2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)
Allerton Park and Retreat Center
Monticello, IL, USA, September 24-27, 2019

A Theory of Uncertainty Variables for State Estimation and Inference

Rajat Talak, Sertac Karaman, and Eytan Modiano

***Abstract*—Probability theory forms an overarching framework for modeling uncertainty, and by extension, also in designing state estimation and inference algorithms. While it provides a good foundation to system modeling, analysis, and an understanding of the real world, its application to algorithm design suffers from computational intractability. In this work, we develop a new framework of uncertainty variables to model uncertainty. A simple uncertainty variable is characterized by an uncertainty set, in which its realization is bound to lie, while the conditional uncertainty is characterized by a set map, from a given realization of a variable to a set of possible realizations of another variable. We prove Bayes' law and the law of total probability equivalents for uncertainty variables. We define a notion of independence, conditional independence, and pairwise independence for a collection of uncertainty variables, and show that this new notion of independence preserves the properties of independence defined over random variables. We then develop a graphical model, namely Bayesian uncertainty network, a Bayesian network equivalent defined over a collection of uncertainty variables, and show that all the natural conditional independence properties, expected out of a Bayesian network, hold for the Bayesian uncertainty network. We also define the notion of point estimate, and show its relation with the maximum a posteriori estimate.**

I. INTRODUCTION

Probability theory, developed over the last three centuries, has provided an overarching framework for modeling uncertainty in the real-world. As a result, it has become a key mathematical tool used in designing state estimation and inference algorithms. While Pascal, Fermat, and Huygens were some of the first contributors to the development of probability theory, Pierre-Simon Laplace and Thomas Bayes were among the first to formulate the notion of conditional probability, and use it to estimate an unknown parameter from the observed data [1], [2]. Since its mathematical formulation, and more so after the axiomatic foundations laid by Kolmogorov [3], the theory of probability has formed the basis for inference and state estimation algorithms.

In Bayesian inference, for example, the goal is to successively improve an estimate of a model parameter or an evolving state variable, such as the pose of a robot [4], by incorporating the observed information [5], [6]. A prior probability distribution is assigned to the initial state variable or the model parameter, and this distribution is successively improved by computing the posteriori distribution, using the observed data. Bayes' law and the law of total probability form the theoretical basis for this computation.

One of the main difficulties in such state estimation and inference procedures is its computational tractability. Computing the posteriori distribution and the maximum a posteriori (MAP) estimate is hard in most problems of practical significance. Several approximation algorithms have been considered to overcome this limitation, and is an active field of research.

Probability theory, in its simplest form, characterizes an uncertain quantity by a distribution function, which assigns a weight over the set of all possible outcomes; more generally it defines a probability measure over the set of all possible outcomes. This distribution function, or the weight function, is often too much information to carry for computation, and as a result leads to computationally intractable solutions. The difficulty in computing the posteriori distribution is a manifestation of such intractability.

Another major issue with using distribution functions is that they are chosen mostly to ensure easier analysis and algorithm design. In robotic perception, for example, additive Gaussian noise is often assumed in the motion and sensing/observation model [4]. Although, this produces a simple Kalman filter like solution to the robotic perception problem, it can cause severe degradation in performance due to the non-linearities inherent in motion and sensing. In several such applications, and in robotic perception in particular, a bounded noise model is more appropriate.

A. Contribution

In this work, we provide an alternate framework of *uncertainty variables* to model uncertainty. A *simple* uncertainty variable X is characterized by an *uncertainty set* U_X. A realization of an uncertainty variable can lie only in its uncertainty set. Conditional uncertainty is characterized by a *conditional uncertainty map* $P_{Y|X} : x \rightarrow P_{Y|X}(x) \subset D_Y$, that maps every realization $x \in D_X$ of X to a subset of D_Y, which is the set of all realizations of Y, i.e. given a realization $X = x$, a realization of Y can only lie in the set $P_{Y|X}(x)$. Thus, the larger the set $P_{Y|X}(x) \subset D_Y$, the larger is the uncertainty in Y given $X = x$.

Using this notion of uncertainty variables and conditional uncertainty maps, we first prove Bayes' law and the law of total probability equivalents for uncertainty variables. We then define the notions of independence, conditional independence, and pairwise independence for a collection of uncertainty variables. We argue that this new notion of independence preserves the properties of independence

The authors are with the Laboratory for Information and Decision Systems (LIDS) at the Massachusetts Institute of Technology (MIT), Cambridge, MA. {talak, sertac, modiano}@mit.edu

This work was supported by NSF Grants AST-1547331, CNS-1713725, and CNS-1701964, and by Army Research Office (ARO) grant number W911NF- 17-1-0508.

defined over random variables. For example, we show that the independence between a collection of uncertainty variables does not imply pairwise independence.

Graphical models over random variables have been very useful in designing exact and approximate inference algorithms [7], [8]. We extend the theory of uncertainty variables, developed in the first part of the paper, to define a graphical model over uncertainty variables. We define *Bayesian uncertainty network*, as a directed graphical model over a collection of uncertainty variables. As the name suggests, this is equivalent to the Bayesian network defined over random variables. We show that all the conditional independence properties, expected out of a Bayesian network, also hold for the Bayesian uncertainty network.

In many state estimation and inference problems, one is interested in a point estimate. We, therefore, define the notion of a point estimate for a Bayesian uncertainty network. We prove that the point estimate equals the MAP estimate for a Bayesian network, under an appropriate representational relation between the Bayesian network and the Bayesian uncertainty network. This illustrates the generality of this new approach of characterizing uncertainty.

