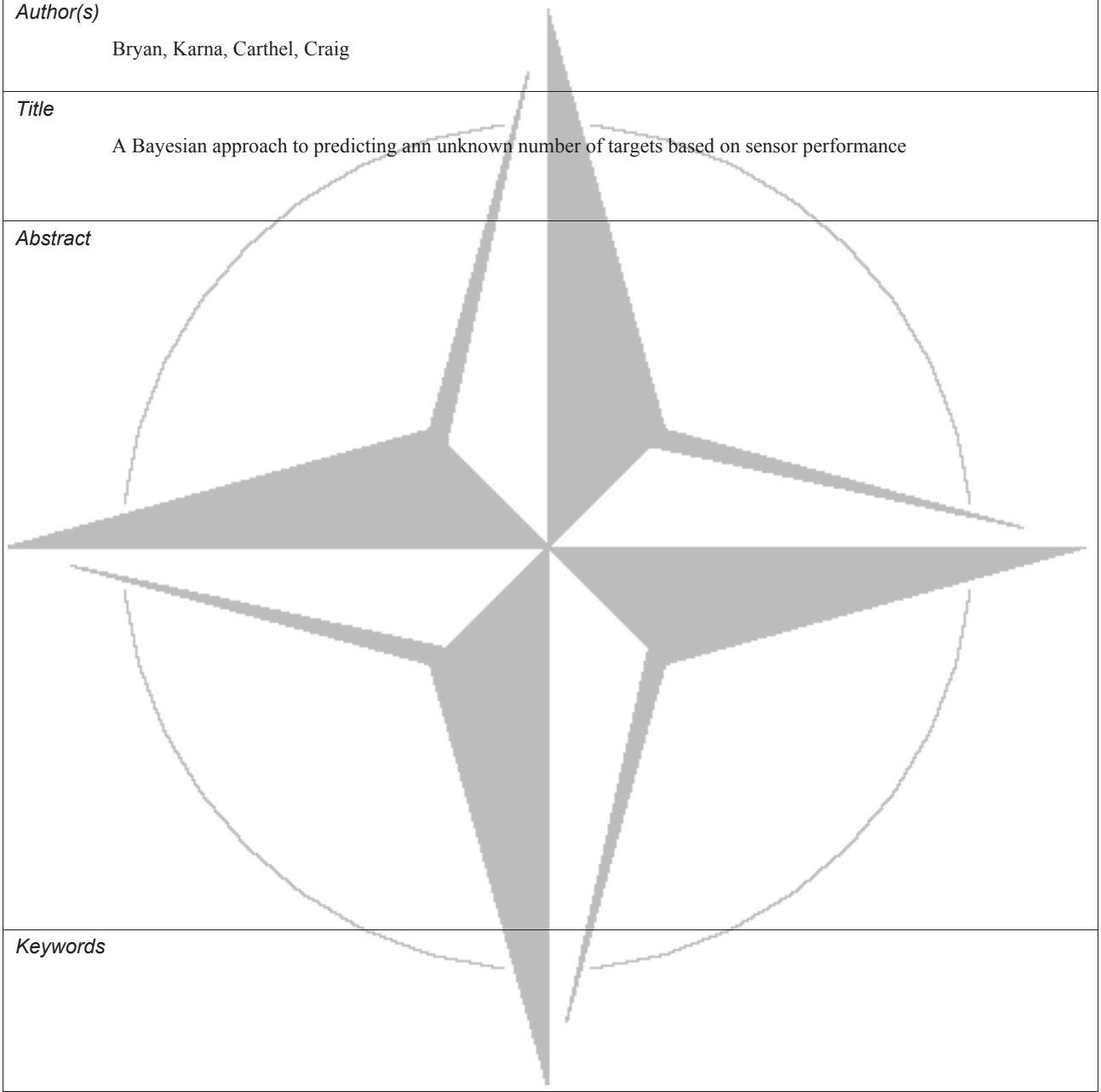
Author(s)

Bryan, Karna, Carthel, Craig

Title

A Bayesian approach to predicting ann unknown number of targets based on sensor performance

Abstract



Keywords

Issuing Organization

NATO Undersea Research Centre
Viale San Bartolomeo 400, 19138 La Spezia, Italy

[From N. America:
NATO Undersea Research Centre
(New York) APO AE 09613-5000]

Tel: +39 0187 527 361
Fax:+39 0187 527 700

E-mail: library@nurc.nato.int