B. A Brief Literature Survey

Using bounded sets instead of probability distributions is not a new idea, and has been explored in the control systems literature [9]–[11]. Some of these early works on bounded noise models in control theory, also inspired the formulation of set-estimation in the signal processing literature [9], [12], [13]. The motivation here was that the point estimate, such as MAP or ML, is not good enough, and a confidence region, namely a set, would be useful. A set estimate, for say a model parameter, was defined as an intersection of sets, each of which, corresponded to an observation. To help compute such an intersection, especially of ellipsoidal sets, several approximating methods were proposed [14], [15].

A notion of uncertainty sets has also been used in the robust optimization literature [16], [17]. Robust optimization also begins with the same premise as ours, that the way probability theory characterizes uncertainty results in computational intractability. As a recourse, when many uncertain quantities are involved, robust optimization constructs uncertainty sets over these uncertain quantities, using the law of large numbers and the central limit theorems [18].

Our work is inspired, and yet different, from these earlier works in the literature. It uses uncertainty sets to develop an alternative framework for modeling uncertainty. The notion of conditional uncertainty maps is novel, and so is the fact that useful notions of probability, such as independence, can be extended to uncertainty variables.

C. Organization and Notations

In Section II, we develop the notion of uncertainty variables, and establish Bayes' law and the law of total probability. In Section III, we define independence between a collection of uncertainty variables. Bayesian uncertainty network and point estimates are discussed in Sections IV and V, respectively. We conclude in Section VI. Proof are omitted due to space constraints, and are available in an extended version in [19].

We use the following notation. For an indexed set A, X_A or β_A denotes the collection $(X_i)_{i\in A}$ and $(\beta_i)_{i\in A}$. We use $1:N$ to denote the set of integers $\{1, 2, \dots N\}$. Uncertainty variables are usually denoted by X, Y, and Z, while random variables are denoted by $\bar{X}$, $\bar{Y}$, and $\bar{Z}$.

II. Theory of Uncertainty Variables

Suppose, we want to characterize an uncertain quantity such as the temperature in a room, or the position of a robot, or an atmospheric condition measured by several variables. All such uncertain quantities have an implicitly defined underlying domain. For example, a temperature measurement can take any real values, a pose of a robot is a point in a d dimensional space of poses. Moreover, any such uncertain quantity is more likely to lie in certain region of this domain, and not spread out everywhere.

Motivated by this, we define the notion of an uncertainty variable and a simple uncertainty variable. It is the simple uncertainty variable that will be of interest to us.

Definition *1:* An *uncertainty variable* (UV) is defined as a tuple $X = (D_X, P_X)$, where D_X is the domain of the UV and P_X is uncertainty map, that maps every point in D_X to a subset of D_X:

$$P_X : D_X \to 2^{D_X}.$$

We say that an UV X is *simple* if for all $x \in D_X$, either $P_X(x) = \{x\}$ or $P_X(x) = \emptyset$.

A simple UV can always be represented by an *uncertainty set*:

$$U_X = \bigcup_{x\in D_X} P_X(x).$$

Simple or not, any realization of the UV X lies in the uncertainty set U_X. For ease of notation, we will use $X = (D_X, U_X)$ to denote a simple UV. We provide the definitions and proofs for UVs, in most generality, in Appendix **??**. Here, we summarize some of the results for the specific case of simple uncertainty variables.

Consider the following examples of simple uncertainty variables:

1) *Elliptic UV*: An Elliptic UV is defined as

$$X = \left(\mathbb{R}^n, \{x \in \mathbb{R}^n \mid (x-\bar{x})^T Q^{-1}(x-\bar{x}) \leq \eta\}\right), \quad (1)$$

where Q is a positive definite matrix and $\bar{x}$ is a vector in $D_X = \mathbb{R}^n$. This UV can be used to model noisy measurement of a location $\bar{x}$.

2) *Polytopic UV*: A polytopic UV is defined as

$$X = (\mathbb{R}^n, \{x \in \mathbb{R}^n \mid H(x-\bar{x}) \leq h\}), \quad (2)$$

where H is a matrix, and h and $\bar{x}$ are vectors in $\mathbb{R}^n$.

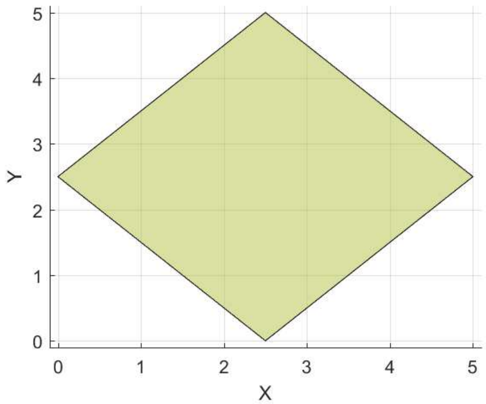

Fig. 1. Plots joint uncertainity set $U_{X,Y}$ of two UVs that are not independent.

3) *Canonical UV*: For every random variable $\bar{X}$, taking values in D_X with a probability density function $f_{\bar{X}}(x)$, we can construct a simple canonical UV $X = (D_X, U_X)$. We call it the canonical UV, canonical to the random variable $\bar{X}$. The simple canonical UV X is given by

$$X = (D_X, \{x \in D_X \mid -\log f_{\bar{X}}(x) \leq \eta\}), \quad (3)$$

for some $\eta \in \mathbb{R}$. Note that the Elliptic UV in (1) is a Canonical UV for the Gaussian random variable $\mathcal{N}(\bar{x}, Q)$ and the polytopic UV in (2) is a Canonical UV for a uniformly distributed random variable over the polytope.

A joint simple uncertainty variable (X, Y) can similarly be defined with a domain $D_X \times D_Y$, an uncertainty map $P_{X,Y}$, and an uncertainty set $U_{X,Y}$, where

$$P_{X,Y}(x,y) = \begin{cases} (x,y) & \text{if } (x,y) \in U_{X,Y} \\ \emptyset & \text{if } (x,y) \notin U_{X,Y} \end{cases} \quad (4)$$

We now state an equivalent of the *law of total probability*.

Theorem *1:* The marginal uncertainty sets U_X and U_Y are given by the projections of $U_{X,Y}$:

$$U_X = \Pi_X(U_{X,Y}) \quad \text{and} \quad U_Y = \Pi_Y(U_{X,Y}). \quad (5)$$

Proof: See [19]. ■

It is not necessarily true that the joint uncertainty set $U_{X,Y}$ is a cross product of the marginal sets U_X and U_Y. In Figure 1, we have plotted the joint uncertainty set $U_{X,Y}$ of two variables. Here, $U_X = U_Y = [0, 5]$, but $U_{X,Y} \neq [0, 5] \times [0, 5]$. We, therefore, define a conditional uncertainty map.

Definition *2:* Let X and Y be two UVs. The conditional uncertainty map of Y given X is a set function $P_{Y|X} : D_X \to 2^{D_Y}$ given by

$$P_{Y|X}(x) = \Pi_Y\left(U_{X,Y} \cap \{X = x\}\right),$$

for all $x \in D_X$, where $\{X = x\}$ denotes the set $\{(x', y') \in D_X \times D_Y \mid \text{s.t. } x' = x\}$.

The conditional uncertainty $P_{Y|X}$ maps each $x \in D_X$ to a set in the collection 2^{D_Y}. Unlike the marginal maps P_X and P_Y, the conditional uncertainty map $P_{Y|X}$ can map an $x \in D_X$ to any subset in D_Y. The larger the set $P_{Y|X}(x)$, the greater is the uncertainty in UV Y, given $X = x$.

A. Bayes' Rule

We now prove an equivalent of Bayes' rule. We first define a convenient operation between a set and a set function.

Definition *3:* For a set $A \subset D_X$ and a set function $B : D_X \to 2^{D_Y}$ we define a cross product $A \otimes B$ to be a set in $D_X \times D_Y$ given by

$$A \otimes B \triangleq \bigcup_{x' \in A} \{x'\} \times B(x').$$

The Bayes rule for the uncertainty variables is as follows.

Theorem *2:* For the joint UV $Z = (X, Y)$, the uncertainty set is given by

$$U_{X,Y} = U_X \otimes P_{Y|X} = \mathcal{R}_{YX \leftrightarrow XY}\left(U_Y \otimes P_{X|Y}\right), \quad (6)$$

where $\mathcal{R}_{YX \leftrightarrow XY}(y, x) = (x, y)$ for all $x \in D_X$ and $y \in D_Y$.

Proof: See [19]. ■

In the next section, we argue that the uncertainty sets and conditional uncertainty maps, can be represented as sub-level sets of some functions. This representation will be useful in proving some of the results later in the paper.

B. Representation

We have represented uncertainty variables and conditional uncertainties as either sets or set maps. It is, at times, useful to deal with functions rather than sets. In this small section, we present a result, that states that every such uncertainty set or a conditional uncertainty map can be represented as a sub-level set of a function.

Lemma *1:* The following statements are true:
1) An uncertainty set $U_X \subset D_X$ can be written as

$$U_X = \{x \in D_X \mid H_X(x) \leq h_X\}, \quad (7)$$

for some function $H_X : D_X \to \mathbb{R}^m$, some positive integer m, and $h_X \in \mathbb{R}^m$.
2) A conditional uncertainty map $P_{X|Y} : D_Y \to 2^{D_X}$ can be written as

$$P_{X|Y}(y) = \left\{x \in D_X \mid H_{X|Y}(x, y) \leq h_{X|Y}\right\}, \quad (8)$$

for some function $H_{X|Y} : D_X \times D_Y \rightarrow \mathbb{R}^m$, some positive integer m, and $h_{X|Y} \in \mathbb{R}^m$.

Proof: The proof is trivial, as such functions, namely H_X and $H_{X|Y}$, can always be obtained by a simple construction. For the first part, given a set $U_X \subset D_X$, take $H_X(x) = 1 - \mathbb{I}_{U_X}(x)$, for all $x \in D_X$. Here, $\mathbb{I}_{U_X}(x)$ is the indicator function for the set U_X. Take $m = 1$ and $h_X = 1/2$. Then, $U_X = \{x \in D_X \mid H_X(x) \leq h_X\}$. Similarly, for the second part, take $H_{X|Y}(x,y) = 1 - \mathbb{I}_{P_{X|Y}(y)}(x)$, $m = 1$, and $h_{X|Y} = 1/2$. ■

Note that we have not imposed any conditions on the functions H_X and $H_{X|Y}$ in Lemma 1, except that they take values in some Euclidean space $\mathbb{R}^m$.

In the next section, we illustrate the usefulness of the uncertainty variables in computing the posteriori distribution. The main tool is the application of the law of total probability (Theorem 1) and the Bayes' rule (Theorem 2) developed here.

C. Computing the Posteriori Map

The main advantage of this formulation is that it can be easier to compute the posteriori uncertainty map. For example, in many machine learning applications, we are given a model for the data, say Y, and a model for the prior parameters, say X. This is equivalent to knowing the conditional uncertainty map $P_{Y|X}$ and the uncertainty set U_X. With this, the joint uncertainty set can be computed as

$$U_{X,Y} = U_X \otimes P_{Y|X}. \tag{9}$$

Then, the posteriori uncertainty map $P_{X|Y} : D_Y \rightarrow 2^{D_X}$ can be computed by a simple projection on D_X:

$$P_{X|Y}(y) = \Pi_X \left(U_{X,Y} \cap \{Y = y\}\right). \tag{10}$$

This posteriori map, for a given observed data $Y = y$, will produce a set in D_X that tells us about the uncertainty in X.

Let us use the sub-level set representations of Lemma 1. Let $U_X = \{x \in D_X \mid H_X(x) \leq h_X\}$ and $P_{Y|X}(x) = \{y \in D_Y \mid H_{Y|X}(y,x) \leq h_{Y|X}\}$. Then the posteriori uncertainty map $P_{X|Y}(y)$, for a given observed data $Y = y$, is given by

$$P_{X|Y}(y) = \{x \in D_X \mid H_X(x) \leq h_X \text{ and } H_{Y|X}(y,x) \leq h_{Y|X} \}. \tag{11}$$

From (11), we see that this is nothing but the projection (on D_X) of intersection of two sub-level sets.

The idea of obtaining set-estimates, as intersection of sets, existed in the set-estimation literature [9], [12], [13]. However, the literature mostly limited itself to linear models, in which, the observed data Y and the underlying state variable X were related by a linear equation. The notion of uncertainty variables generalizes this idea to any such X and Y, and their relation may not be linear.

To see that there is indeed a difference between the posteriori distribution and the posteriori uncertainty map, in the sense that one cannot be trivially constructed from the other, consider the following example: Let $U_X = \{x \in D_X \mid -\log f_X(x) \leq \eta\}$ and $P_{Y|X}(x) = \{y \in D_Y \mid -\log f_{Y|X}(y|x) \leq \eta'\}$ be the canonical uncertainty sets and conditional uncertainty maps, for a given marginal and conditional density functions f_X and $f_{Y|X}$, respectively. Then, the joint uncertainty set is given by

$$U_{X,Y} = U_X \otimes P_{Y|X}, \tag{12}$$
$$= \{(x,y) \mid -\log f_X(x) \leq \eta \text{ and} \tag{13}$$
$$-\log f_{Y|X}(y|x) \leq \eta'\}, \tag{14}$$

and the posteriori uncertainty map is given by

$$P_{X|Y}(y) = \Pi_X \left(U_{X,Y} \cap \{Y = y\}\right), \tag{15}$$
$$= \{x \in D_X \mid -\log f_X(x) \leq \eta \text{ and} \tag{16}$$
$$-\log f_{Y|X}(y|x) \leq \eta'\}.$$

Note that this set is not same as $\{x \in D_X \mid -\log f_{X|Y}(x|y) \leq \eta''\}$, for some constant η''.

In the next section, we define the notion of independence and conditional independence for a given set of uncertainty variables. We show that all the independence properties that are true for random variables, such a total independence not implying pairwise independence and more, are retained for the uncertainty variables.

III. Independence

We first define independence between two simple uncertainty variables.

Definition *4:* We say that the two UVs, X and Y, are independent if $P_{Y|X}(x) = P_{Y|X}(x')$ for all $x, x' \in U_X$.

It is trivial to see that for independent uncertainty variables X and Y, the joint uncertainty set also factors into the product of the marginal uncertainty set. We articulate this in the following lemma.

Lemma *2:* Uncertainty variables X and Y are independent if and only if $U_{X,Y} = U_X \times U_Y$, where U_X, U_Y, and $U_{X,Y}$ are uncertainty sets for X, Y, and (X,Y), respectively.

Proof: This should be intuitively clear as if $P_{Y|X}(x) = P_{Y|X}(x')$ for all $x, x' \in U_X$, then we must have $P_{Y|X}(x) = U_Y$ for all $x \in U_X$. From this, the result can be inferred. The detailed arguments are given in [19]. ■

Conditional independence can be similarly defined. We do so in terms of factorization of the uncertainty maps.

Definition *5:* We say that the UVs $X = (D_X, U_X)$ and $Y = (D_Y, U_Y)$ are independent, given a UV $Z = (D_Z, U_Z)$, if

$$P_{X,Y|Z}(z) = P_{X|Z}(z) \times P_{Y|Z}(z), \tag{17}$$

for all $z \in U_Z$.

We will use the notation $X \perp\!\!\!\perp Y$ to denote that X and Y are independent, and $X \perp\!\!\!\perp Y|Z$ to denote that X and Y are conditionally independent, given Z.

When it comes to several uncertainty variables, defining independence is as tricky as it is for the random variables. However, it turns out that the independence and conditional independence properties that hold for random variables also hold for uncertainty variables. In Section IV, we will introduce Bayesian network models on a collection of uncertainty variables. We will see that the set of uncertainty variables preserve the conditional independence properties, which hold for the Bayesian network defined over random variables [7].

To provide a prelude, we define pairwise and total independence between a collection of uncertainty variables. In traditional probability, pairwise independence does not imply total independence between a collection of random variables. The same is true for the uncertainty variables. Let us first define pairwise and total independence for the uncertainty variables.

Definition *6:* A collection of uncertainty variables $X_{1:N}$ is said to be
1) pairwise independent if for each $i, j \in [N]$, $i \neq j$, we have

$$U_{X_i,X_j} = U_{X_i} \times U_{X_j}, \tag{18}$$

where U_{X_i}, U_{X_j}, and U_{X_i,X_j} are uncertainty sets for X_i, X_j, and (X_i, X_j), respectively.
2) totally independent if

$$U_{X_{1:N}} = \times_{i=1}^{N} U_{X_i}, \tag{19}$$

where $U_{X_{1:N}}$ and U_{X_i} are the uncertainty sets of $X_{1:N}$ and X_i, respectively.

In the following lemma, we prove that pairwise independence does not implies total independence.

Theorem *3:* If $X_{1:N}$ are totally independent then they are also pairwise independent, but the converse is not true.

Proof: (a) Let $X_{1:N}$ be totally independent uncertainty variables. Then we have $U_{X_{1:N}} = \times_{k=1}^{N} U_{X_k}$. Take $i, j \in [N]$ such that $i \neq j$. We know that the uncertainty set U_{X_i,X_j} of (X_i, X_j) is given by a simple projection of $U_{X_{1:N}}$ on (X_i, X_j). Therefore,

$$\begin{aligned} U_{X_i,X_j} &= \Pi_{(X_i,X_j)}\left(U_{X_{1:N}}\right), && (20)\\ &= \Pi_{(X_i,X_j)}\left(\times_{k=1}^{N} U_{X_k}\right), && (21)\\ &= U_{X_i} \times U_{X_j}. && (22) \end{aligned}$$

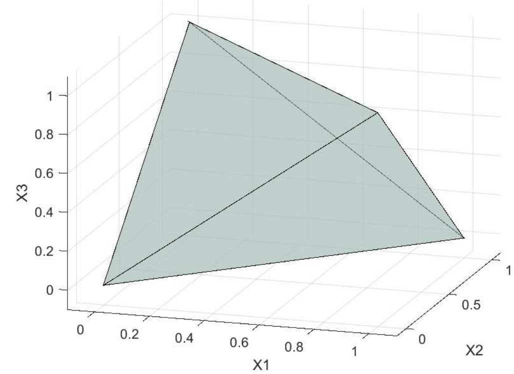

Fig. 2. Plot of the joint uncertainty set $U_{X_{1:3}}$ given by (23).

We assumed j and j to be any $i, j \in [N]$, $i \neq j$. Thus, $X_{1:N}$ is also pairwise independent.

(b) We prove that the converse is not true by constructing a counter-example. Take three uncertainty variables $X_{1:3}$ such that $U_{X_i} = [0, 1]$ and $U_{X_i,X_j} = [0, 1] \times [0, 1]$, for all $i, j \in [N]$ and $i \neq j$. However, the joint uncertainty set $U_{X_1,X_2,X_3} \neq [0, 1] \times [0, 1] \times [0, 1]$. Such a joint uncertainty set U_{X_1,X_2,X_3} is given by

$$U_{X_{1:3}} = \left\{ x_{1:3} \;\middle|\; \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \leq \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}, \tag{23}$$

which is shown in Figure 2. ■

In the next section, we define the Bayesian uncertainty network, in which we extend the concept of Bayesian network for random variables to a collection of uncertainty variables. We will see that the independence properties are preserved in this extension from random variables to uncertainty variables. In Section V, we will define a point estimate and show that in special cases, it reduces to the MAP estimate.

IV. Bayesian Uncertainty Networks

We now extend the notion of Bayeian network, defined for a collection of random variables, to a collection of uncertainty variables. We call it the Bayeian uncertainty network.

Let $G = (V, E)$ be a directed acyclic graph (DAG). For each node $i \in V$, let Pa_i denote the set of parents of node i, i.e. for each $j \in \text{Pa}_i$ there exists a link $(j, i) \in E$. A node $k \in V$ is said to be descendant of i if there exists a directed path from node i to node k in G. We use NonDes_i denote the set of nodes that are non-descendants of i. Also, we will use R to denote the set of all nodes that have no parents, i.e. $R = \{i \in V \mid \text{Pa}_i = \emptyset\}$. Typically, we would need to order the nodes in V in a sequence. A canonical ordering of nodes in V is an ordering such that parents are indexed before their children, i.e., for all $j \in \text{Pa}_i$, we have $j < i$. We know that such an ordering of nodes in a DAG is always possible.

A collection of uncertainty variables is characterized by its joint uncertainty set. We now formally define the notion of *Bayesian uncertainty network*, in which the uncertainty set of a collection of uncertainty variables factorizes according to an underlying DAG.

Definition *7:* A *Bayesian uncertainty network* is the tuple $\mathcal{BN} = (X_V, G)$ of uncertainty variables X_V and a DAG $G = (V, E)$, such that X_V *factorizes* according to G, namely, every node $i \in V$ is associated with a unique uncertainty variable X_i, and there exists conditional uncertainty maps

$$P_{X_i|X_{\text{Pa}_i}} : D_{\text{Pa}_i} \to 2^{D_i}, \tag{24}$$

for each $i \in V \backslash R$, such that, for any canonical ordering of nodes in V, the joint uncertainty set of X_V is given by

$$U_{X_V} = U_{X_R} \otimes P_{X_{|R|+1}|\text{Pa}_{|R|+1}} \otimes \cdots \otimes P_{X_{|V|}|\text{Pa}_{|V|}}, \tag{25}$$

where U_{X_R} is a simple cross product of U_{X_i}, over $i \in R$, namely

$$U_{X_R} = \times_{i \in R} U_{X_i}. \tag{26}$$

Note that the factorization in (25) is well defined, provided we ignore the ordering of variables in the tuple. To see this, let us make use of Lemma 1 in Section II-B. Let for each $i \in R, U_{X_i} = \{x_i \in D_i \mid H_i(x_i) \leq h_i\}$, and for all $i \in V \backslash R$ let

$$P_{X_i|X_{\text{Pa}_i}}(x_{\text{Pa}_i}) = \{x_i \in D_{X_i} \mid H_i(x_i, x_{\text{Pa}_i}) \leq h_i\}, \tag{27}$$

for some functions H_i and vectors h_i. Then, the factorization in (25) implies that the joint uncertainty set U_{X_V} equals

$$U_{X_V} = \{x_V \mid H_i(x_i, x_{\text{Pa}_i}) \leq h_i \ \forall\ i \in V \backslash R, \text{ and } H_i(x_i) \leq h_i \ \forall\ i \in R\}. \tag{28}$$

This set remains the same for any canonical ordering of the set V, except for the ordering of variables in the tuple x_V. With a slight abuse of notation, we denote the cross product in (25) as

$$U_{X_V} = U_{X_R} \otimes \prod_{k \in V \backslash R} P_{X_k|\text{Pa}_k}. \tag{29}$$

A Bayesian network, defined over random variables, satisfies many conditional independence properties. In the next section, we show that these independence properties are retained for the Baysian uncertainty network.

A. Conditional Independence Properties

We first define the local Markov property. These are a set of conditional independence properties that are satisfied by the Bayesian network, defined over a collection of random variables. We will show that these independence properties are also valid for the Bayesian uncertainty network.

Definition *8:* We say that the uncertainty variables X_V satisfy *local Markov property* according to a directed acyclic graph $G = (V, E)$ if
(1) Each node $i \in V$ is associated with a unique UV X_i.
(2) For every $i \in V$, we have $X_i \perp\!\!\!\perp \text{NonDes}_i | \text{Pa}_i$.

We now show that the Bayesian uncertainty network satisfies the local Markov property. The detailed proof is given in [19].

Theorem *4:* Let $G = (V, E)$ be a DAG. The following three statements are equivalent:

1) $\mathcal{BN} = (X_V, G)$ is a Bayesian uncertainty network.
2) X_V satisfies all the local Markov properties according to G.

This theorem implies that the conditional independence properties of the Bayesian network also hold for the Bayesian uncertainty network. In the next section, we define a point estimate for Bayesian uncertainty networks, and show its equivalence with the MAP estimate under certain conditions.

V. Point Estimates

In practice, we are generally interested in point estimates. For example, in the robotic estimation problem, we would like to learn the true trajectory of a robot or a map of its surrounding. In the regression problem or classification problem, we would like to estimate the model parameters. In this section, we define point estimate for a Bayesian uncertainty network.

In the Bayesian uncertainty network, we have some uncertainty variables that we observe, and some others which we want to estimate, given the observed variables. Let $\mathcal{BN} = (X_V, G)$ be a Bayesian uncertainty network, where $G = (V, E)$ is a DAG. Let the joint uncertainty set for X_V be given by (28). Let $J \subset V$ denote the set of nodes, which correspond to the observed data. Namely, we have $x_j = y_j$ for all $j \in J$, and that we know y_J. Let $I \subset V$ be the set of nodes, which correspond to the uncertainty variables that are of interest to us, and we would like to estimate. We assume I and J to be disjoint, and that $I \cup J = V$.

From the joint uncertainty set, we can compute the posteriori uncertainty map $P_{X_I|X_J}(x_J)$ by projection; see Definition 2. Evaluating $P_{X_I|X_J}(x_J)$ at the observed data $x_J = y_J$, yields a posteriori uncertainty set for X_I, given $X_J = y_J$. This set is given by

$$P_{X_I|X_J}(y_J) = \Pi_{X_I}\left[U_{X_V} \bigcap \{X_J = y_J\}\right], \tag{30}$$

$$= \left\{ x_I \in D_{X_I} \;\middle|\; \begin{array}{c} H_i(x_i, x_{\text{Pa}_i}) \leq h_i \ \forall\ i \in V \backslash R, \\ H_i(x_i) \leq h_i \ \forall\ i \in R, x_J = y_J \end{array} \right\}. \tag{31}$$

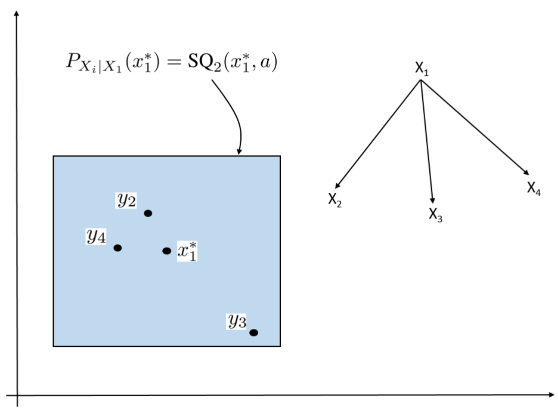

Fig. 3. A Bayesian uncertainty network of four variables $X_{1:4}$. Here, $X_1 = (\mathbb{R}^2, U_{X_1} = \mathbb{R}^2)$, and the conditional uncertainty map $P_{X_i|X_1}(x_1^*) = \mathrm{SQ}_2(x_1^*, a)$ is illustrated. Also, shown is the true value x_1^* of X_1, and the observations $y_{2:4}$ of $X_{2:4}$.

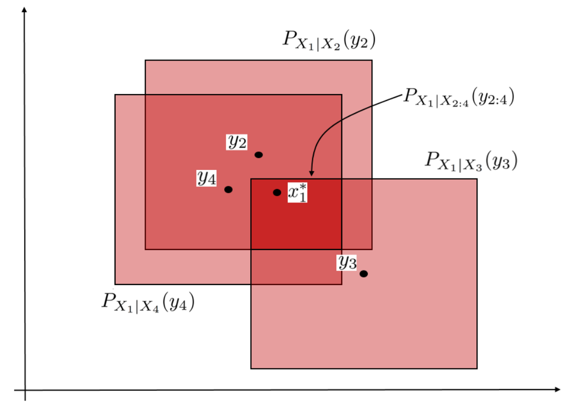

Fig. 4. Illustration of the posteriori invertibility set $P_{X_1|X_{2:4}}(y_{2:4})$ as the intersection of three sets, one for each observation.

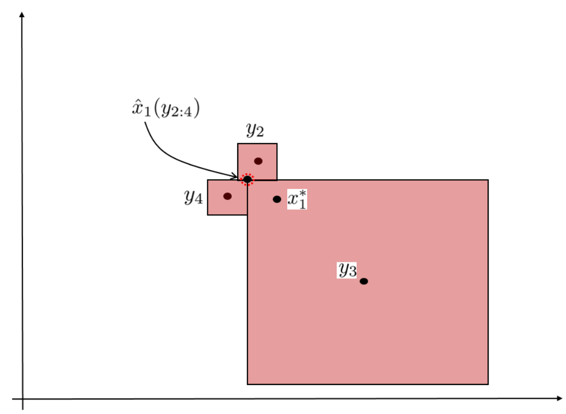

Fig. 5. Shows the point estimate $\hat{x}_1(y_{2:4})$ at the intersection of new, minimally scaled rectangles, obtained by solving (32).

This set gives us a sense of how uncertain we are about the variables of interest, namely X_I. However, it is generally required to come up with a point estimate. We define a point estimate by introducing a *scaling variable* for each constraint in the posteriori set (31). These scaling variable adjust the size of each set, so as to yield an estimate. The point estimate for X_I, given $X_J = y_J$, is defined as

$$\begin{aligned} \hat{x}_I(y_J) = \quad & \underset{x_I,\ \beta_V}{\text{ArgMinimize}} && \sum_{i \in V} \beta_i, \\ & \text{subject to} && H_i(x_i, x_{\mathrm{Pa}_i}) \leq \beta_i h_i \ \forall\ i \in V \backslash R, \\ &&& H_i(x_i) \leq \beta_i h_i \ \forall\ i \in R, \\ &&& x_J = y_J \text{ and } \beta_i \geq 0. \end{aligned} \tag{32}$$

The optimization problem in (32) is over all the variables x_I and the scaling variables β_V. However, as the output of the argminimization, we have only shown a subset of these variables, namely x_I, for notational convenience.

To illustrate the point estimate generated by the optimization problem (32), and the result of scaling variables β_i, we consider a simple example. Consider a Bayesian uncertainty network of four variables $X_{1:4}$ shown in Figure 3. Here, $X_i = (\mathbb{R}^2, U_{X_i})$ for all i. The uncertainty set for X_1 is $U_{X_1} = \mathbb{R}^2$, and the conditional uncertainty maps $P_{X_i|X_1}(x_1) = \mathrm{SQ}_2(x_1, a)$ for all $x_1 \in \mathbb{R}^2$, where $\mathrm{SQ}_2(z, a)$ denotes a square centered at $z \in \mathbb{R}^2$ with side length a. The true value of the uncertainty X_1, namely, x_1^* and the set $P_{X_i|X_1}(x_1^*) = \mathrm{SQ}_2(x_1^*, a)$ is illustrated in Figure 3.

We do not know the true value x_1^* for X_1, and wish to estimate it by observing the variables $X_{2:4}$. Let $y_{2:4}$ be the observations of the uncertainty variables $X_{2:4}$. Using these, we can construct a posteriori uncertainty set for X_1, by evaluating the posteriori uncertainty map $P_{X_1|X_{2:4}}(x_{2:4})$ at $x_{2:4} = y_{2:4}$. This gives the dark-red region shown in Figure 4, which is the posteriori uncertainty set.

To obtain the point estimate we introduce scaling parameters β_is, which scale the size of each of the red-colored rectangles in Figure 4, so that they intersect only at the boundary points. The estimate $\hat{x}_1(y_{2:4})$ is shown in Figure 5. We see that the rectangle corresponding to the one 'far away' observation is enlarged, where as those corresponding to the other observations, that are more closer to one another, are shrunk. This is a process implicit in the definition of the point estimate (32), by which, in computing the point estimate, it weighs more in favor of observations that are closer to one another, than the one that is farther away.

A. Relation with MAP

In this section, we show a relation between the MAP and ML estimate of a Bayesian network, and the point estimate. A Bayesian network $\mathcal{B} = (\bar{X}_V, G)$ is a tuple of a collection of random variables $\bar{X}_V$ and a DAG $G = (V, E)$. For each $i \in V$, is associated a unique random variable $\bar{X}_i$ in $\bar{X}_V$. Further, for each $i \in V$, a conditional probability density[1] function

[1]We will restrict here to the case of continuous distributions for the ease of presentation. However, these results can be extended to discrete valued random variables as well.

$Q_{\bar{X}_i|\bar{X}_{\text{Pa}_i}}(x_i \mid x_{\text{Pa}_i})$ is defined. The joint density function for $\bar{X}_V$ is given by the product factorization

$$Q_{\bar{X}_V}(x_V) = \prod_{i \in V} Q_{\bar{X}_i|\bar{X}_{\text{Pa}_i}}(x_i \mid x_{\text{Pa}_i}). \quad (33)$$

In what follows, we will use Q to denote the probabilities.

For a given Bayesian network $\mathcal{B} = (\bar{X}_V, G)$, defined over the collection of random variables, we construct a *canonical Bayesian uncertainty network* $\mathcal{BN} = (X_V, G)$, such that the underlying DAG is the same, and the functions H_i and h_i in (28) are given by

$$H_i(x_i, x_{\text{Pa}_i}) = -\log\left(Q_{\bar{X}_i|\bar{X}_{\text{Pa}_i}}(x_i \mid x_{\text{Pa}_i})\right), \quad (34)$$

and $h_i = \eta \in \mathbb{R}$, for all $i \in V$. Note that for all $i \in R$, $\text{Pa}_i = \emptyset$, and therefore H_i reduces to a function of just x_i.

The following result shows that the point estimate for the canonical Bayesian uncertainty network, equals the MAP estimate for the corresponding Bayesian network.

Theorem *5:* For the canonical Bayesian uncertainty network $\mathcal{BN} = (X_V, G)$,

$$\hat{x}_I(y_J) = \arg\max_{x_I} Q_{\bar{X}_I|\bar{X}_J}(x_I|y_J), \quad (35)$$

where $Q_{\bar{X}_I|\bar{X}_J}(x_I|y_J)$ denotes the probability density function of $\bar{X}_I$ given $\bar{X}_J$.

Proof: See [19]. ■

This result shows that the point estimate indeed equals the MAP estimate for a canonically defined Bayesian uncertainty network. It is possible to show a similar relation between the point estimate and the maximum-likelihood estimate, by omitting certain constraints in (32). We leave this discussion for our extended work.

VI. Conclusion

We developed a new framework of uncertainty variables to model uncertainty in the real world. We proved Bayes' law and the law of total probability equivalents for uncertainty variables, and showed how this could be used in computing the posteriori uncertainty maps. We defined a notion of independence, conditional independence, and pairwise independence for a given collection of uncertainty variables. We showed that this new notion of independence preserves the properties of independence defined over random variables.

In the second part, we developed a graphical model over a collection of uncertainty variables, namely the Bayesian uncertainty network. This was motivated by the Bayesian network defined over a collection of random variables. A Bayesian network satisfies certain natural conditional independence properties, derived out of the graph structure. We showed that all the natural conditional independence properties, expected out of a Bayesian network, hold also for the Bayesian uncertainty network. We defined a notion of point estimate in a Bayesian uncertainty network, and proved that under a certain representational relation between the Bayesian uncertainty network and a Bayesian network, the point estimate equals the maximum a posteriori estimate.

In a follow up work, we develop other graphical models for uncertainty variables, and show the benefits of its applications in problems such as robotic perception, over some of the traditional approaches.

References

[1] A. Hald, *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713-1935*. Springer-Verlag, New York, 1 ed., 2007.

[2] S. M. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900*. Harvard University Press, Cambridge, 1 ed., 1986.

[3] A. Kolmogorov, *Foundations of the Theory of Probability*. 1933.

[4] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, 2005.

[5] C. Robert, *The Bayesian Choice*. Springer-Verlag, New York, 2 ed., 2007.

[6] C. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag, New York, 1 ed., 2006.

[7] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[8] M. I. Jordan, "Graphical models," *Statistical Science*, vol. 19, pp. 140–155, Jul. 2004.

[9] F. Blanchini and S. Miani, *Set-Theoretic Methods in Control*. Birkhäuser Basel, 2 ed., 2015.

[10] D. Bertsekas, *Control of uncertain systems with a set-membership description of the uncertainity*. Ph.D. Dissertation, MIT, 1971.

[11] D. P. Bertsekas and I. B. Rhodes, "On the minimax reachability of target sets and target tubes," *Automatica*, vol. 7, pp. 233–247, Mar. 1971.

[12] J. R. Deller, "Set membership identification in digital signal processing," *IEEE ASSP Magazine*, vol. 6, pp. 4–20, Oct. 1989.

[13] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE*, vol. 81, pp. 182–208, Feb. 1993.

[14] L. Ros, A. Sabater, and F. Thomas, "An ellipsoidal calculus based on propagation and fusion," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 32, pp. 430–442, Aug. 2002.

[15] A. Kurzhanski and I. Valyi, *Ellipsoidal Calculus for Estimation and Control*. Birkhäuser Basel, 1 ed., 1997.

[16] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Review*, vol. 53, no. 3, pp. 464–501, 2011.

[17] A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton University Press, 2009.

[18] C. Bandi and D. Bertsimas, "Tractable stochastic analysis in high dimensions via robust optimization," *SIAM Review*, vol. 134, pp. 23–70, Aug. 2012.

[19] R. Talak, S. Karaman, and E. Modiano, "A theory of uncertainty variables for state estimation and inference," *arXiv e-prints arXiv:1909.10673*, Sep. 2019.